

HANDBUCH

Nina Baur  
Jörg Blasius Hrsg.

# Handbuch Methoden der empirischen Sozialforschung



Springer VS

---

# Handbuch Methoden der empirischen Sozialforschung

---

Nina Baur • Jörg Blasius (Hrsg.)

# Handbuch Methoden der empirischen Sozialforschung

*Herausgeber*

Nina Baur  
TU Berlin  
Berlin, Deutschland

Jörg Blasius  
Universität Bonn  
Bonn, Deutschland

ISBN 978-3-531-17809-7  
DOI 10.1007/978-3-531-18939-0

ISBN 978-3-531-18939-0 (eBook)

Die Deutsche Nationalbibliothek verzeichnetet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer VS

© Springer Fachmedien Wiesbaden 2014

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Ein-speicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

*Lektorat:* Katrin Emmerich

*Satz:* workformedia, Frankfurt am Main

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer VS ist eine Marke von Springer DE. Springer DE ist Teil der Fachverlagsgruppe Springer Science+Business Media.

[www.springer-vs.de](http://www.springer-vs.de)

---

# Inhaltsverzeichnis

Über die Autorinnen und Autoren .....	13
Danksagung.....	37
1. Methoden der empirischen Sozialforschung.....	41
Ein Überblick	
<i>Nina Baur und Jörg Blasius</i>	
<b>Teil 1</b>	
<b>Grundlagen der empirischen Sozialforschung</b>	
2. Empirische Sozialforschung und soziologische Theorie.....	65
<i>Jo Reichertz</i>	
3. Forschungsethik .....	81
<i>Jürgen Friedrichs</i>	
4. Informationelle Selbstbestimmung .....	93
<i>Andreas Mühlchen</i>	
5. Informationsquellen und Informationsaustausch .....	103
<i>H. Peter Ohly und Karsten Weber</i>	
6. Forschungsdesigns für die qualitative Sozialforschung .....	117
<i>Aglaja Przyborski und Monika Wohlrab-Sahr</i>	
7. Forschungsdesigns für die quantitative Sozialforschung.....	135
<i>Petra Stein</i>	

<b>8. Mixed Methods.....</b>	<b>153</b>
<i>Udo Kelle</i>	
<b>9. Evaluationsforschung.....</b>	<b>167</b>
<i>Nicola Döring</i>	
<b>10. Marktforschung.....</b>	<b>183</b>
<i>Markus Ziegler</i>	
<b>11. Experiment .....</b>	<b>195</b>
<i>Stefanie Eifler</i>	
<b>12. Simulation .....</b>	<b>211</b>
<i>Ben Jann und Debra Hevenstone</i>	
<b>13. Qualitative Daten für die Sekundäranalyse.....</b>	<b>223</b>
<i>Irena Medjedović</i>	
<b>14. Quantitative Daten für die Sekundäranalyse.....</b>	<b>233</b>
<i>Ekkehard Mochmann</i>	
<b>15. Ergebnispräsentation in der qualitativen Forschung.....</b>	<b>245</b>
<i>Christian Meyer und Christian Meier zu Verl</i>	
<b>16. Ergebnispräsentation in der quantitativen Forschung.....</b>	<b>259</b>
<i>Jürgen Friedrichs</i>	

**Teil 2****Stichproben, Datenaufbereitung und Güte**

<b>17. Stichprobenziehung in der qualitativen Sozialforschung .....</b>	<b>265</b>
<i>Leila Akremi</i>	
<b>18. Stichprobenziehung in der quantitativen Sozialforschung.....</b>	<b>283</b>
<i>Michael Häder und Sabine Häder</i>	
<b>19. Pretest .....</b>	<b>299</b>
<i>Martin Weichbold</i>	
<b>20. Einstellungen zu Befragungen .....</b>	<b>305</b>
<i>Anja Hlawatsch und Tino Krickl</i>	
<b>21. Interviewereffekte.....</b>	<b>313</b>
<i>Alexander Glantz und Tobias Michael</i>	
<b>22. Fälschungen von Interviews.....</b>	<b>323</b>
<i>Jörg Blasius</i>	

<b>23. Unit- und Item-Nonresponse .....</b>	<b>331</b>
<i>Uwe Engel und Björn Oliver Schmidt</i>	
<b>24. Gewichtung.....</b>	<b>349</b>
<i>Hans Kiesl</i>	
<b>25. Paradaten.....</b>	<b>357</b>
<i>Barbara Felderer, Alexandra Birg und Frauke Kreuter</i>	
<b>26. Data Fusion, Record Linkage und Data Mining.....</b>	<b>367</b>
<i>Julia Cielebak und Susanne Rässler</i>	
<b>27. Datenaufbereitung und Datenbereinigung in der qualitativen Sozialforschung .....</b>	<b>383</b>
<i>Udo Kuckartz und Stefan Rädiker</i>	
<b>28. Datenaufbereitung und Datenbereinigung in der quantitativen Sozialforschung.....</b>	<b>397</b>
<i>Detlev Lück und Uta Landrock</i>	
<b>29. Gütekriterien qualitativer Sozialforschung .....</b>	<b>411</b>
<i>Uwe Flick</i>	
<b>30. Gütekriterien quantitativer Sozialforschung.....</b>	<b>425</b>
<i>Dagmar Krebs und Natalja Menold</i>	
<b>31. Total Survey Error .....</b>	<b>439</b>
<i>Frank Faulbaum</i>	

**Teil 3****Forschungsparadigmen in der qualitativen Sozialforschung**

<b>32. Grounded Theory und Theoretical Sampling.....</b>	<b>457</b>
<i>Jörg Strübing</i>	
<b>33. Sozialwissenschaftliche Hermeneutik und hermeneutische Wissenssoziologie.....</b>	<b>473</b>
<i>Ronald Kurt und Regine Herbrik</i>	
<b>34. Diskursanalyse.....</b>	<b>493</b>
<i>Boris Traue, Lisa Pfahl und Lena Schürmann</i>	
<b>35. Biographieforschung.....</b>	<b>509</b>
<i>Gabriele Rosenthal</i>	
<b>36. Ethnographie .....</b>	<b>521</b>
<i>Hubert Knoblauch</i>	

<b>37. Einzelfallanalyse.....</b>	<b>529</b>
<i>Linda Hering und Robert J. Schmidt</i>	

<b>38. Qualitative Inhaltsanalyse.....</b>	<b>543</b>
<i>Philipp Mayring und Thomas Fenzl</i>	

#### **Teil 4**

##### **Offene Befragung**

<b>39. Leitfaden- und Experteninterviews .....</b>	<b>559</b>
<i>Cornelia Helfferich</i>	

<b>40. Narratives Interview.....</b>	<b>575</b>
<i>Ivonne Küsters</i>	

<b>41. Gruppendiskussion .....</b>	<b>581</b>
<i>Susanne Vogl</i>	

<b>42. DELPHI-Befragung.....</b>	<b>587</b>
<i>Michael Häder und Sabine Häder</i>	

<b>43. Journalistisches Interview.....</b>	<b>593</b>
<i>Jürgen Friedrichs und Ulrich Schwinges</i>	

#### **Teil 5**

##### **Standardisierte Befragung**

<b>44. Grundlagen der standardisierten Befragung.....</b>	<b>601</b>
<i>Jost Reinecke</i>	

<b>45. Persönlich-mündliche Befragung.....</b>	<b>619</b>
<i>Volker Stocké</i>	

<b>46. Telefonische Befragung .....</b>	<b>631</b>
<i>Volker Hüfken</i>	

<b>47. Schriftlich-postalische Befragung.....</b>	<b>643</b>
<i>Karl-Heinz Reuband</i>	

<b>48. Online-Befragung.....</b>	<b>661</b>
<i>Pia Wagner und Linda Hering</i>	

<b>49. Gesamtgestaltung des Fragebogens .....</b>	<b>675</b>
<i>Jennifer Klöckner und Jürgen Friedrichs</i>	

<b>50. Frageformulierung .....</b>	<b>687</b>
<i>Rolf Porst</i>	

<b>51. Antwortskalen in standardisierten Befragungen .....</b>	<b>701</b>
<i>Axel Franzen</i>	
<b>52. Offene Fragen .....</b>	<b>713</b>
<i>Cornelia Züll und Natalja Menold</i>	
<b>53. Vignetten .....</b>	<b>721</b>
<i>Hermann Dülmer</i>	
<b>54. Soziodemographische Standards .....</b>	<b>733</b>
<i>Jürgen H.P. Hoffmeyer-Zlotnik und Uwe Warner</i>	
<b>55. Skalen und Indizes.....</b>	<b>745</b>
<i>Rossalina Latcheva und Eldad Davidov</i>	
<b>56. Interkulturell vergleichende Umfragen .....</b>	<b>757</b>
<i>Michael Braun</i>	
<b>57. Mitarbeiterbefragungen .....</b>	<b>767</b>
<i>Ingwer Borg</i>	
<b>58. Befragungen von Kindern und Jugendlichen .....</b>	<b>775</b>
<i>Julia Heinen und Susanne König</i>	
<b>59. Befragungen von älteren und alten Menschen .....</b>	<b>781</b>
<i>Andreas Motel-Klingebiel, Daniela Klaus und Julia Simonson</i>	
<b>60. Befragung von Migranten .....</b>	<b>787</b>
<i>Yasemin El-Menouar</i>	
<b>61. Befragung von speziellen Populationen .....</b>	<b>799</b>
<i>Miriam Schütte und Tobias Schmies</i>	

## Teil 6

### Weitere Datentypen

<b>62. Natürliche Daten: Dokumente .....</b>	<b>813</b>
<i>Axel Salheiser</i>	
<b>63. Literarische Quellen und persönliche Dokumente .....</b>	<b>829</b>
<i>Stefanie Ernst</i>	
<b>64. Zeitungsartikel.....</b>	<b>841</b>
<i>Harald Klein</i>	
<b>65. Web Server Logs und Logfiles .....</b>	<b>847</b>
<i>Andreas Schmitz und Olga Yanenko</i>	

<b>66. Beobachtung.....</b>	<b>855</b>
<i>Cornelia Thierbach und Grit Petschick</i>	
<b>67. Unbewegte Bilder: Fotografien und Kunstgegenstände .....</b>	<b>867</b>
<i>Ralf Bohnsack</i>	
<b>68. Videographie .....</b>	<b>875</b>
<i>René Tuma und Bernt Schnettler</i>	
<b>69. Filme .....</b>	<b>887</b>
<i>Leila Akremi</i>	
<b>70. Gebrauchsgegenstände und technische Artefakte.....</b>	<b>899</b>
<i>Cornelius Schubert</i>	
<b>71. Verwaltungsdaten und Daten der amtlichen Statistik.....</b>	<b>907</b>
<i>Peter H. Hartmann und Andrea Lengerer</i>	
<b>72. Aggregatdaten.....</b>	<b>915</b>
<i>Peter Graeff</i>	
<b>73. Paneldaten für die Sozialforschung .....</b>	<b>925</b>
<i>Jürgen Schupp</i>	
<b>74. Netzwerkdaten .....</b>	<b>941</b>
<i>Nina Baur</i>	
<b>75. Organisationsdaten .....</b>	<b>959</b>
<i>Alexia Meyermann, Tobias Gebel und Stefan Liebig</i>	
<b>76. Räumliche Daten .....</b>	<b>973</b>
<i>Jens S. Dangschat</i>	
<b>77. Geodaten.....</b>	<b>981</b>
<i>Tobia Lakes</i>	
<b>78. Neurowissenschaftliche Daten .....</b>	<b>989</b>
<i>Monika Pritzel und Hans J. Markowitsch</i>	

**Teil 7****Datenauswertung in der quantitativen Sozialforschung**

<b>79. Multivariate Datenanalyse .....</b>	<b>997</b>
<i>Jörg Blasius und Nina Baur</i>	
<b>80. Kausalität.....</b>	<b>1017</b>
<i>Steffen Kühnel und André Dingelstedt</i>	

<b>81. Indikatoren .....</b>	<b>1029</b>
<i>Nicole Burzan</i>	
<b>82. Messung von sozialer Ungleichheit.....</b>	<b>1037</b>
<i>Johannes Huinink</i>	
<b>83. Skalierungsverfahren .....</b>	<b>1051</b>
<i>Jörg Blasius</i>	
<b>84. Zeitreihenanalyse .....</b>	<b>1063</b>
<i>Rainer Metz und Helmut Thome</i>	
<b>85. Längsschnittanalyse.....</b>	<b>1077</b>
<i>Tatjana Mika und Michael Stegmann</i>	
<b>86. Verlaufsdatenanalyse.....</b>	<b>1089</b>
<i>Ulrich Pötter und Gerald Prein</i>	
<b>87. Mehrebenenanalyse .....</b>	<b>1101</b>
<i>Manuela Pötschke</i>	
<b>88. Meta-Analyse .....</b>	<b>1117</b>
<i>Michael Wagner und Bernd Weiß</i>	

---

## Über die Autorinnen und Autoren

**Leila Akremi** ist wissenschaftliche Mitarbeiterin am Fachgebiet Methoden der empirischen Sozialforschung an der Technischen Universität Berlin. *Ausgewählte Publikationen:* Kommunikative Konstruktion von Zukunftsängsten. Imaginationen zukünftiger Identitäten im dystopischen Spielfilm. Wiesbaden: VS Verlag (2014); Lebensstile und Geschlecht. In: Rössel, Jörg/Otte, Gunnar (Hg.): Lebensstilforschung. Sonderheft 51 der Kölner Zeitschrift für Soziologie und Sozialpsychologie. Wiesbaden: VS Verlag (gemeinsam mit Nina Baur, 2011); Fans und Konsum. In: Jochen Roose, Mike S. Schäfer und Thomas Schmidt-Lux (Hg.): Fans. Soziologische Perspektiven. Wiesbaden: VS Verlag (zusammen mit Kai-Uwe Hellmann, 2010). *Webseite:* [www.mes.tu-berlin.de](http://www.mes.tu-berlin.de). *Kontaktadresse:* leila.akremi@tu-berlin.de.

**Nina Baur** ist Professorin für Methoden der empirischen Sozialforschung an der Technischen Universität Berlin. *Ausgewählte Publikationen:* Verlaufsmusteranalyse. Entwurf einer prozessorientierten Methodologie. Wiesbaden: Springer VS (2015); Comparing Societies and Cultures. Challenges of Cross-Cultural Survey Research as an Approach to Spatial Analysis, in: HSR (2014); Towards a Process-Oriented Methodology. Modern Social Science Research Methods and Nobert Elias' Figurational Sociology, in: The Sociological Review 59, 777 (zusammen mit Stefanie Ernst, 2011). *Webseite:* [www.mes.tu-berlin.de](http://www.mes.tu-berlin.de). *Kontaktadresse:* nina.baur@tu-berlin.de.

**Alexandra Birg** ist Studentin der Wirtschafts- und Sozialstatistik am Statistik Institut der Ludwig-Maximilians-Universität, München. *Kontaktadresse:* Alexandra.Birg@gmx.de.

**Jörg Blasius** ist Professor für Soziologie mit dem Schwerpunkt Methodenlehre an der Universität Bonn. *Ausgewählte Publikationen:* Visualization and Verbalization of Data. Chapman & Hall, Boca Raton, Fl (Hg. gemeinsam mit Michael Greenacre, 2014); Assessing the Quality of Survey Data. Sage, London (gemeinsam mit Victor Thiessen, 2012); Comparing Ranking Techniques in Web Surveys, in: Field Methods, 24 (2012); Mitherausgeber der Reihe "Research Methods for Social Scientists" (zusammen mit Simona Balbi, Cor van Dijkum, Anne Ryen), Sage (London). *Kontaktadresse:* jblasius@uni-bonn.de.

**Ralf Bohnsack** ist Professor a. D. für qualitative Methoden an der Freien Universität Berlin. *Ausgewählte Publikationen:* Rekonstruktive Sozialforschung. Einführung in qualitative Methoden. Opladen/Berlin/Toronto: Barbara Budrich (9. Auflage 2014); Qualitative Bild- und Videointerpretation. Die dokumentarische Methode. Opladen/Farmington Hills: Barbara Budrich (2. Auflage 2011). *Kontaktadresse:* bohnsack@zedat.fu-berlin.de.

**Ingwer Borg** war wissenschaftlicher Leiter bei GESIS – Leibniz-Institut für Sozialwissenschaften und Professor für Psychologie, Uni Gießen. Er ist jetzt Consultant für Org-Vitality. *Ausgewählte Publikationen:* Applied Multidimensional Scaling. New York: Springer (gemeinsam mit Patrick J. F. Groenen und Patrick Mair, 2013); Employee Surveys in Management. Cambridge, MA: Hogrefe-Huber (gemeinsam mit Paul M. Mastrangelo, 2008); Modern Multidimensional Scaling: Theory and Applications. New York: Springer (2. Auflage 2005). *Kontaktadresse:* ingwer.borg@gmail.com.

**Michael Braun** ist Projektberater bei GESIS – Leibniz-Institut für Sozialwissenschaften in Mannheim und außerplanmäßiger Professor an der Universität Mannheim. *Ausgewählte Publikationen:* Assessing Cross-National Equivalence of Measures of Xenophobia. Evidence from Probing in Web Surveys, in: International Journal of Public Opinion 25, 383 (zusammen mit Dorothée Behr und Lars Kaczmarek, 2003); Patterns of Social Integration of Western European Migrants, in: Journal of International Migration and Integration 13, 403 (zusammen mit Angelika Glöckner-Rist, 2012); Foreign Language Proficiency of Intra-European Migrants. A Multilevel Analysis, in: European Sociological Review 26, 603 (2010). *Webseite:* www.gesis.org. *Kontaktadresse:* michael.braun@gesis.org.

**Nicole Burzan** ist Professorin für Soziologie an der TU Dortmund. *Ausgewählte Publikationen:* Zur Gültigkeit der Individualisierungsthese. Eine kritische Systematisierung empirischer Prüfkriterien, in: ZfS 40, 6 (2011); Zur Debatte um die Verknüpfung qualitativer und quantitativer Sozialforschung, in: Honer, Anne et al. (Hg.): Fragile Sozialität, Wiesbaden: VS (2010); Quantitative Forschung in der Sozialstrukturanalyse. Anwendungsbeispiele aus methodischer Perspektive, Wiesbaden: VS (2007). *Webseite:* http://lehrgebiet-soziologie.fk12.tu-dortmund.de. *Kontaktadresse:* nicole.burzan@fk12.tu-dortmund.de.

**Julia Ciebak** ist wissenschaftliche Mitarbeiterin und Doktorandin am Lehrstuhl für Statistik und Ökonometrie an der Otto-Friedrich-Universität Bamberg. *Ausgewählte Publikationen:* Sozioökonomische Untersuchung der Belastungsgrenzen von Familien in Baden-Württemberg. Diplomarbeit: Otto-Friedrich-Universität Bamberg (2007, als Julia Kobert); Grenzen der Belastbarkeit privater Haushalte mit Schulgeld vor dem Hintergrund des Sonderungsverbotes nach Art. 7 GG. Eine Untersuchung für das Land Baden-Württemberg, München (Mitarbeit bei B. Eisinger, P.K. Warndorf, J. Falterbaum, und J. Feldt, 2007). *Webseite:* www.uni-bamberg.de/stat-oek/. *Kontaktadresse:* julia.ciebak@uni-bamberg.de.

**Jens S. Dangschat** ist Professor für Soziologie und Demographie an der Technischen Universität Wien. *Ausgewählte Publikationen:* Soziale Ungleichheit und der (städtische) Raum, in: Peter A. Berger u.a. (Hg.): Urbane Ungleichheiten (2014); Räumliche Entwicklungen in österreichischen Stadtregionen. Handlungsbedarf und Steuerungsmöglichkeiten, in: ÖROK Schriftenreihe 179. Wien (zusammen mit Alexander Hamedinger u.a., 2009); Lebensstile, Soziale Lagen und Siedlungsstrukturen. Forschungs- und Sitzungsberichte, 230. Hannover: Akademie für Raumforschung und Landesplanung (ARL) (zusammen

mit Alexander Hamedinger, 2007); Raumkonzept zwischen struktureller Produktion und individueller Konstruktion, in: Ethnoscritps 9, 1 (2007). *Webseite:* <http://isra.tuwien.ac.at>. *Kontaktadresse:* jens.dangschat@tuwien.ac.at.

**Eldad Davidov** ist Professor für Soziologie an der Universität Zürich. *Ausgewählte Publikationen:* Cross-Cultural Analysis: Methods and Applications. New York: Routledge (zusammen mit Peter Schmidt und Jaak B. Billiet, 2011); Measurement Equivalence of Nationalism and Constructive Patriotism in the ISSP: 34 Countries in a Comparative Perspective, in: Political Analysis 17 (2009); Testing the Stability of an Acquiescence Style Factor Behind Two Interrelated Substantive Variables in a Panel Design, in: Sociological Methods & Research 36 (zusammen mit Jaak B. Billiet, 2008); Measurement equivalence in cross-national research, in: Annual Review of Sociology 40 (2014) (zusammen mit Bart Meuleman, Jan Cieciuch, Peter Schmidt und Jaak Billiet). *Webseite:* [www.suz.uzh.ch/davidov](http://www.suz.uzh.ch/davidov). *Kontaktadresse:* davidov@soziologie.uzh.ch.

**André Dingelstedt** ist Lehrkraft für besondere Aufgaben im Bereich der quantitativen Methoden der empirischen Sozialforschung am Methodenzentrum Sozialwissenschaften an der Universität Göttingen. *Ausgewählte Publikationen:* James S. Coleman: Foundations of Social Theory, in: Klassiker der Sozialwissenschaften. Wiesbaden: Springer VS (2014); Aufgabensammlung zur „Statistik für die Sozialwissenschaften“. Reinbek bei Hamburg: Rowohlt Verlag (zusammen mit Dagmar Krebs, Steffen M. Kühnel und Anja Mays, 2013). *Kontaktadresse:* andre.dingelstedt@sowi.uni-goettingen.de.

**Nicola Döring** ist Professorin für Medienpsychologie und Medienkonzeption an der Technischen Universität Ilmenau. *Ausgewählte Publikationen:* Forschungsmethoden und Evaluation für die Sozial- und Humanwissenschaften. Heidelberg: Springer (zusammen mit Jürgen Bortz, 5. Auflage 2014); Zur Operationalisierung von Geschlecht im Fragebogen: Probleme und Lösungsansätze aus Sicht von Mess-, Umfrage-, Gender- und Queer-Theorie, in: Gender – Zeitschrift für Geschlecht, Kultur und Gesellschaft 2 (2013); Mobile Market Research/Mobile Marktforschung. Köln: Herbert von Halem (zusammen mit Emanuel Maxl und Astrid Walisch, 2009). *Kontaktadresse:* nicola.doering@tu-ilmenau.de.

**Hermann Dülmer** ist Privatdozent an der Universität zu Köln. *Ausgewählte Publikationen:* Using a Multilevel Structural Equation Modeling Approach to Explain Cross-Cultural Measurement Noninvariance, in: Journal of Cross-Cultural Psychology 43, 4 (zusammen mit Eldad Davidov, Elmar Schläter, Peter Schmidt und Bart Meuleman, 2012); A Multilevel Regression Analysis on Work Ethic, in: Eldad Davidov, Peter Schmidt und Jaak Billiet (Hg.): Cross Cultural Analysis. Methods and Applications. New York: Routledge (2011); Experimental Plans in Factorial Surveys: Random or Quota Design?, in: Sociological Methods & Research 35, 3 (2007). *Kontaktadresse:* HDuelmer@uni-koeln.de.

**Stefanie Eifler** ist Professorin für Soziologie und Empirische Sozialforschung an der Katholischen Universität Eichstätt-Ingolstadt. *Ausgewählte Publikationen:* Kriminalität im Alltag. Eine handlungstheoretische Analyse von Gelegenheiten. Wiesbaden: VS (2009); Wer Wind sät, wird Sturm ernten. Die Transmission von Gewalt im empirischen Theorienvergleich, in: KZfSS 63 (zusammen mit Sonja Schulz und Dirk Baier, 2011); Evaluating the Validity of Self-Reported Deviant Behavior Using Vignette Analyses, in: Quality & Quantity 41 (2007). *Webseite:* [www.ku.de/ggf/soziologie/soziologie2/ueberblick/](http://www.ku.de/ggf/soziologie/soziologie2/ueberblick/). *Kontaktadresse:* stefanie.eifler@ku.de.

**Yasemin El-Menouar** ist Project Manager in der Bertelsmann Stiftung in Gütersloh. *Ausgewählte Publikationen:* The Five Dimensions of Muslim Religiosity. Results of an Empirical Study, in: methods, data, analysis 8, 1 (2014); Islam und Sozialkapital. Beispiele muslimischer Gruppierungen in Deutschland, in: Spenlen, Klaus (Hg.): Gehört der Islam zu Deutschland? Düsseldorf: DUP (2013); Is Islam an Obstacle for Integration? A Qualitative Analysis of German Media Discourse, in: Journal of Religion in Europe 5, 2 (zusammen mit Melanie Becker, 2012); Sozioökonomische Entwicklung und Wertvorstellungen in elf Regionen der Türkei, in: KZfSS 61 (zusammen mit Martin Fritz, 2009). *Webseite:* [www.bertelsmann-stiftung.de](http://www.bertelsmann-stiftung.de). *Kontaktadresse:* Yasemin.El-Menouar@bertelsmann-stiftung.de.

**Uwe Engel** ist Professor für Soziologie mit dem Schwerpunkt Statistik und empirische Sozialforschung und leitet das Sozialwissenschaftliche Methodenzentrum der Universität Bremen. *Ausgewählte Publikationen:* Access Panel and Mixed-Mode Internet Survey. PPSM Panel Report. [http://www.sozialforschung.uni-bremen.de/PPSM\\_Panel\\_Report\\_Online.pdf](http://www.sozialforschung.uni-bremen.de/PPSM_Panel_Report_Online.pdf) (2013); Wissenschaftliche Umfragen. Methoden und Fehlerquellen. Frankfurt/New York: Campus (zusammen mit Simone Bartsch, Christiane Schnabel und Helen Vehre, 2012); Nonresponse und Stichprobenqualität. Frankfurt a.M.: HORIZONT productions (zusammen mit Manuela Pötschke et al., 2004). *Webseite:* [www.sozialforschung.uni-bremen.de](http://www.sozialforschung.uni-bremen.de). *Kontaktadresse:* uengel@empas.uni-bremen.de.

**Stefanie Ernst** ist Professorin für Soziologie mit dem Schwerpunkt Arbeit und Wissen an der Westfälischen Wilhelms-Universität Münster. *Ausgewählte Publikationen:* Towards a Process-Oriented Methodology. Modern Social Science Research Methods and Norbert Elias' Figurational Sociology, in: The Sociological Review 59, 777 (zusammen mit Nina Baur, 2011); Prozessorientierte Methoden in der Arbeits- und Organisationsforschung. Wiesbaden: VS (2010); Using Qualitative Content Analysis of Popular Literature for Uncovering Long-Term Social Processes. The Case of Gender Relations in Germany, in: HSR 34, 1 (2009). *Webseite:* [www.uni-muenster.de/Soziologie/personen/ernst](http://www.uni-muenster.de/Soziologie/personen/ernst). *Kontaktadresse:* stefanie.ernst@uni-muenster.de.

**Frank Faulbaum** war bis 2008 Inhaber des Lehrstuhls für Sozialwissenschaftliche Methoden/Empirische Sozialforschung an der Universität Duisburg-Essen und ist u.a. Vorstandsvorsitzender der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute. *Ausgewählte Publikationen:* Qualitätssicherung in der Umfrageforschung. Wiesbaden: Springer VS (zusammen mit M. Stahl und E. Wiegand, 2012); Gesellschaftliche Entwicklungen im Spiegel der empirischen Sozialforschung. Wiesbaden: Springer VS (zusammen mit C. Wolf, 2010); Was ist eine gute Frage? Wiesbaden: Springer VS (zusammen mit P. Prüfer & M. Rexroth, 2009). *Webseite:* [www.uni-due.de/soziologie/](http://www.uni-due.de/soziologie/). *Kontaktadresse:* frank.faulbaum@uni-due.de.

**Barbara Felderer** ist wissenschaftliche Mitarbeiterin am Institut für Arbeitsmarkt – und Berufsforschung (IAB) in Nürnberg und Stipendiatin des gemeinsamen Graduiertenprograms (GradAB) des IABs und der Universität Nürnberg-Erlangen. *Kontaktadresse:* Barbara.Felderer@iab.de.

**Thomas Fenzl** ist Post-Doc Senior Scientist an der Abteilung für Angewandte Psychologie und Methodenforschung und Projektassistent am Zentrum für Evaluation und

Forschungsberatung ZEF der Alpen-Adria Universität Klagenfurt. *Ausgewählte Publikationen:* Qualitative Content Analysis Programm www.qcamap.org (zusammen mit Philipp Mayring, 2013); Usability gesichtserkennungsbasierter Authentifizierung. Ergebnisse eines Feldtests, in: Datenschutz und Datensicherheit, 07, (zusammen mit Christian Kollmitzer, 2012); Die Massenpsychologie der Finanzmarktkrise. US-Immobilienblase, Subprime Desaster, Schulden-Bubble und ihre Auswirkungen. Wien/New York: Springer (2009). *Webseite:* [www.thomasfenzl.com/Academic.html](http://www.thomasfenzl.com/Academic.html). *Kontaktadresse:* Thomas.Fenzl@aau.at.

**Uwe Flick** ist Professor für Qualitative Sozial- und Bildungsforschung an der Freien Universität Berlin. *Ausgewählte Publikationen:* An Introduction to Qualitative Research. London: Sage (5. Auflage, 2014); The SAGE Handbook of Qualitative Data Analysis. London: Sage (2014); Introducing Research Methodology. A Beginners' Guide to Doing a Research Project. London: Sage (2011); The SAGE Qualitive Research Kit London: Sage (8 Volumes, 2007); Triangulation – Eine Einführung. Wiesbaden: VS Verlag für Sozialwissenschaften (3. Auflage, 2011) *Webseite:* <http://www.ewi-psy.fu-berlin.de/qsb>. *Kontaktadresse:* Uwe.Flick@FU-Berlin.de.

**Axel Franzen** ist Professor für Methoden der empirischen Sozialforschung am Institut für Soziologie der Universität Bern. *Ausgewählte Publikationen:* Two Decades of Measuring Environmental Attitudes: A Comparative Analysis of 33 Countries, in: Global Environmental Change 23: 1001-1008 (zusammen mit Dominikus Vogl, 2013); The External Validity of Giving in the Dictator Game: A Field Experiment Using the Misdirected Letter Technique, in: Experimental Economics 16: 155-169 (gemeinsam mit Sonja Pointner, 2013); Calling Social Capital: An Analysis of the Determinants of Success on the TV Quiz Show "Who Wants to Be a Millionaire?", in: Social Networks 33: 79-87 (zusammen mit Sonja Pointner, 2011). *Webseite:* [www.soz.unibe.ch](http://www.soz.unibe.ch). *Kontaktadresse:* axel.franzen@soz.unibe.ch.

**Jürgen Friedrichs** ist Professor Emeritus am Institut für Soziologie und Sozialpsychologie der Universität zu Köln. *Ausgewählte Publikationen:* Sozialräumliche Kontexteffekte der Armut, in: D. Oberwittler et al. (Hg.): Städtische Armutsquartiere – Kriminelle Lebenswelten? Wiesbaden: Springer (2013); Stadtsoziologie, Opladen: Leske + Budrich (1995); Methoden empirischer Sozialforschung. Opladen: Westdeutscher Verlag (14. Auflage 1990). *Webseite:* [www.iss-wiso.uni-koeln.de](http://www.iss-wiso.uni-koeln.de). *Kontaktadresse:* friedrichs@wiso.uni-koeln.de.

**Tobias Gebel** ist wissenschaftlicher Mitarbeiter am Datenservicezentrum Betriebs- und Organisationsdaten (DSZ-BO) der Universität Bielefeld und im SFB 882 „Von Heterogenitäten zu Ungleichheiten“, Teilprojekt B4 „Betriebe und Ungleichheit: Synchrone und diachrone Ungleichheitseffekte zeitweiser Entlassungen (Recalls)“. *Ausgewählte Publikationen:* Die Dokumentation qualitativer Interviews im Rahmen von Organisationsstudien, in: Denis Huschka et al. (Hg.): Forschungsinfrastruktur für die qualitative Sozialforschung. Berlin: Scivero (zusammen mit Stefan Liebig, 2013); The German Data Service Center for Business and Organizational Data (DSC-BO), in: Schmollers Jahrbuch 132, 4 (zusammen mit Susanne Edler et al. 2012). *Webseite:* [www.uni-bielefeld.de/dsz-bo](http://www.uni-bielefeld.de/dsz-bo). *Kontaktadresse:* tobias.gebel@uni-bielefeld.de.

**Alexander Glantz** ist Projektleiter im Bereich Politik- und Sozialforschung bei Ipsos Public Affairs. *Ausgewählte Publikationen:* Individuelle Unterschiede in der Verarbeitung politischer Informationen, in: Torsten Faas et al. (Hg.), Koalitionen, Kandidaten, Kommunikation. Wiesbaden: Springer (2013); Wahlentscheidungen auf der Spur. Wiesbaden. Springer VS (2012). *Webseite:* [www.ipsos.de](http://www.ipsos.de). *Kontaktadresse:* alexander.glantz@ipsos.com.

**Peter Graeff** ist Professor für Soziologie und empirische Forschungsmethoden an der Christian-Albrechts Universität Kiel. *Ausgewählte Publikationen:* Method Bias in Comparative Research: Problems of Construct Validity as Exemplified by the Measurement of Ethnic Diversity, in: The Journal of Mathematical Sociology, 37, 2 (gemeinsam mit Robert Neumann, 2013); A Multi Trait Multi Method approach to pinpoint the validity of aggregated governance indicators, in: Quality and Quantity, 44 (gemeinsam mit Robert Neumann, 2010). *Kontaktadresse:* pgraeff@soziologie.uni-kiel.de.

**Michael Häder** ist Professor für Methoden der empirischen Sozialforschung an der Technischen Universität Dresden. *Ausgewählte Publikationen:* Telephone Surveys in Europe: Research and Practice. Heidelberg: Springer (zusammen mit S. Häder und M. Kühne, 2012); Empirische Sozialforschung: Eine Einführung. Wiesbaden: VS (2010); Delphi-Befragungen: Ein Arbeitsbuch. Wiesbaden: Springer VS (3. Auflage 2014). *Kontaktadresse:* michael.haeder@tu-dresden.de.

**Sabine Häder** ist Senior Statistician bei GESIS, Leibniz-Institut für Sozialwissenschaften. *Ausgewählte Publikationen:* Telephone Surveys in Europe: Research and Practice. Heidelberg: Springer (zusammen mit M. Häder und M. Kühne, 2012); Telefonbefragungen über das Mobilfunknetz: Konzept, Design und Umsetzung einer Strategie zur Datenerhebung. Wiesbaden: VS (zusammen mit M. Häder, 2009); Die Kombination von Mobilfunk- und Festnetztichproben in Deutschland, in: Weichbold, Martin et al. (Hg.): Umfrageforschung. Wiesbaden: VS (zusammen mit S. Gabler, 2009). *Kontaktadresse:* sabine.haeder@gesis.org.

**Peter H. Hartmann** ist Professor für Soziologie an der Heinrich-Heine-Universität Düsseldorf. *Ausgewählte Publikationen:* Methodische und Methodologische Probleme der Lebensstilforschung, in: Lebensstilforschung. Sonderheft 51 der KZfSS. Wiesbaden: VS Verlag (2011); Der Mikrozensus als Datenquelle für die Sozialwissenschaften, in: ZUMA-Nachrichten 24 (1989). *Webseite:* [www.phil-fak.uni-duesseldorf.de/soz-hartmann](http://www.phil-fak.uni-duesseldorf.de/soz-hartmann). *Kontaktadresse:* peter.hartmann@phil-fak.uni-duesseldorf.de.

**Julia Heinen** ist Referentin im Statistischen Bundesamt. *Ausgewählte Publikationen:* Internetkinder. Eine Untersuchung der Lebensstile junger Nutzergruppen. Opladen, Berlin & Toronto: Budrich UniPress Ltd. (2012). *Kontaktadresse:* jheinen2014@gmx.de.

**Cornelia Helfferich** ist Professorin für Soziologie an der Ev. Hochschule Freiburg, Privatdozentin am Institut für Soziologie der Universität Freiburg und Leiterin des Sozialwissenschaftlichen FrauenForschungsInstituts SoFFI F., Freiburg. *Ausgewählte Publikationen:* Mental Maps und Narrative Raumkarten – Erhebung und Auswertung, in: Bischoff, Christine et al. (Hg.): Methoden der Kulturanthropologie. Stuttgart: UTB (2013); Messen, Rekonstruieren, Verstehen. Begründungen und Reichweite von Forschungsmethoden, in: Heuser, Andreas/Hoffmann, Claudia (Hg.): Erfassen, Deuten, Urteilen. Empirische Zugänge zur Religionsforschung (2013); Die Qualität qualitativer Daten. Manual für die

Durchführung qualitativer Interviews. Wiesbaden: Springer VS (4. Auflage, 2011). *Webseite:* [www.eh-freiburg.de/hochschule/personenverzeichnis/prof-dr-cornelia--helfferich/11](http://www.eh-freiburg.de/hochschule/personenverzeichnis/prof-dr-cornelia--helfferich/11). *Kontaktadresse:* helfferich@eh-freiburg.de.

**Regine Herbrik** ist Juniorprofessorin für qualitative und kulturwissenschaftliche Methoden an der Leuphana Universität Lüneburg. *Ausgewählte Publikationen:* „Du sollst Dir (k)ein Bildnis machen“ – Zur Verwendung von Video- und Bildmaterial und ihrer Bedeutung für die emotionalen Stile christlicher Gemeinden heute, in Petra Lucht et al. (Hg.): Visuelles Wissen und Bilder des Sozialen. Wiesbaden: VS Verlag (2013); Die kommunikative Konstruktion imaginärer Welten. Wiesbaden: VS Verlag (2011). *Webseite:* [www.leuphana.de/regine-herbrik.html](http://www.leuphana.de/regine-herbrik.html). *Kontaktadresse:* regine.herbrik@leuphana.de.

**Linda Hering** ist wissenschaftliche Mitarbeiterin am Fachgebiet Methoden der empirischen Sozialforschung an der Technischen Universität Berlin. *Ausgewählte Publikationen:* Spatial Methods. Special Issue der Zeitschrift „Historical Social Research“ (HSR) (Hg. zusammen mit Cornelia Thierbach, Anna Laura Raschke und Nina Baur, 2014); Das Buch „Städtische Wirtschaftspfade. Auf der Spur der Eigenlogik der Städte.“ erscheint voraussichtlich 2015; Die Rationalität lokaler Wirtschaftspraktiken im Friseurwesen. Der Beitrag der „Ökonomie der Konventionen“ zur Erklärung räumlicher Unterschiede wirtschaftlichen Handelns, in: Bögenhold, Dieter (Hg.): Soziologie des Wirtschaftlichen (zusammen mit Nina Baur, Martina Löw, Anna Laura Raschke und Florian Stoll, 2014). *Webseite:* [www.mes.tu-berlin.de](http://mes.tu-berlin.de). *Kontaktadresse:* linda.hering@tu-berlin.de.

**Debra Hevenstone** ist Oberassistentin am Institut für Soziologie der Universität Bern sowie am Soziologischen Institut der Universität Zürich. *Ausgewählte Publikationen:* Flex-icity, Happiness and Satisfaction, in: International Journal of Sociology 41, 3 (2011); National Context and Atypical Employment, in: International Sociology 25, 3 (2010); Employment Intermediaries: A Model of Firm Incentives, in: The Journal of Mathematical Sociology 33, 1 (2009). *Webseite:* <http://hevenstone.wordpress.com>. *Kontaktadresse:* hevenstone@soz.unibe.ch.

**Anja Hlawatsch** ist wissenschaftliche Mitarbeiterin im Forschungsdatenzentrum des Amtes für Statistik Berlin-Brandenburg. *Ausgewählte Publikationen:* The Shift from Teaching to Learning. Eine überfällige Anpassung der Evaluationskonzepte des Hochschulcontrollings, n: die hochschule 1/2011 (zusammen mit Cornelia Raue, 2011). *Kontaktadresse:* anja.hlawatsch@web.de.

**Jürgen H.P. Hoffmeyer-Zlotnik** ist außerordentlicher Professor am Institut für Politikwissenschaft der Justus Liebig-Universität Gießen und Leiter der gemeinsamen Arbeitsgruppe „Demographische Standards“ im Auftrag der ASI, des ADM und des Statistischen Bu Ausgewählte Publikationen: Soziodemographische Standards für Umfragen in Europa. München und Mering: Rainer Hampp Verlag (zusammen mit Uwe Warner, 2013); Harmonisierung demographischer und sozio-ökonomischer Variablen. Instrumente für die international vergleichende Sozialforschung. Wiesbaden: Springer VS (zusammen mit Uwe Warner, 2012); Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables. New York: Kluwer Academic/Plenum Publishers (zusammen mit Christof Wolf, 2003). *Kontaktadresse:* juergen.hoffmeyer-zlotnik@sowi.uni-giessen.de.

**Volker Hüfken** ist wissenschaftlicher Mitarbeiter an der Heinrich-Heine-Universität Düsseldorf. *Ausgewählte Publikationen:* Supplementary Questionnaire and Nonresponse. Results from the German ISSP Survey, in: Quality and Quantity, 44 (2010); Public Opinion Polls and Estimates in some European Countries. Non-Coverage Bias in Telephone Surveys, in JSM Proceedings, Statistical Computing Section. Alexandria, VA: American Statistical Association. <http://www.amstat.org/sections/srms/Proceedings/y2010/Files/400092.pdf> (2010). *Webseite:* [www.phil-fak.uni-duesseldorf.de/sozwiss/](http://www.phil-fak.uni-duesseldorf.de/sozwiss/). *Kontaktadresse:* volker.huefken@phil.hhu.de.

**Johannes Huinink** ist Professor für Soziologie mit Schwerpunkt „Theorie und Empirie der Sozialstruktur“ am Institut für empirische und angewandte Soziologie (EMPAS) der Universität Bremen. *Ausgewählte Publikationen:* Panel Analysis of Intimate Relationships and Family Dynamics (pairfam). Conceptual Framework and Design, in: Zeitschrift für Familienforschung 23, 77 (zusammen mit Josef Brüderl, Bernhard Nauck, Sabine Walper, Laura Castiglioni und Michael Feldhaus, 2011); Family Research from the Life Course Perspective, in: International Sociology 24, 299 (zusammen mit Michael Feldhaus, 2009); Sozialstruktur Deutschlands. Konstanz: UVK (zusammen mit Torsten Schröder, 2008). *Webseite:* [www.tess.uni-bremen.de](http://www.tess.uni-bremen.de). *Kontaktadresse:* huinink@empas.uni-bremen.de.

**Ben Jann** ist Professor für Soziologie, insbesondere Sozialstrukturanalyse, an der Universität Bern. *Ausgewählte Publikationen:* Reputation Formation and the Evolution of Cooperation in Anonymous Online Markets, in: ASR (gemeinsam mit Andreas Diekmann, Wojtek Przepiorka und Stefan Wehrli, 2014); Asking Sensitive Questions Using the Crosswise Model. An Experimental Survey Measuring Plagiarism, in: Public Opinion Quarterly 71 (2012); Erwerbsarbeit, Einkommen und Geschlecht. Studien zum Schweizer Arbeitsmarkt. Wiesbaden: VS-Verlag (2008). *Webseite:* [www.soz.unibe.ch](http://www.soz.unibe.ch). *Kontaktadresse:* ben.jann@soz.unibe.ch.

**Udo Kelle** ist Professor für Methoden der empirischen Sozialforschung und Statistik an der Helmut-Schmidt-Universität der Bundeswehr in Hamburg. *Ausgewählte Publikationen:* Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung: Theoretische Grundlagen und methodologische Konzepte. Wiesbaden: VS (2008); Handlung, Ordnung und Kultur und das Mehrebenenproblem der empirischen Sozialwissenschaften, in: Gert Albert/Steffen Sigmund (Hg.): Soziologische Theorie kontrovers. Sonderheft 50 der KZfSS (zusammen mit Alex Nolte, 2010); Kontingente Strukturen. Theoretische Grundlagen der Integration qualitativer und quantitativer Methoden, in: Zeitschrift für Betriebswirtschaft. Special Issue: Mixed Methods (2010). *Webseite:* [www.hsu-hh.de/methoden](http://www.hsu-hh.de/methoden). *Kontaktadresse:* kelle@hsu-hh.de.

**Hans Kiesl** ist Professor für Mathematik mit Schwerpunkt Statistik an der Ostbayerischen Technischen Hochschule Regensburg. *Ausgewählte Publikationen:* Sampling Designs of the National Educational Panel Study: Challenges and Solutions, H.-P. Blossfeld et al. (Hg.): Education as a Lifelong Process. Sonderheft 14 der Zeitschrift für Erziehungswissenschaft (zusammen mit C. Alßmann et al., 2011); Selecting Kindergarten Children by Three Stage Indirect Sampling, in: Proceedings of the Joint Statistical Meetings 2010, American Statistical Association (2010); The German Job Vacancy Survey: Estimation of sampling

errors, in: 1st and 2nd International Workshops on Methodologies for Job Vacancy Statistics – Proceedings. Luxembourg: Eurostat Methodologies and Working Papers (2010). *Webseite:* [www.oth-regensburg.de/fakultaeten/informatik-und-mathematik.html](http://www.oth-regensburg.de/fakultaeten/informatik-und-mathematik.html). *Kontaktadresse:* hans.kiesl@oth-regensburg.de.

**Daniela Klaus** ist wissenschaftliche Mitarbeiterin am Deutschen Zentrum für Altersfragen (DZA), Berlin. *Ausgewählte Publikationen:* Relationships to Stepfathers and Biological Fathers in Adulthood: Complementary, Substitutional, or Neglected?, in: Advances in Life Course Research 17, 3 (zusammen mit Bernhard Nauck und Anja Steinbach, 2012); Kinderkosten und Familiengründung: Erste Befunde einer Prüfung der Neuen Haushaltsökonomie unter Verwendung von Paardaten, in: Zeitschrift für Familienforschung 22, 1 (2010); The Value of Children in Palestine and Turkey – Differences and its Consequences for Fertility, in: Current Sociology 55, 4 (zusammen mit Jana Suckow und Bernhard Nauck, 2007). *Webseite:* [www.dza.de/dza/mitarbeiterinnen/klaus.html](http://www.dza.de/dza/mitarbeiterinnen/klaus.html). *Kontaktadresse:* daniela.klaus@dza.de.

**Harald Klein** ist Gründer und Inhaber der Firma Social Science Consulting und war an verschiedenen Universitäten tätig. *Ausgewählte Publikationen:* Computerunterstützte Textanalysen mit TextQuest. Eine Einführung in Methoden und Arbeitstechniken. München/Mering: Hampp (2013); Correspondence Analysis of Text Data with INTEXT/PC, in: Greenacre, Michael/Blasius, Jörg (Hg.): Correspondence Analysis in the Social Sciences. London (zusammen mit Zusammen mit Helmut Giegler (1994). *Kontaktadresse:* hklein@textquest.de.

**Jennifer Klöckner** ist wissenschaftliche Mitarbeiterin am und stellvertretende Geschäftsführerin des Instituts für Soziologie und Sozialpsychologie (ISS) der Universität zu Köln. *Ausgewählte Publikationen:* Partizipation, in: Wagner, Michael/Motel-Klingebiel, Andreas (Hg.): Die Lage älterer Menschen in Nordrhein-Westfalen. Bericht für das Ministerium für Gesundheit, Emanzipation, Pflege und Alter (MGEPA) des Landes Nordrhein-Westfalen (zusammen mit Julia Simonson, 2013); Turkish Islamic Organizations in Europe. A Comparative Study, in: Clooke, Paul et al. (Hg.): Faith, Welfare and Exclusion in European Cities: The FBO phenomenon. Policy Press (zusammen mit Jürgen Friedrichs, Mustafa Şen und Nynke De Witte, 2012). *Webseite:* [www.iss-wiso.uni-koeln.de](http://www.iss-wiso.uni-koeln.de). *Kontaktadresse:* jennifer.kloeckner@wiso.uni-koeln.de.

**Hubert Knoblauch**, 1997 Christa-Hoffmann-Riem-Preisträger für Qualitative Sozialforschung, ist Professor für Allgemeine Soziologie an der Technischen Universität Berlin mit den Arbeitsschwerpunkten Wissenssoziologie, Religionssoziologie, Kommunikation, Thanatosoziologie. *Ausgewählte Publikationen:* Videography. New York: Lang (gemeinsam mit Bernt Schnettler und Rene Tuma, 2014); Qualitative Methods in Europe: The Variety of Social Research. Sonderheft des Forum Qualitative Sozialforschung/ Forum Qualitative Social Research 6, 3 (Hg. gemeinsam mit Uwe Flick und Christoph Maeder, 2005); Qualitative Religionsforschung. Paderborn: Schöningh (2003). *Kontaktadresse:* hubert.knoblauch@tu-berlin.de.

**Susanne König** arbeitet als Marktforscherin in der freien Wirtschaft. *Ausgewählte Publikationen:* Online-Forschung mit Kindern. Opladen, Berlin & Farmington Hills: Budrich UniPress Ltd. (2012). *Kontaktadresse:* susanne\_koenig@yahoo.de.

**Dagmar Krebs** war bis 2008 Professorin für Methoden der Empirischen Sozialforschung am Institut für Soziologie der Justus-Liebig-Universität Gießen. *Ausgewählte Publikationen:* Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen. Reinbek: rowohlt (zusammen mit Steffen Kühnel, 6. Auflage 2012); The Impact of Response Format on Response Behavior, in: S. Salzborn et al. (Hg.): Methods, Theories and Empirical Applications in the Social Sciences. Wiesbaden: VS Verlag (2012); Positive First or Negative First? Effects of the Order of Answering Categories on Response Behavior, in: European Journal of Research Methods for the Behavioral and Social Sciences 6, 3 (zusammen mit Jürgen Hoffmeyer-Zlotnik, 2010). *Kontaktadresse:* dagmar.krebs@sowi.uni-giessen.de.

**Frauke Kreuter** ist Professorin für Survey Methodology an der University of Maryland, College Park und Forschungsbereichsleiterin des Kompetenzzentrum Empirische Methoden am IAB. *Ausgewählte Publikationen:* Improving Surveys with Paradata. New York: Wiley (2013); Using Paradata to Explore Item-Level Response Times in Surveys, in: Journal of the Royal Statistical Society (Series A), 176, 1 (zusammen mit M. Couper, 2013); Evaluating the Measurement Error of Interviewer Observed Paradata, in: Public Opinion Quarterly, 77 (zusammen mit J. Sinibaldi und G. Durrant, 2013). *Kontaktadresse:* frauke.kreuter@stat.uni-muenchen.de.

**Tino Krickl** ist Mitarbeiter im Bereich „Statistische Analysen“ bei der Deutschen Rentenversicherung Bund. *Ausgewählte Publikationen:* Vorgezogene Altersrenten – wer geht früher? In: RVaktuell 11/2013 (zusammen mit Wolfgang Keck, 2013); Rentenzugang 2012: Die Anhebung der Regelaltersgrenze hat begonnen, in: RVaktuell 9/2013 (zusammen mit Jürgen Hofmann, 2013); Rentenzugang 2011: Regelaltersrente auf dem Vormarsch – Aktuelle Entwicklungen im Kohortenvergleich, in: RVaktuell 9/2012 (zusammen mit Jürgen Hofmann, 2012). *Kontaktadresse:* tinokrickl@hotmail.com.

**Udo Kuckartz** ist Professor für empirische Erziehungswissenschaft und Methoden der Sozialforschung an der Philipps-Universität Marburg und leitet die Marburger Arbeitsgruppe für Methoden und Evaluation (MAGMA). *Ausgewählte Publikationen:* Mixed Methods. Methodologie, Forschungsdesigns und Analyseverfahren. Wiesbaden: Springer VS (2014); Statistik - Eine verständliche Einführung. Wiesbaden: Springer VS (2. Auflage, 2013); Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Weinheim: Beltz Juventa (2. Auflage, 2014); Einführung in die computergestützte Analyse qualitativer Daten. Wiesbaden: VS Verlag (3. Auflage 2010). *Webseite:* www.methoden-evaluation.de. *Kontaktadresse:* kuckartz@uni-marburg.de.

**Steffen M. Kühnel** ist Professor für quantitative Methoden der empirischen Sozialforschung am Methodenzentrum Sozialwissenschaften an der Universität Göttingen. *Ausgewählte Publikationen:* Die gegenseitigen Wahrnehmungen und Einstellungen von Einheimischen und MigrantInnen. Ergebnisse aus der Umfrageforschung im Zeitverlauf bis 2011, in: Dabeisein und Dazugehören. Wiesbaden: Springer VS (zusammen mit Jürgen Leibold und Anja Mays, 2013); Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen. Reinbek bei Hamburg, Rowohlt Verlag (zusammen mit Dagmar Krebs, 2012); Analyse von Tabellen und kategorialen Daten. Log-lineare Modelle, latente

Klassenanalyse, logistische Regression und GSK-Ansatz. Berlin/Heidelberg/New York: Springer Verlag (zusammen mit Hans-Jürgen Andreß und Jacques A. Hagenaars, 1997). *Kontaktadresse:* steffen.kuehnel@sowi.uni-goettingen.de.

**Ronald Kurt** ist Professor für Soziologie an der Evangelischen Fachhochschule Rheinland-Westfalen-Lippe in Bochum. *Ausgewählte Publikationen:* Indien und Europa. Ein kultur- und musiksoziologischer Verstehensversuch. Inkl. einer DVD mit drei Dokumentarfilmen. Bielefeld: transcript (2009); Hermeneutik. Eine sozialwissenschaftliche Einführung. Konstanz: UTB/UVK (2004); Menschenbild und Methode der Sozialphänomenologie. Konstanz: UVK (2002). *Kontaktadresse:* kurt@efh-bochum.de.

**Margarethe Kusenbach** ist Professorin (Associate Professor) für Soziologie und Leiterin des Doktorandenprogramms an der University of South Florida, Tampa, USA. *Ausgewählte Publikationen:* Home: International Perspectives on Culture, Identity, and Belonging. Frankfurt a.M.: Peter Lang (Hg. mit Krista Paulsen, 2003); Mobile Methods, in: Sara Delamont (Hrgs.): Handbook of Qualitative Research in Education. Cheltenham, UK: Elgar (2012); Street Phenomenology: The Go-Along as Ethnographic Research Tool, in: Ethnography 4, 3 (2003). *Webseite:* <http://sociology.usf.edu>. *Kontaktadresse:* mkusenba@usf.edu.

**Ivonne Küsters** ist freie Mitarbeiterin bei der triple innova GmbH, Wuppertal. *Ausgewählte Publikationen:* Die Produktion von Autonomie durch Subjektivierung des Heteronomen. Organisation und Management autonomer Kunst, in: Franzen, Martina u.a. (Hg.): Autonomie revisited. 2. Sonderband der Zeitschrift für Theoretische Soziologie. Weinheim: Beltz Juventa 2014; Die Mitte der Gesellschaft. Sicherer als erwartet? Weinheim: Beltz Juventa 2014 (zusammen mit Nicole Burzan und Silke Kohrs). *Webseite:* [www.triple-innova.de](http://www.triple-innova.de). *Kontaktadresse:* ivonne.kuesters@gmail.com.

**Tobia Lakes** ist Juniorprofessorin für Angewandte Geoinformatik an der Humboldt-Universität zu Berlin. *Ausgewählte Publikationen:* Development of an Environmental Justice Index to Determine Socio-Economic Disparities of Noise Pollution and Green Space in Residential Areas in Berlin, in: Journal for Environmental Planning and Management (zusammen mit M. Brückner und A. Krämer, 2013); The Urban Environmental Indicator “Biotope Area Ratio”. An Enhanced Approach to Assess and Manage the Urban Ecosystem Services Using High Resolution Remote-Sensing, in: Ecological Indicator (zusammen mit H.O. Kim, 2012); Geographically Referenced Data in Social Science. Berlin: DIW (zusammen mit P. Hintze, 2009). *Webseite:* [www.geographie.hu-berlin.de](http://www.geographie.hu-berlin.de). *Kontaktadresse:* tobia.lakes@geo.hu-berlin.de.

**Uta Landrock** ist wissenschaftliche Mitarbeiterin in der Abteilung „Survey Design and Methodology“ der GESIS, Leibniz-Institut für Sozialwissenschaften, in Mannheim. *Ausgewählte Publikationen:* Steuerung durch LOM? Eine Analyse zur leistungsorientierten Mittelvergabe an medizinischen Fakultäten in Deutschland, in: Wilkesmann, Uwe/Christian Schmidt (Hg.): Hochschule als Organisation, Münster: VS (zusammen mit René Krempkow und Patricia Schulz, 2012); Matthäus-Effekte oder Governance-Effekte? Eine Analyse zur leistungsorientierten Mittelvergabe an den Medizinischen Fakultäten Deutschlands, in: Forschung 3/2011 (zusammen mit René Krempkow, 2011). *Webseite:* [www.gesis.org/das-institut/mitarbeiterverzeichnis](http://www.gesis.org/das-institut/mitarbeiterverzeichnis). *Kontaktadresse:* Uta.Landrock@gesis.org.

**Rossalina Latcheva** ist Programme Manager – Statistics and Surveys an der Agentur der Europäischen Union für Grundrechte, Abteilung Gleichheit und Bürgerrechte. *Ausgewählte Publikationen:* Measuring National Identity, in: Alex Michalos (Hrgs.): Encyclopedia of Quality of Life Research. New York: Springer (2013); Cognitive Interviewing and Factor-Analytic Techniques: a Mixed Method Approach to Validity of Survey Items Measuring National Identity, in: Quality & Quantity, 45, 6 (2011/2009); Integration trajectories – a mixed method approach, in: M. Wingens et al. (Hrgs.): A life course perspective on migration and integration. Houten: Springer (zusammen mit Barbara Herzog-Punzenberger, 2011) *Webseite:* <http://fra.europa.eu/en>. *Kontaktadresse:* Rossalina.LATCHEVA@fra.europa.eu.

**Andrea Lengerer** ist wissenschaftliche Mitarbeiterin am GESIS – Leibniz-Institut für Sozialwissenschaften in Mannheim. *Ausgewählte Publikationen:* Die soziale Selektivität des partnerschaftlichen Zusammenlebens im Wandel. Eine kohortenbezogene Analyse kumulierter Mikrozensen, in: KZfSS 64, 2 (2012); The GESIS Microcensus-Trendfile. A New Database for the Study of Social Change, in: Schmollers Jahrbuch 132 (zusammen mit Julia H. Schroedter, Mara Boehle und Christof Wolf, 2012); Partnerlosigkeit in Deutschland. Entwicklung und soziale Unterschiede. Wiesbaden: VS (2011). *Webseite:* [www.gesis.org/gml](http://www.gesis.org/gml). *Kontaktadresse:* andrea.lengerer@gesis.org.

**Stefan Liebig** ist Professor für Soziale Ungleichheit und Sozialstrukturanalyse an der Universität Bielefeld. *Ausgewählte Publikationen:* Replicability and Comprehensibility of Social Research and its Technical Implementation. RatSWD Working Paper 219. Berlin: SCIVERO Press (zusammen mit Friedhoff Stefan et al, 2013); Interdisciplinary Longitudinal Surveys. Linking Individual Data to Organizational Data in Life-Course Analysis, in: RatSWD German Data Forum (Hg.): Building on Progress 2. Opladen & Farmington Hills: Budrich UniPress Ltd. (2010); Arbeitsorganisationen als Kontexte der Einstellungsbildung, in Hans J. Hummell (Hg.): Die Analyse von Gesellschaften, Organisationen und Individuen in ihrem Zusammenhang. Bonn: GESIS (zusammen mit Alexandra Krause 2007). *Webseite:* [www.uni-bielefeld.de/dsz-bo](http://www.uni-bielefeld.de/dsz-bo). *Kontaktadresse:* stefan.liebig@uni-bielefeld.de.

**Detlev Lück** ist Mitarbeiter am Bundesinstitut für Bevölkerungsforschung (BiB) in Wiesbaden und Koordinator des Forschungsschwerpunktes „Differentielle Fertilitätsanalysen und Wandel der Familie“. *Ausgewählte Publikationen:* Mängel im Datensatz beseitigen, in: Akremi, Leila et al. (Hg.): Datenanalyse mit SPSS für Fortgeschrittene 1. Wiesbaden: VS (2011); Walking the Tightrope. Combining Family Life, Career and Job Mobility, in: Norbert F. Schneider/Detlev Lück (Hg.): Mobility and Family. Schwerpunktheft der ZfF 22, 2 (2010); Der zögernde Abschied vom Patriarchat. Der Wandel von Geschlechterrollen im internationalen Vergleich. Berlin: edition sigma (2009). *Webseite:* [www.bib-demografie.de/lueck](http://www.bib-demografie.de/lueck). *Kontaktadresse:* detlev.lueck@bib.bund.de.

**Hans J. Markowitsch** ist Professor für Physiologische Psychologie an der Universität Bielefeld. *Ausgewählte Publikationen:* The Development of Autobiographical Memory. Hove: Psychology Press (zusammen mit H. Welzer, 2010); Das Gedächtnis: Entwicklung – Funktionen – Störungen. München: C.H. Beck (2009); Tatort Gehirn. Auf der Suche nach

dem Ursprung des Verbrechens. München: Piper Verlag (zusammen mit W. Siefer, 2009). *Webseite:* [www.uni-bielefeld.de/psychologie/ae/AE14/HOMEPAGE/Markowitsch\\_home.html](http://www.uni-bielefeld.de/psychologie/ae/AE14/HOMEPAGE/Markowitsch_home.html). *Kontaktadresse:* hjmarkowitsch@uni-bielefeld.de.

**Philipp Mayring** ist Professor für Psychologische Methodenlehre, Leiter der Abteilung Angewandte Psychologie und Methodenforschung und Leiter des Zentrums für Evaluation und Forschungsberatung ZEF der Alpen-Adria Universität Klagenfurt. *Ausgewählte Publikationen:* Qualitative Content Analysis. Theoretical Foundation and Basic Procedures (2013); Qualitative Content Analysis Programm. [www.qcamap.org](http://www.qcamap.org) (zusammen mit Thomas Fenzl, 2013); Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim: Beltz (11. Auflage 2010). *Webseite:* [www.uni-klu.ac.at/pmayring/](http://www.uni-klu.ac.at/pmayring/). *Kontaktadresse:* Philipp.Mayring@aau.at.

**Irena Medjedović** ist Leiterin der Forschungseinheit „Qualifikationsforschung und Kompetenzerwerb“ des Instituts Arbeit und Wirtschaft an der Universität Bremen. *Ausgewählte Publikationen:* Qualitative Sekundäranalyse. Zum Potenzial einer neuen Forschungsstrategie in der empirischen Sozialforschung. Wiesbaden: Springer VS (2014); Wiederverwendung qualitativer Daten. Archivierung und Sekundärnutzung qualitativer Interviewtranskripte. Wiesbaden: VS (zusammen mit Andreas Witzel, 2010); Sekundäranalyse, in: Mey/Mruck (Hg.): Handbuch qualitative Forschung in der Psychologie. Wiesbaden: VS (2010); Sekundäranalyse qualitativer Interviewdaten. Problemkreise und offene Fragen einer neuen Forschungsstrategie, in: HSR 33, 3 (2008). *Webseite:* [www.iaw.uni-bremen.de](http://www.iaw.uni-bremen.de). *Kontaktadresse:* imedjedovic@uni-bremen.de.

**Christian Meier zu Verl** ist wissenschaftlicher Mitarbeiter am DFG-Graduiertenkolleg „Locating Media“ der Universität Siegen. *Ausgewählte Publikationen:* Hermeneutische Praxis. Eine ethnomethodologische Rekonstruktion sozialwissenschaftlichen Sinnrekonstruierens, in: sozialersinn (zusammen mit C. Meyer, 2013). *Webseite:* [www.uni-siegen.de/locatingmedia/personen/christian\\_meier\\_zu\\_verl.html](http://www.uni-siegen.de/locatingmedia/personen/christian_meier_zu_verl.html). *Kontaktadresse:* meier\_zu\_verl@locatingmedia.uni-siegen.de.

**Natalja Menold** ist Projektberaterin und Teamleiterin bei GESIS, Leibniz-Institut für Sozialwissenschaften. *Ausgewählte Publikationen:* Survey Standardization and Interviewers' Deviations. Impact, Reasons, Detection and Prevention. Peter Lang: Frankfurt (zusammen mit P. Winker und Rolf Porst, 2013). How Do Real and Falsified Data Differ? Psychology of Survey Response as a Source of Falsification Indicators in Face-to-Face Surveys, in: International Journal for Public Opinion Research (zusammen mit C. J. Kemper, 2013); How Do Respondents Attend to Verbal Labels in Rating Scales?, in: Field Methods (zusammen mit L. Kaczmarek, T. Lenzner und A. Neusar 2013). *Webseite:* [www.gesis.org](http://www.gesis.org). *Kontaktadresse:* natalja.menold@gesis.org.

**Rainer Metz** ist Titularprofessor für Wirtschaftsgeschichte und Methoden der empirischen Wirtschaftsforschung an der Universität St. Gallen und Mitarbeiter im Team „Historische Studien“ der Abteilung Datenarchiv bei GESIS, Leibniz-Institut für Sozialwissen. *Ausgewählte Publikationen:* Do Kondratieff Waves Exist? How Time Series Techniques Can Help to Solve the Problem, in: Cliometrica 5 (2011); Zeitreihenanalyse, in: Christof Wolf

und Henning Best (Hg.): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: Springer VS (2010); Auf der Suche nach den Langen Wellen der Konjunktur. Stuttgart: Steiner (2008). *Webseite:* [www.gesis.org](http://www.gesis.org). *Kontaktadresse:* rainer.metz@gesis.org.

**Christian Meyer** ist akademischer Oberrat an der Fakultät für Soziologie der Universität Bielefeld. Gegenwärtig vertritt er die Professur für Medientheorie an der Universität Siegen. *Ausgewählte Publikationen:* Sozialität in Slow Motion: Theoretische und empirische Perspektiven. Wiesbaden: Springer VS (zusammen mit R. Ayaß, 2012); The Rhetorical Emergence of Culture. New York: Berghahn (zusammen mit F. Girke, 2011); Ereignisethnographie und methodologischer Situationalismus: Auswege aus der Krise der ethnographischen Repräsentation?, in: Peter Berger et al. (Hg.): Feldforschung. Berlin: Weißensee Verlag (2009). *Kontaktadresse:* christian.meyer5@uni-bielefeld.de.

**Alexia Meyermann** ist wissenschaftliche Mitarbeiterin am Datenservicezentrum Betriebs- und Organisationsdaten (DSZ-BO) der Universität Bielefeld und derzeit tätig im Projekt „SOEP-LEE: Die Betriebsbefragung des Sozio-ökonomischen Panels“. *Ausgewählte Publikationen:* DSZ-BO-Item-Datenbank der empirischen Organisationsforschung, Version 1.0 – Hintergrund und Hinweise zur Benutzung. DSZ-BO Technical Report 6 (zusammen mit Annemarie Pankow und Maik Dammann, 2013); Datenmanagement. Eine zentrale Voraussetzung für den Erfolg der Data Sharing-Idee in den Sozialwissenschaften. DSZ-BO Working 2 (2012); The German Data Service Center for Business and Organizational Data (DSC-BO), in: Schmollers Jahrbuch 132, 4 (zusammen mit Susanne Edler et al., 2012). *Webseite:* [www.uni-bielefeld.de/dsz-bo](http://www.uni-bielefeld.de/dsz-bo). *Kontaktadresse:* alexia.meyermann@uni-bielefeld.de.

**Tobias Michael** ist Director des Bereichs Politik- und Sozialforschung bei Ipsos Public Affairs. *Ausgewählte Publikationen:* Innovationen im Rolling-Cross-Section-Design: Neue Ansätze zur Analyse des Wahlverhaltens, in: Zeitschrift für Politikforschung (zusammen mit C. Hoops und A. Glantz, 2013); Effizienzmessung und Steuerung des RCS-Designs. Eine Betrachtung im Hinblick auf die Bundestagswahl, in: DAGStat (zusammen mit C. Hoops, 2013); Die Qual mit der Wahl, in: Research & Results 5/2013 (zusammen mit H.-J. Friess, , 2013). *Webseite:* [www.ipsos.de](http://www.ipsos.de). *Kontaktadresse:* tobias.michael@ipsos.com.

**Tatjana Mika** ist Referatsleiterin im Bereich Forschung der Deutschen Rentenversicherung Bund und zuständig für das Forschungsdatenzentrum. *Ausgewählte Publikationen:* Risiken für eine Erwerbsminderung bei unterschiedlichen Berufsgruppen, in: Bundesgesundheitsblatt 56 (2013); Steigende Ungleichheit unter Aussiedlern und Spätaussiedlern im Alter, in: Baykara-Krumme et al. (Hg.): Viele Welten des Alterns: Wiesbaden Springer VS (zusammen mit Jochen Baumann, 2012); Kinderlosigkeit, Kindererziehung und Erwerbstätigkeitsmuster von Frauen in der Bundesrepublik und der DDR und ihre Auswirkungen auf das Alterseinkommen, in: Dirk Konietzka und Michaela Kreyenfeld (Hg.): Ein Leben ohne Kinder. Wiesbaden: Springer VS (zusammen mit Michael Stegmann, 2007). *Webseite:* [www.fdz-rv.de](http://www.fdz-rv.de). *Kontaktadresse:* Tatjana.Mika@drv-bund.de.

**Ekkehard Mochmann** war Direktor der GESIS- Leibniz-Institut für Sozialwissenschaften und Geschäftsführer des Zentralarchiv für Empirische Sozialforschung (ZA) an der Universität zu Köln. Er ist Mitglied der DGS-Sektion „Methoden der empirischen

Sozialforschung“. *Ausgewählte Publikationen*: Comparaisons interculturelles et recherches comparatives internationales, in: Alain Chenu und Laurent Lesnard (Hg.): La France dans les comparaisons internationales. Paris: Presses de Sciences Po (zusammen mit Dominique Joye 2011); e-Infrastructure for the Social Sciences, in: RatSWD (Hg.): Building on Progress. Opladen: Budrich UniPress (2010); Improving the evidence base for international comparative research, in: International Social Science Journal 59 (2008). *Kontaktadresse*: E.Mochmann@web.de.

**Andreas Motel-Klingebiel** ist Professor für Alternsforschung am National Institute for the Study of Ageing and Later Life (NISAL), Department of Social and Welfare Studies (ISV), Linköping University (LiU), Schweden. *Ausgewählte Publikationen*: Hochaltrigkeit in der Gesellschaft des langen Lebens: Theoretische Herausforderung, empirisches Problem und sozialpolitische Aufgabe, in: Zeitschrift für Gerontologie und Geriatrie, 46 (zusammen mit Jochen Ziegelmann und Maja Wiest, 2013); Altersarmut und die Lebensphase Alter, in: Claudia Vogel und Andreas Motel-Klingebiel (Hg.): Altern im sozialen Wandel: Rückkehr der Altersarmut?. Wiesbaden: Springer VS (zusammen mit Claudia Vogel, 2013); Welfare States and Quality of Later Life: Distributions and Predictions in a Comparative Perspective, in: European Journal of Ageing 6 (zusammen mit Laura Romeu-Gordo und Jörg Betzin, 2009). *Webseite*: <https://www.isv.liu.se/nisal/staff/motel-klingebiel-andreas>. *Kontaktadresse*: andreas.motel-klingebiel@liu.se.

**Andreas Mühlchen** ist Lehrbeauftragter am Institut für Politische Wissenschaft und Soziologie der Rheinischen Friedrich-Wilhelms-Universität Bonn. *Webseite*: [www.politik-soziologie.uni-bonn.de/institut/lehrkoerper/andreas-muehlichen-m.a](http://www.politik-soziologie.uni-bonn.de/institut/lehrkoerper/andreas-muehlichen-m.a). *Kontaktadresse*: a.muehlichen@uni-bonn.de.

**H. Peter Ohly** war bis 2010 beim GESIS – Leibniz-Institut für Sozialwissenschaften u.a. zuständig für die Reihe Sozialwissenschaften im Überblick sowie für sozialwissenschaftliche Internetquellen (SocioGuide) und war 2010-2014 internationaler Präsident der International Society for Knowledge Organization (ISKO). *Ausgewählte Publikationen*: Knowledge-Sociological and Information-Sociological Aspects of Knowledge Organization, in: Scire 19, 2 (2013); Challenges of Knowledge Organization and ISKO, in: SRELS 50, 6 (2013); Wissen – Wissenschaft – Organisation. Würzburg: Ergon (Hg., 2013); Zitationsanalyse: Beschreibung und Evaluation von Wissenschaft. In: Stegbauer und Häußling (Hg.): Handbuch Netzwerkforschung. Wiesbaden: Springer VS (2010). *Webseite*: <http://www.isko.org>. *Kontaktadresse*: peter.ohly@gmx.de.

**Grit Petschick** ist wissenschaftliche Mitarbeiterin am Fachgebiet Methoden der empirischen Sozialforschung an der Technischen Universität. *Ausgewählte Publikationen*: Einflussfaktoren auf die Karrieren von jungen Wissenschaftlerinnen und Wissenschaftlern während der Phase ihrer Promotion, in: K. Grabarz et al. (Hg.): Geschlechterunterschiede in der Selbstpräsentation und der Vernetzung als ausgewählte Karrierestrategien von MathematikerInnen und PhysikerInnen in Wissenschaft und Wirtschaft. Berlin: Logos (2014); Geschlechterpraktiken in neuen Formen der Forschungsorganisation am Beispiel der Physik, in: Ulrike Busolt et al. (Hg.): Karriereverläufe in Forschung und Entwicklung. Logos: Berlin (2014); Frauenförderung zwischen heterogenen Logiken, in: Swiss Journal

of Sociology 39, 2 (zusammen mit Robert J. Schmidt und Maria Norkus, 2013). *Webseite:* [www.tu-berlin.de/?id=72718](http://www.tu-berlin.de/?id=72718). *Kontaktadresse:* grit.petschick@tu-berlin.de.

**Lisa Pfahl** ist Juniorprofessorin für Disability Studies an der Humboldt-Universität zu Berlin. *Ausgewählte Publikationen:* Das Fleisch der Diskurse. Zur Verbindung von Biographie- und Diskursforschung in der wissenssoziologischen Subjektivierungsanalyse am Beispiel der Behindertenpädagogik, in: Susann Fegter et al. (Hg.): Erziehungswissenschaftliche Diskursforschung. Wiesbaden: VS (zusammen mit Lena Schürmann und Boris Traue, 2014); Die Erfahrung des Diskurses. Zur Methode der Subjektivierungsanalyse in der Untersuchung von Bildungsprozessen, in: Reiner Keller und Inga Truschkat (Hg.): Methodologie und Praxis der wissenssoziologischen Diskursanalyse. Wiesbaden: VS-Verlag (zusammen mit Boris Traue, 2012); Bildung, Behinderung und Agency. Eine wissenssoziologisch-diskursanalytische Perspektive zur Untersuchung der Folgen schulischer Segregation und Inklusion, in: Rolf Becker und Heike Solga (Hg.): Soziologische Bildungsforschung. Sonderheft der KZfSS 52, 12 (2011). *Webseite:* <http://www.reha.hu-berlin.de/personal/mitarbeiter>. *Kontaktadresse:* pfahl@hu-berlin.de.

**Rolf Porst** ist Inhaber der Firma DER.PORST – Service für Markt- und Sozialforschung und war mehr als 25 Jahre lang wissenschaftlicher Mitarbeiter bei GESIS-Leibniz-Institut für Sozialwissenschaften (vormals ZUMA). *Ausgewählte Publikationen:* Praxis der Umfrageforschung. Studienkripten zur Soziologie, Stuttgart u.a.: Teubner (2. Auflage, 2000); Interviewers' Deviation in Surveys. Impact, Reasons, Detection and Prevention, Frankfurt a.M.: Peter Lang (Hrgs. zusammen mit Peter Winker und Natalja Menold, 2013); Fragebogen – ein Arbeitsbuch. Studienkripten zur Soziologie. Wiesbaden: Springer VS (4. Auflage, 2014). *Kontaktadresse:* der.porst@t-online.de.

**Manuela Pötschke** ist akademische Oberrätin für Angewandte Statistik am Fachbereich 05 der Universität Kassel. *Ausgewählte Publikationen:* Datengewinnung, Datenaufbereitung und explorative Datenanalyse, in: Christoph Wolf und Henning Best (Hg.): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS-Verlag (2010); Mehrebenenmodelle: angemessene Modellierung von Evaluationsdaten, in: Philipp Pohlenz und Antje Oppermann (Hg.): Lehre und Studium professionell evaluieren. Bielfeld: UVW (2010); Mehrebenenanalyse, in: Joachim Behnke et al. (Hg.): Methoden der Politikwissenschaft. Baden-Baden: Nomos (2006). *Webseite:* [www.uni-kassel.de/go/angst](http://www.uni-kassel.de/go/angst). *Kontaktadresse:* manuela.poetschke@uni-kassel.de.

**Ulrich Pötter** ist wissenschaftlicher Mitarbeiter am Deutschen Jugendinstitut in München. *Ausgewählte Publikationen:* Nachträgliche Chancen? Der Beitrag von Übergangsmaßnahmen und Ausbildungsförderung für den Berufseinstieg im „Laboratorium Ostdeutschland“. In: Solga, Heike/Becker, Rolf (Hg.): Sonderband der Kölner Zeitschrift für Soziologie und Sozialpsychologie (2012 mit G. Prein, Ch. Steiner); Einführung in die Statistik mit R. München: Vahlen (2010 mit A. Behr); Wahrscheinlichkeit. Begriff und Rhetorik in der Sozialwissenschaft. Weinheim: Juventa (2002 mit G. Rohwer) *Webseite:* [www.dji.de/cgi-bin/Mitarbeiter/homepage/mitarbeiterseite.php?mitarbeiter=1472](http://www.dji.de/cgi-bin/Mitarbeiter/homepage/mitarbeiterseite.php?mitarbeiter=1472). *Kontaktadresse:* poetter@dji.de.

**Gerald Prein** ist wissenschaftlicher Mitarbeiter am Deutschen Jugendinstitut in München. *Ausgewählte Publikationen:* Effekte der Angebotsdichte sozialstaatlicher Leistungen und Angebote auf die soziale Selektivität der Inanspruchnahme. Empirische Hinweise an den Beispielen Kinderbetreuung und Jugendzentren, in: Zeitschrift für Sozialreform 59, 1 (zusammen mit E. van Santen, 2013); Nachträgliche Chancen? Der Beitrag von Übergangsmaßnahmen und Ausbildungsförderung für den Berufseinstieg im „Laboratorium Ostdeutschland“, in: Solga, Heike/Becker, Rolf (Hg.): Sonderband der KZfSS (zusammen mit Ulrich Pötter und Ch. Steiner, 2012). *Webseite:* [www.dji.de/cgi-bin/Mitarbeiter/home-page/mitarbeiterseite.php?mitarbeiter=522](http://www.dji.de/cgi-bin/Mitarbeiter/home-page/mitarbeiterseite.php?mitarbeiter=522). *Kontaktadresse:* prein@dji.de.

**Monika Pritzel** ist Professorin für Physiologische Psychologie an der Universität Koblenz-Landau, Campus Landau. *Ausgewählte Publikationen:* 150 Jahre Pfalzklinikum. Psychiatrie, Psychotherapie und Nervenheilkunde in Klingenmünster. Stuttgart: Steiner Verlag (zusammen mit R. Steinberg, 2012); Die „Traumzeit“ im kollektiven Gedächtnis australischer Ureinwohner. Kröning: Asanger Verlag. (2006). Gehirn und Verhalten. Ein Grundkurs der physiologischen Psychologie. Heidelberg: Spektrum Akademischer Verlag (zusammen mit M. Brand, M. und H.J. Markowitsch, 2003). *Webseite:* [www.uni-koblenz-landau.de/landau](http://www.uni-koblenz-landau.de/landau). *Kontaktadresse:* pritzel@uni-landau.de.

**Aglaja Przyborski** lehrt an der Universität Wien und der Sigmund Freud Privat Universität. *Ausgewählte Publikationen:* Qualitative Sozialforschung. Ein Arbeitsbuch. Berlin: De Gruyter Oldenbourg (zusammen mit Monika Wohlrab-Sahr, 4. Auflage 2014); Learning to Think Iconically in the Human and Social Sciences. Iconic Standards of Understanding as a Pivotal Challenge for Method Development, in: Integrative Psychological and Behavioral Science 4, 1 (zusammen mit Thomas Slunecko, 2011); Das politische Bild. Situation Room: Ein Foto – vier Analysen. Opladen, Berlin, Toronto: Verlag Barbara Budrich (zusammen mit Günther Haller Hg., 2014). *Webseite:* [iconicom.univie.ac.at/](http://iconicom.univie.ac.at/). *Kontaktadresse:* aglaja.przyborski@univie.ac.at.

**Stefan Rädiker** war bis 2013 wissenschaftlicher Mitarbeiter der Marburger Arbeitsgruppe für Methoden & Evaluation an der Philipps-Universität Marburg und ist inzwischen Chief Technology Officer der VERBI GmbH in Berlin. *Ausgewählte Publikationen:* Statistik. Eine verständliche Einführung. Wiesbaden: VS Verlag für Sozialwissenschaften (zusammen mit Udo Kuckartz, Thomas Ebert und Julia Schehl, 2013); Computergestützte Verfahren, in: Bock, Karin/Miethe, Ingrid (Hg.): Handbuch qualitative Methoden in der Sozialen Arbeit. Opladen/Farmington Hills: Barbara Budrich (zusammen mit Udo Kuckartz, 2010). *Webseite:* [www.methoden-evaluation.de](http://www.methoden-evaluation.de). *Kontaktadresse:* stefan.raediker@maxqda.de.

**Susanne Rässler** ist Professorin für Statistik und Ökonometrie an der Otto-Friedrich-Universität Bamberg. *Ausgewählte Publikationen:* Imputation, in: Computational Statistics 5, 1 (zusammen mit D.B. Rubin und E.R. Zell, 2013); Statistical Matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches. New York: Springer (2002); Incomplete Data: Diagnosis, Imputation, and Estimation, in: E.D. de Leeuw et al. (Hg.): International Handbook of Survey Methodology. Hillsdale: Erlbaum (zusammen mit D.B. Rubin und N. Schenker, 2008). *Webseite:* [www.uni-bamberg.de/stat-oek/](http://www.uni-bamberg.de/stat-oek/). *Kontaktadresse:* susanne.raessler@uni-bamberg.de.

**Jo Reichertz** ist Professor für Kommunikationswissenschaft an der Universität Duisburg-Essen. *Ausgewählte Publikationen:* Gemeinsam Interpretieren oder: Die Gruppeninterpretation als kommunikativer Prozess. Wiesbaden: Springer (2013); Kommunikationsmacht. Was ist Kommunikation und was vermag sie?. Wiesbaden: Springer (2010); Einführung in die qualitative Videoanalyse. Wiesbaden: VS Verlag (zusammen mit Carina Englert, 2010). *Webseite:* [www.uni-due.de/kowi/Jreichertz.html](http://www.uni-due.de/kowi/Jreichertz.html). *Kontaktadresse:* Jo.Reichert@uni-due.de.

**Jost Reinecke** ist Professor für Quantitative Methoden der empirischen Sozialforschung an der Fakultät für Soziologie der Universität Bielefeld. *Ausgewählte Publikationen:* The Development of Delinquency During Adolescence: A Comparison of Missing Data Techniques, in: Quality & Quantity, 47, 6 (zusammen mit C. Weins, 2013); Multiple Imputation of Incomplete Zero-inflated Count Data, in: Statistica Neerlandica 67, 3 (zusammen mit K. Kleinke, 2013); Wachstumsmodelle. Mehring: Rainer Hampp Verlag (2012). *Webseite:* [www.uni-bielefeld.de/soz/personen/reinecke/index.html](http://www.uni-bielefeld.de/soz/personen/reinecke/index.html). *Kontaktadresse:* jost.reinecke@uni-bielefeld.de.

**Karl-Heinz Reuband** ist Professor (em.) für Soziologie an der Heinrich-Heine Universität Düsseldorf. *Ausgewählte Publikationen:* What We Knew. Terror, Mass Murder and Everyday Life in Nazi Germany. An Oral History. New York: Basic Books (zusammen mit Eric Johnson, 2005); How People Learned About the September 11 2001 Terrorist Attack and How It Affected Them. A Study in News Diffusion and Psychosocial Reactions in Germany, in: Tilo Beckers, Klaus Birkelbach, Jörg Hagenau und Ulrich Rosar (Hg.): Komparative empirische Sozialforschung. Wiesbaden: VS Verlag (2010); Indirect and „Hidden“ Surveys: An Almost Forgotten Survey Technique from the Early Years, in: Hannes Haas, Hynek Jerabek und Thomas Petersen (Hg.) The Early Days of Survey Research and their Importance Today. Wien: Braumüller Verlag (2012). *Kontaktadresse:* reuband@phil.uni-duesseldorf.de.

**Gabriele Rosenthal** ist Professorin für Qualitative Methoden an der Georg-August Universität Göttingen. *Ausgewählte Publikationen:* Interpretative Sozialforschung. Weinheim und München: Juventa (2011); Erlebte und erzählte Lebensgeschichte. Frankfurt a. M.: Campus (1995). *Webseite:* [www.uni-goettingen.de/de/sh/27012.html](http://www.uni-goettingen.de/de/sh/27012.html). *Kontaktadresse:* g.rosenthal@gmx.de.

**Axel Salheiser** ist wissenschaftlicher Mitarbeiter am Institut für Soziologie an der Friedrich-Schiller-Universität Jena. *Ausgewählte Publikationen:* Parteidrei, plangemäß, professionell? Rekrutierungsmuster und Karriereverläufe von DDR-Industriekadern. Wiesbaden: VS Verlag (2009); Elites and Social Change – The Socialist and Post-Socialist Experience. Hamburg: Kraemer Publishers (Hg. zusammen mit Heinrich Best und Ronald Gebauer, 2009); Handling Ideological Bias and Shifting Validity of Longitudinal Data: The Case of Process-Generated Data on GDR Elites, in: Nina Baur (Hrgs.): Linking Theory and Data. Process-Generated and Longitudinal Data for Analyzing Long-Term Social Processes, in: Sonderheft Historical Social Research 34, 1 (2009). *Webseite:* [www.soziologie.uni-jena.de](http://www.soziologie.uni-jena.de). *Kontaktadresse:* axel.salheiser@uni-jena.de.

**Björn Oliver Schmidt** ist wissenschaftlicher Mitarbeiter am Sozialwissenschaftlichen Methodenzentrum und Arbeitsgebiet für Statistik und empirische Sozialforschung der Universität Bremen. *Webseite:* [www.sozialforschung.uni-bremen.de](http://www.sozialforschung.uni-bremen.de). *Kontaktadresse:* bschmidt@empas.uni-bremen.de.

**Robert J. Schmidt** ist wissenschaftlicher Mitarbeiter am Fachgebiet Organisationssoziologie an der Technischen Universität Berlin. *Ausgewählte Publikationen:* Regulation of Bodies and Things: A Structurationist Note on Robust Innovation in Regional Clusters, in: Bianca M. Pirani und Thomas S. Smith (Hg.): Body and Time. Newcastle upon Tyne: Cambridge Scholars Publishing (2013); Frauenförderung zwischen heterogenen Logiken: Der Fall eines Nachwuchsförderprogramms in der deutschen Exzellenzinitiative, in: Swiss Journal of Sociology 39, 2 (zusammen mit Grit Petschick und Maria Norkus, 2013); Versprechen der Innovation, in: Bormann, Inka et al. (Hg.): Indikatoren des Neuen. Wiesbaden: Springer VS (zusammen mit Andrea Pronzini und Cristina Besio, 2012). *Webseite:* [www.os.tu-berlin.de](http://www.os.tu-berlin.de). *Kontaktadresse:* robert.schmidt@tu-berlin.de.

**Tobias Schmies** ist wissenschaftlicher Mitarbeiter am Institut für Politische Wissenschaft und Soziologie an der Rheinischen Friedrich-Wilhelms-Universität Bonn. *Webseite:* <http://www.soziologie.uni-bonn.de>. *Kontaktadresse:* t.schmies@uni-bonn.de.

**Andreas Schmitz** ist wissenschaftlicher Mitarbeiter in der Abteilung für Soziologie an der Rheinischen Friedrich-Wilhelms-Universität Bonn. *Ausgewählte Publikationen:* The Structure of Mate Choice. Theory, Methodology, and Empirical Findings (2013); Survey Topic and Unit Nonresponse – Evidence from an Online Survey on Mating, in: Quality & Quantity (zusammen Doreen Zillmann und Hans-Peter Blossfeld, 2013); Elective Affinities 2.0? A Bourdieusian Approach to Couple Formation and the Methodology of E-Dating, in: RESET 1, 1 (2012). *Webseite:* [www.partnerwahlforschung.de](http://www.partnerwahlforschung.de). *Kontaktadresse:* andreas.schmitz@uni-bonn.de.

**Bernt Schnettler** ist Professor für Kultur- und Religionssoziologie an der Universität Bayreuth und war von 2011 bis 2013 Chair des ESA-Research Network Qualitative Methods *Ausgewählte Publikationen:* Videographie, Wiesbaden: Springer VS (zusammen mit René Tuma und Hubert Knoblauch, 2013). Phänomenologie und Soziologie, Wiesbaden: VS (zusammen mit Jürgen Raab, Michaela Pfadenhauer, Peter Stegmaier und Jochen Dreher, 2008), Video-Analysis. Methodology and Methods, Frankfurt a.M.: Peter Lang (zusammen mit Hubert Knoblauch, Jürgen Raab und Hans-Georg Soeffner, 2006). *Webseite:* [www.soz.uni-bayreuth.de](http://www.soz.uni-bayreuth.de). *Kontaktadresse:* schnettler@uni-bayreuth.de.

**Cornelius Schubert** ist PostDoc im DFG Graduiertenkolleg „Locating Media“ an der Universität Siegen. *Ausgewählte Publikationen:* The Means of Managing Momentum. Bridging Technological Paths and Organisational Fields, in: Research Policy 42 (zusammen mit Jörg Sydow und Arnold Windeler 2013); Die Technik operiert mit. Zur Mikroanalyse ärztlicher Arbeit, in: ZfS 40, 4 (2011); Videographic Elicitation Interviews. Exploring Technologies, Practices and Narratives in organisations, in: Kissmann, Ulrike T. (Hg.): Video Interaction Analysis. Frankfurt a.M.: Peter Lang (2009). *Webseite:* [www.uni-siegen.de/locatingmedia/](http://www.uni-siegen.de/locatingmedia/). *Kontaktadresse:* cornelius.schubert@uni-siegen.de.

**Jürgen Schupp** ist Professor für Soziologie, insbesondere empirische Sozialforschung an der Freien Universität Berlin und Direktor des Sozio-ökonomischen Panels (SOEP) im Deutschen Institut für Wirtschaftsforschung (DIW Berlin). *Ausgewählte Publikationen:* Angst und Ärger: Zur Relevanz emotionaler Dimensionen sozialer Ungleichheit, in: ZfS 41, 5 (zusammen mit Katja Rackow und Christian von Scheve, 2012); Short Assessment of the Big Five: Roboust Across Survey Methods Except Telephone Interviewing, in: Behavior Research Methods 43, 2 (zusammen mit Frieder R. Lang et al., 2011); 25 Jahre Sozio-ökonomisches Panel. Ein Infrastrukturprojekt der empirischen Sozial- und Wirtschaftsforschung in Deutschland, in: ZfS 38, 5 (2009). *Webseite:* [www.diw.de/de/cv/jschupp](http://www.diw.de/de/cv/jschupp). *Kontaktadresse:* jschupp@diw.de.

**Lena Schürmann** ist wissenschaftliche Mitarbeiterin an der Hochschule für Wirtschaft und Recht Berlin. *Ausgewählte Publikationen:* Schmutz als Beruf. Prekarisierung, Klasse und Geschlecht in der Reinigungsbranche. Münster: Westfälisches Dampfboot (2013). *Kontaktadresse:* lschuer@hwr-berlin.de.

**Miriam Schütte** ist wissenschaftliche Mitarbeiterin in der Abteilung Soziologie an der Rheinischen Friedrich-Wilhelms Universität Bonn. *Webseite:* <http://www.politik-soziologie.uni-bonn.de/institut/lehrkoerper/miriam-schuette-m.a>. *Kontaktadresse:* miriam.schuette@uni-bonn.de.

**Ulrich Schwinges** arbeitet als freier Medientrainer. *Ausgewählte Publikationen:* Vor Mikrofon und Kamera. Zur entspannten Zusammenarbeit mit Journalisten. Berlin und Hamburg: VZBV (2011); Das Journalistische Interview. VS: Wiesbaden (zusammen mit Jürgen Friedrichs, 2009); Städtebau und Kriminalität in Solingen. Solingen. [www.gesis.org/sowiport/search/id/iz-solis-90044349](http://www.gesis.org/sowiport/search/id/iz-solis-90044349) (1984). *Webseite:* [www.medientraining.tv](http://www.medientraining.tv). *Kontaktadresse:* mail@schwinges.de.

**Julia Simonson** ist Leiterin des Arbeitsbereichs Forschung und stellvertretende Institutsleiterin am Deutschen Zentrum für Altersfragen (DZA) in Berlin. *Ausgewählte Publikationen:* Ungleichheit sozialer Teilhabe im Alter, in: Zeitschrift für Gerontologie + Geriatrie, 46, 5, (zusammen mit Christine Hagen, Claudia Vogel und Andreas Motel-Klingebiel, 2013); Erwerbsverläufe im Wandel – Konsequenzen und Risiken für die Alterssicherung der Baby-boomer, in: Claudia Vogel und Andreas Motel-Klingebiel (Hg.): Altern im sozialen Wandel. Wiesbaden: Springer VS (2013); Changing Employment Patterns of Women in Germany, in: Advances in Life Course Research 16, 2 (zusammen mit Laura Romeu Gordo und Nadiya Titova, 2011). *Webseite:* <http://www.dza.de>. *Kontaktadresse:* Julia.Simonson@dza.de.

**Michael Stegmann** ist Regierungsdirektor im Bereich Informationsverarbeitung der Deutschen Rentenversicherung Bund und dort Referatsleiter Statistisches Berichtswesen und Forschungsdaten. *Ausgewählte Publikationen:* Kinderlosigkeit, Kindererziehung und Erwerbstätigkeitsmuster von Frauen in der Bundesrepublik und der DDR und ihre Auswirkungen auf das Alterseinkommen, in: Dirk Konietzka und Michaela Kreyenfeld (Hg.): Ein Leben ohne Kinder. Wiesbaden: Springer VS (zusammen mit Tatjana Mika, 2007); Längsschnittdaten der Rentenversicherung für die Erwerbsbiografieforschung, in: Deutsche Rentenversicherung 2, 218 (2010). *Webseite:* [www.fdz-rv.de](http://www.fdz-rv.de). *Kontaktadresse:* dr.michael.stegmann@drv-bund.de.

**Petra Stein** ist Professorin für Empirischen Sozialforschung an der Universität Duisburg-Essen. *Ausgewählte Publikationen:* A Nonlinear Simultaneous Probit-Model for the Investigation of Decision-Making Processes: Modelling the Process of Setting up a Family in Partnerships, in: Quality and Quantity 47, 3 (gemeinsam mit Monika Pavetic, 2013); Mixture of Covariance Structure Models to Identify Different Types of Live Style, in: Methodology 2, 3 (2006); Mixtures of Conditional Mean- and Covariance Structure Models, in: Psychometrika 64, 4 (zusammen mit Jörg Wittenberg und Gerhard Arminger 1999). *Kontaktadresse:* petra\_stein@uni-due.de.

**Volker Stocké** ist Professor für Methoden der empirischen Sozialforschung an der Universität Kassel. *Ausgewählte Publikationen:* Inequality in Transitions to Secondary School and Tertiary Education in the German School System, in Jackson, M. (Hg.), Determined to Succeed? Stanford: Stanford University Press (zusammen mit M. Neugebauer, D. Reimer und S. Schindler, 2013). Response Privacy and Elapsed Time since Election Day as Determinants for Vote Overreporting, in: International Journal of Public Opinion Research 19 (2007). Determinants and Consequences of Survey Respondents' Social Desirability Beliefs about Racial Attitudes, in: Methodology 3 (2007). *Webseite:* www.uni-kassel.de/fb05/fachgruppen/soziologie/chair-for-methods-of-empirical-social-research.html. *Kontaktadresse:* volker.stocke@uni-kassel.de.

**Jörg Strübing** ist Professor für Soziologie und lehrt Methoden der qualitativen Sozialforschung an der Eberhard Karls Universität Tübingen. *Ausgewählte Publikationen:* Qualitative Sozialforschung. Eine komprimierte Einführung für Studierende. München: Oldenbourg (2013); Grounded Theory. Zur sozialtheoretischen und epistemologischen Fundierung des Verfahrens der empirisch begründeten Theoriebildung. Wiesbaden: Springer VS (3. Auflage, 2014); Research as Pragmatic Problem-Solving. The Pragmatist Roots of Empirically-Grounded Theorizing, in Bryant, A./Charmaz, K. (Hg.): The Sage Handbook of Grounded Theory London: Sage (2007). *Webseite:* www.soziologie.uni-tuebingen.de/struebing. *Kontaktadresse:* joerg.struebing@uni-tuebingen.de.

**Cornelia Thierbach** ist wissenschaftliche Mitarbeiterin am Fachgebiet Methoden der empirischen Sozialforschung an der Technischen Universität Berlin. *Ausgewählte Publikationen:* Exploring the Orientation in Space. Mixing Focused Ethnography and Surveys in Social Experiments, in: HSR (zusammen mit Alexandra Lorenz, 2014); Theory and Methods in Spatial Analysis. Towards Integrating Qualitative, Quantitative and Cartographic Approaches in the Social Sciences and Humanities, in HSR (zusammen mit Nina Baur, Linda Hering und Anna Laura Raschke, 2014); Map Design Aspects, Route Complexity, or Social Background? Factors Influencing User Satisfaction with Indoor Navigation Maps, in: CaGis 40, 3 (zusammen mit Alexandra Lorenz et al., 2013). *Webseite:* www.mes.tu-berlin.de. *Kontaktadresse:* cornelia.thierbach@tu-berlin.de.

**Helmut Thome** ist Professor i. R. für Soziologie, insbes. Methoden der empirischen Sozialforschung an der Martin-Luther-Universität Halle-Wittenberg. *Ausgewählte Publikationen:* Violent Crime (and Suicide) in Imperial Germany, 1883-1902: Quantitative Analyses and a Durkheimian Interpretation, in: International Criminal Justice Review 20, 1 (2010). Sozialer Wandel und Gewaltkriminalität. Deutschland, England und Schweden

im Vergleich, 1950-2000. Wiesbaden: VS Verlag (zusammen mit Christoph Birkel, 2007). Zeitreihenanalyse. Eine Einführung für Sozialwissenschaftler und Historiker. München: Oldenbourg (2005). *Kontaktadresse*: helmut.thome@soziologie.uni-halle.de.

**Boris Traue** ist wissenschaftlicher Mitarbeiter und Leiter des DFG-Projekts „Audiovisuelle Kulturen der Selbstthematisierung“ an der Technischen Universität Berlin, (Fachgebiet Allgemeine Soziologie) sowie Fellow am Digital Cultures Research Lab (DCRL) der Leuphana Universität Lüneburg. *Ausgewählte Publikationen*: Visuelle Diskursanalyse. Ein programmatischer Vorschlag zur Untersuchung von Sicht- und Sagbarkeiten im Medienwandel, in: Zeitschrift für Diskursforschung, 2, 1 (2013); Kommunikationsregime, in: Reiner Kelle et al. (Hg.): Kommunikativer Konstruktivismus. Wiesbaden: Springer VS (2012); Das Subjekt der Beratung. Zur Soziologie einer Psycho-Technik. Bielefeld: transcript (2010). *Webseite*: videosoziologie.net. *Kontaktadresse*: boris.traue@tu-berlin.de.

**René Tuma** ist wissenschaftlicher Mitarbeiter am Institut für Soziologie der TU-Berlin. *Ausgewählte Publikationen*: Videographie. Wiesbaden: Springer VS (zusammen mit Bernt Schnettler und Hubert Knoblauch, 2013); Visuelles Wissen und Bilder des Sozialen. Aktuelle Entwicklungen in der Soziologie des Visuellen, Wiesbaden: VS-Springer (zusammen mit Petra Lucht und Lisa Marian Schmidt, 2012); Videography: An Interpretative Approach to Video-Recorded Micro-Social Interaction, in: The SAGE Handbook of Visual Research Methods. Sage Publications Ltd. (zusammen mit Hubert Knoblauch, 2011). *Webseite*: www.tu-berlin.de/?id=74301. *Kontaktadresse*: rene.tuma@tu-berlin.de.

**Susanne Vogl** ist wissenschaftliche Mitarbeiterin am Lehrstuhl für Soziologie und empirische Sozialforschung an der Katholischen Universität Eichstätt-Ingolstadt. *Ausgewählte Publikationen*: Telephone versus Face-to-Face Interviews: Mode Effect on Semi-Structured Interviews with Children, in: Sociological Methodology 43, 1 (2013); Alter und Methode: Ein Vergleich telefonischer und persönlicher Leitfadeninterviews mit Kindern. Wiesbaden: VS Verlag (2012); Gruppendiskussionen mit Kindern – methodische und methodologische Besonderheiten, in: ZA-Information 57 (2005). *Webseite*: www.susanne-vogl.com. *Kontaktadresse*: Susanne.vogl@ku-eichstaett.de.

**Michael Wagner** ist Professor für Soziologie am Institut für Soziologie und Sozialpsychologie der Universität zu Köln. *Ausgewählte Publikationen*: On the Links between Employment, Partnership Quality and the Intention to have a First Child: The Case of West Germany, in: Demographic Research 24, Article 24 (zusammen mit Ina Berninger und Bernd Weiß, 2011); Union Dissolution and Mobility: Who Moves From the Family Home After Separation? In: Journal of Marriage and Family 72, 5 (zusammen mit Clara H. Mulder, 2010); Bilanz der deutschen Scheidungsforschung. Versuch einer Meta-Analyse, in: ZfS 32 1 (zusammen mit Bernd Weiß, 2003). *Webseite*: www.iss-wiso.uni-koeln.de/wagner.html. *Kontaktadresse*: mwagner@wiso.uni-koeln.de.

**Pia Wagner** ist Referentin für Studium und Lehre an der Fakultät VI Planen Bauen Umwelt an der Technischen Universität Berlin. *Ausgewählte Publikationen*: Die moderne Sozialstrukturanalyse und das Problem der Operationalisierung von Intersektionalität.

Eine messtheoretische Perspektive, in: Erwägen – Wissen – Ethik (EWE) 24, 3 (zusammen mit Nina Baur, 2014). *Webseite:* [www.planen-bauen-umwelt.tu-berlin.de/menue/studium\\_und\\_lehre](http://www.planen-bauen-umwelt.tu-berlin.de/menue/studium_und_lehre). *Kontaktadresse:* pia.wagner@tu-berlin.de.

**Uwe Warner** ist Sozialforscher bei CEPS/INSTEAD (Centre d'Etudes de Populations, de Pauvreté et de Politiques SocioEconomiques/International Network for Studies in Technology, Environment, Alternatives, Development) in Esch/Alzette in Luxembourg. *Ausgewählte Publikationen:* Soziodemographische Standards für Umfragen in Europa. München und Mering: Rainer Hampp Verlag (zusammen mit Jürgen Hoffmeyer-Zlotnik, 2013); Harmonisierung demographischer und sozio-ökonomischer Variablen. Instrumente für die international vergleichende Sozialforschung. Wiesbaden: Springer VS (zusammen mit Jürgen Hoffmeyer-Zlotnik, 2012); Die Messung von Einkommen in der international vergleichenden Umfrageforschung. Mannheim: FRG (2009). *Kontaktadresse:* uwe.warner@ceps.lu.

**Karsten Weber** vertritt derzeit den Lehrstuhl für Allgemeine Technikwissenschaften an der Brandenburgischen Technischen Universität Cottbus – Senftenberg und lehrt außerdem Technikfolgenabschätzung an der OTH Regensburg. *Ausgewählte Publikationen:* Erklärung historischer Abläufe mit Computersimulationen, in: Neue Politische Ökonomie, Sonderheft Historical Social Research/Historische Sozialforschung, 32, 4 (2007); Simulationen in den Sozialwissenschaften, in: Journal for General Philosophy of Science, 38, 1 (2007); Search Engine Bias, in: Dirk Lewandowski (Hg.): Handbuch Internet-Suchmaschinen, Band 2 Heidelberg: Akademische Verlagsanstalt AKA (2011). *Webseite:* [www.tu-cottbus.de/fakultaet1/de/technikwissenschaften/](http://www.tu-cottbus.de/fakultaet1/de/technikwissenschaften/). *Kontaktadresse:* Karsten.Weber@tu-cottbus.de.

**Martin Weichbold** ist außerordentlicher Professor für Soziologie und empirische Sozialforschung an der Universität Salzburg. *Ausgewählte Publikationen:* Methodik und Methodologie interkultureller Umfrageforschung. Wiesbaden: Springer VS (zusammen mit Reinhard Bachleitner et al., 2014); Die Befragung im Kontext von Raum, Zeit und Befindlichkeit. Beiträge zu einer prozessorientierten Theorie der Umfrageforschung. Wiesbaden: Springer VS (zusammen mit Reinhard Bachleitner und Wolfgang Aschauer, 2010); Zur Bestimmung und Sicherung der Qualität von Umfragen, in: Unfrageforschung, Sonderband 9 der ÖZS. Wiesbaden: Springer VS (2009). *Webseite:* [www.uni-salzburg.at/soziologie/weichbold](http://www.uni-salzburg.at/soziologie/weichbold). *Kontaktadresse:* martin.weichbold@sbg.ac.at.

**Bernd Weiß** ist wissenschaftlicher Mitarbeiter am Institut für Soziologie und Sozialpsychologie der Universität zu Köln. *Ausgewählte Publikationen:* On the Links between Employment, Partnership Quality and the Intention to have a First Child: The Case of West Germany, in: Demographic Research 24, Article 24 (zusammen mit Ina Berninger und Michael Wagner, 2011); The Identification and Prevention of Publication Bias in the Social Sciences and Economics, in: Journal of Economics and Statistics 231 (zusammen mit Michael Wagner, 2011); Potentiale und Probleme von Meta-Analysen in der Soziologie, in: Sozialer Fortschritt 57 (zusammen mit Michael Wagner, 2008). *Webseite:* [www.berndweiss.net](http://www.berndweiss.net). *Kontaktadresse:* bernd.weiss@uni-koeln.de.

**Monika Wohlrab-Sahr** ist Professorin für Kultursoziologie an der Universität Leipzig. *Ausgewählte Publikationen:* Qualitative Sozialforschung. Ein Arbeitsbuch. München: Oldenbourg (zusammen mit Aglaja Przyborski, 4. Auflage 2014); Multiple Secularities: Towards a Cultural Sociology of Secular Modernities, in: Comparative Sociology 11 (zusammen mit Marian Burchardt, 2013); Forcierte Säkularität. Frankfurt a.M.: Campus (zusammen mit Uta Karstein und Thomas Schmidt-Lux, 2009). *Kontaktadresse:* wohlram@uni-leipzig.de.

**Olga Yanenko** ist wissenschaftliche Mitarbeiterin am Lehrstuhl für Angewandte Informatik in den Kultur-, Geschichts- und Geowissenschaften an der Otto-Friedrich-Universität Bamberg. *Ausgewählte Publikationen:* Enhancing the Quality of Volunteered Geographic Information: A Constraint-Based Approach, in: Bridging the Geographic Information Sciences. Lecture Notes in Geoinformation and Cartography 8 (zusammen mit Christoph Schlieder, 2012); Identifying Artificial Actors in E-Dating: A Probabilistic Segmentation Based on Interactional Pattern Analysis, in: Business and Economics Challenges at the Interface of Data Analysis, Computer Science, and Optimization Studies in Classification, Data Analysis, and Knowledge Organization 5, 319-327 (zusammen mit Andreas Schmitz und Marcel Hebing, 2012); Spatio-Temporal Proximity and Social Distance: A Confirmation Framework for Social Reporting, in: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks (LBSN '10), New York (zusammen mit Christoph Schlieder, 2010). *Webseite:* www.uni-bamberg.de/kinf/mitarbeiter/yanenko-olga/. *Kontaktadresse:* olga.yanenko@uni-bamberg.de.

**Markus Ziegler** ist Mitarbeiter im Bereich Marketing & Data Sciences der GfK SE. *Ausgewählte Publikationen:* Skalenkonstruktion nach Mokken für mehrdimensionale Variablenstrukturen. Ein Anwendungsbeispiel mit SPSS, in: Bamberger Beiträge zur empirischen Sozialforschung 14 (zusammen mit Leila Akremi 2007). *Webseite:* www.mes.tu-berlin.de. *Kontaktadresse:* m\_ziegler@mailbox.tu-berlin.de.

**Cornelia Züll** ist Mitarbeiterin bei GESIS, Leibniz-Institut für Sozialwissenschaften, in der Abteilung Survey Design and Methodology. *Ausgewählte Publikationen:* Item Non-Response in Open-Ended Questions: Who Does Not Answer on the Meaning of Left and Right?, in: Social Science Research 41, 6 (zusammen mit Evi Scholz, 2012); Standardisierte Erfassung von Verweigerungsgründen in Face-to-Face-Umfragen, in: MDA 5, 1 (zusammen mit Natalja Menold, 2010); Identifying Events Using Computer-Assisted Text Analysis, in: Social Science Computer Review 26, 4 (zusammen mit Juliane Landmann 2008). *Webseite:* www.gesis.org. *Kontaktadresse:* cornelia.zuell@gesis.org.

---

## Danksagung

Das hier vorliegende Handbuch ist längst nicht der erste Versuch einer Systematisierung der sozialwissenschaftlichen Erhebungsmethoden – bereits vor mehr als fünfzig Jahren gab René König (1962) das erste systematische „Handbuch der empirischen Sozialforschung“ heraus (Stuttgart: Enke). Behandelt wurden Themen wie „Die Beobachtung“, „Das Interview in der Sozialforschung“, „Analyse von Panel-Befragungen“, „Gruppendiskussion“ und „Die biographische Methode“, womit sowohl die Bereiche der quantitativen als auch die der qualitativen Sozialforschung berücksichtigt wurden. Damals wie heute stellte sich das Problem der Komplexität und damit die Frage, wie ein solches Handbuch strukturiert werden kann, damit der Leser noch den Überblick über das Ganze behält, ohne dass die benötigte Tiefe verloren geht.

Obwohl uns von Anfang an bewusst war, dass wir nur eine eingeschränkte Auswahl aus der Gesamtheit der Erhebungsmethoden berücksichtigen können, wuchs dieser Band nach und nach von ursprünglich geplanten 60 auf 88 Beiträge an, da sich bei der Ausarbeitung der Beiträge zeigte, dass bestimmte Themen wesentlich ausgiebiger beschrieben werden mussten als ursprünglich geplant. Um die einzelnen Beiträge nicht allzu lang werden zu lassen, mussten bei vielen Beiträgen ursprünglich geplante Themen herausgenommen und als neue Einzelbeiträge bearbeitet werden. Aber auch nach dieser Aufstockung der Beiträge – womit der Band deutlich umfangreicher wurde, was diejenigen, die noch eine gedruckte Fassung dieses Handbuchs in den Händen halten, schon allein am Gewicht merken werden – bleibt auch dieses Handbuch der empirischen Sozialforschung notgedrungen (wie es vermutlich bei jeder Auswahl ist) lückenhaft.

Wir haben uns entschieden, dieses Handbuch nach den unterschiedlichen Methoden bzw. methodologischen Problemen zu sortieren. Damit weichen wir ab von der Sortierung der dritten Auflage des Handbuchs von König (deren einzelne Bände zwischen 1973 und 1978 erschienen), die überwiegend nach inhaltlichen Themen erfolgte (Baur/Blasius, Kapitel 1 in diesem Band). Aber wie damals René König strebten auch wir eine gleichgewichtige Darstellung von qualitativen und quantitativen Methoden an, was insofern ein Problem für

den Aufbau des gesamten Bandes war, als dass die Forschungs- und Darstellungslogiken beider Traditionen doch sehr unterschiedlich sind, weshalb der jetzt gewählte Aufbau der Versuch eines Kompromisses ist, von dem wir hoffen, dass er einigermaßen gelungen ist.

In dem hier vorliegenden Band wird zum einen der Stand der Forschung der qualitativen und quantitativen sozialwissenschaftlichen Methoden dargestellt, zum anderen werden grundlegende Begriffe und Techniken der empirischen Sozialforschung erläutert. Die Beiträge dieses Bandes sollen sowohl Studierenden eine Einführung in die dargestellten Verfahren der empirischen Sozialforschung geben, als auch Forschenden in Universitäten, Forschungseinrichtungen, Markt- und Sozialforschungsinstituten einen schnellen Zugriff auf notwendige Details gewähren. Den außer in Deutschland noch in einigen anderen Ländern bestehenden „Streit“, welches Paradigma der Sozialforschung – qualitativ oder quantitativ – richtig oder besser ist, können und wollen wir nicht beantworten, da die Antwort unseres Erachtens immer von der Forschungsfrage abhängt. Längst nicht alle Fragen können mit qualitativen Methoden beantwortet und längst nicht alle Fragen können mit quantitativen Methoden beantwortet werden, bei vielen Forschungsgegenständen ist auch eine Kombination von beiden der optimale Lösungsweg. Gerade im Fall der gemeinsamen Verwendungen beider Verfahren können viele Synergieeffekte genutzt werden (Baur/Blasius, Kapitel 1 in diesem Band), zudem gibt es viel mehr Ähnlichkeiten zwischen den Verfahren, als die Verfechter der einen Seite der jeweils anderen gewöhnlich zugestehen.

Da gerade in der qualitativen Forschung betont wird, dass die Methodenwahl von der Forschungsfrage, der Theorie, den Eigenheiten des Gegenstandsbereiches und der Daten abhängt, gibt es hier eine ständig wachsende Fülle von Ansätzen – der Pluralismus wird hier bewusst und aus gutem Grund zum methodologischen Prinzip gemacht, weshalb auch die Unterschiede innerhalb der qualitativen Forschung teils sogar größer sind als jene zwischen qualitativer und quantitativer Forschung. Dies wird deutlich, wenn man etwa das von Uwe Flick et al. herausgegebene „Handbuch Qualitative Sozialforschung“ (6. Auflage 2008), „The SAGE Qualitative Research Kit“ (2007), „The SAGE Handbook of Qualitative Data Analysis“ (2014) oder die Beiträge im „Forum Qualitative Sozialforschung“ (FQS, [www.qualitative-research.net](http://www.qualitative-research.net)) betrachtet. Diese Fülle an Ansätzen kann in einem Handbuch, in dem qualitative und quantitative Methoden gleichberechtigt vertreten sein sollen, nicht vollständig dargestellt werden.

Ebenso wenig können wir die gesamte Bandbreite der quantitativen Ansätze wiedergeben, hier sei auf das von Christof Wolf und Henning Best (2010, VS Verlag) herausgegebene Buch „Handbuch der sozialwissenschaftlichen Datenanalyse“ verwiesen, in dem auf 1.198 Seiten die quantitative Analyse von bereits erhobenen Daten behandelt wird. In insgesamt 40 Beiträgen beschreiben die mitwirkenden Autoren die gängigen Verfahren, die es in diesem Bereich gibt.

Auf Anregung von Frank Engelhardt, dem damaligen Lektor für Soziologie des VS Verlags, haben wir im Anschluss an die öffentliche Vorstellung des Bandes von Wolf und Best auf dem DGS-Kongress 2010 in Frankfurt ein erstes Konzept des hier vorliegenden Bandes diskutiert, in dem qualitative und quantitative Methoden gleichberechtigt dargestellt

werden sollen. Um einen optimalen Überblick zu geben, haben wir den Fokus auf die paradigmübergreifenden Prinzipien der Gestaltung des Forschungsprozesses und der Datenauswahl und vor allem auf den Bereich der Datenerhebung gelegt. Da sich sowohl in der qualitativen als auch in der quantitativen Tradition Datenerhebung und Auswertung allenfalls analytisch, aber nicht in der Forschungspraxis trennen lassen, gibt es auch Beiträge zur Auswertung der Daten.

Nach dem ersten Entwurf des Konzeptes für dieses Handbuch fragten wir im Herbst 2010 die ersten Autorinnen und Autoren, ob sie sich beteiligen möchten – nahezu alle haben diese Frage mit „ja“ beantwortet. Es sollte aber etwa dreieinhalb Jahre dauern, bis wir alle Beiträge in der finalen Version vorliegen hatten, die wir jetzt dem Verlag zum Setzen übermitteln können. In dieser Zeit wurden wir von zahlreichen Personen unterstützt, ohne deren Mithilfe das Handbuch nie fertig geworden wäre.

Zuallererst danken wir unseren Autorinnen und Autoren, die nicht nur sehr viel Geduld mit uns hatten, sondern auch auf unsere teilweise sehr zahlreichen Änderungswünsche eingegangen sind und die Beiträge immer wieder überarbeiteten. Dies war leider notwendig, da wir zum einen so wenig thematische Überschneidungen wie möglich haben wollten und zum anderen das Niveau der Beiträge auf ähnlicher Höhe sein musste: Der Band soll sowohl für BA- und MA-Studierende verständlich sein als auch gleichzeitig Nachschlagwerk für gestandene Wissenschaftlerinnen und Wissenschaftler sowie für Praktikerinnen und Praktiker in den kommerziellen Markt- und Sozialforschungsinstituten sein, die nur den einen oder anderen Hinweis auf eine Vorgehensweise benötigen.

Ein besonderer Dank geht an Silvia Arnold (Bonn), die alle Beiträge Korrektur gelesen hat, an Isabel Liebener und Michael Salwasser (Berlin), welche die Informationen über die Autoren zusammengestellt haben, sowie an unsere studentischen Hilfskräfte Tobias Drewlani, Nico Schäfer, Michael Salwasser und Josef Steilen, die die Literaturlisten und Zitationen vereinheitlicht haben. Des Weiteren danken wir Frank Engelhardt und Andreas Beierwaltes vom Verlag Springer VS, die die ganze Zeit das Projekt begleiteten und uns großzügig einen immer größer werdenden Umfang des Handbuchs zustanden. Ein weiterer besonderer Dank gilt unseren Partnern, Beate Blasius und Patrik Budenz, welche auch an den Wochenenden Gespräche und Informationen über Stichproben, Qualität der Daten und Interviewtechniken als mehr oder weniger teilnehmende Beobachter erduldeten.

Berlin und Bonn im April 2014

Nina Baur und Jörg Blasius

## Ein Überblick

Nina Baur und Jörg Blasius

---

### 1.1 Qualitative und quantitative Methoden im Forschungsprozess

Die Sozialwissenschaften gelten als Wirklichkeitswissenschaften, d.h. theoretische Aussagen und Prognosen müssen der Überprüfung an der Empirie statthalten. Ohne die „Methoden der empirischen Sozialforschung“ kann nicht empirisch geforscht werden, da diese die Regeln festschreiben, nach denen Daten erhoben, mit Theorien verknüpft und anschließend ausgewertet werden. Nicht umsonst sind daher die „Methoden der empirischen Sozialforschung“ unverzichtbarer Bestandteil der Ausbildung in vermutlich jedem sozialwissenschaftlichen Studiengang, sei es die Soziologie, die Politikwissenschaft oder die Erziehungswissenschaft. Und auch in anderen Studiengängen wie der Psychologie, der Anthropogeographie, der Ökonomie, den Kommunikations- und Planungswissenschaften gehört die empirische Sozialforschung zum Standardrepertoire der Disziplin.

Die sozialwissenschaftliche Methodenausbildung erfolgt entsprechend den Bedürfnissen des jeweiligen Studiengang und den vorhandenen personellen Ressourcen an den Universitäten. So ist sie in einigen Fächern an einigen Universitäten schon im ersten BA-Semester verpflichtend, während sie an anderen Universitäten und in anderen Fächern nur sehr rudimentär ist und nur ein oder zwei Veranstaltungen umfasst – in der Politik- und Erziehungswissenschaft z.T. nicht einmal die. Je nach Universität und je nach Fach liegt der Ausbildungsfokus mal auf den qualitativen Methoden, mal auf den quantitativen Methoden, und nur im Idealfall wird beides in Form eines breiten und gleichberechtigten Ausbildungsprogramms angeboten.

Des Weiteren hängt der Schwerpunkt und Umfang der Methodenausbildung vom jeweiligen Land ab. Im europäischen Raum nimmt derzeit Großbritannien den Spitzenplatz bei der qualitativen, die Niederlande bei der quantitativen Methodenlehre ein. Gerade in den Niederlanden haben die meisten Universitäten eine große Abteilung, die ausschließlich für die Methodenausbildung aller Fächer verantwortlich ist. Deutschland

liegt irgendwo im mittleren Bereich, an vielen Universitäten sind die Methoden aber nur durch eine einzige Professur vertreten, die in der Lehre das komplette Methodenspektrum von der Wissenschaftstheorie über die qualitativen und quantitativen Methoden bis zur qualitativen Dateninterpretation und statistischen Auswertung abdecken soll, was aber wegen der Vielfalt der Themen und dem daraus abzuleitenden Umfang in der Lehre kaum gelingen kann. Trotz aller erreichten Standards besteht daher unseres Erachtens in der deutschen Methodenforschung und -ausbildung noch viel Verbesserungsbedarf.

Ein Aspekt, bei dem wir noch besonders große Lücken sehen, ist die Vermittlung eines adäquaten Wissens um das Verhältnis von qualitativen und quantitativen Methoden. Wird die Methodendiskussion in Deutschland betrachtet, so scheint ein Gegensatz von qualitativen und quantitativen Verfahren quasi natürlich: Beide haben ihre eigene Sektion im Rahmen der *Deutschen Gesellschaft für Soziologie (DGS)*, und sowohl in der Literatur, als auch im Lehrprogramm werden sie i.d.R. getrennt behandelt. In der internationalen Fachdiskussion ist dies nicht der Fall, hier gibt es z.B. die RC33, das *Research Committee on Logic and Methodology in Sociology* ([www.rc33.org](http://www.rc33.org)), eine der größten Sektionen der *International Sociological Association (ISA)*. Eine solche institutionell gestützte gemeinsame Diskussion von qualitativen und quantitativen Methoden ist u.E. die deutlich bessere Alternative, denn primär sollte es in den Sozialwissenschaften darum gehen, Fragen zu beantworten – mit den für die jeweilige Fragestellung und den Gegenstandsbereich bestmöglich geeigneten Methoden. Und das können je nach Kontext entweder qualitative oder quantitative oder eine Kombination aus beiden Methoden sein. Daher sind quantitative und qualitative Forschung keine Gegensätze, sie sollten sich vielmehr ergänzen. Dies entspricht nicht nur dem Stand der internationalen Methodendebatte, sondern es wurde auch schon von den soziologischen Klassikern praktiziert – man denke etwa an die Studien von Max Weber, Norbert Elias oder Émile Durkheim (Baur 2005: 24–39). Ein etwas aktuelleres und in nahezu allen Bereichen der Sozialwissenschaften bekanntes Beispiel für die Kombination qualitativer und quantitativer Verfahren gibt der weltweit meistzitierte Soziologe Pierre Bourdieu in seinem Buch „Die feinen Unterschiede“ (1982). Bei diesem und seinen anderen Büchern ist es vermutlich unmöglich zu sagen, von wem er häufiger zitiert wird – von eher qualitativ oder von eher quantitativ arbeitenden Sozialwissenschaftlern.

Dieser Intention der gleichgewichteten Darstellung beider Verfahrenstypen folgt auch dieses Handbuch. Dass es sehr viele Gemeinsamkeiten zwischen qualitativen und quantitativen Methoden gibt, wird in vielen der Beiträge deutlich, angefangen gleich bei den ersten vier Kapiteln: Jede Datenerhebung weist eine enge *Verbindung zur sozialwissenschaftlichen Theorie* (Reichertz, Kapitel 2 in diesem Band) auf, denn ohne eine Theorie oder zumindest eine fundierte Fragestellung – und dies gilt für qualitative und quantitative Methoden gleichermaßen – kann weder eine Entscheidung über die Datenauswahl noch über ein geeignetes Datenerhebungsverfahren getroffen, geschweige denn das konkrete Erhebungsinstrument konstruiert werden: Was soll gefragt werden? Worauf soll der Beobachtungsfokus liegen? Welche Textpassagen sind relevant?

In jedem Forschungsprozess stellen sich auch *ethische Fragen* (Friedrichs, Kapitel 3 in diesem Band), wie z.B.: Wie vermeide ich, dass meine Forschungssubjekte durch den

Forschungsprozess oder die Publikation der Forschungsergebnisse zu Schaden kommen? Wie genau muss ich meine Forschungssubjekte darüber informieren, was ich in der Forschung mache? Darf man – wenn man etwa in der rechtsextremen Szene forscht – verdeckt beobachten? Wie gehe ich damit um, wenn ich im Feld ethisch und rechtlich zweifelhafte Praktiken sehe? Darf wirklich alles erforscht und publiziert werden, was erforscht werden kann – wie wäre z.B. mit einem potentiellen Ergebnis umzugehen, das besagt „dass eine Erhöhung des Sozialhilfesatzes dazu führt, dass die Nutznießer dieser Maßnahme dadurch davon abgehalten werden, ein gutes Arbeitsplatzangebot anzunehmen“? Zu diesen ethischen Fragen gehören auch Fragen der informationellen Selbstbestimmung (Mühlichen, Kapitel 4 in diesem Band), die durch die Entwicklung des Internets und der damit verbundenen „Social Media“, durch „Big Data“ und durch die NSA-Affäre umso virulenter geworden sind. Weiterhin eingeschlossen in diesem Fragekomplex sind die Regeln guter wissenschaftlicher Praxis, die den sorgfältigen *Umgang mit Literatur* und das Verbot von Plagiaten mit einschließen – und es muss geklärt werden, wie man verlässlich *Informationen* beschafft (Ohly/Weber, Kapitel 5 in diesem Band). Ebenso wie die Fragen zum Verhältnis von Empirie und Theorie müssen Forschende auch diese Fragen beantworten, unabhängig davon ob sie qualitativ oder quantitativ arbeiten.

Das Darstellen dieser Gemeinsamkeiten, fließender Übergänge und Synergieeffekte zwischen qualitativer und quantitativer Forschung – und dieses in möglichst vielen Beiträgen – war ein zentrales Leitmotiv bei der Konzeption dieses Handbuchs, da wir fest davon überzeugt sind, dass es das ist, was für die Forschungspraxis gebraucht wird. Dies entspricht dem internationalen Standard und sollte u.E. auch der Standard der empirischen Sozialforschung im gesamten deutschsprachigen Raum sein. Diese Forderung nach einem verbindlichen Standard entspricht auch der wissenschaftlichen und kommerziellen Forschung, wo immer stärker ein professioneller Umgang mit verschiedenen, der jeweiligen Forschungsfrage angemessenen Methoden und Verfahren erwartet wird.

Verfolgt man dieses Ziel, stellt sich die Frage, wie der Gesamtband aufgebaut sein sollte – ein Unterfangen, das in der praktischen Umsetzung nicht ganz einfach ist, weil – trotz aller Gemeinsamkeiten – die Forschungslogik (Reichert, Stein, Przyborski/Wohlrab-Sahr, Strübing, Kapitel 2, 7, 6 und 32 in diesem Band) und die typischen Darstellungstraditionen (Meyer/Meier zu Verl, Friedrichs, Kapitel 15 und 16 in diesem Band) doch sehr unterschiedlich sind, weshalb jede gemeinsame Darstellung der beiden Traditionen ein Kompromiss bleiben muss.

Dass dieses Problem unterschiedlich gelöst werden kann, zeigt der Blick auf frühere Handbücher der empirischen Sozialforschung, – das hier vorliegende Handbuch ist bei weitem nicht das erste, welches den Anspruch verfolgt, quantitative und qualitative Methoden zu verbinden. So erschien vor nunmehr über 50 Jahren das (übrigens auch heute noch sehr lesenswerte) „Handbuch der empirischen Sozialforschung“ von René König (1962) zunächst als Einzelband beim Enke Verlag in Stuttgart und bereits mit einem Umfang von 651 Seiten. In der fünf Jahre später erscheinenden Neuauflage wurden alle Texte aus dem Jahr 1962 wiederabgedruckt und am Ende des Buches um einen „Anhang“ ergänzt, „in dem über die wichtigsten Entwicklungen seit der ersten Veröffentlichung dieses

Handbuches berichtet wird“ (König 1967: v) Das gesamte Buch umfasste damit bereits 841 Seiten. Die Neuauflage, die 1973 bei dtv erschien, wurde in vier Bände aufgeteilt (die dann in fünf Büchern publiziert wurden, da Band 3 in zwei Teilbände aufgeteilt wurde) und wurde bis 1979 um weitere zehn Bände ergänzt, die thematisch orientiert aufgebaut sind. So befasst sich Band 5 mit Methoden zur Erfassung sozialer Schichtung und Mobilität, Band 6 mit der Jugend, Band 7 mit Familie und Alter, Band 8 mit Beruf, Industrie und sozialem Wandel in unterentwickelten Ländern, Band 9 mit Organisation und Militär, Band 10 mit Großstadt, Massenkommunikation und Stadt-Land-Beziehungen, Band 11 mit Freizeit und Konsum, Band 12 mit Wahlverhalten, Vorurteilen und Kriminalität, Band 13 mit Sprache und Kunst, und Band 14 mit Religion, Bildung und Medizin.

Diesem Gliederungsprinzip von König (1973-1979) sind wir nicht gefolgt: Obwohl es in einzelnen Themenfeldern der Sozialwissenschaften Präferenzen für bestimmte Methoden gibt (z.B. die größere Neigung zu standardisierten Verfahren in der Ungleichheitsforschung, zu Panelstudien in der Lebenslaufforschung oder zu Ethnografien in der Stadt- und Techniksoziologie), entspricht die von König präferierte thematische Strukturierung nicht mehr dem aktuellen Stand der Methodenforschung, die einen Methodenpluralismus propagiert, da heute in fast keinem Themenfeld *alle* Fragen mit Hilfe eines eng begrenzten Methodenkanons beantwortet werden können. Dies liegt zum einen an den großen Veränderungen des gesamten Wissenschaftsbereichs, aber vor allem an den theoretischen und methodologischen Entwicklungen in den Sozialwissenschaften in den letzten 35 bis 50 Jahren:

Die bis Mitte der 1930er Jahre in Deutschland starke qualitativ-historische Tradition brach durch den 2. Weltkrieg ab. In der Nachkriegszeit dominierten theoretisch der Parson'sche Strukturfunktionalismus und Leopold von Wieses formale Soziologie, was nicht nur eine Trennung von Theorie- und Methodenausbildung zur Folge hatte, sondern auch eine methodologische Fokussierung auf die neu entdeckten standardisierten Querschnittsbefragungen mit Hilfe von Zufallsstichproben (Baur 2005: 39-84). In den 1960ern und 1970ern entwickelten sich langsam in den verschiedensten Themen- und Theoriefeldern Alternativen, aus denen sich daran anschließend die verschiedenen Traditionen der qualitativen Sozialforschung entwickelten (Knoblauch et al. 2005).

Auch in der quantitativen Sozialforschung waren viele der heutigen Selbstverständlichkeiten noch nicht absehbar: In den 1960er Jahren wurden quantitative Daten noch mit elektrischen Zählmaschinen (Hollerith-Maschinen) ausgewertet, die erstmals 1890 zur Auswertung von amerikanischen Volkszählungsdaten verwendet wurden. Die Firma SPSS Inc. (SPSS stand damals für „Statistical Package for the Social Sciences“) wurde erst 1968 gegründet – und damals mussten Daten und Anweisungen per Lochkarte auf den Großrechner übertragen werden, die erste PC-Version erschien erst 1983. Die Firma StataCorp, die das heute vermutlich wichtigste Konkurrenzprodukt von sozialwissenschaftlicher Statistiksoftware vertreibt (STATA), wurde erst 1985 gegründet. Während man also bis Mitte der 1980er Jahre komplexe multivariate Analysen mit 1.000 und mehr Befragten nur sehr aufwändig auf den Großrechnern der damaligen Rechenzentren analysieren konnte, so spielt heute die Größe des Datensatzes praktisch keine Rolle mehr, sondern allein die

Statistik- und Methodenkenntnisse der Forschenden. So können etwa die PISA-Daten mit mehreren 100.000 Befragten inzwischen mit jedem PC ausgewertet werden. Ähnliches gilt für den Aufwand, den die Datenerhebung kostete – computergestützte Telefonumfragen (CATI) wurden zu Beginn der 1990er Jahre modern, Online-Befragungen in der zweiten Hälfte der 2000er Jahre.

---

## 1.2 **Organisation des Forschungsprozesses und zu lösende Grundprobleme**

Insgesamt verfügen die Sozialwissenschaften heute über ein sehr breites Methodenarsenal, das kontinuierlich erweitert und verbessert wird. Trotz dieser Bandbreite sind in jedem sozialwissenschaftlichen Forschungsprozess einige grundlegende Probleme zu lösen: Die Forschenden müssen eine der Fragestellung, dem Gegenstand und der Theorie angemessene Methode wählen und ein Untersuchungsdesign aufstellen, also den Gesamtfor-schungsprozess planen. Da der Gegenstand nur selten in seiner Gesamtheit erfasst werden kann, muss eine Stichprobe gezogen werden – in der Sprache der qualitativen Sozialfor-schung müssen Daten (oder ein/mehrere Datenträger) ausgewählt und ein Datenkorpus muss zusammengestellt werden. Es müssen Forschungsinstrumente entwickelt werden, die Daten müssen erhoben, aufbereitet, ausgewertet und (aus Zwecken der Nachvollzieh-barkeit und um sie für Reanalysen und Sekundäranalysen verfügbar zu machen) archiviert werden, und am Ende müssen die Ergebnisse in Form von (wissenschaftlichen) Vorträgen präsentiert und publiziert werden. Während des gesamten Prozesses sind dabei for-schungsethische Gesichtspunkte zu beachten, die von dem jeweiligen gesellschaftlichen Kontext als Anforderung an die Forschenden herangetragen werden. Der Aufbau des hier vorliegenden Handbuchs folgt dieser Forschungslogik, also den Fragen, die Forschende im Zuge des Forschungsprozesses zu lösen haben.

Da es sowohl für die qualitativen Methoden (Flick et al. 2008, Flick 2007, 2014) als auch für die quantitativen Methoden (Wolf/Best 2010) sehr gute Handbücher gibt, die sich auf die vergleichende Darstellung der Paradigmen bzw. Traditionen innerhalb der qualitativen bzw. quantitativen Sozialforschung konzentrieren, wobei der Fokus in diesen Handbüchern auf der *Datenauswertung* liegt, konzentriert sich dieses Handbuch auf die *Datenerhebung*. Dabei gilt auch heute noch die Aussage von René König, dass die Befra-gung der Königsweg der empirischen Sozialforschung sei, insofern, als dass sie die in der Forschungspraxis am häufigsten angewandte Methode ist. Ob das immer sinnvoll ist, sei dahingestellt – jedenfalls findet hier mit Abstand die meiste Methodenforschung statt, und zwar sowohl in der qualitativen als auch in der quantitativen Tradition. Entsprechend sind die Teile 4 und 5 dieses Handbuchs der *offenen Befragung* und der *standardisierten Befra-gung* gewidmet. Da die Methoden der empirischen Sozialforschung aber weitaus mehr Datenformen zu bieten haben als Interviews, werden im sechsten Teil *weitere Datentypen* vorgestellt.

Wie man dem Inhaltsverzeichnis dieses Handbuchs entnehmen kann, liegt zwar ein Fokus auf der Datenerhebung – aber es gibt auch zahlreiche Kapitel zu anderen Phasen des Forschungsprozesses. Der Grund ist, dass sich die Datenerhebung nicht von den übrigen Forschungsphasen trennen lässt. Dabei gelten insbesondere bei den *Grundlagen der empirischen Sozialforschung* (Teil 1) sowie bei den *Stichproben, der Datenaufbereitung und der Güte* (Teil 2) die meisten Prinzipien für die qualitative und quantitative Forschung gleichermaßen, weshalb wir hier in der Gliederung nicht zwischen qualitativen und quantitativen Methoden trennen.

Schwieriger gestaltet sich eine gemeinsame Darstellung qualitativer und quantitativer Methoden bei dem Thema „Auswertung“, da diese mit einigen ungelösten und möglicherweise nie zu lösenden Grundproblemen sozialwissenschaftlicher Forschung verbunden ist. Diese Probleme sind – wie Reichertz (Kapitel 2 in diesem Band) argumentiert – zwar allen Forschungsparadigmen gemeinsam, sie werden in den beiden Forschungstraditionen doch sehr unterschiedlich gelöst.

Ein erstes zentrales Problem ist der Umgang mit der *Subjektivität der Forschenden*: Die Forschenden sind selbst auch Teil jener Gesellschaft, die sie untersuchen. Aufgrund dessen können sie dazu neigen, aus einer spezifischen Perspektive zu argumentieren, bestimmte Aspekte zu übersehen und ihre eigenen Vorurteile in den Forschungsprozess zu tragen. Ein zweites Problem sind die Eigenheiten des Untersuchungsgegenstands; da sich die *Gesellschaft über die Zeit wandelt*, verändern sich sowohl der Forschungsgegenstand selbst, als auch die Fragen, die an ihn gestellt werden (Baur 2005). Nicht umsonst gehören die Sozialwissenschaften zu den ewig „jugendlichen Wissenschaften“ (Weber 1904: 206) – und damit geht auch immer in Frage einher, ob man dieselben Methoden wie früher anwenden kann oder ob neue Methoden entwickelt bzw. vorhandene Methoden weiterentwickelt werden müssen. Die qualitative und quantitative Forschung gehen mit diesen Problemen unterschiedlich um.

Die *quantitative Forschung* versucht, die Subjektivität der Forschenden möglichst weitergehend auszuschalten, indem sie den Forschungsprozess linear organisiert, die theoretische Reflexion an den Anfang des Forschungsprozesses stellt und dann den Forschungsprozess in einzelne Arbeitsschritte aufteilt. Dabei sind die einzelnen Phasen des Forschungsprozesses aufeinander bezogen: So erfordern viele multivariate Auswertungsverfahren bestimmte Strukturen von Daten (z.B. dichotom oder ordinal skaliert) – wenn dies nicht zu Beginn des Forschungsprozesses, z.B. bei der Konstruktion des Fragebogens, berücksichtigt wird, dann können die entsprechenden Forschungsfragen später nicht beantwortet werden. Dennoch versucht die quantitative Forschung die einzelnen Forschungsphasen von der theoretische Reflexion über den Forschungsgegenstand, über die Instrumentenentwicklung, die Stichprobeneziehung, Datenerhebung, Datenaufbereitung und abschließend die Datenauswertung soweit wie möglich nacheinander abzuarbeiten. Jeder dieser Arbeitsschritte wird für sich optimiert und soll möglichst so gestaltet werden, dass er unabhängig von der Person der Forschenden ist, die ganze Untersuchung soll so objektiv wie möglich sein. Dabei bedeutet Objektivität hier, dass die Untersuchung unabhängig von der Person der Forschenden die gleichen Ergebnisse liefert und dass darauf

zu achten ist, dass immer der aktuelle Forschungsstand gewahrt bleibt. Verändert sich der Gegenstand der Forschung bzw. werden neue Fragen entwickelt, dann müssen für die einzelnen Arbeitsschritte (meistens) neue Methoden entwickelt werden. So werden z.B. seit der Verbreitung des Internets die Methoden der Online-Befragung (weiter) entwickelt (Wagner/Hering, Kapitel 48 in diesem Band), z.B. wie Online-Befragungen grafisch gestaltet werden sollten und wie repräsentative Stichproben gezogen werden können. Sind die Daten erhoben und liegen sie aufbereitet vor, z.B. als SPSS-File, so werden sie mit denselben statistischen Methoden ausgewertet wie jene, die mit anderen Modi der Erhebung gesammelt wurden. Dieser Logik der einzelnen Forschungsphasen folgend, werden im siebten Teil dieses Handbuchs, *Datenauswertung in der quantitativen Sozialforschung*, die wichtigsten Methoden und Probleme erläutert.

Die *qualitative Forschung* geht den umgekehrten Weg und setzt bei der Erkenntnis an, dass eine vollkommene Ausschaltung der Subjektivität der Forschenden nicht möglich ist und sich Gegenstände permanent verändern. Weiterhin betont sie, dass die Subjektivität der Forschenden auch eine wichtige Forschungsressource ist – ohne ein Grundverständnis von einer Gesellschaft zu haben, ist es (im wissenssoziologischen Sinne) gar nicht möglich zu verstehen (Kurt/Herbrik, Kapitel 33 in diesem Band), was in eben dieser Gesellschaft passiert. Daher wird betont, dass das Forschungsinstrument „Forschende(r)“ trainiert werden muss. Objektivität heißt hier, dass die Forschenden ihr methodologisches Vorgehen reflektieren und nachvollziehbar offenlegen. Statt zu versuchen, die Subjektivität der Forschenden auszuschalten, wird (zumindest in der sogenannten interpretativen Variante der qualitativen Forschung) der Forschungsprozess zirkulär gestaltet, d.h. Phasen der Datenauswahl, Datenerhebung und Datenanalyse wechseln sich iterativ mit Phasen der theoretischen Reflexion ab, um den Forschungsprozess möglichst effizient zu gestalten. Die Methoden werden dabei flexibel an die jeweiligen Erfordernisse von Theorie, Fragestellung und des Gegenstandes angepasst (Przyborski/Wohlrab-Sahr, Strübing, Kapitel 6 und 32 in diesem Band). Allerdings gibt es auch Ausnahmen, so z.B. die qualitative Inhaltsanalyse (Mayring/Fenzl, Kapitel 38 in diesem Band), die hinsichtlich ihrer Forschungslogik eher dem quantitativen Vorgehen folgt. Methodenentwicklung bedeutet in der qualitativen Sozialforschung ansonsten meistens die Entwicklung neuer Methoden für neue Gegenstände. Das Problem hierbei ist, dass es erheblicher Erfahrung bedarf, um den Wechsel zwischen Theorie und Empirie sachgemäß durchzuführen. Da in der qualitativen Forschung Datenerhebung und Datenauswertung oft zirkulär organisiert und miteinander verwoben sind, wobei die spezifische Art, wie dies geschieht, von der jeweiligen qualitativen Tradition abhängt, stellt dieses Handbuch im dritten Teil die *Forschungsparadigmen in der qualitativen Sozialforschung* gemeinsam mit den Prinzipien der Auswertung anhand einiger ausgewählter Traditionen exemplarisch vor.

Unabhängig davon, ob qualitativ oder quantitativ gearbeitet wird, ist die Forschungserfahrung ein zentraler Parameter für den Erfolg der Studie, wobei kontinuierliche Selbstreflexion, Prozess- und Qualitätskontrolle eine Selbstverständlichkeit in jedem Forschungsprozess sein sollten. Daher ersetzt auch dieses Handbuch *nicht* – das sei hier ausdrücklich gesagt – das gemeinsame Arbeiten in einem Forschungsteam, in dem erfahrene Personen

vertreten sein sollten. Die Grundlagen der empirischen Sozialforschung können zwar aufgeschrieben und die Inhalte können angelesen werden, die Forschungspraxis kann aber nur über die Anwendung des Gelernten im Feld kommen.

---

### 1.3 Grundlagen der empirischen Sozialforschung

Diesen grundsätzlichen Überlegungen folgend, ist der erste Teil dieses Handbuchs der Frage gewidmet, wie sozialtheoretische Hintergrundannahmen, Gegenstandsbereich und Forschungsfrage miteinander und mit dem Forschungsprozess verwoben sind. Hierzu gehört insbesondere die Explikation der Fragestellung sowie die Klärung, welches *Theorie-Empirie-Verständnis* man für das geplante Projekt zugrunde legt, da dies den weiteren Forschungsablauf maßgeblich bestimmt (Reichertz, Kapitel 2 in diesem Band).

Bevor die Datenerhebung beginnen kann, sind, wie bereits erwähnt, oft *forschungsethische Fragen* (Friedrichs, Kapitel 3 in diesem Band) zu stellen, z.B. die Frage, ob und inwie weit Drogenabhängige in einem entsprechenden Projekt überhaupt befragt werden sollten, da zumindest ein Teil der Antworten vermutlich strafrechtlich relevant ist. Mit diesen Fragen verbunden sind *Fragen des Datenschutzes und der informationellen Selbstbestimmung* (Mühlichen, Kapitel 4 in diesem Band). Es ist etwa zu klären, wie die Anonymität der Personen gewahrt bleiben kann, die mit Konsequenzen rechnen müssen, oder diese auch nur vage befürchten. Parallel dazu muss geprüft werden, auf welche zusätzlichen *Informationsquellen* man mit welchen Methoden zurückgreifen kann, um das geplante Projekt erfolgreich beginnen und beenden zu können. Dazu gehören u.a. die Kenntnis von Literaturdatenbanken und Fundorte für benötigte Daten. Ohly und Weber (Kapitel 5 in diesem Band) geben dazu viele praktische Tipps und benennen die wichtigsten Quellen.

Auf der Basis einer sehr guten Kenntnis der aktuellen Literaturlage, bereits vorhandener Daten, einer Theorie und der Fragestellung ist das *Forschungsdesign* zu entwickeln. Damit sollte der Ablauf des Forschungsprozesses so genau wie möglich geplant und festgelegt werden. Des Weiteren sollten in diesem Rahmen alle wesentlichen Entscheidungen bezüglich der Datenauswahl, der Datenerhebung und der Auswertung getroffen werden. Die entsprechenden Forschungsdesigns können sich je nach Forschungsansatz grundlegend voreinander unterscheiden. In diesem Band unterscheiden wir zwischen *Forschungsdesigns für die qualitative Sozialforschung* (Przyborski/Wohlrab-Sahr, Kapitel 6), für die *quantitative Sozialforschung* (Stein, Kapitel 7), für *Mixed Methods Designs* (Kelle, Kapitel 8), für die *Evaluationsforschung* (Döring, Kapitel 9), für die *Marktforschung* (Ziegler, Kapitel 10), für *Experimente* (Eifler, Kapitel 11) sowie für *Simulationsstudien* (Jann/Hevenstone, Kapitel 12).

Da die Datenerhebung meistens sehr aufwändig und teuer ist, greifen Sozialforscherinnen und Sozialforscher zunehmend auf vorhandene Datenbestände zurück, um diese einer *Sekundäranalyse* zu unterziehen. Für die Analyse von sozialen Prozessen ist dies teilweise sogar erforderlich, weil sich diese erst über Zeit entfalten (Baur 2005). So können einige der Daten, die bereits vor vielen Jahren oder gar vor mehreren Jahrzehnten erhoben wurden,

erst heute in ihrem vollen Umfang ausgewertet werden. Des Weiteren ist oft auch im Sinne der Nachvollziehbarkeit und Überprüfung von früheren Forschungsergebnissen eine Re-Analyse von Daten wünschenswert. Im Vergleich zu anderen Ländern ist in Deutschland die hierfür benötigte Dateninfrastruktur relativ gut, d.h. es stehen relativ viele Daten für Re- und Sekundäranalysen zur Verfügung. Einen Überblick über Möglichkeiten und Fundorte für Daten für *Sekundäranalysen in der qualitativen Forschung* gibt Medjedović (Kapitel 13 in diesem Band), einen für die *quantitative Forschung* gibt Mochmann (Kapitel 14 in diesem Band). Weitere Hinweise zu möglichen Fundorten für Daten werden insbesondere in den Kapiteln zu *weiteren Datentypen* (Teil 6 dieses Handbuchs) gegeben. Die letzten beiden Kapitel dieses Abschnittes sind den Konventionen der *Ergebnispräsentation in der qualitativen Forschung* (Meyer/Meier zu Verl, Kapitel 15 in diesem Band) und in der *quantitativen Forschung* (Friedrichs, Kapitel 16 in diesem Band) gewidmet.

---

## 1.4 Stichproben, Datenaufbereitung und Güte

Obwohl es zweifelsohne Unterschiede zwischen den Paradigmen der qualitativen und der quantitativen Sozialforschung gibt, glauben wir, dass sich die beiden Paradigmen in der Forschungspraxis deutlich ähnlicher sind als es oft behauptet wird. Dies zeigt sich u.a. darin, dass trotz all der real vorhandenen Unterschiede viele grundsätzliche Probleme vorhanden sind, die in jedem Forschungsprojekt gelöst werden müssen. Hierzu gehören die Auswahl und Beschreibung der Stichprobe sowie der dazu gehörigen Grundgesamtheit, die Datenaufbereitung sowie die Güte der Daten und die Art und Weise, wie diese beschrieben werden kann.

In der qualitativen Forschung existieren relativ unterschiedliche, theoriegeleitete Strategien der *Datenauswahl und der Stichprobenziehung*. Dabei müssen auf den unterschiedlichen Ebenen Entscheidungen getroffen werden: Der genaue Untersuchungsgegenstand muss definiert und abgegrenzt werden und es muss bestimmt werden, wie Daten ausgewählt werden und ein Datenkorpus zusammengestellt wird – und wie innerhalb dieses Datenkorpus die Analyseeinheiten definiert werden (Akremi, Kapitel 17 in diesem Band). Eine Variante der bewussten Auswahl ist das *Theoretical Sampling* (Strübing, Kapitel 32 in diesem Band), das von der Grounded Theory entwickelt und von vielen anderen qualitativen Forschungsparadigmen übernommen wurde. Sowohl bei der bewussten Auswahl als auch beim Theoretical Sampling ist das Ziel eine theoriegesteuerte Verallgemeinerung bzw. eine Übertragbarkeit auf andere Kontexte. Die quantitative Forschung ist dagegen am Ideal der *Zufallsstichprobe* orientiert (Häder/Häder, Kapitel 18 in diesem Band) und hat das Ziel der Verallgemeinbarkeit auf die zuvor definierte Grundgesamtheit mit Hilfe (zumeist) multivariater statistischer Verfahren (Blasius/Baur, Kapitel 79 in diesem Band).

Da die quantitative Forschung stärker als die qualitative Forschung einer linearen Logik folgt, müssen alle Forschungsinstrumente *vor* der Datenerhebung festgelegt werden, weil möglicherweise auftretende Probleme im Lauf des Forschungsprozesses wesentlich schwerer (und oft auch gar nicht) korrigiert werden können, und dies ist dann auch noch

mit deutlich höheren Kosten verbunden. Wurde z.B. in einer standardisierten Befragung eine wichtige Frage vergessen oder eine fehlerhafte Frage gestellt, dann kann dieser Fehler nicht im Verlauf der Studie korrigiert werden; bei Face-to-Face Befragungen sind die Fragebögen dann schon gedruckt und an die Interviewer verteilt, und schon aus Kostengründen kann die Umfrage kein zweites Mal durchgeführt werden. Deshalb sollten die Erhebungsinstrumente (der Fragebogen) insbesondere in der quantitativen Forschung im Rahmen des *Pretests* (Weichbold, Kapitel 19 in diesem Band) vor der eigentlichen Untersuchung sehr sorgfältig getestet werden, hier können etwaige Mängel in der Frageformulierung bemerkt und korrigiert werden. In der qualitativen Forschung kann zwar ebenfalls eine Voruntersuchung durchgeführt werden, oft wird diese aber in den Forschungsprozess integriert, d.h. Probleme bei der Datenerhebung werden in späteren Phasen des Forschungsprozesses korrigiert, z.B. indem der Stichprobenplan modifiziert wird, Fragen in einem Leitfaden ergänzt oder alternative Beobachtungssituationen gewählt werden. Dies ist insbesondere bei neu zu erarbeitenden Gegenstandsbereichen von Vorteil, weil sie eine sukzessive Annäherung an das Forschungsfeld ermöglicht.

Bei Befragungen kann die *Einstellung der Befragten zu Umfragen* (Hlawatsch/Krickl, Kapitel 20 in diesem Band) einen Einfluss auf die Antwortbereitschaft und insbesondere auf die Sorgfalt der Antwort haben – und damit einen Einfluss auf das Ergebnis der Untersuchung. Ebenso wie es Effekte von Seiten der Interviewten auf die Antwort gibt, muss immer mit *Interviewereffekten* (Glantz/Michael, Kapitel 21 in diesem Band) gerechnet werden. Die Motivation der Interviewerinnen und Interviewer hat sowohl einen Einfluss darauf, ob das Interview überhaupt zustande kommt, als auch auf die Qualität der Interviews (Blasius/Thiessen 2012). Dieses Problem dürfte die quantitative Sozialforschung stärker betreffen – bei qualitativen Interviews können geübte Interviewer Probleme im Antwortverhalten der Befragten schon während des Interviews geschickt umgehen, und eine schlechte Interviewführung erkennt man i.d.R. sofort am Transkript. In der Literatur selten diskutiert, aber anhand von neuerer Forschung belegt, ist, dass *Fälschungen von Interviews* (Blasius, Kapitel 22 in diesem Band) wesentlich öfter vorkommen als bisher angenommen.

Ein weiteres Problem, mit dem die quantitative Sozialforschung seit einigen Jahren kämpft, sind sinkende Ausschöpfungsquoten, d.h. dass immer weniger Personen bereit sind, an Befragungen teilzunehmen. Des Weiteren beantworten die Befragten oft nicht alle Fragen, bei Fragen wie etwa zur Parteipräferenz oder zum Einkommen gibt es meistens eine hohe Nonresponse-Rate. Beide *Nonresponse-Probleme* (Engel/Schmidt, Kapitel 23 in diesem Band) – zum einen die generelle Verweigerung an einer Befragung teilzunehmen und zum anderen die Antwortverweigerung bei bestimmten Fragen – können zu systematischen Verzerrungen führen. Eine Möglichkeit, diese Verzerrungen zu korrigieren, besteht darin, die vorhandenen *Daten zu gewichten* (Kiesl, Kapitel 24 in diesem Band), damit sie als repräsentativ für die zuvor definierte Grundgesamtheit gelten können. Im Umkehrschluss heißt dies: Sind die Daten nicht repräsentativ, dann können, mit welchen statistischen Verfahren auch immer, keine allgemein gültigen Aussagen über die Grundgesamtheit gemacht werden – genau dies ist aber meistens das Ziel der quantitativen Sozialforschung. Bei den für die qualitative Forschung typischen bewussten Auswahlen

ist dieses Problem allenfalls sekundär, z.B. da potentielle Interviewpartner, die sich nicht befragen lassen möchten bzw. die viele Fragen nicht beantworten, durch strukturähnliche Fälle ersetzt werden können (Akremi, Kapitel 17 in diesem Band).

Ein relativ neuer Forschungszweig innerhalb der empirischen Sozialforschung sind *Paradaten* (Felderer et al., Kapitel 25 in diesem Band). Dabei handelt es sich um Daten, mit denen der Forschungsprozess dokumentiert werden kann. Bei Face-to-Face-Erhebungen zählen dazu u.a. die Anzahl der vergeblichen Kontakte und die Uhrzeiten, zu denen die Zielperson (vergeblich) aufgesucht wurde. In der quantitativen Forschung werden in den letzten Jahren außerdem verstärkt Daten aus verschiedenen Stichproben bzw. Datensätzen mittels *Data Fusion*, *Record Linkage* und *Data Mining* miteinander verknüpft (Cielebak/Rässler, Kapitel 26 in diesem Band), um damit das potentielle Auswertungsspektrum zu erweitern. Wurden in einer großen Studie z.B. viele Daten zu den politischen Einstellungen erhoben und in einer anderen ähnlich großen Studie viele Daten über das Nutzen von Medien, so können beide Studien über Befragte, die als *statistische Zwillinge* angesehen werden können (d.h. jeweils zwei Personen haben die gleichen sozio-demografischen Merkmale, gleiche Einstellungen, die in beiden Studien abgefragt wurden, usw.), miteinander verbunden werden. War das Matching erfolgreich, so kann auf der individuellen Ebene der Zusammenhang zwischen Mediennutzung und politischer Einstellung untersucht werden, obwohl die entsprechenden Fragen unterschiedlichen Personen gestellt wurden.

Unabhängig davon, wie sorgfältig die Daten erhoben wurden, müssen sie sowohl in der qualitativen (Kuckartz/Rädiker, Kapitel 27 in diesem Band), als auch in der quantitativen Forschung (Lück/Landrock, Kapitel 28 in diesem Band) vor der Auswertung sorgfältig *aufbereitet* und *bereinigt* werden. So müssen qualitative Interviews transkribiert, und Fragebögen aus standardisierten Befragungen müssen derart in einem Datensatz abgelegt werden, dass sie mit Hilfe eines Programmpaketes wie SPSS oder STATA statistisch auswertbar sind. Obwohl die Datenaufbereitung in der Forschungspraxis mit am meisten Zeit in Anspruch nimmt, wird dieses Thema in den meisten Einführungen in die empirische Sozialforschung vernachlässigt – und dies, obwohl gerade bei der Aufbereitung der Daten meistens viele Entscheidungen getroffen werden müssen, die einen unmittelbaren Einfluss auf die späteren Ergebnisse haben.

Um die *Qualität der Forschung* beurteilen zu können, gehört zu jedem Forschungsprojekt eine Bestimmung der Güte des Forschungsprozesses. Dabei kann z.B. bestimmt werden, wie stark die tatsächliche Messung von den wahren Werten abweicht, d.h. wie nahe die gegebenen Antworten der Befragten an deren tatsächlich vorhandenen Einstellungen liegen. Für diese Bestimmung gibt es *Gütekriterien*, wobei aufgrund der unterschiedlichen Forschungslogiken für die qualitative Forschung (Flick, Kapitel 29 in diesem Band) und für die quantitative Forschung (Krebs/Menold, Kapitel 30 in diesem Band) unterschiedliche Kriterien gelten. Insbesondere für die quantitative Forschung ist zu dabei zu beachten, dass die verschiedenen möglichen Fehlerarten (z.B. Stichprobenfehler, Messfehler, Fehler bei der Datenaufbereitung) aufeinander bezogen und voneinander abhängig sind; diese können zu einem Gesamtfehler zusammengefasst werden, dem *Total Survey Error (TSE)* (Faulbaum, Kapitel 31 in diesem Band).

## 1.5 Forschungsparadigmen in der qualitativen Sozialforschung

In der qualitativen Forschung iterieren Datenauswahl, -erhebung und -analyse mit Phasen der theoretischen Reflexion. Sozialtheoretische Grundannahmen, Forschungsparadigma und die Gesamtgestaltung des Forschungsprozesses sind so miteinander verwoben, dass sie immer in ihrer Gesamtheit betrachtet werden sollten. Um diesen Wechselbezügen gerecht zu werden, sind die hier publizierten Beiträge für die qualitative Sozialforschung relativ lang, da sowohl die theoretischen Grundannahmen als auch die Gestaltung des Forschungsprozesses sowie die der Auswertung – und, soweit für das Verständnis erforderlich, paradigmenspezifische Kriterien der Gestaltung der Datenauswahl und Datenerhebung sowie der Güte – dargestellt werden müssen. Es sei nochmals darauf hingewiesen, dass qualitative Paradigmen nicht nur ständig weiterentwickelt werden, sondern auch, dass mehr als fünfzig verschiedene Verfahren unterschieden werden können. Es ist daher nicht möglich, in einem Handbuch, in dem die grundlegenden qualitativen *und* quantitativen Methoden diskutiert werden, einen vollständigen Überblick zu geben – für weitere Verfahren verweisen wir auf Flick et al. (2008) und Flick (2007, 2014). Bei der für dieses Handbuch getroffenen Auswahl soll die Bandbreite dessen, was qualitative Sozialforschung sein kann und sein will, exemplarisch dargestellt werden. Die ausgewählten Paradigmen lassen dabei verschiedene *Vergleichsdimensionen* zu:

- Es gibt qualitative Methoden, die relativ stark der quantitativen, sequenziellen Logik folgen, z.B. die *qualitative Inhaltsanalyse* (Mayring/Fenzl, Kapitel 38 in diesem Band), und solche, die diese Logik in der Zirkularität auflösen, z.B. die *Grounded Theory* (Strübing, Kapitel 32 in diesem Band).
- Bei manchen Ansätzen wird der Gegenstand in einzelne Teilfragen zerlegt, Beispiele sind die *qualitative Inhaltsanalyse* (Mayring/Fenzl, Kapitel 38 in diesem Band) und die *Einzelfallanalyse* (Hering/Schmidt, Kapitel 37 in diesem Band). Dagegen wird mit anderen Verfahren ein ganzheitlicher Ansatz verfolgt, z.B. in der *Ethnografie* (Knoblauch, Kapitel 36 in diesem Band).
- Bei manchen Verfahren werden nur sehr wenige Fälle, im Extremfall nur ein einziger analysiert, z.B. in der *sozialwissenschaftlichen Hermeneutik*, der *hermeneutischen Wissenssoziologie* (Kurt/Herbrik, Kapitel 33 in diesem Band) und der *Einzelfallanalyse* (Hering/Schmidt, Kapitel 37, in diesem Band). Bei anderen Verfahren wird z.T. sehr viel Material verarbeitet, z.B. bei der *qualitativen Inhaltsanalyse* (Mayring/Fenzl, Kapitel 38 in diesem Band) oder der *Ethnografie* (Knoblauch, Kapitel 36 in diesem Band).
- Es gibt eine Reihe von Verfahren, mit denen explizit die Analyse sozialen Wandels und sozialer Prozesse beschrieben wird, sei es auf der Ebene des einzelnen Individuums wie bei der *Biographieforschung* (Rosenthal, Kapitel 35 in diesem Band), sei es auf der Ebene von Diskursen, wie bei der *Diskursanalyse* (Traue et al., Kapitel 34 in diesem Band), oder sei es auf der Ebene von Organisationen, Städten oder Nationen wie bei der *Einzelfallanalyse* (Hering/Schmidt, Kapitel 37 in diesem Band).
- Mit den unterschiedlichen qualitativen Paradigmen sind auch unterschiedliche Datensorten verbunden. In der *Ethnografie* (Knoblauch, Kapitel 36 in diesem Band) handelt

es sich oft um eine Kombination aus schwach strukturierter Beobachtung (Thierbach/Petschick, Kapitel 66 in diesem Band) und anderen Daten. In der *Biographieforschung* (Rosenthal, Kapitel 35 in diesem Band) werden meistens narrative Interviews (Küsters, Kapitel 40 in diesem Band) verwendet. Des Weiteren ist zu beachten, dass dieselben Datenquellen je nach Forschungsparadigma sehr unterschiedlich interpretiert werden (können).

---

## 1.6 Offene Befragung

So sehr sich die meisten qualitativen Verfahren unterschiedlicher Datentypen bedienen können, so sehr dominiert in der Forschungspraxis die offene Befragung – entweder als Hauptinformationsquelle oder als Ergänzung zu anderen Daten. Die offene Befragung tritt in zahlreichen verschiedenen Varianten auf, von denen hier einige exemplarisch herausgegriffen wurden.

Eine sehr häufig vorkommende Variante sind *Leitfaden- und Experten-Interviews* (Helfferich, Kapitel 39 in diesem Band). Eine überwiegend in der nichtwissenschaftlichen Praxis vorkommende und mit ihr eng verwandte Variante ist das *journalistische Interview* (Friedrichs/Schwinges, Kapitel 43 in diesem Band). Bei diesen Befragungsformen wird vor dem Interview eine Frageliste erstellt, die zwar während des Interviewverlaufs um weitere Fragen ergänzt werden kann, bei der aber versucht wird, alle Fragen im Lauf des Interviews auch abzuarbeiten. Der Hauptunterschied zwischen Leitfaden- und Experteninterviews ist dabei, welcher Status den Interviewpartnern zugewiesen wird: Beim Leitfadeninterview interessieren die persönliche Perspektive der Interviewten und ihre Erfahrungen mit dem zu untersuchenden Phänomen. Dem entgegen wird von Experten erwartet, dass sie einen möglichst neutralen und breiten Blick auf das Geschehen haben. Ob eine Person nun Betroffenen- oder Expertenstatus einnimmt, hängt somit auch von der Einschätzung der Forschenden hinsichtlich deren Neutralität und Wissen ab – was aber weniger für die Interviewführung, sondern stärker für die Behandlung der im Interview vermittelten Informationen während der Auswertung relevant ist. *DELPHI-Befragungen* (Häder/Häder, Kapitel 42 in diesem Band) sind eine Variante der Expertenbefragung, in denen offene mit standardisierten Elementen kombiniert werden.

Eine andere Variante der offenen Befragung ist das *narrative Interview* (Küsters, Kapitel 40 in diesem Band), das im Rahmen der Forschungstradition der Biographieforschung (Rosenthal, Kapitel 35 in diesem Band) entwickelt wurde und besonders gut geeignet ist, soziale Prozesse und Biographien zu erfassen, da das narrative Interview auf der Idee basiert, dass einmal begonnene Geschichten zu Ende erzählt werden müssen, wobei gezielt Techniken eingesetzt werden, um die Befragten zum Erzählen zu bringen. Bei *Gruppendiskussion und Fokusgruppen* (Vogl, Kapitel 41 in diesem Band) werden – anders als bei den eben genannten Befragungsformen – Gruppen von Personen gefragt, was besonders gut geeignet ist, wenn Diskussionen zu einem Thema provoziert oder Gruppenmeinungen erfasst werden sollen.

## 1.7 Standardisierte Befragung

Ebenso wie in der qualitativen Sozialforschung ist in der quantitativen Sozialforschung die Befragung nach wie vor die mit Abstand häufigste Datenerhebungsmethode. Aufgrund der Vielfalt der Möglichkeiten des Einsatzes der standardisierten Befragung ist es nicht verwunderlich, dass sich die *Survey Methodology* mittlerweile zu einem eigenen Strang der Methodenforschung entwickelt hat, in dem zahlreiche Einzelaspekte thematisiert werden, die bei der Durchführung einer Umfrage zu beachten sind. Entsprechend ist diesem Thema ein eigener Bereich dieses Handbuchs gewidmet, in das Reinecke (in diesem Band, Kapitel 44) mit den *Grundlagen der standardisierten Befragung* einführt.

Bei allen Formen der standardisierten Befragung müssen die Art der Erhebung (Face-to-Face, telefonisch, usw.) und die *Ziehung der Stichprobe* (Häder/Häder, Kapitel 18 in diesem Band) aufeinander abgestimmt sein. Als wichtigste *Befragungsmodi* sind die *persönlich-mündlichen Befragungen* (auch als Face-to-Face-Befragungen bezeichnet, Stocké, Kapitel 45 in diesem Band), die *telefonische Befragung* (Hüfken, Kapitel 46 in diesem Band), die *schriftlich-postalische Befragung* (Reuband, Kapitel 47 in diesem Band) und die *Online-Befragung* (Wagner/Hering, Kapitel 48 in diesem Band) zu nennen. Insbesondere wegen der zunehmenden Verweigerungsraten (Engel/Schmidt, Kapitel 23 in diesem Band) werden immer öfter verschiedene Modi zu sogenannten „Mixed Mode-Designs“ kombiniert, d.h. innerhalb eines einzelnen Projektes werden verschiedene Modi der Datenerhebung verwendet (Reuband/Blasius 1996).

Zentral für alle Umfragen sind der *Aufbau und die Gestaltung des Fragebogens* (Klöckner/Friedrichs, Kapitel 49 in diesem Band): Sowohl die Interviewerinnen und Interviewer (bei der Face-to-Face- und telefonischen Befragung), als auch die Befragten (bei selbst-administrativen Befragungen wie schriftlich-postalisch und Online-Befragungen) müssen das Instrument „Fragebogen“ ohne Probleme „fehlerfrei“ bedienen können. Des Weiteren muss die Reihenfolge der Fragen beachtet werden. So sollten sensible Fragen, wie z.B. nach dem Einkommen oder der Parteipräferenz, erst gegen Ende der Befragung gestellt werden.

Aber auch bei der *Formulierung der einzelnen Fragen* (Porst, Kapitel 50 in diesem Band) und der *Antwortvorgaben* (Franzen, Kapitel 51 in diesem Band) gilt es, eine Vielzahl von Regeln zu beachten, angefangen von der genauen Formulierung der einzelnen Fragen bis hin zu der Frage, ob die Kategorie „keine Angabe“ explizit berücksichtigt werden soll. Wie die Frage formuliert werden muss, hängt dabei auch vom Frageformat ab. So werden nicht nur bei qualitativen Interviews, sondern auch bei standardisierten Befragungen oft *offene Fragen* (Züll/Menold, Kapitel 52 in diesem Band) verwendet, auf welche die Befragten frei antworten können. Eng an das experimentelle Design geknüpft (Eifler, Kapitel 11 in diesem Band) ist die Verwendung von *Vignetten* (Dülmer, Kapitel 53 in diesem Band), das sind variierende Personen- bzw. Situationsbeschreibungen, die von den Befragten unter verschiedenen Gesichtspunkten beurteilt werden sollen. Bestandteil einer jeden Befragung ist die Erhebung der *sozio-demografischen Merkmale* der Befragten, so z.B. nach dem Alter und dem Einkommen (Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band). Die Abfrage derartiger Indikatoren ist längst nicht immer so einfach, wie man sich das

vorstellen könnte, so z.B. die Frage, wieviel jemand mit welchen Einnahmequellen (Lohn, Kindergeld, Wohngeld, Unterhalt usw.) verdient. Um später *Skalen* und *Indizes* (Blasius, Kapitel 83 in diesem Band) bilden zu können, müssen Fragen gestellt werden, die einen gemeinsamen Bezug zu der zu *konstruierenden Skala* bzw. dem zu *konstruierenden Index* haben (Latcheva/Davidov, Kapitel 55 in diesem Band), so z.B. um Anomie oder Politikverdrossenheit zu messen. Es ist ersichtlich, dass eine direkte Frage nach einem Konstrukt wie Anomie wenig sinnvoll ist.

Quer hierzu müssen die Besonderheiten bei der *Befragung bestimmter Kulturreiche oder Bevölkerungsgruppen* (oder spezieller Populationen) beachtet werden. Bei den zunehmend häufiger durchgeführten *interkulturell vergleichenden Umfragen* (Braun, Kapitel 56 in diesem Band) ist etwa zu beachten, welche Fragen identisch sind, welche funktional äquivalent sind und welche in welchen Ländern gar nicht gestellt werden sollten. Bei der Befragung von speziellen Bevölkerungsgruppen bzw. bei speziellen Themen oder speziellen Befragungsformaten kann es erforderlich sein, spezielle Maßnahmen zum Erhalt des Datenschutzes oder auch zum Schutz der eigenen Meinung zu ergreifen. Derartige Maßnahmen sind z.B. bei *Mitarbeiterbefragungen* (Borg, Kapitel 57 in diesem Band), bei *Befragungen von Kindern und Jugendlichen* (Heinen/König, Kapitel 58 in diesem Band), von *älteren und alten Menschen* (Motel-Klingebiel et al., Kapitel 59 in diesem Band), von *Migranten* (El-Menouar, Kapitel 60 in diesem Band) sowie bei anderen *speziellen Populationen* wie Eliten und Menschen mit längeren Auslandsaufenthalten (Schütte/Schmies, Kapitel 61 in diesem Band) erforderlich. Hier zeigt sich, dass nicht nur die qualitative, sondern auch die quantitative Forschung kontextspezifische Methoden entwickelt (hat).

---

## 1.8 Weitere Datentypen

Auch wenn in der sozialwissenschaftlichen Forschungspraxis nach wie vor die Befragung die häufigste Datenerhebungsmethode ist, so hat sich die Sozialforschung seit jeher einer *Vielzahl von anderen Datentypen* (in der qualitativen Forschung auch *Datensorten* genannt) bedient. Die wichtigsten werden im sechsten Teil dieses Handbuchs vorgestellt. Zu nennen sind zunächst die vielen Varianten der *natürlichen* bzw. *prozessproduzierten Daten* (Salheiser, Kapitel 62 in diesem Band), die in irgendeiner Form verschriftlicht sind bzw. die *verbale Informationen* enthalten – sei es als *Dokument* (Salheiser, Kapitel 62 in diesem Band), sei es in Form von *Verwaltungsdaten* bzw. *Daten der amtlichen Statistik* (Hartmann/Lengerer, Kapitel 71 in diesem Band), sei es als *literarische Quellen und persönliche Dokumente* (Ernst, Kapitel 63 in diesem Band), sei es als *Zeitungsaufnahmen* (Klein, Kapitel 64 in diesem Band), oder seien es *Web Server Logs und Logfiles* (Schmitz/Yanenko, Kapitel 65 in diesem Band), um nur einige Beispiele zu nennen.

Bei diesen Beispielen ist eine Unterscheidung zwischen qualitativer und quantitativer Forschung wenig zweckmäßig. Diese Daten können entweder im Rahmen qualitativer Paradigmen bearbeitet oder alternativ standardisiert bzw. teilstandardisiert erhoben werden, so z.B. bei den Formen der computerunterstützten Inhaltsanalyse (Klein, Kapitel 64

in diesem Band), oder sie werden im Rahmen der Datenaufbereitung (Kuckartz/Rädicker, Kapitel 27 in diesem Band), etwa mit Hilfe eines inhaltsanalytischen Verfahrens (Mayring/Fenzl, Kapitel 38 in diesem Band), erst standardisiert und dann statistisch weitergehend ausgewertet (Blasius/Baur, Kapitel 79 in diesem Band). In diesen Bereich fallen auch die *Big Data*, die nicht als eigener Beitrag aufgenommen wurden, weil damit lediglich gemeint ist, dass die betreffenden Daten in großer Menge vorliegen.

Eine weitere Alternative zur Befragung ist die *teilnehmende und nicht-teilnehmende Beobachtung* (Thierbach/Petschick, Kapitel 66 in diesem Band), die insbesondere in ethnografischen Studien eingesetzt wird (Knoblauch, Kapitel 36 in diesem Band). Sie eignet sich insbesondere zur Analyse von sozialen Praktiken, Interaktionen und Situationen – und zwar in den Fällen, bei denen die Kommunikation zwischen Forschenden und Beforschten schwierig ist (z.B. mit kleinen Kindern), und dort, wo bei einer Befragung kaum mit validen Ergebnissen zu rechnen ist, z.B. zu den Lebensgewohnheiten von Obdachlosen. Zu dem Bereich der *visuellen Daten* gehören auch *unbewegte Bilder* wie Fotografien und Kunstgegenstände (Bohnsack, Kapitel 67 in diesem Band), *bewegte Bilder* wie Videodaten (Tuma/Schnettler, Kapitel 68 in diesem Band) und *Filme* (Akremi, Kapitel 69 in diesem Band), aber auch *Gebrauchsgegenstände und technische Artefakte* (Schubert, Kapitel 70 in diesem Band). Insbesondere in der Ethnografie (Knoblauch, Kapitel 36 in diesem Band) wurde hier in den vergangenen zwei Jahrzehnten die Methodenentwicklung stark vorangetrieben: So war das Kernproblem der teilnehmenden Beobachtung früher, dass die Daten nur unzureichend aufgezeichnet werden konnten und damit die Nachvollziehbarkeit der Interpretation sehr schwer war. Heutzutage kann mit Kameras und Smartphones die Aufzeichnung von Beobachtungsdaten vorangetrieben werden, was u.a. dazu führte, dass die Ethnografie zur Videografie (Tuma/Schnettler, Kapitel 68 in diesem Band) ausgebaut wurde.

Eine Reihe von Daten adressiert spezifische Themenfelder bzw. sie sind „Abfallprodukte“ sozialer Prozesse. So entstehen *Verwaltungsdaten* (Hartmann/Lengerer, Kapitel 71 in diesem Band) im Zuge von bürokratischen Prozessen. Hierzu gehören etwa Daten der amtlichen Statistik aus der Kranken-, Renten-, und Arbeitslosenversicherung, Studierendenstatistiken und vieles andere mehr. Sie enthalten häufig vielfältige sozialwissenschaftlich relevante Informationen und werden seit etwa 15 Jahren über Forschungsdatenzentren (FDZs) verstärkt für die Forschung zugänglich gemacht. So können aus den Daten der Rentenversicherung nicht nur Hinweise über das Rentenniveau in einer Stadt gewonnen werden, sondern man kann auch Berufsverläufe und Krankheitsgeschichten über einen langen Zeitraum analysieren. Diese Daten geben potenziell ein sehr detailliertes Bild über vielfältige soziale Prozesse, sind aber durch die Perspektive der jeweiligen Institution verzerrt – so enthalten die Daten der Rentenversicherung keine Informationen über alle Erwerbstätigen, sondern nur über die, die mindestens einmal in ihrem Leben Rentenbeiträge gezahlt haben.

Ein anderer Datentypus sind *Aggregatdaten* (Graeff, Kapitel 72 in diesem Band), auf deren Basis soziale Kollektive, Regionen oder Städte eines Landes und, auf der obersten Aggregatstufe, auch Länder miteinander verglichen werden können. Mit *Paneldaten*

können individuelle Veränderungen über Zeit beobachtet werden, so z.B. kann die Entwicklung der Langzeitarbeitslosigkeit in Deutschland mit Hilfe des Sozio-ökonomischen Panels (SOEP) von 1984 (Beginn des Panels) bis heute beschrieben werden (Schupp, Kapitel 73 in diesem Band). *Netzwerkdaten* (Baur, Kapitel 74 in diesem Band) ermöglichen die Abbildung relationaler Beziehungen zwischen Personen oder anderen Elementen, zwischen denen ein Netzwerk besteht. Mit *Organisationsdaten* (Liebig et al., Kapitel 75 in diesem Band) kann z.B. beschrieben werden, inwiefern Organisationen als Kontext oder als Akteure sozialer Ungleichheit fungieren. Um den *geographischen Raum* sozialwissenschaftlich zu analysieren kann sowohl auf *qualitative Daten* (Dangschat, Kapitel 76 in diesem Band), als auch auf *quantitative Daten*, die sogenannten *Geodaten* (Lakes, Kapitel 77 in diesem Band), zurückgegriffen werden. Schließlich werden über *neurowissenschaftliche Daten* (Pritzel/Markowitsch, Kapitel 78 in diesem Band) in den letzten Jahren zunehmend Sozialwissenschaften und Neurowissenschaften miteinander verknüpft.

---

## 1.9 Datenauswertung in der quantitativen Sozialforschung

Der letzte Abschnitt des Handbuchs ist der Datenauswertung in der quantitativen Sozialforschung gewidmet – insbesondere den *Varianten der multivariaten Statistik* (Blasius/Baur, Kapitel 79 in diesem Band). Dabei soll noch einmal betont werden, dass schon bei der Datenerhebung die Datenauswertung berücksichtigt werden muss. Werden bestimmte Variablen nicht abgefragt bzw. werden sie nicht in der Form abgefragt, wie sie später ausgewertet werden sollen, dann können die entsprechenden Auswertungen nicht vorgenommen werden. So ist es für die Konstruktion von Skalen unerlässlich, mehrere Items zum gleichen Thema abzufragen (Latcheva/Davidov, Blasius, Kapitel 55 und 83 in diesem Band), um damit die zugrundeliegende(n) latente(n) Variable(n) zu bestimmen.

Ähnlich wie bei den qualitativen Forschungsparadigmen gibt es eine Vielzahl von statistischen Auswertungsverfahren, die im Rahmen dieses Handbuchs nicht oder nur sehr kurz dargestellt werden können, für ausführlichere Beschreibungen verweisen wir auf das Handbuch von Wolf und Best (2010). Der letzte Teil dieses Handbuchs beginnt mit einer Übersicht über die multivariaten Verfahren (Blasius/Baur, Kapitel 79 in diesem Band), inklusive einer kurzen Wiederholung der uni- und bivariaten Verfahren.

Sehr beliebt in den Sozialwissenschaften sind *kausale Interpretationen*, z.B. innerhalb einer Regressionsanalyse, bei der postuliert wird, dass die unabhängigen Variablen  $x_1$  bis  $x_m$  die abhängige Variable Y erklären. Dass diese Annahme längst nicht immer richtig ist und dass ein kausaler Effekt, wie im linearen Regressionsmodell implizit angenommen, eigentlich nur in relativ wenigen Analysen inhaltlich richtig sein kann, zeigen Kühnel und Dingelstedt (Kapitel 80 in diesem Band).

Ein anderes wichtiges allgemeines Konzept der empirischen Sozialforschung sind *Indikatoren*, die im Vorfeld der Untersuchung festgelegt werden. Ein für viele sozialwissenschaftliche Studien zentraler Indikator ist die Einordnung einer Person bzw. eines Haushaltes auf einer Skala, die von „reich“ bis „arm“ verläuft. Dabei muss ein Schwellenwert

festgelegt werden, ab wann eine Person bzw. ein Haushalt als „arm“ bezeichnet werden kann (Burzan, Kapitel 81 in diesem Band). Als Indikator für diese Zuordnung wird häufig das Äquivalenzeinkommen verwendet, als Variablen für diesen Indikator werden das Einkommen und die Zusammensetzung des Haushaltes benötigt – zur Frage, wie diese soziodemografischen Merkmale erhoben werden können, verweisen wir auf Hoffmeyer-Zlotnik und Warner (Kapitel 54 in diesem Band). Noch etwas komplexer ist die *Messung der sozialen Ungleichheit*, wofür es verschiedene Konzepte gibt, so z.B. den Gini-Index (Huinink, Kapitel 82 in diesem Band). Zusätzlich zu den Indikatoren werden sehr oft Skalen benötigt, z.B. zur Messung von Religiosität, die mit Hilfe von *Skalierungsverfahren* (Blasius, Kapitel 83 in diesem Band) gebildet werden können.

Die nächsten fünf Kapitel sind ebenfalls den Datenstrukturen gewidmet, wobei die ersten drei einen *zeitlichen Aspekt* haben. Bei der *Zeitreihenanalyse* (Metz/Thome, Kapitel 84 in diesem Band) werden in der Regel Aggregatdaten (Graeff, Kapitel 72 in diesem Band) über einen längeren (bzw. über einen langen) Zeitraum betrachtet. Zeitreihenanalysen werden überwiegend in der Ökonomie verwendet, wo Wirtschaftsdaten (Bruttonsozialprodukt, Arbeitslosenquoten, ...) über einen langen Zeitraum analysiert werden, um daraus die zukünftige Entwicklung zu prognostizieren. In den Sozialwissenschaften kann mit Hilfe von Zeitreihenanalysen z.B. der Zusammenhang von Straftätern und Strafverdächtigen über Zeit beschrieben werden. *Längsschnittanalysen* (Mika/Stegemann, Kapitel 85 in diesem Band) arbeiten mit Daten, die kontinuierlich, z.B. monatlich, über einen längeren (langen) Zeitraum erfasst wurden. So kann z.B. auf Monatsebene erfasst werden, wann eine Person schwanger wurde (abgefragt wird der Geburtsmonat des Kindes) und wann diese Person geheiratet hat. Hier ist das Ziel herauszufinden, wann im Zeitverlauf die größte Wahrscheinlichkeit besteht, dass zwei Personen heiraten. Mit *Verlaufsdatenanalysen* (Pötter/Prein in diesem Band) werden Ereignisse und Folgen von Ereignissen analysiert, die sich beliebig oft wiederholen können, so z.B. Berufswchsel und Zeiten der Arbeitslosigkeit. In diesen Daten werden Strukturen gesucht, einander ähnliche Verlaufsmuster, die dann z.B. mit Hilfe der Clusteranalyse (Blasius/Baur, Kapitel 79 in diesem Band) zu Gruppen zusammengefasst werden.

*Mehrebenanalysen* (Pötschke, Kapitel 87 in diesem Band) haben einen Raum- bzw. einen anderen Kontextbezug. So hängt z.B. die Leistung der Schüler einer Klasse von deren individuellen Faktoren ab (Fleiß, Interesse am Thema, Einfluss des Elternhauses usw.), aber auf der zweiten Ebene auch von dem Klassenlehrer (seinen Fähigkeiten den Unterrichtsstoff zu vermitteln usw.). Da dieser Einfluss auf der zweiten Ebene alle Schüler einer Klasse betrifft, müssen relativ viele Klassen in die Analyse einbezogen werden, um eine Aussage auf dieser höheren Ebene machen zu können. Auf der nächst höheren Stufe, der dritten Ebene, kann die Schule genannt werden (deren Ausstattung usw.) – und auch auf dieser Ebene müssen wiederum ausreichend Schulen in die Analyse einbezogen werden, um signifikante Effekte bestimmen zu können; die nächsthöhere Ebene könnte das Bundesland sein (durchschnittlicher Unterrichtsausfall, Ressourcen, die an die Schulen gehen usw.).

Das letzte Kapitel des Handbuch ist den *Metaanalysen* (Wagner/Weiß, Kapitel 88 in diesem Band) gewidmet. Mit diesen können unterschiedliche empirische

Forschungsergebnisse, z.B. bezogen auf den Zusammenhang von Armut und Kriminalität, verglichen werden, um damit herauszufinden, welche der einzelnen Ergebnisse in allen (den meisten) einbezogenen Studien nachgewiesen werden konnten.

Wir wissen, dass – trotz aller Bemühungen, einen breiten Überblick zu vermitteln – dieses Handbuch noch Lücken hat, hoffen aber, dass wir zeigen können, dass sowohl qualitative, als auch quantitative Methoden für die Beantwortung sozialwissenschaftlicher Fragen wertvolle Beiträge leisten können. Für diejenigen, denen dieses Handbuch Lust auf empirische Sozialforschung und auf „mehr Methoden“ gemacht hat, haben wir (ebenfalls ohne Anspruch auf Vollständigkeit) im Anhang eine Auswahl wichtiger Methoden-Zeitschriften (Anhang 1.1), Berufsvereinigungen, Forschungsnetzwerke und Beratungsstellen (Anhang 1.2) zusammengestellt.

## Literatur

- Baur, Nina (2005): Verlaufsmusteranalyse. Wiesbaden: VS Verlag
- Blasius, Jörg/Thiessen, Victor (2012): Assessing the Quality of Survey Data. London: Sage.
- Bourdieu, Pierre (1982): Die feinen Unterschiede. Frankfurt a.M.: Suhrkamp
- Flick, Uwe (Hg.) (2007): The SAGE Qualitative Research Kit. 8 Bände. London/Thousand Oaks/New Dehli: Sage. URL: <http://www.sagepub.co.uk/booksProdDesc.nav?prodId=Book225013>
- Flick, Uwe (Hg.) (2014): The SAGE Handbook of Qualitative Data Analysis. London/ Thousand Oaks/Dehli: Sage
- Flick, Uwe/Kardorff, Ernst von/Steinke, Ines (Hg.) (2008): Qualitative Forschung – Ein Handbuch. Reinbek: Rowohlt
- Knoblauch, Hubert/Flick, Uwe/Maeder, Christoph (Hg.) (2005): The State of the Art of Qualitative Research in Europe. In: FQS: Forum Qualitative Social Research 6 (3). Art. 5. URL: <http://www.qualitative-research.net/>
- König, René (Hg.) (1962, 1967): Handbuch der empirischen Sozialforschung. Stuttgart: Enke
- König, René (Hg.) (1973-1979): Handbuch der empirischen Sozialforschung. 14 Bände. Stuttgart: dtv
- Scheuch, Erwin K. (1989/1993): Theoretical implicatons of comparative survey research. Nachgedruckt in: Historical Social Reserach 18 (2): 172-195. URL: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-32794>
- Reuband, Karl-Heinz/Blasius, Jörg (1996): Face-to-face, telefonische und postalische Befragungen. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 48 (2): 296-318
- Weber, Max (1904): Die „Objektivität“ sozialwissenschaftlicher und sozialpolitischer Erkenntnis. Nachgedruckt in: ibidem (1988): Gesammelte Aufsätze zur Wissenschaftslehre. 146-214. Tübingen: Mohr-Siebeck
- Wolf, Christof/Best, Henning (Hg.) (2010): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS Verlag

## Anhang

### Anhang 1.1: Eine Auswahl wichtiger Methoden-Zeitschriften

Zeitschriftenname	Link
BIOS. Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen	<a href="http://www.budrich-journals.de/index.php/bios">www.budrich-journals.de/index.php/bios</a>
BMS. Bulletin of Sociological Methodology	<a href="http://www.cmh.pro.ens.fr/bms/Welcome.htm">www.cmh.pro.ens.fr/bms/Welcome.htm</a>
Ethnography	<a href="http://www.uk.sagepub.com/journals/Journal200906">www.uk.sagepub.com/journals/Journal200906</a>
Field Methods	<a href="http://www.sagepub.com/journals/Journal200810/title">http://www.sagepub.com/journals/Journal200810/title</a>
FQS. Forum qualitative Sozialforschung	<a href="http://www.qualitative-research.net">www.qualitative-research.net</a>
HSR. Historical Social Research / Historische Sozialforschung	<a href="http://www.gesis.org/hsr">www.gesis.org/hsr</a>
IJPOR. International Journal of Public Opinion Research	<a href="http://ijpor.oxfordjournals.org">ijpor.oxfordjournals.org</a>
JCE. Journal of Contemporary Ethnography	<a href="http://jce.sagepub.com">jce.sagepub.com</a>
JMMR. Journal of Mixed Methods Research	<a href="http://mmr.sagepub.com">mmr.sagepub.com</a>
JOS. Journal of Official Statistics	<a href="http://www.degruyter.com/view/j/jos">http://www.degruyter.com/view/j/jos</a>
JSSAM. Journal of Survey Statistics and Methodology	<a href="http://jssam.oxfordjournals.org">jssam.oxfordjournals.org</a>
MDA. Methods, Data, Analyses. A Journal for Quantitative Methods and Survey Methodology	<a href="http://www.gesis.org/publikationen/zeitschriften/mda">www.gesis.org/publikationen/zeitschriften/mda</a>
QRJ. Qualitative Research	<a href="http://qrj.sagepub.com">qrj.sagepub.com</a>
Qualitative Inquiry	<a href="http://www.sagepub.com/journals/Journal200797">www.sagepub.com/journals/Journal200797</a>
Qualitative Sociology	<a href="http://link.springer.com/journal/11133">link.springer.com/journal/11133</a>
Quality & Quantity. International Journal of Methodology	<a href="http://www.springer.com/social+sciences/journal/11135">www.springer.com/social+sciences/journal/11135</a>
POQ. Public Opinion Quarterly	<a href="http://poq.oxfordjournals.org/">http://poq.oxfordjournals.org/</a>
SM. Sociological Methodology	<a href="http://www.asanet.org/journals/sm">www.asanet.org/journals/sm</a>
SMR. Sociological Methods & Research	<a href="http://smr.sagepub.com">smr.sagepub.com</a>
sozialersinn. Zeitschrift für hermeneutische Sozialforschung	<a href="http://www.sozialer-sinn.com">www.sozialer-sinn.com</a>
SRM. Survey Research Methods	<a href="http://www.europeansurveyresearch.org/journal">www.europeansurveyresearch.org/journal</a>
Studies in Qualitative Methodology	<a href="http://www.emeraldinsight.com/books.htm?issn=1042-3192">www.emeraldinsight.com/books.htm?issn=1042-3192</a>
Survey Methodology	<a href="http://www.statcan.gc.ca/ads-annonces/12-001-x/index-eng.htm">www.statcan.gc.ca/ads-annonces/12-001-x/index-eng.htm</a>

## Anhang 1.2: Eine Auswahl von Berufsvereinigungen, Forschungsnetzwerken und Beratungsstellen

Netzwerk bzw. Vereinigung	Link
AAPOR. American Association for Public Opinion Research	<a href="http://www.aapor.org">www.aapor.org</a>
ADM. Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.	<a href="http://www.adm-ev.de">www.adm-ev.de</a>
ALLF. Archiv für Lebenslaufforschung	<a href="http://www.lebenslaufarchiv.uni-bremen.de">www.lebenslaufarchiv.uni-bremen.de</a>
Archiv „Deutsches Gedächtnis“	<a href="http://www.fernuni-hagen.de/geschichteundbiographie/deutschessgedaechtnis/">www.fernuni-hagen.de/geschichteundbiographie/deutschessgedaechtnis/</a>
ASI. Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V.	<a href="http://www.asi-ev.org">www.asi-ev.org</a>
BVM. Berufsverband Deutscher Markt- und Sozialforscher e.V.	<a href="http://bvm.org">bvm.org</a>
DeGEval. Gesellschaft für Evaluation e.V.	<a href="http://www.degeval.de">www.degeval.de</a>
Destatis. Statistisches Bundesamt (mit Links zu Eurostat und den Statistischen Ämtern der Bundesländer)	<a href="http://www.destatis.de">www.destatis.de</a>
DGOF. Deutsche Gesellschaft für Online-Forschung e.V.	<a href="http://www.dgof.de">www.dgof.de</a>
ESOMAR. European Society for Opinion and Market Research	<a href="http://www.esomar.org">www.esomar.org</a>
ESRA. European Survey Research Association	<a href="http://www.europeansurveyresearch.org">www.europeansurveyresearch.org</a>
GESIS. Leibniz-Institut für Sozialwissenschaften	<a href="http://www.gesis.org">www.gesis.org</a>
INSNA. International Network for Social Network Analysis	<a href="http://www.insna.org">www.insna.org</a>
Methodensektion DGS	
Methodensektionen der DGS (Deutsche Gesellschaft für Soziologie), insbesondere Sektionen „Methoden der empirischen Sozialforschung“ (quantitative Methoden), „Methoden der qualitativen Sozialforschung“ (qualitative Methoden), „Wissenssoziologie“ (Ethnografie), „Soziologische Netzwerkforschung“ (Netzwerkanalyse) und „Modellbildung und Simulation“ (Simulation)	<a href="http://www.soziologie.de">www.soziologie.de</a>
Methodensektionen der ESA (European Sociology Association), insbesondere RN20 (Qualitative Methods) und RN21 (Quantitative Methods)	<a href="http://www.europeansociology.org">www.europeansociology.org</a>
Methodensektionen der ISA (International Sociology Association), insbesondere RC 33 (Research Committee on Logic and Methodology in Sociology)	<a href="http://www.rc33.org">www.rc33.org</a>
MMIRA. Mixed Methods International Research Association	<a href="http://mmira.wildapricot.org">mmira.wildapricot.org</a>
Rat der Deutschen Markt- und Sozialforschung e.V.	<a href="http://rat-marktforschung.de">rat-marktforschung.de</a>
RatSWD. Rat für Sozial- und Wirtschaftsdaten	<a href="http://www.ratswd.de">www.ratswd.de</a>

---

## **Teil 1**

### **Grundlagen der empirischen Sozialforschung**

Jo Reichertz

---

## 2.1 Entstehung und Prüfung wissenschaftlicher Theorien – eine kleine Geschichte

Der (Wieder-)Aufstieg der europäischen Wissenschaft im 17./18. Jahrhundert verdankt sich ganz wesentlich dem Abstieg der christlichen Religion. Galt die Wissenschaft der Religion zu Beginn der *christlichen* Zeitrechnung nicht wirklich als ernstzunehmende Konkurrenz, so wandelte sich dieses Verhältnis in der Aufklärung grundlegend: Die Kultur- und auch die Naturwissenschaft lösten mit Einsetzen der Aufklärung die Religion(en) in Bezug auf die Bereitstellung von Weltdeutungen und Theorien Schritt für Schritt ab, und dies gleich in zweifachem Sinne: Einerseits „erledigten“ sie die Religion, indem sie den Glauben an einen Gott und dessen Botschaft als vermeidbaren Irrtum bzw. als selbstgewollte oder böswillige Täuschung entlarvten, andererseits beerbten sie die Religion. Dem Wissenschaftler oblag demnach die Pflicht, das Wahre, das Vernünftige zu suchen und von ihm in Theorien zu künden – wissenschaftliche Theorien sind mithin Ausdruck einer innerweltlichen Religion, welche die Welt ohne das Wirken Gottes erklärt, und der Wissenschaftler dient als Priester dieser Vernunft der Diesseitigkeit.

Diese neuen „Wahrheitskünder“ benutzen heute, verallgemeinert man sehr stark, im Wesentlichen *vier* Methoden, Theorien und deren Validität zu fundieren – und zwar die Absicherung

- mit Hilfe des Hinweises auf eine anerkannte Autorität,
- mit Hilfe des Gebrauchs der Vernunft,
- mit Hilfe der Inanspruchnahme persönlicher Hellsichtigkeit und schließlich
- mittels eigener empirischer Forschung.

### **2.1.1 Absicherung mit Hilfe anerkannter Autoritäten**

Die Methode der Absicherung der Aussagen-Gültigkeit mit dem Hinweis auf vergleichbare Aussagen anderer, anerkannter *Autoritäten* hat eine sehr lange Tradition, und sie kann ihre religiöse Abstammung nicht leugnen, fundiert sie doch die Gültigkeit einer Aussage mit der hervorgehobenen Position seines Aus- und Fürsprechers. Lange Zeit wurden Nachfragen nach der Gültigkeit von Aussagen mit dem Hinweis auf die Bücher von (meist griechischen) Autoritäten gemeistert, und so blieb Meister Lampe (dank Aristoteles) für viele ein Wiederkäuer. Diese Methode, Gültigkeit zu begründen, endete nicht mit dem Niedergang der Renaissance, sondern ist auch heute noch auf fast allen Ebenen wissenschaftlicher Auseinandersetzung anzutreffen. Hatten jedoch vor einigen Jahren noch Habermas und Luhmann das erste und letzte Wort, so sind es heute vornehmlich Bourdieu, Foucault und Latour, denen diese zweifelhafte Ehre zukommt.

### **2.1.2 Absicherung mit Hilfe des Gebrauchs der Vernunft**

Auch die zweite Methode, nämlich Gültigkeit aufgrund des regelgerechten Einsatzes von *Vernunft* für sich in Anspruch zu nehmen, ist religiöser Abstammung, auch wenn sie sich sehr viel mehr als legitimes Kind der klassischen europäischen Philosophie wähnt. Die platonische Konstruktion eines Chorismos, der Kluft zwischen einer zeitlosen geistigen Ideenwelt und einer sinnlich erfahrbaren geschichtlichen Faktenwelt, etablierte zugleich die Überzeugung, wahre Erkenntnis habe sich von der sinnlichen (menschlichen) Wahrnehmung zu lösen, und Gültiges sei nur in der geistigen Schau zu erlangen. Diese Art des Philosophierens war und ist durchgängig gekennzeichnet durch den Gebrauch der Vernunft oder (in einer anderen, modernerer Ausdrucksweise) des Verstandes. Implizite Prämissen dieses Arguments ist jedoch, dass der Gebrauch der Vernunft deshalb so sinnvoll ist, weil diese ein Geschenk Gottes ist. Nicht das Instrument, sondern seine göttliche Herkunft verbürgt letztlich seine Güte.

### **2.1.3 Absicherung mit Hilfe persönlicher Hellsichtigkeit**

Im Kern ebenfalls religiös ist die dritte Methode zur Fundierung von Gültigkeit – nämlich der Hinweis auf eine dem individuellen Wissenschaftler eigene, besondere und gesteigerte *Hellsichtigkeit*. Validität wird in diesem Falle an die Person des Wissenschaftlers gebunden bzw. an dessen außerordentliche intellektuelle Kompetenz. Diese wird dann oft als Kunstfertigkeit entworfen, die folgerichtig in der Tradition des Genieglaubens, also der Vorstellung vom kreativen, also Neues schaffenden Potential der Künstler, steht. Dieses besondere künstlerische Vermögen, Neues zu erkennen und Neues zu bilden, ist (auch dann, wenn es an profane kognitive Fähigkeiten gebunden wird) letztlich eine „*Gabe Gottes*“ oder moderner: *guter Gene*.

### 2.1.4 Absicherung mit Hilfe empirischer Forschung

Mit dem Aufkommen des Empirismus im 16. Jahrhundert tauchte ein neues und sehr langlebiges Argument zur Begründung von Gültigkeit auf, das auf jede „göttliche“ Hilfe oder Unterstützung verzichtet – die *Beobachtung*. Sie gilt seitdem für viele Wissenschaftler als der beste und sicherste Weg zur Erlangung gültiger Aussagen und Theorien. Vernunft ohne Sinnesdaten erscheint den Vertretern empirischer Forschung blind, nur die systematische Erkundung der inneren und äußeren Welt mit Hilfe menschlicher Sinne kann (so der Glaube) Licht ins Dunkel bringen. Allein gestellt auf sich selbst (da ja von einem höheren Wesen keine Aufklärung mehr erhofft werden kann), nutzen menschliche Forscher und Forscherinnen das, was der Gattung „Mensch“ an Wahrnehmungsmöglichkeiten gegeben ist. Zusätzlich verfeinern und erweitern sie ihre Sinne mit einer Vielzahl von Medien: Manche dieser Medien erweiterten die Reichweite der Sinne, andere erhöhten deren Sensibilität, andere vergrößerten deren Speichervermögen und wieder andere verstärken deren Aufnahme- und Verarbeitungsgeschwindigkeit. Diese systematische Ausdehnung des Sinnesapparates soll die Grenzen der beschränkten menschlichen Wahrnehmung überschreiten und auf diese Weise gültiges Wissen erzeugen, das zu komplexen wissenschaftlichen Theorien verdichtet wird. *Das Standardargument aller empirischen Forscher und Forscherinnen lautet dabei in etwa so: Wissenschaftliche Aussagen sind gültig, weil sie dem Beobachteten entsprechen – die Aussagen sind letztlich nichts anderes als verallgemeinerte Beobachtungen.*

Schon sehr früh wurde diesem Anspruch (und dieser Hoffnung) des *Induktionismus* widersprochen. So machte Kant auf die unhintergehbare Selektivität des menschlichen Erkenntnisapparates aufmerksam, Marx (und später Mannheim) zeigten dessen Bindung an die soziale Position und Freud die an das individuelle Schicksal. Die massive Kritik an dem Induktionismus des Wiener Kreises, der (vergeblich) versuchte, deutungsfreie Protokollsätze von der Wirklichkeit zu formulieren, erschütterte zu Beginn des 20. Jahrhunderts weiter die Gültigkeit empirischer Forschung. Ein Übriges leistete der Hinweis der Sprachphilosophie, dass es sich bei wissenschaftlichen Erkenntnissen allein um sprachliche Äußerungen handele und nicht um geronnene Beobachtungsdaten und dass beides kategorial voneinander zu trennen sei. Weder die Aussagenlogik noch der Aufbau einer „idealen Sprache“ (beides Versuche, die Empirie durch das Mittel der Vernunft zu ergänzen und zu verbessern) konnten jedoch die Gültigkeitslücke schließen. Stattdessen setzten viele Forscher und Forscherinnen (z.B. Popper 1935; Reichenbach 1938) auf eine empirisch fundierte Intuition bei der *Entdeckung* („*Logic of Discovery*“) und eine streng empirisch-logisch vorgehende *Rechtfertigung* („*Logic of Justification*“). Am Beginn der Forschung steht der gute und kreative Einfall, der in der Psyche des Entdeckenden verankert ist und nicht zur Wissenschaft gehört, dann erst folgt die systematische wissenschaftliche Überprüfung und Begründung des Einfalls. Diese „Lösung“ geht einher mit einer scharfen Trennung zwischen einer Logik oder besser: einer individuellen *Kunst* der Entdeckung und einer systematischen *Logik* der Begründung. Auch heute noch ist in der *quantitativen Sozialforschung* ein beachtlicher Nachhall dieser Position vernehmbar.

Die neue qualitative Sozialforschung, entstanden in den 1970er Jahren, hat sich der Trennung von Entdeckungs- und Rechtfertigungszusammenhang von Beginn an vehement verweigert und statt dessen im Laufe der letzten vier Jahrzehnte neue Plausibilitäten sozialwissenschaftlichen Forschens (Methodologien und Methoden) entwickelt, erprobt und teilweise auch schon kanonisiert. Diese neuen Plausibilitäten sind selbst aus der Kritik klassischer Standards hervorgegangen und die qualitative Forschung hat auch durch diese Kritik ihre Kontur und ihr Selbstverständnis gefunden. Dabei beanspruchte diese Art der Forschung, die mal „*qualitative Forschung*“, mal „*interpretatives Paradigma*“, mal „*verstehende Sozialforschung*“ genannt wird, stets für sich, die bislang von der Wissenschaft wenig beachteten oder sozial verhüllten Gegenstände (z.B. subjektive Perspektive, latente Muster oder die Hinterbühne von Organisationen) angemessen(er) ausleuchten zu können.

Dieses qualitative Forschungsprogramm hat in den letzten Jahrzehnten nicht nur enorm an Bedeutung gewonnen, sondern hat sich zweifellos in den deutschen Landen etabliert (Reichertz 2007). Allerdings sind die kämpferischen Tage der Pionierzeit vorbei und fast schon vergessen – die Zeit der metatheoretischen und paradigmatischen Abgrenzungsbemühungen gegenüber quantifizierenden Ansätzen scheint vorerst vorüber. Die „lange erstarrten Fronten [sind] (...) – unverkennbar – in Bewegung geraten“ (Esser 2007: 351).

Seit den späten 1990er Jahren spricht man in beiden Gruppen verstärkt von zwei *Forschungsstrategien*, die beide ihre Berechtigung hätten – es käme auf die jeweilige Frage und das jeweilige Ziel der Forschung an, ob man eine qualitative oder eine quantitative Strategie einzusetzen habe. Andere propagieren bereits die „*Mixed Methods*“ – also die Verbindung von qualitativen mit quantitativen Methoden (z.B. Kelle 2008, Kapitel 8 in diesem Band) – im Übrigen mit beachtlichem Erfolg. Die Frage ist jedoch, ob sich die beiden Positionen wirklich mit einander verbinden lassen oder ob sie nur isoliert nebeneinander stehen können.

---

## 2.2 Zwei Forschungsparadigmata, zwei Forschungsprozesse

Oft wird mit Rückgriff auf die Bestimmung von Weber, der Soziologie ginge es um „erklärendes Verstehen“ (Weber 1973), der qualitativen Forschung das „*Verstehen*“ und der quantitativen Forschung das „*Erklären*“ zugeordnet. Diese simple Zuordnung ist gewiss ungenau oder besser: undifferenziert. Angebrachter ist die Sicht, dass es das Ziel jeder wissenschaftlichen Forschung ist, zu allgemeinen Aussagen zu kommen. Wissenschaft kann nämlich (will sie ernst genommen werden) nicht im Besonderen verbleiben, sondern Wissenschaft muss auch das Allgemeine wollen. Deshalb beschränken sich nur sehr wenige qualitativ arbeitende Forscher und Forscherinnen auf die Deskription und/oder Dokumentation des Einzelfalles. Fast allen geht es auch um das Muster, das verbindet, das Muster, das verstehen lässt, das Muster, das erklärt. Es gibt also keinen generellen Verzicht der Qualitativen auf Erklären – wie dies immer wieder behauptet und auch befürchtet wird.

Einzuräumen ist allerdings, dass innerhalb qualitativer Forschung „Erklären“ nicht mehr mit einem einfachen Determinismus verbunden wird, sondern nur noch mit Wahrscheinlichkeiten. Aber das hat etwas mit dem Gegenstand der qualitativen Forschung zu tun. Dieser Gegenstand (= menschliches sinnhaftes Handeln) ist einerseits sehr komplex, aber wichtiger ist, dass er nicht nur auf die eigene Praxis reagiert, indem er das Vergangene interpretiert und sein Handeln danach neu orientiert, sondern diese Praxis reagiert auch sehr sensibel auf Deutungen dieses Handelns z.B. durch die Wissenschaften. Und je mehr die Wissenschaftler/innen sich in ihren Publikationen an die Öffentlichkeit wenden und je leichter diese an die Ergebnisse der Wissenschaft gelangen kann, desto leichter und nachhaltiger wird die Wissenschaft das von ihr Untersuchte auch irritieren und verändern. Insofern konstituieren die Wissenschaften ihren Gegenstand immer auch mit.

## 2.2.1 Das Feld qualitativer Methoden

Qualitative Methoden weisen in gewisser Hinsicht Ähnlichkeiten mit Südfrüchten auf – nicht weil sie wohlschmeckend oder gar gesund wären. Nein, ähnlich wie bei Südfrüchten gibt es auch bei den qualitativen Methoden nicht etwas Bestimmtes, Festes, das (bei näherer Betrachtung) allen gemeinsam wäre – etwas, das es rechtfertigen würde, einerseits Feigen, Bananen und Zitronen und andererseits Inhaltsanalyse (Mayring/Fenzl, Kapitel 38 in diesem Band), Grounded Theory (Strübing, Kapitel 32 in diesem Band), Biographieforschung (Rosenthal, Kapitel 35 in diesem Band), Diskursanalyse (Traue et al, Kapitel 34 in diesem Band), hermeneutische Wissenssoziologie (Kurt/Herbrik, Kapitel 33 in diesem Band), Ethnografie (Knoblauch, Kapitel 36 in diesem Band) und Methode der Einzelfallanalyse (Hering/Schmidt, Kapitel 37 in diesem Band) unter einen jeweils eigenen Begriff zu fassen (Reichertz 2007).

Betrachtet man das aktuelle Feld der qualitativen Methoden, dann sieht man vor allem „ein kompliziertes Netz von Ähnlichkeiten, die einander übergreifen und kreuzen – Ähnlichkeiten im Kleinen wie im Großen“ (Wittgenstein 1977: 57). Was man gerade nicht sieht, das sind klare Grenzen und abgeschlossene Gebiete. Es gibt also aus meiner Sicht keine (kleine) Schnittmenge, die allen qualitativen Methoden gemein ist (z.B. die Ausrichtung auf den Akteur und seine Intentionen), sondern es gibt Ähnlichkeiten und Überschneidungen, aber auch Widersprüche und Gegensätze. Zumal darin sind sich die aktuellen Versuche einig, die qualitative Sozialforschung fassbar zu machen (Flick et al. 2004; Reichertz 2007; Przyborski/Wohlrab-Sahr 2008; Strübing 2013).

Weil das so ist, macht es keinen Sinn, von *der* qualitativen Sozialforschung zu sprechen, sondern, wenn überhaupt, sollte man das Ganze das *Feld* der qualitativen Methoden nennen. Dieses Feld der qualitativen Methoden spannt sich im Wesentlichen entlang zweier Prämissen auf: Die eine bezieht sich auf den *Gegenstandsbereich* der Wissenschaften, die sich mit dem Menschen beschäftigen, die andere bezieht sich auf die *Arbeit der Erforschung* menschlichen Handelns. Beide Prämissen bestehen aus einer Reihe von Annahmen über die Besonderheiten (a) menschlichen Handelns und (b) deren Erforschung. Diese Bündel

theoretischer Aussagen zum Gegenstandsbereich und zur Methodologie resultieren so nicht aus empirischer Forschung, sondern liegen ihr zugrunde. Sie beinhalten im Wesentlichen folgende Aussagen:

### 2.2.1.1 Der Gegenstandsbereich

*Es sind immer konkrete Menschen, die handeln.* Stets nehmen konkrete, allerdings in die Geschichte und in die Gesellschaft eingebettete und eingebundene Menschen etwas wahr, bewerten es, messen ihm Sinn zu, ordnen sich dann (aufgrund der vorgenommenen Sinnzuschreibung) unter, oder lassen alles beim alten, oder entscheiden sich dafür, etwas zu verändern oder Neues zu entwickeln. Das tun sie in erworbenen, gesellschaftlich erarbeiteten Formen, Gattungen und Rahmen. Handeln hat immer eine *Geschichte* und einen *Kontext*. Auf dieses Handeln wirkt das Äußere – die Natur, die Sozialität, der Kontext, die Geschichte – nicht direkt und unmittelbar ein, sondern das Außen wird von der implizit deutenden *Wahrnehmung und der (bewussten oder routinisierten) Deutung des Handelnden* gebrochen. Das Äußere besitzt nur dann (einschränkende oder ermögliche) Kraft und manchmal auch Macht über den Handelnden, wenn es *durch* ihn und damit *für* ihn Bedeutung erhalten hat.

### 2.2.1.2 Methodologie

In und mit ihrer Lebenspraxis schaffen (konstruieren) Menschen arbeitsteilig Gesellschaft und soziale Ordnung: nicht jeder mit der gleichen Möglichkeit, seine Vorstellungen umzusetzen und viele nicht freiwillig. Ihre Deutungen der gemeinsamen Welt werden im Anschluss an die Arbeiten von Alfred Schütz „*Konstruktionen erster Ordnung*“ (Schütz 2010) genannt. Sie sind Ausdruck der jeweiligen sozialen Lage und zugleich auch Mittel der „*Reflexion*“ und Veränderung dieser sozialen Lage. Wissenschaftliche Arbeit ist immer und notwendig Teil dieser arbeitsteilig organisierten Schaffung von Gesellschaft und der sozialen Lagen. Sie ist ein spezifischer Arbeitsbereich mit spezifischen Methoden, Arbeitsstilen und Zielen. Wissenschaftliche Arbeit, die wesentlich durch *Forschung, Kommunikation, Reflexion, Lehre und Prüfung* gekennzeichnet ist, produziert auf diese Weise „*Konstruktionen zweiter Ordnung*“. Dies sind also (*Re-)Konstruktionen der sozialen Konstruktionen erster Ordnung*). Da Wissenschaftler auch immer in der Praxis leben, die sie untersuchen, sind sie Produzenten von Konstruktionen erster wie zweiter Ordnung. Das bringt sie in eine verzwickte Lage.

### 2.2.1.3 Vielfalt ohne rechte Einheit

*Vielfalt ohne rechte Einheit* – so lautete oben der Befund hinsichtlich der Geschlossenheit der deutschen qualitativen Sozialforschung. Dennoch ist es gewiss keine Geschmacksfrage, welche Forschungsstrategie gewählt wird, besteht doch eine enge Verbindung zwischen vorab entwickelter Theorie über den Gegenstand, der Datenerhebung und der Datenanalyse. Mit einigen Daten lässt sich mehr anfangen, mit anderen weniger. Mit welchen Daten sich nun mehr und mit welchen sich nun weniger anfangen lässt, ist letztlich eine Frage des Verwendungszweckes oder anders: Ob Daten nützlich sind, hängt von der Frage ab, auf

die man mit ihrer Hilfe eine Antwort (re)konstruieren will (Reichertz/Englert 2011). Das scheint mir im Übrigen der Punkt zu sein, über den sich alle Forscher und Forscherinnen, egal welche Methoden sie bevorzugen, einig sind: *Die Auswahl der Methoden hängt von der Frage ab, auf welche die Forschung eine Antwort produzieren soll.* Und meines Erachtens lassen sich zur Zeit innerhalb der qualitativen Forschung insgesamt vier Großfragerichtungen unterscheiden (Lüders/Reichertz 1986):

- *Frage nach den subjektiven Sinnwelten von Handlungen* (Was war der subjektiv gemeinte Sinn einer Handlung?)
- *Deskription sozialen Handelns und sozialer Milieus* (Wie sieht das Milieu aus, in dem das Handeln situiert ist?)
- *Rekonstruktion deutungs- und handlungsgenerierender Strukturen* (Welche latenten Strukturen determinieren das Handeln der Subjekte?)
- *(Re)Konstruktion historisch und sozial vortypisierter Deutungsarbeit* (An welchen sozial verteilten Wissensordnungen orientieren sich die Handelnden, indem sie diese deuten?)

Da qualitative Forschung ein zukunftsoffener Prozess ist, bleibt abzuwarten, ob es bei diesen vier Forschungsperspektiven bleiben wird. Auch hier ist zu erwarten, dass neue Entwicklungen (Medien, Daten, Gegenstandsbereiche) es erforderlich machen, neue Fragen zu stellen und neue Theorien und Verfahren zu entwickeln.

## 2.2.2 Quantitative Methoden – ein Versuch der Umgrenzung

Die These, dass auch im Feld der quantitativen Forschung vor allem Vielfalt anzutreffen ist, ist nicht gewagt. So verweist etwa Ziegler (Kapitel 10 in diesem Band) auf die Unterschiede zwischen wissenschaftlicher Forschung und kommerzieller Markt- und Meinungsforschung. Gemeinsam dürfte allen Vertretern und Vertreterinnen der „quantitativen Sozialforschung“ lediglich der stabile Glaube an drei *Glaubenssätze (Axiome)* sein, nämlich *dass (a) auch die Welt der sozialen Ordnung und die Welt der handelnden Akteure recht stabil, (b) von Gesetzmäßigkeiten bestimmt und (c) direkt oder indirekt beobachtbar sind* (z.B. Diekmann 2007, Häder 2010, Kromrey 2009). Und weil dies so ist, können diese Welten genau vermessen und auch Theorien zu ihnen entwickelt werden. Und – so die Annahme – die Theorien können auch in einem „dichten“ Verhältnis zu der jeweils untersuchten Welt stehen – entweder weil man aufgrund von Beobachtungen (also induktiv) Wissen von der Welt erlangt hat (das ist eine etwas ältere und heute weitgehend überholte Sicht) oder weil man aufgrund der Unterstellung von Gesetzen Annahmen über die Beschaffenheit von Welt abgeleitet und mittels empirischer Forschung entweder widerlegt oder bestätigt hat.

Diese Verfahrensweise, die (wie Popper es einmal formuliert hat) ein aus Theorien geknüpftes Netz auswirft, um Stück für Stück die Welt einzufangen, wird *nomologisch-deduktiv* genannt und leitet einen zu erklärenden Sachverhalt aus Gesetzen und Bedingungen ab. Auch heute gilt sie bei vielen (noch) als Inbegriff empirischer Sozialforschung. Ziel

dieser Forschungsstrategie, die vor allem auf die Beseitigung falscher Hypothesen setzt, ist, möglichst nahe an die „Wahrheit“ heranzukommen.

Quantitativ orientierte Forscher und Forscherinnen verfügen über vermeintlich scharfe *Gütekriterien* – einerseits um sich von bestimmten Kollegen abzugrenzen, andererseits um bei konkurrierenden Theorien zwischen den Böcken und den Schafen unterscheiden zu können. Zentral hierfür sind neben den Kriterien der *Repräsentativität* der Datenauswahl die *Objektivität*, die *Validität* und die *Reliabilität* der Datenauswertung (Stein und Krebs/Menold, Kapitel 7 und 30 in diesem Band).

Diese Gütekriterien beziehen sich nun auf den Teil des Forschungsprozesses, der als durchaus heikel gilt – nämlich auf den Teil der Forschung, in dem die „harte“ Wirklichkeit durch bestimmte (Mess-)Operationen in wissenschaftlich verwertbare (in diesem Fall: numerische) Daten verwandelt wird. Hat man erst einmal die numerischen Daten, dann kann man mit der Hilfe von Logik und Mathematik Hypothesen und auch Theorien testen. Da die Güte von Logik und Mathematik außer Zweifel stehen, sind die entscheidenden Fragen alleine die, ob die Daten gut ausgewählt wurden (Repräsentativität) und ob „gut“ gemessen wurde. Objektivität, Validität und Reliabilität beziehen sich nun genau und nur auf diesen Messvorgang. Die strategische Bedeutung der beiden Begriffe innerhalb einer quantifizierenden Sozialforschung besteht also darin, dass sie im wahrsten Sinne des Wortes das Fundament legen für die Glaubwürdigkeit wissenschaftlicher Forschung: Ist das Fundament brüchig, wird das darauf ruhende Theorie-Gebäude bald einstürzen – ist dagegen das Fundament stabil, lassen sich selbst gewagte und weit gespannte Brücken bauen.

### **2.2.3 Lassen sich qualitative und quantitative Methoden miteinander verbinden?**

Wissenschaftliche Methoden sind bestimmte Praktiken, mit Daten umzugehen – und zwar solche Praktiken, von denen bestimmte Wissenschaftler zu bestimmten Zeiten erhoffen, dass mit ihrer Hilfe das Offensichtliche deutlich überschritten werden kann. Methoden gründen stets, und diese Einsicht ist weder neu noch originell, auf einer oft impliziten und (zu) selten explizierten Vorstellung davon, was die Daten „sind“ bzw. repräsentieren, was wir wie erkennen und wie sich Daten erheben und auswerten lassen. Forschung „schafft“ somit ihren Gegenstand (in gewisser Weise) und damit auch die zu erhebenden Daten – weder das Eine noch das Andere findet sie einfach in der Außenwelt vor (Breuer 2005). *Deshalb sind Methoden immer theoretisch geleitet und deshalb sind die ausgearbeiteten „Methoden“ immer auch Theorie:* Gesellschaftstheorie, Sozialtheorie, Handlungstheorie, Kommunikations- und Medientheorie und Erkenntnistheorie in einem.

Qualitative wie quantitative Methodenlehren – und das kennzeichnet sie – sind wegen dieser Lage der Dinge grundsätzlich der *Selbstreflexion* verpflichtet und zugleich darum bemüht, die eigene *theoretische und methodische Arbeit immer wieder selbst in die Forschung mit einzubeziehen*.

Wer dagegen ohne vorab entworfene Grundlagentheorie seine Forschungsarbeiten beginnt, liefert sich (meist ohne dass es von den Forschenden selbst wahrgenommen wird) seinen *Alltagstheorien über den Gegenstand* aus – seinem Common Sense: Statt kontrollierter und reflektierter Erkenntnisse liefern nicht-bewusste, meist ad-hoc entworfene Vor-Urteile die relevanten Ansichten über die noch zu untersuchenden Gegenstände und verlängern sie auf diese Weise.

Mit dieser Forschungspraxis einher geht oft die Ansicht, die Methoden seien neutrale *Tools* (im Sinne von „Hilfsmittel“), die sich je nach Geschmack und Arbeitszeit beliebig für jede Fragestellung nutzen lassen. Hauptsache sie sind „praxisnah“ oder auch: „praxistauglich“. Hier ergibt sich die Auswahl der Verfahren der Datenerhebung und Datenauswahl eher aus situativen Erwägungen als aus systematischen Gründen. Eine solche Methodenwahl gleicht in wesentlichen Punkten dem Zusammenstellen des Essens in Kantinen und Menschen. Hier wählt man das aus den angebotenen Angeboten aus, was einem gerade schmackhaft, gesund oder bezahlbar erscheint. So sammelt sich dann auf dem Tablett ein buntes Gemenge von Vor-, Haupt- und Nachspeisen, deren jeweilige Spezifik sich daraus ergibt, dass sie gerade verfügbar und günstig sind und „passend“ erscheinen.

Methoden sind jedoch (wie oben dargelegt) keine Mittel, die man sich nach aktuellem Geschmack und Verfügbarkeit zu einem Kantinenmenü zusammenstellen kann, sondern sie sind im engen Sinne des Wortes „Tools“, also *Handwerkszeuge*. Jedes Handwerkszeug enthält in seiner Form und seiner Materialität bereits eine „*Theorie*“ seines Gegenstandes: Der Hammer ist so wie er ist, weil er sich aus der Praxis des Nagelleinschlagens ergeben hat und das gilt vergleichbar auch für den Meißel, die Säge, den Füller, den Anspitzer und alle Dinge, die geeignet sind, bestimmte Aufgaben effektiv zu erledigen. Wer mit einer Spitzhacke einem Zahn im Mund zu Leibe rückt wird ebenso scheitern wie der, welcher das Fieberthermometer nutzt, um Erbsen zu zählen. Das ist trivial – ohne Zweifel. Nicht trivial ist dagegen der Befund, dass diese Trivialität oft vergessen wird.

## 2.2.4 Es gibt sie doch – die zwei Kulturen

Aber es gibt nicht nur Unterschiede, es gibt auch *Gemeinsamkeiten zwischen den Gruppen* der qualitativen und quantitativen Forscher und Forscherinnen: Auffällig ist erst einmal das *Fehlen einer rechten Einheit in beiden Gruppen*. So gibt es keine Einheit bei den Qualitativen: Es gibt die, die vor allem verstehen wollen, und die, die den Prozess des Verstehens und das Verstandene reflektieren wollen. Und es gibt auch keine rechte Einheit bei den Quantitativen: Es gibt die, die vor allem zählen und vermessen, und die, die den Prozess des Zählens und Vermessens und das Vermessene reflektieren wollen.

In beiden Gruppen (auch das ist ihnen gemeinsam) arbeiten Forscher und Forscherinnen offen, kommunikativ, flexibel, intersubjektiv und reproduzierbar. In beiden Gruppen gibt es welche, die sich auf Aussagen über Zusammenhänge kleiner Reichweite beschränken, und andere, die zudem noch Zusammenhänge mittlerer und großer Reichweite erklären wollen. In beiden Gruppen finden sich Forscher und Forscherinnen, welche

die Mikroperspektive, und andere, welche die Makroperspektive bevorzugen. Manchmal liefern bei Forschungsarbeiten die Quantitativen die Hypothesen, manchmal tun das die Qualitativen, und manchmal überprüfen die Quantitativen und manchmal tun das die Qualitativen.

Und natürlich ruht jede quantitative Untersuchung auf einer qualitativen Basis und natürlich muss sie im Verlauf der Arbeit immer wieder interpretieren – weshalb in quantitativen Untersuchungen immer und notwendigerweise mit den Prämissen qualitativer Forschung gearbeitet wird. Und natürlich ruht jede qualitative Untersuchung auf einer quantitativen Basis (z.B. dann, wenn sie „Normalität“ bestimmen will) und natürlich muss sie im Verlauf der Arbeit immer wieder die Relevanz, die Wichtigkeit, die Häufigkeit „intuitiv“ ermitteln – weshalb in qualitativen Untersuchungen immer und notwendigerweise mit den Prämissen quantitativer Forschung gearbeitet wird.

In der konkreten Forschung durchdringen sich also das Feststellen von Häufigkeiten und die Ausdeutung von Sachverhalten – und zwar in jeder Phase der Forschung. Es gibt sie also nicht, die klare Trennung der quantitativen und qualitativen Forschung entlang von bestimmten *Merkmalen, Phasen, Perspektiven* oder *Reichenweitenansprüchen*. Gründe genug, darauf zu hoffen, dass es auch eine echte Verbindung der beiden Methodenverständnisse geben könnte.

*Was die beiden Forschungsrichtungen allerdings trennt, das ist der Umstand, dass jede Richtung eine eigene Kultur besitzt und dass diese Kulturen nicht so viele inhaltliche Gemeinsamkeiten aufweisen.* Das gilt insbesondere, wenn man im Anschluss an Hans-Georg Soeffner unter „Kultur“ jenen „Bedeutungsrahmen [versteht – J.R.], in dem Ereignisse, Dinge, Handlungen, Motive, Institutionen und gesellschaftliche Prozesse dem Verstehen zugänglich, verständlich beschreibbar und darstellbar sind“ (Soeffner 1988: 12). Obwohl manche sprachlichen Formulierungen ähnlich klingen, ist eine Reise von der qualitativen Forschung zur quantitativen nicht mit einer Reise von Dortmund nach Essen, sondern mit einer von Dortmund nach Detroit zu vergleichen: Ein Ozean trennt die beiden Kulturen – um eine alte Idee und Metaphorik von Snow (1965) aufzugreifen. Quantitative und qualitative Forschung sind nicht nur durch die Methoden getrennt, sondern vor allem und wesentlich: durch die Kultur, deren Ausdruck die Methoden sind (zu der Idee der verschiedenen Forschungskulturen auch im Hinblick auf die Sozialwissenschaften siehe Lepenies 1985).

Wollte man beide Kulturen zum Zwecke der Optimierung, die sowohl die Qualität des Endproduktes erhöht, als auch bleibend eine konstante Qualität erreicht, wirklich miteinander verbinden, dann bräuchte man als erstes eine „neue“ Sprache, mit der die Fragen und Probleme aus einer *neutralen* und übergeordneten Perspektive formuliert werden könnten. Die beiden methodischen Zugangsweisen zur Wirklichkeit müssten zugleich als Perspektiven begriffen werden und das Neue bestünde dann in einer Perspektivenüberschreitung, die jedoch die beiden vorhandenen Perspektiven bewahrt und zugleich überschreitet. Ich bin mir allerdings nicht sicher, ob es überhaupt eine Sprache geben kann, die *methodenneutral* ist, die also nicht bereits Ausdruck eines Paradigmas ist. Unterhalb der Entwicklung einer neuen gemeinsamen Sprache und einer Perspektivenüberschreitung ist

---

jedoch jede Kombination von qualitativen und quantitativen Methoden nur „Kantinenessen“ oder moderner: ein *Crossover*, ein Gemenge ohne ernsthafte Qualitätsverbesserung.

Ohne Zweifel bringen ein *Crossover* und ein *Mix* der Methoden ein neues, *bunteres* Bild der untersuchten gesellschaftlichen Prozesse. Wenn man aber daran festhält, dass (allen sozialkonstruktivistischen Einsprüchen zum Trotz) Wissenschaft auf eine Verbesserung von Validität hinarbeitet, dann ist fraglich, ob das *Crossover* und der *Mix* mehr an Validität erbringen kann – immer unterstellt, man ist sich einig, dass „Validität“ nicht eine messtechnisch definierte Gültigkeit meint, sondern statt dessen, dass der untersuchte Wirklichkeitsbereich in irgendeiner Weise „besser“, „differenzierter“ erfasst wird. Praxistaugliche Handreichungen zum *Crossover* von Methoden führen dann eher in eine (neue) Sackgasse, als dass sie Wege eröffnen.

---

## 2.3 Über die Logik der Forschung

Die gesellschaftliche Ordnung, an der sich Menschen (oft, aber nicht immer) in ihrem Handeln orientieren, ist stets im Fluss und ist zudem „subkulturell fragmentiert“. Aber auch die Ordnungen, die Menschen ihrem Handeln bewusst oder unbewusst geben, oder die sie übernehmen, sind im Fluss und sozial verteilt. Die Ordnung(en) besitzen deshalb immer nur einen lokalen Geltungsbereich und werden ständig und (seit dem Aufkommen der Moderne und der Postmoderne) immer schneller von eben diesen Menschen geändert, die ihnen zuvor (in Maßen) noch folgten. Hinzu kommt, dass sowohl die Gestaltung als auch die Geltung dieser Ordnung an die Sinnzuschreibungen und Interpretationsleistungen der handelnden Subjekte gebunden sind. Sozialwissenschaftliche Handlungserklärungen, unabhängig davon, ob sie qualitativ oder quantitativ vorgehen, zielen nämlich auf die (Re)Konstruktion der für die handelnden Subjekte relevanten Ordnungen.

Allerdings können solche Ordnungen nicht mehr aus klassischen und bewährten Großtheorien abgeleitet werden, da diese zum einen in der Regel nicht „lokal“ genug, zum anderen diese durch den gesellschaftlichen Wandel bereits überholt sind. Weil dies so ist, müssen neue „passende“ Ansichten über die Beschaffenheit sozialer Ordnung und sozialen Handelns stets aufs Neue von der Sozialwissenschaft generiert werden. Deshalb ist es ausgesprochen sinnvoll, sich die zu verstehende Lebenspraxis möglichst genau anzusehen und aufgrund dieser Daten die neuen Ordnungen zu (re)konstruieren.

Wenn man nun in der (qualitativen wie der quantitativen) Forschung ernsthaft damit beginnt, erhobene Daten auszuwerten, also diese entlang bestimmter Merkmale und Merkmalsordnungen zu vermessen und zu typisieren, dann stellt sich sehr schnell die Frage, wie man Ordnung in sein Datenchaos bringen kann. Das ist nur zu einem geringen Teil eine arbeitsorganisatorische Frage (Sortieren der Daten), sondern sehr viel mehr die Frage, wie (also mit welcher gedanklichen Operation oder Logik) die unüberschaubare Mannigfaltigkeit der Daten mit (vorhandenen oder noch zu findenden) Theorien in Verbindung gebracht werden können.

Bei diesem Unternehmen sind idealtypisch drei Verfahren zu unterscheiden (Peirce 1976; Reichertz 2013), wobei das zweite Verfahren im Weiteren in zwei Untergruppen geteilt wird – jedoch nicht, weil es zwischen den beiden gravierende Unterschiede gibt, sondern weil so die wichtige Unterscheidung zwischen Abduktion und Hypothese bzw. qualitativer Induktion klarer gemacht werden kann.

### **2.3.1 Deduktion**

Eine Art der Datenauswertung besteht in dem Verfahren der Subsumtion. Die Subsumtion geht von einem bereits bekannten Merkmalszusammenhang, also einer bekannten Regel aus (z.B.: Alle Einbrecher, die auch den Medizinschrank plündern, sind drogenabhängig.) und versucht diesen allgemeinen Zusammenhang in den Daten wiederzufinden (z.B. Der unbekannte Einbrecher hat den Medizinschrank geplündert.), um dann über den Einzelfall Kenntnisse zu erlangen (z.B.: Der unbekannte Einbrecher ist drogenabhängig.).

Die logische Form dieser gedanklichen Operation ist die der Deduktion: Der in Frage stehende Einzel-Fall wird einer bereits bekannten Regel untergeordnet. Hier wird eine vertraute und bewährte Ordnung auf einen neuen Fall angewendet. Neues (über die Ordnung der Welt) erfährt man auf diese Weise nicht – halt nur, dass der noch unbekannte Einbrecher drogenabhängig ist (ein Wissen, das für die Polizei ganz nützlich sein kann – wenn die Regel stimmt). Deduktionen sind also tautologisch, sie besagen nicht Neues. Deduktionen sind jedoch nicht nur tautologisch, sondern auch wahrheitsübertragend: ist die zur Anwendung gebrachte Regel gültig, dann ist nämlich auch das Ergebnis der Regelanwendung gültig. Deduziert man, dann hat man sich entschlossen, das zu Untersuchende als Wiederkehr des Bekannten und Bewährten anzusehen. Obwohl die Deduktion innerhalb empirischer Forschung (das gilt nicht für die Theorie) offiziell als wenig sinnvoll angesehen wird, freut sie sich dennoch großer Beliebtheit, sowohl bei den quantitativen wie auch bei den qualitativen Forschern und Forscherinnen.

### **2.3.2 Quantitative und Qualitative Induktion**

Eine zweite Art der Auswertung besteht darin, die im Datenmaterial vorgefundenen Merkmalskombinationen zu einer Ordnung oder Regel zu „verlängern“, zu generalisieren. Ausgehend von der Beobachtung: „Bei den Einbrüchen a, b und c ist der Medizinschrank geplündert worden.“ und der Fallkenntnis: „Herr Müller beging die Einbrüche a, b und c.“ wird der Schluss gezogen: „Herr Müller plündert bei Einbrüchen immer den Medizinschrank“.

Die logische Form dieser gedanklichen Operation ist die der quantitativen Induktion. Sie überträgt die quantitativen Eigenschaften einer Stichprobe auf die Gesamtheit, sie „verlängert“ den Einzelfall zu einer Regel. Quantitative Induktionen sind also (streng genommen) ebenfalls tautologisch (da sie keine neue Idee in die Welt bringen), jedoch

nicht wahrheitsübertragend. Die Resultate dieser Form des Schlussfolgerns sind lediglich wahrscheinlich.

Eine besondere Variante der induktiven Bearbeitung der Daten besteht nun darin, bestimmte qualitative Merkmale der untersuchten Stichprobe so zusammenzustellen, das diese Merkmalskombination einer anderen (bereits im Wissensrepertoire der Interaktionsgemeinschaft vorhandenen) in wesentlichen Punkten gleicht. In diesem Fall kann man den bereits existierenden Begriff für diese Kombination benutzen, um die ‚eigene‘ Form zu benennen. Die logische Form dieser Operation ist die der qualitativen Induktion. Sie schließt von der Existenz bestimmter qualitativer Merkmale einer Stichprobe auf das Vorhandensein anderer Merkmale (z.B.: Ich sehe hier am Tatort eine bestimmte Spurenlage. In sehr vielen Elementen stimmt sie mit dem Spurenmuster von Müller überein. Schluss: Müller ist der Spurenleger). Die qualitative Induktion schließt also, und das ist entscheidend, von zwei bekannten Größen, nämlich Resultat und Regel, auf den Fall. Der beobachtete Fall (token) ist ein Exemplar einer bekannten Ordnung (type).

Kurz: Schließt die quantitative Induktion von den quantitativen Eigenschaften einer Stichprobe auf die Gesamtheit, so ergänzt die qualitative Induktion dagegen die wahrgenommenen Merkmale einer Stichprobe mit anderen, nicht wahrgenommenen. Nur in diesem Sinne überschreitet diese Art der Induktion die Grenzen der Erfahrung – nämlich lediglich die Erfahrung mit der in Frage stehenden Stichprobe. Kenntniserweiternd ist dieser Schluss nur insofern, als er von einer begrenzten Auswahl auf eine größere Gesamtheit schließt. Neues Wissen (im strengen Sinne) wird auf diese Weise nicht gewonnen, bekanntes lediglich ausgeweitet. Die qualitative Induktion ist ebenfalls kein gültiger, sondern ein nur wahrscheinlicher Schluss – allerdings besitzt sie den Vorteil, dass sie sich (wenn auch nicht einfach) operationalisieren lässt. Allen wissenschaftlichen Verfahren, die in den erhobenen Daten nur neue Formen des bereits Bekannten erkennen, liegt die qualitative Induktion zugrunde.

### 2.3.3 Abduktion

Die dritte (der qualitativen Induktion um Verwechseln ähnliche, aber dennoch völlig verschiedene) Art der Datenbearbeitung besteht nun darin, aufgrund der Ausdeutung der erhobenen Daten solche Merkmalskombinationen zusammenzustellen bzw. zu entdecken, für die sich im bereits existierenden Wissensvorratslager keine entsprechende Erklärung oder Regel findet. Dann ist man überrascht. Wirkliche Überraschung löst echtes Erschrecken aus – und die Suche nach der (neuen) Erklärung. Da kein passender „type“ zu finden ist, muss in einem geistigen Prozess ein neuer er- bzw. gefunden werden. Manchmal erlangt man aufgrund eines gedanklichen Prozesses eine solche neue Erkenntnis, und wenn, dann stellt sie sich „blitzartig“ ein, und der gedankliche Prozess ist nur „sehr wenig von logischen Regeln behindert“ (Peirce 1976).

Eine Ordnung, eine Regel ist bei diesem Verfahren also erst noch zu (er)finden – und zwar mit Hilfe einer geistigen Anstrengung. Etwas Unverständliches wird in den Daten

vorgefunden und aufgrund des geistigen Entwurfs einer neuen Regel wird die Regel gefunden bzw. erfunden und zugleich klar, was der Fall ist. Die logische Form dieser Operation ist die der Abduktion. Hier hat man sich (wie bewusst auch immer und aus welchen Motiven auch immer) entschlossen, der bewährten Sicht der Dinge nicht mehr zu folgen.

Eine solche Bildung eines neuen „types“, also die Zusammenstellung einer neuen typischen Merkmalskombination ist ein kreativer Schluss, der eine neue Idee in die Welt bringt. Diese Art der Zusammenschließung ist nicht zwingend, eher sehr waghalsig. Die Abduktion „schlussfolgert“ also aus einer bekannten Größe (=Resultat) auf zwei unbekannte (=Regel und Fall). Die Abduktion ist also ein mentaler Prozess, ein geistiger Akt, ein gedanklicher Sprung, der das zusammenbringt, von dem man nie dachte, dass es zusammengehört. Das Ergebnis einer Abduktion ist eine sprachliche Hypothese. Solche Hypothesen können wahrscheinlich oder naheliegend sein. Aber die Wahrscheinlichste ist nicht die Beste.

Abduktionen ereignen sich, sie kommen so unerwartet wie ein Blitz („flash“), sie lassen sich nicht willentlich herbeizwingen, und sie stellen sich nicht ein, wenn man gewissenhaft einem operationalisierten Verfahrensprogramm folgt. Begleitet wird die Abduktion von einem angenehmen Gefühl, das überzeugender ist als jede Wahrscheinlichkeitsrechnung. Leider irrt dieses gute Gefühl nur allzu oft. Abduktionen resultieren also aus Prozessen, die nicht rational begründ- und kritisierbar sind. Aber auf dem Weg zu neuen Erkenntnissen gibt es zur Abduktion keine Alternative.

### **2.3.4    Abduktion, Deduktion, Induktion – oder: die Logik der Forschung**

Die Abduktion sucht (wie oben gesagt) angesichts überraschender Fakten nach einer sinnstiftenden Regel, nach einer möglicherweise gültigen bzw. passenden Erklärung, welche das Überraschende an den Fakten beseitigt, weil wir es jetzt verstehen. Endpunkt dieser Suche ist eine (sprachliche) Hypothese. Ist diese gefunden, beginnt in der Regel (sowohl in der quantitativen als auch in der qualitativen Forschung) ein mehrstufiger Überprüfungsprozess.

Besteht die erste Stufe des wissenschaftlichen Erkenntnisprozesses in der Findung einer Hypothese mittels Abduktion, dann besteht die zweite aus der Ableitung von Voraussagen aus der Hypothese, also einer Deduktion, und die dritte in der Suche nach Fakten, welche die Vorannahmen „verifizieren“, also einer Induktion. Sollten sich die Fakten nicht finden lassen, beginnt der Prozess von neuem, und dies wiederholt sich so oft, bis die „passenden“ Fakten erreicht sind. Die Logik der Forschung ist also immer eine dreistufige Forschungslogik – bestehend in dem wiederholten Nacheinander von Abduktion, Deduktion und Induktion.

## 2.4 Abschied vom Elfenbeinturm oder: Wissenschaft ist immer auch Partei

Verschiedene Sozialwissenschaften haben in den letzten Jahrzehnten eine Vielzahl von Disziplinierungen menschlichen Handelns nicht nur nachgewiesen, sondern sie haben gerade in den modernen wie nachmodernen Zeiten auch einen wesentlichen Anteil daran, diese Disziplinierungen zu legitimieren und durchzusetzen. Sie haben deswegen einen beachtlichen Anteil an den Disziplinierungen, weil die Begriffe, Einsichten, Ratschläge und die Politiken z.B. der Linguistik, der Sprachphilosophie, der Sozialwissenschaft, der Psychologie und der Pädagogik (um nur die Wichtigsten zu nennen) ihren Gegenstand nicht nur „wertneutral“ beschreiben, sondern ihn immer wieder durchdringen und somit verändern, ihn an die wissenschaftliche Deutung immer wieder anpassen. Manchmal oder besser: oft waren diese Veränderungen (also z.B. die Bestimmung der kommunikativen Kompetenz und deren Eindringen in die Schulcurricula) keine Kollateralschäden, also nicht eine unbeabsichtigte und (un-)vorhersehbare Folge wissenschaftlichen Tuns, sondern ausdrücklich intendiert. Auch für die Soziologie gibt und gab es in dieser Sache kein „unschuldiges Sprechen“.

Jede Wissenschaft (ob an Universitäten oder in Privatunternehmen betrieben) lebt in der jeweiligen Gesellschaft und lebt von ihr (finanziell wie inhaltlich). Als solche steht sie nicht außerhalb der Gesellschaft, sondern ist einerseits Ergebnis gesellschaftlicher Entwicklungen und gestaltet andererseits immer auch Gesellschaft durch ihre Arbeit und die Publikation ihrer Ergebnisse mit. Jede Wissenschaft kommuniziert mit der Gesellschaft, die sie erforscht und über die sie schreibt. Und: Sie gibt nicht nur, sondern bevor sie anfängt, hat sie schon sehr viel erhalten. Wissenschaft ist auch Ausdruck der Kultur einer Gesellschaft. Kurz: Wissenschaft hat stets auch gesagt, was das Subjekt, die Gesellschaft, Normen, Werte, Biographien sind und sein sollten und hat dabei vor allem eigene Wertvorstellungen und Interessen vertreten. Die Wissenschaft war von Beginn an Partei und nicht uninteressierte Beobachterin im Elfenbeinturm. Als solche hat sie auch, ob sie das will oder nicht, Verantwortung für die Gesellschaft, für die sie schreibt. Der sollte sie sich bewusst sein bzw. wieder bewusst werden.

## Literatur

- Breuer, Franz (2005): Konstruktion des Forschungsobjekts durch methodischen Zugriff.  
In: Mey, Günter (Hg.): Handbuch Qualitative Entwicklungspsychologie. Köln: Kölner  
Studien Verlag. 57-102
- Diekmann, Andreas (2007): Empirische Sozialforschung. Reinbeck: Rowohlt
- Esser, Hartmut (2007): Soll das denn alles (gewesen) sein? In: Soziale Welt 58: 351-358
- Flick, Uwe/von Kardoff, Ernst/Steinke, Ines (Hg.) (2004): Qualitative Forschung. Reinbek:  
Rowohlt
- Häder, Michael (2010): Empirische Sozialforschung. Wiesbaden: VS Verlag
- Kelle, Udo(2008): Die Integration qualitativer und quantitativer Methoden in der empiri-  
schen Sozialforschung. Wiesbaden: VS Verlag
- Kromrey, Helmut (2009): Empirische Sozialforschung. Stuttgart: UTB
- Lepenies, Wolf (1985): Die drei Kulturen. München: Hanser
- Lüders, Christian/Reichertz, Jo (1986): Wissenschaftliche Praxis ist, wenn alles funktioniert  
und keiner weiß warum. Bemerkungen zur Entwicklung qualitativer Sozialforschung.  
In: Sozialwissenschaftliche Literaturrundschau 12: 90-102
- Peirce, Charles S. (1976): Schriften zum Pragmatismus und Pragmatizismus. Frankfurt  
a.M.: Suhrkamp
- Popper, Karl (1935): Logik der Forschung. Wien
- Przyborski, Aglaja/Wohlrab-Sahr, Monika (2008): Qualitative Sozialforschung. München:  
Oldenbourg
- Reichenbach, Hans (1938): Erfahrung und Prognose. Braunschweig: Vieweg
- Reichertz, Jo (2007): Qualitative Sozialforschung. In: EWE 18: 195-208
- Reichertz, Jo (2013): Die Abduktion in der qualitativen Sozialforschung. Wiesbaden: VS  
Verlag
- Reichertz, Jo/Englert, Carina J. (2011): Einführung in die qualitative Videoanalyse.  
Wiesbaden: VS Verlag
- Schütz, Alfred (2010): Zur Methodologie der Sozialwissenschaften. Konstanz: UVK
- Snow, Charles P. (1965): The Two Cultures: and A Second Look. Cambridge: Cambridge  
University Press
- Soeffner, Hans-Georg (1988): Kulturmythos und kulturelle Realität(en). In: ibidem (Hg.):  
Kultur und Alltag. Göttingen: Schwarz Verlag. 3-20
- Strübing, Jörg (2013): Qualitative Sozialforschung. München: Oldenbourg Verlag
- Weber, Max (1973): Wirtschaft und Gesellschaft. Tübingen: Mohr Siebeck
- Wittgenstein, Ludwig (1977): Philosophische Untersuchungen. Frankfurt a.M.: Suhrkamp

Jürgen Friedrichs

Ethische Probleme treten vor allem bei der Befragung, dem Experiment und der teilnehmenden Beobachtung auf. Ich beschränke mich im Folgenden weitgehend auf die Befragung, weil sie am häufigsten verwendet wird.

---

## 3.1 Informierte Einwilligung

Bei der Befragung, sei sie face-to-face, telefonisch oder schriftlich, tritt das ethische Problem bereits bei der Formulierung des Anschreibens auf. Gemeinhin führen wir allgemeine Gründe für die jeweilige Studie an, z.B. den, die „Lebensbedingungen in Köln-Mülheim“ erforschen zu wollen; sagen aber nicht, dass wir auch nach Kriminalitätsfurcht und beobachtetem abweichendem Verhalten fragen werden. Wir fürchten, die vollständige Liste der Gründe, warum wir die Studie durchführen, würde die Befragten abschrecken und nicht kooperieren lassen. Wir informieren aber auch nicht über Skalen zur sozialen Erwünschtheit. Die halbe Wahrheit ist aber auch ein unethisches Verhalten. Das alles geschieht im Namen der höheren „Wahrheit“, der Validität der Ergebnisse der Studie.

Nun ist die vollständige oder zumindest angemessene Information der Befragten die Voraussetzung dafür, dass sie entscheiden können, ob sie an der Befragung, der (teilnehmenden) Beobachtung oder dem Experiment teilnehmen wollen (zu dem damit verbundenen Problem der Vertraulichkeit s. weiter unten). Der Willkür der Forscher/innen sind aber deutliche Grenzen gesetzt. Die Standes-Organisationen haben längst Statuten verabschiedet, die für die Disziplin bindend sind. So schreibt der Ethik-Kodex der Deutschen Gesellschaft für Soziologie und des Berufsverbandes Deutscher Soziologen vom 27. 11. 1992 vor (ähnlich ICC/ESOMAR 2007: Art. 3a, 4b): „Generell gilt für die Beteiligung an sozialwissenschaftlichen Untersuchungen, dass diese freiwillig ist und auf der Grundlage einer möglichst ausführlichen Information über Ziele und Methoden des entsprechenden Forschungsvorhabens erfolgt. Nicht immer kann das Prinzip der informierten Einwilligung

in die Praxis umgesetzt werden, z.B. wenn durch eine umfassende Vorabinformation die Forschungsergebnisse in nicht vertretbarer Weise verzerrt würden. In solchen Fällen muss versucht werden, andere Möglichkeiten der informierten Einwilligung zu nutzen.“ (I B)

Der Text lässt offen, wann genau eine vollständige Information nicht erforderlich ist. Deutlicher wird dies in den „Ethischen Richtlinien der Deutschen Gesellschaft für Psychologie e.V. und des Berufsverbands Deutscher Psychologinnen und Psychologen e.V.“ von 1998 (mit Änderungen von 2005): „Beim Einholen der auf Aufklärung basierenden Einwilligung klären Psychologen die teilnehmenden Personen über folgende Sachverhalte auf: (1) den Zweck der Forschung, die erwartete Dauer der Untersuchung und das Vorgehen; (2) ihr Recht darauf, die Teilnahme abzulehnen oder sie zu beenden, auch wenn die Untersuchung schon begonnen hat; (3) absehbare Konsequenzen der Nicht-Teilnahme oder der vorzeitigen Beendigung der Teilnahme; (4) absehbare Faktoren, von denen man vernünftigerweise erwarten kann, dass sie die Teilnahmebereitschaft beeinflussen, wie z.B. potenzielle Risiken, Unbehagen oder mögliche anderweitige negative Auswirkungen, die über alltägliche Befindlichkeitsschwankungen hinausgehen; (5) den voraussichtlichen Erkenntnisgewinn durch die Forschungsarbeit; (6) die Gewährleistung von Vertraulichkeit und Anonymität sowie ggf. deren Grenzen; (7) Bonus für die Teilnahme und (8) an wen sie sich mit Fragen zum Forschungsvorhaben und zu ihren Rechten als Forschungsteilnehmer wenden können. Den potenziellen Teilnehmern und Teilnehmerinnen wird die Gelegenheit gegeben, Antworten auf ihre Fragen zum Forschungsvorhaben zu erhalten.“ (C III, 3, d)

Sehr ähnlich lautet die Formulierung der American Sociological Association (ASA): „When informed consent is required, sociologists inform research participants or their legal representatives of the nature of the research; they indicate to participants that their participation or continued participation is voluntary; they inform participants of significant factors that may be expected to influence their willingness to participate (e.g., possible risks and benefits of their participation); and they explain other aspects of the research and respond to questions from prospective participants. Also, if relevant, sociologists explain that refusal to participate or withdrawal from participation in the research involves no penalty, and they explain any foreseeable consequences of declining or withdrawing sociologists explicitly discuss confidentiality ...“ (12.02d)

Ein Problem, das genauerer Bearbeitung bedarf, ist das der Einwilligungen. Bei Umfragen in Schulen müssen nicht nur die Schulen, sondern auch die Eltern der Schüler/innen einwilligen, dass ihre Kinder befragt werden. Noch komplizierter wird es, wenn man ein Panel mit Schülern und Jugendlichen aufbaut, und Jugendliche, die das 18. Lebensjahr erreicht haben, nun verlangen, dass nachträglich ihre Daten gelöscht werden.<sup>1</sup>

---

<sup>1</sup> Persönliche Information durch Walter Bien, Deutsches Jugendinstitut, München.

### 3.2 Täuschungen

Wenn wir diese Regeln streng auslegen, dann wären z.B. die Elektroschock-Experimente von Milgram (1963, 1974) nicht ethisch gerechtfertigt gewesen. In dem Experiment wurden Personen aufgefordert, an einem Lernexperiment teilzunehmen. Die Personen sollten den Lernfortschritt bei Wortpaaren kontrollieren. Die kontrollierende Person (P) war in Wirklichkeit die eigentliche Versuchsperson (Vp), die lernende Person ein Konfident (K) des Versuchsleiters (VL). Machte K einen Fehler, erhielt er einen Elektroschock; die Skala reichte von 15 bis XXX (450 Volt). Nach jedem Fehler wurde der Schlag um 15 Volt erhöht. Entsprechend regierte K mit zunehmenden Körperverrenkungen und Schreien, ab 300 Volt verstummte er/sie. Willigte die Vp nicht ein, höhere Schläge anzuwenden, forderte ihn der VL auf, weiter zu machen. In Wirklichkeit erhielt K keine Schläge sondern simulierte nach einem mit dem VL abgesprochenen Ablauf. Von den 40 Probanden des ersten Experiments gingen 65 % der Vp bis zur höchsten Voltzahl – bei der K längst verstummt war. Das Experiment wurde unter drei Versuchsbedingungen durchgeführt: a) Vp hört und sieht die Reaktionen von K, b) Vp hört nur die Reaktionen, c) Vp hört und sieht keine Reaktionen von K. Die Bereitschaft, extreme Schläge zu vergeben, stieg deutlich von (a) nach (c). Folgerung: Je weniger der Aggressor die Reaktionen eines Opfers erfährt, desto bereitwilliger führt er aggressive Akte (legitimiert durch eine Autorität) aus.

Eine korrekte Information der Vp hätte nie zu den (erschreckenden) Ergebnissen – also der „Realität“ - geführt. Dennoch sind wohl die allermeisten Forscher/innen der Ansicht, die Experimente hätten sehr wichtige Ergebnisse erbracht. Wir müssen demnach festlegen, unter welchen Bedingungen wir Ausnahmen von den Regeln zulassen, ohne dass generell der Zweck die Mittel heiligt. Solche Ausnahmen sehen die Regeln der DGP vor: „Psychologen führen keine Studie auf der Basis von Täuschung durch, es sei denn, sie sind nach gründlicher Überlegung zu dem Schluss gekommen, dass der Einsatz von Täuschungs-techniken durch den voraussichtlichen bedeutsamen wissenschaftlichen, pädagogischen oder praktischen Erkenntnisgewinn gerechtfertigt ist und dass geeignete alternative Vor-gehensweisen ohne Täuschung nicht zur Verfügung stehen.“ (C, III, 8)

Noch deutlicher wird es in den Richtlinien der ASA formuliert: „Despite the para-mount importance of consent, sociologists may seek waivers of this standard when (1) the research involves no more than minimal risk for research participants, and (2) the research could not practicably be carried out were informed consent to be required. Sociologists recognize that waivers of consent require approval from institutional review boards or, in the absence of such boards, from another authoritative body with expertise on the ethics of research.“ (12.01b)

Es sind also Ausnahmen dann möglich, wenn a) die Ergebnisse ungewöhnlich, b) gesellschaftlich nützlich sind und c) auf andere Weise nicht gewonnen werden können. Das wird man sicherlich von Fall zu Fall entscheiden müssen. Hilfreich ist in diesem Zusam-menhang hierzu das Urteil des Bundesverfassungsgerichts vom 25.1.1984. Hier hatte der Springer-Verlag gegen Günter Wallraff geklagt, weil er unter dem Decknamen Hans Esser drei Monate in der Redaktion von Bild-Hannover gearbeitet und dann in einer Publikation

wörtlich aus Redaktionssitzungen zitiert hatte. Das BVerfG urteilte: „Derart gewichtige Einbrüche in die Grundrechtssphäre könnten nicht schon durch das bloße Interesse der Öffentlichkeit an der Aufdeckung gewichtiger Missstände, sondern nur bei Vorliegen gravierender Rechtsbrüche gerechtfertigt werden.“ (Telemedicus)

Wenn demnach die Forscher „gravierende Rechtsbrüche“ vermuten können, ist es nicht erforderlich, die Versuchspersonen oder Befragten (angemessen) aufzuklären. Dieser „informed consent“ fehlte im folgenden Fall. Der Pädagoge und Soziologe Heinz Geißlinger hat 1992 seine Dissertation mit dem Titel „Die Imagination der Wirklichkeit“ veröffentlicht (vgl. auch: Stenger/Geißlinger 1991). Darin werden Reisen von Jugendlichen im Alter von 10 - 12 Jahren geschildert, die in ein Feriendorf an der tschechischen Grenze fahren, finanziert vom Bezirksamt Berlin-Kreuzberg, organisiert von einer Firma namens Story Deal AG. Sie sollten etwas erleben, wozu die begleitenden Pädagogen eine Reihe von Täuschungen vornahmen, z.B. in einer Studie, die Sie in der Kölner Zeitschrift aus dem Jahre 1991 nachlesen können, wird den Jugendlichen vorgegaukelt, sie hätten Träume, in denen drei Jugendliche instruiert werden, sie hätten von merkwürdigen Lebewesen geträumt. Außerdem sei ein merkwürdiger Geruch zu spüren. Dieser Geruch ist künstlich erzeugt worden, den Jugendlichen wurde jedoch gesagt, er käme von einem bestimmten Stein. Daraufhin haben mehr Jugendliche berichtet, sie träumten davon. Dann hat man einen Wünschelrutengänger geholt, der sagen sollte, hier seien Wasserleitungen, schließlich hat man die Jugendlichen, die immer mehr eine Realität konstruierten, in den Wald geführt, hat in einen Baum ein Licht eingebaut, so dass die Jugendlichen den Eindruck hatten, hier gäbe es ein unheimliches Tier. In anderen dieser Reisen hat man den Jugendlichen vorgegaukelt, es gäbe Saurier. Diese Reisen sind dann von der Berliner Presse aufgegriffen und Herr Geißlinger ist dafür kritisiert worden.

Mit diesem Fall hat sich dann die Ethikkommission der Deutschen Gesellschaft für Soziologie unter dem Vorsitz von Stefan Lamnek beschäftigt. Es hat eine unangenehme Kontroverse zwischen dem Autor Geißlinger und der Ethikkommission gegeben (Geißlinger 1994, Lamnek 1994a, 1994b). Geißlinger hat ausgeführt, er habe hier die Jugendlichen ja nur als Pädagoge getäuscht und es wäre eine Erlebnisreise für sie gewesen, man könne dagegen nichts einwenden. Er hat des Weiteren argumentiert, und zwar auf der wissenschaftlichen Ebene, hier könne man die These des Konstruktivismus nachvollziehen, die ja besagt, dass Realität von uns konstruiert wird. Man habe die Jugendlichen nur dazu gebracht, eine solche Realität zu konstruieren, zudem habe es doch auch einen hohen Erlebniswert für sie gehabt.

Die Ethikkommission hat drei Einwände gegen diese Studie formuliert. 1. Weder die Schüler noch die Eltern waren über die Täuschungsmanöver der Story Deal AG informiert. Es gab nicht das, was man im Code of Ethics als „informed consent“ bezeichnet. 2. Weder die Eltern noch die Schüler sind nachträglich darüber informiert worden, dass hier eine Täuschung vorlag. Dies ist bedeutsam, weil einige Schüler immer noch ihren Freunden erzählten, es gäbe Saurier. 3. Die Argumentation von Geißlinger, er habe als Pädagoge gehandelt und später als Wissenschaftler seine Dissertation geschrieben, könne man nicht akzeptieren, denn es gäbe eine zeitliche Überlappung. Er könne daher wohl keine

Rollentrennung derart vornehmen, dass er zunächst als Pädagoge Experimente durchführt, die sozusagen keinem Code of Ethics unterliegen, um sie dann als Wissenschaftler neutral unter den Rahmen eines konstruktivistischen Ansatzes zu stellen. Hier sind zweifellos ethische Verpflichtungen verletzt worden. Zudem kann man auch fragen, welches der Erkenntnisgewinn dieser Studie ist.

---

### 3.3 Unethische Fragen?

Aber auch einzelne Fragen können als „unethisch“ interpretiert werden. Ist es unethisch, Sozialhilfe-Empfänger (SGB II oder SGB XII) nach der Höhe ihres Budgets und dessen Verwendung zu fragen? Sollen wir Muslime nicht nach dem Grad ihrer Religiosität mit einer eigens dafür entwickelten Skala befragen? Ein einschlägiges Beispiel ist eine Befragung von Schüler/innen der achten Klasse in 40 allgemeinbildenden Schulen in Nürnberg im Jahre 2004.<sup>2</sup> Ein Elternteil wendet sich an den Landesbeauftragten für den Datenschutz in Bayern; der findet in der Studie erhebliche Mängel. Zuerst kritisiert er das Vorgehen: Es würden personenbezogene Daten erhoben, darunter Alter, Geburtsmonat, Nationalität, Religionszugehörigkeit, Körpergröße und Körbergewicht. Dies erlaube, durch „Zusatzwissen“ die Befragten zu „reidentifizieren“. Zudem wurden Daten über die Zahl der Räume in der Wohnung, die Zahl der Urlaubsreisen und die der Autos in der Familie erhoben. Das bringt den Beauftragten zu dem Schluss: „Auf Grund dieser detaillierten Einzelangaben konnte ich – jedenfalls in Einzelfällen – nicht ausschließen, dass eine Reidentifikationsmöglichkeit zu Lasten einer(s/r) Betroffenen und damit auch seiner/ihrer Eltern bestand.“

Beanstandet wurden auch Fragen an Mädchen wie „Hast Du schon einmal mit einem Jungen geschlafen?“ – „Wie habt Ihr beim ersten Mal verhütet?“ – „Wissen Deine Eltern davon, dass Du mit einem Jungen geschlafen hast?“ (Wahrscheinlich hätte der bayerische Datenschutzbeauftragte bei der Lektüre der Kinsey-Fragebögen aus den Jahren 1948 und 1955 beide Studien verboten.)

Ferner erhebt er Einwände gegen die Einwilligung. „Zunächst ist festzustellen, dass die den Eltern eingeräumte Möglichkeit, Widerspruch gegen die Teilnahme ihres Kindes an der Klassenraumbefragung zu erheben, die fehlende Einwilligung nicht ersetzen kann. Im Übrigen wurden die Eltern in der ‚Elterninformation‘ gerade nicht darüber informiert, dass mit der Umfrage auch Daten über sie selbst erhoben wurden. Folglich bezog sich das Widerspruchsformular gerade nicht hierauf; eine rechtswirksame Einwilligung der Eltern lag somit schon deswegen nicht vor. Auch wurde keine datenschutzrechtlich wirksame Einwilligung der Schüler eingeholt. Bei Schülerinnen und Schülern der Achten Klassen ist grundsätzlich anzunehmen, dass sie selbst über ihr Recht auf informationelle Selbstbestimmung entscheiden können. Damit war auch ihre Einwilligung erforderlich.“

---

<sup>2</sup> Persönliche Information von Reinhard Wittenberg, Universität Erlangen-Nürnberg, Lehrstuhl für Soziologie und empirische Sozialforschung.

Nicht immer wird ein Datenschutzbeauftragter so rigide urteilen. Schon gar nicht wird man das Argument, durch „Zusatzwissen“ könne man die Befragten identifizieren, ohne weiteres zulassen dürfen. Es sind sicherlich viele Wege denkbar, sich „Zusatzwissen“ zu beschaffen – nur wer tut das? Wenn die Forscher für die Daten verantwortlich sind, dann ist ein solches Vorgehen ethisch nicht vertretbar. Schon deshalb kann die bloße Möglichkeit nicht als Grund dafür herhalten, im Prinzip jede Studie zu unterbinden. Hier müssen die Standesverbände sich gegen die Datenschützer durchsetzen.

---

### 3.4 Vertraulichkeit

Ob Befragte einwilligen, an einer Studie teilzunehmen, hängt auch davon ab, ob man ihnen Vertraulichkeit zusichert. Die Forderung tritt häufig für Studien zu abweichendem Verhalten auf. Die Forscher befinden sich nun in einer schwierigen rechtlichen Situation, denn sie könnten vor Gericht gezwungen werden, ihre Informanten preis zu geben – oder aber mit nachteiligen Folgen für sich diese Informationen zu schützen.

Dieses Problem wird besonders deutlich an folgender Richtlinie des Ethik-Codes der ASA: „Confidential information provided by research participants, students, employees, clients, or others is treated as such by sociologists even if there is no legal protection or privilege to do so. Sociologists have an obligation to protect confidential information and not allow information gained in confidence from being used in ways that would unfairly compromise research participants, students, employees, clients, or others.“ (11.01b)

In der ursprünglichen Fassung von 1989 endete der erste Satz mit „...or privilege or legal force is applied“. Das hätte bedeutet, dass der Forscher sich auch nicht einer gerichtlichen Anweisung verpflichtet fühlt, also gegen das Gesetz verstößt und die Vertraulichkeit wahrt. Ohne den Zusatz (also in der Version von 1997) bleibt es dem Forscher überlassen, wie weit er sich an sein Versprechen, Informationen vertraulich zu behandeln, gegenüber den Teilnehmer/innen gebunden fühlt. Hierüber ist es zu einer Kontroverse zwischen Soziologen gekommen (Palys/Lowman 2002, Lindgren 2002, Stone 2002, Lowman/Palys 2003). Palys/ Lowman fordern, die Teilnehmer/innen stärker zu schützen. Sie plädieren daher für ein Forscher-Teilnehmer-Privileg, das die Forscher auch schon in der Anlage ihrer Studie – also auch dem versprochenen Vertrauen – berücksichtigen müssen (Palys/ Lowman 2002: 7, 13). Ferner solle von Fall zu Fall darüber entschieden werden, ob das Privileg gelte.

Dazu könnte man auch die Kriterien des Wigmore-Tests heranziehen, die Bundes- und Bundesstaaten-Gerichte anwenden. Die vier Kriterien dieses Tests sind: 1. Die Kommunikation muss im Vertrauen darauf erfolgt sein, dass sie nicht offengelegt wird. 2. Die Vertraulichkeit ist ein wesentliches Merkmal der Beziehung zwischen den Parteien. 3. Die Beziehung sollte in den Augen der Gemeinschaft stark unterstützt werden. 4. Der Nachteil, den eine Veröffentlichung der Kommunikation nach sich zöge, muss größer sein als der Nutzen der korrekten Durchführung des Gerichtsverfahrens (übersetzt nach Palys/Lowman 2002: 7). Die Gegner dieser Position haben den Autoren u.a. vorgeworfen, sie stellten

sich über das Recht (Lindgren 2002, Stone 2002) und gäben gegenüber den Teilnehmern vor, sie hätten die rechtliche Autorität, Vertraulichkeit zuzusichern (Stone 2002).

Diese Kontroverse zeigt sehr deutlich, dass die Forscherin oder der Forscher die Verantwortung für seine Zusagen übernehmen muss. Streng genommen kann sie oder er nur dann Vertraulichkeit zusichern, wenn sie oder er auch vor Gericht nicht bereit ist, Namen oder Informationen preiszugeben. Offensichtlich fehlt der Soziologie eine Regelung (oder ein Privileg) analog dem Informantenschutz im Journalismus oder dem Beichtgeheimnis in der katholischen Kirche.

Besonders riskant wird diese Gratwanderung, wenn man eine teilnehmende Beobachtung vornimmt. Begleitet ein Beobachter die Polizei auf dem Revier und ihren Einsatzfahrten; so gilt er nachträglich als „Spion“, wenn er seine Ergebnisse publiziert und damit die (vollen) Absichten des teilnehmenden Beobachters offengelegt werden (Feest/Blankenburg 1972). Ist das ethisch zu rechtfertigen?

Ebenso kann der Forscher sich einer Randgruppe widmen. So führte Rüdiger Lautmann im Jahre 1990 in einer Dunkelfeld-Studie Interviews mit 60 pädophilen Männern durch. Der Fall kam vor die Ethik-Kommission der DGS, die es aus formalen Gründen ablehnte, sich damit zu beschäftigen; dennoch gab es einen Brief des Vorsitzenden an Lautmann, er solle alle Ämter niederlegen.<sup>3</sup>

Hier ist der/die Forscher/in ja Teil des Geschehens, damit unter Umständen auch Mitwisser oder gar Mitschuldiger krimineller Handlungen oder deren Vorbereitung. Auch hier gilt, dass eine Handlung vorliegen muss, die eine besondere gesellschaftliche Bedeutung hat, über die die Forscher aufklären wollen: Ein gesellschaftlicher Missstand soll durch die Publikation der Ergebnisse behoben werden. Unter welchen Umständen genau das gerechtfertigt ist, kann vermutlich nur von Fall zu Fall durch Ethik-Kommissionen der jeweiligen Disziplin entschieden werden.

Diese Regel dürfte auch für die Beobachtung gelten. Auf diese Methode wird in den Ethik-Codes kaum eingegangen, vermutlich, weil sie zumindest in der Soziologie sehr selten angewendet wird. Eine Beobachtung auf privatem Gelände ist ohne Einwilligung der Grundstücksbesitzer nicht möglich, mithin kann man nicht ohne weiteres auf einem Schulhof, einem militärischen Übungsplatz oder in einer Kantine beobachten. Es tritt das unter Punkt (1) aufgeführte Dilemma von Information und Einwilligung auf. Auch hier ist zu fragen, ob der Nutzen der Ergebnisse so groß ist, dass er eine unzureichende Information rechtfertigt – mit dem zusätzlichen Risiko, dass die Betroffenen sich „betrogen“ fühlen, wenn die Ergebnisse veröffentlicht werden.

Ein Beispiel: „In den Jahren 2004 - 2006 wurde eine Studie bei Landwirten durchgeführt, in der der Verdacht untersucht werden sollte, ob die Landwirte bei der Ausbringung von Düngemitteln u.a. die Fehler machen, zu dicht an Randgewässer zu fahren oder die Düngemittel bei starkem Wind auszubringen, so dass Pestizide in Gewässer gelangen. Dies sollte durch eine Beobachtung z.T. auf den Äckern ermittelt werden, ohne dass die Landwirte darüber informiert wurden. Durch eine Indiskretion der an den Projektsitzungen

<sup>3</sup> Persönliche Information durch Rüdiger Lautmann, Universität Bremen.

Beteiligten gelangte die Information sowohl an die Landwirte (die sich empörten) als auch an die Bild-Zeitung, die unter der Schlagzeile „Bauernspione“ über das Projekt berichtete. Trotz formal korrektem Verlauf der bisherigen Projektarbeiten wurde das Projekt aufgrund der auch projektintern geführten Diskussion um das Für und Wider nach diesem öffentlichem Bekanntwerden schließlich eingestellt und die bis dahin erzielten Ergebnisse nicht verwertet.<sup>4</sup> Handelt es sich hier um unethisches Verhalten?

Wenn aber die Beobachtung außerhalb des Grundstücks oder aus der Ferne und ohne Kenntnis der Beobachteten erfolgt, ist sie dann ethisch gerechtfertigt? Kann man das Verhalten von Schüler/innen auf dem Pausenhof oder das Erziehungsverhalten von Müttern auf einem Kinderspielplatz von einem benachbarten Grundstück aus beobachten? Hängt die Antwort auf diese ethische Frage auch mit der Art der Aufzeichnung (Film, Notizen in Beobachtungsschema) zusammen? Hier sind sicherlich noch rechtliche Klärungen und Regeln einer professionellen Ethik erforderlich, oder rechtfertigt der erwartete Nutzen das Vorgehen?

---

### 3.5 Publikation

Die ethischen Prinzipien beziehen sich selbstverständlich nicht nur auf den Prozess der Forschung sondern auch auf die Publikation. Dazu gehört, alle an dem Projekt beteiligten Mitarbeiter/innen aufzuführen. Die ASA führt folgende Regeln an:

„Sociologists ensure that principal authorship and other publication credits are based on the relative scientific or professional contributions of the individuals involved, regardless of their status.“ (15b)

„When sociologists publish data or findings that they have previously published elsewhere, they accompany these publications by proper acknowledgment.“ (16.02)

Diese ethische Forderung hat weit reichende Konsequenzen. Man könnte aus ihr ableiten, dass Autoren frühere und ähnliche Fassungen ihrer Ergebnisse als Sonderdruck ihrem neu eingereichten Manuskript beifügen. In meiner 20-jährigen Tätigkeit als Herausgeber der Kölner Zeitschrift hat das kein Autor getan – es war Aufgabe der Herausgeber, den Innovationswert des eingereichten Manuskriptes gegenüber den früher publizierten Aufsätzen zu bestimmen. Mit dem steigenden Publikationsdruck wird sich die Praxis verstärken, Aufsätze mit relativ geringen Veränderungen bei verschiedenen Zeitschriften einzureichen, deshalb ist eine solche Kontrolle umso wichtiger. Wer sich scheut, die anderen einschlägigen deutschen oder englischen Publikationen mit einzureichen, sollte vom Begutachtungsverfahren ausgeschlossen werden.

---

<sup>4</sup> Persönliche Information von Volker Müller-Benedict, Universität Flensburg.

### 3.6 Nicht-Diskriminierung

Eine weitere Forderung richtet sich darauf, offen für Kritik – auch an den eigenen Publikationen - zu sein. Dies ist besonders deutlich in den Richtlinien der DGP formuliert: „Die Anerkennung der wissenschaftlichen Leistungen Andersdenkender, Andersgläubiger, Angehöriger anderer Altersgruppen und des anderen Geschlechts, anderer sozialer Schichten und Kulturen, und die Bereitschaft, eigene Irrtümer durch überzeugende Argumente, welcher Herkunft auch immer, zu korrigieren, kennzeichnen das Berufsethos der in Forschung und Lehre tätigen Psychologen in besonderem Maße.“ (C, I, 2)

Wissenschaftliche Ethik muss sich aber auch gegen political correctness behaupten. Als James S. Coleman 1975 seine Forschungsergebnisse veröffentlichte, denen zufolge das als fortschrittlich gelobte schwarz-weiße Schulbus-Programm nicht die Integration in den öffentlichen Schulen förderte, sondern vielmehr zu einem Exodus von weißen Schülern führte, wurde Coleman von liberalen Soziologen und Politikern angegriffen. Es gab sogar eine Bestrebung, ihn aus der American Sociological Association auszuschließen. Das ist dann doch nicht geschehen. Und wie sich die Zeiten ändern: Coleman wurde 1991 Präsident der ASA.

## Literatur

- ADM (2008): Erklärung für das Gebiet der Bundesrepublik Deutschland zum ICC/ESOMAR Internationalen Kodex für die Markt- und Sozialforschung, [http://www.adm-ev.de/fileadmin/user\\_upload/PDFS/Kodex\\_D.pdf](http://www.adm-ev.de/fileadmin/user_upload/PDFS/Kodex_D.pdf)
- American Sociological Association (1999): Code of Ethics <http://www.asanet.org/images/asa/docs/pdf/CodeofEthics.pdf>
- Deutsche Gesellschaft für Psychologie (1998/2005): Ethische Richtlinien der Deutschen Gesellschaft für Psychologie e.V. und des Berufsverbands Deutscher Psychologinnen und Psychologen e.V. <http://www.dgps.de/dgps/aufgaben/003.php>
- Deutsche Gesellschaft für Soziologie (1992): Ethik-Kodex <http://www.soziologie.de/index.php?id=19>
- Feest, Johannes und Erhard Blankenburg (1972): Die Definitionsmacht der Polizei. Düsseldorf: Bertelsmann Universitätsverlag
- Geißlinger, Hans (1992): Die Imagination der Wirklichkeit. Frankfurt/M.-New York: Campus
- Geißlinger, Hans (1994): Zum Einzug der „Political Correctness“ durch die Hintertür des Ethik-Kodex: Richtigstellung einiger falscher Behauptungen der Ethik-Kommission im Bericht „Konfliktregelung im außerrechtlichen Raum“ in Heft 2/94: 29, 30 und in: Soziologie 4/1994: 72-75
- ICC/ESOMAR (2007): ICC/ESOMAR Internationalen Kodex für die Markt- und Sozialforschung [http://www.iccwbo.org/uploadedFiles/ICC/policy/marketing/Statements/ICCESOMAR\\_Code\\_German.pdf](http://www.iccwbo.org/uploadedFiles/ICC/policy/marketing/Statements/ICCESOMAR_Code_German.pdf)
- Lautmann, Rüdiger (1994): Die Lust am Kind. Portrait des Pädophilen. Hamburg: Ingrid Klein
- Lamnek, Siegfried (1994a): Bericht der Ethik-Kommission für das Jahr 1994. Sozialwissenschaften und Berufspraxis 17: 341-352
- Lamnek, Siegfried (1994b): Konfliktregelung im außerrechtlichen Raum. Ein Jahr Arbeit der Ethik-Kommission. Bericht der Ethik-Kommission für das Jahr 1994. In: Soziologie 2/1994: 22-42
- Lindgren, James (2002): Anticipating Problems: Doing Social Science Research in the Shadow of Law. In: Sociological Methodology 32: 29-32
- Lowman, John und Ted Palys (2003): Subject to the Law: Civil Disobedience, Research Ethics, and the Law of Privilege. In: Sociological Methodology 33: 381-389
- Milgram, Stanley (1963): Behavioral Study of Obedience. Journal of Abnormal and Social Psychology, 67: 371-378
- Milgram, Stanley (1974): Obedience to Authority. An Experimental View. New York. Harper
- Palys, Ted und John Lowman (2002): Anticipating Law: Research Methods, Ethics, and the Law of Privilege. In: Sociological Methodology 32: 1-17

- Stenger, Horst, und Hans Geißlinger (1991): Die Transformation sozialer Realität. Ein Beitrag zur empirischen Wissenssoziologie. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 43: 247-234
- Stone, Geoffrey R. (2002): Above the Law: Research Methods, Ethics, and the Law of Privilege. In: Sociological Methodology 32: 19-27
- Telemedicus: <http://www.telemedicus.info/urteile/Presserecht/Vertraulichkeit-der-Redaktionsarbeit/181-BVerfG-Az-1-BvR-27281-Wallraff-I.html> (BverfG 66,116)

Andreas Mühlichen

## 4.1 Einleitung

Im Jahre 1983 begab es sich, dass in ganz (West-)Deutschland das Volk gezählt werden sollte – doch dazu sollte es erst vier Jahre später kommen: Anstatt sich brav zählen zu lassen, probten die Bürger den Aufstand und Boykottaufrufe durchzogen das Land. Die politisch sensible Stimmung der Zeit, Angst vor dem gläsernen Bürger, Sorge, die Uhr ticken zu hören bis zum drohenden, nahenden „1984“ eines Orwellschen, dystopischen Überwachungsstaates, führte zu Klagen vor dem Bundesverfassungsgericht. Auch wenn die Volkszählung dann 1987 trotzdem durchgeführt wurde – immer noch unter Protest und Boykottaufrufen, aber wenigstens mit verändertem Fragebogen und besserem (und trotzdem unzureichendem) Schutz vor einer Deanonymisierung –, so wurde sie doch 1983 erst einmal gestoppt und in Folge des politischen Debakels, als wohl wichtigstes Ergebnis der Affäre, das sogenannte „Volkszählungsurteil“ von höchstrichterlicher Stelle gefällt. Um Zensusdaten ärmer, war die Bundesrepublik nun um das „Grundrecht auf informationelle Selbstbestimmung“ reicher – und um die Erkenntnis, dass es „unter den Bedingungen der automatischen Datenverarbeitung kein ‚belangloses‘ Datum mehr“ (BVerfG 1983: 45) gibt. Spätestens seit diesem Zeitpunkt sind in Deutschland der Datenschutz und die empirische Sozialforschung untrennbar miteinander verbunden.

All dies geschah, obwohl es nur staatlichen Institutionen und wenigen großen Unternehmen überhaupt möglich war, große Mengen maschinenlesbarer Daten zu gewinnen, zu verknüpfen und schließlich sinnvoll auszuwerten, und so etwas wie das Internet einer breiten Öffentlichkeit kaum bekannt war, geschweige denn, dass diese Zugang gehabt hätte. Trotz dieser, gerade aus heutiger Perspektive, massiven technischen Limitationen, wurde die Möglichkeit, maschinell Daten zu sammeln, als beängstigend empfunden, gerade in Deutschland, wo die Furcht vor dem „Big Brother“ eines übermächtigen Überwachungsstaates vielleicht wegen der Erfahrungen während der Zeit des NS-Regimes besonders ausgeprägt war. Ein wichtiges Argument der Gegner bezieht sich dabei auf die mangelnde

Anonymität der Zensusdaten: Eine Informatikerin aus dem Kreis von Klaus Brunnstein, einer prominenten Figur auf Seiten der Volkszählungsgegner und einer der Beschwerdeführer vor dem Bundesverfassungsgericht, konnte nachweisen, dass die Daten bis auf das enge Umfeld einzelner Personen zurückgeführt werden könnten (Brunnstein 2007) – was gerade durch die computergestützte Verknüpfung von Daten und weitergehender Analyse praktikabel wird.

---

## 4.2 Moores Law

Seither hat sich viel getan. Bereits seit Mitte der 1960er-Jahre ist bekannt, dass die Entwicklung von Mikroprozessortechnologien rasant und nichtlinear ist. Die Chips folgen (zumindest bisher) brav dem Mooreschen Gesetz (Mack 2011) und werden nach wie vor noch etwa alle 18 bis 24 Monate doppelt so leistungsfähig – was natürlich gleichzeitig bedeutet, dass immer komplexere Operationen mit immer mehr Daten durchgeführt werden können – man denke nur an die schneller durchführbaren, ausgefeilten multivariaten Verfahren (Blasius/Baur, Kapitel 79 in diesem Band), die empirischer sozialwissenschaftlicher Forschung zur Verfügung stehen. Analysen, die so in den 1980er-Jahren kaum durchführbar gewesen wären oder einen ganzen Mainframe lahmgelegt hätten, können heute von der Rechenkapazität her in Sekunden vom Prozessor eines Mobiltelefons bearbeitet werden. Neben der gestiegenen Leistungsfähigkeit stehen größer und schneller werdende Speichermedien zur Verfügung, die es heute jedem erlauben, ein Vielfaches der Informationen festzuhalten, die früher ein ganzes Rechenzentrum beschäftigt hätten. Da die Technologien gleichzeitig immer billiger werden, können immer mehr Institutionen immer mehr Informationen immer schneller miteinander verknüpfen, was aus Perspektive des Datenschutzes natürlich das Missbrauchspotential massiv erhöht und die gesetzliche Kontrolle erheblich erschwert.

---

## 4.3 Ubiquitäre Netze

Noch schwerwiegender sind aber die Veränderungen, die sich durch die fortschreitende Vernetzung ergeben. Das Internet, aus seinen frühen, eher militärischen Zwängen befreit, ist von einem Instrument, mit dem eine zunächst recht überschaubare Anzahl von Forschern und Institutionen Daten austauschte, von einem Werkzeug für Spezialisten also, erst zu einem Spielzeug junger, gebildeter, eher wohlhabender Männer geworden (wie eine der ersten repräsentativen Onlinestudien 1997 zeigt; van Eimeren/Frees 2011), bevor es sich im Laufe einer überraschend kurzen Zeit von einer Infrastruktur für Nerds zu einem (fast) „ubiquitären Netz“ entwickelt hat. Zwar gibt es in Deutschland immer noch ca. ein Viertel Internetabstinenten (van Eimeren/Frees 2011), darunter nach wie vor mehr Frauen, aber die Gruppe wird trotz sinkender Internettumszahlen kleiner – und älter. Bei den unter 50-Jährigen hingegen spielt das Geschlecht keine Rolle mehr und es finden sich

insgesamt wenige, die nicht wenigstens ab und zu online sind und zumindest ein paar der vielfältigen Anwendungen des Netzes nutzen, das in immer mehr Lebensbereichen immer größere Veränderungen mit sich bringt. Ubiquitär werden die Netze auch dadurch, dass auch diejenigen, die das Internet nicht explizit in dem Sinne nutzen, dass sie einen Computer einschalten und einen Browser oder ein E-Mail-Programm öffnen, mehr und mehr an diese Infrastruktur angebunden werden. Zum einen, weil traditionelle Medien wie Radio, Fernsehen oder Telefon immer mehr zum Internet-Radio/-Fernsehen/-Telefon und damit auf den Betrieb via IP (Internet Protocol) umgestellt werden – was man den Endgeräten nicht immer ansehen kann. Zum anderen, weil zunehmend technische Geräte, die früher ganz ohne Anbindung an Netzstrukturen auskamen, gleich ob kabelgebunden oder per Funk, zu „intelligenten“ Geräten werden – und selbst wenn hier der immer wieder angebrochene internetfähige Kühlschrank bisher glücklicherweise eher die Ausnahme ist, so ist das doch beispielsweise bei Heizungssystemen und Stromzählern gerade recht aktuell.

---

## 4.4 Social Media

Unter der Vorherrschaft insbesondere traditioneller Massenmedien wie Fernsehen oder Radio ist die Möglichkeit von Einzelnen, insbesondere Nicht-Prominenten, Sichtbarkeit für ein größeres Publikum zu erlangen, sehr begrenzt auf so etwas wie den Anruf in einer Live-Radiosendung oder die Teilnahme an einer (meist) drittklassigen Talkshow. Die Medieninhalte werden unter professionellen Bedingungen aufwändig produziert und müssen, um erfolgreich zu sein, ein entsprechend großes Publikum erreichen, nicht zuletzt, um wirtschaftlichen Interessen Genüge zu tun. Gleichzeitig ist eine direkte Rückkopplung strukturell nicht vorgesehen und die Sender funken ihre Nachrichten blind auf ihnen Unbekannte, von denen sie keine Rückmeldung erhalten (außer vielleicht über empirische Forschung). Eine der bemerkenswertesten Eigenschaften der Infrastruktur Internet ist, dass diese klassische Beziehung zwischen Sender und Empfänger grundsätzlich aufgehoben wird und das, was gerne als „social media“ bezeichnet wird, möglich wird. Wer sendet und wer empfängt, ist zunächst einmal egal, Rückmeldungen sind relativ einfach möglich (so existiert kaum noch ein Netzauftritt, der keine Kommentare zulässt), jeder kann mit beliebig großen und kleinen, professionellen wie unprofessionellen Mitteln produzieren – und hat grundsätzlich die Chance, gehört zu werden. Trotz dieses grundsätzlichen Unterschiedes waren die Strukturen traditioneller Massenmedien gerade in den 1990er-Jahren auch im Internet dominant. Zwar war Austausch durchaus möglich, aber die meisten Angebote waren weiterhin Produkte kommerzieller Medienunternehmen, die von vielen Nutzern konsumiert wurden. Dies veränderte sich aber zunehmend in den 2000er-Jahren – insbesondere seit 2003/2004 mit dem Begriff Web 2.0 (für einen Überblick siehe Cormode/Krishnamurthy 2008) das Konzept social media die Öffentlichkeit mit einem eingängigen Slogan trifft. Mehr Nutzer werden gleichzeitig zu Produzenten nutzergenerierter Inhalte, während parallel (Netz-)Anwendungen florieren, deren Kern gerade solche Inhalte darstellen, sei dies Wikipedia, Flickr, YouTube oder ein beliebiges soziales Netzwerk bzw. ein Blog.

## 4.5 Datensammler

In dieser schönen, neuen, vernetzten Welt scheint vieles von dem, was die Gegner der Volkszählung in den 1980er-Jahren als Grundlage für eine mögliche Überwachungsgesellschaft fürchteten, eingetreten und übertroffen zu sein. Ein Großteil der Bevölkerung ist nicht nur immer mobil telefonisch zu erreichen, sondern hat in der gleichen Technologie verankert auch den eigenen Peilsender dabei – und eine Kamera, die samt Geotagging und Zeitstempel immerzu und überall zielstrebig die peinlichsten Entgleisungen festzuhalten und ins Netz zu stellen bereit ist (und dabei ist die Welt nach der Einführung von Google-Glass noch nicht einmal mitgedacht). Und so weiß Apple dank dem iPhone, wo wir sind, und kann Bewegungsprofile anlegen, Facebook kennt unsere Freunde und weiß, wenn wir die entsprechende Seite ausfüllen, auch einiges mehr über uns, Amazon registriert, was wir kaufen und lesen, und Google weiß so viel über uns, dass es auch gleich antizipiert, was wir wissen wollen – und lässt dies subtil in unsere Suchanfragen (und selbstredend in die personalisierte Werbung) einfließen. Natürlich besteht hier im Vergleich zu der Volkszählung keine Auskunftspflicht und eine detaillierte Abfrage einer durchaus sensiblen Information wie dem Migrationshintergrund ist auch nicht unbedingt Standard. Tatsächlich ist aber völlig unüberschaubar, welche zunächst vielleicht sogar harmlos erscheinenden Einzelinformationen über die verschiedenen Aktivitäten und genutzten Dienste verschiedener Anbieter wo und wie verknüpft und weiterverarbeitet werden und welche Erkenntnisse sich daraus ergeben. Angaben zur Familie sind beispielsweise gerade in sozialen Netzwerken durchaus vorgesehen, die eigenen Netzwerke werden durch die Nutzung der Seite generiert, woraus sich unter Umständen deutlich mehr ableiten lässt als nur der Migrationshintergrund.

---

## 4.6 Die veränderten Kosten der Überwachung

So vielfältig, lückenlos und besorgniserregend die Möglichkeiten auch sind, die sich dank ubiquitärer Netze zur Überwachung ergeben, so hat es doch bereits in analogen Zeiten sowohl Möglichkeiten wie Begehrlichkeiten der Informationsbeschaffung und Überwachung gegeben, man erinnere sich nur an die Rasterfahndung zu Zeiten der RAF; was also ist anders? Das Problem besteht darin, dass sich die Grundbedingungen, die für verschiedene Formen der Überwachung gelten, grundlegend geändert haben – man denkt, es sei das gleiche Spiel, aber die Spielregeln sind auf so subtile Weise umgeschrieben worden, dass ein schon immer problematisches Äquilibrium strukturell aus den Fugen zu geraten droht. Dies wird deutlich, wenn bedacht wird, dass in einer prävernetzten Ära zwar auch schon vieles von dem überwacht werden konnte, was sich heute einem neugierigen Dritten ergibt, die Kosten der Durchführung, nicht nur monetär, sondern insbesondere auch bezüglich des Faktors Zeit, waren aber so hoch, dass sich eine Art natürlicher Grenze dessen, was und wie viel überwacht werden konnte, ergab. Dies wird besonders deutlich im Kontext der systematischen Überwachung von Audiodaten, Textinformationen und Bewegungsprofilen: Während ein Telefon zu analogen Zeiten tatsächlich physisch „angezapft“

ein Brief abgefangen, vorsichtig geöffnet und wieder verschlossen, eine Person von einer anderen verfolgt werden musste, so reicht heute ein Mitschnitt der Datenbündel des VoIP-Gespräches (Internettelefonie), das Auslesen eines E-Mail-Kontos und die Auswertung der Bewegungsdaten von Mobiltelefonen – und in Zukunft vermutlich die Erfassung von RFID (Radio Frequency Identification, beinahe beliebig kleinen Transpondern, die je nach Funktionsumfang mindestens eine eineindeutige Identifikation senden und z. B. Waren kennzeichnen können, im Gegensatz zum Barcode aber ebenfalls eineindeutig) und die automatische Gesichtserkennung von Überwachungskameras. Steht hier die Infrastruktur erst einmal, dann sind die Kosten, insbesondere was den Faktor Zeit angeht, beinahe vernachlässigbar, vor allem angesichts der Möglichkeit, die Erfassung verdachtsunabhängig und damit flächendeckend durchzuführen – und weitestgehend im Verborgenen, da es kein Knacken in der Leitung und keinen angerissenen Briefumschlag mehr gibt. An die Stelle einer Reihe von Ermittlern, die unter Aufwendung vieler Überstunden versuchen, kleinteilig die gewonnenen Informationen manuell zu sichten und zu einem vernünftigen Überblick zusammenzuführen, rückt ebenfalls eine Automatisierung, und über einen leistungsfähigen Data-Mining-Algorithmus (Cielebak/Rässler, Kapitel 26 in diesem Band) wird aus „Big Data“ wie von Zauberhand „Big Surveillance“. Wo, trotz der frühen Erfahrung mit Programmen wie Echelon, diesen Zeilen noch wenige Wochen vor ihrer Niederschrift der Nimbus des Paranoiden angehängen hätte, ist seit der NSA-Affäre Snowden um PRISM und Tempora klar, dass all dies aus der Sphäre des nur theoretisch Machbaren bereits längst in die Praxis überführt wurde, dass eine flächendeckende Überwachung, Verknüpfung und Weiterverarbeitung grundsätzlich aller im Rahmen ubiquitärer Netzwerke anfallender Daten längst stattfindet.

---

#### 4.7 Deterministisch oder doch probabilistisch?

Die Zuverlässigkeit der so erlangten Ergebnisse (Krebs/Menold, Kapitel 30 in diesem Band) dürfte dabei in verschiedener Hinsicht besonders problematisch sein. Wenn es darum geht, grundsätzliche Zusammenhänge zwischen Variablen zu erforschen, werden die so gewonnenen Daten vermutlich sogar recht gut einsetzbar sein und zu halbwegs stabilen Ergebnissen führen können – mit den in der empirischen Sozialforschung üblichen Irrtumswahrscheinlichkeiten (Häder/Häder Kapitel 18 in diesem Band) und Reichweiten der Erkenntnisse. Ansätze, die so versuchen, mittels nutzergenerierter Daten, aber auch mittels Paradataen (also Informationen, die nebenbei zusätzlich anfallen, wie Logindauer, Häufigkeit und Länge von Telefonanrufen, verbrauchte Zeit beim Ausfüllen eines Onlineformulars; Feldecker/Birg/Kreuter, Schmitz/Yanenko, Kapitel 25 und 65 in diesem Band) Zusammenhänge zwischen Nutzereingaben und Nutzerverhalten zu Nutzereigenschaften herzustellen, sind sicherlich aus einer Forschungsperspektive sehr interessant. Solche Proxy-Variablen haben zwar die Eigenschaft, probabilistisch einigermaßen stabil zu sein, lassen aber nicht unbedingt Rückschlüsse auf das Individuum zu. Je nach Kontext ist aber durchaus denkbar, dass sich so gewonnene Zuschreibungen wie beispielsweise Geschlecht (z. B. über den Namen),

Schwangerschaft (z. B. über Suchanfragen oder getätigte Onlinekäufe) oder auch sexuelle Orientierung (mit Hilfe von Freunden auf Facebook; Jernigan/Mistree 2009) verselbständigen können, weil sie tatsächlich dem Individuum zugeordnet werden. Solange die Auswirkung ist, dass „personalisierte“ Werbung für eine unzutreffende Zielgruppe geschaltet wird, mag das noch verschmerzbar sein. Je nach Personenkreis, an den die Informationen gehen, kann dies aber massiv negative Konsequenzen nach sich ziehen – und das kann auch schon bei auf den ersten Blick harmlos wirkenden (vermeintlichen) Eigenschaften so sein, wenn beispielsweise aufgrund einer Reihe entsprechender Suchanfragen plötzlich Werbung für Babynahrung ins Haus kommt und der Partner noch nichts von seinem Glück weiß. Und mehr noch: Um die Reichweite der Folgen beurteilen zu können, darf nicht vergessen werden, dass diese Techniken ja gerade auch im Überwachungskontext zur Anwendung kommen, in dem die meisten Daten zusammengeführt werden und besonders große staatliche Begehrlichkeiten in einer Welt post 9/11 aufkommen – und hier geht es ja gerade darum, einzelne Individuen aus der Masse als deviante Elemente auszuweisen.

---

## 4.8 Das Problem der Identität

Neben der Verlässlichkeit der statistischen Zusammenhänge, auf denen solche Zuschreibungen beruhen, ist zu bedenken, dass die Identität der Personen, der die Datengenerierung zugeschrieben wird und auf die damit auch die Rückschlüsse gezogen werden, aber keineswegs sichergestellt werden kann. Nicht alle Eingaben, Suchanfragen, Einkäufe und E-Mails, die von einem Rechner, sogar von einem Nutzerkonto aus geführt werden, müssen tatsächlich von dem Eigner sein, sondern können genauso gut vom Partner, von Freunden oder Bekannten vorgenommen worden sein, die „nur mal eben schnell“ den Rechner oder das Tablet benötigten. Auch sind gezielte Identitätsdiebstähle oder durch Trojaner befallene und damit von Dritten fernsteuerbare Rechner keine Seltenheit. Paradoxerweise verschlimmern unter Umständen bessere Authentifizierungsverfahren die Situation. Denn gerade dann, wenn dieses in sich als besonders ausgefeilt, sicher und technisiert erscheint, wie dies beispielsweise bei biometrischen Systemen der Fall ist, wird dem Testverfahren leicht absolute Glaubwürdigkeit unterstellt. Die Komplexität des Identitätstests selbst hat aber keinen Einfluss darauf, dass allein schon bei der Eingabe immer noch sowohl die Möglichkeit von Manipulationen als auch von simplen Fehlern besteht – gerade, wenn flächendeckend ein neues Verfahren implementiert wird und viele Informationen in kurzer Zeit eingepflegt werden müssen. So reicht auch bei einem tollen, neuen, biometrischen Identitätstest immer noch ein Bug aus, um, um mit Terry Gilliams „Brazil“ zu sprechen, aus Mr. Tuttle Mr. Buttle zu machen, und den einen für die Sünden des anderen zu strafen, weil sich gerade durch die vermeintliche Zuverlässigkeit des Identitätstests die Beweislast bei einem Fehler umkehrt.

Dies relativiert auch sehr stark die Sinnhaftigkeit des Arguments „Ich habe nichts zu verheimlichen.“ Selbst wenn dies zutreffen sollte, schützt dies weder vor einer Verwechslung auf Personenebene, noch davor, dass aufgrund einer Reihe von automatisch erfassten

und automatisch analysierten, vermeintlich einer Person zuschreibbaren Variablenausprägungen dieser eine (unerwünschte) Eigenschaft zugeordnet wird, die zwar probabilistisch stark mit diesen Variablenausprägungen zusammenhängen kann, aber keineswegs auf der Individualebene zutreffen muss – und so auch der Unbescholtene plötzlich zum Abweichler werden kann.

---

#### **4.9 Habe ich doch etwas zu verheimlichen?**

Zusätzlich zu dem „ich habe nichts zu verheimlichen“ sind noch zwei andere Positionen populär, die entweder den Verlust von Privatsphäre als relativ unproblematisch ansehen oder diesen sogar als wünschenswert herbeisehnen: Zum einen ist dies eine defätistisch-zynisch erscheinende, die der ehemaligen Cisco-CEO Scott McNealy mit seiner Aussage „You have zero privacy anyway. Get over it.“ (Manes 2000) sehr prägnant umrissen hat; die zweite ist die eher opportunistisch-naive Utopievorstellung einer Welt, die von Post-Privacy-Befürwortern ins Feld geführt wird, in der alle alles über alle wissen und diese Informationen aber nicht missbrauchen. Alle drei Positionen sind aus mindestens zwei Gründen problematisch: Erstens kann Privatheit im Rahmen des Liberalismus als notwendiges Fundament eines autonomen und damit freien Lebens gesehen werden (Rössler 2001). Zweitens setzt insbesondere die Position der Post-Privacy-Befürworter nicht geringe soziale Veränderungen voraus. Selbst, wenn ich nichts zu verheimlichen habe, so sind doch zumindest die meisten sozialen Situationen so konstruiert, dass bestimmte Informationen, die in sich völlig unproblematisch sind, in anderen Kontexten als unangemessen bewertet werden. Ein Bild, das einen Vollrausch in der Freizeit dokumentiert, gehört nicht in die Hände des Arbeitgebers, so wie eine Liste der sexuellen Präferenzen nicht in die der Kollegen (oder gar in die Untergebener, die sich vielleicht schon durch die reine Kenntnis dieser Informationen belästigt fühlen könnten). Bisher erscheint es zumindest hochgradig unwahrscheinlich, dass eine Reihe einschlägiger Bilder bei Facebook von einem potentiellen Arbeitgeber als Einstellungskriterium einfach ignoriert werden würden. Natürlich kann sich dies ändern, so wie sozialer Wandel immer möglich ist, aber fraglich ist, ob es einem Primat des technisch Machbaren überlassen sein muss, in welche Richtung dieser Wandel geht – und ob damit nur aus der Situation, dass in einer Welt ubiquitärer Netze ubiquitäre Überwachung möglich ist, auch folgen sollte, dass sich die Gesellschaft dahingehend verändert, diese Überwachung zu akzeptieren.

---

#### **4.10 Indiskretionen**

Derzeit lässt sich jedenfalls feststellen, dass sich die Rahmenbedingungen einer vernetzten Welt weitreichend auf viele Lebensbereiche auswirken können, im Zweifelsfall auch ohne die Mitarbeit des Einzelnen. Denn auch ein datensparsames Leben ohne Mobiltelefon und Internet ist kein Schutz vor unliebsamen Überraschungen. Insofern es nicht mit einem

Dasein als Eremit einhergeht, ist auch der Vorsichtigste nicht vor einem Informationsleck geschützt, so es denn seine liebe Umgebung nicht will. Dank des Internets hat jede Person (in einer westlichen Industrienation zumindest) Zugriff auf eine potentiell globale Informationsplattform, nicht nur zum Konsum, sondern auch für die Distribution von Informationen. So gibt es keinen echten Schutz davor, dass die eigene informationelle Selbstbestimmung durch die nutzergenerierten Inhalte Anderer kompromittiert wird. Selbst wenn vorsätzliche Übergriffe, vielleicht sogar Falschaussagen und Mobbing, nicht betrachtet werden, so sind Indiskretionen kaum vermeidbar, und sei dies auch „nur“ die Versendung einer E-Mail an einen ganzen Verteiler im cc – der sogleich alle Mailadressen allen Adressaten offenlegt. Das mag im Rahmen des Kegelclubs noch unproblematisch sein, darf aber keinem Arzt, Juristen, Journalisten, Politiker usw. passieren.

---

#### **4.11 Verzicht ist auch keine Lösung**

Es erscheint paradox, dass zu einer Zeit, in der die empirische Sozialforschung mit immer größeren Nonresponse-Raten zu kämpfen hat (Engel/Schmidt, Kapitel 23 in diesem Band), die Menschen an anderer Stelle als Nutzer freiwillig Unmengen an Daten generieren, twittern, liken, unverschlüsselt mailen, online shoppen, sozial netzwerken – und dies vor dem Hintergrund, dass inzwischen fast jede große Branche, die diese Informationen sammelt und verwaltet, bereits in Skandale verwickelt war – ob Apples Übertragung von Bewegungsprofilen, Googles Auslesen von WLAN-Informationen, Verlust von Kundendaten und Passwörtern durch, ja, wen nicht? Aber während die Nichtbeantwortung eines Fragebogens einer wissenschaftlichen Forschungseinrichtung in der Regel weder den Verlust einer direkten (und wirklich interessanten) Incentivierung, noch eine reduzierte Teilhabe am sozialen Leben bedeutet, so kann der Verlust des Facebook-Accounts sehr wohl deutlich schmerzlichere Spuren hinterlassen. Welche Aktivitäten der Freunde würden ohne Facebook verpasst, welche potentiellen Arbeitgeber einen ohne Eintrag in Xing nicht finden? Wie umständlich (und teuer) wäre es, jede E-Mail wieder als Brief zu verfassen, auszudrucken oder gar von Hand zu schreiben und zur Post zu bringen – und vielleicht Tage auf die Antwort warten zu müssen? Während es sehr gute Gründe gibt, alle möglichen Dienstleistungen zu nutzen, die durch ubiquitäre Netze ermöglicht werden, und eine solche Nutzung nicht mit sofort sichtbaren Gratifikationen geizt, so sind die Sorge um den Verlust der Privatsphäre und die möglichen Konsequenzen, die ein Kontrollverlust über ein auf den ersten Blick unbedeutsam erscheinendes Datum nach sich zieht, deutlich schwerer zu fassen, weniger direkt – und zum Teil auch einfach aufgrund der doch eher komplexeren technischen Komponente (beispielsweise bei RFID) schlecht nachvollziehbar. Somit überrascht es nicht, dass es so erscheint, als ob im Zweifelsfall Dienste eher verwendet und Datenschutzbedenken dabei erst einmal ignoriert und dann vergessen werden. Fraglich ist indes aber, ob dies bedeutet, dass den Nutzern der Wert des Privaten (Rössler 2001) tatsächlich unbewusst ist oder das Private für sie an Bedeutung verloren.

## 4.12 Und nun?

Was bedeutet das nun alles konkret für die sozialwissenschaftliche, insbesondere die empirische Forschung? Derzeit sind die Konsequenzen gerade der neueren Entwicklungen kaum absehbar. Es erscheint aber unwahrscheinlich, dass selbst so weitreichende Skandale wie die um PRISM und Tempora plötzlich eine flächendeckende Abkehr der Nutzer von Datenkraken nach sich ziehen wird – zu sehr sind deren Anwendungen in den Alltag eingebunden, zu sehr an soziale Teilhabe geknüpft, als dass sie gänzlich aufgegeben werden. Wahrscheinlich wird schon allein zur Vermeidung kognitiver Dissonanz nicht mehr als ein kurzer Moment des Unbehagens bleiben, bevor munter weiter gepostet wird. Es wäre aber durchaus denkbar, dass in den Augenblicken, in denen das eigene Verhalten bewusst reflektiert wird, weil es eben gerade nicht alltäglich ist, wie beispielsweise in einer Befragungssituation, eine vorsichtigere Haltung angenommen, eine Anpassung an sozial erwünschte Verhaltensweisen vorgenommen wird – und das könnte sich durchaus negativ auf die Antwortbereitschaft und die Datenqualität von Online-Umfragen auswirken. Deren Vorteil, dass gerade kein Interviewer präsent ist, könnte dadurch konterkariert werden, dass stattdessen „Big Brother“ mitliest, und so das Gefühl der Anonymität im Netz gänzlich aufgehoben wird. Es wird also noch schwieriger, den Befragten glaubhaft zu vermitteln (und dies auch tatsächlich einzuhalten), dass ihre Daten, wie es immer so schön heißt, „selbstverständlich anonym“ und nur für die Forschung verwendet werden.

Wenn es damit einerseits immer schwieriger wird, selbsterhobene Befragungsdaten zu generieren, aber andererseits immer größere Datenschätzungen in großen Unternehmen vorliegen, ist es durchaus denkbar, dass Begehrlichkeiten auch auf Seiten empirischer Forscher auftreten, diese zu nutzen. Das erste, pragmatische Problem dürfte hier erst einmal das des Zugangs sein – denn für die Unternehmen liegen diese Daten irgendwo zwischen Kapital und Geschäftsgeheimnis, und Scientific-Use-Files dürften wohl eher die Ausnahme bleiben. Bleibt also noch die Möglichkeit einer (mehr oder weniger unheiligen) Allianz zwischen Wirtschaft und Forschung, in der die Forschung die Daten entweder kauft oder aber mit ihren Forschungsergebnissen und -schwerpunkten bezahlt. Damit stellt sich aber gleich das zweite Problem über die damit verbundenen rechtlichen und ethischen Fragen, die von der Forschung erst beantwortet werden müssen (Friedrichs, Kapitel 3 in diesem Band): Woher kommen die Daten? Sind sie an einen Verwendungszweck gebunden und ist die Forschung, Verknüpfung, Weiterverarbeitung etc. darin enthalten? Wurden sie überhaupt legal generiert? Sind sie auch in Deutschland legal? Nicht zuletzt: Welche Gegenleistung erwartet der Datenlieferant? Und natürlich: Ist es eine gute Idee, diese zu erbringen? Denn wenn probabilistische Aussagenmodelle dazu verwendet werden, Zustandsbeschreibungen an Individuen zu knüpfen, auf die diese gar nicht zutreffen müssen, muss sich die empirische Sozialforschung fragen, ob sie nicht von ihrem Weg abgekommen ist, wenn sie bei diesem Spiel mitspielen will.

## Literatur

- Brunnstein, Klaus (2007): Zur Entwicklung des Datenschutzes in der Informationsgesellschaft. In: *it – Information Technology* 49. 2.: 132-35
- BVerfG (1983): Volkszählungsurteil. Urteil vom 15. Dezember 1983. BVerfGE 65: 1-71
- Cormode, Graham/Krishnamurthy, Balachander (2008): Key Differences Between Web 1.0 and Web 2.0. In: *First Monday* 13. 6. Quelle: <http://journals.uic.edu/ojs/index.php/fm/article/view/2125/1972> Am: 02.08.2013
- Jernigan, Carter/Mistree, Behram F. T. (2009): Gaydar. Facebook Friendships Expose Sexual Orientation. In: *First Monday* 14.10. Quelle: <http://firstmonday.org/ojs/index.php/fm/article/view/2611/2302>. Am: 02.08.2013
- Mack, Chris A. (2011): Fifty Years of Moore's Law. In: *IEEE Transactions on Semiconductor Manufacturing* 24. 2.: 202-207
- Manes, Stephen (2000): Private Lives? Not Ours! In: *PCWorld*. Quelle: <http://www.pcworld.com/article/16331/article.html>. Am: 02.08.2013
- Rössler, Beate (2001): Der Wert des Privaten. Frankfurt am Main: Suhrkamp
- van Eimeren, Birgit/Frees, Beate (2011): Drei von vier Deutschen im Netz. Ein Ende des digitalen Grabens in Sicht? In: *Media Perspektiven* (7-8): 334-349

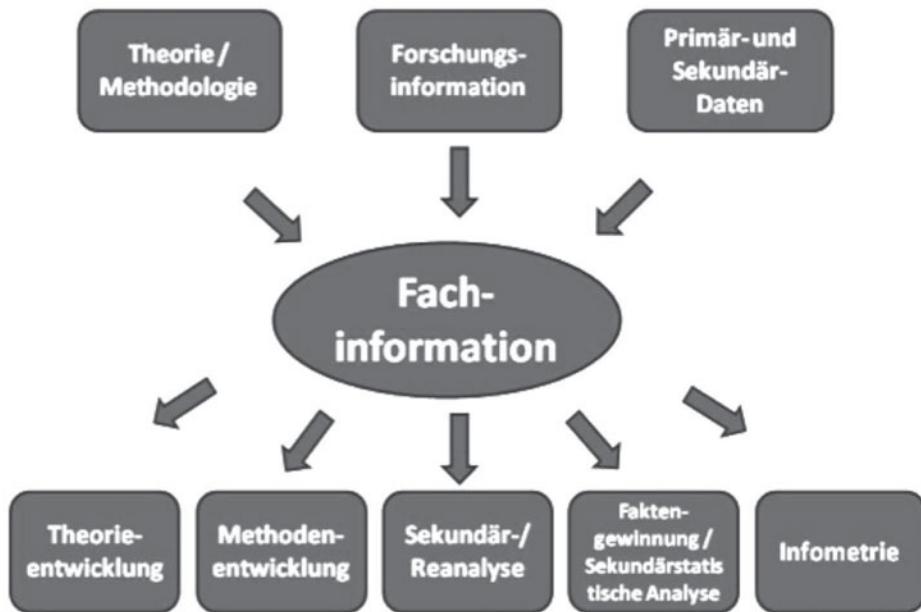
H. Peter Ohly und Karsten Weber

---

## 5.1 Die Bedeutung von Information für die Forschung

Bereits im Planungsstadium eines Forschungsvorhabens ist es wichtig, dass man die Forschungsergebnisse auf dem untersuchten Gebiet kennt. Zum einen sollten bereits geklärte Fragestellungen nicht erneut aufgegriffen werden, andererseits muss, gerade in den Sozialwissenschaften, das bereits Erforschte oft in anderen zeitlichen und räumlichen Settings erneut abgesichert werden. Auch um Wissen zu produzieren, benötigt man selbst bereits Wissen, z.B. Theorien, Werkzeuge zur Auswertung von Daten und natürlich Informationen, Daten und Statistiken über soziale Phänomene. Mit der zunehmenden internationalen und interdisziplinären Ausrichtung der Forschung ist eine professionelle Recherche in und die Zulieferung zu fachlichen Informationsdatenbanken unabdingbar (siehe Abb. 5.1; Kaminsky 1977, Kuhlen 1989).

Im Folgenden wird beschrieben, wie man in den Sozialwissenschaften Zugang zu Informationen findet und diese aufbereitet, aber auch, wie man solche Informationen publiziert und damit der (wissenschaftlichen) Öffentlichkeit zur Verfügung stellt. Dieses vollzieht sich heutzutage weitgehend über das Medium „Internet“ und mit Techniken der Informationswissenschaft und Informatik. Suchen per Freitexteingaben in WWW-Suchmaschinen wie Google oder Bing sind inzwischen aus dem Alltag nicht mehr wegzudenken. Bei anspruchsvoller Nutzung dieser sollte aber beachtet werden, dass es auch Alternativen (z.B. Yahoo, ASK und vor allem die Suchmöglichkeiten der Fachportale) gibt, Daten aus dem sogenannten Deep-Web zu sammeln, Gewichtungsverfahren anzuwenden oder Suchergebnisse zu präsentieren. So werden mehrere Suchmaschinen gleichzeitig von Metasuchmaschinen, wie z.B. MetaCrawler, angestoßen (Sherman 2005). Google benutzt PageRank, der wesentlich die Verlinkung mit prominenten Seiten berücksichtigt (Fries et al. 2001). Vor allem aber sollten bei einer Fachrecherche auch die Möglichkeiten für „Erweitertes Suchen“ (ULB Münster o.J.) und Zusatzleistungen – wie Bestellmöglichkeiten, Mitteilungsdienste, Übersetzungen oder Personalisierungen – ausgeschöpft werden.



**Abb. 5.1** Fachinformation im Forschungskontext (Die Autoren danken Sebastian Netscher, GESIS, für die Anregung zu diesem Schaubild)

## 5.2 Informationssuche und Informationsbeschaffung

### 5.2.1 Literaturangebote

In den Wissenschaften hat die Suche nach Literatur nach wie vor einen hohen Stellenwert (Heidtmann 1985). Insbesondere bei theoretischen und geisteswissenschaftlichen Fragestellungen sind aber allgemeine Suchmaschinen hier wenig zielgerichtet (Fröhlich 1998). Vielmehr sind Suchen mit expliziten wissenschaftlichen Suchmaschinen, die auf entsprechende Datenbestände zurückgreifen – wie *Google Scholar* (inkl. *Google Books*), *Bielefeld Academic Search Engine (BASE)*, *Web of Science* (bzw. *ISI Web of Knowledge*) oder *Scopus* – zu empfehlen (siehe: *ULB Münster 2010*). Diese beziehen sich weniger auf Bücher als vielmehr auf *Zeitschriftenaufsätze* (inkl. Open-Access-Zeitschriften), *Buchkapitel* und z.T. auf *Repositorien mit Universitätsarbeiten* sowie weitere Materialien (*CiteSeer* bezieht sich z.B. ausschließlich auf Internet-Veröffentlichungen) – je nach Anbieter allerdings *eher auf den angloamerikanischen Bereich* bezogen. Ein wesentlicher Vorteil der oben genannten wissenschaftlichen Suchmaschinen liegt darin, dass sie Zitationsstrukturen abbilden und damit sowohl für eine Ausweitung der Suche, als auch für eine bewertende Analyse offen stehen. Ein weiterer Service ist das Angebot der unmittelbaren Lieferung des Volltextes oder ersatzweise eines Nachweises von Bezugsmöglichkeiten der Literaturstellen (Lehnhard-Bruch 2009).

Eine stärkere Fokussierung auf *deutsche Fachbücher und -zeitschriften* ermöglichen die *Online-Kataloge* (OPACs) der Hochschulbibliotheken, die in Deutschland zu Verbünden (GBV, KOBV, HBZ, HeBIS, SWB, BSZ, BVB) zusammengefasst sind (s.a.: Plassmann et al. 2011). Per Fernleihe sind Ausleihen über die Verbundgrenzen möglich. In der Regel sind Dokumentenlieferdienste (z.B. Subito) angeschlossen.

Einen *kombinierbaren Zugang zu deutschen und internationalen Bibliothekskatalogen* bietet der *Karlsruher Virtuelle Katalog* (KVK). Hierüber sind auch die *Zeitschriftendatenbank* (ZDB) aus deutschen und österreichischen Bibliotheken und das *Verzeichnis lieferbarer Bücher* (VLB) durchsuchbar.

*Dandelon* durchsucht über digitalisierte Inhaltsverzeichnisse und mit Hilfe von Fachthesauri (strukturierte Fachbegriffsverzeichnisse) wissenschaftliche Bücher und Aufsätze aus 20 Hochschul-Bibliotheken im deutschsprachigen Raum.

Eine umfangreiche Sammlung elektronischer Zeitschriften (mit Volltexten) aller Fachgebiete stellt die *Elektronische Zeitschriftenbibliothek* (EZB) in Deutschland dar, welche auch auf Bestände von JSTOR (enthält u.a. Bestände der British Library) verlinkt, worin eine Globalsuche nach einzelnen Aufsätzen möglich ist. Für den *englischsprachigen Bereich* können *IngentaConnect* und *ScienceDirect* als Metasuchmaschinen für die Angebote von Fachverlagen genannt werden, abgesehen von Recherchen direkt bei den Zeitschriftenherausgebern wie Elsevier, Sage, Springer, Wiley u.a. selbst.

Für die Sozialwissenschaften ist das Open-Access-Angebot in den letzten Jahren sehr gewachsen. So bietet z.B. das *Directory of Open Access Journals* zahlreiche Einträge für Journale der Geistes-, Sozial- und Kulturwissenschaften (auch: [http://open-access.net/de/oa\\_in\\_verschiedenen\\_faechern/sozialwissenschaften/#c462](http://open-access.net/de/oa_in_verschiedenen_faechern/sozialwissenschaften/#c462)), die wie herkömmliche Zeitschriften oft eine Begutachtung (mit *Peer-Review*) durchführen. Ähnliches zeichnet sich nun auch für Bücher mit dem europäischen Verbundprojekt *Open Access Publishing in European Networks* (OAPEN) ab.

Hervorzuheben sind auch *Sondersammelgebiete der Bibliotheken* (SSGs), die einschlägige Literatur unabhängig von ihrer Erscheinungsform so umfassend wie möglich sammeln, so Sozialwissenschaften an der *Universitäts- und Stadtbibliothek Köln*, Volkswirtschaft und Weltwirtschaft bei der *Deutschen Zentralbibliothek für Wirtschaftswissenschaften* (ZBW) in Kiel und Hamburg, Psychologie an der *Saarländischen Universitäts- und Landesbibliothek* (SULB). Die SSGs sind u.a. an der Zeitschriftendatenbank ZDB beteiligt.

## 5.2.2 Fachinformation

Neben den meist von Bibliotheken verwalteten Angeboten zur Fachliteratur haben sich Fachinformationszentren und andere fachliche Dokumentationsstellen etabliert, die grundsätzlich Fachinformation über die im Verlagshandel erhältlichen Quellen hinausgehend anbieten, wie *Informationen über Forschungsstellen, Forschungsprojekte, Internetauftritte, Veranstaltungen, Daten und Fakten, Artefakte, Bild-, Audio-, Videodokumente u.a., inkl. weiterem Fachservice*.

Kennzeichnend sind hierbei die Erschließung der Information über eine formalisierte Beschreibung der Inhalte (*Metadaten*, z.B. inhaltliche Deskriptoren), wie auch die möglichst informative Sekundärbeschreibung über *Abstracts*. Welche Datenbanken für ein Themengebiet von Interesse sind, kann das *DatenBankInfoSystem (DBIS)* zeigen, das zentrale Nachweissystem für frei zugängliche und lizenzpflchtige Datenbanken. In der Regel werden diese Informationen inzwischen über *Fachportale* angeboten, die mehrere Datenarten und Datenanbieter unter einer gemeinsamen Oberfläche im WWW anbieten (Hohlfeld 2012). Oft besteht auch eine Vernetzung zu Angeboten der Bibliotheksverbünde. Von den *Universitätsbibliotheken* wird meist ein auf ihre Fachbereiche bezogenes Spektrum an Fachdatenbanken in ihrem *Hochschulnetz* angeboten.

Zu den Informationsquellen, die im Internet mehr oder weniger frei verfügbar sind, zählt eine Reihe von fachlichen Angeboten, die sogenannten *Gateways*, *Clearinghouses* oder *Kataloge*, die oft auch systematisch geordnet sind. Diese sind besonders auf ihr Fachgebiet fokussiert und listen nur Quellen auf, die sorgsam von Experten geprüft und beschrieben wurden. Systematisch geordnete Fachkataloge sind die *WWW Virtual Library Social and Behavioural Sciences*, die *Social Sciences Virtual Library*, *WWW Virtual Library: Sociology*, *Sozialwissenschaften FAQ* und *SocioSite*. Der umfangreiche britische Dienst *Intute* (ehemals: SOSIG) ist bis 2014 weiter online verfügbar, wird aber seit 2011 nicht mehr aktualisiert. Ähnlich sieht es mit dem *Arts and Humanities Data Service (AHDS)* aus.

### **5.2.3 Sozialwissenschaftliche Fachinformation**

In den Sozialwissenschaften erstellt die Abteilung *Fachinformation für Sozialwissenschaften (FIS)* des GESIS Leibniz-Institut für Sozialwissenschaften bezogen auf die deutschsprachigen Länder eine *Datenbank zu Literatur (SOLIS)* und eine zu *Forschungsprojekten (SOFIS)*. In SOLIS werden (zusätzlich zu Forschungsberichten, Monographien und Sammelwerken) rund 300 Zeitschriften ausgewertet, wie: *Berliner Journal für Soziologie*, *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *Soziale Welt*, *Zeitschrift für Soziologie*.

Weiter bietet GESIS in Kooperation mit anderen Institutionen das Fachportal SOWI-PORT an, das folgende Datenbanken vereint: *Literaturdatenbank FIS Bildung*, *Katalog des Sondersammelgebietes Sozialwissenschaften der Universitäts- und Stadtbibliothek Köln*, *Sozialwissenschaftliches Literaturinformationssystem (SOLIS)*, *Sozialwissenschaftliches Forschungsinformationssystem (SOFIS)*, *Datenbestandskatalog der GESIS (DBK)*, *Social Sciences Open Access Repository (SSOAR)*, *Fachinformationsführer SocioGuide* der GESIS, *Bibliothekskatalog des Wissenschaftszentrums Berlin für Sozialforschung*, *Literaturdatenbank des DZI (SoLit)*, *Datenbank des Deutschen Zentrums für Altersfragen (GeroLit)*, *Katalog der Bibliothek der Friedrich-Ebert-Stiftung*, *Datenbank der Bertelsmann Stiftung*.

Für die Hochschulen in Deutschland frei zugänglich sind darüber hinaus (Nationalallianzen): *CSA Sociological Abstracts*, *CSA Social Services Abstracts*, *Applied Social Sciences*

*Index and Abstracts* (ASSIA), *CSA Worldwide Political Science Abstracts* (WPSA), *Database of the Public Affairs Information Services* (PAIS International), *Physical Education Index* (PEI).

In der empirischen Sozialwissenschaft werden Umfragedaten und zunehmend auch qualitative Daten oft einer Zweitauswertung (*Sekundäranalyse*) unterworfen. Die Suche nach und die Aufbereitungs- und Analysemöglichkeiten von entsprechend geeigneten Datenbeständen stellen besonders hohe Anforderungen an die Informationsversorgung. Empirisch ausgerichtete wissenschaftliche Arbeiten können nur so gut sein wie die zugrundeliegenden Daten. Hierzu verweisen wir auf die ausführlichen Beiträge über Sekundäranalysen mit qualitativen Daten (Medjedović, Kapitel 13 in diesem Band) sowie quantitative Trend- und Querschnittsdaten (Mochmann, Kapitel 14 in diesem Band) und Längsschnittdaten (Schupp, Mika/Stegemann und Pötter/Prein, Kapitel 73, 85 und 86 in diesem Band).

---

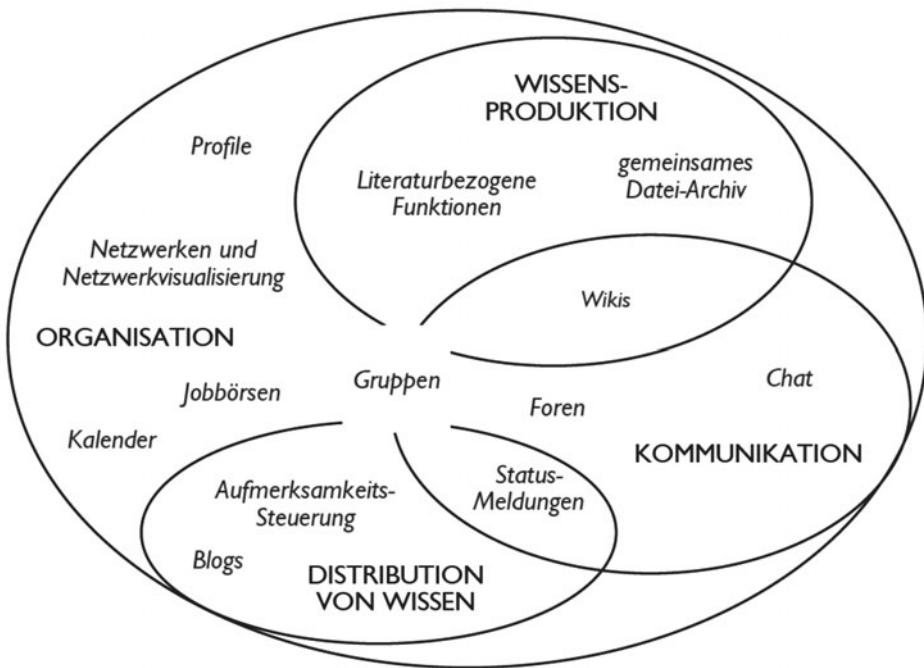
## 5.3 Kollaborative Informationsmedien

Mit dem Ausbau eines interaktiven Internet, dem „Web 2.0“ (O'Reilly 2005), gewinnen zunehmend Möglichkeiten des kooperativen Erarbeitens und der maßgeschneiderten Anpassung von Informationen auch für die Forschung an Bedeutung. Die Potenziale der virtuellen sozialen Netze können im Wesentlichen den Bereichen Produktion, Kommunikation (im Sinne von Diskurs und Kollaboration) und Distribution von Wissen, sowie der Organisation der wissenschaftlichen Arbeit zugerechnet werden (siehe Abb. 5.2).

### 5.3.1 Blogs

*Blogs* (oder: *Weblogs*) dienen der dynamischen Berichterstattung Einzelner und der Kommentierung durch einen engeren Benutzerkreis von herausgegriffenen Themenbereichen. Beispiele sind der Blog der Deutschen Gesellschaft für Soziologie (*SozBlog*), der *Theorieblog*, das *Soziologieforum*, *ScienceBlogs*, die *SciLogs* vom Spektrum der Wissenschaft oder auch die *Sociopedia* der ISA für Reviews und zugehörige Kommentare.

*E-Mail-Diskussionslisten* und *Web-Foren* stellen demgegenüber kontrollierte, längerfristige Medien für Meinungsaustausch oder Informationsstreuung zu einzelnen Diskussionssträngen von inhaltlich definierten Communities dar (Beispiele: *H-Net – Humanities and Social Sciences Online*; Liste der *Methodensektion der DGS*, *SOCNET – The Social Networks Listserv*, *Superlist of the European Sociological Association*, *Liste für Mittel Europäische Soziologie LIMES*, *Forum RC33 – Research Committee on Logic/Methodology of ISA*). Blogs und Foren können hilfreich sein, wenn Know-how gesucht wird, das wenig standarisert ist.



**Abb. 5.2** Typische Funktionalitäten von Social Network Software (nach Nentwich/König 2011)

### 5.3.2 Wikis

Wikis (oder: *WikiWebs*) sind Content-Management-Systeme, die gemeinsame Textwerke in Hypertextstrukturen mit Online-Änderungsmöglichkeiten für alle Teilnehmer kombinieren, wie *EduCamp*, das „offen“ Tagungen organisiert, oder die *Wikipedia*, die stellenweise herkömmliche Lexika abgelöst hat, da sie aktueller und differenzierter sein kann. Umgekehrt ist der Inhalt von Wikis weniger autorisiert und es sind fachliche Mehrdeutigkeiten zu beachten.

### 5.3.3 Virtuelle Personennetzwerke

Über virtuelle Personennetzwerke (z.B. *Facebook* oder *Google+*) können Selbstdarstellungen, Aktivitäten und Materialien einem selbstgewählten Personenkreis mitgeteilt werden. In *LinkedIn* finden sich mehrere Gruppen zum Framework Programme FP7 The Future of European Union Research Policy, wie z.B. zu Socio-economic Sciences and Humanities (SSH). Die Gesellschaft für Informationswissenschaft und -praxis betreibt verschiedene Foren über das berufliche Netzwerk XING, u.a. zu Terminen und Veranstaltungen sowie Ausschreibungen.

*Mendeley* wurde zum Organisieren, Austauschen und Zitieren von wissenschaftlichen Artikeln geschaffen. *ResearchGate* und *Academia.edu* sind soziale Netzwerke, die sich an Forscher aus allen Bereichen der Wissenschaft richten, *ScienceStage* vernetzt speziell Vortragsmedien (Nentwich/König 2011; Bittner/Müller 2011). *Kurznachrichten (Tweets)* und *Mediendateien (Podcasts)* können per *Twitter* gezielt verteilt und auf mobilen Endgeräten empfangen werden (Kwak et al. 2010).

Bei den Personennetzwerken ist einerseits das Erreichen einer großen Teilnehmerzahl von Bedeutung, andererseits sind gerade kommerzielle Angebote hinsichtlich ihres Datenschutzes zu hinterfragen. Social Media Guidelines wären deshalb auch für die Wissenschaft wünschenswert (Buggisch 2011), zudem sind sie wertvolle Quellen für Inhaltsanalysen (Bruns/Burgess 2012).

### 5.3.4 Mediasharing

Wenn Referenzen zu Internet-Quellen zusammen mit anderen Personen verwaltet werden sollen, empfiehlt sich der Einsatz von *Social Bookmarking* wie *mister-wong.de*, *del.icio.us* oder *BibSonomy* (ein Ranking hierzu gibt es u.a. auf <http://social-bookmarking-sites-list.com>). Ähnliches kooperatives Mediasharing durch *Sharehoster* ergibt es für Bilder, Filme oder auch Vortragsfolien (Slideshows).

Es ist jeweils abzuwägen, wieweit das Einbringen eigenen Materials erwünschte Ergänzungen und Kommentare erbringt und ob das von Anderen angebotene Material in qualitativer oder rechtlicher Hinsicht überhaupt entsprechend genutzt werden kann. In *Flickr* befinden sich etwa dokumentarische Fotos zur Invasion der Alliierten in der Normandie: Kommentare hierzu können quellenkritische Funktion haben. Podcasts mit Luhmann und anderen Soziologen in *YouTube* sind Informationen aus erster Hand, die für die Ausbildung genutzt werden können.

Die Vorteile hierbei sind gleichzeitig auch die Nachteile: Die Vernetzung der Medien geschieht assoziativ durch die Nutzer (*Social Tagging*) und nicht zentral ordnend durch eine Fachinstanz. Oft sind diese Social Media-Angebote (aber auch fachliche Informationsangebote) mit Alertingdiensten versehen, die zu gewählten Rubriken Neuaufnahmen per Feedreader aktiv melden, etwa an das eigene E-Mail-Programm oder den Internet-Browser. So bietet Sowiport ausgewählte Themen-Feeds, etwa zu „Web 2.0“, an. Eine systematische Auswertung von Social Media kann allerdings nur mit einem recht komplexen Information Mining durchgeführt werden, wie z.B. mit *web2monitor* (Sen 2010).

### 5.3.5 Web 3.0 (Semantic Web)

Inhalte verschiedener Websites in einen logischen Zusammenhang zu bringen ist das Ziel des „Web 3.0“ (Semantic Web). Alleine schnittstellenbasierte „Mashups“, die Projektion von verschiedenartiger Information ineinander (oft auf Landkarten, wie <http://climateviewer.com/3D>), ergeben einen Mehrwert.

Basierend auf mehr formalen Links wird mit *InfoRapid Wissensportal*, *Beats Biblionetz*, *TouchGraph*, *Semantic Media Wiki* oder *OpenCalais* auf engeren Gebieten die semantische Strukturierung der gesammelten Informationen bereits jetzt erreicht. Graphische Übersichten ermöglichen eine gute Überschaubarkeit und Verortung der Fragestellung und der verfügbaren Quellen. Auch wenn dies von einer inhaltlichen Synthese noch weit entfernt ist, erlaubt es doch, Aspekte des eigenen Interessengebietes zu erkennen, die man vielleicht aus Fachblindheit heraus ausgeklammert hätte.

---

## 5.4 Verwalten und Auswerten von Dokumenten

Vor der wissenschaftlichen Auswertung und Synthese gefundener Information mag eine formale Informationsaufbereitung infrage kommen. Da der klassische Zettelkasten ausgedient hat, empfiehlt es sich, stattdessen mit einem Textverwaltungssystem wichtige Quellen und Notizen festzuhalten und zu strukturieren. Grundsätzlich können hierfür Word, Excel und Access genutzt werden. Angemessener sind aber Verwaltungssysteme, die unterschiedliche Dateitypen, Annotationen, inhaltliche Auszählungen und Strukturänderungen explizit zulassen. Im Prinzip wird dies bereits mit Textanalysesystemen wie *Maxqda*, *Atlas/ti*, *ASKSAM* oder *AutoMap* ermöglicht (siehe auch: Biste/Hohls 2000, Kuckartz/Rädiker, Kapitel 27 in diesem Band).

Spezielle *Literaturverwaltungssoftware*, die auch ein leichtes Abgleichen mit bibliographischen Angaben aus dem Internet erlaubt, wie *Zotero*, *MS OneNote*, *Citavi*, *BibTeX*, *EndNote*, *RefWorks*, *Lidos* sind aber bei einer umfangreichen Arbeit zu empfehlen.

Begriffsnetzwerke, Begriffskarrieren und weitere *bibliometrische Auswertungen* (siehe etwa: Delling 2010, Tunger 2007) oder graphische Verdichtungen können, wie vom *Digital Humanities Observatory* (<http://discovery.dho.ie/discover.php>) oder von Börner/Chen (2002) veranschaulicht, für die Bewertung der Informationsquellen von Nutzen sein, sollen hier aber nicht weiter ausgeführt werden.

Bevor eine Literaturliste erstellt wird, sollten die Regeln des betreffenden Institutes, der Fachzeitschrift oder des Verlegers eingesehen werden. Wer hiervon unabhängig ist, kann die internationalen Empfehlungen der *Modern Language Association (MLA)*, der *American Psychological Association (APA)* oder die entsprechenden Normen (z.B. ISO 690-2 (1997), DIN 1505-2 (1984)) zurate ziehen. Dies ist vor allem hilfreich, wenn graue Literatur, Materialien oder Internet-Links nachgewiesen werden müssen (Sawatzky 2011).

---

## 5.5 Wissenschaftliches Publizieren und Open Access

Forschungsergebnisse haben nur Bestand, wenn sie so publiziert werden (siehe auch Meyer/Meier und Friedrichs, Kapitel 15 und 16 in diesem Band), dass die wissenschaftliche Gemeinschaft sie auch künftig noch wahrnehmen und diskutieren kann. Wenn die eigene (z.B. studentische) Arbeit einen Wert auch für andere haben könnte, ist es sinnvoll,

die eigenen Gedanken zunächst auf einer Tagung oder in einer Fachzeitschrift anderen Wissenschaftlern mitzuteilen. So bekommt man fachliche Kritik, Anregungen und Hinweise sowie eine Rückmeldung, ob der Stand der eigenen Arbeit für andere hinreichend interessant ist. Wie oben erwähnt, führen auch Open Access-Fachzeitschriften oft ein Peer-Review durch. Auch ohne dieses ist im Falle der Ablehnung eine kurze Begründung zu erwarten. Hierdurch lernt man als Wissenschaftler mehr als durch jede bezahlte Publikation. Um zu sehen, wie eine Veröffentlichung sich letztlich niederschlägt, ist es ratsam, sich oder Kollegen in Referenz- oder Rezensionsdatenbanken abzufragen, z.B. mit der Auswertungssoftware *PublishorPerish*. Hier sieht man, wie ungleich verteilte („Matthäus-Effekt“ nach Merton 1985), aber auch ungeahnte Nutzungen von publizierten Arbeiten zustande kommen.

*Aufsätze in Fachzeitschriften* dienen einer möglichst raschen Veröffentlichung. *Sammelände* werden benutzt, um die Beiträge einer Tagung zu dokumentieren oder eine bestimmte Forschungsfrage von verschiedenen Autoren aus unterschiedlichen Blickwinkeln zu untersuchen. *Monographien* dokumentieren häufig die umfassende und detaillierte Darstellung des oder der Autoren zu einem geschlossenen Thema. Eine Sonderform ist das *Lehrbuch*: Der Inhalt soll nicht zuletzt Grundlagen des wissenschaftlichen Arbeitens vermitteln.

In Deutschland müssen alle *Doktorarbeiten* publiziert werden. Inzwischen bieten die meisten Universitätsbibliotheken eine kostenfreie Online-Veröffentlichung dieser Arbeiten an, ebenso für Sammelwerke und Aufsätze (oft Preprints) in fachlichen Repositorien.

Die auf Dokumentenservern online publizierte „*Graue Literatur*“ wird nach bibliothekarischen Standards katalogisiert und langzeitarchiviert. Auch für die Sozialwissenschaften wird eine fachlicher Dokumentenserver angeboten: das *Social Science Open Access Repository* (SSOAR), das auch die online-Version der Zeitschrift *Methoden, Daten, Analysen* (MDA) enthält. Darüber hinaus sind ECONSTOR und das *Social Science Research Network* (SSRN) zu nennen.

Das *Forum Qualitative Sozialforschung* und *Sociological Research Online* sind Beispiele für reine Online-Zeitschriften. Mit *Archivalia*, *OpenDOAR*, *MDPI*, *Hindawi* oder *Social Science Research Network* finden sich zahlreiche frei zugängliche Dokumentenserver. *PubMed Central* ist für Medizinozoologen und -statistiker, Epidemiologen, Gesundheitsökonomen usw. von großem Interesse. Nähere Bedingungen für die Weiterverwendung von frei zugänglichen Werken werden meist nach Standard-Lizenzverträgen auf Basis der *Creative Commons* definiert. Ebenso gibt es Initiativen zur freien Weitergabe von Daten (*Open Data*, bzw. *Linked Open Data*; Aschenbrenner 2011).

---

## 5.6 Quellen und Qualität

Oft stellt sich die Frage, welche Quellen wichtig sind und welche nicht. Qualität von Information lässt sich aber nur indirekt erfassen. So haben bestimmte Verlage einen guten Ruf in Bezug auf ihre sozialwissenschaftlichen Publikationen (dies gilt z.B. für *Sage* oder

*Springer VS). Sammelbände*, deren Beiträge einer Begutachtung unterzogen wurden, sind qualitativ höher einzuschätzen als solche ohne Begutachtung. *Fachzeitschriften* mit einem *Begutachtungsverfahren (Peer Review)* enthalten in der Regel bessere Artikel.

Mit dem „*Journal Impact Factor*“ des Web of Science wird versucht, Qualität auf quantitative Angaben zu reduzieren. Er gibt standardisiert an, wie oft Beiträge einer Zeitschrift von solchen in anderen Zeitschriften zitiert werden. Viele Fachzeitschriften werben mit ihrem Impact-Factor. Konkrete Hinweise zur Anwendung und Alternativen gibt die Universität Heidelberg (<http://www.umm.uni-heidelberg.de/bibl/emedien/jcr.html>).

Oft wird von diesem Impact-Factor für Zeitschriften auch auf die Qualität der darin veröffentlichten Autoren geschlossen, was aber nicht korrekt ist – hierzu wären eher der sog. *Hirsch-Index* (Egghe/Rousseau 2006) oder der *Science Impact Index* (Lehrl 1991) geeignet. Neu in der Diskussion sind auch Verfahren, die andere als zitationsmäßige Wirkungen messen, wie der *UsageFactor* von COUNTER ([http://www.projectcounter.org/usage\\_factor.html](http://www.projectcounter.org/usage_factor.html)).

Ein Argument gegen die Nutzung von *Websites als Quellen* ist deren ungesicherte Persistenz. Auftritte von staatlichen Institutionen, Nichtregierungsorganisationen, Unternehmen, Universitäten und anderen Forschungseinrichtungen sind in der Regel als stabiler einzuschätzen als Angebote von Privatpersonen. In diesen Fällen können solche Quellen relativ problemlos verwendet werden, doch ist es sinnvoll, gerade bei wissenschaftlichen Arbeiten immer nach anderweitig veröffentlichten Quellen zu recherchieren. Vorsicht ist aber auch bei Quellen von Unternehmen und zuweilen auch von Nichtregierungsorganisationen und staatlichen Institutionen geboten: Nicht immer ist hier Objektivität gegeben, sondern es spielen partikuläre Interessen eine Rolle.

Es ist einfach, bei *Wikipedia* zu schauen, was z.B. „soziales Kapital“ bedeutet. Doch hierzu – wie zu jedem beliebigen (sozial)wissenschaftlichen Thema – existiert eine Unmenge qualitativ hochwertiger Literatur, die weitaus präziser beschreibt, um was es geht.

*Lexika* sind zwar nützlich bei der ersten Klärung und Sichtung eines Themas, aber je nach Autorenschaft keine zitierfähigen Quellen. Als renommierte Lexika der Soziologie zu nennen sind dagegen Fuchs-Heinritz et al. (2011), Hillmann (2007), Endrweit/Trommsdorff (2002), Lamnek et al. (2000), bedingt auch Online-Versionen, wie: Internet-Lexikon der Methoden der empirischen Sozialforschung, Wörterbuch der Sozialpolitik „socialinfo“, Wirtschaftslexikon Gabler, „50 Klassiker der Soziologie“, das Lexikon-Portal der Bundeszentrale für politische Bildung (bpb).

Ähnliches gilt für *populärwissenschaftliche Zeitschriften*; sie zu nutzen ist nur sinnvoll, wenn es z.B. darum geht, den Verlauf einer öffentlichen Debatte, wie jene um Migration und Integration in Deutschland, nachzuzeichnen. Wenn es aber um theoretische Erläuterungen oder empirische Belege geht, sollte auf andere Quellen zurückgegriffen werden.

Wichtig ist es auch, die *Originalquellen* zu haben, und sei es als gescannte Datei aus dem Internet. Oft werden nämlich Schreib- und Interpretationsfehler von einer Quelle zur anderen weitergereicht, ohne dass es selbst Wissenschaftlern auffällt. Zu denken ist etwa an die „Sokal-Affäre“ (Sokal 1998), wo ein Nonsense-Text abgedruckt und oft zitiert wurde,

oder an die Tradierung des Fehlers des Weltklimarats IPCC bzgl. des Abschmelzens der Gletscher (Traufetter 2010). Von daher sind Projekte wie Google Books, Vorschauen in Bestelldatenbanken und das Gutenberg-Projekt sehr zu begrüßen.

---

## 5.7 Schlussbemerkungen

Der Zugriff auf und die Bereitstellung von Informationen ist für jede wissenschaftliche Disziplin essentiell – so auch für die Sozialwissenschaften. Open-Access-Angebote können heute in Umfang und Qualität durchaus mit kommerziellen Angeboten mithalten, so dass sie andere Angebote sinnvoll ergänzen und zuweilen sogar ersetzen können. Für angehende Wissenschaftler ist aber immer noch von Bedeutung, dass Open Access noch nicht jene Reputation genießt wie viele Fachzeitschriften kommerzieller Verlage.

Trotz der elektronischen Recherchemöglichkeiten sollte man nicht vergessen, reale Bibliotheken, Fachinformationsstellen und Forschungsinstitute zu besuchen. Denn so findet man oft Quellen, die man durch eine gezielte Katalogsuche vielleicht nicht erschlossen hätte. Die ‚freundlichen Regalnachbarn‘, also die Bücher neben den Büchern, die man eigentlich suchte, bieten häufig Einblicke in den weiteren Kontext eines Themas. Zudem bedeutet persönliche Kommunikation immer auch eine soziale Absicherung der eigenen Sichtweisen eines Themas.

## Literatur

- Aschenbrenner, Andreas/Neuroth, Heike (2011): 2.5 Forschungsdaten-Repositorien. In: Büttner, Stephan/Hobohm, Hans-Christoph/Müller, Lars (Hg.): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock+Herchen. 101-114
- Biste, Bärbel/Hohls, Rüdiger (Hg.) (2000): Fachinformationen und EDV – Arbeitstechniken für Historiker. Einführung und Arbeitsbuch. Historical Social Research. Supplement 12. Köln: Zentrum für Historische Sozialforschung
- Bittner, Sven/Müller, Andre (2011): Social Networking Tools and Research Information Systems: In: Webology 8 (1). URL: <http://www.webology.org/2011/v8n1/a82.html> am 12.9.2013
- Börner, Katy/Chen, Chaomei (2002): Visual Interfaces to Digital Libraries. Berlin/Heidelberg: Springer
- Buggisch, Christian (2011): Deutsche Social Media Guidelines. URL: <http://buggisch.wordpress.com/2011/10/12/deutsche-social-media-guidelines> am 12.9.2013
- Delling, Alexandra (2010): Bibliometrie als Bewertungsgrundlage für Wissenschaftler. Diplomarbeit. Potsdam: Fachhochschule Potsdam. URL: [http://pubman.mpdl.mpg.de/pubman/item/escidoc:825572:2/component/escidoc:825571/Diplomarbeit\\_Alexandra%20Delling.pdf](http://pubman.mpdl.mpg.de/pubman/item/escidoc:825572:2/component/escidoc:825571/Diplomarbeit_Alexandra%20Delling.pdf). am 12.9.2013
- DIN 1505-2 (1984): Titelangaben von Dokumenten: Zitierregeln. Berlin: Beuth
- Egghe, Leo/Rousseau, Ronald (2006): An informetric model for the Hirsch-index. In: Scientometrics 69 (1). 121-129
- Endruweit, Günter/Trommsdorff, Gisela (2002): Wörterbuch der Soziologie. Stuttgart: Lucius & Lucius
- Fries, Rüdiger/Kinstler, Till/Schweibenz, Werner/Strobel, Johannes/Weiland, Peter (2001): Was indexieren Suchmaschinen? In: BIT online 1: 49-56
- Fröhlich, Gerhard (1998): Online Classic. URL: [http://eprints.rclis.org/8840/1/online\\_classic.pdf](http://eprints.rclis.org/8840/1/online_classic.pdf) am 12.9.2013
- Fuchs-Heinritz, Werner/Klimke, Daniela/Lautmann, Rüdiger/Rammstedt, Otthein/Stäheli, Urs/Weischer, Christoph/Wienold, Hanns (2011): Lexikon zur Soziologie. Wiesbaden: VS Verlag
- Heidtmann, Frank (1985): Wie finde ich Literatur zur Volkswirtschaft, Betriebswirtschaft, Psychologie, Soziologie, Politikwissenschaft, Publikationswissenschaft, Statistik? Berlin: Berlin-Verlag
- Hillmann, Karl-Heinz (2007): Wörterbuch der Soziologie. Stuttgart: Kröner
- Hohlfeld, Michael (2012): Fachportale. URL: <http://blogs.sub.uni-hamburg.de/webis/fachportale> am 12.9.2013
- ISO 690-2 (1997): Information and documentation – Bibliographic references – Part 2: Electronic documents or parts thereof. Berlin: Beuth
- Kaminsky, Reiner (1977): Information und Informationsverarbeitung als ökonomisches Problem. München: Verlag Dokumentation

- Kuhlen, Rainer (1989): Pragmatischer Mehrwert von Information: Sprachspiele mit informationswissenschaftlichen Grundbegriffen. Bericht 1/89. Konstanz: Informationswissenschaft an der Universität Konstanz
- Kwak, Haewoon/Lee, Changhyun/Park, Hosung/Moon, Sue (2010): What is Twitter, a Social Network or a News Media? In: International World Wide Web Conference Committee (Hg.): WWW 2010. Proceedings of the 19th International World Wide Web (WWW) Conference. April 26-30, 2010. Raleigh (NC): ACM. 591-600. URL: <http://an.kaist.ac.kr/~haewoon/papers/2010-www-twitter.pdf> am 12.9.2013
- Lamnek, Siegfried/Recker, Helga/Reinhold, Gerd (2000): Soziologie-Lexikon. München: Oldenbourg Wissenschaftsverlag
- Lehnhard-Bruch, Susanne (2009): „Wie komme ich dran?“ URL: [http://www.hbz-nrw.de/dokumentencenter/produkte/digitale\\_bibliothek/jubilaeum/vortraege/lehnardbruch.pdf](http://www.hbz-nrw.de/dokumentencenter/produkte/digitale_bibliothek/jubilaeum/vortraege/lehnardbruch.pdf) am 12.9.2013
- Lehrl, Siegfried (1991): Der Science Impact Index als Maß der Durchsetzung sowie der Forscherqualität und -kapazität von Wissenschaftlern. Nürnberg: Media Point Verlagsgesellschaft
- Merton, Robert K. (1985): Der Matthäus-Effekt in der Wissenschaft. In: Merton, Robert K. (Hg.): Entwicklung und Wandel von Forschungsinteressen. Frankfurt a.M.: Suhrkamp. 147-171
- Nentwich, Michael/König, René (2011): Wissenschaft und Social Network Sites – Steckbrief 5 im Rahmen des Projekts Interactive Science. ITA-Projektbericht Nr. 52-5. Wien: Institut für Technikfolgen-Abschätzung. URL: [http://epub.oeaw.ac.at/0xc1aa-500d\\_0x002530f9.pdf](http://epub.oeaw.ac.at/0xc1aa-500d_0x002530f9.pdf) am 12.9.2013
- O'Reilly, Tim (2005): What Is Web 2.0? URL: <http://oreilly.com/web2/archive/what-is-web-20.html> am 12.9.2013
- Plassmann, Engelbert/Rösch, Hermann/Seefeldt, Jürgen/Umlauf, Konrad (2011): Bibliotheken und Informationsgesellschaft in Deutschland. Wiesbaden: Harrassowitz
- Sawatzky, Elisabeth (2011): Zitieren elektronischer Quellen. URL: <http://lotse.uni-muenster.de/toolbox/00035270.pdf> am 12.9.2013
- Sen, William (2010): Social Media Measurement. URL: <http://www.social-media-magazin.de/index.php/inhalt/social-media-measurement.html> am 12.9.2013
- Sherman, Chris (2005): Metacrawlers and Metasearch Engines. URL: <http://searchenginewatch.com/article/2066974/Metacrawlers-and-Metasearch-Engines> am 12.9.2013
- Sokal, Alan D. (1998): What the Social Text Affair Does and Does Not Prove. Oxford: Oxford University Press
- Traufetter, Gerald (2010): Weltklimarat schlampete bei Gletscher-Prognosen. In: Spiegel Online. URL: <http://www.spiegel.de/wissenschaft/natur/0,1518,672709,00.html> am 19.01.2010
- Tunger, Dirk (2009): Bibliometrische Verfahren und Methoden als Beitrag zu Trendbeobachtung und -erkennung in den Naturwissenschaften. Jülich: Forschungszentrum Jülich. URL: <http://juwel.fz-juelich.de:8080/dspace/bitstream/2128/3593/1/9783893365500.pdf> am 12.9.2013

ULB Münster (2010): Vom Thema zu den Suchbegriffen. URL: <http://lotse.uni-muenster.de/toolbox/00035241.pdf> am 12.9.2013

ULB Münster (2013) : „In den Tiefen des Internet“. Teil 2: Wo finden Sie was? URL: <http://lotse.uni-muenster.de/toolbox/00035190pdf>. am 12.9.2013

Aglaja Przyborski und Monika Wohlrab-Sahr

---

## 6.1 Einleitung: Warum braucht die qualitative Sozialforschung ein besonderes Forschungsdesign?

Bei der Verwendung qualitativer Methoden ist nicht immer klar, was die Durchführung qualitativer Forschung im Einzelnen bedeutet und welchen methodischen Grundlagen und Standards sie genügen muss. Das mag auch daran liegen, dass wir erst am Beginn einer systematischen Integration qualitativer Methoden in die Curricula human- und sozialwissenschaftlicher Studiengänge stehen, von der Besetzung entsprechender Professuren ganz abgesehen.

Auch in Forschungsprojekten trifft man nicht selten auf solche Probleme. Die Forschenden haben oft nicht oder nur am Rande gelernt, wie eine qualitative Untersuchung im Einzelnen abläuft und von welchen wissenschaftstheoretischen Annahmen sie ausgeht. Auch glauben immer noch viele, bei Projektanträgen oder Qualifikationsarbeiten gegenüber Gutachtern Kompromisse machen zu müssen, so dass am Ende ein wenig erfreuliches Durcheinander von „qualitativem“ Material, paraphrasierendem Nachvollzug und quantifizierender Interpretation steht. Daraus resultiert oft unter der Hand eine Forschungslogik, die letztlich an standardisierten Verfahren orientiert ist, ohne freilich deren Kriterien Genüge tun zu können.

Dieser Beitrag will helfen, ein „Durcheinander“ dieser Art zu vermeiden. Für die qualitative wie auch für die quantitative Forschung gilt: Untersuchungen müssen auf dem Niveau durchgeführt werden, das auf dem entsprechenden methodischen Gebiet erreicht ist. Es gilt, Fragestellungen zu wählen, die mit einer qualitativen Vorgehensweise sinnvoll bearbeitet werden können, die ihr angemessen sind und die Ergebnisse zutage fördern, die so mit anderen Verfahren nicht hätten hervorgebracht werden können. Qualitative Arbeiten sind keine „kleinen“ Varianten von Untersuchungen, die man im Prinzip auch als standardisierte Befragung (Reinecke, Kapitel 44 in diesem Band) oder mit Hilfe eines experimentellen Designs (Eifler, Kapitel 11 in diesem Band) durchführen könnte. Sie sind

aber auch keine „weiche“ Forschung, bei der es nur gilt, möglichst nahe an den Aussagen der befragten oder beobachteten Personen zu bleiben und bei der ein Generalisierungsspruch von vornherein ausgeschlossen wird.

Das *Forschungsdesign* macht die Art und Weise, wie eine Untersuchung angelegt ist, explizit. Es hilft, sich das Ineinandergreifen der verschiedenen Schritte des Forschungsablaufs klar zu machen und diesen auf seine Stimmigkeit hin zu überprüfen. Damit bietet es eine Leitlinie, auf die im Verlauf der Forschung immer wieder zurückgegriffen werden kann. Und schließlich macht es die Ressourcen, die für ein Forschungsvorhaben aufgewendet werden müssen, planbar und hilft diese – im Fall eines Forschungsantrags – zu legitimieren.

Eine Besonderheit von qualitativen Forschungsdesigns liegt darin, dass sie mehr zirkuläre Elemente beinhalten als quantitative Designs, die in ihrer Grundstruktur in der Regel linear-chronologisch aufgebaut sind. Dies betrifft vor allem das Ineinandergreifen von Formulierung (und ggf. Re-Formulierung) der Fragestellung, Erhebung und Auswertung bis hin zu Generalisierung und Theoriebildung, die nicht in säuberlich getrennten Schritten nacheinander absolviert werden können.

Ein weiterer zentraler Unterschied liegt in der Forschungslogik begründet: Avancierte qualitative Verfahren sind rekonstruktiv ausgerichtet. Dies gilt zunächst für den Gegenstand der Forschung, bei dem es darum geht, eine Forschungsfrage auf der Grundlage der *Rekonstruktion von sozialem Sinn* (Schütz 2004a) zu beantworten: Wir rekonstruieren also etwas, das bereits in sich sinnhaft ist und dessen Sinn es zu erschließen und in wissenschaftliche Konzepte zu übertragen gilt. Aber auch hinsichtlich ihrer Methode sind qualitative Ansätze rekonstruktiv. Man fragt immer auch, mit welcher Vorgehensweise man zu bestimmten Ergebnissen gekommen ist, d.h. man rekonstruiert auch das Forschungshandeln selbst und fasst es theoretisch. Das hat wichtige Konsequenzen für das Forschungsdesign.

Das Entwerfen eines Forschungsdesigns ist das Ergebnis

1. einer Auseinandersetzung mit dem eigenen Erkenntnisinteresse und dessen Übersetzung in eine Fragestellung,
2. einer entsprechenden methodologischen Positionierung,
3. der Bestimmung des Forschungsfeldes,
4. einer im Hinblick auf ihre Implikationen reflektierten Wahl der Erhebungs- und Auswertungsverfahren,
5. von Entscheidungen über das Sampling und die damit möglichen Schritte der Generalisierung bis hin zur Theoriebildung sowie
6. der grundlagentheoretischen Einbettung der Forschung.

Die immer wieder zu hörenden verzweifelten Fragen – „Ich habe Daten erhoben, aber wie kann ich sie nun auswerten?“ oder „Ich habe Daten erhoben, aber welche Theorie passt‘ nun dazu?“ – kommen so gar nicht erst auf.

Wir werden im Folgenden anhand von praktischen Beispielen die Grundentscheidungen behandeln, die im Hinblick auf eine geplante qualitative Forschung zu treffen sind und die Schritt für Schritt zu einem Forschungsdesign führen.

## 6.2 Erster Schritt: Formulierung des Erkenntnisinteresses und der Fragestellung

Das Interesse für qualitative Forschung entzündet sich oft an der Neugier in Bezug auf konkrete Phänomene und Beobachtungen, seien es das ostentative Zur-Schau-Stellen von Reichtum, der auffällige Wechsel in eine andere, ‚fremde‘ Religion oder bestimmte Formen des Umgangs mit Mode und Körperschmuck. Diese Neugier ist ohne Zweifel eine wichtige erste Grundlage, um sozialen Prozessen auf die Spur zu kommen: Mit der Frage „What the hell is going on here?“ hat Erving Goffman (1980) die daraus resultierende Haltung charakterisiert. Aber natürlich reicht bloße Neugier nicht aus. Sie muss zu einer differenzierten Bestimmung des Phänomens, das von Interesse ist, führen und schließlich in eine präzise Fragestellung einmünden. Wenn hier nachlässig verfahren wird, wird sich dies im späteren Verlauf der Forschung rächen. Phänomene sind noch keine Fragestellungen, und ohne klar formulierte Fragestellung lässt sich kein Forschungsdesign entwickeln.

Wenn eine Doktorandin z.B. vorhat, sich in ihrer Dissertation mit *Maßnahmen zur Integration von Migranten* zu befassen, hat sie damit zwar einen Phänomenbereich bestimmt, aber noch keine Fragestellung formuliert. Zunächst bedarf es einer – wenn auch vorläufigen – Definition des Phänomens. Womit hat man es zu tun, wenn man solche Maßnahmen in den Blick nimmt, und welche Maßnahmen kommen in Frage? Zu denken wäre beispielsweise an Kurse verschiedener Art: Sprachkurse, Integrationskurse, speziell auf Migranten zugeschnittene kulturelle oder religiös gerahmte Angebote und anderes mehr.

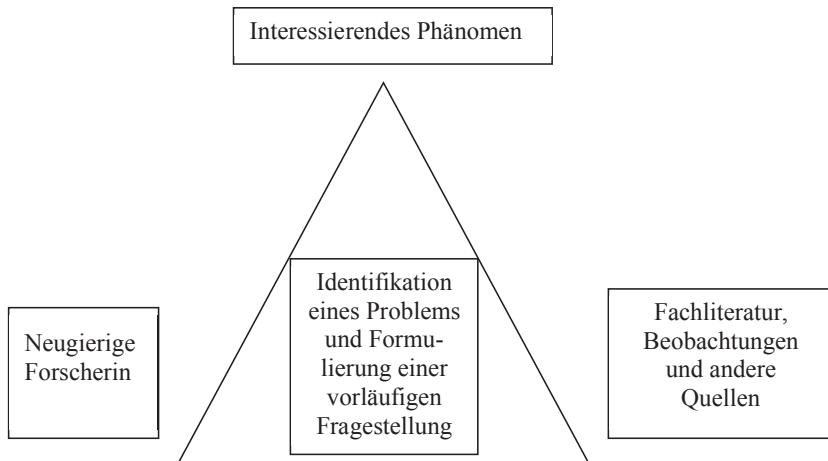
Hier lassen sich bereits erste Differenzierungen vornehmen: Handelt es sich um Angebote, die dem Erlernen der Sprache und damit dem *Kompetenzerwerb* dienen oder um solche, in denen es vor allem um die rechtliche, politische und kulturelle *Orientierung* im Aufnahmeland geht? Handelt es sich um Kurse, die *freiwillig* in Anspruch genommen werden oder um solche, die *verpflichtend* sind und deren Absolvieren *Bedingung* ist für etwas anderes, beispielsweise die Gewährung eines bestimmten aufenthaltsrechtlichen Status? Oder trifft in den avisierten Kursen beides über verschiedene Personengruppen aufeinander: Personen, die den Kurs freiwillig absolvieren und solche, die ihn absolvieren müssen? Gibt es *Verbindungen* zwischen Sprachkursen und sog. Orientierungskursen für Personen, die die Staatsbürgerschaft des Aufenthaltslandes erwerben wollen? Wer bietet die Kurse an? Wer unterrichtet in ihnen? Gibt es staatliche Förderformate für Angebote dieser Art? Sind die Angebote kostenfrei oder kostenpflichtig? Gibt es Personen, die solche Kurse mehrfach absolvieren bzw. Veranstaltungen mehrfach oder regelmäßig besuchen? Gibt es spezifizierte Angebote für bestimmte Interessenlagen und Vorkenntnisse? Sind die Gruppen nach bestimmten Kriterien (z.B. Sprachkompetenz, Herkunftsländer, Geschlecht, Alter) zusammengesetzt? usw.

Diese vorläufige Phänomenbestimmung gibt schon erste Aufschlüsse über mögliche Konfliktlinien: *Freiwilligkeit vs. Verpflichtung, Hilfe vs. Kontrolle* (der Integrationswilligkeit), *Unterhaltung vs. Erziehung, intrinsische vs. strategische Motivation, soziale vs. sachliche Interessenlagen, divergierende Kompetenzen und Interessen* u.a.m. Vor diesem Hintergrund gilt es nun, eine Fragestellung zu formulieren, die einerseits einem rekonstruktiven Vorgehen (Przyborski/Wohlrab-Sahr 2014; Meuser 2003) adäquat ist und die andererseits die richtige Verbindung von Spezifität und Offenheit beinhaltet.

Es ist klar, dass die Spezifikation der Fragestellung in verschiedener Hinsicht vorgenommen werden kann: im Hinblick auf diejenigen, die solche Angebote wahrnehmen, im Hinblick auf die Institutionen und Personen, die Angebote machen oder sie durch ihre Vorgaben mit prägen, oder im Hinblick auf die soziale Situation in den entsprechenden Maßnahmen selber, also dem Ort, an dem die unterschiedlichen Perspektiven aufeinandertreffen.

Vor dem Hintergrund dieser Überlegungen lässt sich nun das Thema näher eingrenzen und eine Fragestellung formulieren: Gehen wir einmal davon aus, dass wir uns für „Kurse“ interessieren, die zum Zwecke der Förderung von Integration eingerichtet wurden, und zwar als sozialem Ort, an dem unterschiedliche Perspektiven und Interessen aufeinandertreffen. Besonders interessieren uns dabei die expliziten und impliziten „Botschaften“, die in den entsprechenden Kursen transportiert werden bzw. von den Teilnehmern aufgenommen werden. Was ist es, was neben der Sprachkompetenz oder den sachlichen Informationen in solchen Kursen noch „gelernt“ und „gelehrt“ wird? Oder – anders formuliert – in welche „Botschaften“ transferieren die Teilnehmer das, was ihnen im Kurs begegnet? Dabei interessieren uns nicht primär die länderkundlichen Gehalte der Lehrbücher, sondern *die informellen und impliziten Botschaften, die in den Kursen vermittelt oder „verstanden“ werden bzw. die bei den Teilnehmern „ankommen“*: über das Aufnahmeland und seine Regeln und Selbstverständlichkeiten, über den Umgang mit Neuankömmlingen und das, was man für ihre „Kultur“ hält, über das, was von ihnen erwartet wird usw. In welchem Zusammenhang stehen das Sprachenlernen und solche anderen – impliziten – „Botschaften“, Orientierungen und Wissensbestände der Kurse? Und vor welchem Hintergrund wird auf der Seite der Kursteilnehmerinnen das eine oder andere fokussiert?

In diesem Zusammenhang fällt bereits auf, dass das Formulieren einer Fragestellung im Rahmen qualitativer Sozialforschung komplexer „gebaut“ ist und weniger „exakt“ auf eine abhängige (zu erklärende) und einige wenige unabhängige (erklärende) Variablen reduziert ist. Im Rahmen quantitativer Forschung könnte es im Zusammenhang mit den genannten Angeboten, etwa bei Sprachkursen, z.B. um die Frage gehen, wovon der erfolgreiche Erwerb von Sprachkompetenz bei Migranten (abhängige Variable) abhängt. Hier üblicherweise ins Spiel gebrachte erklärende Variablen sind: die Anzahl deutscher Freunde, das im Verlauf des Lebens angesammelte Humankapital oder die gefühlte „Nähe“ zum Aufnahmeland. Ging es im Rahmen standardisierter Forschung also darum zu erforschen, welcher Einflussfaktor *letztlich* ein bestimmtes Resultat *erklärt*, geht es im Rahmen qualitativer Forschung darum, einen Sinnzusammenhang und die (persönlichen, institutionellen oder situativen) *Mechanismen*, die in ihm wirken, zu erfassen (d.h. zu *verstehen*).



**Abb. 6.1** Forschungstrias zur Identifikation von Fragestellungen

Allerdings bedeutet dies nicht, dass ein solcher verstehender Zugang nichts zur Erklärung beiträgt. Wie bereits Max Weber (1988) in seinen methodologischen Schriften deutlich gemacht hat, ist das Sinnverstehen geradezu Voraussetzung für die Möglichkeit, etwas zu erklären. Wenn man rekonstruieren kann, welche „Botschaften“ bei Kursteilnehmern von Sprachkursen „ankommen“, kann man auch die Wirkungen solcher Maßnahmen und damit ihren Erfolg oder Misserfolg besser erklären.

Entsprechend anders wäre auch bei beiden paradigmatischen Zugängen die Vorbereitung mit Hilfe von Forschungsliteratur. Beim standardisierten Vorgehen dient sie dazu einen noch nicht hinreichend erklärten Kausalzusammenhang zu identifizieren, der dann Gegenstand der Forschung sein soll. Bei der qualitativen Forschung dient sie – neben Beobachtungen, Gesprächen und anderem mehr – dazu, ein Problem zu identifizieren, bei dem man komplexere Zusammenhänge vermutet, die es mit rekonstruktiven Verfahren zu erfassen gilt.

Das Beispiel der Integrationsmaßnahmen zeigt aber auch, dass zu ein- und demselben Phänomenbereich durchaus unterschiedliche Fragestellungen entwickelt werden können. Die möglichst genaue Formulierung der Fragestellung und des Erkenntnisinteresses ist ein wesentlicher erster Schritt der Forschung, in dem auch bereits vorläufige theoretische Konzepte auftauchen.

Das bedeutet nicht, dass sich an dieser ersten Formulierung der Fragestellung im Verlauf der Forschung nichts mehr ändern kann. Qualitative Forschung zeichnet sich gerade dadurch aus, dass sie ihre Fragestellungen, Konzepte und Instrumente in Interaktion mit dem Forschungsfeld immer wieder überprüft und anpasst. Auf Grundlage einer präzise formulierten Fragestellung zu Beginn lässt sich aber später genauer bestimmen und dokumentieren, wo und aus welchen Gründen sich Perspektiven im Verlauf der Forschung verändert haben.

### **6.3 Zweiter Schritt: Methodologische Positionierung und metatheoretische Verortung**

Mit der Formulierung der Fragestellung eng verschränkt ist etwas, das wir *methodologische Positionierung* nennen wollen. Damit ist kein die andere Seite ideologisch ausgrenzendes „Bekenntnis“ zu qualitativen Verfahren gemeint. Wenn man sich für das interessiert, was in Kursen, die der Integration dienen sollen, jenseits der offiziellen Programme gelernt und gelehrt wird, könnte man natürlich auch eine standardisierte Befragung unter allen Teilnehmern durchführen und sie auf einer Skala ankreuzen lassen, was sie in dem Kurs glauben gelernt zu haben und was eher nicht. Parallel dazu könnte man die Kursleiter nach ihren Lernzielen und dem vermuteten Lernerfolg befragen und beides vergleichen. Hier möglicherweise aufgedeckte Diskrepanzen wären zweifellos hoch aufschlussreich für die Frage nach dem, was Teilnehmer aus einem solchen Kurs mitnehmen und für die Illusionen, die sich die Lehrenden unter Umständen davon machen.

Es wäre für manchen Absolventen sicherlich verführerisch, ein solches Design einfach zu „verkleinern“, indem er mehrere halbstandardisierte Leitfadeninterviews (Przyborski 2010: 138ff, Hopf 1978) mit Vertretern beider Gruppen führt. Einem qualitativen Design entspräche eine solche methodologische Orientierung allerdings noch nicht.

Generell scheint es uns eine gute Leitlinie zu sein, sich zu Beginn der Forschung zu fragen, ob man für das, was man machen will, qualitative Verfahren *braucht*, oder ob sich dieselben Ergebnisse, noch dazu vielleicht umfassender und aussagekräftiger, mit quantitativen Verfahren erzielen ließen. Um eines der standardisierten Erhebungsinstrumente herauszugreifen: Gibt es in dem Forschungsfeld etwas, das sich nicht ohne weiteres oder nicht sinnvoll mittels Fragebogenerhebung erforschen lässt? Gibt es an dem, was uns interessiert, etwas, das z.B. offene Beobachtungs- oder Befragungsformen *braucht*, das man also ohne diese nicht herausbekommen würde. Auch wenn diese Formulierung nach Abgrenzung klingt, ist sie nicht polemisch gemeint, sondern dient der klaren Aufgabenbestimmung der paradigmatischen Zugänge. Nur dann, wenn die jeweiligen Zuständigkeiten – damit aber auch die Grenzen von deren Reichweite – klar benannt sind, können die standardisierten und die rekonstruktiven Verfahren in eine produktive – aber eben auch einander ergänzende – Aufgabenteilung eintreten.

Für unser Beispiel heißt das: Von Interesse ist nicht primär das, was man in einem Fragebogen abfragen könnte. Von Interesse ist also nicht, wie sich positive oder kritische Einschätzungen des Angebots in verschiedenen Kursen darstellen und von welchen Faktoren (z.B. Eigenschaften der Teilnehmer, Eigenschaften der Kursleiter) diese Einschätzungen vorrangig beeinflusst werden. Von Interesse sind eher die *impliziten Organisationsprinzipien und Wirkungen* sozialer Situationen, die Arten des aufeinander Reagierens, die Typisierungen, die im Hinblick auf die soziale Situation, auf die Anderen und einen selbst vorgenommen werden und die gerade dort zum Ausdruck kommen, wo die Akteure versuchen, sich in ihr Gegenüber hineinzuversetzen.

Diese Prozesse sind, da sie in Alltagsroutinen eingelassen und den Beteiligten selbst nur in Grenzen bewusst sind, nicht einfach abzufragen, sondern müssen beobachtet werden. Gleichwohl könnte man eine solche Beobachtung durch offene Befragungen mit Kursteilnehmern und -veranstaltern ergänzen (Heinig 2010).

Auf die Unterscheidung verschiedener methodologischer Paradigmen zu pochen, bedeutet nun aber umgekehrt keinen Methodenpurismus. Es kann ausgesprochen sinnvoll sein, in aufeinander folgenden Arbeitsschritten verschiedene methodische Zugänge zu verknüpfen: etwa auf der Grundlage einer standardisierten Befragung unter Kursteilnehmern gezielt Kurse auszuwählen, in denen dann solche informellen Mechanismen näher untersucht werden. Dies könnten etwa Kurse sein, bei denen sich in der Befragung zeigt, dass die Einschätzungen der Teilnehmer und die der Kursleiter extrem auseinanderfallen und – im Kontrast – solche, bei denen beides in hohem Maße übereinstimmt. Aber auch bei einer solchen Verknüpfung unterschiedlicher methodischer Vorgehensweisen, die man als „Triangulation“ bezeichnet, wäre bei jedem Arbeitsschritt die „Logik“ des dabei zugrunde gelegten methodischen Vorgehens zu berücksichtigen. Die Verknüpfung unterschiedlicher methodischer Zugänge erfordert daher weitgehende methodische Kompetenzen: Man muss sich in den standardisierten Verfahren ebenso zu Hause fühlen wie in den interpretativen, man muss wissen, an welcher Stelle man Aussagen welchen Typs trifft und wie beide aufeinander zu beziehen sind.

Mit dieser methodologischen Positionierung einher geht zu Beginn eines rekonstruktiven Forschungsprozesses auch die Entscheidung für eine formale bzw. *Metatheorie*. Darin werden begrifflich-theoretische Grundlagen erfasst, die mit dem Gegenstand, auf welchen sich das Erkenntnisinteresse richtet, nur mittelbar etwas zu tun haben, wie z.B. die Konzepte des Fremdverständehens (Schütz 2004b) und der reziproken Typisierung (Schütz 1971), das Konzept der Identität (Goffman 1975, Wohlrab-Sahr 2006), des Handelns (Weber 1980: 1-30), der Kollektivität (Mannheim 1980), des Motivs (Schütz 2004), der Kommunikation (Garfinkel 2004) oder der Orientierung (Bohnsack 2007). Phänomen, Erkenntnisinteresse, Methode und Metatheorie müssen aufeinander bezogen sein. Die Auseinandersetzung mit Grundlagentheorien strukturiert die Wahl der Methoden sowie die der Erhebungs- und Auswertungsverfahren. Ein theoretischer Fokus auf Kollektivität und Interaktion wird sich mit anderen methodischen Instrumenten verbinden als einer, der auf biographischer Identität und Sozialisation liegt.

---

## 6.4 Dritter Schritt: Identifikation des Forschungsfeldes

Wenn man sich für Angebote interessiert, in denen es um die Integration von Zuwanderern geht, kommt ein breites Forschungsfeld in den Blick, in dem sich verschiedene Angebote untersuchen ließen: Integrationskurse (Heinig 2010), Einbürgerungsfeiern (Jakob 2012), besondere, auf Migranten ausgerichtete Kulturangebote oder auch Deutschkurse für Ausländer. In all diesen Feldern ließe sich gut untersuchen, ob es neben den

Sachinformationen und formalen Kompetenzen auch „implizite Botschaften“ gibt, die eher im kommunikativen Bereich anzusiedeln sind, und zu analysieren, wie diese in der Kommunikation generiert werden. Auch Sprachkurse, in denen es vorrangig um das Erlernen einer formalen Kompetenz geht, enthalten Hinweise auf kulturelle Regeln und Selbstverständlichkeiten und bestehen aus Formen der Kommunikation, in denen Einschlüsse und Ausschlüsse generiert werden. Überdies sind sie bisweilen auch mit anderen Angeboten – etwa mit sog. „Integrationskursen“ – gekoppelt und in dieser Verknüpfung wiederum Bedingung für die Möglichkeit der Einbürgerung, die dann bei entsprechenden „Einbürgerungsfeiern“ gewürdigt werden kann. Kulturträger wiederum entwickeln Angebote, etwa im Bereich von Musik und Theater, die auf die Interessen von bestimmten Zuwanderergruppen zugeschnitten sind und sich aus der Logik des „Forderns und Förderns“ der staatlichen Angebote zu befreien versuchen.

Insgesamt eröffnet sich hier also ein großes Forschungsfeld, in dem verschiedene Maßnahmen ineinander greifen. Dieses Feld kann sicherlich nicht „komplett“ untersucht werden, allerdings ist es wichtig, sich diese Verbindungen klar zu machen. So haben Teilnehmer an Sprachkursen möglicherweise bereits Erfahrung mit anderen Formaten und interpretieren das, was sie im Sprachkurs erleben, vor diesem Hintergrund.

Grenzt man nun das Forschungsfeld – aus pragmatischen Gründen – ein auf *Deutschkurse*, gilt es diesen Horizont von „Integrationsangeboten“ mit zu bedenken. Allerdings ist dies nicht der einzige Horizont. In den Kursen könnten auch Personen sein, die für ein paar Jahre in Deutschland studieren oder die aus beruflichen oder familiären Gründen für eine gewisse Zeit nach Deutschland gekommen sind. Diese Personen werden sich unter Umständen in ihrer Motivation, bisweilen aber auch in ihrem sozialstrukturellen Hintergrund, von anderen Teilnehmergruppen unterscheiden. Ein wichtiger Unterschied könnte zum Beispiel sein, dass sie freiwillig – und nicht gezwungenermaßen – an den Kursen teilnehmen. Im Forschungsfeld überlagern sich also verschiedene Horizonte, die es etwa beim Sampling zu berücksichtigen gilt. Gleichzeitig stellt sich die Frage, ob die unterschiedlichen Personengruppen die Angebote anders wahrnehmen oder ob sie von Seiten der Kursleiter anders eingeschätzt und behandelt werden. Gibt es unterschiedliche Botschaften für verschiedene Gruppen von Teilnehmern und – wenn ja – entlang welcher Kriterien werden in der Kommunikation die Grenzen gezogen?

---

## **6.5 Vierter Schritt: Entscheidung für ein konkretes methodisches Verfahren**

Damit ist man nun gut vorbereitet auf die Frage nach der Methode im engeren Sinn: Über welche konkreten methodischen Operationen – d.h. welche Formen der Erhebung und Auswertung – fördert man das aufschlussreichste Material und entsprechend aussagekräftige Ergebnisse zutage? Entscheidungen, die es hier zu treffen gilt, sind etwa, ob eine Form der Befragung (Przyborski/Wohlrab-Sahr 2014: 53ff) oder eher die teilnehmende Beobachtung (Lüders 2003) das Instrument der Wahl sind; oder ob es möglicherweise

Dokumente oder Abbildungen (Bohnsack 2008) auszuwerten gilt, die in dem jeweiligen Forschungsfeld eine Rolle spielen.

Spielen wir – wieder mit Bezug auf unser Thema – die möglichen Instrumente durch: Eine Variante wäre die teilnehmende Beobachtung. Das hieße, dass die Forscherin – in diesem Fall nur nach Absprache mit der Kursleitung – als eine Art Praktikantin am Kurs teilnimmt und die Abläufe in Beobachtungsprotokollen festhält (Heinig 2010). Dies wäre sicherlich das Instrument der Wahl, wenn es darum geht, die Kommunikation im Kurs zu interpretieren: Wie werden Personen angesprochen, wie bringen sie sich selbst ins Spiel, zu welchen Beispielen greifen die Interagierenden, wenn sie etwas erläutern wollen? Auf welches „Wissen“ über die jeweils andere Seite und den gegebenen Kontext wird in der Kommunikation rekurriert, welche Erfahrungen werden geltend gemacht, und welche Schlussfolgerungen werden daraus gezogen? Dabei dürfte allerdings die Forschungsfrage gegenüber der Kursleitung nur in relativ allgemeiner Weise kommuniziert werden, um Artefakte zu vermeiden. Im Prinzip jedenfalls wäre hier die Beobachtung ein geeignetes Instrument, um in der Kommunikation transportierte implizite Botschaften erkennbar zu machen.

Nun kann es sein, dass eine solche Beobachtung Anhaltspunkte für herausgehobene Positionen im Kurs oder für Gruppenbildungen gibt. Insofern könnten sich an die Beobachtung verschiedene weitere Formen der Erhebung anschließen: Einzelinterviews mit Kursteilnehmern, die im Kurs in irgendeiner Weise „aufgefallen“ sind: weil sie sich oft äußern oder gar nicht, weil sie sich kritisch positionieren oder besonders angepasst erscheinen, weil sie über gute oder über wenig Vorbildung verfügen usw. Es können auch Gruppen erkennbar werden, weil sie wiederholt in räumlicher Nähe zueinander sitzen, sich aufeinander beziehen oder von der Kursleiterin gleichsam „kollektiv“ angesprochen werden. Solche Beobachtungen könnten dann ein Anhaltspunkt dafür sein, Befragungen – etwa in Form von Gruppendiskussionen (Vogl, Kapitel 41 in diesem Band) – in der jeweiligen Gruppe durchzuführen, weil sich im Gruppenkontext Gemeinsamkeiten des Erlebens oft besser artikulieren lassen. Allerdings gehen dabei herausgehobene individuelle Positionen eher unter als beim Einzelinterview. Diese wenigen Beispiele machen bereits deutlich, dass die Entscheidung für bestimmte Formen der Erhebung eine voraussetzungsreiche Entscheidung ist, die es im Hinblick auf das, was sie zutage fördert, aber auch im Hinblick darauf, welche Wirkungen sie erzeugt, zu reflektieren gilt.

In engem Zusammenhang mit der Entscheidung für Formen der Erhebung steht auch die Frage nach den geeigneten Auswertungsverfahren (Przyborski/Wohlrab-Sahr 2014: 189ff, Kleemann et al. 2009). Dabei wollen wir keinem Purismus der Ansätze das Wort reden. Bei genauerer Betrachtung teilen die diversen Kunstreihen im Feld qualitativer Forschung – etwa Objektive Hermeneutik (Wernet 2000), Narrationsanalyse (Schütze 1983), Dokumentarische Methode (Bohnsack 2011), Biographieforschung (Rosenthal, Kapitel 35 in diesem Band), sozialwissenschaftliche Hermeneutik (Kurt/Herbrik, Kapitel 33 in diesem Band), Ethnografie (Knoblauch, Kapitel 36 in diesem Band) und Grounded Theory-Methodologie (Strübing, Kapitel 32 in diesem Band) – eine ganze Reihe von Grundannahmen und darauf bezogenen Regeln und Praktiken. Wesentlich ist aber, dass Formen der

Erhebung und Auswertung eng aufeinander bezogen sind. Oft sind bestimmte hermeneutische Auswertungsverfahren nicht zu praktizieren, weil während der Erhebung den Befragten kein Spielraum für selbstläufige Darstellungen gelassen wurde. Ein Auswertungsverfahren, bei dem es auf die Interaktion in Gruppen und die Analyse von Gruppendiskursen ankommt, läuft ins Leere, wenn während der Gruppendiskussion ständig einzelne Personen angesprochen wurden, die dann nacheinander – aber eben nicht aufeinander bezogen – antworten. Ein Auswertungsverfahren, das für die Rekonstruktion biographischer Prozesse entwickelt wurde, kann nicht zur Anwendung kommen, wenn keine Interviews geführt wurden, in denen solche Prozesse selbstläufig zur Darstellung kommen. Demnach wäre die letzte, zu Beginn der Forschung zu beantwortende Frage: Sind Erhebungsform und Auswertungsverfahren so aufeinander abgestimmt, dass das erhobene Material eine geeignete Grundlage für das gewählte Auswertungsverfahren bildet?

---

## 6.6 Fünfter Schritt: Sampling

Auch in qualitativen Studien ist eine Entscheidung darüber zu treffen, unter Einbeziehung welcher *Erhebungseinheiten* die Forschung durchgeführt werden soll. Damit eng verbunden ist die Frage, worüber man Aussagen zu treffen beabsichtigt, in welche Richtung also eine spätere Generalisierung (oder Theoretisierung) möglich ist. Sicherlich sind bei kleineren – insbesondere studentischen – Arbeiten die Möglichkeiten der Generalisierung begrenzt. Dennoch sind über ein kluges Sampling (Przyborski/Wohlrab-Sahr 2014: 177ff) gleichwohl sinnvolle und relevante Schritte in diese Richtung möglich.

Wichtige Überlegungen zum Sampling in qualitativen Studien wurden im Rahmen der Grounded Theory Methodologie (GTM) (Przyborski/Wohlrab-Sahr 2014: 190ff, Strauss 1991) entwickelt. Wir wollen diese hier nicht im Einzelnen darstellen (Strübing: Kapitel 32 in diesem Band), sondern nur einige Grundprinzipien festhalten:

1. Zum einen orientiert sich das Sampling an Kontrasten. Im Hinblick auf das genannte Beispiel könnte das heißen, sich einerseits Kurse anzusehen, die gegen Geld angeboten und freiwillig besucht werden, und andererseits solche, die verpflichtend, aber kostenfrei sind. Solche grundlegenden Kontrastierungen sind auch in kleineren Studien sinnvoll und möglich und führen schnell dazu, dass man verschiedene Bedingungen und Konstellationen in den Blick bekommt. Im Verlauf der Forschung werden sich weitere Kontraste herauskristallisieren, anhand derer das Sample vervollständigt wird.
2. Es geht beim Sampling in letzter Instanz nicht um „Fälle“ (in unserem Fall also um Sprachkurse), sondern um die daraus zu generierenden Konzepte (Corbin/Strauss 1990). Letztlich sind es also nicht die Deutschkurse, die miteinander verglichen werden, sondern die Mechanismen, die in ihnen zur Geltung kommen bzw. die Sinnstrukturen, die sich in ihnen zeigen. Diese werden miteinander verglichen und auf ihr Zustandekommen hin untersucht.

3. Das Sampling ist in qualitativen Untersuchungen sinnvollerweise nicht bereits in einem ersten Untersuchungsschritt abzuschließen. Unter Umständen wird man auch in späteren Untersuchungsphasen noch einmal Material erheben, um die Varianten, die man in einem Sample erfasst, zu vervollständigen.
4. Anders als beim Sampling in standardisierten Untersuchungen versucht man in qualitativen Verfahren nicht, ein verkleinertes Abbild der *Verteilung* in einer Grundgesamtheit zu erreichen, sondern die *Vielfalt der in einem Untersuchungsfeld vorhandenen Konstellationen* zu erfassen und unterschiedliche Mechanismen auf ihre Bedingungen und Wirkungen hin zu befragen. Die Frage zu beantworten, wie häufig ein Zusammenhang auftritt und ob er mit größerer Wahrscheinlichkeit im Kontext A oder B auftritt, bliebe dann anderen Methoden vorbehalten. Auch hier gilt, dass insbesondere in Abschlussarbeiten eine „Sättigung“ des Samples (d.h. dass durch neue Fälle keine grundsätzlich neuen, theoretisch fassbaren Aspekte hinzukommen), oft nicht realisiert werden kann. Gleichwohl ist es auch unter diesen Umständen möglich, sich über die Grenzen des eigenen Samples Rechenschaft abzulegen und dies entsprechend zu dokumentieren. Beispielsweise kann man nichts über die Spezifik verpflichtender Kurse sagen, wenn man sie nicht mit freiwillig besuchten kontrastiert hat. Oft wird es auch für die Interpretation des vorhandenen Materials helfen, den (in diesem Fall gedanklichen) Kontrast gegenüber unter anderen Bedingungen zustande gekommenen Sprachkursen mit zu reflektieren. Es ist in jedem Fall besser, die Grenzen der Reichweite der eigenen Untersuchung auszuweisen und zu reflektieren als sie zu verdecken und Aussagen über Dinge zu treffen, die man nicht in den Blick genommen hat.

---

## 6.7 Sechster Schritt: Bestimmung der Richtung der Generalisierung oder Theoriebildung

Auch qualitative Studien untersuchen Phänomene – welcher Art auch immer – nicht allein aus Interesse an der sozialen Vielfalt oder am einzelnen, interessanten Fall. Sie wollen vielmehr Aussagen über *soziale Sachverhalte* treffen, die über die konkret untersuchten Fälle hinaus Gültigkeit beanspruchen. Das ist nur durch ein grundlagentheoretisch begründetes und systematisch durchgeführtes methodisches Vorgehen möglich. Nur durch empirische Arbeit, die über den einzelnen Fall hinaus zu theoretischer Abstraktion kommt, ist wissenschaftlicher Fortschritt möglich. Nur so platziert man sich mit der eigenen Forschung in einem wissenschaftlichen Feld und gibt anderen die Möglichkeit, daran anzuschließen oder sich kritisch damit auseinanderzusetzen. Und nur so entsteht die Möglichkeit, ein Forschungsergebnis als „falsch“ oder „unzureichend“ zurückzuweisen. Wenn ich mich darauf beschränke, nur etwas über die individuelle Besonderheit der von mir untersuchten Fälle auszusagen, verhindere ich faktisch, dass meine Forschung als Beitrag zu einer allgemeinen Problemstellung wahrgenommen wird. Damit stellt sich aber die Frage, was diese allgemeine Problemstellung ist. Woraufhin will ich meine Befunde generalisieren, in welche Richtung will ich eine Theorie über einen empirischen Zusammenhang entwickeln?

Hier sind nun unterschiedliche Antworten denkbar. Gehen wir wieder von unserem Beispiel aus: Unsere Generalisierungsabsicht könnte in Richtung einer Evaluation von Deutschkursen für Migranten zielen, und zwar nicht primär im Hinblick auf die erlernte Sprache, sondern im Hinblick auf das, was in den Kursen – gleichsam „nebenbei“ – noch vermittelt wird. Welche Bilder von eigenen und fremden Kulturen werden entworfen, welche Typisierungen (Schütz 2004a) werden vorgenommen und wem werden sie zugeschrieben, welches Orientierungswissen wird im Zusammenhang mit dem neuen kulturellen Kontext aufgebaut? Wie unterscheiden sich verschiedene Kursangebote im Hinblick auf diese „latenten Botschaften“ und mit welchen Folgen auf Seiten der Teilnehmerinnen? Eine solche – gegenstandsbezogene – Auswertung und Generalisierung wäre sicherlich hilfreich für eine Einschätzung von Instrumenten, die dem Kompetenzerwerb und der Integration dienen sollen, aber unter Umständen – aus Gründen, die den Beteiligten vielleicht gar nicht klar sind – ganz andere, nicht intendierte Folgen haben.

Eine andere Richtung der Generalisierung, die von diesem Beispiel ausgehend möglich wäre, setzt abstrakter, man könnte auch sagen ‚grundlagentheoretischer‘, an. Der Deutschkurs wäre dann nur ein Beispieldfall für das allgemeinere Problem der Intersubjektivität und des Fremdverständens (Schütz 2004a, Garfinkel 2004), das von den Beteiligten zu lösen ist: Wie verdeutlicht man etwas, das selbstverständlicher Alltag ist, gegenüber jemandem, dem dieser Alltag nicht in gleicher Weise vertraut ist? Welche Beispiele wählt man aus, welches Wissen setzt man voraus und welches Nichtwissen unterstellt man, von welcher Gemeinsamkeit und welcher Differenz geht man aus und woraus speisen sich diese Annahmen? Eine Generalisierung dieser Art würde in Richtung einer Theorie des Fremdverständens gehen, die hier an einem Migrationskontext untersucht wird, in dem bestimmte Arten der Gruppenbildung und (Stereo-)Typisierung deutlicher zutage treten als anderswo und vielfältige (insbesondere medial vermittelte) Anhaltspunkte finden.

Wenn man den Gegenstandsbereich der Migration, den wir hier beispielhaft behandelt haben, einmal verlässt und die ganze Breite von Studien aus dem Feld der qualitativen Sozialforschung in den Blick nimmt, findet man beide Formen der Verallgemeinerung und Theoretisierung: gegenstandsbezogene und grundlagentheoretische. Und beide haben zweifellos ihren Wert. Narrationsanalysen, wie sie Fritz Schütze und andere durchgeführt haben, können helfen, die „Karrieren“ in einem bestimmten sozialen Feld und deren Bedingungen zu rekonstruieren – etwa die Karrieren von Personen, die eine Zeitlang in der Psychiatrie waren (Riemann 1987). Sie können aber auch sehr viel allgemeiner ansetzen und – auf der Ebene der Grundlagentheorie – typische Prozessstrukturen des Lebensablaufs und deren narrative Repräsentation herausarbeiten (Schütze 1983). Oder um ein anderes Beispiel zu nennen: Untersuchungen über den Umgang mit Demenz können etwas zum Gegenstandsbereich der Pflege, des Umgangs mit Kranken in Familien oder der medizinischen Versorgung beitragen. Sie können aber auch einen Beitrag leisten zum Zusammenbruch der Selbstverständlichkeiten sozialer Kommunikation und den Reparaturmechanismen, die in einem solchen Fall greifen (Radvanszky 2010), vergleichbar mit dem Zusammenbruch der Kommunikation, wie ihn Garfinkel (2004) am Beispiel seiner

„Krisenexperimente“ untersucht hat. Sie tragen damit entweder etwas zum Erkenntnisgewinn in einem konkreten Gegenstandsbereich oder zu unserem Wissen über grundlegende Prozesse und Mechanismen des gesellschaftlichen Lebens bei.

Generalisierung und Theoriebildung setzen die Bildung von Typen voraus. Die Möglichkeit der Typenbildung (Przyborski/Wohlrab-Sahr 2014: 359ff) und die Richtung, in der sie erfolgt, sind wiederum sowohl vom Sampling als auch von der Forschungsfrage abhängig.

Wir haben oben unter Verweis auf das Problem des Fremdverständens bereits eine Richtung der Typenbildung angedeutet. Daran anschließend, könnte eine Typologie den Zusammenhang zwischen der Bearbeitung des Problems des Fremdverständens und damit verbundenen Mechanismen von Integration und Grenzziehung theoretisch verdichten. Beispielsweise können im Prozess der Bearbeitung des Problems des Fremdverständens solche Grenzziehungen entlang unterstellter kultureller oder religiöser Besonderheiten vorgenommen werden, im Hinblick auf die man bestimmte Sachverhalte in besonderer Weise erläutern muss. Sie können sich aber auch an der Unterscheidung von Neuem vs. bereits Vertrautem orientieren. Als eine Dimension der Typologie ließen sich hier etwa Partikularismus und Universalismus unterscheiden: Neues lässt sich im Prinzip von jedem im Lauf der Zeit in Bekanntes verwandeln, das In-Rechnung-Stellen ethnisch-religiöser Differenz dagegen setzt eher auf dauerhafte Unterschiede. Damit ist noch keine Aussage darüber verbunden, welche Form der Grenzziehung „besser“ oder der Realität „angemessener“ ist. Es geht vielmehr um die in einer Situation wirksamen Ordnungsmuster und deren Konsequenzen. Gleichwohl entsteht hier eine Aufmerksamkeit dafür, dass die etwa im Kontext „interkulturellen Lernens“ implizierten kulturellen Unterscheidungen nicht die einzigen möglichen Differenzunterstellungen im Prozess des Fremdverständens sind.

Das bedeutet aber auch, dass bereits beim Entwerfen eines Forschungsdesigns die Aufmerksamkeit darauf gerichtet sein sollte in welche Richtung die Theoriebildung, ausgehend vom konkreten Forschungsgegenstand, gehen soll. Zum einen fließen diese Überlegungen in das Sampling ein. Zum anderen werden Auswahl und Rezeption der Forschungsliteratur sich entsprechend anders gestalteten.

---

## 6.8 Umgang mit auftauchenden Problemen

Wir wollen nun zum Abschluss in lockerer Reihenfolge noch einige Probleme ansprechen, die im Rahmen qualitativer Forschung immer wieder auftauchen, und mögliche Lösungen skizzieren. Für manche Probleme bieten sich eher „institutionelle“ Lösungen an, andere lassen sich von der Forscherin oder dem Forscher selbst beheben.

### **6.8.1 Revisionen sind bisweilen nötig und erlaubt!**

Die Zuspitzung des Erkenntnisinteresses und die endgültige Formulierung der Fragestellung können sich auch nach Beginn der empirischen Arbeit noch in gewissem Maße verändern. Das ist an sich kein Problem, sondern in der Regel ein Qualitätsmerkmal qualitativer Forschung. Wichtig ist es aber, nicht einfach drauf los zu forschen. Wenn man keine Forschungsfrage hat, kann man sie später auch nicht anpassen, und sie ergibt sich auch nicht einfach „von selbst“. Die Veränderung der Forschungsfrage wiederum geht oft mit beträchtlichem Erkenntnisgewinn einher: Erstens kann darin zum Ausdruck kommen, dass man den Relevanzen der Untersuchten Rechnung trägt. Zweitens heißt es, dass man bereits in der Lage ist, diese Relevanzen zu explizieren. Und drittens bedeutet es, dass man begonnen hat, die eigenen Vorannahmen in Relation zu den empirisch gegebenen Sachverhalten zu setzen. Eine solche Veränderung der ursprünglichen Perspektive ist aber oft ein schwieriger Prozess.

Ähnliches gilt für das Sampling. Es kann z.B. passieren, dass man von bestimmten Vergleichsdimensionen ausgegangen ist, die sich aber bei genauerer Kenntnis des Feldes als nicht besonders relevant herausstellen oder für die sich schlicht keine Beispiele finden. Im ersten Fall sollte man sich an dem orientieren, was im Feld relevant ist. Im zweiten Fall muss man zunächst überlegen, ob das Problem unter Umständen auf eine falsche oder zu enge Suchstrategie zurückzuführen ist oder ob es sich unter Umständen durch „Vermittler“ lösen lässt, die den Zugang zu einem Forschungsfeld erleichtern. Wenn man all dies ausschließen kann oder muss, ist es besser, sich an verfügbare Fälle zu halten und an den ursprünglichen Ideen nicht allzu lange festzuhalten.

### **6.8.2 Die theoretische Anbindung nicht zu lange hinausschieben!**

Qualitative Forschung erfordert eine eingehende Auseinandersetzung mit Grundlagentheorien. Da diese oft recht abstrakt sind, gestaltet sich deren Aneignung manchmal mühsam. Am besten gelingt sie in direkter Konfrontation mit Forschungsanliegen gleichsam aus der Forschungspraxis heraus. Hier entsteht in der Regel auch ein Interesse an grundlagentheoretischen Fragen. Das kann für den konkreten Arbeitsprozess manchmal allerdings zu spät sein oder den Forschungsprozess unnötig in die Länge ziehen. Vorbeugen lässt sich dem etwa durch die Teilnahme an geeigneten Lehrforschungsprojekten oder anderen kooperativen Forschungskontexten, in denen die Theorie-Empirie-Verbindung kontinuierlich besprochen wird.

### **6.8.3 Feldforschung braucht Zeit!**

Problematisch gestaltet sich bisweilen auch der Zugang zum Feld bzw. zu Interviewpartnern: Man findet keine Gesprächspartner oder hat keine Kontakt Personen, die den Zugang zum Feld erleichtern. Das zugrunde liegende Problem ist fast immer mangelnde

Zeit. Einen tragfähigen Feldzugang erarbeitet man sich durch persönliches Engagement und durch Interesse an den Untersuchten, und das kostet Zeit. Sobald Engagement und Interesse sichtbar werden, öffnen sich in der Regel auch die Türen. Eine unbekannte Person lediglich einmal anzusprechen, sich ihre Telefonnummer zu notieren und später zu versuchen, per Telefon einen Interviewtermin zu vereinbaren, führt oft zu Enttäuschungen. Regelmäßige Präsenz an einem Ort, an dem sich die potentiellen Untersuchungspersonen immer wieder aufhalten (sei es ein Jugendclub, ein Sprachkurs oder eine Moschee), dokumentiert das eigene Interesse und führt eher zur Bereitschaft, sich auf ein Interview einzulassen.

#### **6.8.4 Methoden lassen sich nicht allein abstrakt erlernen, sondern brauchen Übung!**

Bei manchen unerfahrenen Forschern kursiert die Einschätzung, dass sich qualitative Methoden gleichsam im Selbststudium aus Büchern erlernen ließen, während man für statistische Verfahren eine lange Ausbildung brauche. Auch wenn Methodenbücher sicherlich helfen können (und sollten), sollte man nicht vergessen, dass alle methodischen Verfahren, auch die qualitativen, Übung verlangen. Dabei wird es in der Regel hilfreich sein, sich bei der eigenen Forschung zunächst an einem Verfahren zu orientieren und sich dieses gründlich anzueignen. Im zweiten Schritt lassen sich dann Elemente verschiedener Methoden durchaus miteinander verbinden. Das fällt aber leichter, wenn bereits eine gewisse Sicherheit im Umgang mit einem Verfahren erworben wurde.

#### **6.8.5 Kooperative Forschungszusammenhänge nutzen oder schaffen!**

Qualitative Forschung gelingt am besten, wenn sie durch Forschungswerkstätten unterstützt und begleitet wird bzw. wenn im Rahmen von Projektzusammenhängen gemeinsame Arbeit am Material stattfindet. Hier reflektiert man die Passung zwischen Phänomen, Fragestellung, konkreten Methoden und grundlagentheoretischer Einbettung, die Erfahrungen, Fortschritte und Rückschläge im Feld, diskutiert die Interpretation des empirischen Materials und erarbeitet Strategien für das Sampling. Das größte Problem für Nachwuchsforscherinnen ist es, alleine arbeiten zu müssen und keine Anbindung an einen derartigen Forschungszusammenhang zu haben. Eine gute Alternative zu Forschungswerkstätten, die von einer erfahrenen Forscherin angeleitet werden, sind Interpretationsrunden in einer Peergroup. Sie ergänzen auch Forschungswerkstätten optimal.

## Literatur

- Bohnsack, Ralf (2007): „Orientierungsmuster“. Ein Grundbegriff qualitativer Sozialforschung. In: Schmidt, Folker (Hg.): Methodische Probleme der empirischen Erziehungswissenschaft. Baltmannsweiler: Schneider. 49-61
- Bohnsack, Ralf (2008). Qualitative Bild- und Videointerpretation. Opladen/Farmington Hills: Verlag Barbara Budrich/UTB
- Bohnsack, Ralf (2011): Dokumentarische Methode. In: Bohnsack, Ralf/Marotzki, Winfried/Meuser, Michael (Hg.): Hauptbegriffe qualitativer Forschung. Opladen/Farmington Hills: Verlag Barbara Budrich/UTB. 40-44
- Corbin, Juliet/Strauss, Anselm (1990): Grounded Theory Research. In: Zfs 19 (6). 418-427
- Garfinkel, Harold (2004 [1967]): Studies in Ethnomethodology. Cambridge: Polity Press
- Goffman, Erving (1975): Stigma. Frankfurt a.M.: Suhrkamp
- Goffman, Erving (1980): Rahmenanalyse. Frankfurt a.M.: Suhrkamp
- Heinig, Kristin (2010): Doing Integration. Wie Integration gemacht wird. Eine qualitative Studie über die Praxis der Integrationskurse in Leipzig. Magisterarbeit. Leipzig: Institut für Kulturwissenschaften der Universität Leipzig
- Hopf, Christel (1978): Die Pseudo-Exploration. In: Zeitschrift für Soziologie 7: 97-115
- Jakob, Maria (2012): Einbürgerungsfeiern in Deutschland. Willkommensritual oder kulturelle Abwehr? In: sinnprovinz. kultursoziologische working papers. 3. Leipzig: Institut für Kulturwissenschaften
- Kleemann, Frank/Krähnke, Uwe/Matuschek, Ingo (2009): Interpretative Sozialforschung. Wiesbaden: VS Verlag
- Lüders, Christian (2003): Beobachten im Feld und Ethnographie. In: Flick, Uwe/Kardorff, Ernst von/Steinke, Ines (Hg.): Qualitative Forschung. Reinbek: Rowohlt. 384-401
- Mannheim, Karl (1980 [1922-1925]): Strukturen des Denkens. Frankfurt a.M.: Suhrkamp
- Meuser, Michael (2003): Rekonstruktive Sozialforschung. In: Bohnsack, Ralf/Marotzki, Winfried/Meuser, Michael (Hg.): Hauptbegriffe qualitativer Forschung. Opladen/Farmington Hills: Verlag Barbara Budrich/UTB. 140-142
- Przyborski, Aglaja/Wohlrab-Sahr, Monika (2014): Qualitative Sozialforschung. 4. erweiterte Auflage. München: Oldenbourg
- Radvanszky, Andrea (2010): Die Alzheimer Demenz als soziologische Diagnose. In: Jahrbuch für Kritische Medizin und Gesundheitswissenschaften 46: 122-142
- Riemann, Fritz (1987): Das Fremdwerden der eigenen Biographie. München: Fink
- Schütz, Alfred (1971): Das Problem der sozialen Wirklichkeit. Den Haag: Nijhoff
- Schütz, Alfred (2004a): Der sinnhafte Aufbau der sozialen Welt. Konstanz: UVK
- Schütz, Alfred (2004b): Das Problem der Relevanz. In: Schütz, Alfred (Hg.): Relevanz und Handeln 1. Zur Phänomenologie des Alltagswissens. Konstanz: UVK. 57-249
- Schütze, Fritz (1983): Biographieforschung und narratives Interview. In: Neue Praxis 3. 283-293
- Strauss, Anselm L. (1991): Grundlagen qualitativer Sozialforschung. München: Fink

- Weber, Max (1980 [1922]): Wirtschaft und Gesellschaft. Tübingen: Mohr Siebeck
- Weber, Max (1988 [1904]): Die „Objektivität“ sozialwissenschaftlicher und sozialpolitischer Erkenntnis. In: Weber, Max (Hg.): Gesammelte Aufsätze zur Wissenschaftslehre. Tübingen: UTB. 146-214
- Wernet, Andreas (2000): Einführung in die Interpretationstechnik der Objektiven Hermeneutik. Opladen: Leske + Budrich
- Wohlrab-Sahr, Monika (2006): Die Realität des Subjekts. Überlegungen zu einer Theorie biographischer Identität. In: Keupp, Heiner/Hohl, Joachim (Hg.): Subjektdiskurse im gesellschaftlichen Wandel. Bielefeld: transcript. 75-97

# Forschungsdesigns für die quantitative Sozialforschung

Petra Stein

Ein Forschungsvorhaben setzt sich aus einer Reihe von Entscheidungen zusammen. Dieser Vorgang wird als Forschungsprozess bezeichnet. Als Ergebnis der verschiedenen Entscheidungen entsteht ein spezifisches *Forschungsdesign* (auch *Untersuchungsdesign* oder *Untersuchungsanordnung* genannt). Dieses bildet die Grundlage einer sozialwissenschaftlichen Untersuchung. Das Forschungsdesign ist für die Aussagekraft der erzielten Ergebnisse von entscheidender Bedeutung, da es festlegt, wie eine theoretische Forschungsfrage konkret in eine empirische Untersuchung umgesetzt werden soll. Im Rahmen des Forschungsprozesses wird die Art der Forschung festgelegt, die Grundgesamtheit definiert (Häder/Häder, Kapitel 18 in diesem Band), das Stichprobenziehungsverfahren bestimmt (Häder/Häder, Kapitel 18 in diesem Band), die Messinstrumente festgesetzt (Burzan, Huinink, Latcheva/Davidov, Kapitel 81, 82 und 54 in diesem Band) sowie über die Art der Untersuchungsmethode (Reinecke, Salheiser, Hartmann/Lengerer, Schmitz/Yanenko, Kapitel 44, 62, 71 und 65 in diesem Band) und der Auswertung entschieden. Im Folgenden wird ein typischer Forschungsverlauf einer empirischen Untersuchung dargestellt. Den Schwerpunkt stellt dabei die Planungsphase der Untersuchung.

## 7.1 Formulierung des Forschungsproblems

Der Forschungsprozess lässt sich grob in vier Prozessphasen unterteilen. Dies sind erstens die *theoretische Phase*, zweitens die *empirische Phase*, drittens die *Auswertungsphase* und viertens die *praktische Phase*. Jedes Forschungsprojekt beginnt mit der Formulierung eines Forschungsproblems bzw. der Festlegung der Untersuchungsthematik. Diese Phase wird in der Wissenschaftstheorie auch als *Entdeckungszusammenhang* bezeichnet. Voraussetzung für eine empirische Arbeit ist ein klar definiertes Forschungsproblem, aus dem dann die konkrete Forschungsfrage formuliert wird. Bei der *Konkretisierung der Forschungsfrage* handelt es sich um einen iterativen Prozess, in dem ein zunächst vage formuliertes

Forschungsproblem durch vertiefende Auseinandersetzung mit der Materie in Form von Literaturrecherchen, Sichten und Systematisieren der Fachliteratur (Ohly/Weber, Kapitel 5, in diesem Band) zunehmend präzisiert wird.

---

## 7.2 Theoriebildung

Nach der Entscheidung über die Themenfindung beginnt die Phase der Theoriebildung (Reichertz, Kapitel 2 in diesem Band). Häufig liegen für ein bestimmtes Thema bereits ausgearbeitete Theorien in der Fachliteratur vor. Ist dies nicht der Fall, muss eine neue Theorie zur Erklärung der behandelten Forschungsthematik entwickelt werden. In der Regel bieten die verschiedenen theoretischen Ansätze in den Sozialwissenschaften die Ausgangsbasis für eine Theoriebildung. Die theoretische Einbettung dient der Begründung der Forschungsfrage. Ein großer Teil der Aufgaben des Forschers in der Phase der Theoriebildung besteht in der Literaturrecherche, der Sichtung, der Auswahl und der umfassenden Bearbeitung der zum Thema existierenden Fachliteratur, insbesondere in Form von Fachzeitschriftenaufsätzen (Ohly/Weber, Kapitel 5 in diesem Band). Aus der theoretischen Begründung werden einzelne zentrale Forschungsfragen der übergeordneten Forschungsfrage abgeleitet.

---

## 7.3 Festlegung der Art der Forschung

Aus der Forschungsfrage und ihrer theoretischen Begründung ergibt sich die Art der Forschung. Eine Hauptaufgabe der empirischen Forschung liegt in der Überprüfung von Theorien bzw. Hypothesen. Hierzu zählt allgemein die Überprüfung von vorab aufgestellten Vermutungen über gesellschaftliche Gegebenheiten. Einen wichtigen Bestandteil nimmt die Überprüfung von solchen Hypothesen ein, die einen vermuteten Kausalzusammenhang zwischen zwei Gegebenheiten postulieren (Kühnel/Dingelstedt, Eifler, Kapitel 80 und 11 in diesem Band), so genannte Ursache-Wirkungs-Zusammenhänge. Das Interesse liegt hierbei bei der ursächlichen Erklärung von sozialen Phänomenen.

Neben den hypothesentestenden Untersuchungen spielen in der empirischen Sozialforschung auch explorative Untersuchungen, deskriptive Untersuchungen, Prognosen und Evaluationsstudien eine Rolle.

So werden *explorative Studien* durchgeführt, wenn der Kenntnisstand in einem Untersuchungsbereich noch sehr gering ist. Sie zielen darauf ab, erste Einblicke in einen bestimmten Gegenstandsbereich zu erhalten.

*Deskriptive Untersuchungen* werden durchgeführt, um eine umfassende Beschreibung eines definierten Untersuchungsgebiets zu erhalten. Im Mittelpunkt stehen Fragen nach der Verteilung bestimmter Merkmale. Voraussetzung hierfür ist eine klar definierte Grundgesamtheit und Stichprobe (Häder/Häder, Kapitel 18 in diesem Band).

*Prognosen* werden durchgeführt, um möglichst treffende Aussagen über zukünftige Entwicklungen, basierend auf bisherigen Entwicklungen, zu ermöglichen. Voraussetzung hierfür sind Längsschnittdaten (Schupp, Pötter/Prein, Kapitel 73 und 86 in diesem Band).

*Evaluationsstudien* (Döhring, Kapitel 9 in diesem Band) finden hingegen als wissenschaftliche Begleitforschung von (z.B. arbeitsmarktpolitischen) Programmen statt und zielen auf die Überprüfung der Wirksamkeit von neu eingeführten Maßnahmen. Evaluationsstudien erfordern ebenfalls ein Längsschnittdesign, so dass die Wirksamkeitsüberprüfung einer Maßnahme durch eine Vorher-Nachher-Messung (Kühnel/Dingelstedt, Eifler, Kapitel 80 und 11 in diesem Band) erfolgen kann.

---

## 7.4 Konzeptspezifikation und Operationalisierung

Nach der Formulierung und Präzisierung des Forschungsproblems folgt die empirische Phase. Diese besteht aus der *Planung und Vorbereitung der Erhebung* sowie der anschließenden Feldarbeit, der Datenerhebung.

In dieser Phase wird zunächst das Erhebungsinstrument konstruiert. Der erste Schritt hierzu ist die Konzeptspezifikation und Operationalisierung. Die Aufgabe besteht hierbei darin, die in den Forschungshypothesen auftretenden Begriffe zu definieren und zu operationalisieren, d.h. einer Messung zugänglich zu machen.

Allgemein arbeiten Wissenschaften nicht mit konkreten Ereignissen an sich, sondern einer in Sprache gefassten Realität bzw. mit Aussagen über die Realität. Wenn solche Aussagen nicht nur von dem jeweiligen Wissenschaftler verstanden werden sollen, sondern deren Inhalt all jenen, an die sich die Erkenntnisse richten, zugänglich sein sollte, dann ist es unumgänglich, dass der gemeinte Sinn der verwendeten Begriffe auch von allen Beteiligten verstanden werden kann. Entsprechend enthalten auch die meisten Theorien in den Sozialwissenschaften Begriffe, die zunächst noch nicht exakt definiert sind. Ein Beispiel für einen solchen Begriff ist der Begriff „Integration“.

Eine notwendige Voraussetzung zur Beschreibung und Erklärung eines sozialen Ereignisses ist die *begriffliche Präzisierung der empirischen Begriffe* (Burzan, Huinink, Latcheva/ Davidov, Kapitel 81, 82 und 55, alle in diesem Band). Die Präzisierung dieser Begriffe geschieht mit Hilfe von Definitionen, mit denen der Vorstellungsinhalt der Begriffe festgelegt wird. In dieser Phase des Forschungsprozesses wird somit entschieden, dass ein bestimmtes sprachliches Zeichen nur noch in einer bestimmten Weise verwendet werden soll. Die Präzisierung von empirischen Begriffen geschieht mit Hilfe einer *Nominaldefinition*, also einer Festsetzung darüber, dass ein neuer Begriff (Definiendum) gleichbedeutend mit einem anderen Begriff sein soll, wobei die Bedeutung des letzteren Begriffes bekannt ist und der zu definierende Begriff den Vorstellungsinhalt des letzteren Begriffs annehmen soll (Definiens). Da eine Nominaldefinition nur eine Festsetzung über die Verwendung eines sprachlichen Ausdrucks ist, ergibt sich daraus, dass Nominaldefinitionen weder wahr noch falsch sein können. Sie können sich nur als brauchbar oder unbrauchbar erweisen, denn sie gelten allein aufgrund des Willens desjenigen, der sie festsetzt.

Handelt es sich um komplexe, mehrdimensionale Begriffe, müssen die einzelnen Dimensionen des Begriffs herausgearbeitet werden. Nach der exakten Definition aller verwendeten Begriffe und der Angabe der Dimensionen, auf die sich die Begriffe beziehen, kann die Konzeptspezifikation als abgeschlossen angesehen werden.

Es bleibt die Frage, wie die theoretischen Begriffe erfasst werden können, konkret, wie sie gemessen werden können. Dabei sind Begriffe, die sich auf einen unmittelbar wahrnehmbaren Sachverhalt beziehen (*manifeste Variablen*), von Begriffen zu unterscheiden, die einen Sachverhalt beschreiben, welcher nicht direkt beobachtbar ist (*latente Variablen*). Bei den manifesten Variablen besteht die Möglichkeit, dass die Ausprägungen der Variablen direkt wahrnehmbar sind. Im Gegensatz zu manifesten Variablen heißen Variablen, die nicht unmittelbar (direkt) wahrnehmbar sind, latente Variablen. Hierbei handelt es sich um theoretische Konstrukte, wie z.B. Einstellungen und Haltungen. Aus dieser Perspektive sind theoretische Begriffe latente Variablen, da sie nicht direkt beobachtbar sind.

Den theoretischen Begriffen müssen somit beobachtbare Sachverhalte zugeordnet werden, so dass diese auch erfassbar werden. Dies geschieht über *Indikatoren* (Burzan, Kapitel 81 in diesem Band), also Merkmale, die als Anzeichen für etwas Latentes dienen. Diese Umsetzung von theoretischen Begriffen in so genannte Indikatoren wird als *Operationalisierung* bezeichnet. Je nach Informationsgehalt der Variablen bieten sich unterschiedliche *Messniveaus* an (nominal-, ordinal-, intervall- und ratioskaliert). Die Zusammenstellung der Messoperationen für alle Variablen, die in die Erhebung einbezogen werden sollen, ergibt das Erhebungsinstrument. In den Sozialwissenschaften sind diese Erhebungsinstrumente zumeist Fragebögen (Klöckner/Friedrichs, Reinecke, Kapitel 49 und 44 in diesem Band), aber durchaus auch Dokumente (Salheiser, Kapitel 62 in diesem Band) oder Logfiles (Schmitz/Yanenko, Kapitel 60 in diesem Band).

---

## 7.5 Festlegung des Forschungsdesigns

In Zusammenhang mit der Konstruktion des Erhebungsinstruments steht die Festlegung der jeweiligen Untersuchungsform und somit die Bestimmung des Forschungsdesigns. Bei der Festlegung eines Forschungsdesigns sind verschiedene Entscheidungen zu treffen, die teilweise nacheinander, teilweise aber auch zeitlich parallel in einem wechselseitigen Zusammenhang getroffen werden müssen.

Eine Hauptaufgabe der empirischen Forschung liegt in der Überprüfung von Theorien bzw. Hypothesen. Wie bereits erläutert findet die Überprüfung auf der Ebene von empirischen Indikatoren statt. Entsprechend sind die folgenden Grundfragen zu beantworten:

1. In welchem Zeitraum sollen die manifesten Indikatoren zur Überprüfung der Hypothesen erhoben werden?
2. Mit welchem Messinstrument soll die Überprüfung durchgeführt werden?
3. Wie häufig soll die Erhebung erfolgen?
4. An welchen Objekten soll die Messung stattfinden?

Im Rahmen des Forschungsdesigns wird hierzu die Art der Untersuchungsmethode (z.B. echtes experimentelles Design, Querschnitts- oder Längsschnittsdesign) festgelegt.

Der Aufbau des Untersuchungsdesigns ist für die Sicherheit von Aussagen über die vermuteten Ursache-Wirkungszusammenhänge (Kühnel/Dingelstedt, Kapitel 80 in diesem Band) entscheidend. Das Ziel besteht dabei in dem Ausschluss von alternativen Erklärungsmöglichkeiten neben dem angenommenen gerichteten Zusammenhang. Allgemein können Untersuchungsdesigns, die die Beziehung zwischen den vermuteten Ursachen (*unabhängigen Variablen*) und Wirkungen (*abhängigen Variablen*) untersuchen, als Experimente bezeichnet werden. Hierbei wird zwischen den echten experimentellen Designs und den Ex-post-facto-Designs unterschieden.

## 7.5.1 Experimentelle Designs

### 7.5.1.1 Vorexperimentelle Versuchsanordnung

Die simpelste Untersuchungsform zur Beurteilung einer Hypothese ist das Design der einmaligen Messung. Hierbei wird eine Gruppe von Personen (Versuchspersonen) der Gegebenheit ausgesetzt, die in der Hypothese die unabhängige Variable bildet (Eifler, Kapitel 11 in diesem Band). Danach wird bei der gesamten Gruppe die abhängige Variable gemessen. Der gemessene Durchschnittswert müsste dann mit einem unbekannten Vorherwert verglichen werden. Da dieser nicht vorliegt, muss der gemessene Wert mit einem Wert aus einer anderen Untersuchung verglichen werden. Weicht der gemessene Durchschnittswert signifikant von dem Vergleichswert ab, könnte daraus gefolgert werden, dass die Hypothese zutrifft.

### 7.5.1.2 Einfaches experimentelles Design

Im Unterschied zu dem Design einer einmaligen Messung werden bei einem einfachen experimentellen Design (Eifler, Kapitel 11 in diesem Band) zwei Gruppen gebildet, wobei Personen nach dem Zufallsprinzip einer *Experimental- und einer Kontrollgruppe* zugeordnet werden. Diese zufällige Zuordnung wird als *Randomisieren* bezeichnet. Vor der experimentellen Manipulation, also der Einführung des *Stimulus*, wird in beiden Gruppen eine Vorhermessung der abhängigen Variable durchgeführt. Anschließend wird nur in der Experimentalgruppe der Stimulus eingeführt. Daraufhin werden erneut Messungen der abhängigen Variablen, sowohl bei der Experimental- als auch bei der Kontrollgruppe, durchgeführt (Nachhermessung). Im Unterschied zu der einmaligen Messung ist für die Experimentalgruppe somit ein direkter Vergleich der Werte von *Vorher- und Nachher-Messung* möglich, die zusätzlich mit den Werten der Kontrollgruppe verglichen werden können.

### 7.5.1.3 Echtes experimentelles Design

Ein Problem bei der Prüfung von Hypothesen über Ursache-Wirkungszusammenhänge liegt darin, dass ein und dasselbe Objekt nicht gleichzeitig der vermuteten Ursache zum einen ausgesetzt und zum anderen nicht ausgesetzt werden kann. Aus diesem Grund

werden die Objekte durch Zufallszuweisungen in eine Experimental- und eine Kontrollgruppe eingeteilt (Randomisierung). Die signifikante Differenz der Mittelwerte der relevanten Variablen bei den beiden Gruppen ergibt den kausalen Effekt der unabhängigen Variablen.

Die Mittelwertdifferenz muss jedoch nicht zwingend durch die unabhängige Variable ausgelöst worden sein, sondern es könnten auch andere Faktoren dafür verantwortlich sein. Damit ausgeschlossen werden kann, dass diese Faktoren nicht für den Effekt der unabhängigen Variablen verantwortlich sind, müssen diese *Störfaktoren* kontrolliert werden. Es wird unterschieden zwischen Störfaktoren, die die *interne Validität* (also die Gültigkeit, dass tatsächlich der Stimulus für die Veränderung der abhängigen Variablen verantwortlich ist) beeinflussen und Störfaktoren, die die *externe Validität* (also die Gültigkeit, die sich auf die Fähigkeit der Verallgemeinerung der durch das Experiment erzielten Ergebnisse bezieht) beeinflussen (Krebs/Menold, Kapitel 30 in diesem Band).

Zu den Ersteren zählt *zwischenzeitliches Geschehen*, womit Ereignisse gemeint sind, die zwischen Vorher- und Nachhermessung zusätzlich zur betrachteten unabhängigen Variablen die abhängige Variable beeinflussen. Ein weiterer Störfaktor ist die *Reifung der Versuchspersonen*, bei welcher interpersonelle Prozesse die Veränderung der abhängigen Variablen beeinflussen. Des Weiteren zählen zu den Störfaktoren, die die interne Validität in Frage stellen, *Messeffekte* (Veränderungen der abhängigen Variablen werden als Auswirkung des ersten Messvorgangs auf den Zweiten erzeugt), *Hilfsmittel* (die abhängige Variable wird durch veränderte Messinstrumente und -bedingungen beeinflusst) sowie *verzerrte Auswahlen* (Experimental- und Kontrollgruppe unterscheiden sich zuzüglich zu der eigentlichen unabhängigen Variablen noch hinsichtlich anderer Merkmale, welche die abhängige Variable beeinflussen).

Als Störfaktoren, die die externe Validität in Frage stellen, zählen *reaktive Effekte des Messens* (Durchführung der ersten Messung hat zu einer Sensibilisierung der Versuchsperson in Hinblick auf den Stimulus geführt) und *reaktive Effekte* durch die experimentelle Situation (experimentelle Situation weicht von der Alltagssituation der Versuchspersonen exorbitant ab).

Zur *Kontrolle dieser Störfaktoren* liegen verschiedene Techniken vor. Im Einzelnen handelt es sich dabei um Elimination, Konstanthaltung, Matching und Randomisierung. Unter *Elimination* ist das Ausschalten von möglichen Störfaktoren bei der Durchführung eines Experimentes in einer Laborsituation gemeint. *Konstanthaltung* bezeichnet die größtmögliche Angleichung der Versuchsbedingungen bei Experimental- und Kontrollgruppe bis auf die relevante unabhängige Variable. Eine weitere Technik zur Kontrolle der Versuchsbedingungen ist die stichhaltige Bildung von Experimental- und Kontrollgruppen. Dies kann erstens durch *Matchen* erfolgen. Darunter wird die Bildung von zwei *Parallelgruppen* verstanden, die aus Personen bestehen, die in bestimmten Merkmalen gleiche Ausprägungen aufweisen. Eine zweite Möglichkeit ist die bereits angesprochene *Randomisierung* bei der Gruppenzuordnung.

Experimentelle Forschungsdesigns prüfen eine Hypothese, indem die unabhängige Variable durch den Versuchsleiter manipuliert wird und der Einfluss von Störfaktoren durch Konstanthaltung der Versuchsbedingungen, Elimination, Parallelisierung oder Randomisierung kontrolliert wird. Nur in diesem Fall handelt es sich um ein echtes Experiment.

Dabei werden zwei *Arten von Experimenten* unterschieden. Dies sind Labor- und Feldexperimente. Der Vorteil von *Laborexperimenten* liegt darin, dass die Versuchsbedingungen in hohem Maße kontrolliert werden können, was zu einer hohen internen Validität führt. *Feldexperimente* haben häufig den Vorteil, dass sie eine hohe externe Validität aufweisen, was auf die natürliche Umgebung, in der sie durchgeführt werden, zurückgeführt werden kann.

#### 7.5.1.4 Quasi-experimentelles Design

In der Forschungspraxis der empirischen Sozialforschung sind echte Experimente nur selten möglich. Beispielsweise kann man nicht bei der Überprüfung der Auswirkungen verschiedener Schulformen auf den Lernerfolg die Kinder zufällig auf die Schulformen verteilen. In solchen Fällen ist der Forscher auf andere Designs angewiesen. Ein solches Design ist das quasi-experimentelle Design. Im Unterschied zu den echten Experimenten erfolgt die Zuordnung der Versuchspersonen zu den Experimental- und Kontrollgruppen nicht durch Randomisierung oder Matchen, sondern aufgrund bereits vorhandener Merkmale der Untersuchungseinheiten bzw. erfolgt durch die Untersuchungseinheiten selbst. Der Stimulus wird jedoch wie bei einem echten Experiment gesetzt. Quasi-experimentelle Untersuchungen sind aufgrund der Selbstselektion der Versuchspersonen besonders stark von Ausfällen und verzerrten Auswahlen bedroht. Quasi-experimentelle Untersuchungen ermöglichen keine Rückschlüsse auf kausale Zusammenhänge, da in diesem Rahmen nicht festgestellt werden kann, ob tatsächlich ein Effekt von der unabhängigen auf die abhängige Variable ausgeht oder ob beide konfundiert sind.

#### 7.5.2 Ex-post-facto-Design

Häufig sind bei Fragestellungen in der empirischen Sozialforschung die Anforderungen, welche für die Umsetzung in Form eines Experiments oder eines Quasi-Experiments nötig sind nicht erfüllt. In diesem Fall wird ein Ex-post-facto-Design angewendet. Diese Anordnung zeichnet sich dadurch aus, dass der Forscher keinen Stimulus manipuliert und auch keine Randomisierung der unabhängigen Variablen auf die Untersuchungseinheiten vornimmt. Dagegen findet eine *Ex-post-Randomisierung* statt. Die abhängigen und unabhängigen Variablen werden zuerst gemessen und erst im darauffolgenden Schritt werden die Untersuchungseinheiten aufgrund der gemessenen Werte in Versuchs- und Kontrollgruppe eingeteilt. Aus diesem Grund sind aus den daraus resultierenden Ergebnissen nur korrelative Aussagen möglich. Ex-post-facto-Designs sind in den Sozialwissenschaften die verbreitetste Untersuchungsanordnung. Der Vorteil liegt darin, dass mit verhältnismäßig

geringem finanziellem und personellem Aufwand eine *große Menge an Daten* (meistens durch Befragungen) erhoben werden kann und durch die Verwendung geeigneter Auswahlverfahren eine *Generalisierung* möglich ist.

Ex-post-facto-Anordnungen sind gleichsam mit einer Reihe von Problemen behaftet:

1. Das erste Problem besteht in der *Varianz der unabhängigen Variablen*. Die nachträgliche Klassifizierung der Daten in Experimental- und Kontrollgruppe erweist sich als schwierig, wenn mindestens eine der interessierenden Merkmalsausprägungen nur selten anzutreffen ist.
2. Das zweite Problem betrifft die *kausale Reihenfolge der Variablen*. Diese ist dann nicht herstellbar, wenn die Daten in einem Querschnittsdesign, also gleichzeitig erhoben worden sind. Eine Lösung hierzu bildet die Erhebung von Daten im Längsschnittsdesign (siehe unten).
3. Drittens ist die *Kontrolle von Störfaktoren* komplizierter, da der Effekt der unabhängigen Variablen nicht nach dem Zufallsprinzip auf die Untersuchungseinheiten verteilt werden kann.

Ex-post-facto-Anordnungen lassen sich anhand des Erhebungsdesigns unterteilen, wobei zwischen dem Querschnitts- und dem Längsschnittsdesign (Trendstudie, Kohortenstudie oder Panelstudie) Unterschiede werden kann. Der Forscher hat in dieser Phase des Forschungsprozesses somit festzulegen, ob das Forschungsvorhaben in Form einer Querschnittserhebung oder einer Längsschnitterhebung durchgeführt werden soll.

### 7.5.2.1 Querschnittsdesign

Eine Querschnittuntersuchung (Mochmann, Kapitel 14 in diesem Band) impliziert eine einmalige und gleichzeitige Messung aller für das Forschungsvorhaben relevanter Merkmale. Dies ist immer dann ausreichend, wenn lediglich eine aktuelle Bestandsaufnahme innerhalb einer Population vorgenommen werden soll. Aussagen über kausale Beziehungen zwischen den Merkmalen sind dabei nicht möglich.

Das Querschnittsdesign erfasst ein Phänomen lediglich zu einem Zeitpunkt und ermöglicht somit auch nur eine Momentaufnahme.

### 7.5.2.2 Längsschnittsdesign

Ein häufiges Anliegen innerhalb der empirischen Sozialforschung ist jedoch die Erforschung von Veränderungen und des sozialen Wandels. In diesen Fällen ist eine Längsschnittuntersuchung notwendig, bei der mehrere, zeitlich gestaffelte Untersuchungen zu einem Themenkomplex erfolgen. Hierbei lassen sich verschiedene Möglichkeiten des Designs einer Längsschnittuntersuchung unterscheiden. Im Einzelnen handelt es sich dabei um Trend-, Kohorten- und Paneldesign.

### 7.5.2.2.1 Trenddesign

Bei Trendanalysen (auch replikatives Survey oder unechtes Panel genannt) werden zum gleichen Thema unter Verwendung des gleichen Fragebogens mehrere Querschnittsuntersuchungen in einem gewissen zeitlichen Abstand hintereinander an nicht unbedingt denselben Objekten durchgeführt (Mochmann, Kapitel 14 in diesem Band). Grundgesamtheit und Auswahlmethode bleiben dabei unverändert. Trendanalysen haben einen höheren Informationsgehalt gegenüber Querschnittsuntersuchungen, da zusätzlich auch Aussagen über Wandel und Stabilität auf der Aggregatebene getätigten werden können. Wie der Name dieses Designs verrät, eignet es sich insbesondere zur Erforschung von Trends. Beispielsweise erfolgt im Rahmen der ALLBUS-Studien u.a. eine Erhebung verschiedener Einstellungen und Werte, welche in regelmäßigen Abständen wiederholt wird. Intraindividuelle Veränderungen können mit Trenddaten hingegen nicht analysiert werden, da eine direkte Verknüpfung der Personen zwischen den Datensätzen aufgrund der unterschiedlichen Stichprobenzusammensetzungen nicht möglich ist, was zu einem beträchtlichen Informationsverlust führen kann. Eine weitere Problematik besteht in der Trennung von tatsächlichen Veränderungen und Stichprobenfehlern (Faulbaum, Kapitel 31 in diesem Band), beispielsweise durch unterschiedliche Auswahlwahrscheinlichkeiten von Untersuchungseinheiten zwischen zwei Erhebungszeitpunkten.

### 7.5.2.2.2 Kohortenanalyse

Eine weitere Form der Längsschnittuntersuchung ist die Kohortenstudie. Kohortenstudien werden innerhalb der empirischen Sozialforschung vorwiegend zur Analyse von Lebensverläufen angewendet. Unter Kohorten werden Personengruppen gefasst, die zu annähernd gleichen Zeitpunkten in ihrem Leben dieselben Ereignisse durchliefen, wie z.B. Geburt, Heirat oder Scheidung. Zu der gewählten Kohorte, z.B. Personen desselben Jahrgangs, werden dann Informationen zu mehreren Messzeitpunkten erhoben. Kohortenstudien sind somit gleichzeitig auch Panelstudien und weisen entsprechend die gesamten Vorteile einer Panelstudie auf (siehe unten). Jedoch existieren auch dieselben Probleme. Bei Kohortenanalysen lassen sich zwei Vorgehensweisen unterscheiden: Beim Intra-Kohorten-Vergleich sind die Objekte zu verschiedenen Messzeitpunkten identisch (z.B. ein Geburtsjahrgang), während beim Inter-Kohorten-Vergleich Angehörige verschiedener Kohorten untersucht werden, die den gleichen Zeitabstand zu dem Ereignis aufweisen, das die jeweilige Kohorte definiert, wie z.B. Geschiedene unterschiedlichen Alters ein Jahr nach ihrer Scheidung. Ein häufig im Zusammenhang mit Kohortenanalysen diskutiertes Problem ist die Unterscheidung zwischen Alters-, Kohorten- und Periodeneffekten. Alterseffekte beziehen sich dabei auf individuelle Veränderungen, welche auf den Prozess des Alterns zurückzuführen sind, z.B. aufgrund der Akkumulation von Erfahrungen im Zeitverlauf. Kohorteneffekte beschreiben dagegen Eigenschaften, welche aufgrund der Zugehörigkeit zu einer bestimmten Kohorte vorliegen, z.B. durch ähnliche Erfahrungen in der Sozialisation. Periodeneffekte beziehen sich schließlich auf individuelle Veränderungen, welche durch historische Ereignisse oder gesellschaftliche Umwälzungen, wie z.B.

wirtschaftliche oder politische Krisen, hervorgerufen werden. Liegt nun der Forschungsfokus auf der Analyse von Alterseffekten, kann dieser mit Kohorten- oder Periodeneffekten konfundiert sein, welches eine eindeutige Effektzuweisung erschwert. Eine Zerlegung dieser Effekte kann dabei aus designbasierten (z.B. aufgrund der Beobachtung von nur einer einzelnen Kohorte) sowie analysetechnischen (aufgrund perfekter Kollinearität bei der gleichzeitigen Aufnahme der drei Aspekte in statistischen Modellen) Gründen problematisch sein. Auch der Vorschlag der direkten Integration der hinter den jeweiligen Effekten vermuteten Einflussgrößen – anstelle der Verwendung der Kohorten- und Periodeninformationen als Proxyvariablen – ist aufgrund der eingeschränkten Verfügbarkeit dieser Größen zumeist nur selten praktikabel.

### 7.5.2.2.3 Paneldesign

Eine große und wachsende Bedeutung nehmen in der empirischen Sozialforschung Längsschnittuntersuchungen ein, denen ein Paneldesign unterliegt (Schupp, Kapitel 73, in diesem Band).

Panelstudien haben gemeinsam, dass sie *zu mehreren Zeitpunkten an denselben Objekten dieselben Variablen erheben*. Diese drei Aspekte, die sich auf die Zeit-Dimension, die erhobenen Merkmale und auf die Untersuchungseinheiten beziehen, charakterisieren Panelstudien. Die Messzeitpunkte einer Panelstudie werden üblicherweise als *Wellen* bezeichnet. Der Vorteil von Paneldaten gegenüber anderen Daten besteht erstens darin, dass die Möglichkeit besteht, Entwicklungen zu analysieren, da Daten zu mehreren Messzeitpunkten vorliegen. Die Unterscheidbarkeit von Variablen wird somit komplexer, da nun Unterscheidungen in Hinblick auf den Inhalt der Variablen und in Hinblick auf den Messzeitpunkt der Variablen möglich werden. Veränderungen, die im Zeitverlauf festgestellt werden, können dabei eindeutig bestimmten Untersuchungseinheiten zugerechnet werden, so dass es möglich wird, auch tatsächlich *kausale Zusammenhänge* zu analysieren. Zweitens besteht die Möglichkeit, *zeitversetzte Informationen zu erhalten*. Drittens besteht die Möglichkeit, *intraindividuelle Prozesse* zu analysieren (*Analyse individueller Dynamiken*) und viertens die Möglichkeit, *Heterogenität zu kontrollieren*.

Längsschnittuntersuchungen im Paneldesign unterliegen jedoch auch einer Reihe von Problemen. Es ist einleuchtend, dass eine Erhebung, welche zu mehreren Zeitpunkten durchgeführt wird, aufwendiger, kostenintensiver und fehleranfälliger ist. Vor allem macht sich diese Problematik durch die spezifische Eigenart von Panelstudien bemerkbar, an denselben Personen (Objekten) zu mehreren Zeitpunkten Messungen durchzuführen. Ein Problem bezieht sich auf die Verwendung derselben Variablen, wobei sich im Laufe der Zeit Bedeutungsverschiebungen bezüglich der verwendeten Begriffe ergeben können. Ein weiteres sehr großes Problem ist die *Panelmortalität*. Damit wird die Ausfallrate bezeichnet, die sich dadurch ergibt, dass nicht alle Teilnehmer, die einmal ausgewählt wurden, auch tatsächlich bis zum Ende der Untersuchung teilnehmen. Ursachen hierfür können neben dem tatsächlichen Tod des Teilnehmers auch schwere Krankheiten, eine fehlende Bereitschaft, weiterhin an der Untersuchung teilzunehmen oder unbekannte Umzüge sein.

Die Panelmortalität wird umso größer, je länger die Untersuchung angelegt ist, d.h. sie wächst mit zunehmender Anzahl an Befragungswellen. Das Dramatische daran ist, dass solche Ausfälle in Panelstudien eine größere Tendenz haben systematisch zu sein, als dies in Querschnittsuntersuchungen der Fall ist. Der fehlenden Bereitschaft, an der Untersuchung teilzunehmen kann jedoch durch geeignete Panelpflege entgegengewirkt werden.

Das zweite Problem von Panelstudien sind Paneleffekte. Darunter werden Veränderungen der Teilnehmer durch die laufende Teilnahme an der Untersuchung verstanden. Diese entstehen insbesondere dann, wenn Personen durch die Teilnahme an der Untersuchung sich verstärkt mit der in der Studie erhobenen Thematik auseinandersetzen und in Folge dessen erst eine Einstellung zu dieser entwickeln, verändern oder verstetigen. Im Falle einer Befragung (welche immer die Gefahr der Reaktivität in sich birgt) könnten Befragte auf die Befragungssituation späterer Zeitpunkte somit anders reagieren, als dies im Falle einmaliger Befragung geschehen wäre. Daraus könnten verzerrte, invalide Messungen resultieren. Diese Besonderheit führt zu weiteren Maßnahmen, die Phase der Datenerhebung zu optimieren und an Datensätze mit fehlenden Werten bei der Analyse adäquat heranzugehen. Auch treten Probleme bei der Datenanalyse auf. Diese betreffen vor allem den Sachverhalt, dass die in zwei verschiedenen Wellen erhobenen Variablenwerte ein und derselben Person zum inhaltlich gleichen Merkmal nicht mehr als zwei unabhängige Realisierungen eines Zufallsexperimentes verstanden werden können. Die Annahme der stochastischen Unabhängigkeit ist jedoch zentral im Kontext verschiedener statistischer Verfahren. Neben den oben dargestellten Gründen können Paneeleffekte auch durch die persönliche Beziehung zwischen Interviewer und Interviewten entstehen, die sich durch einen wiederholten Kontakt zwischen den beteiligten Personen herausbilden kann.

Zur Kontrolle von Panelmortalität und Paneeleffekten sind eine Reihe von Design-Varianten entwickelt worden. Im Einzelnen handelt es sich dabei um das *alternierende Panel*, das *rotierende Panel* und das *geteilte Panel*. Während beim alternierenden Panel nach der Unterteilung der Gesamtstichprobe in Gruppen diese Substichproben abwechselnd befragt werden, wird beim rotierenden Panel zu jedem Erhebungszeitpunkt eine Gruppe der Ausgangsstichprobe mit einer nacherhobenen Gruppe ersetzt. Beim Design des geteilten Panels erfolgt eine Panelbefragung mit mehreren Messzeitpunkten an einer Gruppe, während als zweite Komponente eine replikative Querschnittsbefragung oder ein rotierendes Panel aufgenommen wird.

Streng genommen sind die oben aufgeführten Kriterien zu hart formuliert, denn kleine Abweichungen, z.B. bezüglich der Items im Fragebogen von Zeitpunkt zu Zeitpunkt und eine geringe Fluktuation der Teilnehmer lässt es immer noch zu, den Begriff Panelstudie zu verwenden. In der Umfragepraxis hat es sich ferner eingebürgert, von einem Panel schon dann zu sprechen, wenn dieselben Personen mehrmals befragt werden, auch wenn das Messinstrument von Welle zu Welle modifiziert wird oder wenn es völlig unterschiedlich ist. Dies ist nicht zuletzt auf die immer stärkere Verbreitung von so genannten (*Online*)-*Access-Panels* (Wagner/Hering, Kapitel 48 in diesem Band), welche zumeist in der Marktforschung Anwendung finden (Ziegler, Kapitel 10 in diesem Band), jedoch auch in der empirischen Sozialforschung eine zunehmende Beliebtheit erfahren, zurückzuführen. Der

Betreiber eines Access-Panels rekrutiert dabei einen möglichst großen und heterogenen Pool von Teilnehmern, welche über Bonus- und Incentive-Systeme zum zukünftigen teilnehmen an diversen Befragungen motiviert werden. Verschiedene Auftraggeber können nun an den Betreiber herantreten, um Befragungen mit einer (meist nach verschiedenen Merkmalen gezielt ausgewählten) Teilmenge der Probanden des Access-Panels durchzuführen. Diese Praxis muss allerdings von der klassischen Vorgehensweise bei der Realisierung eines Paneldesigns deutlich abgegrenzt werden.

---

## 7.6 Auswahl der Untersuchungseinheiten

Nach der Bestimmung der Untersuchungsform geht der Forschungsprozess in die Entscheidung über die Auswahl der Untersuchungseinheiten über. Hierbei ist zunächst zu klären, über welchen Objektbereich etwas in Erfahrung gebracht werden soll. Sollen Aussagen getroffen werden, die über die untersuchte Objektmenge hinausgehen, ist die Gültigkeit der Aussagen abhängig von der Definition des Objektbereichs und der Auswahl der Untersuchungseinheiten. Voraussetzung für die Durchführung einer empirischen Untersuchung ist die Definition der Grundgesamtheit, also der Menge, über die die Aussagen getroffen werden sollen. Aus der Grundgesamtheit (Population) werden Elemente ausgewählt, die dann die Untersuchungseinheiten bilden. Eine Untersuchung anhand der Menge der Elemente, die ausgewählt wurde, ergibt eine Teilerhebung (Häder/Häder; Kapitel 18 in diesem Band). Eine *Vollerhebung* bedeutet eine gesamte Erhebung aller Mitglieder der Population und ist aus Kosten- und Zeitgründen in der Regel nicht realisierbar. Sie eignet sich nur bei kleinen Grundgesamtheiten oder wenn das zu betrachtende Merkmal in der Grundgesamtheit sehr heterogen verteilt ist. Eine Teilerhebung wird *Stichprobe* genannt, wenn sie nach festgelegten Regeln aus der Population, über die sie Aussagen treffen soll, gezogen wird.

*Zufallsstichproben* werden Stichproben genannt, bei der jedes Element der Grundgesamtheit die gleiche oder eine anzugebende Wahrscheinlichkeit besitzt, in die Stichprobe zu gelangen. Die Grundgesamtheit muss dabei bekannt und exakt definiert sein sowie physisch oder symbolisch präsent und manipulierbar sein. Ferner muss die Auswahlwahrscheinlichkeit der einzelnen Elemente der Grundgesamtheit größer als Null sein. Ergebnisse von Stichprobenuntersuchungen können unter Verwendung inferenzstatistischer Verfahren auf die Grundgesamtheit verallgemeinert werden, wenn die Elemente der Stichprobe durch eine Zufallsauswahl aus der Grundgesamtheit gezogen worden sind (Induktionsschluss, Repräsentationsschluss). Fehler, die durch den Induktionsschluss entstehen, sind nur bei Zufallsstichproben berechenbar.

Im Allgemeinen wird zwischen Zufallsauswahlen und Auswahlen, die nicht zufällig sind, differenziert. Welches Auswahlverfahren geeignet ist, hängt von der jeweiligen Forschungsfrage ab. Generell erlauben nur Zufallsauswahlen eine Generalisierung der Stichprobenergebnisse auf die Grundgesamtheit. Jedoch gibt es auch eine Reihe von Gründen, unter denen eine Zufallsauswahl entweder nicht realisierbar ist oder es nicht notwendig ist, eine Zufallsauswahl durchzuführen.

## 7.6.1 Zufallsauswahlen

### 7.6.1.1 Einfache Zufallsauswahl

Von einer einfachen Zufallsauswahl (Häder/Häder, Kapitel 18 in diesem Band) wird dann gesprochen, wenn jede mögliche Stichprobe mit einer bestimmten Anzahl von Elementen dieselbe Chance der Realisierung hat. Dies führt dazu, dass jedes Element die gleiche Wahrscheinlichkeit hat in eine Stichprobe zu gelangen. Die aus der einfachen Zufallsauswahl resultierende Stichprobe wird als einfache Zufallsstichprobe bezeichnet.

### 7.6.1.2 Komplexe Zufallsauswahl

Bei komplexeren Zufallsauswahlen hat jedes Element der Grundgesamtheit nicht mehr die gleiche Chance, in die Grundgesamtheit zu gelangen sondern nur noch eine berechenbare Chance. Hierbei wird unterschieden zwischen geschichteten Zufallsauswahlen, Klumpenauswahlen und mehrstufigen Auswahlen (Häder/Häder, Kapitel 18 in diesem Band).

#### 7.6.1.2.1 Geschichtete Auswahlen

Bei geschichteten Zufallsauswahlen werden die Elemente der Grundgesamtheit in Gruppen bzw. Schichten unterteilt, wobei jedes Element der Grundgesamtheit zu genau einer Schicht gehört. Schichten sollten dabei in Hinblick auf das verwendete Schichtungsmerkmal intern möglichst homogen und zwischen den Schichten möglichst heterogen sein. Aus jeder Schicht werden dann mittels einfacher Zufallsauswahl die Elemente gezogen. Wenn die Umfänge der einzelnen Stichproben den jeweiligen Anteilen der Schichten in der Grundgesamtheit entsprechen, wird die Auswahl als proportional geschichtete Zufallsauswahl bezeichnet. Falls die Fallzahlen der Stichproben nicht den jeweiligen Anteilen der Schichten in der Grundgesamtheit entsprechen, wird die Auswahl hingegen als disproportional geschichtete Zufallsauswahl bezeichnet. Geschichtete Zufallsauswahlen eignen sich immer dann besonders, wenn Gruppen verglichen werden sollen, die unterschiedlich viele Elemente enthalten. Der Vorteil gegenüber einfachen Zufallsstichproben liegt darin, dass die Kosten der Erhebung reduziert werden können, da sich der Gesamtstichprobenumfang minimieren lässt. Ferner lassen sich unabhängige Schätzungen für jede Schicht durchführen, was bei Forschungsfragen, bei denen die Schichten der Grundgesamtheit im Zentrum des Interesses stehen, relevant ist.

Bei der Datenauswertung einer disproportional geschichteten Zufallsauswahl ist zu beachten, dass die Mitglieder der Grundgesamtheit nicht die gleichen Chancen hatten ausgewählt zu werden. Aus diesem Grund können die Daten entweder nur für die einzelnen Schichten getrennt analysiert werden oder die Mitglieder der Schichten müssen unterschiedlich gewichtet werden, um Aussagen über die Grundgesamtheit treffen zu können. Die Bestimmung geeigneter Gewichtungsfaktoren ist jedoch häufig problematisch und kann zu noch stärkeren Verzerrungen der Ergebnisse führen als dies ohne Gewichtung der Fall wäre.

### 7.6.1.2.2 Klumpenauswahl

Als Klumpenauswahl wird eine einfache Zufallsauswahl verstanden, die zunächst nicht auf die einzelnen Elemente der Grundgesamtheit zielt, sondern auf zusammengefasste Elemente, die als Klumpen oder Cluster bezeichnet werden. Dabei werden die Daten aller Mitglieder eines Klumpen erhoben. Im Gegensatz zu den Schichten sollen die Mitglieder innerhalb eines Klumpen dabei möglichst heterogen und die Klumpen möglichst homogen sein. Die Klumpenauswahl wird vor allem dann angewendet, wenn die Grundgesamtheit nicht bekannt ist, also keine Liste der Elemente der Grundgesamtheit existiert bzw. nicht zu Verfügung steht, jedoch eine Liste der Klumpen vorliegt. Ferner kann das Verfahren zur Kostenminimierung einer Erhebung durchgeführt werden, beispielsweise wenn die Elemente der Grundgesamtheit räumlich sehr weit voneinander entfernt sind. Die Klumpenauswahl hat jedoch einen gravierenden Nachteil, der darin liegt, dass sich unter Umständen die einzelnen Klumpen sehr stark voneinander unterscheiden und zugleich die Elemente innerhalb der Klumpen sehr ähnlich sind. Dieses Problem, auch bekannt unter dem Begriff Klumpeneffekt, bedeutet einen Verlust der Genauigkeit einer Schätzung auf Basis einer Klumpenstichprobe gegenüber einer Schätzung auf Basis einer einfachen Zufallsstichprobe.

### 7.6.1.2.3 Mehrstufige Auswahlverfahren

Häufig besteht jedoch der Fall, dass weder eine einfache Zufallsauswahl noch eine geschichtete Auswahl oder eine Klumpenauswahl für die vorliegende Fragestellung angemessen oder durchführbar ist. Insbesondere bei Untersuchungen, die sich auf die allgemeine Bevölkerung beziehen, sind komplexere Auswahlverfahren anzuwenden, die aus einer stufenweisen Abfolge mehrere Zufallsauswahlen bestehen. Bei diesen mehrstufigen Auswahlverfahren bildet die jeweilige Zufallsstichprobe der vorherigen Stufe die Grundlage für die Auswahl der nachfolgenden Zufallsstichprobe. Mehrstufige Auswahlverfahren werden dann angewendet, wenn keine Listen über die Grundgesamtheit zur Verfügung stehen, allerdings Listen von zusammengefassten Elementen der Grundgesamtheit.

Das gebräuchlichste mehrstufige Auswahlverfahren, das seit 1978 in fast allen repräsentativen allgemeinen Bevölkerungsumfragen verwendet wird, ist das „ADM-Master-Sample“ (ADM-Design), welches vom Arbeitskreis Deutscher Marktforschungsinstitute (ADM) entwickelt wurde. Die Grundgesamtheit ist dabei die erwachsene, deutschsprachige Wohnbevölkerung in Deutschland. Das Verfahren umfasst drei Stufen: Die erste Stufe stellen Stimmbezirke zum Deutschen Bundestag dar (Primäreinheiten), wobei kleinere Bezirke zu synthetischen Stimmbezirken zusammengefasst werden. Die Zufallsauswahl aus der Gesamtheit der Stimmbezirke stellt die Auswahlgrundlage für die zweite Stufe dar. Auf dieser Stufe werden die Privathaushalte ausgewählt (Sekundäreinheiten). Deren Auswahl erfolgt in der Marktforschung sehr oft nach dem Random-Route-Verfahren (Ziegler, Kapitel 10 in diesem Band). Auf der dritten Stufe werden schließlich aus den ausgewählten Privathaushalten die Untersuchungseinheiten ausgewählt (Tertiäreinheiten). Soweit es sich nicht um einen Ein-Personen-Haushalt handelt, erfolgt die Auswahl über Tabellen mit Zufallszahlen, dem so genannten Schwedenschlüssel (Kish-Selection-Grid).

## 7.6.2 Nicht-zufällige Auswahlverfahren

In einigen Untersuchungen ist es nicht möglich, eine Zufallsauswahl durchzuführen. Gründe hierfür können darin liegen, dass die Grundgesamtheit unbekannt oder nicht klar abgrenzbar ist oder die Angehörigen sich nicht erfassen lassen. In solchen Fällen kann auf nicht-zufällige Auswahlverfahren zurückgegriffen werden. Dabei ist zwischen willkürlichen Auswahlen und bewussten Auswahlen zu unterscheiden.

### 7.6.2.1 Willkürliche Auswahlen

Willkürliche Auswahlen (Auswahl auf Geratewohl) sollten in der Wissenschaft keine Rolle spielen, da diesen weder ein Auswahlplan noch eine Definition der Grundgesamtheit zugrunde liegt. Die Entscheidung, welches Element in die Stichprobe gerät, liegt lediglich im Ermessen der auswählenden Person.

### 7.6.2.2 Bewusste Auswahlen

Im Unterschied zur willkürlichen Auswahl ist bei einer bewussten Auswahl (Akremi, Kapitel 17 in diesem Band) sowohl eine Definition der Grundgesamtheit als auch ein Auswahlplan vorhanden. Die Auswahl der Untersuchungseinheiten erfolgt jedoch nicht wie bei den Zufallsauswahlen nach dem Zufallsprinzip, sondern durch bewusste Entscheidungen, die zuvor festgelegt werden. So werden z.B. bei der Auswahl typischer Fälle Untersuchungseinheiten ausgewählt, die als besonders charakteristisch für eine definierte Grundgesamtheit angesehen werden. Die Kriterien für die Entscheidung der Aufnahme werden durch theoretische Überlegungen abgeleitet. Dennoch ist auch dieses Verfahren problematisch, da häufig unklar ist, wie stark sich die verschiedenen typischen Fälle ähneln und inwieweit sich die in die Stichprobe einbezogenen Fälle verallgemeinern lassen.

Das wohl am häufigsten eingesetzte Verfahren bei einer bewussten Auswahl ist das *Quotaverfahren*. Dabei werden Personen so ausgewählt, dass die Verteilung bestimmter Merkmale in der Stichprobe exakt der Verteilung dieser Merkmale in der Grundgesamtheit entspricht. Die Auswahl einer bestimmten Person wird dabei den Interviewern überlassen. Kenntnisse über die Verteilung bestimmter Merkmale in der Grundgesamtheit werden dabei als bekannt vorausgesetzt (z.B. aus der amtlichen Statistik). Der Interviewer verfährt bei der Auswahl seiner Zielobjekte nach einem Quotenplan, der genaue Angaben über die Anzahl der zu befragenden Personen mit bestimmten Merkmalen oder Merkmalskombinationen enthält.

Während bei einfachen Quotenverfahren unabhängige Quoten (d.h. die Merkmale müssen in gleicher Weise in Stichprobe und Grundgesamtheit vorhanden sein) verwendet werden, werden bei komplexeren Verfahren kombinierte Quoten (d.h. die Häufigkeit von Kombinationen von Merkmalen in der Stichprobe müssen dem Vorkommen der gleichen Kombinationen in der Grundgesamtheit entsprechen) eingesetzt. Der große Vorteil der Quotenverfahren liegt in der schnellen und preiswerten Durchführbarkeit, weswegen dieses Verfahren in der Marktforschung relativ beliebt ist. Das Quotaverfahren weist gleichsam eine Reihe von Problemen auf, aufgrund dessen die Anwendung dieses Verfahrens in der empirischen Sozialforschung nicht zum Regelfall werden sollte.

## 7.7 Datenerhebung

Nachdem die Entscheidung für ein Forschungsdesign getroffen worden ist, unter Berücksichtigung der Entscheidungen, welche Erhebungsmethode angewendet werden soll und wie die Untersuchungseinheiten ausgewählt werden sollen, ist die eigentliche Planungsphase abgeschlossen. Im zweiten Teil der empirischen Phase erfolgt die eigentliche Feldphase der empirischen Untersuchung, die Datenerhebung. Dazu wird die jeweils gewählte Datenerhebungsmethode auf die Untersuchungseinheiten angewendet. Hierzu stehen eine Reihe verschiedener Datenerhebungsverfahren zur Verfügung, grob unterteilt in Befragung (Reinecke, Kapitel 44 in diesem Band), Beobachtung (Thierbach/Petschick, Kapitel 66 in diesem Band) und Dokumentenanalyse (Salheiser, Kapitel 62 in diesem Band), auf die in dieser Abhandlung nicht weiter eingegangen werden soll. In dieser Phase wird auch das konstruierte Erhebungsinstrument getestet (einem Pretest unterzogen), bevor es eingesetzt wird (Weichbold, Kapitel 19 in diesem Band).

---

## 7.8 Datenerfassung und Datenaufbereitung

Nach der Datenerhebung folgt die Datenerfassung, Datenaufbereitung sowie Plausibilitätskontrollen (Lück/Landrock, Kapitel 28 in diesem Band). In dieser Phase werden die erhobenen Daten für das geplante Auswertungsverfahren vorbereitet, z.B. in der Form von Variablentransformationen, dem Bilden von Subgruppen oder dem Zusammenführen von Datensätzen. Eine gewissenhafte Durchführung der anfallenden Operationen ist hierbei essentiell für die Gültigkeit der darauf basierenden Analysen.

---

## 7.9 Auswertung

Danach ist die empirische Phase des Forschungsprozesses abgeschlossen und das Forschungsprojekt geht in die Auswertungsphase, die Datenanalyse über. In dieser Phase werden die erhobenen Daten unter Einsatz statistischer Verfahren und unter Verwendung spezieller Analysesoftware (z.B. SPSS, Stata oder Mplus) in Hinblick auf die Untersuchungsfrage ausgewertet. Neben deskriptiven Verfahren steht dazu eine große Anzahl von multivariaten Analyseverfahren (Blasius/Baur, Kapitel 79 in diesem Band) zur Verfügung, auf die an dieser Stelle nicht weiter eingegangen werden kann. Auf Basis der Analyseergebnisse können schließlich die im Rahmen der theoretischen Überlegungen aufgestellten Hypothesen angenommen oder abgelehnt werden. Der typische Forschungsablauf endet mit der Verfassung eines Forschungsberichtes (Friedrichs, Kapitel 16 in diesem Band).

## Literatur

- Allison, Paul D. (2009): Fixed Effects Regression Models. Thousand Oaks: Sage
- Engel, Uwe/Reinecke, Jost (1994): Panelanalyse: Grundlagen, Techniken, Beispiele. Berlin: DeGruyter
- Faulbaum, Frank (2014): Methodische Grundlagen der Umfrageforschung. Wiesbaden: VS Verlag
- Gabler, Siegfried (1992): Schneeballverfahren und verwandte Stichprobendesigns. In: ZUMA-Nachrichten 31: 47-69
- Gabler, Siegfried/Hoffmeyer-Zlotnik, Jürgen H. P. (1997) (Hg.): Stichproben in der Umfragepraxis. Opladen: Westdeutscher Verlag
- Greenwood, Ernest (1972): Das Experiment in der Soziologie. In: König, René (1972): 177-220
- Groves, Robert M./Fowler, Floyd J./Couper, Mick P/Lepkowski, James M./Singer, Eleanor/Tourangeau, Roger (2004): Survey Methodology. Hoboken: Wiley
- Häder, Michael (2006): Empirische Sozialforschung. Eine Einführung. Wiesbaden: VS Verlag
- König, René (1972) (Hg.): Beobachtung und Experiment in der Sozialforschung. Köln, Berlin: Kiepenheuer & Witsch
- Koolwijk, Jürgen van (1974): Das Quotenverfahren. In: Koolwijk, Jürgen van/Wieken-Mayser, Maria (1974) (Hg.): 81-99
- Koolwijk, Jürgen van/Wieken-Mayser, Maria (1974) (Hg.): Techniken der empirischen Sozialforschung. München: Oldenburg
- Lohr, Sharon L. (2010): Sampling: Design and Analysis. Boston: Brooks/Cole
- Lynn, Peter (2009) (Hg.): Methodology of Longitudinal Surveys. Hoboken: Wiley
- Opp, Karl-Dieter (2005): Methodologie der Sozialwissenschaften. Einführung in Probleme ihrer Theoriebildung und praktischen Anwendung. Wiesbaden: VS Verlag
- Opp, Karl-Dieter (2010): Kausalität als Gegenstand der Sozialwissenschaften und der multivariaten Statistik. In: Opp, Karl-Dieter (Hg.): 9-38
- Schnell, Rainer (2012): Survey-Interviews. Methoden standardisierter Befragungen. Wiesbaden: VS Verlag
- Schnell, Rainer/Hill, Paul B./Esser, Elke (2013): Methoden der empirischen Sozialforschung, 10. Edition, München: Oldenbourg
- Wolf, Christof/Best, Henning (2010) (Hg.): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS Verlag

Udo Kelle

---

## 8.1 Was sind „Mixed Methods“?

Unter „Mixed Methods“ wird üblicherweise die Kombination qualitativer und quantitativer Forschungsmethoden in einem Untersuchungsdesign verstanden. Es handelt sich um einen Begriff aus der anglo-amerikanischen Methodendebatte in den Sozial- und Erziehungswissenschaften, der seit dem Ende der 1990er-Jahre, konkret seit dem Erscheinen der Monographie „Mixed Methodology“ von Abbas Tashakkori und Charles Teddlie (1998) große Prominenz erlangt hat. Von den amerikanischen Erziehungswissenschaften ausgehend hat sich eine eigene Mixed Methods-Bewegung gebildet – mittlerweile existieren eine ganze Reihe von Lehrbüchern (etwa Creswell/Plano Clark 2007; Morse/Niehaus 2009; Kuckartz/Cresswell 2014), ein in zweiter Auflage erschienenes umfangreiches Handbuch (Tashakkori/Teddlie 2010), seit 2007 eine Zeitschrift mit Namen „Journal of Mixed Methods Research“ (JMMR) und eine internationale Fachgesellschaft unter dem Namen „Mixed Methods International Research Association“ (MMIRA).

Die Kombination von qualitativen und quantitativen Methoden in einer empirischen Studie ist eine seit vielen Jahren geübte Praxis in der Sozialforschung. Berühmte Forschungsprojekte, die zur Entwicklung der Sozialwissenschaften und zu ihrer Theoriebildung viel beigetragen haben, haben qualitative und quantitative Methoden kombiniert. Hierzu zählt etwa die *Marienthal-Studie* (Jahoda et al. [1933], 1975), wo Sozialforscher bei der Untersuchung der Massenarbeitslosigkeit in einem österreichischen Dorf der frühen 1930er-Jahre die „Lücke zwischen den nackten Ziffern der Statistik und den zufälligen Eindrücken der sozialen Reportage“ (Jahoda et al. [1933], 1975: 15) mit einem großen Methodeninstrumentarium, bestehend aus Fragebögen (Reinecke, Kapitel 44 in diesem Band), „natürlichen“ oder „prozessproduzierten“ Daten und Dokumenten wie Bibliotheksstatistiken und Wahlziffern (Salheiser, Kapitel 62 in diesem Band), standardisierten Beobachtungen (Thierbach/Petschick, Kapitel 66 in diesem Band), persönlichen Interviews (Küsters, Kapitel 40 in diesem Band), Schulaufgaben von Kindern (Ernst, Kapitel

63 in diesem Band), informellen Gelegenheitsgesprächen mit Beteiligten u.v.a.m. zu füllen suchten. Aber auch die *Hawthorne-Studie* (Roethlisberger/Dickson 1939) wenige Jahre später, die Studie von Festinger und Kollegen über *Weltuntergangssekt*en (Festinger et al. 1956) oder Zimbardos (1969) bekanntes Gefangenensexperiment liefern gute Beispiele für solche frühen Mixed Methods Studien (siehe auch Kelle 2008: 46).

---

## 8.2 Die Debatte um Mixed Methods

Dass Methodenkombination oder -integration erst in den letzten Jahren zu einem lebhaft diskutierten methodologischen Thema werden konnte, mittlerweile als ein eigener methodologischer Ansatz oder sogar als ein selbständiges „Paradigma“ (Johnson/Onwuegbuzie 2004) betrachtet werden, lässt sich dabei nur unter Bezug auf historische Entwicklungen in der empirischen Sozialforschung angemessen verstehen: qualitative und quantitative Methoden haben sich seit den 1920er-Jahren sehr stark unabhängig voneinander im Rahmen eigener methodologischer Traditionen und Denkschulen entwickelt. Unterschiedliche Ziele von Sozialforschern gaben dabei den Anlass für die Entstehung dieser Methodentraditionen, und aus diesen wurden wiederum verschiedene Kriterien für die Qualität von Forschungsmethoden abgeleitet, die sehr leicht miteinander in Konflikt geraten können (Reichertz, Kapitel 2 in diesem Band).

### 8.2.1 Die quantitative Methodentradition

Am Beginn der quantitativen Methodentradition im 19. Jahrhundert stand das Ziel, soziale Makrophänomene, wie etwa die Struktur und Entwicklung der Bevölkerung, die Verteilung von Suiziden, Scheidungs- und Kriminalitätsraten oder die öffentliche Meinung empirisch zu untersuchen. Die hierzu notwendigen statistischen Erhebungs- und Analysemethoden wurden zuerst im Kontext der Demographie und seit den 1930er-Jahren zunehmend anhand von Fragestellungen der sich damals akademisch etablierenden Fächern der (Sozial)psychologie und Soziologie entwickelt. Da die Erfassung von Makrophänomenen durch Mikrodaten nicht möglich ist ohne eine starke Standardisierung der Datenerhebung, wurde die quantitative Methodentradition durch die Konstruktion entsprechender Mess- und Skalierungsverfahren wie die Guttman-, Thurstone- oder Likertskalierung (Blasius, Kapitel 83 in diesem Band) und der ihnen zugrundeliegenden Theorien sowie durch die Entwicklung von angemessenen Methoden zur möglichst verzerrungsfreien Ziehung großer Stichproben (Häder/Häder, Kapitel 18 in diesem Band) geprägt. Seit den 1950er-Jahren wurden zudem intensiv Anschlussmöglichkeiten an zeitgleich laufende Debatten in der analytischen Wissenschaftsphilosophie über die angemessene Methodologie der empirischen Wissenschaften gesucht und auf dieser Basis ein bis heute sehr einflussreiches Modell sozialwissenschaftlicher Forschung, das „hypothetiko-deduktive“ Modell des Forschungsprozesses (Reichertz, Stein, Kapitel 2 und 7 in diesem Band) ausgearbeitet.

### 8.2.2 Die qualitative Methodentradition

Die qualitative Methodentradition (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) nahm ihren Anfang von empirischen Studien zur Migrations- und Stadtsoziologie, die in den 1910er- und 20er-Jahren am Institut für Soziologie und Sozialanthropologie der Universität von Chicago begonnen wurden. Ziel dieser Arbeiten, die sich ursprünglich stark an der ethnologischen Feldforschung (Knoblauch, Kapitel 36 in diesem Band) und der Tradition des sozialkritischen Investigativjournalismus in den USA orientierten, war eine „naturalistische“ Erforschung sozialer Lebenswelten „aus erster Hand“, welche den Sozialforschern, zumeist Angehörige der Mittelschicht, zuvor kaum vertraut waren: soziale Milieus an der Peripherie der städtischen Gesellschaft, die von sozialer Exklusion, Marginalisierung und abweichendem Verhalten gekennzeichnet waren. Eine theoretische Fundierung fanden diese empirischen und forschungsmethodischen Ansätze in jenen Arbeiten zur soziologischen Theorie, die seit den 1930er-Jahren eine zum damals v.a. in den USA dominierenden Strukturfunktionalismus konkurrierende, „alternative Klassik“ der Sozialtheorie begründeten, die auch als „interpretative Soziologie“ bezeichnet wird – das waren vor allem der symbolische Interaktionismus, die soziologische Phänomenologie und der Sozialkonstruktivismus. Der von diesen Ansätzen postulierten „Prozesshaftigkeit“, das heißt der Flexibilität und Wandelbarkeit sozialer Strukturen musste mit Hilfe einer Forschungsstrategie Rechnung getragen werden, mit deren Hilfe sich rasch wandelnde und heterogene Muster und Regeln in den Blick genommen werden konnten – und dies schien nur durch eine „offene“, wenig hypothesesgeleitete Erhebung nicht standardisierter Daten möglich.

### 8.2.3 Methodenmix und Methodentriangulation

Diese jeweils unterschiedlichen Ziele quantitativer und qualitativer Forschung führte zur Entwicklung von mittlerweile sehr weit ausgearbeiteten verschiedenen Standards und Kriterien für die Qualität von Forschung in beiden Traditionen (Krebs/Menold, Flick, Kapitel 30 und 29 in diesem Band), zwischen denen leicht Konflikte entstehen:

Das sind auf der quantitativen Seite die Postulate der Objektivität (das ist die Beobachterunabhängigkeit) und Reliabilität (also der prinzipiellen Wiederholbarkeit) der Datenerhebung (Krebs/Menold, Kapitel 30 in diesem Band), der statistischen Verallgemeinerbarkeit von Befunden (Häder/Häder, Kapitel 18 in diesem Band) und der theoriegeleiteten Forschung (Reichert, Stein, Kapitel 2 und 7 in diesem Band).

Die Aufgabe, vor die sich demgegenüber Sozialforscher der qualitativen Tradition gestellt sahen, nämlich die Exploration bislang unbekannter sozialer Lebensformen und eine detaillierte Beschreibung von Einzelfällen (Hering/Schmidt, Kapitel 37 in diesem Band), lokaler Milieus usw., machte eine methodische Offenheit (Strübing, Kapitel 32 in diesem Band) dem Forschungsgegenstand gegenüber erforderlich, die sich kaum in Einklang bringen lässt mit der Formulierung exakter Hypothesen ex ante, der Konstruktion

von standardisierten Instrumenten und der Erhebung großer Fallzahlen. Ausgehend von Grundannahmen der interpretativen Soziologie lässt sich nämlich argumentieren, dass den Forschenden das Wissen über die im Feld relevanten Regeln der Sinnkonstitution sozialen Handelns (Kurt/Herbrik, Kapitel 33 in diesem Band), welches sie für die Hypothesenformulierung und Instrumentenkonstruktion benötigen würden, oft gar nicht zur Verfügung steht.

Diese unterschiedlichen Forschungsziele und Qualitätskriterien bilden den eigentlichen Nährboden für die seit Jahrzehnten andauernde Kontroverse zwischen der quantitativen und der qualitativen Methodentradition, in der sich die Kontrahenten oft wechselseitig die Validität und Wissenschaftlichkeit ihres Vorgehens abgesprochen haben. Beide Traditionen neigen dabei dazu, voneinander abgeschottete Lager zu bilden, dem sich besonders methodisch und methodologisch interessierte Sozialwissenschaftler oft zuordnen. Das hat u.a. zur Folge, dass die im deutschen Sprachraum verfügbaren Lehr- und Handbücher sich entweder fast ausschließlich den quantitativen Methoden (bspw. Schnell et al. 2011; Diekmann 2007) oder der qualitativen Forschung (bspw. Flick 2007; Lamnek 2010) widmen. Dieses Methodenschisma, welches z.B. seinen Niederschlag vor einigen Jahren in der Gründung einer eigenen Sektion „Methoden der qualitativen Sozialforschung“ neben der schon seit langem bestehenden Sektion „Methoden der empirischen Sozialforschung“ in der Deutschen Gesellschaft für Soziologie (DGS) fand, wird von den Vertretern der einer methodenintegrativen Position manchmal auch ironisch als „Paradigmenkriege“ (Gage 1989) bezeichnet. Hierzu steht die schon beschriebene Tatsache in Kontrast, dass in der sozialwissenschaftlichen Forschungspraxis, und dabei auch in zahlreichen berühmten Forschungsarbeiten, mit großem Gewinn gleichzeitig auf beide Ansätze zurückgegriffen wurde und wird.

Die Formulierung einer methodologischen Begründung für einen solchen Methodenmix ist allerdings keine triviale Aufgabe. Bereits vor dem Entstehen der Mixed Methods-Bewegung wurde hierzu der Begriff der „*Triangulation*“ (Flick, Kapitel 29 in diesem Band) verwendet. Hiermit wird ursprünglich ein in der Navigation, Landvermessung und Astronomie häufig eingesetztes Verfahren bezeichnet, mit dessen Hilfe der Abstand zweier Punkte A und B zu einem dritten unbekannten Punkt C durch zwei Winkelmessungen (bei bekanntem Abstand zwischen A und B) berechnet werden kann. Diesen Begriff haben ursprünglich quantitativ orientierte Psychologen Campbell/Fiske (1959) für die Validierung psychologischer Testergebnisse durch die Korrelation mit anderen (ähnlichen und unterschiedlichen) Tests („multitrait-multimethod matrices“), also für die Kombination zweier quantitativer Verfahren, eingesetzt. In der Folge haben andere Autoren den Begriff übernommen für eine Kombination unterschiedlicher (auch, aber nicht nur, qualitativer und quantitativer) Methoden, so etwa Denzin (1978) mit seinem Konzept der „*Methodentriangulation*“.

Den Ideen von Campbell und Fiske und auch von Denzin (der diesen Ansatz aber später selber verworfen hat, Flick 2004: 20), wonach der Einsatz verschiedener Methoden zu einer Maximierung der Validität von Daten und Ergebnissen führt, wurde allerdings von anderen Autoren widersprochen, die darauf hingewiesen haben, dass sich qualitative und quantitative

Methoden oftmals auf ganz unterschiedliche Gegenstände und Gegenstandsbereiche beziehen. Um ein Beispiel zu nehmen: in der Lebenslaufforschung können etwa quantitative Methoden verwendet werden, um objektive Lebenslaufdaten (bspw. Wechsel der Arbeitsstelle, Familiengründung) zu erfassen (Pötter/Prein, Kapitel 86 in diesem Band), qualitative Daten, um die subjektive Deutung dieser Lebensereignisse durch die Befragten in den Blick zu nehmen (Rosenthal, Kapitel 35 in diesem Band). Bei einer solchen Methodenkombination, bei denen sich die verschiedenen Methoden auf jeweils verschiedene Phänomene und Sachverhalte beziehen, können sich qualitative und quantitative Forschungsergebnisse im besten Fall ergänzen, aber nicht unbedingt entsprechen oder sich widersprechen (was ja eine Voraussetzung wäre für eine Validierung von Daten und Methoden, siehe hierzu auch Fielding/Fielding 1986: 32 ff.).

### 8.2.4 Ergebnisse und Funktionen von Methodentriangulation

Ein Blick auf empirische Forschungsprojekte, in denen qualitative und quantitative Methoden gemeinsam eingesetzt werden, zeigt, dass Methodentriangulation tatsächlich in unterschiedlicher Weise genutzt werden kann:

- Sie kann einerseits zu einer *Validierung oder Validitätskritik* (Krebs/Menold, Flick, Kapitel 30 und 29 in diesem Band) von (quantitativen oder qualitativen) Daten und Methoden verwendet werden,
- in anderen Fällen aber auch zu *unterschiedlichen Forschungsergebnissen* führen, die sich *wechselseitig ergänzen* (Kelle 2008: 49 ff.).

Werden qualitative und quantitative Methoden in einem gemeinsamen Forschungsprojekt eingesetzt, können die mit den beiden methodischen Ansätzen gewonnenen Forschungsergebnisse nämlich sowohl

- *konvergent* sein, d.h. übereinstimmen (was Argumente zugunsten einer Validität der Daten und Methoden liefert) – das wäre etwa der Fall, wenn in einer Mixed Methods Befragung über Einstellungen gegenüber Minderheiten die Befragten sowohl in standardisierten Fragebögen (Reinecke, Kapitel 44 in diesem Band) als auch in qualitativen Leitfadeninterviews dieselben (Helfferich, Kapitel 39 in diesem Band) oder ähnliche antisemitische Einstellungen äußern;
- *divergieren*, d.h. sich widersprechen (wodurch Validitätsprobleme entdeckt werden können) – das wäre bspw. der Fall, wenn Befragte, die in standardisierten Befragungen keinerlei antisemitische Einstellungen zeigen, in qualitativen Tiefeninterviews dann aber judentreibende Äußerungen machen;
- *komplementär* sein, d.h. sich wechselseitig ergänzen (so dass aus der Verbindung aus den Ergebnissen beider Methoden ein umfassenderes Bild des Forschungsgegenstandes entsteht) – ein Beispiel hierfür würde eine Lebenslaufstudie liefern, in der die quantitativen Daten zeigen, dass Angehörige einer bestimmten Berufsgruppe ein spezifisches Muster hinsichtlich ihrer Weiterbildungsbemühungen zeigen und hierfür Erklärungen gesucht werden in Daten aus qualitativen Interviews, in denen die Befragten begründen, warum sie sich für bestimmte Weiterbildungsmaßnahmen entschieden haben.

Triangulation selbst ist demnach noch gar kein methodologisches Konzept oder Modell der Methodenintegration, sondern nur eine deskriptive Metapher mit einem weiten Bedeutungsfeld.

Auch die Mixed Methods-Bewegung hat bislang noch keine allgemein akzeptierten methodologischen Modelle zur Verbindung qualitativer und quantitativer Forschung hervorgebracht. Konsens existiert hier nur über die grundlegende Idee, dass sowohl qualitative als auch quantitative Methoden spezifische Schwächen aufweisen, die durch die Stärken des jeweils anderen Ansatzes ausgeglichen werden können. Dies führt zu dem „*fundamentalen Prinzip*“ von *Mixed Methods Designs*: „.... Methoden sollen in einer solchen Weise kombiniert werden, dass ihre komplementären Stärken und Schwächen genutzt werden können (...) das impliziert freilich die Anerkennung der Tatsache, dass alle Methoden sowohl Beschränkungen als auch Stärken aufweisen“ (Johnson/Turner 2003: 299; Übersetzung UK).

Hinsichtlich der Frage jedoch, welche Schwächen welcher Methodentraditionen in welchen Forschungsfeldern durch welche Stärken der jeweils anderen Tradition ausgeglichen werden können, ist die Debatte noch sehr stark im Fluss. Vor allem fehlt bislang eine systematische Taxonomie typischer Methodenprobleme und Validitätsbedrohungen qualitativer und quantitativer Forschung und von Mixed Methods Designs, mit deren Hilfe solche Probleme in konkreten Forschungsprojekten aufgedeckt und bearbeitet werden könnten. Die Diskussion und die Methodenforschung konzentriert sich demgegenüber aktuell stark auf die Beschreibung, Darstellung und deskriptive Systematisierung von Designformen. Dabei konnte bislang aber noch kein einheitliches System entwickelt werden, auf das sich alle Diskutanten verständigen konnten, vielmehr konkurrieren eine Vielzahl unterschiedlicher Taxonomien miteinander (für einen Überblick siehe Cresswell/Plano-Clark 2007: 59ff.).

---

### 8.3 Mixed Methods Designs – Typen und Funktionen

Vorschläge, systematisch zwischen Typen von Mixed Methods Designs zu differenzieren, orientieren sich an unterschiedlichen Kriterien:

- an der Frage, ob *Methoden bei ihrer Anwendung kombiniert* werden (also ob etwa qualitative und quantitative Daten in einem Fragebogen zusammen eingesetzt werden; ob qualitative Daten mit quantitativen Methoden ausgewertet werden o.ä.) oder ob Methoden getrennt voneinander eingesetzt und dann nur ihre *Ergebnisse aufeinander bezogen* werden,
- anhand der *Reihenfolge*, in der qualitative und quantitative Methoden in einem Forschungsdesign miteinander kombiniert werden,
- anhand der *relativen Bedeutung*, die der quantitative oder der qualitative Untersuchungsteil im Kontext des gesamten Forschungsprojektes einnimmt,
- oder anhand der *Funktion* des jeweiligen Methodenstranges innerhalb des Gesamtprojektes.

Im Folgenden sollen anhand dieser vier Kriterien unterschiedliche Mixed Methods Designs erörtert werden.

### **8.3.1 Methodenmix in der Methodenanwendung vs. Methodenmix als Ergebnisintegration**

*Qualitative und quantitative Methoden können während der Datenerhebung kombiniert werden.* Dies ist etwa bei der gemeinsamen Erhebung von standardisierten und verbalen Daten in einem halbstrukturierten Interview der Fall, eine seit langem geübte Praxis in der Umfrageforschung. Informationen können auch zweifach, qualitativ und quantitativ, aufgezeichnet werden: das ist etwa dann der Fall, wenn bei einem standardisierten Interview nicht nur die Antworten des Befragten auf dem Fragebogen vermerkt werden, sondern der gesamte Gesprächsverlauf auf Tonträger aufgezeichnet oder gefilmt (Tuma/Schnettler, Kapitel 68 in diesem Band) wird – man erhält dann zusätzlich zu den standardisierten Daten ein qualitatives Interaktionsprotokoll.

Eine weitere Möglichkeit der Methodenkombination besteht darin, *qualitative Methoden der Datenerhebung mit quantitativen Methoden der Datenauswertung zu kombinieren*. In der sozialwissenschaftlichen Medienforschung wird ein solches Verfahren schon seit mehr als 50 Jahren unter der Bezeichnung „quantitative Inhaltsanalyse“ eingesetzt: Unstrukturierte verbale Daten werden entweder direkt quantitativ ausgewertet (indem bspw. die Häufigkeit bestimmter Wörter in Zeitungsartikeln o.ä. gezählt wird) oder zuerst durch menschliche Kodierer kodiert und anschließend die Verteilung der Kodes in den Texten statistisch analysiert. Die sich gegenwärtig rapide entwickelnden Verfahren der Videographie (Tuma/Schnettler, Kapitel 68 in diesem Band) ermöglichen ebenfalls eine quantitative Weiterverarbeitung qualitativer, unstrukturierter Daten sowohl in automatisierter Form (indem etwa Gesprächspausen oder die Länge von Wortbeiträgen von Teilnehmern einer Gruppendiskussion (Vogl, Kapitel 41 in diesem Band) gemessen werden) als auch nach einer Kodierung anhand eines entsprechenden Kategorienschemas (Mayring/Fenzl, Kapitel 38 in diesem Band).

In einer solchen Methodenkombination liegen allerdings auch *Gefahren*. Der Begriff der „Kodierung“ ist etwa sowohl in der qualitativen als auch in der quantitativen Methodentradition gebräuchlich, bedeutet aber dort jeweils etwas anderes. In der qualitativen Forschung werden Kodes oft ad hoc und anhand des Datenmaterials entwickelt und dienen nur der *Indizierung* des Datenmaterials (d.h. mit der Hilfe der Kodierungen können Textstellen wiedergefunden werden, die dann vergleichend interpretiert werden), während die in einer quantitativen Inhaltsanalyse eingesetzten Kodes *zählbare Sachverhalte* repräsentieren und deshalb strenger Objektivitäts- und Reliabilitätsanforderungen unterliegen. Beide Formen der Kodierung unreflektiert zu vermischen kann zu erheblichen Fehlinterpretationen führen: qualitativen Kodes kann dann etwa eine Aussagekraft unterstellt werden, die sie in konkreten Fällen (wo sie ggfs. nur dem Wiederfinden von Textstellen dienen) gar nicht haben sollen. Wird aber, um eine adäquate statistische Analyse von

Kodierkategorien zu ermöglichen, ein vorab formuliertes Kategorienschema eingesetzt, dessen Reliabilität und Objektivität auch statistisch gesichert werden kann, wird die spezifische Stärke qualitativer Kodierungsverfahren (welche ja eine Entwicklung von Kategorien anhand von empirischen Daten ermöglichen sollen) aufgegeben.

Die Alternative zur Kombination von Methoden während ihrer Anwendung besteht darin, eine Studie aus unterschiedlichen (qualitativen und quantitativen) Teilprojekten aufzubauen, in denen die Datenerhebung und Datenauswertung konsequent mit nur jeweils einem der beiden methodischen Ansätze durchgeführt wird. Mixed Methods bedeutet in diesem Fall, dass die *Ergebnisse dieser Teilstudien systematisch aufeinander bezogen werden*. Hierbei muss allerdings darauf geachtet werden, dass qualitative und quantitative Forschungsergebnisse unterschiedlich große Reichweite besitzen und sich jeweils verschiedene Geltungs- und Verallgemeinerungsansprüche (Häder/Häder, Akremi, Kapitel 18 und 17 in diesem Band) mit ihnen verbinden. Dieses Problem wird in der Mixed Methods-Community auch unter dem Begriff der „*Meta-Inferenzen*“ diskutiert (Tashakkori/Teddlie 2008) – Schlussfolgerungen aus qualitativen und quantitativen Teilstudien eines Mixed Methods Designs können und müssen auf einer übergeordneten („Meta-“)ebene integriert werden.

### **8.3.2 Dominanz vs. Gleichwertigkeit qualitativer und quantitativer Methoden in einem Mixed Methods-Design**

Hinsichtlich des relativen Stellenwerts qualitativer und quantitativer Methoden in einem Mixed Methods-Design werden in der Literatur unterschiedliche Positionen vertreten. Während etwa Tashakkori/Teddlie (1998: 43) unterscheiden zwischen (a) *Designs mit einem gleichgewichtigen Status beider Methodenansätze* und (b) *Designs, bei denen einer der Ansätze (qualitativ oder quantitativ) dominant ist*, vertreten Morse/Niehaus (2009: 23 f.) die Auffassung, dass ein Mixed Methods-Projekt aus einer (qualitativen oder quantitativen) „*Kernkomponente*“ und einer Ergänzung bestehen sollte, bei der dann der jeweils andere methodische Ansatz eingesetzt wird. Während die Kernkomponente im Prinzip auch für sich allein stehen könnte, diene die ergänzende Komponente nur zur Erhebung von zusätzlichen Informationen, die für die Beantwortung der Forschungsfrage noch benötigt würden – die methodischen Werkzeuge, die im Rahmen der ergänzenden Komponente des Designs eingesetzt werden, sollten deshalb eher als Strategien denn als Methoden bezeichnet werden.

Die Differenzierung nach Wertigkeit oder Bedeutung des jeweiligen methodischen Ansatzes in einem Mixed Methods-Design kann allerdings den Methodenstreit leicht wieder auflöben lassen. Die Idee der Methodenkombination ist ja, wie bereits erwähnt, keineswegs neu: In einem schon in den 1950er-Jahren erschienenen und später häufig zitierten Aufsatz hatten Barton und Lazarsfeld vorgeschlagen, *qualitative Vorstudien zur Generierung von Hypothesen* zu nutzen, welche in darauf folgenden quantitativen Hauptuntersuchungen geprüft werden sollten. Qualitative Methoden haben hier also eine Hilfsfunktion

in einem hypothetiko-deduktiven Design: „Qualitatives Datenmaterial eignet sich besonders für die exploratorische Phase eines Forschungsprojekts“ (Barton/Lazarsfeld [1955] 1984: 82). Dieses besondere Konzept der Methodenkombination wurde auch oft in der quantitativen Methodenliteratur aufgegriffen, dort aber weiter verwässert – qualitative Methoden sollen demnach im Vorfeld einer quantitativen Studie so eingesetzt werden, dass die Forschenden „in mehr oder weniger impressionistischer Form“ durch „Ideen, Gespräche und Explorationen“ das untersuchte „Problem (...) strukturieren“ (Friedrichs 1973 : 52).

Hier muss nun aus Sicht der qualitativen Forschungstradition mit Nachdruck darauf hingewiesen werden, dass qualitative Methoden als Verfahren methodisch kontrollierten Fremdverstehens keineswegs impressionistisch durchgeführt werden dürfen, sondern ein systematisches, aufwändiges Vorgehen erfordern, welches durchaus auch Schritte der Hypothesenprüfung enthalten kann (Kurt/Herbrik, Kapitel 33 in diesem Band). Die Durchführung von Feldbeobachtungen (Thierbach/Petschick, Kapitel 66 in diesem Band) und die Erhebung von qualitativen Interviews (Helfferich, Kapitel 39 in diesem Band) zur explorativen und empiriegestützten Hypothesen- und Theoriebildung kann ein sinnvoller erster Schritt in einem Mixed Methods-Design sein, lässt sich aber keineswegs nebenbei erledigen.

### 8.3.3 Sequentielle vs. parallele Mixed Methods-Designs

Die Reihenfolge, in denen die qualitativen und quantitativen Untersuchungsschritte abgearbeitet werden, wird in der Literatur häufig als zentrales formales Kriterium zur Unterscheidung zwischen Mixed Methods-Designs herangezogen (etwa bei Morse 1991 oder bei Tashakkori/Teddlie 1998). Dabei kann grob zwischen einer *sequentiellen* und einer *parallelen Kombination* qualitativer und quantitativer Methoden unterschieden werden:

- Die bereits erwähnte, ursprünglich von Lazarsfeld und Barton vorgeschlagene Strategie, eine qualitative Pilotstudie zur Generierung von Hypothesen einzusetzen, die in einer anschließenden quantitativen Hauptstudie geprüft werden, beschreibt ein *sequentielles qualitativ-quantitatives Design*. Zur Kennzeichnung von Mixed Methods-Designs wird oft eine abkürzende Schreibweise verwendet, die ursprünglich auf Morse (1991) zurückgeht: die qualitativen und quantitativen Untersuchungsschritte werden mit „quan“ und „qual“, ihre Bedeutung durch Groß- und Kleinschreibung deutlich gemacht und ein sequentielles Design durch den Gebrauch eines Pfeils „→“, ein paralleles durch ein Pluszeichen „+“. Bei Verwendung dieser Notation würde man das Barton-Lazarsfeld-Design mit „qual → QUAN“ abkürzen.
- Daneben werden in Mixed Methods Forschung häufig auch *sequentielle quantitative-qualitative Designs* eingesetzt, in denen also eine quantitative Teilstudie am Anfang steht und beide Methodentraditionen gleiches Gewicht besitzen (QUAL → QUAN) oder in denen sogar der qualitative Methodenstrang bedeutsamer ist (QUAL → quan).

- Weiterhin können auch beide methodischen Ansätze zur gleichen Zeit, also in einem *parallelen Design* (QUAN + QUAL, oder z.B. QUAN + *qual*) eingesetzt werden.

Nun können die beschriebenen Designs in einer Mixed Methods-Studie auch mehrfach miteinander kombiniert werden (etwa *qual* → QUAN + *qual* → QUAL). Das hat zu einer verwirrenden Vielfalt von Designformen und von deren Bezeichnungen in der Literatur geführt, wobei manche Autoren ähnliche Designs oft mit unterschiedlichen Namen versehen (ein „parallel design“ heißt z.B. bei bestimmten Autoren auch „concurrent design“) oder sprachliche Differenzierungen treffen (wie die schon erwähnte Unterscheidung von Morse zwischen „Methode“ und „Strategie“), die in anderen Teilen der Community nicht nachvollzogen werden.

### **8.3.4 Funktionen qualitativer und quantitativen Methoden in Mixed Methods-Designs**

Entsprechend der pragmatischen Orientierung der Mixed Methods-Community ist für die Auswahl eines Designs die Funktion des jeweiligen Ansatzes zur Beantwortung einer bestimmten Forschungsfrage bzw. zur Bearbeitung eines Untersuchungsgegenstands entscheidend. Auch hinsichtlich dieses Aspekts gibt es bislang nur wenige umfassende systematische Arbeiten – ein bedeutsamer Grund hierfür ist sicher darin zu suchen, dass die Diskussion über Mixed Methods nur selten Bezüge zur Theoriedebatte in den Sozialwissenschaften aufgenommen hat (Kelle 2008).

Grundlegender Konsens besteht aber darüber, dass durch eine Kombination von Methoden die Schwächen und Stärken der beiden Traditionen wechselseitig ausgeglichen werden sollen (s.o.). Im Sinne der bereits vorgetragenen Überlegungen zum Begriff Triangulation lassen sich Methoden also einerseits kombinieren, um *typischen Validitätsbedrohungen qualitativer und quantitativer Forschung entgegenzuwirken*. Das können z.B. Probleme des Nonresponse (Engels/Schmidt, Kapitel 23 in diesem Band), der „sozialen Erwünschtheit oder der „Meinungslosigkeit“ (Hlwatsch/Krickl, Kapitel 20 in diesem Band) sein, die in standardisierten Interviews auftreten, oder Probleme der Verallgemeinerbarkeit oder Übertragbarkeit von solchen Befunden, die an kleinen qualitativen Stichproben gewonnen wurden.

oder die Ergebnisse von qualitativen und quantitativen Untersuchungsschritten können sich komplementär ergänzen, so dass „blinde Flecken“ oder Erklärungslücken, die die qualitativen oder quantitativen Forschungsergebnisse jeweils offenlassen, durch die Resultate der anderen Methode ausgeglichen werden.

Im Einzelnen kann ein Mixed Methods Design bspw. eine oder mehrere der folgenden *Funktionen* übernehmen (wobei diese Liste keinesfalls vollständig ist):

- Prüfung der Geltungsreichweite qualitativer Forschungsergebnisse, die in qualitativen Studien anhand empirischen Datenmaterials entwickelten Kategorien, Hypothesen und Theorien können in quantitativen Studien anhand großer Fallzahlen auf ihre Geltungsreichweite hin geprüft werden.

- *Erklärung quantitativer Forschungsergebnisse:* In einem sequentiellen quantitativ-qualitativen Design (QUAN→QUAL) kann der qualitative Untersuchungsteil Erklärungen für schwer verständliche quantitative Ergebnisse liefern (wenn diese Funktion im Vordergrund steht, sprechen manche Autoren auch von einem „explanativen Design“, Creswell/Plano-Clark 2007: 71 ff.).
- *Instrumentenkonstruktion:* Qualitative Methoden können die Entwicklung quantitativer Instrumente unterstützen, etwa indem in einem sequentiellen qualitativ-quantitativen Design (qual→QUAN) standardisierte Items aufgrund qualitativer Interviewdaten entwickelt werden.
- *Pretest von standardisierten Instrumenten:* Qualitative Methoden können Validitätsbedrohungen standardisierter Instrumente aufzudecken helfen, indem in einem sequentiellen qualitativ-quantitativen Design (QUAL→QUAN) standardisierte Instrumente in qualitativen Pretestinterviews (Weichbold, Kapitel 19 in diesem Band) eingesetzt werden oder indem in parallelen quantitativ-qualitativen Designs (QUAN + qual, oder auch quan + QUAL) zusätzlich zu den standardisierten Daten qualitative Daten über die Erhebungssituation (Interaktionsprotokolle o.ä.) gesammelt werden.
- *Aufklärung unerklärter Varianzen in quantitativen Daten:* Sowohl in einem parallelen als auch in einem sequentiellen quantitativ-qualitativen Design (QUAN → QUAL) können unerklärte Varianzen und unbeobachtete Heterogenität (Latcheva/Davidov, Faulbaum, Kapitel 55 und 31 in diesem Band) im quantitativen Datenmaterial durch qualitative Daten und Methoden aufgeklärt werden (Kelle 2008: 237).
- *Unterstützung des qualitativen Sampling durch quantitative Daten:* die Fallauswahl und das Sampling einer qualitativen Studie (Akremi, Kapitel 17 in diesem Band) können in einem sequentiellen quantitativ-qualitativen Design (quan→QUAL) durch die Ergebnisse einer quantitativen Vorstudie angeleitet werden, die die Forschenden über die Verteilung von für die qualitative Fallauswahl relevanten Merkmalen im Forschungsfeld informiert.

---

## 8.4 Fazit

Gegenwärtig ist Mixed Methods ein in der sozialwissenschaftlichen Methodendiskussion weithin ernst genommener und in der empirischen Forschungspraxis beliebter Ansatz. Die methodologische Debatte jedoch über eine trennscharfe Systematik möglicher Designformen und ihrer Funktionen ist, wie bereits dargestellt, noch im Fluss und gekennzeichnet von einem Nebeneinander verschiedener Terminologien und Taxonomien.

Auch über methodologische Begründungen für Mixed Methods wird nach wie vor gestritten – wobei die Frage diskutiert wird, inwieweit Mixed Methods ein eigenes erkenntnistheoretisches Paradigma (gegründet etwa im philosophischen Pragmatismus) benötigt, welches dann gleichberechtigt neben Paradigmen aus der qualitativen Methodentradition (etwa: den Konstruktivismus) und aus der quantitativen Tradition (wie den Postpositivismus oder den kritischen Realismus) treten kann. Gegen diese vor allem in

der anglo-amerikanischen Debatte beliebte Sichtweise ließe sich einwenden, dass die komplexe Aufgabe (qualitatives und/oder quantitatives) Forschungshandeln zu begründen, sich nicht mit dem Verweis auf philosophische Schlagworte erledigen lässt. Die Wahl der passenden Forschungsmethode hat sich vielmehr an der Fragestellung, dem untersuchten Gegenstandsbereich und den inhaltlich-theoretischen Vorannahmen der Forschenden zu orientieren. Wer bspw. demographische Phänomene, etwa Geburtenraten im internationalen Vergleich, untersuchen möchte, wird auch dann, wenn er erkenntnistheoretisch dem Konstruktivismus nahe steht (philosophisch gesehen also „qualitativ orientiert“ wäre), ohne quantitative Daten und statistische Analysen kaum zureckkommen. Wer dahingegen subjektive Wahrnehmungs- und Deutungsmuster von Mitgliedern einer gesellschaftlich marginalisierten Subkultur explorieren will, wird durch ein dogmatisches Festhalten an der Methode standardisierter Befragung als vermeintlichem Königsweg der Sozialforschung kaum brauchbare Resultate erzielen können.

Von der Mixed Methods Bewegung werden vor allem dann weiterhin innovative Impulse für die Entwicklung von Forschungsmethoden ausgehen, wenn verstärkt solche gegenstandsbezogenen Fragen thematisiert und dabei auch intensiv Anschlüsse an die Theoriedebatte in den Sozialwissenschaften gesucht werden.

## Literatur

- Arnold, William T./Levine, David (1969): Nebraska Symposium on Motivation 17. Lincoln, NE: University of Nebraska Press
- Barton, Allen H./Lazarsfeld, Paul F. ([1955] 1984): Einige Funktionen von qualitativer Analyse in der Sozialforschung. In: Hopf, Christel/Weingarten, Elmar (Hg.): 41-89
- Bergmann, Manfred M. (Hg.): Advances in Mixed Methods Research. London: Sage
- Campbell, Donald T./Fiske, Donald W. (1959): Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. In: Psychological Bulletin 56: 81-105
- Creswell, John W./Plano Clark, Vicki L. (2007): Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: Sage
- Denzin, Norman K. (1978): The Research Act. A Theoretical Introduction to Sociological Methods. New York: McGraw Hill
- Diekmann, Andreas (2007): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen (18. Auflage). Reinbek bei Hamburg: rowohlt enzyklopädie
- Festinger, Leon/Riecken, Henry W./Schachter, Stanley (1956): When Prophecy Fails. Minneapolis, MI.: University of Minnesota Press
- Fielding, Nigel G./Fielding, Jane L. (1986): Linking Data. (Qualitative Research Methods Vol. 4). London: Sage
- Flick, Uwe (2004): Triangulation. Eine Einführung. Wiesbaden: VS Verlag
- Flick, Uwe (2007): Qualitative Forschung. Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften. Reinbek bei Hamburg: rowohlt enzyklopädie
- Friedrichs, Jürgen (1973): Methoden empirischer Sozialforschung. Opladen: Westdeutscher Verlag
- Gage, Nathaniel L. (1989): The paradigm wars and their aftermath: A "historical" sketch of research and teaching since 1989. In: Educational Researcher 18: 4-10
- Hopf, Christel/Weingarten, Elmar (1984): Qualitative Sozialforschung. Stuttgart: Klett-Cotta
- Jahoda, Marie/Lazarsfeld, Paul F./Zeisel, Hans ([1933]1982). Die Arbeitslosen von Marienthal. Frankfurt: Suhrkamp
- Johnson, Burke/Onwuegbuzie, Anthony (2004): Mixed methods research: A research paradigm whose time has come. In: Educational Researcher 33: 14-26
- Johnson, Burke/Turner, Lisa A. (2003): Data Collection Strategies in Mixed Methods Research. In: Tashakkori, Abbas / Teddlie, Charles (Hg.): 297-319
- Kelle, Udo (2008): Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung. Theoretische Grundlagen und methodologische Konzepte. Wiesbaden: VS Verlag
- Kuckartz, Udo/Cresswell, John (2014): Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren. Wiesbaden: VS Verlag
- Lamnek, Siegfried (2010): Qualitative Sozialforschung: Lehrbuch (5., vollständig überarbeitete Auflage). Weinheim, Basel: Beltz

- Morse, Janice/Niehaus, Linda (2009): Mixed Method Design. Principles and Procedures. Walnut Creek, CA: Left Coast Press
- Morse, Janice M. (1991): Approaches to qualitative-quantitative methodological triangulation. In: *Nursing Research* 40: 120-133
- Roethlisberger, Fritz J./Dickson, William J. (1939): Management and the worker. Cambridge, Ma.: Harvard University Press
- Schnell, Rainer/Hill, Paul B./Esser, Elke (2011): Methoden der empirischen Sozialforschung (9. Auflage). München: Oldenburg
- Tashakkori, Abbas/Teddlie, Charles (1998): Mixed Methodology. Combining qualitative and quantitative approaches. Thousand Oaks, CA: Sage
- Tashakkori, Abbas/Teddlie, Charles (2008): Quality of inferences in Mixed Methods research: calling for an integrative framework. In: Bergmann, Manfred M. (Hg.): 101-120
- Tashakkori, Abbas/Teddlie, Charles (2010) (Hg.): Sage Handbook of Mixed Methods in Social and Behavioral Research. Thousand Oaks, CA: Sage
- Zimbardo, P.G. (1969): The human choice: Individuation, reason and order versus deindividuation, impulse, and chaos. In: Arnold, William T./Levine, David (Hg.): 237-307

Nicola Döring

---

## 9.1 Einleitung

Mit „Evaluationsforschung“, „wissenschaftlicher Evaluation“ oder kurz „Evaluation“ ist die Bewertung eines Gegenstandes mit Hilfe sozialwissenschaftlicher Methoden durch Evaluationsfachleute gemeint (Widmer/De Rocchi 2012: 11). Die Evaluationsforschung ist ein boomendes interdisziplinäres Forschungsfeld, in dem sowohl qualitative als auch quantitative Methoden der empirischen Sozialforschung zum Einsatz kommen (Döring/Bortz 2014). In zahlreichen Bereichen des öffentlichen Lebens werden Evaluationsmaßnahmen inzwischen rechtlich vorgeschrieben. So verlangt beispielsweise das Hochschulgesetz vieler Bundesländer die Evaluation von Lehrveranstaltungen und Studiengängen. Mit eigenen Fachgesellschaften und Fachzeitschriften hat sich die Evaluationsforschung wissenschaftlich fest etabliert. Im deutschsprachigen Raum sind die *Gesellschaft für Evaluation* (DeGEval; [www.dgeval.de](http://www.dgeval.de)) sowie die *Zeitschrift für Evaluation* (ZfEv) einschlägig. Das vorliegende Handbuchkapitel grenzt zunächst die *Evaluations- von der Grundlagenforschung* ab, stellt zentrale *Begriffe der Evaluationsforschung* vor und unterscheidet wichtige *Typen von Evaluationsstudien*. Anschließend präsentiert er die für gute Evaluationsstudien bindenden *Evaluationsstandards*, erläutert verschiedene *Ansätze der Evaluationsforschung* und geht abschließend auf die einzelnen *Phasen einer Evaluationsstudie* ein.

---

## 9.2 Abgrenzung der Evaluations- von der Grundlagenforschung

Die sozialwissenschaftliche Methodenausbildung setzt üblicherweise bei der *akademischen Grundlagenforschung* (Stein, Przyborski/Wohlrab-Sahr, Kelle, Kapitel 7, 6 und 8 in diesem Band) an: Ein selbst gewähltes und als wissenschaftlich relevant erachtetes Forschungsproblem wird mit dem Ziel des Erkenntnisgewinns eigenverantwortlich bearbeitet und gemäß wissenschaftlichen Gütekriterien der Fachöffentlichkeit präsentiert. Nehmen

wir ein *Beispiel aus der Prostitutionsforschung*, die u.a. in Psychologie, Soziologie, Kriminologie, Sozialarbeit, Medizin usw. betrieben wird (Döring, 2014). Vor dem Hintergrund der Beobachtung, dass sich Prostitutionswerbung und Kontaktanbahnung in den letzten Jahren stark in das Internet verlagert haben und hier auch männliche Prostituierte sichtbarer werden, könnte man der Forschungsfrage nachgehen, welche Merkmale und Motive männliche Prostituierte haben, die ihre Dienste über Internet-Plattformen bewerben und die man auf diesem Wege auch für Forschungsinterviews erreichen kann.

### **9.2.1 Evaluationsforschung als Anwendungs- und Auftragsforschung**

Im Unterschied zu der primär wissenschaftlichen Kriterien verpflichteten Grundlagenforschung handelt es sich bei der Evaluationsforschung um *praxisnahe Anwendungs- und meist auch Auftragsforschung*. Themen werden also nicht selbst gewählt, sondern von den Auftraggebenden an die Evaluierenden herangetragen. Und die Ergebnisse müssen nicht nur wissenschaftlich überzeugen, sondern auch nützlich für die Praxis sein. Praxisbezogene Forschungsprobleme sind häufig umfassender und schwerer greifbar (da zahlreiche Betroffenengruppen und Kontextfaktoren eine Rolle spielen) als grundlagenwissenschaftliche Themen, die sich bei Bedarf stark eingrenzen lassen. So kann man bei einer grundlagenwissenschaftlichen Studie selbst entscheiden, ob man – wie im obigen Beispiel – lediglich männliche oder auch weibliche und Transgender-Prostituierte untersucht und ob man sich auf den Status Quo bezieht oder Veränderungen über die Zeit erfasst usw.

Bei einer Evaluationsstudie dagegen geben Auftraggebende bestimmte Zielsetzungen und Rahmenbedingungen des Forschungsprozesses vor. Oft werden hier relativ umfangreiche Analysen und praktisch direkt anwendbare Resultate erwartet. So gab die deutsche Bundesregierung im Jahr 2007 – fünf Jahre nach Einführung des Prostitutionsgesetzes (ProstG) – eine Evaluationsstudie in Auftrag, die beurteilen sollte, ob das Gesetz die Lebens- und Arbeitsbedingungen von Prostituierten in Deutschland tatsächlich wie intendiert verbessert hatte (SOFFI K, 2007). Gemäß dem 2002 in Kraft getretenen ProstG gilt Prostitution in Deutschland nicht länger als sittenwidrig; dementsprechend haben Prostituierte ein Anrecht auf ihre Entlohnung und können am Kranken-, Sozialversicherungs- und Rentensystem teilnehmen. In der Begriffslogik quantitativer Sozialforschung ist hier also die Gesetzesänderung (Einführung des ProstG) die unabhängige Variable, deren Auswirkungen auf die Lebens- und Arbeitsbedingungen von Prostituierten als abhängige Variablen zu betrachten sind.

Eine umfassende Beantwortung der genannten Evaluationsfrage erfordert die Berücksichtigung unterschiedlicher Formen der Prostitution, unterschiedlicher Gruppen von Prostituierten sowie ihrer Angehörigen. Neben den Sichtweisen der unmittelbar Betroffenen sind auch Informationen von weiteren Anspruchsgruppen zu beachten (z.B. Polizei, Staatsanwaltschaften, Gerichte, Fachberatungsstellen, Sozialarbeit). Auch ist zu klären, was im Einzelnen als „Verbesserung“ der Lebens- und Arbeitsbedingungen gelten und empirisch erfasst werden kann (z.B. Gefühl der Entstigmatisierung und gesellschaftlichen

Inklusion; Anzahl der offiziellen Anstellungsverhältnisse von Prostituierten; Anzahl und Art der neu abgeschlossenen Krankenversicherungsverträge; Häufigkeit von Problemen mit Kunden, die den vertraglich vereinbarten Preis nicht bezahlen). Gefordert ist in jedem Fall eine *Veränderungsmessung*, denn es geht um den Vergleich der Situation vor und nach der Einführung des neuen Gesetzes. Dabei muss jede festgestellte Veränderung der Situation der Prostituierten nach 2002 *kausal* eindeutig auf die Gesetzesänderung zurückführbar sein, damit sie zur Bewertung des Gesetzes dienen kann (andere Einflussfaktoren müssen also beispielsweise durch Berücksichtigung einer Kontrollgruppe ausgeschlossen werden). Die Erwartungen an eine aussagekräftige Evaluationsstudie sind typischerweise sehr hoch, können aber oft nur partiell eingelöst werden. Denn für eine kausal möglichst eindeutig interpretierbare Veränderungsmessung sind prospektive Längsschnittdesigns notwendig (Pötter/Prein, Schupp, Kapitel 86 und 73 in diesem Band). Diese sind sehr teuer und aufwändig – und eben definitionsgemäß im Nachhinein nicht zu realisieren: Ein im Jahr 2007 beauftragtes Evaluationsteam kann die Situation vor 2002 nur begrenzt empirisch rekonstruieren.

### 9.2.2 Besondere Anforderungen an Evaluationsforschende

Evaluationsstudien sind also oft wesentlich komplexer als Grundlagenstudien und meist nur in *interdisziplinärer Teamarbeit* zu bewältigen. Nicht zuletzt ist zu beachten, dass Evaluationsforschung als wissenschaftliche Auftragsforschung im Praxisfeld, von deren Ergebnissen weitreichende Entscheidungen abhängen können (z.B. Weiterführung einer Bildungsmaßnahme, Schließung einer sozialen Einrichtung, Änderung eines Gesetzes), besondere soziale Verantwortung verlangt und zahlreiche äußere und innere Konflikte mit sich bringen kann. Die Forschenden sind einerseits *Dienstleistende für die Auftraggebenden* und sollen diesen mit ihren Evaluationsergebnissen bei Entscheidungen helfen. Andererseits sind sie *wissenschaftlicher Unabhängigkeit* verpflichtet und dürfen keine beliebigen Wunschergebnisse produzieren (etwa um von einem zufriedengestellten Auftraggeber dann ein Anschlussprojekt zu erhalten). Sie müssen eine Datenerhebung nach wissenschaftlichen Standards in die Wege leiten und umsetzen (auch wenn das im Praxisfeld Schwierigkeiten bereitet). Letztlich müssen sie auch unliebsame Ergebnisse in die Auswertung einbeziehen, sie in geeigneter Weise nach innen und außen kommunizieren und fair interpretieren.

Indem die Evaluationsforschung sich der wissenschaftlichen Bewertung praxisrelevanter und teilweise auch politisch sehr brisanter Untersuchungsgegenstände widmet, stellt sie an die Forschenden eine Reihe zusätzlicher Anforderungen: Wer das methodische Handwerkszeug der qualitativen und quantitativen Grundlagenforschung beherrscht, ist damit noch lange nicht in der Lage, eine gute Evaluationsstudie erfolgreich zu realisieren. Dafür sind vor allem spezifische *soziale und persönliche Kompetenzen* notwendig: Es geht beispielsweise darum, die Sichtweisen unterschiedlicher (mitunter zerstrittener) Betroffenengruppen überparteiisch einzubeziehen, überzogene Erwartungen von Auftraggebenden

zugunsten einer realistischen Evaluationsfragestellung zu korrigieren, dem mitunter massiven sozialen Druck standzuhalten, bestimmte Wunschergebnisse zu produzieren oder Negativbefunde zu verschweigen sowie mehr oder minder subtile Manipulations- und Vereinnahmungsversuche unterschiedlicher Interessengruppen zu erkennen und geeignet gegenzusteuern usw. Die DeGEval (2012) hat u.a. deswegen konkrete *Anforderungen an Evaluierende* formuliert und daraus auch *Empfehlungen für die Ausbildung* abgeleitet. Wesentlicher Baustein der Ausbildung ist die praktische Evaluationstätigkeit in allen Phasen des Evaluationsprozesses unter Anleitung langjährig erfahrener Evaluationsfachleute. Denn gerade die für Evaluationsforschung (im Unterschied zur Grundlagenforschung) essentiellen *sozial-kommunikativen und ethischen Problemlösungsstrategien*, die zudem passgenau auf den jeweiligen Evaluationskontext und dessen soziale Normen und rechtliche Rahmenbedingungen zuzuschneiden sind (z.B. Schule, Betrieb, Partei, Museum, Bordell), lassen sich allein theoretisch kaum vermitteln.

---

## 9.3 Zentrale Begriffe der Evaluationsforschung

Die Merkmale der Evaluationsforschung werden im Folgenden anhand zentraler Evaluationsbegriffe wie *Evaluationsgegenstand*, *Anspruchsgruppen*, *Evaluationskriterien*, *Evaluationsfunktionen* und *Evaluationsnutzung* verdeutlicht.

### 9.3.1 Evaluationsgegenstand

Während man in der Grundlagenforschung vom Untersuchungsgegenstand bzw. Forschungsobjekt (research object) spricht, ist in der Evaluationsforschung analog vom *Evaluationsgegenstand* bzw. *Evaluationsobjekt* oder auch vom *Evaluandum* (evaluation object, evaluand) die Rede. Das Spektrum der Evaluationsgegenstände ist dabei sehr breit und kann kaum trennscharf und erschöpfend dargestellt werden. Evaluiert werden Bildungseinrichtungen und Lehrveranstaltungen ebenso wie medizinische und psychotherapeutische Behandlungsmethoden, Gesetze und Verordnungen, Bauvorhaben und Entwicklungshilfeprojekte, Methoden der Unternehmens- und Mitarbeiterführung, technologische Innovationen oder politische Kampagnen. Besonders typisch in der Evaluationsforschung ist die Bewertung von größer angelegten Programmen (z.B. Weiterbildungs-, Frauenförder-, Integrations-Programme), die auf bestimmte individuelle und kollektive Veränderungen abzielen, man spricht deswegen auch von *Programm-Evaluation* (program evaluation).

### 9.3.2 Anspruchsgruppen

Da die Evaluationsforschung auf eine wissenschaftlich fundierte Bewertung des jeweiligen Evaluationsgegenstandes hinausläuft, welche oft als Entscheidungsgrundlage dient und

weit reichende praktische Konsequenzen haben kann, ist es wichtig, die Sichtweisen der *Betroffenen* (v.a. der Zielgruppen einer Maßnahme) und der *Beteiligten* (v.a. der Durchführenden einer Maßnahme) bzw. allgemein der zentralen *Anspruchsgruppen* (stakeholder) zu berücksichtigen. Wesentliche Informationen über den Evaluationsgegenstand sind oft nur über die Anspruchsgruppen zu erhalten. Zudem besteht ethisch der Anspruch, die Interessen der Betroffenen und Beteiligten zu berücksichtigen und nicht zu übergehen. Im Rahmen der Evaluationsforschung gilt es, für die jeweilige Evaluationsfrage die relevanten Stakeholder zu identifizieren und deren Anliegen angemessen in den Forschungsprozess einzubeziehen. Ein gravierender Mangel einer Evaluationsstudie besteht darin, wichtige Stakeholder zu übersehen. Bei der angesprochenen Evaluationsstudie zum Prostitutionsgesetz wären u.a. die Krankenkassen als Stakeholder einzubeziehen, etwa um zu eruieren, ob und unter welchen Konditionen sie Prostituierten unter dieser Berufsbezeichnung Versicherungsverträge anbieten (würden).

### 9.3.3 Evaluationskriterien

Ziel der Evaluationsforschung ist die wissenschaftlich fundierte Bewertung des Evaluationsgegenstandes. Dazu müssen klare und messbare Bewertungskriterien definiert, entsprechende Daten erhoben und interpretiert werden. Die *Auswahl und Festlegung der Evaluationskriterien* bestimmt maßgeblich das spätere Evaluationsergebnis. Kriterien sind unbrauchbar, wenn sie sich nicht empirisch erfassen lassen oder wenn sie inhaltlich nicht zur Lösung des Evaluationsproblems beitragen. Es ist zu beachten, dass neben der Definition der Evaluationskriterien auch noch die jeweiligen *Bewertungsmaßstäbe* festzulegen sind: Welche Ausprägungen des Evaluationskriteriums sollen als positives, ausreichendes oder negatives Ergebnis gewertet werden? Solche Maßstäbe können beispielsweise als *Mindest-, Regel- oder Maximalstandards* gefasst werden. Sie können empirisch relativ zu einem Ausgangs- oder anderen Vergleichszustand formuliert sein (z.B. Symptomreduktion im Vorher-Nachher-Vergleich um 10%) oder auch als absolute Werte bzw. Idealnormen festgelegt werden (z.B. Symptomfreiheit aller Patienten als Behandlungsziel; siehe zu Vergleichskriterien Gollwitzer/Jäger 2009: 48).

Wenn also bei der Evaluation des ProstG untersucht werden soll, ob sich die Lebens- und Arbeitsbedingungen von Prostituierten verbessert haben, so müssen zunächst die relevanten Kriterien abgeleitet werden. Da Prostituierte bekanntermaßen stark unter gesellschaftlicher Ausgrenzung und Stigmatisierung leiden, könnte ein Evaluationskriterium die Erfahrung von Stigmatisierung sein. Da Stigmatisierung sehr belastend ist, könnte bereits ein geringer Entstigmatisierungs-Effekt (z.B. von einem halben Punkt auf einer Stigmatisierungsskala) als Erfolg gewertet werden. Derartige Festlegungen von Evaluationskriterien und Bewertungsmaßstäben sind auf der Basis des aktuellen Forschungsstandes sowie mit Rückgriff auf Expertenurteile detailliert zu begründen.

### 9.3.4 Evaluationsfunktionen

Die im Zuge einer Evaluation erarbeitete Bewertung des Evaluationsgegenstandes anhand unterschiedlicher Evaluationskriterien (und Bewertungsmaßstäbe) soll in der Praxis verschiedenen Funktionen, Zwecken bzw. Zielen dienen. Die *Evaluationsfunktionen* (bzw. *Evaluationszwecke* oder *Evaluationsziele*) werden in der Literatur zwar unterschiedlich gegliedert (Stockmann 2002: 3ff., Widmer/De Rocchi 2012: 27f.), sie hängen aber eng zusammen. Es lassen sich pragmatisch fünf Funktionen abgrenzen:

- *Erkenntnisfunktion*: Evaluationsforschung trägt definitionsgemäß dazu bei, wissenschaftliche Erkenntnisse über die Eigenschaften und Wirkungen von Evaluationsgegenständen zu sammeln.
- *Lern- und Dialogfunktion*: Im Zuge der Evaluationsforschung vollziehen die Beteiligten Lernprozesse und es treten verschiedene Anspruchsgruppen miteinander in Dialog. Eine Evaluationsstudie kann die Funktion haben, einen solchen Dialog über den Evaluationsgegenstand zu fördern und fruchtbar zu gestalten.
- *Optimierungsfunktion*: Über die reine Wissenserweiterung und damit verbundene Lern- und Dialogprozesse hinaus soll Evaluation oft Erkenntnisse liefern, die zur zielgerichteten Verbesserung des Evaluationsgegenstandes beitragen können.
- *Entscheidungsfunktion*: Evaluationsergebnisse bilden nicht selten die Basis für Wahlentscheidungen zwischen verschiedenen Evaluationsgegenständen (soll Maßnahme A oder B umgesetzt werden?) oder auch für Entscheidungen über die Implementierung oder Weiterführung einer Maßnahme (lohnt es sich, die Maßnahme fortzusetzen oder sollte sie abgebrochen werden?). Damit eine Evaluationsstudie die Entscheidungsfunktion erfüllt, muss das Evaluationsteam die Evaluationsergebnisse zu klaren Praxis-Empfehlungen verdichten. Die Entscheidung selbst wird indessen nicht vom Evaluationsteam getroffen, es liefert im Zuge der unabhängigen wissenschaftlichen Bewertung des Evaluationsgegenstandes allenfalls eine Entscheidungshilfe.
- *Legitimationsfunktion*: Sowohl die Durchführung von Evaluationsforschung als auch ihre Befunde sollen dazu beitragen, die Entwicklung und Durchführung einer Intervention zu kontrollieren und nach außen zu legitimieren. Vor allem bei öffentlich geförderten Programmen und Maßnahmen dient regelmäßige Evaluation dazu, über die sachgerechte Verwendung öffentlicher Gelder Rechenschaft abzulegen.

### 9.3.5 Evaluationsnutzung

Verantwortung im Hinblick auf die Evaluationsnutzung (evaluation use; Patton 2008) zu übernehmen bedeutet für die Evaluierenden zunächst, ihre Ergebnisse in *Zwischen- und Abschlussberichten* verständlich darzustellen und sich vor eindeutigen Bewertungen und daraus abgeleiteten Praxisempfehlungen nicht zu scheuen. Anders als in der Grundlagenforschung, in der zu Recht vor überinterpretierten Ergebnissen zu warnen ist, kann die Evaluationsforschung ihrer Ratgeberpflicht nur nachkommen, wenn sie ihre Ergebnisse

zu praxisbezogenen Empfehlungen verdichtet. So ist in den vorliegenden Evaluationsberichten zum ProstG (SOFFI K 2007, Renzikowski 2007, Kavemann/Steffan 2013) die Einschätzung zu finden, dass Entkriminalisierung der richtige Weg ist, um die Arbeits- und Lebensbedingungen von freiwilligen Prostituierten zu verbessern, ohne die Verfolgung von Zwangsprostitution zu erschweren. Zudem wird in den Berichten diskutiert, welche zusätzlichen rechtlichen Änderungen (z.B. im Steuer-, Gewerbe- und Baurecht) für eine Gleichstellung notwendig wären und welche sozialen Interventionen (z.B. Ausstiegshilfen) sinnvoll sind.

Evaluationsergebnisse und darauf aufbauende Praxisempfehlungen in Berichten darzustellen, reicht für eine erfolgreiche Nutzung der Evaluationsergebnisse im Sinne der genannten Optimierungs- und Entscheidungsfunktion aber meist nicht aus. Denn oft kann nicht davon ausgegangen werden, dass Evaluationsberichte auch wirklich gelesen, verstanden und akzeptiert werden und dann auch noch zur Ableitung und Umsetzung der richtigen Konsequenzen führen. Die Förderung der Evaluationsnutzung umfasst deswegen weitere Aktivitäten wie *Beratung, Coaching, Workshops, Weiterbildungskurse usw.*, in denen Evaluierende in der Ratgeberrolle Bedeutung und praktische Konsequenzen der Evaluationsergebnisse mit den intendierten Evaluationsnutzenden (also denjenigen Personen und Organisationen, die aus den Evaluationsergebnissen praktische Konsequenzen ziehen sollen) behandeln.

---

## 9.4 Wichtige Typen von Evaluationsstudien

Um Evaluationsstudien griffig zu kennzeichnen, werden sie als *formative oder summative Evaluationen*, als *interne oder externe Evaluationen*, als *Konzept-, Prozess- oder Ergebnisevaluationen* bezeichnet. Damit sind unterschiedliche Aspekte einer Evaluationsstudie angesprochen. Größere Evaluationsprojekte kombinieren nicht selten mehrere dieser Typen miteinander, um umfassendere Erkenntnisse zu generieren.

### 9.4.1 Formative und summative Evaluation

Eine *formative Evaluation* liefert Ergebnisse, die der schrittweisen Verbesserung des Evaluationsgegenstandes dienen sollen (oft wird mit qualitativen Methoden gearbeitet: z.B. offene Befragung zu Verbesserungsvorschlägen). Eine *summative Evaluation* verfolgt das Ziel einer abschließenden Beurteilung (hier sind quantitative Methoden einschlägig: z.B. Bewertung auf einer Schulnotenskala). Im Hinblick auf die Evaluationsfunktionen erfüllt die formative Evaluation eine Optimierungsfunktion, die summative Evaluation dagegen eine Entscheidungs- und/oder Legitimationsfunktion.

### 9.4.2 Interne und externe Evaluation

Bei der *internen Evaluation* wird die Evaluationsforschung von Personen durchgeführt, die derselben Organisation angehören (z.B. die Lehrveranstaltungen und Lehrenden einer Universität werden unter Leitung des universitätseigenen Didaktik-Zentrums evaluiert). Ein Sonderfall der internen Evaluation ist die *Selbstevaluation*, hier organisieren diejenigen, die für den Evaluationsgegenstand verantwortlich sind, auch selbst die Evaluation (z.B. eine Dozentin nutzt für ihre eigenen Lehrveranstaltungen ein spezielles Evaluationsverfahren). Bei der *externen Evaluation* werden Außenstehende tätig (z.B. ein Studiengang einer Hochschule wird durch ein Expertengremium einer anderen Hochschule evaluiert). Die interne Evaluation hat den Vorteil, dass den Beteiligten vor Ort meist mehr detaillierte Informationen über den Evaluationsgegenstand zugänglich sind, die externe Evaluatoren kaum erlangen können. Sie hat gleichzeitig den Nachteil, dass sie eher als parteiisch eingeschätzt wird, die externe Evaluation dagegen als neutraler.

### 9.4.3 Konzept-, Prozess- und Ergebnisevaluation

Ein Evaluationsgegenstand kann in unterschiedlichen Phasen seines Lebenszyklus evaluiert werden. Eine *Konzeptevaluation* bewertet das Konzept einer Maßnahme bereits im Vorfeld einer möglichen Umsetzung. Die *Prozessevaluation* bewertet die Durchführung einer laufenden Maßnahme. Die *Ergebnisevaluation* schließlich konzentriert sich auf die Bewertung der Effekte der Maßnahme. Prozess- und Ergebnisevaluation spielen in der Praxis eine so große Rolle, dass für sie umfassendere Evaluationsansätze entwickelt wurden (siehe Abschn. 9.6.1 für Modelle der Ergebnisevaluation und 9.6.2 für Modelle der Prozessevaluation).

---

## 9.5 Evaluationsstandards

Woran ist eine gute Evaluationsstudie zu erkennen und von einer durchschnittlichen oder gar mangelhaften Studie abzugrenzen? Gute Evaluationsstudien müssen nach aktuellem nationalem und internationalem Diskussionsstand in der Wissenschaftsgemeinschaft vier grundlegende Eigenschaften aufweisen (DeGEval 2008):

1. *Nützlichkeit* (utility): „Die Nützlichkeitsstandards sollen sicherstellen, dass die Evaluation sich an den geklärten Evaluationszwecken sowie am Informationsbedarf der vorgesehenen Nutzer und Nutzerinnen ausrichtet.“
2. *Durchführbarkeit* (feasability): „Die Durchführbarkeitsstandards sollen sicherstellen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst geplant und ausgeführt wird.“

3. *Fairness* (propriety): „Die Fairnessstandards sollen sicherstellen, dass in einer Evaluation respektvoll und fair mit den betroffenen Personen und Gruppen umgegangen wird.“
4. *Genauigkeit* (accuracy): „Die Genauigkeitsstandards sollen sicherstellen, dass eine Evaluation gültige Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt.“

Den vier Qualitätsdimensionen sind insgesamt 25 *Einzelstandards* zugeordnet. Wer selbst eine Evaluationsstudie durchführt, sollte bereits in der Planungsphase prüfen, ob und wie das avisierte Vorgehen die Evaluationsstandards erfüllt, im Zweifelsfall ist das Studienkonzept anzupassen.

Wie jede wissenschaftliche Studie ist auch eine Evaluationsstudie in ihrem Ablauf sorgfältig zu dokumentieren (z.B. Anfertigung und Archivierung von Protokollen aller Stakeholder-Meetings), so dass das Vorgehen im Nachhinein auch für Außenstehende nachvollziehbar und nachprüfbar ist. Zuweilen ist es gewünscht oder erforderlich, eine Evaluationsstudie ihrerseits zu evaluieren (z.B. wenn Auftraggebende Zweifel an der Qualität der Studie anmelden und der Meinung sind, sie sei den vertraglich vereinbarten Preis nicht wert). Die systematische Bewertung einer Evaluationsstudie im Hinblick auf die Erfüllung der Evaluationsstandards wird als *Meta-Evaluation* bezeichnet und typischerweise als externe Evaluation durchgeführt: Fachkollegen bekommen alle Unterlagen zu der betreffenden Evaluationsstudie, halten ggf. Rücksprache mit den Beteiligten und/oder sammeln anderweitig weitere Daten, um die Qualität der Evaluationsstudie zu beurteilen. Eine Meta-Evaluation muss ihrerseits Evaluationsstandards genügen.

---

## 9.6 Verschiedene Ansätze der Evaluationsforschung

Um den vielfältigen praxisbezogenen Anforderungen gerecht zu werden, die an Evaluationsstudien gestellt werden, sind in den letzten Jahren eine Reihe von unterschiedlichen *Evaluationsmodellen* vorgeschlagen und erprobt worden. Die einzelnen Evaluationsmodelle unterscheiden sich in der Ausrichtung des methodischen Vorgehens und im Evaluationsverständnis. Grob lassen sich die einzelnen Modelle vier *Evaluationsansätzen* zuordnen (Döring/Bortz 2014, Kap. 18). Es ist zu beachten, dass die verschiedenen Evaluationsansätze und -modelle wichtige Hilfestellungen zur Durchführung von Evaluationsstudien liefern, allerdings nicht als starre Vorschriften zu verstehen sind. Evaluationsforschende gehen meistens nicht streng modellkonform, sondern *eklektisch* vor, d.h. sie kombinieren Elemente aus unterschiedlichen Evaluationsansätzen miteinander, um das vorliegende Evaluationsproblem optimal bearbeiten zu können.

### 9.6.1 Der ergebnisorientierte Evaluationsansatz

Der ergebnisorientierte Evaluationsansatz konzentriert sich auf einige wenige ergebnisbezogene Merkmale des Evaluationsgegenstandes.

- Bei einer *Wirksamkeitsevaluation* geht es darum, ob und wie stark eine Maßnahme ihre intendierten Wirkungen tatsächlich erzeugt (*Effektivität*). Wichtig ist dabei, dass die gemessenen Wirkungen wirklich kausal auf das Evaluationsobjekt zurückgeführt werden können, was methodisch am eindeutigsten mit experimentellen oder quasi-experimentellen Kontrollgruppenstudien gelingt. So muss die Wirksamkeitsevaluation eines neuen Kopfschmerzmedikaments nachweisen, dass das Medikament besser wirkt als ein Placebo.
- Bei der *ökonomischen Evaluation* setzt man den Nutzen der Maßnahme mit den anfallenden Kosten in Relation (*Effizienz*). Die ökonomische Evaluation steht vor der Aufgabe, sowohl den Nutzen des Evaluationsgegenstandes als auch seine Kosten möglichst exakt zu quantifizieren, am besten in Geldwerteinheiten.

### 9.6.2 Der systemische Evaluationsansatz

In Abgrenzung zu Evaluationen, die sich auf einzelne Merkmale einer Maßnahme – vor allem Ergebnismerkmale der Effektivität und Effizienz – konzentrieren, sind unter dem systemischen Ansatz hier solche Evaluationsmodelle angesprochen, die eine Maßnahme *ganzheitlicher* als die ergebnisorientierten Modelle betrachten: Der systemische Ansatz untersucht den Evaluationsgegenstand also nicht nur hinsichtlich seiner Wirkungen, sondern auch im Hinblick auf Prozesse seiner Konzeption, Umsetzung und Nutzung sowie im Hinblick auf die Bedingungen des Umfeldes.

Das prominenteste Evaluationsmodell im systemischen Ansatz ist das bereits in den 1960er Jahren entwickelte *CIPP-Evaluations-Modell* von Daniel Stufflebeam (Stufflebeam et al. 1971), das aus den vier Komponenten Context, Input, Process und Product Evaluation besteht. Zur Durchführung von Evaluationen nach dem CIPP-Modell liegt eine umfassende Evaluations-Checkliste vor ([www.wmich.edu/evalctr/checklists](http://www.wmich.edu/evalctr/checklists)). Eine konkrete Evaluationsstudie im CIPP-Modell kann dabei stets nur eine kleine Auswahl aus dem großen Spektrum möglicher Evaluationsfragen bearbeiten:

- Context
  - Welche Personengruppen haben Bedarf an einer geplanten (oder bereits realisierten) Intervention?
  - Welche vergleichbaren Maßnahmen existieren bereits?
  - Wie lassen sich Mittel zur kurz- und langfristigen Finanzierung der Intervention beschaffen?
- Input
  - Wie gut entspricht das Konzept der Intervention den Bedürfnissen und Besonderheiten der Zielgruppen?

- Wie berücksichtigt das Konzept mögliche Widerstände gegen seine Umsetzung?
- Wie viel Personal steht zur Umsetzung der Maßnahme zu welchen Zeitpunkten zur Verfügung und welche Qualifikationen bringt es mit?
- Process
  - In welchen Situationen bzw. an welchen Orten wird die Intervention wie intensiv genutzt?
  - Welche Probleme treten bei der Umsetzung der Intervention auf?
  - Wie kann die Intervention verbessert werden?
- Product
  - Wie groß ist die Nutzergruppe der Intervention? In welchem Verhältnis steht die aktuelle Nutzergruppe zur Zielgruppe? (*Impact*) Spezieller: Ist die Intervention Menschen mit körperlichen Beeinträchtigungen zugänglich (*Barrierefreiheit*)? Ist die Intervention von Menschen mit unterschiedlichem kulturellem Hintergrund gleichermaßen nutzbar (*Internationalisierung*)?
  - Erreicht die Intervention die intendierten Wirkungen? In welchem Umfang werden die intendierten Wirkungen erzielt? (*Effektivität*) Spezieller: Gibt es differenzielle Wirkungen in unterschiedlichen Gruppen? Profitieren Frauen und Männer gleichermaßen von der Intervention, oder wird eine Gruppe benachteiligt (*Gender Mainstreaming*)?
  - In welchem Verhältnis stehen Kosten und Nutzen der Intervention? Erreicht die Intervention erwünschte Ziele kostengünstiger als andere Interventionen? (*Effizienz*)
  - Welche Vorkehrungen werden getroffen, um die Intervention an anderen Orten/für andere Zielgruppen anzubieten? (*Portabilität*)
  - Welche Vorkehrungen werden getroffen, um die Intervention langfristig oder dauerhaft zu erhalten? Welchen Status hat die Intervention z.B. zwei Jahre nach Abschluss des ersten Evaluationsprojektes? (*Nachhaltigkeit*)

### 9.6.3 Der theorieorientierte Evaluationsansatz

Bei der so genannten theorieorientierten Evaluation wird über die Frage der reinen Kausalität (die in der wirksamkeitsorientierten Evaluation untersucht wird) hinausgegangen: Man versucht, die Wirkungsmechanismen zu rekonstruieren: *Welche* Interventionskomponenten funktionieren für *welche* Zielgruppe unter *welchen* Umständen *wie* und *warum*? Es geht also darum, die Wirkprozesse des Evaluationsgegenstandes theoretisch zu modellieren (Programm-Theorie) und dann empirisch zu prüfen. Die *theoriegeleitete Evaluation* (theory-driven evalution; Chen 1990, 2005) ist ein bekanntes Evaluationsmodell innerhalb dieses Ansatzes. Es besteht aus einem theoretischen *Aktionsmodell* (das genau beschreibt und erklärt, wie die Maßnahme funktioniert) und einem *Veränderungsmodell* (das genau beschreibt und erklärt, wie und wodurch die Maßnahme die gewünschten Wirkungen erzeugt).

### 9.6.4 Der akteursorientierte Evaluationsansatz

Der akteursorientierte Evaluationsansatz (Widmer/De Rocchi 2012) schließlich ist dadurch gekennzeichnet, dass seine Forschungsaktivitäten in den Dienst einzelner Stakeholder-Gruppen gestellt werden (klientenorientierte Modelle) oder alle Stakeholder gleichermaßen aktiv einbeziehen sollen, bis hin zur Steuerung der Evaluation durch die Evaluierten selbst (stakeholderorientierte Modelle).

- *Klientenorientierte Evaluationsmodelle*: Die Evaluierenden stellen hier ihre Dienste primär auf eine ausgewählte Stakeholder-Gruppe ab. Seien es die Auftraggebenden (Management-orientierte Evaluation), seien es sozial unterprivilegierte Bevölkerungsgruppen, deren Situation durch eine Maßnahme verbessert werden soll (anwaltschaftliche Evaluation) oder seien es die Maßnahmenverantwortlichen, die Evaluationsergebnisse zur Veränderung nutzen sollen (nutzungsfokussierte Evaluation; Patton 2008).
- *Stakeholderorientierte Evaluationsmodelle*: Während sich die klientenorientierten Evaluationsmodelle jeweils primär in den Dienst einzelner Stakeholdergruppen stellen, verfolgen stakeholderorientierte Modelle das Ziel, alle Beteiligten und Betroffenen gleichermaßen so aktiv wie möglich in den Evaluationsforschungsprozess zu involvieren. Die *Partizipative Evaluation* (participatory evaluation) bezieht die Beteiligten und Betroffenen nicht nur als Informanten, sondern auch als *Forschungspartner* in das Evaluationsprojekt mit ein, etwa durch deren Mitwirkung an der Ausformulierung von Evaluationsfragen oder der Planung des Untersuchungsdesigns. Die *Empowerment-Evaluation* (empowerment evaluation) geht noch einen Schritt weiter als die Partizipative Evaluation: Die Stakeholder werden nicht nur zu Forschungspartnern der Evaluierenden, sondern sollen die Evaluation gänzlich selbst in die Hand nehmen, die Evaluierenden agieren von Anfang an nur beratend. Stakeholderorientierte Evaluationsmodelle sind weniger gebräuchlich als klientenorientierte Modelle. Denn zum einen besteht oft das Problem, dass nicht genügend Maßnahmen-Betroffene zu der zeintensiven und anstrengenden Arbeit an der Evaluationsforschung bereit oder in der Lage sind. Zum anderen wird die wissenschaftliche Qualität von Evaluationsstudien in Frage gestellt, deren Leitung (wie im Falle der Empowerment-Evaluation) ausdrücklich nicht in den Händen von Evaluationsfachleuten liegt.

---

### 9.7 Phasen einer Evaluationsstudie

Generell folgt eine wissenschaftliche Evaluationsstudie einerseits den typischen *Phasen des empirischen Forschungsprozesses* (mit entsprechenden Varianten, je nachdem, ob ein quantitativer, qualitativer oder Mixed-Methods-Forschungsprozess ansteht) und kann andererseits aus Projektmanagement-Perspektive in *Projektphasen* eingeteilt werden. Im Folgenden wird ein Fünf-Phasen-Modell vorgestellt, wobei jeweils die Besonderheiten der Evaluationsforschung herausgestellt werden (Döring/Bortz 2014, Kap. 18).

### 9.7.1 Initierungsphase der Evaluation

In der Initierungsphase geht es aus Sicht der Evaluierenden darum, *potenzielle Evaluationsprojekte zu akquirieren*, indem Evaluationsmöglichkeiten identifiziert (z.B. über öffentliche Ausschreibungen von Evaluationsprojekten oder über Kontakte zu den Auftraggebenden) und deren Evaluationsziele kritisch beleuchtet werden (z.B. hinsichtlich Aufwand, eigener Expertise oder politischer und ethischer Brisanz). Da Evaluation meist Auftragsforschung ist, müssen sich Evaluierende darum bemühen, entsprechende Aufträge zu erhalten. Umgekehrt geht es für die Auftraggebenden darum, für ihr jeweiliges Evaluationsanliegen ein geeignetes Evaluationsteam zu finden.

### 9.7.2 Konzeptionsphase der Evaluation

Wurden mögliche Evaluationsprojekte identifiziert, deren Zielsetzungen machbar und verantwortbar erscheinen, so müssen im nächsten Schritt die Evaluationsbedingungen genauer exploriert und das *Evaluationskonzept* bzw. *Projektangebot* ausgearbeitet werden. Dabei ist es notwendig sicherzustellen, dass die gängigen Evaluationsstandards von Anfang an und über den weiteren Evaluationsprozess hinweg eingehalten, das Projektmanagement vorbereitet (Zusammensetzung des Evaluationsteams, Zeit- und Kostenplan) und verbindliche Vereinbarungen zwischen Evaluierenden und Auftraggebenden getroffen werden. Erhält ein Projektangebot den Zuschlag oder nimmt man einen Projektauftrag an, so folgt der Abschluss des *Evaluationsvertrags* (evaluation contract), der die Rechte und Pflichten von Auftraggebenden und Auftragnehmenden fixiert. Insbesondere bei umfangreichen Projekten sind die rechtlichen Bedingungen genau zu prüfen, etwa die Zahlungsmodalitäten, Möglichkeiten des Vertragsrücktritts oder Konsequenzen bei Nichteinhaltung einzelner vertraglicher Regelungen (für Checklisten zur Vorbereitung von Evaluationsverträgen, Kuster et al. 2006; Stufflebeam/Shinkfield 2007: 547f.).

### 9.7.3 Planungsphase der Evaluation

Nachdem der Evaluationsvertrag abgeschlossen wurde, stehen die *Detailplanung* und konkrete Vorbereitung der Evaluationsstudie an. Dies umfasst die Formulierung von detaillierten Evaluationsfragen bzw. -hypothesen, die Untersuchungsplanung inklusive Untersuchungsdesign, Stichprobenpläne und Instrumente sowie die organisatorische Projektplanung (z.B. Rekrutierung des Evaluationsteams). Im Sinne der Evaluationsstandards der Nützlichkeit und Durchführbarkeit ist in der Planungsphase wiederum ein enger Kontakt zu den Stakeholdern notwendig, ebenso eine enge Rückkopplung mit den Abläufen im Praxisfeld (z.B. Identifikation geeigneter Zeitpunkte und Orte für Datenerhebungen). Andernfalls besteht die Gefahr, dass „am grünen Tisch“ ein Studienentwurf ausgearbeitet wird, der forschungspraktisch nicht umsetzbar ist und/oder am Widerstand derjenigen Stakeholder scheitert, die nicht von Anfang an genügend einbezogen wurden.

### **9.7.4 Realisierungsphase der Evaluation**

Auf der Basis der Ergebnisse der Planungsphase werden in der Realisierungs- oder Durchführungsphase die für die Evaluationsziele notwendigen *Evaluationsdaten erhoben, analysiert und interpretiert*, zudem wird der operative Projektablauf durch ein Projektcontrolling begleitet. Je nach Umfang und Dauer einer Intervention sowie der aktuellen Phase im Interventionslebenszyklus werden Erhebungs-, Analyse- und Interpretationsschritte mehrfach nacheinander oder auch parallel stattfinden, wobei es im Sinne eines *iterativen Vorgehens* oft Rücksprünge in die Planungsphase gibt.

Aus dem Satz der Evaluationsstandards sind bei der Datenerhebung vor allem Standards der Durchführbarkeit und Fairness relevant, während bei der Datenauswertung und Dateninterpretation der Schwerpunkt auf Standards der Genauigkeit liegt (DeGEval 2008).

### **9.7.5 Abschlussphase der Evaluation**

Die in der Realisierungsphase der Evaluation gewonnenen Ergebnisse gilt es in der Abschlussphase an unterschiedliche Stakeholder der Intervention zu *kommunizieren* und ihre *praktische Umsetzung* durch die intendierten Evaluationsnutzenden vorzubereiten. Auch eine wissenschaftliche Bewertung der Evaluation in Form von systematischer *Meta-Evaluation* kann sinnvoll oder notwendig sein (z.B. wenn Auftraggebende oder andere Stakeholder die Qualität der Studie anzweifeln). Schließlich ist es wünschenswert, Evaluationsstudien zu *publizieren*, um sie der Fachcommunity vorzustellen und kritisch diskutieren zu können. Dabei geht es zum einen um kumulativen Erkenntnisgewinn über den jeweiligen Evaluationsgegenstand und zum anderen um die methodische Weiterentwicklung des Feldes der Evaluationsforschung.

## Literatur

- Chen, Huey-tsyh (1990): Theory-Driven Evaluations. London: Sage
- Chen, Huey-tsyh (2005): Practical Program Evaluation: Assessing and Improving Program Planning, Implementation, and Effectiveness. London: Sage
- DeGEval (2008): Standards für Evaluation (4. Aufl.). Köln: DeGEval
- DeGEval (2012): Empfehlungen für die Aus- und Weiterbildung in der Evaluation. Anforderungsprofile an Evaluatorinnen und Evaluatoren (2. Aufl.). Mainz: DeGEval
- Döring, Nicola (2014): Prostitution in Deutschland: Eckdaten und Veränderungen durch das Internet. Zeitschrift für Sexualforschung 2: 99-137
- Döring, Nicola/Bortz, Jürgen (2014): Forschungsmethoden und Evaluation für die Sozial- und Humanwissenschaften (5. vollst. überarb., aktual. und erw. Aufl.). Heidelberg: Springer
- Gollwitzer, Mario/Jäger, Reinhold S. (2009): Evaluation kompakt. Weinheim: Beltz
- Kavemann, Barbara/Steffan, Elfriede (2013): Zehn Jahre Prostitutionsgesetz und die Kontroverse um die Auswirkungen. Aus Politik und Zeitgeschichte 9: 9-15
- Kuster, Jürg/Huber, Eugen/Lippmann, Robert/Schmid, Alphons/Schneider, Emil/Witschi, Urs/Wüst, Roger (2006): Handbuch Projektmanagement (3. Aufl.). Berlin: Springer
- Patton, Michael. Q. (2008): Utilization-Focused Evaluation: The New Century Text (4. Auflage). Thousand Oaks: Sage
- Renzikowski, Joachim (2007): Reglementierung von Prostitution. Ziele und Probleme: Eine kritische Betrachtung des Prostitutionsgesetzes. Berlin: Bundesministerium für Familie, Senioren, Frauen und Jugend
- Scriven, Michael (1967): The methodology of Evaluation. In: Tyler, Ralph W./Gagne, Robert M./Scriven, Michael (Hg.): 39-55
- Tyler, Ralph W./Gagne, Robert M./Scriven, Michael: Perspectives of Curriculum Evaluation. Chicago: Rand McNally.
- SOFFI K (Sozialwissenschaftliches Frauenforschungsinstitut, Freiburg Kontaktstelle praxisorientierte Forschung der Evangelischen Fachhochschule Freiburg) (2007): Untersuchung „Auswirkung des Prostitutionsgesetzes“. Berlin: Bundesministerium für Familie, Senioren, Frauen und Jugend
- Stockmann, Reinhart (2002): Was ist eine gute Evaluation?, Ceval-Arbeitspapiere (Vol. 9). Saarbrücken: Centrum für Evaluation
- Stufflebeam, Daniel L./Shinkfield, Anthony J. (2007): Evaluation Theory, Models, and Applications. San Francisco: Wiley
- Stufflebeam, Daniel L./Foley, Walter J./Gephart, W. J./Guba, Egon G./Hammond, Robert L./Marriman, Howard O. et al. (1971): Educational Evaluation and Decision Making. Itasca: Peacock
- Widmer, Thomas/De Rocchi, Thomas (2012): Evaluation. Grundlagen, Ansätze und Anwendungen. Zürich: Rüegger

Markus Ziegler

---

## 10.1 Einführung

Kommt man im Alltag mit empirischer Sozialforschung in Kontakt, so steckt im Normalfall ein Institut der Markt-, Meinungs- oder Sozialforschung hinter der jeweiligen Studie. Es gibt wohl niemanden, der nicht als aktiver Teilnehmer oder passiver Konsument von Studienergebnissen (etwa in Zeitungen) in irgendeiner Form mit dieser Art der praktischen empirischen Sozialforschung in Berührung gekommen ist.

Im Gegensatz zur akademischen Sozialforschung, die in Universitäten oder universitätsnahen Instituten verankert ist, finanziieren sich Institute der kommerziellen Sozialforschung ausschließlich über Auftragsforschung Dritter und sind entsprechend in der Regel als privatwirtschaftliche Unternehmen organisiert. Ihr Ziel ist es, im Auftrag ihrer Auftraggeber auf Basis empirischer Forschung Informationen zu an ökonomischen Märkten tätigen Akteuren (Marktforschung) bzw. öffentlicher Meinung (Meinungsforschung) zu liefern. Wie dies erfolgen soll, haben die im Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. (ADM) organisierten Unternehmen in einem Kodex formuliert (ADM 2013b), der berufsethische Prinzipien und Verhaltensregeln festlegt (zu allgemeinen Ethikrichtlinien siehe Friedrichs und Mühlichen, Kapitel 3 und 4 in diesem Band). Im Wesentlichen verpflichten sich die Institute zu einem strikt an wissenschaftlichen Prinzipien orientierten Handeln, einer ausschließlichen Nutzung anonymisierter Daten sowie einer klaren Abgrenzung gegenüber forschungsfremden Tätigkeiten wie etwa Werbung oder Direktmarketing.

So gerne in der Öffentlichkeit von einer Einheit von Markt-, Meinungs- und Sozialforschung gesprochen wird, so klar muss man auch festhalten, dass der übergroße Anteil des Umsatzvolumens der kommerziellen Sozialforschung auf die Marktforschung entfällt. Daher soll im folgenden Beitrag der Fokus auf der Marktforschung liegen, wobei die zentralen Aussagen auch auf die kommerzielle Meinungs- oder Sozialforschung übertragen werden können.

## 10.2 Marktforschung in Deutschland

Marktforschung in Deutschland und in der ganzen Welt hat ihren Ursprung in den 1930er-Jahren und ist seit jeher eine große Erfolgsgeschichte. Die Branche kann in dieser Zeit einen durchaus beachtlichen Wachstumsprozess vorweisen, der sich in den vergangenen beiden Jahrzehnten noch einmal beschleunigt hat. So stieg etwa der in Deutschland mit Marktforschungsdienstleistungen erzielte Umsatz seit 2000 um ca. 1,1 Mrd. Euro auf 2,5 Mrd. Euro im Jahr 2012, während sich im gleichen Zeitraum die Anzahl festangestellter Mitarbeiter auf aktuell ca. 20.000 fast verdoppelte. Dabei lässt sich ein Trend hin zu immer größeren Einheiten erkennen, denn in dieser Zeit halbierte sich die Zahl der am Markt tätigen Institute auf etwa 140 (ADM 2013a).

### 10.2.1 Institute als Full-Service-Anbieter mit internationaler Ausrichtung

Die Unternehmen agieren immer stärker als sogenannte Full-Service-Anbieter. Das heißt, dass alle Phasen des Forschungsprojekts aus einer Hand angeboten werden. Dies umfasst in der Regel sowohl qualitative (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) als auch quantitative Forschungsansätze (Stein, Kapitel 7 in diesem Band), die Konzeption und Abwicklung von Studien, Felddienstleistungen, Datenanalysen (Blasius/Baur, Kapitel 79 in diesem Band) und Erstellung von Abschlussberichten und Präsentationen (Friedrichs und Meyer/Meier zu Verl, Kapitel 15 und 16 in diesem Band). Die standardisierte Befragung (Reinecke, Kapitel 44 in diesem Band) ist nach wie vor die mit weitem Abstand wichtigste Erhebungsmethode, neben der telefonischen Befragung (Hüfken, Kapitel 46 in diesem Band) hat in den letzten Jahren vor allem das Online-Interview (Wagner/Hering, Kapitel 48 in diesem Band) an Bedeutung gewonnen.

Die große Bedeutung quantitativer Methoden wird auch beim Umsatz deutlich; etwa 90% des in Deutschland verzeichneten Umsatzvolumens (bezogen auf Mitgliedsinstitute des ADM) entfällt auf diese. Neben traditionellen Querschnittsstudien, auch Ad-hoc-Studien genannt, spielt die Panelforschung (Schupp, Kapitel 73 in diesem Band) eine nicht unbedeutliche Rolle. Bei dieser Art der Erhebung wird das Kaufverhalten bestimmter Produktkategorien oder Dienstleistungen bzw. der Konsum von Medieninhalten bei stets den gleichen Personen über einen längeren Zeitraum gemessen. Das vielleicht bekannteste Beispiel ist das Fernsehpanel der GfK, das im Auftrag der „Arbeitsgemeinschaft Fernsehforschung“ (AGF) die Fernsehquoten für Deutschland ermittelt. Bei diesem Panel werden etwa 5000 Haushalte im gesamten Bundesgebiet mit einem Messgerät ausgestattet, das an das Fernsehgerät angeschlossen wird. Per Fernbedienung meldet jedes Haushaltsmitglied bzw. auch ein möglicher Gastseher sich bei diesem Gerät an, sobald es mit dem Fernsehkonsum beginnt und ihn auch wieder beendet. Die gewonnenen Daten darüber, wer wann welches Programm gesehen hat, werden dann in der Nacht elektronisch übertragen und die daraus errechnete Fernsehquoten können bereits wenige Stunden später

Programmmachern, Werbetreibenden, aber auch interessierten Dritten zur Verfügung gestellt werden.

In den vergangenen 20 Jahren hat auch der Trend zur Internationalisierung immer weiter zugenommen. Das betrifft sowohl die Geschäftstätigkeit der Unternehmen, von denen die wenigsten noch lokal spezialisiert agieren, als auch die Studien selbst. Mehrländerstudien (Braun, Kapitel 56 in diesem Band) sind heute eher die Regel als die Ausnahme; gleichzeitig sind bei der Auswahl der zu untersuchenden Ländern quasi kaum noch Grenzen gesetzt. Wo Unternehmen tätig sind, dort wird auch Marktforschungsinformation benötigt.

### **10.2.2 Die wissenschaftliche Verankerung der Marktforschung**

Ihre wissenschaftliche Verankerung erfährt die Marktforschung (oder präziser die Marketingforschung) als Teil des Marketings in den Wirtschaftswissenschaften. Die inhaltlichen Fragen, die die Marktforschung beantworten soll, decken das gesamte Spektrum des Marketing ab und reichen von der erfolgversprechenden Produktkonzeption über das Bestimmen relevanter Zielgruppen bis hin zu Fragen des Vertriebs oder der Preisgestaltung (Nieschlag et al. 2002: 377ff). Man kann an dieser Auswahl bereits recht leicht die Verbindung zu den vier Säulen des Marketings-Mix, der Preis-, Produkt-, Vertriebs- und Kommunikationspolitik, als Kernbestandteile des Marketings erkennen. Diese enge Bindung zeigt auch, womit sich Marktforschung *nicht* beschäftigt. So wird gelegentlich moniert, dass die Unternehmen die Daten nicht breiter analysieren, um damit zu sozialwissenschaftlich relevanten Ergebnissen zu gelangen. Allerdings ist das eben gerade nicht der Gegenstand und das Ziel von Marktforschung und sie darf nicht mit Disziplinen wie der Konsumsoziologie verwechselt werden.

### **10.2.3 Die empirischen Methoden in der Marktforschung**

Methodisch bedient sich Marktforschung der gängigen Methoden und Instrumente der empirischen Sozialforschung, wobei quantitative Methoden – wie bereits angesprochen – eine weitaus wichtigere Rolle spielen als qualitative. Dies hat verschiedene Gründe. Zum einen werden qualitative Methoden in den Wirtschaftswissenschaften kaum gelehrt, entsprechend schwach ist ihre Integration in der Praxis ausgeprägt. Zum anderen ist aber auch ein quantitativer Forschungsprozess von Studiendesign über Datenerhebung bis hin zum Ergebnis in Bezug auf den anfallenden Aufwand und die damit entstehenden Kosten wesentlich planbarer und entsprechend kalkulierbarer. Vor dem Hintergrund unternehmerischer Risikominimierung von Seiten des Marktforschungsunternehmens ist das ein ganz wesentliches Element.

Das Repertoire der in der Praxis angewandten statistischen Verfahren ist grundsätzlich sehr breit. Jedoch zeigt sich die Branche bei den häufig verwendeten Verfahren überaus bodenständig. Kausalanalytische Fragen werden im Regelfall mittels Regressionen

bearbeitet, vor allem in Form der OLS-Regression. Zur Untersuchung latenter Variablen (Latechva/Davidov und Blasius, Kapitel 55 und 83 in diesem Band) wiederum verwendet man häufig eine Faktorenanalyse (zwar in aller Regel als Hauptkomponentenanalyse, aber die akademische Unterscheidung zwischen Faktoren- und Hauptkomponentenanalyse bleibt in der Marktforschung quasi unberücksichtigt). Daneben erleben in den vergangenen Jahren Verfahren der Psychologie, vor allem die Conjoint-Analyse (Orme 2006), einen enormen Aufschwung. Dabei geht es darum, Entscheidungsprozesse mathematisch-statistisch zu erfassen und abzubilden. Bei den qualitativen Verfahren werden in erster Linie Tiefeninterviews (Helfferich, Kapitel 39 in diesem Band) und Gruppendiskussionen (Vogl, Kapitel 41 in diesem Band) verwendet.

Für alle Verfahren gilt, dass sie meistens mit einem gewissen Maß an Pragmatismus Anwendung finden. Konkret heißt das, dass zu einem gewissen Grad auch die Verletzung statistischer Voraussetzungen oder methodischer Regeln akzeptiert wird, solange man der Meinung ist, damit das inhaltliche Problem des Kunden lösen zu können. Eine in der akademischen Welt übliche Reflexion des methodischen Vorgehens findet normalerweise nur in eingeschränktem Umfang statt, sie wird an vielen Stellen ersetzt oder ergänzt durch ein umfangreiches Reservoir an Erfahrungswissen im Umgang mit Daten und mit der Verwendung der Ergebnisse. Wissenschaftlichen Neuerungen begegnet man in der Regel recht zögerlich, und das aus gutem Grund. Verfahren wie die OLS-Regression, die Hauptkomponentenanalyse oder auch eine Gruppendiskussion haben sich über Jahrzehnte im praktischen Einsatz bewährt. Sie sind wissenschaftlich gut erforscht und man kennt ihre theoretischen wie praktischen Vor- und Nachteile, so dass man sicher im Umgang mit ihnen ist. Zudem können sie auch methodisch nicht geschulten Kunden vermittelt werden, ein ganz wesentlicher Aspekt, da dies bei komplexen methodisch-statistischen Ansätzen häufig nicht möglich ist.

Das heißt nun nicht, dass man nicht auch neue Methoden ausprobiert und testet, aber die Praxistauglichkeit eines Verfahrens ist ohne Zweifel für die Marktforschung das wohl wichtigste Kriterium für dessen Einsatz. Die Verwendung eines methodischen Verfahrens in der Marktforschungspraxis kann u.a. daran scheitern, dass Kunden die entsprechenden Forschungsfragen nicht nachfragen (z.B. Zeitreihenanalysen) (Metz/Thome, Kapitel 84 in diesem Band) bzw. die nötigen Daten nicht vorliegen oder nicht wirtschaftlich zu erheben sind (z.B. Netzwerkanalyse) (Baur, Kapitel 74 in diesem Band).

---

### **10.3 Akademische Sozialforschung und Marktforschung – Zwei Kulturen**

Wirft man einen Blick auf Unterschiede und Gemeinsamkeiten von akademisch-universitäter Forschung und Marktforschung, so fällt auf, dass man es mit zwei völlig verschiedenen Forschungskulturen zu tun hat, ähnlich wie es Reichertz (Kapitel 2 in diesem Band) für das Verhältnis von qualitativer und quantitativer Forschung beschreibt. Jeder Bereich verfolgt seine eigene innere Logik und damit auch eine eigene Sichtweise auf Inhalte und Vorgehensweisen.

### 10.3.1 Forschung als Dienstleistung

Als privatwirtschaftliche Unternehmen stellen Marktforschungsinstitute eine Dienstleistung für andere Unternehmen bereit. Damit müssen sie stets bestrebt sein, die Leistungserbringung so zu gestalten, dass die Erträge größer sind als die verursachten Aufwände. Die Forschungstätigkeit muss sich „rechnen“, um damit die wirtschaftliche Existenz sicherstellen. Sie wird damit zu „einem handelbaren und [...] auch zu einem [von einem Auftraggeber, d.A.] konsumierbaren Gut“ (Schrage 2008: 12). Sie unterscheidet sich in ihrem Kern zunächst einmal nicht von der Leistung eines Friseurs, einer Unternehmensberatung oder einer Autowerkstatt. In allen Fällen bieten im Wettbewerb miteinander stehende Unternehmen eine Dienstleistung für ihre Kunden an. Kommerzielle Forschung stellt hier keine Ausnahme dar. Sie löst die alleinige Orientierung der Forschungstätigkeit an den wissenschaftlichen Prinzipien von „Wahrheit“ und „Nützlichkeit“ und versieht ihr Tun mit einem monetären Wert. Das heißt nicht, dass wissenschaftliche Prinzipien damit über Bord geworfen werden. Sie unterwirft sich aber den Gesetzmäßigkeiten des Handelns an Märkten, die die akademische Forschung nicht kennt.

Der wohl prägendste Unterschied ist die Rolle des Konsumenten des Forschungsergebnisses, der als Auftraggeber oder Kunde völlig anders gelagert ist und auch einen viel größeren Einfluss auf das forscherische Tun hat.

Dieser Kunde kann hierbei die unterschiedlichsten Rollen in seinem Unternehmen einnehmen. Je nach Organisation des Auftraggebers kann dies vom methodisch erfahrenen Betriebsmarktforscher über einen Mitarbeiter der Marketingabteilung bis hin zum Produktmanager reichen. Häufig bündeln diese Personen auch nur die Interessen verschiedener Stakeholder innerhalb der Organisation und nehmen sie im Kontakt mit dem Marktforschungsunternehmen wahr.

Den Kunden ist aber eines gemein: Sie sind in der Regel keine Experten für Marktforschung oder Sozialforschung im Allgemeinen, so dass zwischen dem Marktforscher auf der einen Seite und dem Kunden auf der anderen Seite im Regelfall ein gewisses Kompetenzgefälle vorliegt. Während der Kunde im Normalfall in der Lage ist seine inhaltliche Fragestellung zu formulieren, so kann er nur selten Brauchbarkeit und Qualität der möglichen operativen Umsetzung prüfen. So ungewöhnlich dies für den akademischen Bereich erscheint, wo Forscher und Konsument eines Forschungsergebnisses auf Augenhöhe sowohl inhaltliche als auch methodische Details diskutieren können, so sehr kennen wir dies doch aus unserem alltäglichen Umgang mit angebotenen Dienstleistungen, etwa bei einer Autoreparatur. Die Auswahl des Auftragnehmers erfolgt hier nicht über eine Beurteilung der fachlichen Eignung zur Problemlösung, sondern über Ersatzkriterien wie dessen Reputation, dem angebotenen Preis oder auch einer möglichst schnellen Verfügbarkeit der Ergebnisse. In diesem Zusammenhang hat es sich durchgesetzt, Betriebsabläufe und Qualitätssicherungsmaßnahmen unabhängig prüfen zu lassen. Ein wichtiger Weg für die Unternehmen ist es, sich nach den gültigen DIN-Marktforschungsnormen zertifizieren zu lassen (DIN 2013). Dem Kunden gegenüber, der die fachliche Eignung eines Unternehmens zunächst nicht einschätzen kann, soll somit die Prüfung der forscherischen Kompetenz und die Einhaltung von festgelegten Standards nachgewiesen werden.

Man sollte nicht davon ausgehen, dass der Kunde keinen Einfluss auf die operative Ausführung einer Studie hat, das Gegenteil ist der Fall. Als Auftraggeber gehört dem Kunden nach Beendigung der Studie auch dessen Ergebnis. Damit liegt der Letztentscheid über Forschungsinhalt oder -design immer auch bei diesem, selbst wenn die nötige Fachkenntnis fehlt. Der inhaltliche Kern der Forschung wird dabei zwischen den Parteien ausgetauscht, entsprechend gibt es bei einer Marktforschungsstudie – wie bei jeder anderen Dienstleistung auch – einen wahrgenommenen Preis-Leistungsaspekt, der bei der Beauftragung durch den Kunden berücksichtigt wird. Es gibt folglich keine rein fachliche Entscheidung darüber, ob ein Forschungsvorhaben zu gestalten sei wie in der akademischen Welt, sondern es kommt von Kundenseite zu einer Nutzenabwägung der angebotenen Dienstleistung in Bezug auf den aufgerufenen Preis. Kann oder will der Kunde bestimmte Leistungen nicht oder nur in eingeschränkter Form in Anspruch nehmen – so sehr sie auch die Qualität der Studie erhöhen –, so ist das Marktforschungsunternehmen normalerweise gezwungen, dies zu akzeptieren. Es wird dann einen gewissen Pragmatismus in methodischer oder inhaltlicher Hinsicht an den Tag legen, um die Beziehung zu einem zahlenden Kunden nicht aufs Spiel zu setzen. Bei aller Kundenorientierung wird dabei ein seriöser Marktforscher aber stets bestimmte Mindeststandards einhalten. Denn skrupelloses Umsatzstreben gefährdet die Glaubwürdigkeit am Markt, ein nicht nur in der Marktforschung extrem hohes Gut.

Der Druck auf den Marktforscher von Kundenseite hat in den letzten Jahren sicherlich eher zu- als abgenommen. Nicht nur Kosten-, sondern vor allem auch immer höhere Anforderungen an eine möglichst kurze Bearbeitungsdauer stellen für die Unternehmen große Herausforderungen dar. Dass eine innerhalb weniger Tage zu realisierende Studie auf Basis eines (auf Kundenwunsch) 30-minütigen Telefoninterviews bestimmte qualitative Mängel aufweisen kann, liegt dann natürlich auch auf der Hand. Arbeitsschritte, die hauptsächlich der Qualitätssicherung dienen, müssen in solchen Fällen weitgehend unterbleiben.

Grundsätzlich versucht man aber durch standardisierte Qualitätssicherungsmaßnahmen, etwa einer Kontaktkontrolle bei der Datenerhebung, durch das vorhandene Erfahrungswissen und durch Spezialisierung in der Arbeitsorganisation (siehe dazu Abschnitt 10.3.4) durch Zeit- oder Kostendruck entstehende Mängel aufzufangen.

### **10.3.2 Marktforschungsstudien als Produkt**

Ein weiteres Zeichen des Aushandlungsprozesses zwischen Marktforscher und Kunde ist die Art und Weise, wie Marktforschungsinhalte angeboten werden. Dem Kunden fehlt häufig der Überblick darüber, wie Marktforschung ihm bei seinen Problemen des Marketings helfen kann. Entsprechend aktiv versuchen Institute daher, die Perspektive der Kunden einzunehmen und Lösungen für deren Probleme anzubieten. Die Unternehmen gehen dazu über, als Teil ihres Portfolios auch standardisierte Forschungspakete anzubieten, die man als „Produkte“ vermarktet. Diese Produkte werden dergestalt angeboten, dass sie

unabhängig vom thematischen Kontext auf eine fest definierte Frage mit einem klar vorgegebenen Forschungsansatz ein zuvor festgelegtes Ergebnisformat liefern. Für die Kunden bedeutet das eine enorme Komplexitätsreduktion, weil das konkrete Informationsbedürfnis weniger klar definiert werden muss und man vielmehr – wie am Regal im Supermarkt – das für sich (scheinbar) passende Angebot wählen kann. Das Marktforschungsunternehmen entledigt sich mit dieser Art des Angebots der hohen finanziellen Unsicherheit, die ein individuelles, auf die jeweilige Fragestellung zugeschnittenes Forschungsdesigns mit sich bringt. Es wird bei dieser Art des Studiendesigns das wissenschaftliche Prinzip der Orientierung an einem inhaltlichen Problem quasi umgedreht. Nicht der methodische Ansatz folgt der Fragestellung, sondern beliebiger Inhalt wird mit einem festgesetzten methodischen Ansatz abgearbeitet.

### **10.3.3 Das Ergebnis von Marktforschungsstudien**

Neben der großen Orientierung an den Bedürfnissen der Kunden bei der inhaltlichen Fragestellung hat sich auch beim eigentlichen Forschungsergebnis ein gewisser Marktstandard entwickelt, der sich von der akademischen Welt deutlich unterscheidet. Während die akademische Forschung als Ziel eines Forschungsprozesses die Veröffentlichung der Ergebnisse als Zeitschriftenartikel oder auch als Monographie hat, so hat sich in der Marktforschung die Vorstellung der Ergebnisse als Präsentation, Ergebnisbericht oder auch nur als Tabellenband etabliert. Auch hier gilt die Fokussierung darauf, was der Kunde in welcher Form benötigt oder verarbeiten möchte. Nicht selten führt dies auch dazu, dass Studienergebnisse auf einige wenige Kernergebnisse reduziert werden, die dann als „Management Summary“ zur Entscheidungsfindung in den Unternehmen verwendet werden.

### **10.3.4 Arbeitsteilige Organisation**

Wie bereits mehrfach betont, beinhaltet kommerzielle Marktforschung stets die Vorgabe eines effizienten Umgangs mit vorhandenen Ressourcen und ein entsprechend kostenorientiertes Denken. Daher hat sich im Gegensatz zur akademischen Sozialforschung eine streng arbeitsteilige Arbeitsorganisation durchgesetzt wie sie in der freien Wirtschaft üblich ist. Das Hauptgliederungsprinzip orientiert sich nicht an einzelnen Projekten, sondern an den einzelnen Phasen der Projektabwicklung. Liegen in der akademischen Forschung konzeptionelles Arbeiten und operative Umsetzung bei ein und derselben Person oder Team, so sind diese Bereiche in der Marktforschung voneinander getrennt. Die administrative und inhaltliche Leitung verbleibt in der Regel an einer Stelle, die konkrete Abwicklung einzelner operativer Schritte erfolgt bei spezialisierten Teams oder Abteilungen, die wiederum Experten in ihrem eigenen Feld darstellen und auch als interne Berater der administrativen Ebene fungieren können. Deren Aufgabe ist es, einen definierten Teil des kompletten Projektprozesses zu übernehmen. Beispiele für solche eigenständigen

Einheiten sind die Feldorganisation, qualitative Methodik, quantitative Auswertung oder Datenprozessierung. Projektleiter übernehmen in ihrer Tätigkeit sehr viele administrative und koordinierende Tätigkeiten und sind damit eher Projektmanager. Sie behalten zwar den inhaltlichen Überblick und fungieren als Bindeglied zum Auftraggeber, tiefgreifende Einblicke und ein fundiertes methodisches Verständnis der Tätigkeiten der einzelnen Funktionsteams sind dafür aber nicht zwangsläufig nötig. Die Stärke der Ausdifferenzierung der einzelnen Tätigkeitsfelder bzw. auch das Tätigkeitsprofil des Projektleiters hängt stark von der Größe des Unternehmens ab. Je kleiner ein Marktforschungsunternehmen ist, desto geringer ist die Unterstützung spezieller operativer Arbeitseinheiten.

Es ist wichtig festzuhalten, dass nicht jeder Mitarbeiter eines Marktforschungsunternehmens einer aus akademischer Sicht forschenden Tätigkeit nachgeht und über ein hohes Maß an methodischem Wissen verfügen muss. Durch die arbeitsteilige Organisation sind sehr vielfältige Tätigkeiten zu übernehmen, die mit der eigentlichen Forschung nur mittelbar zusammenhängen. So ist ein Projektleiter mit einem großen Organisationstalent ebenso wenig ausschließlich an die Marktforschung gebunden wie der Vertriebsmitarbeiter, der jederzeit statt standardisierter Marktforschungsprodukte auch andere Produkte oder Dienstleistungen verkaufen kann.

### **10.3.5 Forschungsroutinen und Expertenwissen als Kennzeichen der Marktforschung**

Diese Art der Arbeitsorganisation soll eine möglichst effiziente Bearbeitung einzelner Projektschritte in vergleichbarer Qualität ermöglichen, die sowohl bezüglich zeitlicher Vorgaben als auch eines wettbewerbsfähigen Kostenrahmens den Vorstellungen des Auftraggebers entsprechen. Bei dieser Art der Beschäftigung mit bestimmten Projektschritten wird in hohem Maße methodologisch wertvolles Experten- und Erfahrungswissen aufgebaut. Im Gegensatz zur akademischen Forschung, bei der eine personelle Kontinuität aufgrund gesetzlicher Regelungen weitgehend verhindert wird, liegt es im Interesse eines Marktforschungsunternehmens, Mitarbeiter möglichst langfristig an sich zu binden und deren Erfahrung und Wissen zu nutzen. Diese Tatsache in Verbindung mit der Möglichkeit einer immer wiederkehrenden Beschäftigung mit einem bestimmten operativen Schritt des Forschungsprozesses ermöglicht es den Beschäftigten der einzelnen Abteilungen, sich sehr intensiv mit dem jeweiligen Inhaberbereich bzw. mit dem methodischen Problem zu beschäftigen. Da die Gesamtbearbeitungsdauer einer einzelnen Studie in der Regel deutlich kürzer ist als in der akademischen Forschung, kommen Beschäftigte in der Marktforschung mit sehr viel mehr und häufig auch methodisch heterogeneren Studien und Projekten in Berührung. Universitäres Lehrwissen und theoretische Vorbildung werden der direkten Beobachtung der praktischen Forschungsrealität ausgesetzt und auf ihre Bewährung in der konkreten Forschungssituation hin getestet.

Ein konkretes Beispiel, an dem man dies deutlich sehen kann, ist die Erbringung von Feldleistungen. Für die Durchführung von Interviews benötigen Marktforschungsunternehmen eigene Feldabteilungen, die Interviewer anwerben und schulen und auch für Durchführung und Qualitätskontrollen verantwortlich zeichnen. Neben der nötigen Infrastruktur ist es ein wichtiges Ziel, geeignetes Personal zu finden und möglichst langfristig an sich zu binden.

Ein derartiges Vorgehen kennt die akademische Forschung in der Regel nicht. Da die nötige Infrastruktur nicht durchgängig ausgelastet werden kann, wird häufig auf ihre Anschaffung verzichtet und im Bedarfsfall eine provisorische Lösung, etwa über Studierende, gewählt.

Der akademischen Forschung fehlt damit die Möglichkeit der Marktforschung, Expertenwissen und Forschungsroutinen aufzubauen und sich als lernende Organisation zu entwickeln. Da Personalwechsel in der Regel von der universitären in die kommerzielle Forschung erfolgen und nur in sehr seltenen Fällen in die andere Richtung, können neue, auf Erfahrung basierende Kenntnisse und Routinen auch nicht mehr ihren Weg zurück in die universitäre Welt finden. Zudem haben die Unternehmen leider selten ein eigenes Interesse an einem derartigen Austausch, da dieses Expertenwissen als Vorteil im Wettbewerb mit anderen Unternehmen verwendet werden soll.

### **10.3.6 Gemeinsamkeiten von akademischer und kommerzieller Forschung**

So unterschiedlich die Forschungslogiken akademischer und kommerzieller Sozialforschung aber auch sein mögen, so richtig und wichtig ist es dennoch, dass beide Bereiche durchaus eng miteinander verwoben sind. Dies beginnt bei der Ausbildung des Marktforschungspersonals an den Universitäten, findet seine Fortsetzung bei der Festlegung wissenschaftlicher Standards, die, in den Universitäten gesetzt, dann auch in der kommerziellen Praxis Anwendung finden, bis hin zu ganz konkreten Fällen, in denen die akademische Welt auf von Marktforschungsinstituten erhobene Daten zurückgreift. Dies ist etwa der Fall bei der von der GESIS durchgeführten „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) oder auch bei den zur Re-Analyse bereitgestellten Daten der Media-Analyse. Im ersten Fall wird lediglich die Feldinfrastruktur eines Marktforschers genutzt, im letzteren liegt die komplette Konzeption der Studie in den Händen Dritter. Was in diesen Fällen deutlich wird, ist, dass beide Bereiche mit den gleichen Herausforderungen zu kämpfen haben: Ausschöpfungsquoten gehen immer weiter zurück, die Kooperationsbereitschaft von potentiellen Teilnehmern nimmt stetig ab und Feldzeiten dauern deutlich länger als geplant. Entsprechend schwerer wird es aber auch, an gängigen wissenschaftlichen Vorgehensweisen festzuhalten. Die inneren Logiken von Marktforschung und akademischer Forschung sind sehr unterschiedlich, die Berührungspunkte zeigen aber auch, dass sich beide Bereiche nicht ignorieren können und dürfen.

## 10.4 Ein Blick voraus: Herausforderungen der Marktforschung

Wirft man einen Blick voraus, so wird sich die Marktforschungsbranche in den kommenden Jahren wichtigen Herausforderungen stellen müssen, die bis zu einem bestimmten Grad auch über ihre Bedeutung in der betrieblichen Praxis entscheiden werden. Ein wichtiger Punkt wird sein, die Erreichbarkeit von Befragten (Engel/Schmidt, Kapitel 23 in diesem Band) sicherzustellen und Lösungen für die stetig abnehmende Teilnahmebereitschaft an Marktforschungsstudien zu finden. Dieses Problem ist vielschichtig und berührt verschiedene Ebenen. Eine ist die Tatsache, dass schon heute der bisherige Königsweg der empirischen Erhebung, die telefonische Befragung über das Festnetz (Hüfken, Kapitel 46 in diesem Band), eher schlecht als recht funktioniert. Auch der Siegeszug des Onlineinterviews (Wagner/Hering, Kapitel 48 in diesem Band) löst dieses Problem nicht. Sowohl telefonische als auch Onlineinterviews weisen grundsätzliche Probleme bezüglich der Erreichbarkeit von bestimmten Bevölkerungsgruppen auf, die es zu lösen gilt. Neben der reinen Erreichbarkeit müssen auch neue Wege gefunden werden, Zielpersonen zu einer Teilnahme zu motivieren und sich gleichzeitig klar von der Werbewirtschaft abzugrenzen, um das Vertrauen von potentiellen Befragten nicht zu verspielen.

Inhaltlich wird eine der zentralen Herausforderungen die zunehmende Beschäftigung mit prozessproduzierten Daten sein (Salheiser, Hartmann/Lengerer; Schmitz/Yanenko und Felderer et al., Kapitel 62, 71, 65 und 25 in diesem Band). Kommerzielle wie akademische Sozialforschung sind in ihrem Tun maßgeblich darauf ausgerichtet, den Prozess der Datenerhebung zu gestalten und zu kontrollieren. Dies wird sich insofern ändern, als etwa die Beschäftigung mit sozialen Medien im Internet oder die Analyse von riesigen Transaktionsdaten nicht mehr den klassischen Erhebungsmechanismen unterworfen ist und hier neue Forschungsansätze gefunden werden müssen.

## Literatur

- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. [ADM] (2013a): Marktforschung in Zahlen. URL: <https://www.adm-ev.de/index.php?id=zahlen#c232> am 18.04.2013
- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. [ADM] (2013b): Marktforschung in Zahlen, URL: <https://www.adm-ev.de/index.php?id=kodex&L=1%25252525252527%252529> am 14.07.2013
- Deutsches Institut für Normung e.V. [DIN] (2013): Marktforschung. URL:: <http://www.din.de/cmd?level=tpl-artikel&languageid=de&cmstextid=marktforschung> am 17.09.2013
- Nieschlag, Robert et. al. (2002): Marketing. Berlin: Duncker und Humblot
- Orme, Bryan (2006): Getting started with conjoint analysis. Madison: Research Publishers
- Schrage, Dominik (2008): Zur Soziologie der kommerziellen Konsumforschung. In: Schrage, Dominik/Friederici, Markus (Hg.) (2008): Zwischen Methodenpluralismus und Datenhandel. Wiesbaden: VS Verlag. 11-29

Stefanie Eifler

Das Experiment ist keine besondere Form der Datenerhebung, sondern eine besondere Form der Anordnung und des Ablaufs einer empirischen Untersuchung. Das Ziel des Experiments besteht darin, Erkenntnisse über Ursache-Wirkungs-Beziehungen zu gewinnen und auf dieser Grundlage soziale Phänomene zu erklären. Die Besonderheit des Experiments besteht in der Anwendung des *Prinzips der aktiven Erfahrung*: Die zu untersuchenden Vorgänge werden vom Forscher mit Absicht hergestellt, damit bestimmte Phänomene eindeutig als Wirkungen einer bestimmten Ursache betrachtet werden können. In diesem Zusammenhang kommt der *Kontrolle* von Faktoren, die möglicherweise auch als Ursachen wirken können, eine zentrale Bedeutung zu: Nur wenn Einflüsse einer vermuteten Ursache nachgewiesen und zugleich andere Faktoren als potentielle Ursachen ausgeschlossen werden können, ist es möglich, Erklärungen für soziale Phänomene zuungunsten von Alternativerklärungen zu finden.

Im Folgenden sollen zunächst Begriff und Konzeption des Experiments erläutert werden. Anschließend sollen experimentelle Forschungsdesigns vorgestellt und Möglichkeiten aufgezeigt werden, Experimente in verschiedenen Forschungsumgebungen (Labor und Feld) zu realisieren. Da insbesondere in jüngerer Zeit experimentelle Forschungsstrategien mit den Methoden der Umfrageforschung verbunden wurden, sollen abschließend weiterführende Techniken des Experimentierens in den Sozialwissenschaften umrissen werden.

---

## 11.1 Begriff und Konzeption des Experiments

Begriff und Konzeption des Experiments werden verständlich, wenn man sich das folgende Beispiel vergegenwärtigt: Möchte ein Mediziner die Cholesterin senkende Wirkung von grünem Tee untersuchen, so könnte er im Rahmen einer Feldstudie den Cholesterinspiegel seiner Patienten bestimmen und sie fragen, ob und ggf. in welcher Menge sie grünen

Tee konsumieren. Wenn sich zeigen sollte, dass die Patienten, die regelmäßig grünen Tee trinken, auch diejenigen sind, die einen niedrigen Cholesterinspiegel aufweisen, so könnte er jedoch nicht auf die Cholesterin senkende Wirkung von grünem Tee schließen: Möglich wäre nämlich, dass andere Gründe für einen niedrigen Cholesterinspiegel verantwortlich sind, wie etwa Alter und Geschlecht der Patienten oder ihre Lebensgewohnheiten wie Ernährung und Bewegung. In einem Experiment würde der Mediziner deshalb anders vorgehen. Er würde beispielsweise alle Patienten, die einen hohen Cholesterinspiegel aufweisen, zunächst in zwei Gruppen einteilen. Dabei würde er darauf achten, dass sich die beiden Gruppen hinsichtlich ihres Alters und Geschlechts oder ihrer Lebensgewohnheiten wie Ernährung und Bewegung möglichst weitgehend ähneln. Einer Gruppe von Patienten würde er ein aus grünem Tee gewonnenes Präparat verordnen, der anderen Gruppe von Patienten würde er ein Placebo verordnen. Wenn sich nach einer gewissen Zeit nur in der Gruppe der Patienten, die das echte Medikament eingenommen haben, eine Senkung des Cholesterinspiegels beobachten lässt, nicht aber in der Gruppe der Patienten, die das Placebo eingenommen haben, so kann der Mediziner den Unterschied zwischen beiden Gruppen auf die Wirkung des Medikaments zurückführen. Zugleich kann er ausschließen, dass andere Gründe wie etwa Alter oder Geschlecht der Patienten oder ihre Lebensgewohnheiten wie Ernährung und Bewegung zu dem Unterschied zwischen beiden Gruppen geführt haben.

Das Beispiel verdeutlicht, wie es die Anwendung des Prinzips der aktiven Erfahrung ermöglicht, auf das Vorliegen einer kausalen Beziehung (Kühnel/Dingelstedt, Kapitel 80 in diesem Band) zu schließen: Der Zeitpunkt, zu dem eine vermutete Ursache eintritt – der Zeitpunkt der Verordnung des Medikaments – wird vom Forscher bestimmt; das Eintreten der Ursache – die Verordnung des Medikaments – wird vom Forscher aktiv hergestellt; die Ausprägungen der vermuteten Ursache – Verordnung eines echten Medikaments und eines Placebo – werden vom Forscher systematisch variiert. Außerdem kann der Forscher mögliche Einflüsse von anderen Merkmalen – wie etwa Alter und Geschlecht der Patienten oder ihre Lebensgewohnheiten wie Ernährung und Bewegung – kontrollieren.

Die genannten Aspekte kommen auch in der klassischen Definition des Experiments von Wilhelm Wundt (1832-1920) zum Ausdruck: „Das Experiment besteht in einer Beobachtung, die sich mit der *willkürlichen Einwirkung* (Hervorhebung von mir, StE) des Beobachters auf die Entstehung und den Verlauf der zu beobachteten Erscheinungen verbindet. Die Beobachtung im engeren Sinn untersucht die Erscheinungen ohne derartige Einwirkungen, so wie sie sich in dem Zusammenhang der Erfahrung von selbst dem Beobachter darbieten“ (Wundt [1896]/1922: 25). Wesentliche Elemente dieser Definition hat Wundt später im Rahmen seiner *allgemeinen Regeln der experimentellen Methode* ausgearbeitet: Der Forscher muss den zu untersuchenden Vorgang *willkürlich* in Gang setzen und dabei die Bedingungen, die (vermutlich) zu einer bestimmten Wirkung führen, in geplanter Weise *variiieren* lassen können; der zu untersuchende Vorgang muss *protokollierbar* und beliebig oft *wiederholbar* sein (Wundt 1907: 308).

Das Prinzip der aktiven Erfahrung wurde bereits im 13. Jahrhundert von Roger Bacon (1214/20-1292) im Rahmen seiner „scientia experimentalis“ umrissen. Unter dem Einfluss

von Francis Bacon (1561-1626) hat es im 17. Jahrhundert zunächst Eingang in die Naturwissenschaften gefunden. Dort weist das Experiment anfangs eine vergleichsweise einfache Form auf: Es besteht darin, dass der Forscher absichtlich den zu untersuchenden Gegenstand manipuliert, um eine bestimmte Wirkung zu erzielen. Wird die erwartete Wirkung tatsächlich beobachtet, so wird diese auf die Eingriffe des Forschers ursächlich zurückgeführt. Verformt sich beispielsweise Metall unter dem Einfluss von Hitze, so ist diese Verformung als Wirkung der Erhitzung interpretierbar, weil das Metall unter normalen Bedingungen bestimmte konstante Eigenschaften aufweist, die durch die Erhitzung verändert werden (Schulz 1970). Diese einfache Konzeption des Experiments konnte in den Sozialwissenschaften, die sich seit dem 19. Jahrhundert herausbildeten, nicht ohne weiteres Anwendung finden. Wenn absichtliche Eingriffe in einen bestimmten Gegenstand erfolgen, können beobachtete Veränderungen dieses Gegenstandes nur dann als Wirkungen dieser absichtlichen Eingriffe interpretiert werden, wenn im Vorfeld eine „Normalsituation“ (Schulz 1970: 95) bekannt ist, mit der die veränderte Situation verglichen werden kann. Eine solche Normalsituation ist aber in den Sozialwissenschaften in aller Regel nicht gegeben. Veränderungen eines Gegenstandes können im Allgemeinen nicht nur auf Einflüsse einer experimentellen Einwirkung, sondern auch auf Einflüsse von anderen Faktoren zurückgeführt werden.

Sozialwissenschaftler sind beispielsweise häufig an Einflüssen von Medien auf Einstellungsänderungen interessiert. So könnte ein Forscher der Frage nachgehen, ob Medienberichte über den Umweltschutz zu Einstellungsänderungen führen. Angenommen, der Forscher würde einer Gruppe von Probanden Fernsehberichte über den Umweltschutz präsentieren und vor und nach einer solchen Präsentation die Einstellungen zum Umweltschutz erfassen. Mit einer solchen Anordnung wäre es nicht möglich, eine auftretende Einstellungsänderung eindeutig auf die Präsentation der Fernsehberichte zurückzuführen, vielmehr sind für das Auftreten einer Einstellungsänderung mehrere Alternativerklärungen denkbar. So könnte beispielsweise ein Umweltkandal, der sich zeitgleich ereignet hat, zu der beobachteten Einstellungsänderung geführt haben, auch könnte die wiederholte Erfassung der Einstellungen zum Umweltschutz das Ergebnis beeinflusst haben oder die wenig am Thema Umweltschutz interessierten Personen waren nach der Präsentation der Medienberichte über den Umweltschutz nicht mehr bereit, ihre Einstellungen zum Umweltschutz zu berichten.

Das Beispiel verdeutlicht, dass Einstellungen von Menschen zum Umweltschutz nicht als konstante Eigenschaften aufgefasst werden können, die sich unter „normalen Bedingungen“ nicht verändern würden. Einstellungen sind Merkmale, die sich nicht nur durch eine vom Forscher initiierte Einwirkung verändern, sondern auch aus anderen Gründen einem Wandel unterliegen können. Deshalb musste die einfache Konzeption des Experiments im Kontext sozialwissenschaftlicher Fragestellungen erweitert werden. Eine solche Konzeption des Experiments wurde in der Tradition von Donald T. Campbell (1916-1996) entwickelt, die heute in den Sozialwissenschaften als Standard gilt (Shadish et al. 2002).

## 11.2 Experimentelle Designs und Kontrolltechniken

### 11.2.1 Ursache-Wirkungs-Beziehungen

Die gängige Konzeption des Experiments in den Sozialwissenschaften basiert auf einem allgemeinen Bezugsrahmen für die Analyse kausaler Beziehungen (Shadish et al. 2002). Eine *Ursache* wird dabei nach Mackie (1980: 62) als eine *INUS-Bedingung* bezeichnet. Dieses Akronym steht für „(...) insufficient but non-redundant part of an unnecessary but sufficient condition“. Die Idee der INUS-Bedingung verweist darauf, dass in der Regel auch andere Faktoren als die gerade untersuchten hypothetischen Ursachen zu einer bestimmten Wirkung führen können: So können beispielsweise nicht nur Fernsehberichte über den Umweltschutz zu einer Einstellungsänderung führen, sondern ebenso gut auch Umweltkandale. Häufig sind dabei nicht alle Ursachen, die zu einer Einstellungsänderung führen können, im Vorfeld bekannt: So kann beispielsweise die Geburt des ersten Kindes eine im Vorfeld nicht bekannte oder bedachte Ursache für eine Einstellungsänderung sein. Schließlich bedeutet die Idee der INUS-Bedingung, dass Ursache-Wirkungs-Beziehungen in bestimmten Konstellationen auftreten und nicht deterministisch sind: Fernsehberichte über den Umweltschutz führen beispielsweise nicht immer, sondern nur mit einer bestimmten Wahrscheinlichkeit zu einer Einstellungsänderung, und auch nur dann, wenn sie seriös wirken und aus einer glaubwürdigen Quelle stammen.

Der experimentelle Nachweis eines *Effekts* steht vor dem grundsätzlichen Problem, dass immer unbekannt bleibt, ob bzw. wie sich die Untersuchungsgegenstände verändert hätten, wenn die hypothetische Ursache nicht eingeführt worden wäre. Shadish et al. (2002) beziehen sich hier auf die *Idee des Kontrafaktischen*, die auf David Hume (1711–1776) zurückgeht: Sie betrachten einen Effekt als den Unterschied zwischen der Situation, die durch das experimentelle Einwirken einer hypothetischen Ursache erzeugt wurde, und der Situation, die ohne das experimentelle Einwirken dieser hypothetischen Ursache aus derselben Ausgangssituation entstanden wäre. Im oben genannten Beispiel wäre also ein Effekt der Unterschied zwischen der Einstellung der Probanden, nachdem sie den Fernsehbericht über den Umweltschutz gesehen haben, im Vergleich zu der Einstellung derselben Probanden, nachdem sie den Fernsehbericht über den Umweltschutz nicht gesehen haben. Das Kontrafaktische ist also immer unmöglich. Vor dem Hintergrund dieser Überlegungen stellen Shadish et al. (2002) heraus, dass ein Experiment immer so angelegt sein muss, dass es eine möglichst gute Annäherung an das unmögliche Kontrafaktische darstellt.

Für den Nachweis einer kausalen Beziehung zwischen hypothetischer Ursache und hypothetischer Wirkung werden im Anschluss an die Wissenschaftslehre von John Stuart Mill (1806–1873) drei notwendige Bedingungen angegeben: Damit eine hypothetische Ursache A mit einer hypothetischen Wirkung B kausal verknüpft ist, muss die Ursache A der Wirkung B *zeitlich vorangehen* und mit ihr *kovariieren*; außerdem müssen *alternative Erklärungen* für die Beziehung zwischen der Ursache A und der Wirkung B *ausgeschlossen* werden (Shadish et al. 2002). In einem experimentellen Forschungskontext wird die Wirkung, die erklärt werden soll, als die abhängige Variable (AV) bezeichnet;

die hypothetische Ursache, deren Einfluss auf die abhängige Variable bestimmt werden soll, wird als die unabhängige Variable (UV) bezeichnet; schließlich werden alle Faktoren, die neben der unabhängigen Variablen einen Einfluss auf die abhängige Variable haben und zu Alternativerklärungen der abhängigen Variablen herangezogen werden können, als Störvariablen bezeichnet. Einflüsse von Störvariablen müssen in einem Experiment kontrolliert werden.

### 11.2.2 Interne Validität und vorexperimentelle Designs

Das Ausmaß, in dem es möglich ist, eine eingetretene Veränderung eindeutig auf eine bestimmte Ursache zurückzuführen, wird seit Campbell (1957) als die *interne Validität* eines Experiments bezeichnet. Dabei hängt es von der Anordnung der unabhängigen und abhängigen Variablen, mit anderen Worten dem *Design* der empirischen Untersuchung ab, ob es Faktoren gibt, die die Beziehung zwischen einer unabhängigen und einer abhängigen Variablen beeinflussen (können) und in diesem Sinne Bedrohungen der internen Validität darstellen. Campbell (1957) hat die wichtigsten dieser Faktoren zusammengestellt:

1. Unklare zeitliche Abfolge
2. Einflüsse einer selektiven Auswahl (selection)
3. Einflüsse der Zeit (history)
4. Einflüsse der Reifung (maturation)
5. Regression (regression)
6. Systematische Ausfälle von Untersuchungseinheiten (attrition)
7. Einflüsse der Messung (testing)
8. Einflüsse veränderter Messinstrumente (instrumentation)

Worin die Bedeutung dieser Faktoren besteht, kann am besten erläutert werden, wenn man Designs betrachtet, die als nicht intern valide gelten: Insbesondere die sogenannten *vorexperimentellen Designs* (Campbell/Stanley 1963, Cook/Campbell 1976, 1979) sind Anordnungen, in denen Einflüsse von Störfaktoren nicht angemessen kontrolliert werden können; hierzu gehören die *einmalige Untersuchung eines Individuums oder einer Gruppe*, die *Nachher-Untersuchung ohne Kontrollgruppe* und die *Vorher-Nachher-Untersuchung ohne Kontrollgruppe*. Eine *einmalige Untersuchung* ist z.B. eine Studie, bei der zu einem Erhebungszeitpunkt Einstellungen zum Umweltschutz und umweltrelevantes Verhalten erhoben werden. Anhand eines solchen Designs ist die zeitliche Abfolge der unabhängigen und abhängigen Variable nicht eindeutig, es kann also nicht geklärt werden, ob die Einstellungen zum Umweltschutz zu umweltrelevantem Verhalten geführt haben, oder ob umgekehrt das umweltrelevante Verhalten die Einstellungen zum Umweltschutz bewirkt hat (zeitliche Abfolge). Diese Anordnung wird oft in der Umfrageforschung gewählt und als *Ex-Post-Facto-Design* bezeichnet.

Bei einer *Nachher-Untersuchung ohne Kontrollgruppe* wird im Unterschied zu einer einmaligen Untersuchung beispielsweise der Einfluss von Medien auf Einstellungsänderungen

untersucht. Wird etwa einer Gruppe von Probanden ein Fernsehbericht über den Umweltschutz präsentiert, kann aufgrund einer einmaligen Messung keine Einstellungsänderung nachgewiesen werden, weil die Einstellungen, die vor der Präsentation des Fernsehberichts über den Umweltschutz bestanden haben, unbekannt bleiben. Diese Situation wird durch eine *Vorher-Nachher-Untersuchung ohne Kontrollgruppe* nicht wesentlich verbessert: Dabei wird eine Gruppe von Probanden vor und nach einer experimentellen Einwirkung – einem sogenannten *Treatment* – untersucht. Beispielsweise werden an einer Gruppe von Probanden Einstellungen zum Umweltschutz vor und nach der Präsentation eines Fernsehberichts über den Umweltschutz erfasst. Für das Auftreten einer Einstellungsänderung können in einem solchen Kontext gleich mehrere Alternativerklärungen herangezogen werden: Das Ergebnis der Studie kann darauf zurückzuführen sein, dass nur Personen an der Untersuchung teilgenommen haben, die am Thema Umweltschutz besonders interessiert sind (*selection*). Möglicherweise wurden auch Probanden untersucht, die extrem positive oder negative Einstellungen zum Umweltschutz aufweisen. Bei diesen Personen wäre aufgrund eines Effekts, den man als die *Regression zur Mitte* bezeichnet, auch ohne experimentelles Treatment eine Tendenz dahingehend zu erwarten, dass sich die extrem positiven oder negativen Einstellungen eher in Richtung moderaterer Einstellungen entwickeln (*regression*). Zudem kann das Ergebnis der Nachher-Untersuchung von der Vorher-Untersuchung beeinflusst worden sein (*testing*), Veränderungen zwischen den beiden Untersuchungen können auf veränderte Messinstrumente zurückgehen (*instrumentation*), oder aber auf Einflüsse anderer, zeitgleich wirksamer Faktoren wie etwa einen Umweltkandal (*history*) und/oder auf Reifungsvorgänge (*maturity*) zurückzuführen sein; schließlich kann es aufgrund mangelnder Motivation der Probanden selektive Ausfälle von Untersuchungseinheiten zum zweiten Untersuchungszeitpunkt geben, etwa weil die Probanden am Thema Umweltschutz nur mäßig interessiert waren (*attrition*).

Vor diesem Hintergrund war es für die Konzeption des sozialwissenschaftlichen Experiments von zentraler Bedeutung, die Einschränkungen der vorexperimentellen Designs zu überwinden und Anordnungen zu entwickeln, die eine Kontrolle der Faktoren ermöglichen, die die interne Validität beeinträchtigen. Einen sehr wichtigen Schritt in diese Richtung stellte die Entwicklung des *Parallelgruppenmodells* dar: Dabei wird eine Gruppe einer experimentellen Einwirkung ausgesetzt und eine andere Gruppe – die sogenannte Parallelgruppe, nicht. Erstmals wurde diese Anordnung in einer sozialpsychologischen Studie zur Analyse des sozialen Einflusses von Triplett (1897) realisiert: Ausgehend von der Beobachtung, dass Radrennfahrer in Begleitung eines weiteren Fahrers schneller fahren als alleine, hat Triplett die Frage untersucht, ob die Anwesenheit von anderen in einer Wettbewerbssituation zu einer Steigerung der Leistung führt. Hierzu hat er in einer eigens für die Zwecke der Studie hergerichteten Umgebung – einem Labor – eine Wettbewerbssituation hergestellt. Die Aufgabe der Probanden bestand darin, an einer Apparatur Garnrollen aufzuwickeln, wobei die Anwesenheit von Anderen absichtlich hergestellt und in geplanter Weise eingeführt wurde: Triplett teilte die Probanden in eine *Experimental-* und eine *Kontrollgruppe* und ließ sie in jeweils umgekehrter Folge abwechselnd alleine oder im Beisein von anderen Garnrollen aufwickeln. Auf diese Weise konnten Leistungssteigerungen in

Abhängigkeit von der Anwesenheit anderer Personen betrachtet und zugleich Einflüsse der Übung kontrolliert werden.

Ein weiterer sehr wichtiger Schritt bestand in der Entwicklung der so genannten *experimentellen Kontrolltechniken* und ihrer Etablierung als Standardverfahren im Rahmen der Konzeption des Experiments in den Sozialwissenschaften.

### 11.2.3 Kontrolltechniken

Bei den Kontrolltechniken lassen sich Techniken, die auf der Ebene der experimentellen Situation ansetzen, von Techniken, die auf der Ebene der untersuchten Personen ansetzen, voneinander unterscheiden (Huber 2013).

In der experimentellen Situation bestehen die Techniken der *Elimination* und *Konstanthaltung* darin, dass mögliche Störfaktoren vollständig ausgeschaltet bzw. eliminiert und in Experimental- und Kontrollgruppe eine jeweils gleichartige Wirkung entfalten, mit anderen Worten konstant gehalten werden. Bezogen auf das Beispiel der Einflüsse von Medien auf Einstellungsänderungen würde Elimination bedeuten, dass man während der Präsentation des Fernsehberichts über den Umweltschutz Störungen durch Lärm ausschaltet; Konstanthaltung würde bedeuten, dass die Umgebung, in der die Fernsehberichte präsentiert werden, in Experimental- und Kontrollgruppe gleich ist.

Auf der Ebene der untersuchten Personen werden die Techniken der *Randomisierung* und *Parallelisierung* eingesetzt. Dabei wurde das Prinzip der Randomisierung von Sir Ronald A. Fisher (1890–1962) entwickelt. Es besteht darin, dass die Zuweisung von Probanden zu einer Experimental- und einer Kontrollgruppe nach einem Zufallsverfahren – beispielsweise durch ein Losverfahren oder durch Münzwurf – erfolgt. Das Prinzip der Parallelisierung wurde von Chapin (1938) entwickelt und besteht darin, dass Experimental- und Kontrollgruppe hinsichtlich relevanter Merkmalsausprägungen angeglichen werden, z.B. hinsichtlich Geschlecht und Alter(sgruppe). Dies kann entweder als paarweise Gleichsetzung von Probanden in Experimental- und Kontrollgruppe (*matching*) oder aber als Gleichverteilung relevanter Merkmale in Experimental- und Kontrollgruppe (*factor equation*) erfolgen. Ein fiktives Beispiel für die beiden Formen der Parallelisierung findet sich bei Zimmermann (1972: 68). Sowohl das Parallelgruppenmodell als auch die Anwendung der Kontrolltechniken, insbesondere des Prinzips der Randomisierung, sind kennzeichnend für die sogenannten experimentellen Designs.

### 11.2.4 Experimentelle Designs

Die beiden wichtigsten experimentellen Designs sind die Vorher-Nachher-Untersuchung mit Kontrollgruppe und der auf Solomon (1949) zurück gehende Viergruppen-Plan.

#### 11.2.4.1 Vorher-Nachher-Untersuchung mit Kontrollgruppe

Das Design einer Vorher-Nachher-Untersuchung mit Kontrollgruppe, bei der die Probanden zufällig der Experimental- oder Kontrollgruppe zugewiesen werden, ist zur Kontrolle eines großen Teils der potentiellen Störfaktoren geeignet. Abbildung 11.1 zeigt in Anlehnung an Shadish et al. (2002) das Design einer Vorher-Nachher-Untersuchung mit Kontrollgruppe. O steht jeweils für eine Gruppe von Untersuchungseinheiten, X für die experimentelle Einwirkung einer hypothetischen Ursache (treatment) und R für Randomisierung.

Als Beispiel kann erneut die Analyse der Einflüsse von Medien auf Einstellungsänderungen herangezogen werden. In einer Studie, in der Einflüsse von Fernsehberichten über den Umweltschutz auf Einstellungsänderungen untersucht werden sollen, werden zunächst geeignete Probanden zufällig auf die Experimental- und Kontrollgruppe verteilt. Der Experimentalgruppe werden Fernsehberichte über den Umweltschutz präsentiert (treatment), der Kontrollgruppe nicht; Eine Abwandlung könnte darin bestehen, dass in der Kontrollgruppe nicht auf ein Treatment verzichtet wird, sondern ein anderes Treatment erfolgt, beispielsweise könnte ein Fernsehbericht zu einem anderen Thema vorgeführt werden. Wenn zu einem späteren Zeitpunkt Einstellungsänderungen nur in der Experimentalgruppe aufgetreten sind, so kann die Differenz zwischen der ursprünglichen und der späteren Einstellung auf das Treatment zurückgeführt werden. Die Einflüsse zahlreicher Störfaktoren kann kontrolliert werden: Die Kontrollgruppe ist ebenso wie die Experimentalgruppe zwischenzeitlichem Geschehen ausgesetzt, so dass dieser Faktor keinen Unterschied zwischen der Experimental- und der Kontrollgruppe herbeiführen kann (history); es kann angenommen werden, dass Experimental- und Kontrollgruppe vergleichbaren biologischen und psychologischen Veränderungen unterliegen (maturation); sofern in Experimental- und Kontrollgruppe identische Erhebungsinstrumente verwendet werden, sind diesbezügliche Einflüsse kontrolliert (instrumentation). Durch die Anwendung von Kontrolltechniken (Randomisierung und Matching) können außerdem Einflüsse einer selektiven Auswahl vermieden werden (selection). Das Vorhandensein der Kontrollgruppe erlaubt es schließlich, Effekte von selektiven Ausfällen zu bestimmen und statistisch zu kontrollieren (attrition).

Da die Vorher-Nachher-Untersuchung mit Kontrollgruppe viele Alternativerklärungen ausschließt, wird sie als intern valide bewertet. Trotz der genannten Vorteile dieser Anordnung kann mit diesem Design ein Einfluss der Vorher-Untersuchung auf die Nachher-Untersuchung nicht ausgeschlossen werden (testing). Bezogen auf die Analyse von Einstellungsänderungen hieße das, dass die erste Einstellungsmessung einen Einfluss auf die zweite Einstellungsmessung hätte. Sofern dieser Einfluss in Experimental- und Kontrollgruppe gleich wäre, läge keine Beeinträchtigung der internen Validität vor. Möglich wäre es aber, dass sich der Einfluss des Treatments auf die Einstellungen aufgrund der Vorher-Messung von dem tatsächlichen Einfluss, der ohne diese Vorher-Messung hätte beobachtet werden können, unterscheidet.

**Abb. 11.1** Design einer Vorher-Nachher-Untersuchung mit Kontrollgruppe

R	O	X	O
R	O		O

#### 11.2.4.2 Solomon-Viergruppen-Plan

Vor dem Hintergrund dieser Überlegungen hat zuerst Solomon (1949) eine Erweiterung der Vorher-Nachher-Untersuchung mit Kontrollgruppe um jeweils eine weitere Experimental- und Kontrollgruppe vorgeschlagen, bei denen keine Vorher-Messung erfolgt (Abb. 11.2 in Anlehnung an Shadish et al. 2002).

Der Solomon-Viergruppen-Plan ist die einzige experimentelle Anordnung, die es ermöglicht, alle Störfaktoren, die eine Beeinträchtigung der internen Validität darstellen könnten, zu kontrollieren. Bezogen auf die Analyse der Einflüsse von Medien auf Einstellungsänderungen werden zunächst geeignete Probanden zufällig auf die vier Gruppen verteilt: In zwei Gruppen wird ein Fernsehbericht über den Umweltschutz präsentiert, wobei in einer Gruppe die Einstellungen vor und nach der Präsentation des Treatments erhoben werden, in der anderen Gruppe nur nach der Präsentation des Treatments; in zwei Gruppen wird kein Fernsehbericht über den Umweltschutz präsentiert, und die Einstellungen werden wiederum in einer Gruppe vor und nach der Präsentation des Treatments, in der anderen Gruppe nur nach der Präsentation des Treatments erhoben. Zusätzlich zu den Kontrollmöglichkeiten, die bereits im Zusammenhang mit der Vorher-Nachher-Untersuchung mit Kontrollgruppe erläutert wurden, ist es im Solomon-Viergruppen-Plan möglich, Einflüsse der Vorher-Messung zu kontrollieren (*testing*), indem die Differenz zwischen den nachher gemessenen Einstellungen zwischen den beiden Gruppen, denen das Treatment präsentiert wurde und die sich nur durch das Vorhandensein oder Nichtvorhandensein einer Vorher-Messung unterscheiden, berücksichtigt wird.

#### 11.2.4.3 Quasi-Experimentelle Designs

Anders als der Solomon-Viergruppenplan ist die Gruppe der sogenannten *quasi-experimentellen Designs* mit Einschränkungen der internen Validität verbunden. Dies sind experimentelle Designs, bei denen keine Randomisierung der Probanden auf die Experimentalgruppen oder Kontrollgruppen erfolgt. Häufig werden quasi-experimentelle Designs in Studien eingesetzt, die sich mit der Wirksamkeit sozialer Interventionsprogramme beschäftigen. Dies ist beispielsweise der Fall, wenn eine Behörde eine Informationskampagne zum Umweltschutz initiiert und wissen möchte, ob diese Maßnahme zu Einstellungsänderungen in der Bevölkerung geführt hat. In einem solchen Kontext ist es häufig nicht möglich, Bürgerinnen und Bürger zufällig einer Experimental- oder einer Kontrollgruppe zuzuweisen. Die Kontrolle darüber, wer zu welchem Zeitpunkt einer experimentellen Einwirkung ausgesetzt wird, liegt im Rahmen quasi-experimenteller Designs nicht in der Hand des Forschers (Shadish et al. 2002).

**Abb. 11.2** Solomon-Viergruppen-Anordnung

R	O	X	O
R	O		O
R		X	O
R			O

**Abb. 11.3** Mehrfaktorielles Designs mit Kontrollgruppe

R	X <sub>A1B1</sub>	O
R	X <sub>A1B2</sub>	O
R	X <sub>A2B1</sub>	O
R	X <sub>A2B2</sub>	O

#### 11.2.4.4 Faktorielle Designs

Eine Erweiterungsmöglichkeit der experimentellen und quasi-experimentellen Designs besteht im Hinblick auf die Anzahl der untersuchten hypothetischen Ursachen. Sollen mehrere hypothetische Ursachen zugleich untersucht werden, so können diese in ein so genanntes *faktorielles Design* einbezogen werden. Abbildung 11.3 zeigt in Anlehnung an Shadish et al. (2002) ein zweifaktorielles Design, bei dem zwei Faktoren mit jeweils zwei Ausprägungen untersucht werden.

Mit einem faktoriellen Design könnte beispielsweise untersucht werden, welchen Einfluss die Präsentation des Fernsehberichts zum Umweltschutz hat, wenn sie aus verschiedenen Quellen stammt (A), einer glaubwürdigen Quelle (A1) und einer weniger glaubwürdigen Quelle (A2), und wenn der Fernsehbericht in verschiedenen Formen präsentiert wird (B), als positive Darstellung des Umweltschutzes (B1) und als skeptische Darstellung des Umweltschutzes (B2).

#### 11.2.5 Externe Validität

Im Kontext der Bewertung experimenteller Designs wird dem Kriterium der internen Validität eine zentrale Bedeutung beigemessen. Campbell und Stanley (1963) sprechen von der „*sine qua non*“ (Campbell/Stanley 1963: 175). Dem nachgeordnet ist das Kriterium der *externen Validität*, das seit Campbell (1957) als das Ausmaß betrachtet wird, in dem experimentell erzielte Befunde auf andere Gruppen, Umgebungen und Situationen übertragen

werden können. Die externe Validität kann beeinträchtigt werden, wenn wechselseitige Beziehungen zwischen dem untersuchten kausalen Vorgang und ...

1. den Untersuchungseinheiten (units)
2. der Variation der experimentellen Einwirkung (treatment variation)
3. der Art der abhängigen Variablen (outcomes)
4. den Umgebungen (settings)
5. kontextabhängig wirkenden Effekten (context dependent mediation)

bestehen. So ist beispielsweise die externe Validität einer experimentellen Untersuchung der Einflüsse von Medien auf Einstellungsänderungen eingeschränkt, wenn Effekte nur für bestimmte Gruppen von Probanden wie etwa hoch gebildete Personen nachgewiesen werden können (units), nur durch bestimmte Treatments wie etwa allein durch Fernsehberichte, nicht aber durch Zeitungsartikel erzeugt werden (treatment variation), nur hinsichtlich bestimmter abhängiger Variablen wie etwa Einstellungen zur Mülltrennung und nicht Einstellungen zum Verzicht auf einen PKW auftreten (outcomes) und nur in bestimmten räumlichen Umgebungen wie beispielsweise nur in der Stadt, nicht aber auf dem Land beobachtet werden können (settings).

Mit gewissen Einschränkungen können intern und extern valide experimentelle Designs in unterschiedlichen Forschungsumgebungen und mit unterschiedlichen Verfahren der Datenerhebung realisiert werden, wie im Folgenden gezeigt wird.

---

### 11.3 Labor- und Feldexperiment

Experimentelle Designs werden hauptsächlich im Rahmen von Labor- oder Feldexperimenten realisiert. *Laborexperimente* finden dabei in Räumen statt, die eigens für die Durchführung von Experimenten eingerichtet bzw. ausgestattet wurden. *Feldexperimente* finden dagegen in der natürlichen Umgebung der Probanden statt. Dabei wird der Begriff des Feldes vergleichsweise weit definiert und meint jede Umgebung, die nicht eigens für die Durchführung einer empirischen Untersuchung hergerichtet wurde. Anders als in Feldstudien findet im Rahmen von Feldexperimenten ein aktives Eingreifen des Forschers in eine reale Situation mit dem Ziel der Analyse einer kausalen Beziehung statt.

Ein klassisches Laborexperiment ist die bereits berichtete Studie von Triplett (1897), der den Einfluss der Anwesenheit von Anderen in einer Wettbewerbssituation untersucht hat. Als Aufgabe mussten die Probanden in einer Experimentalgruppe im Beisein von Anderen und in einer Kontrollgruppe nicht im Beisein von Anderen Garnrollen aufwickeln, wobei der Faktor Übung kontrolliert wurde. Es zeigten sich die von Triplett (1897) erwarteten Leistungssteigerungen bei Anwesenheit von Anderen. Am Beispiel des Experiments von Triplett (1897) können die Vor- und Nachteile von Laborexperimenten verdeutlicht werden: Der Einsatz von experimentellen Kontrolltechniken ist vergleichsweise leicht möglich

und trägt zu einer hohen internen Validität des Experiments bei. Demgegenüber gibt es zwei Aspekte, die als Beeinträchtigungen der externen Validität gelten: Zum einen wird die relative Künstlichkeit und Alltagsferne von Laborexperimenten bemängelt, die deshalb keine Aussagen über das Verhalten von Menschen in natürlichen Situationen erlaubten. Dies trifft für das Aufwickeln von Garnrollen, wie es im Triplett-Experiment untersucht wird, sicherlich zu. Zum anderen zeigen die Ergebnisse der *Artefaktforschung*, dass experimentelle Einwirkungen im Labor vielfach andere Wirkungen erzielen als die, die ein Versuchsleiter zu erzielen glaubte bzw. beabsichtigte (Patry 1982).

Ein klassisches Feldexperiment ist die Studie von Lefkowitz et al. (1955), in der Nachahmungseffekte im Rahmen von Normverletzungen im Straßenverkehr – Überqueren einer Fußgängerampel bei rotem Signal – untersucht wurden. Das Experiment basierte auf einem zweifaktoriellen Design, bei dem Merkmale eines Modells – einer Person, die ein bestimmtes Verhalten vormacht – systematisch variiert wurden, und zwar zeigte das Modell a) entweder konformes oder abweichendes Verhalten und war b) anhand seiner Kleidung entweder als Status hoch oder Status niedrig zu erkennen. An drei Fußgängerampeln wurde zu verschiedenen Tageszeiten beobachtet, wie häufig Fußgänger unter den vier verschiedenen Bedingungen die Ampel bei rotem Signal überqueren. Es zeigte sich, dass Normverletzungen häufiger in Anwesenheit eines abweichenden Modells auftraten, und zwar insbesondere dann, wenn dieses Modell aufgrund seiner Kleidung als Status hoch erkennbar war. Am Beispiel der Studie von Lefkowitz et al. (1955) können Vor- und Nachteile eines Feldexperiments beleuchtet werden: Die Tatsache, dass Feldexperimente in natürlichen Umgebungen stattfinden, wird im Allgemeinen als ein Vorteil betrachtet. Dabei wird implizit angenommen, dass Feldexperimente deshalb extern valider sind als Laborexperimente. Diese Ansicht ist jedoch nur selten empirisch begründet. Ein wesentlicher Vorteil von Feldexperimenten besteht unabhängig davon aber darin, dass sie das Problem der Reaktivität vermeiden und das Verhalten von Probanden in natürlichen Umgebungen und ohne deren Wissen, Gegenstand einer wissenschaftlichen Untersuchung zu sein, analysieren. Allerdings ist die Anwendung der experimentellen Kontrolltechniken im Feldexperiment vergleichsweise schwieriger, so dass Beeinträchtigungen der internen Validität entstehen können. Die Probanden in der Studie von Lefkowitz et al. (1955) konnten beispielsweise nicht zufällig auf die unterschiedlichen Bedingungen des Experiments verteilt werden. Im Prinzip ist aber die Anwendung von Kontrolltechniken – die Randomisierung eingeschlossen – auch im Feldexperiment möglich. Häufig werden Labor- und Feldexperimente in der Literatur hinsichtlich ihrer internen und externen Validität verglichen. Dabei wird im Allgemeinen das Laborexperiment bevorzugt, wenn es um den Ausschluss möglichst vieler Alternativerklärungen geht, während das Feldexperiment bevorzugt wird, wenn der Aspekt der Verallgemeinerbarkeit im Vordergrund steht.

Mit Feldexperimenten werden häufig die sogenannten *natürlichen Experimente* verwechselt. Natürliche Experimente sind jedoch keine Experimente im eigentlichen Sinne. Dabei werden nämlich natürliche Situationen, in denen ungeplante Veränderungen entstehen, in der Weise eines Experiments – d.h. mit Bezug auf eine Vergleichssituation – untersucht. Natürliche Experimente werden vergleichsweise selten durchgeführt. Eine

Ausnahme bildet die Studie von Mummendey (1986), in der Einflüsse von Verhaltensänderungen auf Einstellungs- und Selbstkonzeptänderungen untersucht wurden. Dabei wurden die Geburt des ersten Kindes und der Eintritt in das Militär im Rahmen des Wehrdienstes als natürliche Treatments betrachtet, die zu einer Einstellungs- und Selbstkonzeptänderung führen sollten. Die Veränderung der alltäglichen Aktivitäten wurde erfasst und auf Veränderungen in Bezug auf konservative Einstellungen, Wertorientierungen und Einstellungen zur eigenen Person betrachtet. Die Ergebnisse der Studie weisen darauf hin, dass es neben dem erwarteten Effekt von Verhalten auf Einstellungen auch Beziehungen in der entgegengesetzten Richtung gibt. In einem natürlichen Experiment ist der Ausschluss von Alternativerklärungen nur schwer möglich, wie auch Mummendey (1986) kritisch diskutiert. Die interne Validität natürlicher Experimente ist daher als gering einzustufen.

---

## 11.4 Weiterführende Techniken

In jüngerer Zeit wird die Strategie, Techniken der Umfrageforschung mit experimentellen Forschungsstrategien zu verbinden, intensiv verfolgt. Hierzu gehören vor allem die Technik der gegabelten Befragung, die Vignettentechnik und die Randomized-Response-Technik.

Im Mittelpunkt gegabelter Befragungen stehen Themen der Methodenforschung wie beispielsweise die Analyse der Einflüsse verschiedener Antwortformate. Bei einer gegabelten Befragung wird die Untersuchungsstichprobe zufällig in mindestens zwei Gruppen geteilt, die im Rahmen einer Umfrage systematisch mit unterschiedlichen Stimuli konfrontiert werden (Petersen 2002).

In der Umfrageforschung werden häufig Techniken verwendet, bei denen *Vignetten* – das sind kurze verbale oder visuelle Darstellungen von hypothetischen Personen oder Situationen – eingesetzt werden. Die Vignetten werden experimentell variiert und von den Befragten beurteilt (Mutz 2011, Dülmer, Kapitel 53 in diesem Band). Die Vignettentechnik wird häufig im Rahmen des sogenannten *Factorial Survey Approach* in der Tradition von Peter H. Rossi (1921–2006) eingesetzt (Rossi 1979). Bei der Randomized-Response-Technik (RRT) sollen Antworten von Befragten zu heiklen Inhalten valide erhoben werden. Dabei wird den Befragten zufällig eine heikle oder eine belanglose Frage vorgelegt. Das Vorgehen wird bei Schnell et al. (2013) beschrieben.

## Literatur

- Campbell, Donald T. (1957): Factors Relevant to the Validity of Experiments in Social Settings. *Psychological Bulletin* 54: 297-312
- Campbell, Donald T./Stanley, Julian C. (1963): Experimental and Quasi-Experimental Designs for Research on Teaching. In: Gage (Hg.): 171-246
- Chapin, F. S. (1938): Design for Social Experiments. In: *American Sociological Review* 3: 786-800
- Cook, Thomas D./Campbell, Donald T. (1976): The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings. In: Dunnette (Hg.): 223-326
- Cook, Thomas D./Campbell, Donald T. (1979): Quasi-Experimentation. Design & Analysis Issues for Field Settings. Boston: Houghton Mifflin Company
- Dunnette, Marvin D. (Hrsg.) (1976): *Handbook of Industrial and Organizational Psychology*. Chicago: Rand McNally & Company
- Gage, Nathaniel L. (Hrsg.) (1963): *Handbook of Research on Teaching*. Chicago: Rand McNally & Company
- Huber, Oswald (2013): Das psychologische Experiment: Eine Einführung (6., überarb. Aufl.). Bern: Huber
- Lefkowitz, Monroe/Blake, Robert R./Mouton, Jane S. (1955): Status Factors in Pedestrian Violation of Traffic Signals. In: *Journal of Abnormal and Social Psychology* 51: 704-706
- Mackie, John L. (1980): *The Cement of the Universe. A Study in Causation*. Cambridge: Clarendon
- Merton, Robert K./Coleman, James. S./Rossi, Peter H. (Hg.) (1979): Qualitative and Quantitative Social Research. Papers in Honor of Paul F. Lazarsfeld. New York: Free Press
- Mummendey, Hans Dieter (1986): Attitude and Self-Concept Change Following Behavior Change: A Research Project on Mothers' and Soldiers' Behavior and Attitudes. In: *German Journal of Psychology* 10: 157-177
- Mutz, Diana C. (2011): Population-Based Survey Experiments. Princeton, NJ: Princeton University Press
- Patry, Jean-Luc (1982): *Feldforschung. Methoden und Probleme sozialwissenschaftlicher Forschung unter natürlichen Bedingungen*. Bern: Huber
- Petersen, Thomas (2002): Das Feldexperiment in der Umfrageforschung. Frankfurt a. M.: Campus
- Rossi, Peter H. (1979): Vignette Analysis. Uncovering the Normative Structure of Complex Judgments. In: Merton/Coleman/Rossi (Hg.)
- Schnell, Rainer/Hill, Paul B./Esser, Elke (2013): Methoden der empirischen Sozialforschung (10., überarb. Aufl.). München: Oldenbourg
- Schulz, Winfried (1970): Kausalität und Experiment in den Sozialwissenschaften. Methodologie und Forschungstechnik. Mainz: v. Hase & Koehler
- Shadish, William R./Cook, Thomas D./Campbell, Donald T. (2002): Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin Company

- Solomon, Richard L. (1949): An Extension of Control Group Design. In: Psychological Bulletin 46: 137-150
- Triplett, Norman (1897): The Dynamogenic Factors in Pacemaking and Competition. In: American Journal of Psychology 4: 507-533
- Wundt, Wilhelm (1907): Über Ausfragemethoden und über die Methoden zur Psychologie des Denkens. In: Psychologische Studien 3: 301-360
- Wundt, Wilhelm (1922/1896): Grundriß der Psychologie (15. Aufl.). Kröner: Leipzig
- Zimmermann, Ekkart (1972): Das Experiment in den Sozialwissenschaften. Stuttgart: Teubner

Ben Jann und Debra Hevenstone

In der empirischen Sozialforschung befasst man sich mit gesellschaftlichen Phänomenen. Es geht darum, wie soziale Akteure durch ihren gesellschaftlichen Kontext beeinflusst werden und wie die Gesellschaft durch das interdependente, individuelle Handeln ihrer Mitglieder geformt wird. Einerseits wirken also gesellschaftliche Strukturen und Institutionen auf die Deutungen und Handlungsweisen der Gesellschaftsmitglieder, andererseits ergeben sich gesellschaftliche Tatbestände letztlich immer aus dem Zusammenspiel der Handlungen von Individuen. Ein einfaches Beispiel ist etwa eine Reform der Regulierung von Kinderbetreuungsmöglichkeiten, die sich auf das Fertilitätsverhalten der Gesellschaftsmitglieder auswirkt, was Folgen für die demographische Entwicklung der Gesellschaft hat und unter Umständen den Kreis durch weiteren Reformbedarf von Kinderbetreuungsmöglichkeiten schließt. Insgesamt erscheinen Gesellschaften somit als komplexe, dynamische Systeme, in denen die Merkmale und Akteure unterschiedlicher konzeptioneller Ebenen in Wechselbeziehung zueinander stehen, was die Erklärung und Vorhersage gesellschaftlicher Prozesse und Entwicklungen vor große Herausforderungen stellt.

Ein hilfreiches Werkzeug, um sich dieser Herausforderungen anzunehmen, ist die Methode der Simulation. Unter einer *Simulation* wird eine modellhafte Imitation eines realen Systems verstanden, die dazu dient, das Verständnis der komplexen Systemprozesse zu verbessern, zu untersuchen, wie sich das dynamische System über die Zeit verändert, sowie zukünftige Zustände des Systems vorherzusagen. Für die empirische Forschung sind Simulationen von Bedeutung, da sie einerseits dazu dienen, Hypothesen über empirische Regularitäten herzuleiten oder deren Plausibilität zu prüfen. Andererseits lassen sich mit Simulationen ausgehend von empirischen Daten und Erkenntnissen über Wirkungsmechanismen Prognosen und Szenarien zukünftiger Entwicklungen erstellen.

## 12.1 System Dynamics

Abstrakte, dynamische Systeme können beispielsweise anhand von mathematischen Gleichungen modelliert werden (sog. Differenzialgleichungen), die die gegenseitigen Abhängigkeiten zwischen den verschiedenen Systemvariablen beschreiben. Eine erste Gleichung könnte etwa beschreiben, wie sich das reproduktive Verhalten ändert, wenn zusätzliche Krippenplätze geschaffen werden. Eine zweite Gleichung könnte fassen, wie das Angebot an Krippenplätzen von der Fertilitätsrate abhängt. Das Ziel ist dann zu ermitteln, wie sich das Betreuungsangebot und die Fertilitätsrate ausgehend von einer anfänglichen Veränderung der Anzahl Krippenplätze über die Zeit entwickeln.

Eine Art von Simulation, die direkt bei solchen Differenzialgleichungssystemen ansetzt, wird mit dem Begriff *System Dynamics* bezeichnet (Forrester 1971). Bei System Dynamics wird die Gesellschaft oder ein anderes System durch deterministische, jedoch zirkuläre und ineinander verzahnte Beziehungen zwischen Systemvariablen modelliert. Durch die Zirkularität der Beziehungen entstehen Feedbackschlaufen, die ein solches System analytisch unlösbar machen. Die Systemprozesse können jedoch durch numerische Methoden untersucht werden – das heißt, indem die Variablenwerte ausgehend von einem Anfangszustand gemäß den durch die Differenzialgleichungen unterstellten Zusammenhängen sequenziell fortgeschrieben werden. Das prominenteste Beispiel von System Dynamics ist die Studie zu den „Grenzen des Wachstums“ des Club of Rome (Meadows 1972), die in den 1970er-Jahren große Wellen schlug. Anhand einer Simulation der Interaktion zwischen Wachstum und Umweltressourcen sagte die Studie der Zivilisation den baldigen Kollaps vorher, sofern keine einschneidenden Maßnahmen bezüglich Bevölkerungswachstum und Konsum bzw. Umweltverschmutzung getroffen werden. Wichtige Variablen des in der Studie modellierten Weltsystems sind die

1. Produktion von Nahrungsmitteln als Funktion von verfügbarem Land und Düngemitteln,
2. die industrielle Produktion (z.B. von Düngemitteln),
3. die Bevölkerungsgröße und ihr Wachstum aufgrund der durch den verfügbaren Bestand an Nahrungsmitteln beeinflussten Geburten- und Sterberaten,
4. die durch die Produktion verursachte Umweltverschmutzung, sowie
5. die Ausbeutung nicht-erneuerbarer Ressourcen.

Ausgehend von unterschiedlichen Annahmen insbesondere über den Bestand an natürlichen Ressourcen wurden verschiedenen Szenarien zur Vorhersage der Entwicklung des Systems berechnet. Die Ergebnisse deuteten darauf hin, dass im Verlauf des 21. Jahrhunderts die Grenzen des Wachstums von Bevölkerung und industrieller Produktion erreicht werden und es danach aufgrund übermäßiger Umweltverschmutzung und der Erschöpfung natürlicher Ressourcen zu einem Systemzusammenbruch mit starkem Rückgang von Bevölkerung und Produktion kommen wird.

## 12.2 Mikrosimulation

An System Dynamics kann kritisiert werden, dass es sich um einen Ansatz handelt, bei dem Beziehungen zwischen Variablen auf der Makroebene (z.B. Bevölkerungsgröße, Geburtenrate, Ausmaß an industrieller Produktion) modelliert werden, ohne jedoch genauer auf die hinter diesen Beziehungen stehenden Mikroprozesse (z.B. das Handeln von Personen und Unternehmen) einzugehen. Ein solcher Ansatz ist begrenzt, da eine realistische Beschreibung von Zusammenhängen auf der Makroebene sehr schnell kompliziert werden kann und viele Annahmen benötigt werden. Anstatt sich auf Zusammenhänge zwischen Makrovariablen zu konzentrieren, die mit Hilfe von Annahmen in Form von Differenzialgleichungen festgelegt werden müssen, kann man jedoch auch versuchen, ein System aufgrund individueller Handlungen auf der Mikroebene zu modellieren. Eine Art von Simulation, die dieser Logik folgt, ist die so genannte *Mikrosimulation* (Caldwell 1997).

Mikrosimulationen werden zumeist für demographische Projektionen verwendet. Als Ausgangspunkt dienen dabei empirische Informationen zu Bevölkerungsstand und -struktur sowie die altersabhängigen Fertilitäts- und Sterbewahrscheinlichkeiten. Je nach Anwendung werden zudem weitere empirische Informationen einbezogen, z.B. zum Heirats- und Scheidungsverhalten oder der Abhängigkeit der Fertilität von Bildung und Einkommen. In einem ersten Schritt wird eine virtuelle Population von Individuen erzeugt, welche die empirisch festgestellte Bevölkerungsstruktur widerspiegelt. Die Individuen werden dann einer schrittweisen Alterung unterzogen, das heißt, das Alter der Individuen wird zum Beispiel auf Monats- oder Jahresbasis fortgeschrieben. In jedem Alterungsschritt können dabei entsprechend empirisch gemessener oder theoretisch festgelegter altersspezifischer Wahrscheinlichkeiten Ereignisse wie Geburten, Eheschließungen oder Todesfälle auftreten. Durch diese Ereignisse verändert sich die Population, die dann wiederum den Ausgangspunkt für den nächsten Alterungsschritt bildet. Über kürzere Zeiträume lassen sich so relativ zuverlässige Vorhersagen über die Entwicklung der Bevölkerungsgröße und -struktur treffen. Bei längeren Zeiträumen werden die Vorhersagen zunehmend ungenau, da in der Regel nicht davon ausgegangen werden kann, dass die altersspezifischen Ereigniswahrscheinlichkeiten konstant bleiben. Üblicherweise werden deshalb verschiedene Prognoseszenarien mit unterschiedlichen Annahmen über die Entwicklung von Ereigniswahrscheinlichkeiten (zum Beispiel der altersspezifischen Fertilität) erstellt.

---

## 12.3 Agentenbasierte Modellierung

Während System Dynamics und klassische Mikrosimulationen ihren Fokus hauptsächlich auf möglichst realistischen Vorhersagen zukünftiger Gesellschaftszustände haben, geht es bei der heute in den Sozialwissenschaften bedeutendsten Spielart von Simulation, der *agentenbasierten Modellierung* (kurz ABM für *Agent Based Model*), stärker um eine grundlegende theoretische Analyse von gesellschaftlichen Prozessen. Wie bei der klassischen

Mikrosimulation handelt es sich bei der agentenbasierten Modellierung um einen Ansatz mit individuellen Akteuren, die Akteure stehen hier jedoch in dynamischer Interaktion zueinander. Das heißt, die agentenbasierte Modellierung ist eine „bottom-up“ Strategie zur Simulation der Interaktionen von autonomen, heterogenen und unter Umständen lernfähigen Agenten. Die Interaktionen der Agenten generieren dabei eine Makroumgebung, die sich kontinuierlich verändert und auf die künftigen Handlungen der Agenten zurückwirkt (Gilbert 2008, Macy/Willer 2002).

Agentenbasierte Modelle schließen eine wichtige Lücke in der sozialwissenschaftlichen Erkenntnisgewinnung, was sich an James S. Colemans „Badewanne“ (*Coleman's boat*) erläutern lässt. Coleman unterscheidet die *Makroebene* der Gesellschaft und die *Mikroebene* der Individuen innerhalb der Gesellschaft. Anhand von Webers Studie zur protestantischen Ethik und dem Geist des Kapitalismus führt Coleman (1990: 6-9) aus, wie sich der Makro-Zusammenhang zwischen der religiösen Doktrin und der ökonomischen Organisation einer Gesellschaft über den Rückgriff auf die Mikroebene erklären lässt. Demnach generiere erstens die protestantische religiöse Doktrin gewisse Werthaltungen bei den Gesellschaftsmitgliedern wie Antitraditionalismus und Verpflichtung gegenüber der Berufung (Makro-Mikro-Beziehung). Zweitens führen diese Werthaltungen zu gewissen ökonomischen Verhaltensweisen wie Fleiß, Gewinnstreben und innerweltlicher Askese (Mikro-Mikro-Beziehung). Drittens ergibt sich aus dem Zusammenspiel dieser ökonomischen Verhaltensweisen eine kapitalistische Organisation der Gesellschaft (Mikro-Makro-Beziehung).

Der Vorteil einer solchen Argumentationskette via der Mikroebene ist, dass der Mechanismus, nach dem protestantische religiöse Doktrin zu Kapitalismus führen soll, explizit gemacht wird, was verschiedene Ansatzpunkte für eine aussagekräftige empirische Prüfung des Zusammenhangs liefert. Kontexteffekte (Makro-Mikro-Beziehungen) und Handlungstheorien (Mikro-Mikro-Beziehungen) werden traditionellerweise mit statistischen Methoden anhand von Beobachtungs- und Befragungsdaten analysiert. Weiterhin kommen experimentelle Methoden (Labor- sowie auch Feldexperimente) zum Einsatz, aufgrund derer sich valide Aussagen über die Kausalstruktur gefundener Zusammenhänge treffen lassen. In traditionellen Ansätzen weitgehend unberücksichtigt bleibt hingegen häufig das letzte Glied der Argumentationskette, der Mikro-Makro-Link. Dies liegt einerseits daran, dass oft zumindest implizit davon ausgegangen wird, die Mikro-Makro-Beziehung lasse sich durch eine einfache Aggregationsregel beschreiben. Andererseits erweisen sich klassische Methoden für die Analyse von Mikro-Makro-Beziehungen oft als ungeeignet, da sie keine Möglichkeit vorsehen, die Dynamik der Aggregation abzubilden (wobei geeignet gestaltete Experimente hier eine Ausnahme bilden können).

Beim Übergang von individuellem Handeln auf der Mikroebene zu einem gesellschaftlichen Phänomen auf der Makroebene handelt es sich in den meisten Fällen nicht um eine lineare, unidirektionale Beziehung. Vielmehr spielt sich ein dynamischer Prozess ab, bei dem sich die Akteure in ihrem Handeln gegenseitig beeinflussen. Das heißt, durch ihr Handeln verändern die Akteure laufend ihre lokale Umwelt, was sich auf das Handeln weiterer Akteure und auf die eigenen zukünftigen Handlungen auswirkt. Es entstehen so

Kettenreaktionen, Feedback-Schleifen und Kaskadeneffekte, die eine Vorhersage der Veränderungen auf der Makroebne erschweren. Man spricht in diesem Zusammenhang von *emergenten* Phänomenen, also eigenständigen Makrophänomenen, die keine direkte Entsprechung in den Eigenschaften der Mikroelemente haben.

Um die Komplexität des Übergangs von der Mikro- zur Makroebene zu illustrieren, beschreiben Macy/Willer (2002) das Beispiel eines Vogelschwams. Die einzelnen Vögel verhalten sich in ihrem Flug gemäß einiger höchst einfacher Regeln (Wahrung eines Mindestabstands zu den anderen Vögeln, Übernahme der Geschwindigkeit und Flugrichtung der umliegenden Vögel, Vermeidung peripherer Positionen) und doch ergibt sich aus diesen simplen und gut vorhersagbaren individuellen Verhaltensregeln ein höchst dynamisches, auf globaler Ebene äußerst schwierig zu modellierendes Flugmuster des ganzen Schwams. Während eine direkte Vorhersage des Schwarmverhaltens fast unmöglich ist, lassen sich durch Simulation einer Gruppe individueller Vögel mit den genannten Verhaltensregeln jedoch höchst realistische Resultate erzielen. Dabei werden ausgehend von einer Ausgangskonfiguration die Fluggeschwindigkeiten und Flugrichtungen der einzelnen Vögel schrittweise aufeinander angepasst. Die komplexe Dynamik entsteht dadurch, dass die Vögel durch ihre individuellen Verhaltensänderungen laufend ihre gegenseitigen lokalen Umgebungen beeinflussen, so dass sich Kettenreaktionen ergeben.

Auch wenn ein Vogelschwarm nur bedingt mit einer menschlichen Gesellschaft verglichen werden kann, so funktioniert letztere doch häufig nach ähnlichen grundlegenden Prinzipien. Das heißt, aufgrund relativ einfacher Verhaltensmuster der einzelnen Menschen können sich höchst komplexe Phänomene auf gesellschaftlicher Ebene ergeben. Ein wichtiges Ziel der Sozialwissenschaften ist es zu versuchen, solche Phänomene zu verstehen, und das Zusammenspiel von Mikro- und Makroebenen in einer dynamischen und nicht-deterministischen, probabilistischen Umwelt zu entschlüsseln. Simulationen, insbesondere in Form von agentenbasierten Modellierungen, sind ein geeignetes Werkzeug zur Erreichung dieses Ziels. Sie werden eingesetzt, um eine Vielzahl makrostruktureller Phänomene zu studieren, wie etwa die Entstehung von Normen und sozialer Ordnung, die Evolution von Kooperation, die Entwicklung von Preisen, die Evolution von Gemeindegrenzen oder die Entstehung und Auflösung von Verkehrsstaus, um nur einige Beispiele zu nennen.

Agentenbasierte Simulationsmodelle werden grundsätzlich zu drei verschiedenen Zwecken eingesetzt. Erstens können ABM verwendet werden, um theoretische Erklärungen für gesellschaftliche Phänomene zu suchen (Epstein 2007). Es wird dabei geprüft, ob ein bestimmter Mechanismus in der Lage ist, ein empirisch beobachtetes Phänomen wie zum Beispiel soziale Kooperation oder Segregation zu erzeugen. Ein positives Ergebnis – wenn sich also das empirische Phänomen durch die einfachen Annahmen des Simulationsmodells generieren lässt – beweist allerdings nicht, dass die angenommenen Wirkungsmechanismen tatsächlich für das in der Realität beobachtete Phänomen verantwortlich sind. Die Simulation zeigt lediglich auf, was potenzielle Erklärungen sein könnten. In diesem Sinne ist agentenbasierte Modellierung ein Mittel zur sozialwissenschaftlichen Theoriebildung.

Ein zweiter wichtiger Anwendungsbereich liegt in der Durchführung von Simulationsexperimenten, also in der Generierung hypothetischer Szenarien durch gezielte Variation verschiedener Elemente des Simulationsmodells (Cederman 1996). Eine solche Anwendung erfolgt in der Regel erst auf einer zweiten Stufe der Modellentwicklung, nachdem gezeigt wurde, dass sich mit den Annahmen, die dem Modell zugrunde liegen, realistische Makrophänomene erzeugen lassen. An diesem Punkt der Modellentwicklung, wird eine hypothetische Veränderung in das Modell eingeführt, um seine Auswirkungen auf die Simulationsergebnisse zu beobachten. Dies kann beispielsweise ein Weg sein, um die Wirkung vorgeschlagener politischer Interventionen abzuschätzen und – immer im Rahmen der getroffenen Modellannahmen – optimale Lösungen für bestimmte Probleme zu finden.

Ein dritter Verwendungszweck liegt in der realistischen Modellierung von Phänomenen in Korrespondenz mit empirischen Daten. Empirische Informationen können einerseits zur Kalibrierung eines Modells verwendet werden (z.B. durch Festlegung der Anfangsverteilungen von Merkmalen und Methoden der Agenten nach empirischen Gegebenheiten). Andererseits lassen sich empirische Daten zur Validierung von Modellergebnissen nutzen, z.B. vergleicht Epstein (2007) simulierte Migrationsmuster mit empirisch beobachteten Migrationsströmen. Wenn die Dynamik von Mikro-Makro-Prozessen hingegen bereits sehr gut verstanden wird, kann sich die Anwendungslogik von agentenbasierten Modellen auch umdrehen. Simulationen lassen sich dann ähnlich wie herkömmliche statistische Verfahren nutzen, um die Beiträge verschiedener Mechanismen zur Erzeugung beobachteter empirischer Phänomene zu bestimmen. Ein Beispiel für eine solche Anwendung ist im Bereich der Analyse dynamischer Netzwerke zu finden (Baur, Kapitel 74 in diesem Band), wo allgemeine Simulationssoftware zur Abbildung typischer Netzwerkprozesse existiert (Snijders et al. 2010). Die Werte der Parameter der Simulation, die etwas über die relative Bedeutung verschiedener Mechanismen aussagen, werden dabei schrittweise so lange angepasst, bis die Simulationsergebnisse möglichst gut mit den Beobachtungsdaten übereinstimmen. Die am weitesten fortgeschrittenen Beispiele für die Integration von Daten und Simulation findet man indes in Modellierungen wie „Urbansim“ (Waddell 2002), die multiple Mikro-Makro-Szenarien zu einer Simulation vereinen (z.B. Raumplanung und Verkehr) und eine ganze Reihe von Ergebnisvariablen projizieren und mit empirischen Daten abgleichen (bei „Urbansim“ sind das etwa städtebaulich relevante Indikatoren wie Ansiedlungsmuster oder Verkehrsströme von Pendlern).

---

## 12.4 Funktionsweise

Einige Hauptmerkmale von agentenbasierten Modellen sind wie folgt:

1. Ein ABM baut in der Regeln auf einer einzigen, grundlegenden Ebene von unabhängigen Mikroagenten auf (wobei allerdings auch erweiterte Modelle mit zusätzlichen Agenten auf höheren Ebenen möglich sind), die per Zufall oder nach vorgegebenem Muster in einem virtuellen Raum platziert werden (im einfachsten Fall auf einem regelmäßigen Gitter; die Topologie des virtuellen Raumes kann sich jedoch auch erst aus dem Handeln der Agenten ergeben).

2. Die Mikroagenten haben Eigenschaften (Merkmale oder Variablen) und Methoden oder Strategien (Regeln, die ihr Verhalten bestimmen).
3. Die Eigenschaften der Agenten sind i.d.R. heterogen, das heißt, die Agenten unterscheiden sich in ihren Merkmalen voneinander, und die Methoden bzw. Verhaltensregeln sind idealerweise möglichst sparsam. Das heißt, um das Modell handhabbar zu halten, verwendet man in der Regel möglichst einfache, von der Wirklichkeit abstrahierende Verhaltensregeln.
4. Die Agenten werden in ihrem Verhalten durch den Kontext beeinflusst, wobei sie üblicherweise nur über begrenzte Information verfügen und sich so in einer Art „lokalen“ Umwelt bewegen. Diese lokalen Welten ergeben sich beispielsweise über soziale Netzwerke oder räumliche Nähe.
5. Die Methoden der Agenten oder allgemeiner die Prozesse, nach denen die Agenten miteinander interagieren, sind probabilistisch, enthalten also Zufallselemente, so dass es sich bei einem ABM um eine stochastische Form von Simulation handelt.
6. Manche agentenbasierte Modelle sind durch „adaptive“ Agenten gekennzeichnet. Das heißt, die Agenten können ihre Merkmale und/oder Methoden aufgrund erhaltenener Rückmeldungen aus der lokalen Umwelt anpassen. Solche Anpassungen können entweder durch Lernen oder durch evolutionäre Prozesse stattfinden. Ein Spezialfall davon sind Simulationsmodelle, in denen die Verteilungen von Eigenschaften der Agenten mit Hilfe genetischer Algorithmen verändert werden. Ein genetischer Algorithmus ahmt anhand von Mechanismen wie Kreuzung, Mutation und Selektion einen evolutionären Prozess nach. Die Merkmale und präferierten Methoden von Agenten können dabei durch eine Kette von „Genen“ zum Beispiel in Form einer Folge numerischer Codes beschrieben werden. Die „Gene“ von Agenten werden dann im Verlauf der Simulation gekreuzt oder es werden Mutationen eingeführt, wodurch neue Kombinationen von Merkmalen und Methoden entstehen. Gleichzeitig erfolgt eine Selektion von erfolgreichen Agenten oder Strategien aufgrund eines im Rahmen der vorliegenden Simulation geeigneten Fitness-Maßes (das heißt, Agenten mit einer besseren „Fitness“ erhalten höhere Überlebens- oder Reproduktionschancen von Runde zu Runde). Die Simulation wird so lange fortgesetzt, bis sich ein Gleichgewicht ergibt und das Modell nur noch mit den erfolgreichsten Agenten bzw. Strategien bevölkert ist.

Den Ablauf einer agentenbasierten Simulation kann man sich etwa wie folgt vorstellen: In einem ersten Schritt wird das System initialisiert, das heißt, die Agenten werden nach bestimmten Verteilungen generiert und im virtuellen Raum platziert. Danach wird die Simulation in Gang gesetzt. Über eine bestimmte Anzahl Runden (die sich üblicherweise danach richtet, wie lange es dauert, bis die Simulation einen gewissen Gleichgewichtszustand erreicht) wird das Modell mit seinen Agenten, Regeln und Eigenschaften fortgeschrieben, wobei den Agenten in jeder Runde die Möglichkeit gegeben wird, sequenziell in einer vorgegebenen oder zufälligen Reihenfolge Handlungen zu vollziehen oder ihre Merkmale und Methoden anzupassen (je nach Simulation können zudem Agenten dazu kommen oder wegfallen). Der Zustand des Systems wird in jeder Runde (oder zu bestimmten Abständen) erfasst, was Daten produziert, die später analysiert werden können. Sofern es

sich um eine stochastische Simulation handelt, ist es sinnvoll, einen solchen Simulationslauf mit gleichen Anfangsbedingungen mehrmals zu wiederholen. Weiterhin werden häufig mehrere Simulationsläufe mit variierten Anfangsbedingungen durchgeführt, um die Abhängigkeit der Simulationsergebnisse von den Anfangsbedingungen zu untersuchen. Agentenbasierte Modelle sind somit weder strikt deduktiv noch rein induktiv. Ein deduktives Vorgehen beginnt mit einem Satz expliziter Annahmen und leitet dann mit Hilfe logischer Argumentation Schlussfolgerungen aus diesen Annahmen ab. Bei einem induktiven Verfahren trifft man hingegen generalisierende Aussagen aufgrund von empirischen Beobachtungen. Ähnlich wie die deduktive Logik fußt ein ABM auf einem Satz grundlegender Annahmen. Diese Annahmen betreffen die Verhaltens- und Interaktionsregeln sowie die Verteilungen der Merkmale der Agenten. Die durch die Simulation generierten Daten werden jedoch ähnlich wie Daten von einem wissenschaftlichen Experiment oder einer Bevölkerungsbefragung induktiv analysiert. Das heißt, mit Hilfe induktiver, statistischer Methoden werden aus den durch die Simulation erzeugten Daten verallgemeinernde Rückschlüsse über die aus den Annahmen folgenden Phänomene gezogen.

---

## 12.5 Beispiele

Agentenbasierte Modelle werden in vielen verschiedenen Bereichen eingesetzt wie zum Beispiel der Stadtplanung, der Biologie, der Epidemiologie oder den Ingenieurwissenschaften. Anwendungen beziehen sich etwa auf die Untersuchung von Zersiedelungsprozessen und Verkehrsströmen, die Diffusion von Tuberkulose-Ansteckungen oder die optimale Gestaltung von Energieversorgungsnetzen. In den Sozialwissenschaften finden sich ABMs bislang vor allem in der Soziologie, Ökonomie und Politikwissenschaft.

Eine der bekanntesten soziologischen Anwendungen ist Schellings (1971) Modell der Segregation. Dieses einfache Modell, das man gut auch ohne Computer auf einem Schachbrett durchspielen kann, zeigte, dass auch dann eine stark segregierte Gesellschaft entstehen kann, wenn die Gesellschaftsmitglieder nur moderate Integrationstendenzen haben. Das Modell funktioniert wie folgt (Schelling 1971): Man nehme ein Schachbrett und eine Handvoll kupferfarbener 5-Cent- und goldfarbener 10-Cent-Münzen. Die 5-Cent- und 10-Cent-Münzen repräsentieren zwei unterschiedliche soziale Gruppen, zum Beispiel zwei Ethnien. Die Münzen werden nun per Zufall auf dem Schachbrett verteilt (wobei pro Feld nur eine Münze erlaubt ist und einige Felder frei bleiben sollen). Der Reihe nach (z.B. nach Zufall oder festgelegtem Muster) wird dann für jede Münze bestimmt, ob sie „zufrieden“ ist mit ihrer unmittelbaren Nachbarschaft (die acht umliegenden Felder). Eine Münze ist beispielsweise zufrieden, wenn mehr als ein Drittel der maximal acht Nachbarn die gleiche Farbe haben wie sie selbst (z.B. mindestens drei von acht Nachbarn oder mindestens zwei von fünf Nachbarn). Wird dieser Wert unterschritten, ist die Münze unzufrieden und zieht zum nächstgelegenen Feld mit einer zufriedenstellenden Nachbarschaft um. Dieser Prozess wird so lange wiederholt, bis keine Münze mehr umziehen möchte

(oder bis das System ein dynamisches Gleichgewicht annimmt). Auch bei relativ moderaten Präferenzen (z.B. mehr als ein Drittel der Eigengruppe in der Nachbarschaft) zeigt sich, dass in den meisten Fällen nach Abschluss der Simulation ein Ergebnis vorliegt, bei dem die 5-Cent- und 10-Cent-Münzen räumlich stark segregiert sind. Das heißt, es bilden sich Klumpen von 5-Cent- und 10-Cent-Münzen, so dass der Anteil der Eigengruppe in der Nachbarschaft für die meisten Münzen deutlich höher liegt, als man aufgrund der moderaten Präferenz erwarten würde. Der Grund für die starke Segregation im Schelling-Modell liegt an den Kettenreaktionen, die durch die Umzüge der einzelnen Münzen ausgelöst werden. zieht eine Münze nämlich in ein anderes Feld, so verändern sich die Umgebungen der alten und neuen Nachbarn, die dadurch u.U. selbst unzufrieden werden und ebenfalls umziehen, was sich wiederum auf deren Nachbarn auswirkt, usf.

In den Wirtschaftswissenschaften wurden agentenbasierte Modelle u.a. verwendet, um die Evolution ganzer Ökonomien zu erklären. Sehr bekannt ist beispielsweise Epstein/Axtells (1996) Sugarscape-Simulation, einem ursprünglich einfachen Modell des Wettstreits von Agenten um knappe Ressourcen (Zucker). Die Sugarscape-Simulation wurde in der Folge schrittweise ausgebaut, um eine Vielzahl von Phänomenen wie Tausch, kriegerische Auseinandersetzungen, Handel oder die Übertragung von Krankheiten zu untersuchen. Die heute wahrscheinlich häufigsten Anwendungen von agentenbasierten Modellen in der Ökonomie beziehen sich auf Preisbildungsmechanismen und Finanzmärkte.

An der Schnittstelle zwischen Ökonomie und Politikwissenschaft befindet sich die Simulation von Kollman et al. (1997), die sich der klassischen Frage annimmt, wie öffentliche Güter in effizienter Weise bereitgestellt werden können. Tiebout (1956) hatte als Lösung die Einführung von Konkurrenz zwischen politischen Einheiten vorgeschlagen, so dass ein Markt für öffentliche Güter entsteht. Leider lässt sich jedoch zeigen, dass dies zu einer ineffizienten Passung zwischen Präferenzen und Angebot an öffentlichen Gütern führt. In ihrer Simulation untersuchten Kollman et al. (1997), wie politische Mechanismen (Referenden zu politischen Entscheidungen, Konkurrenz zwischen Parteien um Wähler, Kombination von Parteiprogrammen in Abhängigkeit von Parteistärken) genutzt werden können, um effizientere Ergebnisse zu erreichen.

In einer politikwissenschaftlichen Anwendung untersuchte Axelrod (1984) den Erfolg verschiedener Strategien in einem iterierten Gefangenendilemma. Ein Gefangenendilemma beschreibt eine Entscheidungssituation, in der zwei Akteure entweder kooperieren oder defektieren können. Für beide Akteure ist es individuell rational zu defektieren, durch gegenseitige Kooperation könnte jedoch ein besseres Ergebnis erreicht werden. Ein iteriertes Gefangenendilemma ist ein Gefangenendilemma, das zwischen den gleichen Akteuren wiederholt gespielt wird. In Computerturnieren ließ Axelrod verschiedene Strategien gegeneinander antreten, die sich von Runde zu Runde in Abhängigkeit ihres Erfolgs vermehrten konnten. Als erfolgreichste Strategie stellte sich die berühmte Tit-for-Tat-Strategie heraus – eine freundliche und kooperative Strategie, die immer mit Kooperation beginnt und kooperatives Verhalten mit eigener Kooperation belohnt, die sich jedoch nicht ausnutzen lässt, da sie unkooperatives Verhalten immer mit Defektion bestraft. Die

Simulationen von Axelrod zeigten, dass Kooperation in einem iterierten Gefangenendilemma möglich ist, und haben ein breites Forschungsprogramm zur Evolution von Kooperation hervorgerufen.

---

## 12.6 Fazit

Wie man an den aufgeführten Beispielen erkennen kann, handelt es sich bei der Simulation – insbesondere in Form der agentenbasierten Modellierung – um eine Methode mit viel Potenzial in diversen Forschungsbereichen. Eine große Herausforderung für Simulationsmethoden liegt allerdings in der Sicherstellung der Zugänglichkeit und Replizierbarkeit der Ergebnisse. Simulationen benötigen spezifischen Computercode und können häufig nicht mit einem einfachen, standardisierten Werkzeug umgesetzt werden. Einerseits besteht dadurch die Gefahr, dass sich Programmierfehler einschleichen, die die Resultate verfälschen. Andererseits wird für Simulationsprojekte hochspezialisiertes Personal benötigt, das sich nicht nur im entsprechenden inhaltlichen Fachbereich auskennt, sondern auch hohe computertechnische Fähigkeiten mitbringt. Neuere Entwicklungen weisen darauf hin, dass die Probleme der Standardisierung und Replizierbarkeit zumindest teilweise überwunden werden können, da es zwischen den Simulationsmodellen durchaus viele Gemeinsamkeiten gibt, die sich geeignet verallgemeinern und in Form von vorgefertigter Software bereitstellen lassen.

Ferner gibt es verschiedene Einführungen und Referenzwerke, an denen man sich orientieren kann. Einen Einstieg bieten etwa Gilbert (2008), Gilbert/Troitzsch (2005), Epstein (2007) oder Miller/Page (2007). Zur technischen Umsetzung von Simulationen stehen verschiedene Werkzeuge zur Verfügung. Im Allgemeinen lassen sich Simulationen in beliebigen objektorientierten Programmiersprachen realisieren (z.B. Java, Python, MATLAB oder R). Zumindest für den Einstieg kann es jedoch sinnvoll sein, auf vorprogrammierte Simulationsumgebungen wie NetLogo (<http://ccl.northwestern.edu/netlogo/>), Swarm (<http://savannah.nongnu.org/projects/swarm>), RePast (<http://repast.sourceforge.net/>) oder Manson (<http://www.cs.gmu.edu/~eclab/projects/mason/>) zurückzugreifen.

## Literatur

- Axelrod, Robert (1984): *The Evolution of Cooperation*. New York: Basic Books
- Cederman, Lars Eric (1996): *Rerunning History: Counterfactual Simulation in World Politics*. In: Tetlock/Belkin (Hg.)
- Caldwell, Steven B. (1997): *Dynamic Microsimulation and the Corsim 3.0 Model*. Ithaca, NY: Strategic Forecasting
- Coleman, James S. (1990): *Foundations of Social Theory*. Cambridge, MA: Harvard University Press
- Gilbert, Nigel (2008): *Agent-Based Models*. Thousand Oaks, CA: Sage
- Gilbert, Nigel/Trotzsch, Klaus G. (2005): *Simulation for the Social Scientist*. Second Edition. Berkshire: Open University Press
- Epstein, Joshua M. (2007): *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton, N.J.: Princeton University Press
- Epstein, Joshua M./Axtell, Robert (1996): *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, DC: Brookings Institution Press
- Forester, Jay W. (1971): *World Dynamics*. Cambridge, MA: MIT Press
- Kollman, Ken/Miller, John H./Page, Scott F. (1997): Political Institutions and Sorting in a Tiebout Model. In: *American Economic Review* 87: 977-992
- Macy, Michael W./Willer, Robert (2002): From Factors to Actors: Computational Sociology and Agent-Based Modeling. In: *Annual Review of Sociology* 28: 143-166
- Meadows, Dennis (1972): *Die Grenzen des Wachstums*. Bericht des Club of Rome zur Lage der Menschheit. Stuttgart: Deutsche Verlags-Anstalt
- Miller, John H./Page, Scott E. (2007): *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, N.J.: Princeton University Press
- Schelling, Thomas. C. (1971): Dynamic Models of Segregation. In: *Journal of Mathematical Sociology* 1: 143-186.
- Snijders, Tom A.B./Steglich, Christian E. G./van de Bunt, Gerhard G. (2010): Introduction to Actor-Based Models for Network Dynamics. In: *Social Networks* 32: 44-60
- Tetlock, Philip E./Belkin, Aaron (Hg.) (1996): *Counterfactual Thought Experiments in World Politics*. Princeton: Princeton University Press
- Tiebout, Charles (1956): A Pure Theory of Local Expenditures. In: *The Journal of Political Economy* 64: 416-24
- Waddell, Paul (2002): UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning. In: *Journal of the American Planning Association* 68: 297-314

Irena Medjedović

---

## 13.1 Zugänge zu qualitativen Daten für Sekundäranalysen

Während die Sekundäranalyse auch Dank einer positiven Archiventwicklung zum Synonym für die erneute Nutzung statistischer (insbesondere Umfrage-)Daten (Mochmann, Kapitel 14 in diesem Band) geworden ist, zeigt sich nunmehr auch eine internationale und aktuell voranschreitende Entwicklung von Archiven, die interessierten Forschenden einen organisierten Zugang zu qualitativen Daten schaffen.

Umfangreichere Datensammlungen bieten bereits das „*Henry A. Murray Research Archive*“ an der Harvard University und „*Qualidata*“ des UK Data Archive. Weitere Informationen zum internationalen Stand finden sich bei Medjedović und Witzel (2010) und im „*Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*“ (FQS, Heft 12 (3) 2011) sowie im „*IASSIST Quarterly*“ (Heft 34 & 35 2011).

In Deutschland existiert bislang kein zentrales Angebot an qualitativen Daten für die Sozialforschung. Stattdessen findet sich *eine Reihe dezentraler Archive* für recht unterschiedliche qualitative Daten. Zum großen Teil sind diese Archive auf jeweils spezifische Themenbereiche und entsprechend unterschiedliche qualitative Datenformen spezialisiert. Neben einigen Archiven zu politischen Bewegungen und Parteiengeschichte (z.B. „Archiv der sozialen Demokratie“ der Friedrich-Ebert-Stiftung) können interessierte Forschende vor allem auf viele geschichtswissenschaftlich ausgerichtete Archive mit prozessproduzierten Daten, aber auch (auto-)biografischen Dokumenten (Ernst, Kapitel 63 in diesem Band) und Interviews (Küsters, Kapitel 40 in diesem Band) der Oral History (z.B. Archiv „Deutsches Gedächtnis“ an der FernUniversität Hagen) sowie auf einige Serviceeinrichtungen mit Daten aus originär sozialwissenschaftlichen Forschungsprojekten (z.B. „Qualiservice“, vormals „Archiv für Lebenslaufforschung“ an der Universität Bremen; „Archiv Kindheit-Jugend-Biographie“ an der Universität Siegen; „Datenservicezentrum für Betriebs- und Organisationsdaten“ an der Universität Bielefeld) zurückgreifen. Hinweise auf weitere qualitative Datenarchive in Deutschland finden sich in Medjedović und Witzel (2010: 90f.).

In den meisten kleinen Spezialarchiven fehlen internethfähige Datennachweissysteme und es gibt Probleme der Langzeitsicherung des meist nicht digitalisierten Datenbestands. Vor dem Hintergrund dieser defizitären Lage untersuchte das Archiv für Lebenslaufforschung (ALLF) in Kooperation mit dem GESIS-Datenarchiv für Sozialwissenschaften in einer bundesweiten DFG-geförderten Studie erstmals die Machbarkeit einer zentralen Servicestelle für qualitative Daten. Aus den Ergebnissen der Machbarkeitsstudie wurde die Konzeption eines solchen Servicezentrums („Qualiservice“) erarbeitet (Medjedović/Witzel 2010). Als Kooperationsprojekt mit dem eScience Lab an der Universität Bremen, der Staats- und Universitätsbibliothek Bremen und der GESIS ([www.gesis.de](http://www.gesis.de)) soll *Qualiservice* realisiert werden. Dies schließt den Aufbau eines transparenten Online-Datennachweissystems sowie die Vernetzung bestehender dezentraler Spezialarchive ein, womit ein nutzungsfreundlicher Datenzugang entstehen soll. Für den Zugang zu Studien, die Methodenkombinationen angewandt haben, wird ein wechselseitiger Datennachweis im Qualiservice (qualitative Daten) und im Datenarchiv für Sozialwissenschaften (GESIS) (quantitative Daten) aufgebaut.

Die Frage der Schaffung von Infrastrukturen für die qualitative Forschung erhält jüngst eine neue Dynamik. So erkennt auch der *Wissenschaftsrat (WR)* für Deutschland einen Nachholbedarf im Bereich der Archivierung qualitativer Forschungsdaten und empfiehlt eine verstärkte Förderung von Mixed-Methods-Projekten (Kelle, Kapitel 8 in diesem Band), die auch der Langzeitverfügbarmachung von qualitativen Forschungsdaten dienen soll (WR 2011, 2012). Diesen Empfehlungen folgend rücken vor allem die Aktivitäten des *Rat für Sozial- und Wirtschaftsdaten (RatSWD, www.ratswd.de)* die Entwicklung in eine neue Perspektive: Gingen die bisherigen Bemühungen um den Aufbau von qualitativen Datenarchiven vornehmlich von Akteuren aus der qualitativen Forschungscommunity aus, bietet sich nunmehr die Chance, qualitative und quantitative Forschungs- und Dateninfrastrukturen zu integrieren und damit die Bemühungen auf eine breitere Basis zu stellen (Huschka et al. 2013).

---

### 13.2 Aufbau dieser Daten

Suchen interessierte Forschende nach passenden Daten für ihre Forschungsfrage, treffen sie zunächst auf *strukturierte Metadaten* zum archivierten Datensatz. Sie ermöglichen Nutzenden die Identifikation bzw. das Auffinden von Studien und erlauben eine erste Entscheidung über die Nützlichkeit für das jeweilige Forschungsvorhaben. Diese Art von Informationen wird typischerweise in Form von Studienbeschreibungen in Datennachweissystemen bzw. sog. *Bestandskatalogen* bereitgestellt. Sie enthalten in der Regel eine kurze Zusammenfassung der Studie und Angaben zu den wesentlichen Eckdaten, die einen Überblick über die Studie vermitteln. Zu diesen Metadaten gehören auch Informationen über *Zugangsbedingungen* sowie weitere Hinweise etwa auf verlinkte Datensätze.

Die internationalen Vorreiter im Bereich des Service für qualitative Daten orientieren sich bei der Erstellung der strukturierten Metadaten am internationalen Dokumentationsstandard der *Data Documentation Initiative (DDI)*, der bereits seit einigen Jahren im Bereich der quantitativen Datenarchivierung etabliert ist und in angepasster Form auch bei qualitativen Daten Anwendung findet. Wie oben bereits erläutert, finden sich in den deutschen Archiven weitgehend noch keine Online-Datennachweissysteme. Die Datenbestände werden in der Regel grob auf den jeweiligen Internetseiten der Archive beschrieben. Genauere Recherchen können zum Teil vor Ort in Datenbanken vorgenommen werden.

In der empirisch-qualitativen Sozialforschung werden mit unterschiedlichen Methoden und Vorgehensweisen Textinformationen, Audio- und visuelles Material erzeugt. Diese Dokumentationen wiederum liegen in verschiedenen Formaten vor. Texte können auf Papier festgehalten sein oder aber als digitale Textdatei vorliegen. Audio- und Videoaufnahmen können sowohl in analogen als auch digitalen Formaten gespeichert sein. Entsprechend dieser Vielfalt an Datenformen und Formaten liegen auch die Daten in den Archiven unterschiedlich vor.

Ähnlich wie es im Bereich des Datenservice der quantitativ orientierten Sozialforschung seit Jahrzehnten üblich ist, orientieren sich auch die internationalen Datenarchive daran, für die Nutzung gut aufbereitete, dokumentierte und elektronisch verfügbare Daten bereitzustellen. Für die deutsche Situation machen die Ergebnisse der erwähnten Machbarkeitsstudie deutlich, dass der Zustand der Daten eines großen Teils der Forschungsprojekte als so mangelhaft zu kennzeichnen ist, dass zunächst *Standards hinsichtlich einer geeigneten Datenaufbereitung* zu definieren sind, an denen sich die zukünftige Forschungspraxis und die Datenübernahme in die Archive orientieren können. Diese Standards beziehen sich im Wesentlichen auf:

1. die *formale Sicherung der Qualität und Nutzbarkeit der Daten* (Vollständigkeit, technische Qualität, Software-Format, etc.),
2. die *datenschutzrechtliche Sicherung* (Mühlchen, Kapitel 4 in diesem Band) der Archivier- und Nutzbarkeit der Daten sowie
3. die Gewährleistung der *Nachvollziehbarkeit des Erhebungs- und Forschungskontextes* (Meyer/Meier zu Verl und Flick, Kapitel 15 und 29 in diesem Band) der jeweiligen Daten.

Aufgrund der besonderen Sensibilität von qualitativen Daten ist die Umsetzung der *datenschutzrechtlichen Aspekte* schwieriger als bei quantitativen Daten. So erhalten die Nutzenden die Daten in der Regel in anonymisierter Form. Die Problematik der *Anonymisierung qualitativer Daten* besteht darin, dass durch die Lösung oder Veränderung von Informationen die Nutzbarkeit der Daten für die wissenschaftliche Analyse potenziell beeinträchtigt wird (Thomson et al. 2005). Das Archiv für Lebenslaufforschung entwickelte daher ein Anonymisierungskonzept für Interviewtranskripte, das einerseits personenbezogene Informationen aus den Transkripten entfernt, sie andererseits aber auch mit sozialwissenschaftlich relevanten Informationen anreichert (Kluge/Opitz 1999; Medjedović/Witzel 2010: 149-154).

Die Gewährleistung des Zugangs zu *ausreichend Kontextinformationen* ist die zentrale methodologische Prämisse für die sekundäre Verwendung des Datenmaterials. Nutzende erhalten daher ergänzend zum Datensatz eine ausführliche Dokumentation, die dabei hilft, die Daten und ihren Entstehungskontext nachzuvollziehen (siehe hiezu 13.3.2). Darüber hinaus umfasst die Dokumentation auch Informationen darüber, welche Schritte bei der Aufbereitung der Daten stattgefunden haben. Bei Daten in Textform betrifft dies vor allem eine Erläuterung des Transkriptionsverfahrens und der Form der durchgeführten Anonymisierung.

---

### 13.3 Regeln im Umgang mit qualitativen Sekundärdaten

Gegenüber einem konventionellen Forschungsprozess zeichnet sich die Sekundäranalyse vor allem dadurch aus, dass der Prozess der Datenauswertung und -interpretation vom Prozess der Datenerhebung entkoppelt ist (Klingemann/Mochmann 1975). Diese Entkopplung geht mit einigen methodologischen Implikationen einher (ausführlich: Medjedović 2014). So werden „Daten“ nicht als objektive und im Feld vorzufindende Entitäten verstanden, sondern als soziale und kontextuell eingebettete Produkte. Dies gilt insbesondere für qualitative Erhebungsverfahren, in denen die soziale Situation, d.h. die intersubjektive und interaktive Beziehung zwischen forschendem und beforschtem Subjekt, von tragender Bedeutung ist.

#### 13.3.1 Die methodologische Diskussion

Eine Sekundäranalyse lässt sich nur durchführen, wenn unterstellt wird, dass Daten auch außerhalb ihres unmittelbaren Erhebungskontextes ausgewertet und interpretiert werden können. Diese Annahme ist jedoch nicht unumstritten und bietet die Grundlage für eine methodologische Diskussion, die zum Teil stark polarisiert geführt wird.

Die prominenteste Kritik an der Machbarkeit der Sekundäranalyse stammt von Mauthner et al. (1998). Gestützt auf eigene sekundäranalytische Versuche stellen sie die (über historische und methodologische Untersuchungen hinausgehende) erneute Nutzung qualitativer Daten prinzipiell in Frage. Da es unmöglich sei, den ursprünglichen Status, den die Primärforschenden hatten, wieder herzustellen, sei die Sekundäranalyse unvereinbar mit einer interpretativen und reflexiven Epistemologie (Mauthner et al. 1998: 742f.).

Andere Autoren und Autorinnen sind dagegen der Auffassung, dass der Nachvollzug kontextueller Effekte weniger ein epistemologisches als ein praktisches Problem sei, das sich zudem auch in Primäranalysen stelle. Qualitativ Forschende hätten häufig mit unvollständigen (Hintergrund-)Informationen umzugehen und abzuwagen, inwieweit ein Aspekt tatsächlich belegt werden könne oder doch verworfen werden müsse (Fielding 2004: 99).

Eine zweite Form der Replik auf Mauthner et al. (1998) kritisiert deren Verharren in einem Verständnis von Kontext, der statisch und fix in der Vergangenheit angesiedelt

werde. Moore (2006) verweist darauf, dass Forschende in der Auseinandersetzung mit den Daten diese immer auch in einen eigenen Kontext *setzen*. Daher sei es nicht das Ziel, das originäre Forschungsprojekt vollständig nachzubilden. Vielmehr sei die Sekundäranalyse als neuer Prozess der Rekontextualisierung und Rekonstruktion von Daten zu verstehen.

Diese Diskussion reflektiert darauf, dass die *Kontextsensitivität* einen Grundpfeiler qualitativer Forschung darstellt. Die Einsicht in die Kontextabhängigkeit einer sprachlichen Äußerung oder einer Handlung eint alle qualitativen Forschungsansätze und berührt einen wichtigen Punkt im Selbstverständnis dieser Forschungstradition. Hinzu kommt, dass qualitative Forschung häufig damit verbunden wird, sich persönlich ins Feld zu begeben, um mit Kontextwissen aus „erster Hand“ die anschließende Analyse und Interpretation der Daten leisten zu können.

In Sekundäranalysen fehlt den Forschenden dieser unmittelbare Bezug zum Kontext. Alternativ können aber Wege der Kontextualisierung aufgezeigt werden. Für die Sekundäranalyse ist hierbei relevant, welche Art von Kontext überhaupt gemeint ist – denn Kontext wird je nach Forschungsansatz unterschiedlich definiert und ist somit auch je nach Auswertungsinteresse unterschiedlich relevant.

### 13.3.2 Zugang zu Kontextinformationen

Auf der Ebene der einzelnen Interaktion sollte zuallererst der Zugang zu den „Daten selbst“ gegeben sein. In qualitativen Auswertungsverfahren wird der Fallanalyse (Meyer/Meier zu Verl, Kapitel 15 in diesem Band) ein zentraler Stellenwert eingeräumt. Auch in Studien, die allgemeine Aussagen auf der Grundlage von vielen Fällen entwickeln, dient die Rekonstruktion des Einzelfalls in der Regel als Ausgangspunkt. Diese Interpretation setzt den Fokus auf den Kontext, den die Beteiligten durch die wechselseitige Bezugnahme aufeinander in der Interaktion selbst erzeugen (*kommunikativer Kontext der Konversation*). D.h. entscheidend ist, wie die Beteiligten das Gespräch führen. Die detaillierte Arbeit am einzelnen Fall erfordert den Zugriff auf Aufnahmen und/oder das Gespräch möglichst präzise erfassende Transkripte (Kuckartz/Rädiker, Kapitel 27 in diesem Band); die Einbettung einer einzelnen Sequenz in den Gesamtverlauf der Interaktion oder einer einzelnen Äußerung in den Kontext einer längeren Erzählung erfordert die Vollständigkeit von Aufnahme oder Transkript.

Darüber hinaus werden Metainformationen über das Gespräch als soziale Situation (*situationaler Kontext*) relevant. Eine Interaktion und ihre Akteure sind stets verortet in Raum und Zeit, d.h., dass das unmittelbare Setting bedeutsam sein kann: Soziale Interaktionen können etwa zu unterschiedlichen Tageszeitpunkten unterschiedlich verlaufen. Ebenso können räumliche Bedingungen das Gespräch beeinflussen oder selbst empirisches Material für die Forschungsfrage liefern (z.B. Wohnsituation der Befragten). Ferner mögen die Beteiligten ein gemeinsames Hintergrundwissen haben, das die Interaktion rahmt und von Bedeutung ist, aber nicht explizit im Gespräch artikuliert wird. Beispiele hierfür wären: Merkmale der Beteiligten wie Alter, Geschlecht, Ethnie, soziale Klasse;

Informationen über relevante Dritte oder die Anwesenheit Dritter sowie weitere Informationen über die Beziehung zueinander, die etwa durch die Art der Kontaktaufnahme und die Bedingungen, unter denen das Gespräch zustande gekommen ist, beeinflusst wurde (Van den Berg 2005). Derartige Informationen können über Feld- oder Interviewnotizen (sog. „Postskripte“, Witzel 1982) für die Sekundäranalyse zugänglich sein.

Soziales Handeln – und damit auch die Erhebungssituation – findet immer in einem institutionellen, kulturellen, sozio-politischen und historischen Kontext statt. Dieser *extrasituationalen Kontext* (oder auch „*Makro*“-Kontext) meint ein Hintergrundwissen, das über das lokale Gespräch und sein unmittelbares Setting hinausgeht. Doch auch dieser Kontext ist kein objektiver Satz von Umständen, der getrennt von den sozialen Akteuren vorliegt, sondern es geht um diejenigen Bestandteile des äußeren Kontextes, die sich empirisch manifestieren bzw. von den Beteiligten in der Interaktion tatsächlich aufgriffen werden. Wenn eine Studie sich etwa für bestimmte soziale Fragen und politische Debatten interessiert, ist es für die Sekundäranalyse von hohem Wert, diese Verknüpfung auch nachvollziehen zu können (z.B. über „graue Literatur“, Bishop 2006).

Nicht selten werden qualitative Daten in Kontexten erhoben, die durch eine *lokale Kultur* (Holstein/Gubrium 2004) charakterisiert sind. Dies kann beispielsweise eine Praxis oder (Fach-)Sprache sein, die innerhalb einer Institution, einer sozialen oder beruflichen Schicht oder eines geografischen Gebiets geteilt wird. Für Sekundäranalysen kann es daher entscheidend sein, den Zugang zu Dokumentationen zu haben, die die Daten in dieser elementaren Weise erst verständlich machen (z.B. Glossar eines Fachvokabulars).

Bishop (2006) ergänzt „*Projekt*“ als besonderen Teil der Situation, weil Forschungsprojekte spezifische Kontextmerkmale als eigenes Subset des Gesamtsettings einschließen. Heruntergebrochen auf die einzelne Situation bedeutet dies, dass Forschende einen (projekt-)spezifischen Erhebungskontext *produzieren*, vor dessen Hintergrund sich Feldinteraktionen vollziehen. Dies umfasst die methodischen Entscheidungen (wie die Wahl der Erhebungsmethode, des Forschungsdesigns), die theoretischen Vorannahmen, den institutionellen Hintergrund etc. Aufgrund der Kürze und Präzision, die Fachzeitschriften und Verlage erfordern, bieten die in Publikationen üblichen Kapitel zu Methodik und Durchführung einer Untersuchung keine ausreichende Erläuterung der methodischen Details. In Ergänzung sollte auf weitere (meist unveröffentlichte) Projektdokumente zurückgegriffen werden, die die wesentlichen Informationen über das Forschungsprojekt enthalten, wie etwa Anträge, Berichte, Leitfäden, Arbeitspapiere, Forschungstagebücher, Memos (zu den Bestandteilen einer Dokumentation siehe auch: Steinke 1999: 208-214).

### **13.3.3 Analysepotenzial der Daten**

Eine Grundprämisse der Sekundäranalyse ist, dass den im Rahmen eines spezifischen Forschungsprojekts erhobenen Daten genügend Potenzial innewohnt, um weitere Forschungsfragen zu bedienen. Die bislang veröffentlichten Beispiele zeigen, dass eine grundlegende Skepsis hinsichtlich des sekundären Analysepotenzials qualitativer Daten unbegründet ist.

Aufgrund der Offenheit (Hoffmann-Riem 1980) ihrer Erhebungsmethoden zeichnen sich qualitative Daten durch einen inhaltlichen Reichtum aus, der in einer ersten Analyse häufig unausgeschöpft bleibt und die Anwendung neuer Perspektiven fördert.

Die Nutzbarkeit von Daten hängt wesentlich mit ihrem Informationsgehalt zusammen. Auf einer allgemeinen Ebene bestimmt sich dieser durch die *Qualität* der Daten (-erhebung) (Flick, Kapitel 29 in diesem Band), also zum einen durch die Qualität des Erhebungsinstruments und zum anderen durch die Qualität der durch dieses Instrument erhaltenen Daten (Bergman/Coxon 2005). Für Sekundäranalysen muss also beurteilt werden, ob bei vorliegenden Daten die dem Gegenstand angemessenen Methoden ausgewählt und diese valide umgesetzt wurden und ob die auf den Gegenstand bezogenen Sicht- oder Handlungsweisen der Untersuchten in einer angemessenen Tiefe in den Daten repräsentiert sind.

Die Qualität der Daten vorausgesetzt bleibt zu prüfen, ob eine *Passung* der Daten für die konkrete Sekundäranalyse gegeben ist. Hierfür ist entscheidend, dass das Thema der Sekundäranalyse in der Originalstudie abgedeckt ist und deren Methoden die Analyse nicht einschränken.

Im Sinne einer praktischen Anleitung wurden in der quantitativen Forschungstradition Fragen formuliert, die im Rahmen einer Sekundäranalyse an die Daten gestellt werden sollten (Dale et al. 1988; Stewart/Kamins 1993). Folgende Fragen können auch für qualitative Sekundäranalysen übernommen werden (vgl. auch das „Assessment Tool“ in Hinds et al. 1997 sowie Heaton 2004.):

- Was ist die Zielsetzung der Studie und ihr konzeptioneller Rahmen?
- Welche Inhalte werden tatsächlich behandelt?
- Wie wurden die Daten erhoben (Methoden, Sampling)?
- Wann wurden die Daten erhoben (Aktualität)?
- Wer hat die Daten erhoben (Qualität)?

### 13.3.4 Verifikation und Validierung

Vor dem Hintergrund der in den letzten fünfzehn Jahren verstärkt geforderten Transparenz von Forschung und ihren Ergebnissen (DFG 1998; Kaase 1998) gerät die Sekundäranalyse – in der speziellen Form als Re-Analyse – als Instrument zur Validitätsprüfung ins öffentliche Blickfeld. Unter *Re-Analyse* wird dabei eine Sekundäranalyse gefasst, die die Daten erneut unter der gleichen Fragestellung analysiert.

Die jüngere Diskussion in der deutschsprachigen qualitativen Forschung greift diesen Punkt auf. Angesichts eines verschärften Wettbewerbs um Forschungsgelder sieht Reichertz (2007: 200) eine neue Notwendigkeit für die qualitative Forschung, ihre Anerkennung und praktische Relevanz im Wissenschaftsbetrieb zu behaupten. Die Archivierung und Bereitstellung für Sekundäranalysen sei in diesem Zusammenhang ein Instrument, um qualitative Forschung der intersubjektiven Nachvollziehbarkeit zugänglich zu machen und darüber wissenschaftliche Gütestandards qualitativer Forschung fest zu etablieren.

Methodologisch lassen sich berechtigte Einwände gegen das Validierungspotenzial von Re-Analysen vorbringen. Zum einen scheinen Re-Analysen bedenklich, wenn sie ein aus den Naturwissenschaften stammendes Ideal der Replikation auf die in den Sozialwissenschaften untersuchten sozialen und daher zu variablen Phänomene übertragen (Hammersley 1997). Zum anderen zeigt sich speziell für die qualitative Sozialforschung eine Formulierung und Einigung auf Gütekriterien letztlich (noch) nicht ausgemacht (vgl. Diskussion in Erwägen – Wissen – Ethik 18 (2), 2007).

Auch sind die bisherigen empirischen Beispiele (z.B. Gläser/Laudel 2000; Fielding/Fielding 2000) weniger tatsächliche Überprüfungen im Sinne eines schrittweisen Nachvollzugs des originären Forschungs- und Theoriebildungsprozesses. Sie zeigen, dass Re-Analysen zwar wertvolle Hinweise auf methodologische Probleme der Primärstudie geben können, aber kaum geeignete Mittel zur „echten Falsifizierung“ von Forschungsergebnissen darstellen. Vielmehr werden alternative oder neue theoretische Sichtweisen an den Daten entwickelt bzw. aufgedeckt, welche Themen in der Primäranalyse nicht erforscht wurden.

### **13.3.5 Sekundäranalyse und Forschungsethik**

Sekundäranalysen unterliegen den gleichen *datenschutzrechtlichen und ethischen Prinzipien*, wie sie allgemein für die Forschung gelten (Hopf 2005; Metschke/Wellbrock 2002; Friedrichs und Mühlichen, Kapitel 3 und 4 in diesem Band). Ein verantwortungsbewusster Umgang mit den Daten ist auch deshalb geboten, weil der Aufbau einer Vertrauensbeziehung in der qualitativen Forschung eine grundlegende Rolle spielt, um einen Zugang zur Innenperspektive der Forschungssubjekte zu erlangen. Darüber hinaus fühlen sich einige Primärforschende gerade durch den in der qualitativen Erhebungssituation gegebenen persönlichen Kontakt in einer herausgehobenen Stellung gegenüber ihren „Schützlingen“, so dass die (zentrale) Archivierung bzw. Sekundäranalysen durch Dritte als Einfallstor für den *Missbrauch der zugesagten Vertraulichkeit* empfunden wird (Medjedović 2007; Richardson/Godfrey 2003).

Daten bereitstellende Forscherinnen und Forscher sowie Datenarchive sind bemüht, durch das Ergreifen unterschiedlicher Maßnahmen den größtmöglichen *Schutz der Forschungssubjekte* zu sichern (Medjedović/Witzel 2010, Kap. 3.2). Doch auch wenn Archivnutzende etwa (faktisch) anonymisierte Daten nutzen, sollten sie sich ihrer Verantwortung gegenüber denjenigen, die sie zum „Objekt“ der Forschung erklären, bewusst sein und vertraulich mit diesen Daten umgehen.

Darüber hinaus tangiert der forschungsethische Aspekt die Wahrung der Interessen derer, die die Daten erhoben haben (Medjedović 2007). Daher sollten Sekundäranalysen unter der Bedingung stattfinden, dass (auch) Primärforschende einwilligen und ihr „*Urheberrecht*“ an den Daten in angemessener Weise (z.B. Verweis auf die Datenquelle bei Publikationen, Regelungen der Autorenschaft) berücksichtigt wird.

## Literatur

- Bergman, Manfred Max/Coxon, Anthony P.M. (2005): The Quality in Qualitative Methods. In: FQS 6 (2): Art. 34. URL: <http://nbn-resolving.de/urn:nbn:de:0114-fqs0502344> am 03.09.2009
- Bishop, Libby (2006): A Proposal for Archiving Context for Secondary Analysis. Methodological Innovations Online 1 (2): 10-20. URL: [http://erdt.plymouth.ac.uk/mionline/public\\_html/viewarticle.php?id=26](http://erdt.plymouth.ac.uk/mionline/public_html/viewarticle.php?id=26) am 01.09.2009
- Dale, Angela/Arber, Sara/Procter, Michael (1988): Doing Secondary Analysis. London: Unwin Hyman
- Deutsche Forschungsgemeinschaft (DFG) (1998): Sicherung guter wissenschaftlicher Praxis. Weinheim: Wiley
- Fielding, Nigel (2004): Getting the Most from Archived Qualitative Data. In: International Journal of Social Research Methodology 7 (1): 97-104
- Hammersley, Martyn (1997): Qualitative Data Archiving. In: Sociology 31 (1): 31-142
- Heaton, Janet (2004): Reworking Qualitative Data. London: Sage
- Hinds, Pamela/Vogel, Ralph/Clark-Steffen, Laura (1997): The Possibilities and Pitfalls of Doing a Secondary Analysis of a Qualitative Data Set. In: Qualitative Health Research 7 (3): 408-424
- Hoffmann-Riem, Christa (1980): Die Sozialforschung einer interpretativen Soziologie. In: KZfSS 32 (2): 339-372
- Holstein, James A./Gubrium, Jaber F. (2004): Context. In: Seale, Clive/Gombo, Giampietro/Gubrium, Jaber F./Silverman, David (Hg.): Qualitative Research Practice. London: Sage. 297-311
- Hopf, Christel (2005): Forschungsethik und qualitative Forschung. In: Flick, Uwe/Kardorff, Ernst von/Steinke, Ines (Hg.): Qualitative Forschung. Reinbek: Rowohlt. 589-600
- Huschka, Denis/Knoblauch, Hubert/Oellers, Claudia/Solga, Heike (Hg.) (2013): Forschungsinfrastrukturen für die qualitative Sozialforschung. Berlin: Scivero
- James, Jaquelyn B./Sørensen, Annemette (2000): Archiving Longitudinal Data for Future Research. In: FQS 1(3): Art. 23. URL: <http://nbn-resolving.de/urn:nbn:de:0114-fqs0003235> am 01.09.2009
- Kaase, Max (1998): Datendokumentation und Datenzugang in bei sozialwissenschaftlichen Fachzeitschriften eingereichten Beiträgen. In: Soziologie 2: 95-96
- Klingemann, Hans D./Mochmann, Ekkehard (1975): Sekundäranalyse. In: Koolwijk, Jürgen van/Wieken-Mayser, Maria (Hg.): Techniken der Empirischen Sozialforschung. Band 2. München: Oldenbourg. 178-194
- Kluge, Susann/Opitz, Diane (1999): Die Archivierung qualitativer Interviewdaten. In: Soziologie 4: 48-63
- Mauthner, Natascha S./Parry, Odette/Backett-Milburn, Kathryn (1998): The Data are Out There, or are They? In: Sociology 32 (4): 733-745

- Medjedović, Irena (2007): Sekundäranalyse qualitativer Interviewdaten. In: *Journal für Psychologie* 15 (3). URL: <http://www.journal-fuer-psychologie.de/jfp-3-2007-6.html> am 01.09.2009
- Medjedović, Irena (2014): Qualitative Sekundäranalyse. Zum Potenzial einer neuen Forschungsstrategie in der empirischen Sozialforschung. Wiesbaden: Springer VS
- Medjedović, Irena/Witzel, Andreas (2010): Wiederverwendung qualitativer Daten. Wiesbaden: VS Verlag
- Metschke, Rainer/Wellbrock, Rita (2002): Datenschutz in Wissenschaft und Forschung. URL: <http://www.datenschutz-berlin.de/attachments/47/Materialien28.pdf?1166527077> am 01.09.2009
- Moore, Niamh (2006): The Contexts of Context. In: *Methodological Innovations Online* 1 (2): 21-32. URL: [http://erdt.plymouth.ac.uk/mionline/public\\_html/viewarticle.php?id=27](http://erdt.plymouth.ac.uk/mionline/public_html/viewarticle.php?id=27) am 01.09.2009
- Parry, Odette/Mauthner, Natasha S. (2004): Whose Data are They Anyway? In: *Sociology* 38 (2): 139-152
- Reichertz, Jo (2007): Qualitative Sozialforschung. In: *EWE* 18 (2): 195-208
- Richardson, Jane C./Godfrey, Barry S. (2003): Towards Ethical Practice in the Use of Archived Transcribed Interviews. In: *International Journal of Social Research Methodology* 6 (4): 347-355
- Steinke, Ines (1999): Kriterien qualitativer Forschung. Weinheim: Juventa
- Stewart, David W./Kamins, Michael A. (1993): Secondary Research. Newbury Park: Sage
- Thomson, Denise/Bzdel, Lana/Golden-Biddle, Karen/Reay, Trish/Estabrooks, Carole A. (2005): Central Questions of Anonymization. In: *FQS* 6(1): Art. 29. URL: <http://nbn-resolving.de/urn:nbn:de:0114-fqs0501297> am 01.09.2009
- Thorne, Sally (1994): Secondary Analysis in Qualitative Research. In: Morse, Janice M. (Hg.): *Critical Issues in Qualitative Research Methods*. London: Sage. 263-279
- Van den Berg, Harry (2005): Reanalyzing Qualitative Interviews from Different Angles. In: *FQS* 6 (1): Art. 30. URL: <http://nbn-resolving.de/urn:nbn:de:0114-fqs0501305> am 01.09.2009
- WR (Wissenschaftsrat) (2011): Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. Drs. 10465-11. URL: <http://www.wissenschaftsrat.de/download/archiv/10465-11.pdf> am 02.08.2013
- WR (Wissenschaftsrat) (2012): Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020. Drs. 2359-12. <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf> am 02.08.2013
- Witzel, Andreas (1982): Verfahren der qualitativen Sozialforschung. Frankfurt a.M.: Campus
- Witzel, Andreas/Medjedović, Irena/Kretzer, Susanne (Hg.) (2008): Sekundäranalyse qualitativer Daten. *HSR* 33 (3): 10-32

Ekkehard Mochmann

Die Entwicklung der empirischen Sozialforschung ist untrennbar mit dem Fortschritt ihrer Datenbasis verbunden. Obwohl die ersten Untersuchungen bereits im 19. Jahrhundert durchgeführt wurden, waren über lange Zeit die Forschungsmöglichkeiten durch eine karge Datenlage geprägt. Dieses intensiv wahrgenommene Defizit wurde bald zum Motor der Bemühungen, die Datenlage systematisch zu verbessern. Dies gilt nicht nur für die Menge der verfügbaren Daten sondern auch für ihre Qualität und Vergleichbarkeit. Der Zugang zu bereits vorliegenden Forschungsdaten erlaubte die Auswertung auch unter neuen Forschungsfragen (Sekundäranalyse) und erschloss neben weiteren Querschnittsanalysen zugleich eine Datenbasis mit Potential für die Analyse gesellschaftlichen Wandels.

---

## 14.1 Charakterisierung von Querschnittstudien und Trendstudien

Die empirische Sozialforschung verfügt heute über ein reiches Repertoire an Untersuchungsmethoden und Instrumenten. Maßgebend für die Auswahl des geeigneten Verfahrens für das Forschungsprojekt ist die jeweilige Forschungsfrage.

Soll zu einem bestimmten Zeitpunkt die Verteilung von Merkmalen, wie z.B. Einstellungen zu aktuellen politischen Themen und Wahlpräferenzen analysiert werden, dann ist die Querschnittsstudie bei einer Stichprobe der Wahlberechtigten das geeignete Verfahren. Kennzeichnend für *Querschnittstudien* ist, dass die Daten mit einem Erhebungsinstrument nur zu einem Zeitpunkt bei einer Stichprobe der Grundgesamtheit (Häder/Häder, Kapitel 18 in diesem Band) erhoben werden. Zum Beispiel werden in einer standardisierten Befragung (Reinecke, Kapitel 44 in diesem Band) neben den inhaltlich interessierenden Fragen soziodemographische Merkmale der Befragten (Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band) in der Stichprobe erfasst. So können zum jeweiligen Zeitpunkt Einstellungen zu den aktuellen politischen Fragen in Abhängigkeit von Merkmalen wie Alter, Schulbildung oder Einkommen erforscht werden. Die Ergebnisse haben überwiegend deskriptiven Charakter.

Richtet sich die Forschungsfrage hingegen auf Entwicklungen von Verhalten und Einstellungen über Zeit, dann sind Trendstudien das zielführende Verfahren. Kennzeichnend für *Trendstudien* ist, dass das Erhebungsinstrument mit vergleichbaren Fragen bei unterschiedlichen Stichproben derselben Grundgesamtheit zu mehreren Zeitpunkten abgefragt wird und die Antwortverteilungen über die Erhebungszeitpunkte verglichen werden. So werden z.B. zur Analyse des Wahlverhaltens über Zeit jeweils nach den Wahlen die gleichen Fragen erneut gestellt und die Antwortverteilungen mit den Ergebnissen früherer Studien verglichen. Da jedes Mal eine Stichprobe neu gezogen wird, werden für gewöhnlich zu jedem Zeitpunkt *unterschiedliche* Personen befragt. Dabei beruhen die Stichproben in der Regel auf einer repräsentativen Zufallsauswahl (Häder/Häder, Kapitel 18 in diesem Band).

Entsprechendes gilt auch für andere Datenerhebungsverfahren wie Analysen von Daten der amtlichen Statistik und anderen Verwaltungsdaten (Hartmann/Lengerer, Kapitel 71 in diesem Band), Inhaltsanalysen (Mayring/Fenzl, Kapitel 38 in diesem Band) oder Beobachtungsstudien (Thierbach/ Petschick, Kapitel 66 in diesem Band), wo das Kategorien- oder Beobachtungsschema zu den entsprechenden Zeitpunkten eingesetzt wird. Komplexere Studien sind mit Panels (Schupp, Kapitel 73 in diesem Band) und Kohortenstudien (Pötter/Prein, Kapitel 86 in diesem Band) gegeben, bei denen *dieselben* Personen bzw. Kohorten wiederholt befragt werden.

Mit zunehmender Verfügbarkeit von Querschnittsstudien verbessern sich die Möglichkeiten, mit anderen bereits vorliegenden Daten oder speziell unter der betreffenden Thematik erhobenen Daten *ex post* *Trendstudien* zu erstellen und damit alten Studien ein „neues Leben“ zu geben.

---

## 14.2 Zur Entwicklung von Querschnitterhebungen und Trendstudien

In Deutschland entstand die Grundidee zur Durchführung sozialer Umfragen durch die Zusammenarbeit von Repräsentanten der Verwaltungsbürokratie und „Kathedersozialisten“, die eine gemeinsame Basis im Verein für Socialpolitik hatten. Zwar gab es schon früher regional begrenzte Erhebungen im Auftrag von Regionalbehörden, aber die erste landesweite Umfrage in Deutschland wurde auf Beschluss des Bundesrates vom 19. Februar 1875 im Jahr 1876 von den Verwaltungen der Bundesländer durchgeführt. Sie sollte den Zustand des Lehrlings- und Gesellenwesens im Handwerk sowie die Beziehungen zwischen Arbeitern und Arbeitgebern in der Großindustrie prüfen. Dazu konnte man englische, französische und belgische Enqueten als Beispiele heranziehen. Skizzenhaften Aufzeichnungen zum Vorgehen bei der deutschen Studie ist zu entnehmen, dass Beamte in 559 Orten Sprechstage hielten, wo sie etwa 7.000 Menschen mündlich befragten. Die wohl interessanteste Frage zu den Arbeiterverhältnissen war, ob die Fabrikordnungen häufig Bestimmungen enthielten, die für den Arbeitgeber günstigere Bedingungen vorsahen als für die Arbeitnehmer, und ob in diesen oder anderen Punkten gesetzliche Beschränkungen empfehlenswert seien. Obwohl die Umfrage in Deutschland von Fachleuten im

Verein für Socialpolitik konzipiert und ausgewertet wurde, kam es nicht zu einer breiteren Diskussion von Methodenfragen und einer Weiterentwicklung als Instrument der Sozialforschung, für die sich insbesondere Max Weber wiederholt einsetzte (Oberschall 1997: 40 f.). Die Aufmerksamkeit konzentrierte sich vielmehr auf praktische Fragen der Durchführung- sollte man ausführliche stenographische Berichte aufnehmen oder nur Zusammenfassungen? Sollte an einem Ort nur ein Rechercheur eingesetzt werden oder mehrere? Und wie sollten diese geschult werden? Dagegen wurden weitergehende Überlegungen zur Fragebogenkonstruktion und darüber, was mit den gewonnenen Information geschehen sollte, nicht angestellt (Oberschall 1997: 43). Aus heutiger Sicht ist die damals schon vorhandene internationale Orientierung an belgischen, französischen und englischen Enquêtes mit ihren Fragestellungen hervorzuheben, doch es sollte Jahrzehnte dauern bis Umfragen kontinuierlich mit international vergleichender Perspektive durchgeführt wurden. Das neue Instrument der Forschung wurde bei potenziellen Geldgebern skeptisch betrachtet. So mussten die innovativen Ideen erst nach USA exportiert werden, um dort für den Masseneinsatz weiterentwickelt zu werden. Nach dem zweiten Weltkrieg wurden die Instrumente aus den USA re-importiert (Lazarsfeld 1997: 17), wo sie sich zunehmend in repräsentativen Bevölkerungsumfragen bewährt hatten.

Die Entwicklung der empirischen Sozialforschung war nach dem zweiten Weltkrieg von einem hohen Datenbedarf geprägt, der konträr zur tatsächlichen Verfügbarkeit solcher Daten stand. Sehr schnell entstanden nicht nur in Nordamerika sondern auch in Westeuropa Markt- und Sozialforschungsinstitute. In Ergänzung der amtlichen Statistik wurden zu interessierenden Themen Querschnittsstudien durchgeführt. Die Daten wurden bei einer Stichprobe einmal erhoben und eigneten sich so für explorative Studien oder für die Beschreibung des Zustandes der Gesellschaft zum jeweiligen Zeitpunkt. Die damals in statistischen Ämtern noch vorherrschende Skepsis gegenüber Stichprobenerhebungen und ihrer Aussagekraft für die Gesamtheit machte zunehmender Akzeptanz dieser Informationsquellen in der Wirtschaft und der Bevölkerung Platz, und soweit noch vorhanden, auch in den auf Vollerhebungen fokussierten statistischen Ämtern (Scheuch 1967: 728). Hinzu kam eine gewisse Neugier auf das neue Instrument der Meinungsforschung. Dies alles förderte die Produktion von Umfragedaten.

Nach gut informierten Schätzungen wurden 1951 in West-Europa über 500.000 Interviews jährlich erhoben, Ende der 50er-Jahre waren es bereits 4 bis 5 Millionen (Rokkan 1957: 1). Die Daten wurden meistens für kommerzielle Zwecke verwendet und lediglich tabellarisch ausgewertet. Für detailliertere Analysen waren sie nicht zugänglich, da die Datenträger, damals Lochkarten, mangels geeigneter und ausreichender Aufbewahrungsmöglichkeiten binnen Jahresfrist vernichtet wurden. So blieb die Datenlage für die akademisch verfasste Sozialforschung prekär, obwohl seit beinahe 100 Jahren empirische Sozialforschung in Europa betrieben wurde.

### **14.3 Grundlegung einer Datenbasis für die empirische Sozialforschung**

Zunächst standen die frisch erhobenen Daten allein den Erhebungsinstituten und den Primärforschern zur Verfügung, die sie meist nur für ihre Fragestellungen auswerteten. Bereits in den 1950er-Jahren wurde das Potential dieser Daten für weitergehende, insbesondere historische und vergleichende wissenschaftliche Analysen erkannt (Rokkan 1957, 1964: 49-62, Scheuch/Brüning 1964: 77-85). Die Studien der Markt- und Sozialforschung enthielten Daten, die über die Beantwortung kommerzieller Fragestellungen hinaus auch reichhaltige Informationen für die akademisch verfasste Sozialforschung lieferten, so z.B. für die Erforschung des Konsum- und Freizeitverhaltens oder der politischen Einstellungen. Hinzu kam die Erkenntnis, dass einmal erhobene Daten nach der Erstauswertung auch für weitere Forschungsfragen ausgewertet werden können. Dank der wachsenden Bereitschaft einiger Erhebungsinstitute, ihre Daten in Originalform der Wissenschaft zur Verfügung zu stellen, konnten nach 1960 zunehmend detaillierte Beschreibungen des gesellschaftlichen Lebens und der Einstellungen auf Basis repräsentativer Bevölkerungs-Umfragen gewonnen werden. Schließlich wurden auch Umfragen von der akademisch verfassten Sozialforschung bei den Erhebungsinstituten in Auftrag gegeben. Diese ergänzten die soziostrukturrellen Bilder der Gesellschaft, die auf Basis der Erhebungen statistischer Ämter in aggregierter Form veröffentlicht wurden (Scheuch 1967: 655-685).

Noch Mitte der 1970er-Jahre sind die meisten Studien auf einen Zeitpunkt beschränkt und kommen somit nicht über einfache statische Analysen hinaus. Eine prominente Ausnahme ist z.B. die Kölner Wahlstudie 1961 mit Vor- und Nachwahlstudien (Scheuch/Wildenmann 1965). Zunehmend wurden dann dynamische Analysen mit zeitlicher Tiefendimension gefordert, bei denen eine Variablenauswahl aus den Querschnitten über die Zeit verglichen wird, indem die Zunahme, Abnahme oder Konstanz der Werte festgestellt wird.

Die wachsende Zahl von Studien mit zeitlicher Dimension wurde in dieser Zeit hauptsächlich vier Einflussfaktoren zugeschrieben: einer wachsenden Zahl vergleichbarer oder annähernd vergleichbarer Daten, der Sozialindikatorenbewegung, einem wachsendem Interesse an sozialem Wandel und sozialen Prozessen und schließlich dem verbesserten Zugang zu bereits vorliegenden Daten. In dieser Zeit werden Probleme mangelnder Vergleichbarkeit der Erhebungsinstrumente bei der Analyse oft nicht ausreichend berücksichtigt, und die Überlegenheit longitudinaler Studien über Querschnittsstudien zur Aufdeckung kausaler Beziehungen wird häufig überschätzt (Glenn/Frisbie 1977).

---

### **14.4 Datenarchive und internationale Dateninfrastruktur**

Mit wachsendem Datenbedarf stellte sich die Aufgabe, den Zugang zu den Daten zu organisieren und die Daten so zur Verfügung zu stellen, dass bei späteren Auswertungen die Forscher alle relevanten Informationen über die ursprüngliche Forschungsfrage, das Studiendesign (Reinecke, Stein, Kapitel 44 und 7 in diesem Band), die Stichprobe (Häder/

Häder Kapitel 18 in diesem Band), den Fragebogen (Klöckner/Friedrichs, Kapitel 49 in diesem Band) und die Feldarbeit erhalten konnten, um einen ähnlichen Informationsstand und Vertrautheit mit den Daten herzustellen, wie sie beim Primärforscher gegeben waren. Wesentliche Impulse kamen aus der international vergleichenden Sozialforschung, die zur Gründung von sozialwissenschaftlichen Datenarchiven und ihrer internationalen Vernetzung führten. Diese Ideen wurden in den Standing Committees on „Comparative Research“ und „Social Science Data“ des International Social Science Council (ISSC) in Paris insbesondere von Stein Rokkan, Warren Miller, Erwin K. Scheuch und Alexander Szalai entwickelt und vorangetrieben (Scheuch 2003, Mochmann 2008).

Mit dem Roper Center wurde 1957 in den USA das erste Datenarchiv der Welt gegründet, 1960 folgte als erstes Archiv in Europa das Zentralarchiv für empirische Sozialforschung an der Universität zu Köln, heute Datenarchiv für Sozialwissenschaften des GESIS-Leibniz-Instituts für Sozialwissenschaften (GESIS-DAS) in Köln. Weitere Archivgründungen folgten in den westlichen Industriegesellschaften. Heute verfügen die meisten Länder der Europäischen Gemeinschaft und weitere osteuropäische Länder über ein sozialwissenschaftliches Datenarchiv.

Im Jahre 1976 wurde der europäische Verbund sozialwissenschaftlicher Datenarchive gegründet, der sich dann als Council of European Social Science Data Archives (CESSDA) etablierte. Insbesondere zur Unterstützung der transatlantischen Zusammenarbeit, zunehmend auch unter Einbeziehung von Datenserviceeinrichtungen in Australien, Südafrika und Asien wurde 1977 die International Federation of Data Organisations for the Social Sciences (IFDO) gegründet ([www.ifdo.org/](http://www.ifdo.org/)). Als Organisation für die „data professionals“ entstand 1974 IASSIST, die International Association for Social Science Information Services and Technologies, um Forschung und Lehre in den Sozialwissenschaften zu unterstützen ([www.iassistdata.org/](http://www.iassistdata.org/)).

Die sozialwissenschaftlichen Datenarchive haben weltweit erheblich zur Herausbildung einer Kultur des „data sharing“ beigetragen, die die Ausbildung in fortgeschrittenen Techniken der Datenanalyse unterstützt und Standards für die Dokumentation, Aufbereitung, langfristige Speicherung und Bereitstellung sozialwissenschaftlicher Daten entwickelt. Jährlich werden derzeit über CESSDA rund 40.000 Datensätze der sozialwissenschaftlichen Community in Europa zur Verfügung gestellt. Darüber hinaus unterstützt CESSDA die Forscher beim Datenmanagement. Viele europäische Forschungsprojekte kooperieren mit einem der nationalen CESSDA-Archive, um die notwendigen Arbeiten der Datendokumentation und der Datenarchivierung für ihre Projekte durchzuführen. Inzwischen sind diese Aufgaben als essentiell für die sozialwissenschaftliche Dateninfrastruktur anerkannt. CESSDA wurde vom European Strategy Forum on Research Infrastructures (ESFRI) als ein European Research Infrastructure Center (CESSDA-ERIC) in die erste „European Roadmap for Research Infrastructures“ aufgenommen und 2013 in Bergen (Norwegen) als CESSDA-AS institutionalisiert ([www.cessda.net/](http://www.cessda.net/)).

## 14.5 Von der Sekundäranalyse einzelner Studien zu kontinuierlichen Datenerhebungsprogrammen

Bis Ende der siebziger Jahre war die Sozialforschung für Sekundäranalysen, die erneute Auswertung für andere als die ursprünglichen Forschungsfragen, auf vorliegende Einzelstudien angewiesen (Klingemann/Mochmann 1975). Die Auswertungsmöglichkeiten insbesondere für Längsschnittanalysen waren aber häufig sehr eingeschränkt, da es an Vergleichbarkeit der Fragestellungen und Antwortkategorien in den verschiedenen Umfragen mangelte. Zudem gab es Differenzen bei der Zusammensetzung der Stichproben, mitunter waren sie unterschiedlichen Grundgesamtheiten zuzuordnen.

Zur Verbesserung der Datenlage begannen die Datenarchive mit der Erstellung von Übersichten mit vergleichbaren Frageformulierungen über verschiedene Erhebungszeitpunkte (continuity guides) und konzentrierten erhebliche Ressourcen auf die Harmonisierung und Integration von vorliegenden Daten in kumulierten Analysedateien. Die damit verbundene Methodendiskussion führte zu weiteren Verbesserungen bei den Frageformulierungen und sensibilisierte für die Grenzen der Interpretierbarkeit der Fragestellungen unter einem anderen als dem ursprünglichen Bezugsrahmen. Die Notwendigkeit einer umfassenden Dokumentation und Aufbereitung der Daten führte zur Aufnahme von Methoden- und Feldberichten. Relevante methodische und technische Informationen wie Primärforscher, Titel der Studie, Erhebungsverfahren, Art und Umfang der Stichprobe, Struktur des Datensatzes etc. werden in einer standardisierten Studienbeschreibung erfasst, die Messinstrumente mit den Frageformulierungen und Merkmalsausprägungen werden in Codebüchern dokumentiert. Zugleich kommt diesen Metadaten, die jeden Datensatz in seinen verschiedenen Aspekten beschreiben, besondere Bedeutung für die Wiederauffindbarkeit zu, da diese Informationen in verteilten Datenbanksystemen der Archive unter Wahrung der Datenschutzinteressen der Befragten für Recherchen und weitere Auswertungen zur Verfügung stehen, z.B. der Datenkatalog von CESSDA ([www.cessda.net/catalogue/](http://www.cessda.net/catalogue/)) oder die Datenbestände der GESIS ([www.gesis.org/unser-angebot/recherchieren/datenbestandskatalog/](http://www.gesis.org/unser-angebot/recherchieren/datenbestandskatalog/)). Zudem konnte vielfach die Datenlage durch Verknüpfung (record linkage) mit Aggregatdaten, z.B. zur Wirtschaftslage, zum Wahlsystem oder zur kleinräumigen Charakterisierung angereichert werden.

Zur eindeutigen Identifizierung bietet GESIS in Zusammenarbeit mit dem Leibniz-Informationszentrum Wirtschaft (ZBW) den Service der Registrierungsagentur da|ra zur Registrierung von Sozial- und Wirtschaftsdaten in Deutschland an ([www.da-ra.de/](http://www.da-ra.de/)). Die Vergabe von Digital Object Identifiers (DOI-Namen) schafft die Voraussetzungen für eine dauerhafte Identifizierung, Lokalisierung und verlässliche Zitierbarkeit von Forschungsdaten. Durch die Vergabe von „persistent identifiers“ wird die Verknüpfung von Daten und zugehöriger Literatur in den Netzwerken dauerhaft möglich (<http://www.da-ra.de/de/home/>).

Während einzelne international vergleichende Studien, wie z.B. die „Civic Culture Study“ (Almond/ Verba 1963), die internationale Zeitbudgetstudie (Szalai 1972) oder die Political Action Studie (Barnes et al. 1979) hohe Sichtbarkeit gewannen, blieb ein frühes

Projekt zur kontinuierlichen Gesellschaftsbeobachtung lange Zeit unbemerkt. Die United States Information Agency (USIA) begann ihre Datenerhebungen in Frankreich, Italien, Großbritannien und Westdeutschland 1952 und setzte sie bis 1976 fort (Merrit 1968). Viele Studien aus der Nachkriegszeit sind heute in aufbereiterter Form (Kaase/Newton 1995) über das Datenarchiv der GESIS in Köln zugänglich.

Im Frühjahr 1970 begann Jacques-René Rabier in Kooperation mit Ronald Inglehart die erste Studie in der Europäischen Gemeinschaft in Belgien, Frankreich, Italien, Holland und Deutschland. Diese Studie war der Anfang der Eurobarometer, die seither zweimal im Jahr durchgeführt werden. Die Kontinuität der Erhebung mit vergleichbaren Fragestellungen ermöglicht es längerfristige Trends und Veränderungen in den Einstellungen zu erkennen. Hinzu kommen die Flash-Eurobarometer, sowie die Osteuropa- und Kandidatenländer- Eurobarometer und Spezial- Eurobarometer zu zeitgemäß besonder interessierenden Fragestellungen, wie z.B. zu „Energie und Umwelt“ oder „Wissenschaft und Technologie“. Die Europäische Kommission stellt die erhobenen Daten nach Erstauswertung für die eigene politische Orientierung der Wissenschaft für weitere Auswertungen zur Verfügung. Sie werden von GESIS und dem Inter University Consortium for Political and Social Research (ICPSR) in Michigan, Ann Arbor ([www.icpsr.umich.edu/](http://www.icpsr.umich.edu/)) aufbereitet und der Wissenschaft weltweit über das internationale Archivnetz zur Verfügung gestellt. Damit sind die Eurobarometer bezüglich Datenaufbereitung und Vertrieb zum Modell für andere kontinuierliche Datenerhebungsprogramme geworden. Die Wissenschaft hat inzwischen eine Reihe kontinuierlicher Datenerhebungsprogramme etabliert. Dazu gehören der „European Values Survey“ ([www.europeanvaluesstudy.eu/](http://www.europeanvaluesstudy.eu/)) und die „World Values Surveys“ ([www.worldvaluessurvey.org/](http://www.worldvaluessurvey.org/)), die erstmals 1981 erhoben wurden, das „International Social Survey Programme“ (ISSP) seit 1985, die nationalen Wahlstudien zum Bundestag und den Landesparlamenten im „Forschungsdatenzentrum Wahlen“ der GESIS ([www.gesis.org/das-institut/kompetenzzentren/fdz-wahlen/wahlstudien/](http://www.gesis.org/das-institut/kompetenzzentren/fdz-wahlen/wahlstudien/)), sowie die „Comparative Study of Electoral Systems“ ([www.cses.org/](http://www.cses.org/)).

In einem europaweiten Sekundäranalyseprojekt, das von der European Science Foundation unterstützt die Einstellungen zur und das Vertrauen in die Regierungen untersuchte (Beliefs in Government BIG – Projekt) wurden die Bestände der CESSDA-Archive ausgewertet. Obwohl das Projekt mit fünf umfangreichen Bänden sehr ertragreich war, zeigten sich doch auch die Grenzen der Vergleichbarkeit vorliegender Daten, die nicht bereits als komparative Studien konzipiert waren und häufig auch nicht ex post harmonisiert werden konnten (Kaase/Newton 1995). Aus dieser Einsicht entsprang der Plan für einen „European Social Survey (ESS)“. Dieser wird derzeit mit einem einheitlichen Fragebogen als repräsentative Bevölkerungsumfrage zu verschiedenen Problemen des politischen und gesellschaftlichen Zusammenlebens in über 30 europäischen Ländern durchgeführt ([www.europeansocialsurvey.org/](http://www.europeansocialsurvey.org/)). Dabei wird besonderer Wert auf die Einhaltung möglichst hoher methodischer Standards der Umfrageforschung gelegt. Seit 2002/2003 finden die Erhebungen in Zweijahresintervallen statt.

Der ESS misst Einstellungen, Wertorientierungen und Verhaltensmuster der Bevölkerung in zahlreichen Gesellschaften Europas. Der Fragebogen besteht aus über die Wellen

gleichbleibenden Fragen (sogenannte Kernmodule), aus verschiedenen Wechselmodulen, die sich auf spezielle Themen beziehen und von Welle zu Welle variieren, sowie aus einem zusätzlichen Teil, der experimentelle Testfragen und Fragen zum individuellen Wertesystem enthält.

Die Kernmodule enthalten Fragen zu Themen wie: Mediennutzung, Politik, Vertrauen in Institutionen, Immigration, Angst vor Verbrechen, Gesundheit, Religion, Wohlbefinden, individuelle Wertorientierungen und Demographie. Die Wechselmodule befassen sich bisher mit folgenden Themen: Immigration, Citizenship und Demokratie (ESS 1, 2002), Wirtschaftsmoral, Arbeit, Familie und Wohlbefinden, Gesundheit und Pflege (ESS 2, 2004), Lebensplanung, persönliches und soziales Wohlbefinden (ESS 3, 2006). Altersdiskriminierung und Einstellungen zum Wohlfahrtsstaat (ESS 4, 2008), Vertrauen in Strafjustiz und Polizei, Arbeit und Familie (ESS 5, 2010), Verständnis und Bewertungen der Demokratie, persönliches und soziales Wohlbefinden (ESS 6, 2012) und Einstellungen zu Immigration, gesundheitliche Ungleichheiten (ESS 7, 2014). Mit diesen Daten können Stabilität und Wandel in der gesellschaftlichen Struktur, in den Lebensumständen und Einstellungen der Bürger in Europa analysiert werden, zugleich dienen sie der Erklärung von Veränderungen in Europas politischer, sozialer und moralischer Verfasstheit. Die Daten werden vom norwegischen sozialwissenschaftlichen Datenservice (NSD) aufbereitet und dokumentiert. Sie stehen für wissenschaftliche Analysen im Internet zur Verfügung ([www.europeansocialsurvey.org/data](http://www.europeansocialsurvey.org/data)).

In Deutschland erhebt „Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) aktuelle Daten über Einstellungen, Verhaltensweisen und Sozialstruktur der Bevölkerung in der Bundesrepublik Deutschland. Seit 1980 wird alle zwei Jahre ein repräsentativer Querschnitt der Bevölkerung in persönlichen Interviews mit einem teils konstanten, teils variablen Fragenprogramm befragt. Die Daten stehen unmittelbar nach ihrer benutzungsgerechten Aufbereitung und Dokumentation über GESIS allen Interessenten für Forschung und Lehre zur Verfügung ([www.gesis.org/allbus/datenzugang](http://www.gesis.org/allbus/datenzugang)). Die einzelnen Querschnittsdatensätze, die jeweils einen eigenen inhaltlichen Schwerpunkt haben, dienen der Untersuchung von Einstellungen und Verhaltensweisen der deutschen Bevölkerung. Durch die Replikation von Fragekomplexen aus dem ALLBUS und aus anderen Studien wird die Analyse von Entwicklungsprozessen und gesellschaftlichem Wandel ermöglicht. Für die Analyse von Trends steht ein kumulierter Datensatz zur Verfügung, der sämtliche Merkmale enthält, die mindestens zweimal im ALLBUS erhoben wurden.

Die Kooperation mit dem US-amerikanischen General Social Survey (GSS) und mit Forschergruppen aus weiteren Ländern, wie z.B. England (British Social Attitudes Survey) oder Japan (JGSS), und insbesondere die Einbindung des International Social Survey Programme (ISSP) (<http://www.issp.org/>), bieten die Möglichkeit, international vergleichende Gesellschaftsanalysen durchzuführen. Methodische Fragestellungen können zum Teil mit den ALLBUS-Umfragen selbst untersucht werden (z.B. Tests auf Reihenfolgeeffekte oder Effekte unterschiedlicher Frageformulierungen durch den Einsatz von Split-Versionen), zum Teil anhand begleitend durchgeföhrter Methodenstudien, für die gesonderte

Datensätze im GESIS-Datenarchiv verfügbar sind (z.B. zu Interviewereffekten, zur Test-Retest-Reliabilität, zur Nonresponse-Problematik), GESIS stellt ausführliche Dokumentationen zur Methode, zu Metadaten und zu den Forschungsdaten zur Verfügung ([www.gesis.org/allbus/](http://www.gesis.org/allbus/)).

Neben diesen Daten der quantitativen Sozialforschung werden nun auch Daten der qualitativen Forschung angeboten. Dieser Service wurde zunächst am englischen Datenarchiv im Qualidata Project entwickelt ([www.esds.ac.uk/qualidata/about/introduction.asp](http://www.esds.ac.uk/qualidata/about/introduction.asp)). In Deutschland gibt es an der Universität Bremen das erste Archiv für die qualitative Forschung. Das Archiv für Lebenslaufforschung (ALLF) ist Teil des eScience Center an der Universität Bremen. Es stellt umfangreiches digitalisiertes Datenmaterial aus empirischen Forschungsprojekten und dazugehörige Dokumentationen zur Sekundärnutzung in Forschung und Lehre zur Verfügung (Medjedović/Witzel 2010). Mit bislang etwa 700 qualitativen, dokumentierten und anonymisierten Interviews, die zum großen Teil aus Längsschnittuntersuchungen des Sonderforschungsbereiches (Sfb) 186 „Statuspassagen und Risikolagen im Lebenslauf“ (1988-2001, DFG-finanziert) zu verschiedenen Übergängen und Statuspassagen im Lebenslauf stammen, ist es wohl das bundesweit größte Archiv qualitativer Interviewdaten aus den Sozialwissenschaften ([www.lebenslaufarchiv.uni-bremen.de/](http://www.lebenslaufarchiv.uni-bremen.de/)).

Nachdem in den vergangenen Jahrzehnten die Weiterentwicklung des Datenanalyse-Instrumentariums im Vordergrund stand, wird zunehmend Gewicht auf die Qualität der Datenerhebung und den Zustand der Daten gelegt (Kaase 1999, Blasius/Thiessen 2012, RatSWD 2014). Ziel ist es, kostenintensive Untersuchungen methodisch kontrolliert durchzuführen und damit eine Datenqualität zu fördern, die auch beim Einsatz anspruchsvoller Analysen zu validen Ergebnissen führt und kritischen Überprüfungen standhält.

Konnten früher die Primärforscher ein Auswertungsprivileg für die erhobenen Daten beanspruchen, so wurde die unmittelbare Bereitstellung der Daten für alle interessierten Forscher nach Aufbereitung und Freigabe durch die Archive zur Regel. Damit wurde faktisch praktiziert, was in den letzten Jahren intensiv unter „Open Access to Research Data“ diskutiert wird: die Freigabe öffentlich finanzierter Daten für weitere Auswertungen durch andere Forscher. Die Open-Access-Bewegung hat die freie Zugänglichkeit von Informationen wesentlich gefördert und auch den Zugang zu Forschungsdaten explizit in die Forderungen aufgenommen, so die Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen im Jahr 2003 ([http://openaccess.mpg.de/3515/Berliner\\_Erklarung](http://openaccess.mpg.de/3515/Berliner_Erklarung)). Das Internet hat die Arbeitsweise auch der Sozialwissenschaftler in vieler Hinsicht verändert. Die Bereitstellung von Wissen, bibliographischen Nachweisen und Forschungsdaten ist wesentlich vereinfacht und beschleunigt. Benutzerfreundlich aufbereitete Forschungsdaten stehen für den interaktiven Zugriff in virtuell verteilten Datenbanken weltweit zur Verfügung. Zunehmende Interoperabilität der Systeme erlaubt den simultanen Zugriff auf Daten aus unterschiedlichen Quellen und deren vergleichende Analyse. Damit wird die Einbeziehung von Daten anderer Disziplinen und eine problembezogene integrierte Auswertung erleichtert (Mochmann: 2010).

Folgt man Visionen für globale Infrastrukturen im Jahre 2020, wie sie derzeit in europäischen Großprojekten erarbeitet werden, dann sind in diesem Bereich auch für sozialwissenschaftliche Dateninfrastrukturen dynamische Entwicklungen zu erwarten (<http://www.grdi2020.eu>).

## Literatur

- Almond, Gabriel A./Verba, Sydney (1963): The Civic Culture. Political Attitudes and Democracy in Five Nations. Princeton: Princeton University Press
- Barnes, Samuel H./Kaase, Max et al. (1979): Political Action. Mass Participation in Five Western Democracies. Beverly Hills/London: Sage
- Blasius, Jörg /Thiessen, Victor (2012): Assessing the Quality of Survey Data. London: Sage
- Glenn, Norval D./Frisbie, Parker W. (1977): Trend Studies with Survey Sample and Census Data. In: *Annual Review of Sociology* 3: 79-104
- Kaase, Max (Hrsg.) (1999): Qualitätskriterien der Umfrageforschung, Denkschrift. Berlin: Akademie Verlag
- Kaase, Max/Newton, Kenneth (Hrsg.) (1995): Beliefs in Government. Oxford: Oxford University Press
- Klingemann, Hans-Dieter/Mochmann, Ekkehard (1975): Sekundäranalyse. In: Koolwijk/Wieken-Mayser (Hg.): 178-195
- König, René (1967): Handbuch der empirischen Sozialforschung, Band 1. Stuttgart: Ferdinand Enke Verlag
- Koolwijk, Jürgen van/Wieken-Mayser, Maria (Hg.) (1975): Techniken der empirischen Sozialforschung. Band 2. München: Oldenbourg Verlag
- Lazarsfeld, Paul (1997): Vorwort. In: Oberschall (Hg.): 17-20
- Lucci, York/Rokkan, Stein (1957): A Library Center of Survey Research Data: A Report on an Inquiry and a Proposal. New York: Columbia University School of Library Services
- Medjedović, Irena/Witzel, Andreas (2010): Wiederverwendung qualitativer Daten. Archivierung und Sekundärnutzung qualitativer Interviewtranskripte. (unter Mitarbeit von Reiner Mauer und Oliver Watteler) Wiesbaden: VS Verlag
- Merritt, Richard L. (1968): The USIA Surveys. In: Merritt/Puchala (Hg.): 3-31
- Merritt, Richard L./Puchala, Donald J. (1968): Western European Perspectives on International Affairs. Public Opinion Studies and Evaluations. New York: Frederick A. Praeger
- Mochmann, Ekkehard (2008 ): Improving the Evidence Base for International Comparative Research. In: International Social Science Journal 59: 489-506
- Mochmann, Ekkehard (2010): e-Infrastructure for the Social Sciences. In: RatSWD (Hg.): 265-285
- Oberschall, Anthony (1997): Empirische Sozialforschung in Deutschland 1848–1914. Karl Alber GmbH: Freiburg
- RatSWD (Hg.) (2014): Qualitätsstandards zur Entwicklung, Anwendung und Bewertung von Messinstrumenten in der sozialwissenschaftlichen Umfrageforschung. Working paper 230
- RatSWD (Hg.) (2010): Building on Progress: Expanding the Research Infrastructure for the Social, Economic and Behavioral Sciences. Opladen: Barbara Budrich

- Rokkan, Stein (1957): Archives for Secondary Analysis of Sample Survey Data. Vervielfältigtes Manuskript
- Rokkan, Stein (1964): Archives for Secondary Analysis of Sample Survey Data: An Early Inquiry into the Prospects for Western Europe. In: International Social Science Journal 16: 49-62
- Scheuch, Erwin K./Brüning, Iris (1964): The Zentralarchiv at the University of Cologne. In: International Social Science Journal 16: 77-85
- Scheuch, Erwin K./Wildenmann, Rudolf (Hg.) (1965): Zur Soziologie der Wahl. Köln: Westdeutscher Verlag (Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 9)
- Scheuch, Erwin K. (2003): History and Visions in the Development of Data Services for the Social Sciences . In: International Social Science Journal 55: 385-399
- Scheuch, Erwin K. (1990): From a Data Archive to an Infrastructure for the Social Sciences. In: International Social Science Journal 42: 93-111
- Scheuch, Erwin K. (1967): Entwicklungsrichtungen bei der Analyse sozialwissenschaftlicher Daten. In: König (Hg.): 655-685
- Szalai, Alexander (1972): The Use of Time. Daily Activities of Urban and Suburban Populations in Twelve Countries, (in Cooperation with Philip E. Converse, Pierre Feldheim, Erwin K. Scheuch, and Philip J. Stone). Den Haag/Paris: Mouton

Christian Meyer und Christian Meier zu Verl

---

## 15.1 Einleitung

Wissenschaftliche Erkenntnisse, die in der Sozialforschung gewonnen werden, können in Vorträgen, Berichten und wissenschaftlichen Texten auf sehr unterschiedliche Weise dargestellt werden. Dieser Aufsatz beschreibt und diskutiert verschiedene solcher Darstellungsformen und ihre theoretischen Begründungen in der qualitativen Forschung.

Unter *Darstellungsformen* verstehen wir die *unterschiedlichen Arten und Weisen, in denen im Forschungsprozess empirisches Wissen materialisiert und objektiviert wird*. So können etwa qualitative Interviews (Helfferich, Küsters, Vogl, und Friedrichs/Schwinges, Kapitel 39, 40, 41 und 43 in diesem Band) selektiv und zusammenfassend protokolliert oder per Audiorekorder oder Videokamera aufgezeichnet und im Anschluss daran mehr oder weniger detailliert transkribiert werden. Im wissenschaftlichen Text können Passagen daraus als Einzelbegriffe oder kurze paraphrasierte Ausschnitte oder als kurze oder längere Zitate aufgenommen werden.

Formen der Darstellung sind in der qualitativen Sozialforschung historisch nicht zufällig entstanden, sondern spiegeln theoretische Debatten (Reichert, Strübing, Herbrik/Kurt, Traue et al., Rosenthal, Knoblauch, Hering/Schmidt und Mayring/Fenzl, Kapitel 2, 32, 33, 34, 35, 36, 37 und 38 in diesem Band) wider und übernehmen im Forschungsprozess (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) spezifische Funktionen. Der vorliegende Text gibt hierzu einen ersten Überblick und ermöglicht weitere Recherchen im weitläufigen Feld der qualitativen Sozialforschung.

## 15.2 Die Suche nach adäquaten Darstellungsformen in der qualitativen Sozialforschung

Die Frage, welche Formen angemessen sind, um Erkenntnisse der qualitativen Sozialforschung zu präsentieren, stellt sich im Wesentlichen auf zwei Ebenen.

### 15.2.1 Bedeutung von Daten in wissenschaftlichen Texten

Erstens stellt sich die Frage, was aus Sicht der qualitativen Sozialforschung als „Datum“ angesehen werden kann, wenn Daten – Beschreibungen, Protokolle, Dokumente oder Aufzeichnungen – nicht als Fenster zur oder Spiegel der Welt angesehen werden können. Hintergrund für diesen Gedanken war ein aus der pragmatistischen und poststrukturalistischen Philosophie übernommener und als „*Krise der Repräsentation*“ (Berg/Fuchs 1993) bekannt gewordener Skeptizismus gegenüber der Annahme, die Welt könne sprachlich und mental wie mit einem Spiegel abgebildet werden, so dass Phänomene und Text Eins-zu-Eins korrespondieren (Rorty 1981). Einer solchen *Korrespondenztheorie der Wahrheit* wurde entgegnet, dass es im Verhältnis zwischen Welt und sprachlicher Repräsentation eine unausweichliche Differenz gebe und das Phänomen daher eher durch stilistische Mittel nachgebildet werden müsse, statt es wie eine Kopie abbilden zu wollen. Außerdem wurde darauf hingewiesen, dass sprachliche Repräsentationspraktiken von Wissenschaftlern ebenso Teil der sozialen und kulturellen Welt seien wie die erforschten Phänomene selbst, und daher ein Interaktionsverhältnis zwischen Beschreibungssprache und sozialer Welt bestehe (Rabinow 1993).

Welche Funktion können Daten in der Wissenschaft noch übernehmen, und wie können wissenschaftliche Aussagen als in irgendeiner Form mit der sozialen und kulturellen Wirklichkeit in Zusammenhang stehend begründet und plausibilisiert werden? *Innerhalb der qualitativen Sozialforschung herrscht Uneinigkeit über den wissenschaftstheoretischen Status von Daten und deren Bedeutung im wissenschaftlichen Text.*

Obwohl „Datum“ wörtlich das „Gegebene“ bedeutet, sind sich die meisten Forscher darüber einig, dass „Daten“ erst im Forschungsprozess erzeugt werden und daher als Artefakte wissenschaftlichen Arbeitens anzusehen sind (Cicourel 1964). Sie sind keine naturalistischen Abbilder aus der sozialen Welt, sondern basieren auf medial (technisch, textlich, mnemonisch, etc.) konservierten Dokumentationen sozialer Wirklichkeit. Erst der Konserverungs-, Selektions-, Bewertungs-, Analyse-, Interpretations- und Darstellungsprozess während der Forschung macht diese im Feld gesammelten Dokumentationen dann schrittweise zu Daten, was ihnen nicht nur in der qualitativen Forschung einen grundlegend sozialen und rhetorischen Charakter verleiht (Gitelman 2013).

Dementsprechend herrscht in der qualitativen Sozialforschung auch weitgehend Konsens darüber, dass es vermieden werden soll, Daten allein zur „gekonnten Illustration“ (Gluckman 1961: 7-8) oder zum „Exampling“ (Glaser/Strauss 1967: 5) wissenschaftlicher Aussagen zu verwenden. D.h., sie sollten keinesfalls einfach isoliert dazu verwendet

werden, Analyseergebnisse oder gar theoretische Annahmen zu veranschaulichen. Vielmehr sollte gerade der Prozess, in dem gesammelten Dokumenten Bedeutung durch den Forscher zugeordnet und zugeschrieben wird, an den im wissenschaftlichen Text dargestellten Daten gezeigt und dem kritischen Leser *nachvollziehbar* (Flick, Kapitel 29 in diesem Band) gemacht werden.

### 15.2.2 Adäquate Textgattung(en) für qualitative Forschungsergebnisse

Zweitens stellt sich die Frage nach einer der qualitativen Forschungshaltung angemessenen textlichen Gattung. Der Ethnologe Clifford Geertz (1980) z.B. hat den *Essay* als die der ethnographischen dichten Beschreibung adäquate Textgattung favorisiert. Die *dichte Beschreibung* setzt auf das Entschlüsseln und Beschreiben kulturell und gesellschaftlich geprägter subjektiver Sinnstrukturen, die das Handeln von Personen in einem bestimmten Kontext motivieren. Eine solche Methode, die sich mit Sinn (und nicht Kausalität) befasst, müsse, so Geertz, Motivlagen von Akteuren und Lesarten von sozialen und kulturellen Phänomenen (sozialen Handlungen, Artefakten, Objektivationen) mittels sinnhaft strukturierter narrativer Praktiken nachvollziehen.

Im *ethnographischen Essay* werden unterschiedliche wissenschaftliche Genres aus Geschichts- und Ingenieurwissenschaft, Recht und Mathematik, Ethik und Erkenntnistheorie vom Autor gekonnt miteinander verschränkt, um der Vielfalt und dem Konstruktionscharakter kultureller Phänomene gerecht zu werden (ein berühmtes Beispiel hierfür ist Geertz 1972).

Dem wurde entgegnet, dass das Geertzsche Unternehmen der essayistischen dichten Beschreibung den Autor des ethnographischen Textes übermäßig privilegiere und den Beschriebenen jegliche Stimme und Möglichkeit zur Selbstdarstellung raube (Crapanzano 1996). Zudem wurde kritisiert, dass es dem Wissenschaftscharakter der Kultur- und Sozialwissenschaften generell schade, wenn sie ihre Forschungspraxis als literarische Kunst positioniere und damit die Grenzen zur Dichtung verwische, statt plausible, datenbasierte wissenschaftliche Begründungen für ihre Aussagen zu fordern (Shankman 1984).

Die Frage, welche Textgattung der qualitativen Forschungshaltung adäquat ist, hängt letztlich ebenso vom *Gegenstand* ab wie von *individuellen Präferenzen* und *institutionellen Kontexten des Forschers*. Eine *Vielheit der Darstellungsformen* sollte angesichts der Pluralität sozialer und kultureller Phänomene aber ausdrücklich zugelassen und als Stärke begriffen werden.

Die genannten Probleme gelten nicht nur für die ethnographische Methode (Knoblauch, Kapitel 36 in diesem Band), sondern für die qualitative Sozialforschung im Allgemeinen. Diese umfasst jedoch zahlreiche unterschiedliche Ansätze mit jeweils sehr verschiedenen wissenschafts- und erkenntnistheoretischen Grundpositionen. Während es z.B. strukturalistisch orientierten Ansätzen (Diskursanalyse, Objektive Hermeneutik) um objektive und zugleich latente Sinnstrukturen geht, befassen sich phänomenologische Zugänge mit subjektiven Deutungsmustern und Sinnzuschreibungen (Lebensweltanalyse)

oder mit Konstitutionsbedingungen und Produktionsformen von Sozialität (Ethnomethodologie). Der am Pragmatismus orientierten Grounded Theory (Strübing, Kapitel 32 in diesem Band) wiederum geht es meist um Handlungsperspektiven und den gemeinten und erlebten Sinn sozialer Situationen aus der Perspektive der Akteure sowie daraus resultierender Konfliktpotentiale.

### **15.2.3 Datensorte und Darstellungsform**

Entsprechend der divergierenden Grundorientierungen besitzen die einzelnen Ansätze sehr verschiedene Auffassungen über die *Aussagekraft einzelner Datensorten*, was zur Konsequenz hat, dass jeder Ansatz bestimmte methodische Vorentscheidungen beinhaltet. Während die einen das *Interview* (Helfferich, Küsters, Vogl und Friedrichs/Schwinges, Kapitel 39, 40, 41 und 43 in diesem Band) als Methode der Wahl ansehen, um Daten über soziale Wirklichkeitsbereiche zu sammeln, verstehen andere unter Interviewaussagen allenfalls „Spuren“, die nur sehr undeutlich auf eine dahinter liegende soziale Realität verweisen. Die Kritiker halten nicht selten *Beobachtungen* (Thierbach/Petschick und Schnettler/Tuma, Kapitel 66–68 und in diesem Band) oder *technische Aufzeichnungen umgestellter sozialer Prozesse* (Salheiser, Klein, Ernst, Bohnsack, Akremi und Schubert, Kapitel 62, 64, 63, 67, 69 und 70 in diesem Band) für aussagekräftiger, was ihnen bisweilen wiederum den Vorwurf eines naiven Naturalismus einhandelt. Insofern kann sowohl von Präferenzen für bestimmte Datentypen als auch von einer unterschiedlichen Bewertung des Bezugs von Dokumenten sozialer Wirklichkeit zu ihrer jeweiligen Vorlage in den einzelnen Ansätzen der qualitativen Sozialforschung gesprochen werden.

---

## **15.3 Die Darstellung von Daten und die Güte qualitativer Forschung**

Historisch hat die Methodologie der qualitativen Sozialforschung zum einen von alltäglichen Methoden der Wissensproduktion und -vermittlung profitiert, wie z.B. der *Reiseberichterstattung*, der *journalistischen Reportage* oder der *kriminalistischen Befragung*. Zum anderen wurde sie durch wissenschafts- und erkenntnistheoretische Debatten über die angemessene Güte und Form von Erkenntnissen geprägt. Die aktuelle Debatte unterscheidet mehr als zehn *Gütekriterien qualitativer Sozialforschung* (Flick, Kapitel 29 in diesem Band). Mit Blick auf die Frage nach Darstellungsformen sind fünf Kriterien besonders relevant: Datenanreicherung, Kontextorientierung, Einzelfallorientierung, Reflexivität und intersubjektive Nachvollziehbarkeit. Allgemeine Qualitäten guter wissenschaftlicher Praxis, die ebenfalls Einfluss auf Darstellungsformen besitzen, wie z.B. diejenige, dass Erkenntnisse sich nicht intern widersprechen, sondern kohärent sein sollten, werden an dieser Stelle nicht erörtert.

### 15.3.1 Datenanreicherung

Die Datenanreicherung in der qualitativen Sozialforschung impliziert, *soziale Phänomene gegenstandsangemessen zu beschreiben, d.h., sie in ihrer Vielfalt, Ambivalenz und Dynamik zu erhalten*. Dies steht im Gegensatz zur quantitativen Sozialforschung, die soziale Phänomene in isolierte undzählbare Einheiten zerlegt (Bergmann 2006: 17). Für die Darstellung bedeutet dies, dass Darstellungsoptionen gefunden werden müssen, die je nach Gegenstand und analytischem Interesse das Phänomen in seiner Partikularität dem Leser oder Zuhörer begreiflich machen. Hierfür wurden, wie unten gezeigt wird, sehr unterschiedliche Lösungen gefunden.

### 15.3.2 Kontextorientierung

Unter der Perspektive der Kontextorientierung entfaltet die *Datenanreicherung* ihren analytischen Wert. Eine Äußerung innerhalb eines ethnographischen Gesprächs wie z.B. ein „Nein“ als Antwort auf die Frage, ob eine Person gesellschaftlich integriert ist, kann z.B. im Laufe des Gesprächs durch weitere Äußerungen oder anhand von Beobachtungen während eines Feldaufenthaltes relativiert werden. Dabei ist das Ausmaß des Kontextes variabel; aber im Wesentlichen wird die Bedeutung einer Handlung oder Äußerung über ihre kontextuelle Einbettung erfasst (Bergmann 2006: 18). Es ist also für die qualitative Darstellung wesentlich, dass *ausreichend Kontextinformationen* bereitgestellt werden, damit der Leser oder Zuhörer den Schritt der Interpretation eines Datums nachvollziehen kann.

### 15.3.3 Einzelfallorientierung

Dabei fokussieren qualitative Forschungen oft auf einzelne Fälle (Einzelfallorientierung) (Hering/Schmidt, Kapitel 37 in diesem Band), die es gilt, „exhaustiv“ (erschöpfend) in Hinblick auf allgemeine Bedeutungsstrukturen und generative Mechanismen zu untersuchen. Die Fallbeschreibung eines Gegenstands nimmt immer auch Mehrdeutigkeiten und Widersprüchlichkeiten auf. D.h. es werden in besonderem Maße abweichende Fälle („deviant cases“), die in Widerspruch zu bisher formulierten Theorie stehen, in die Analyse einbezogen, um die aus der Analyse hervorgehende fallspezifische Theorie als allgemeine Theorie formulieren zu können (Bergmann 2006: 21). Um einen Einzelfall angemessen beurteilen zu können, müssen Leser und Zuhörer über *ausreichende Informationen über die Fallstruktur und -logik* verfügen.

### 15.3.4 Reflexivität

Ferner hebt Bergmann (2006: 23-24) für die qualitative Sozialforschung drei Ebenen der Reflexivität hervor:

1. Die Daten sind durch den *Forscher* geprägt, so dass der Forscher sich in seinen Daten selbst begegnet; umgekehrt inkorporiert der Forscher Wissen über das Forschungsfeld.
2. Im Sinne von Schütz (1953) zeichnet sich die *soziale Welt* dadurch aus, dass die daran beteiligten Personen selbst bereits Fragen stellen, Beobachtungen machen und Interpretationen vornehmen. Daher handelt es sich bei sozialwissenschaftlichen Aussagen um *Konstrukte zweiter Ordnung*, die auf den alltäglichen Konstrukten der Akteure aufbauen.
3. *Soziale Praktiken* sind nach Garfinkel (1967) selbst insofern reflexiv, als sie von vornherein in einer Form vollzogen werden, die sie für Dritte (Akteure wie Forscher) erkenn- und interpretierbar macht.

Die reflexive Grundhaltung qualitativer Forschung impliziert, dass sowohl *Forschungsprozess* und *Position des Forschers* als auch *der Bezug zu Konstrukten erster Ordnung der Akteure* sowie die *Reflexivität* von deren Äußerungen *dokumentiert* und den Lesern und Zuhörern zugänglich gemacht werden müssen.

### **15.3.5 Intersubjektive Nachvollziehbarkeit**

Mit intersubjektiver Nachvollziehbarkeit wird der bereits angeklungene Aspekt hervorgehoben, dass in der qualitativen Sozialforschung der Forschungsprozess und die Ergebnisse der Forschung in allen genannten Dimensionen generell *auch für andere Forscher (Leser, Zuhörer) transparent und nachvollziehbar* gemacht werden, damit Anspruch auf Geltung erhoben werden kann. Neben der angemessenen *Dokumentation des Forschungsprozesses* werden hierzu die *Reflexion der angewendeten Methoden* und die *frühe Integration von anderen Forschern in den Forschungsprozess* (z.B. in Form von *Datensitzungen*) empfohlen (Steinke 2000: 324-326).

### **15.3.6 Folgen für die Darstellungspraxis in der qualitativen Sozialforschung**

Aus den angeführten Gütekriterien folgen im Wesentlichen zwei Herausforderungen für die Darstellungspraxis in der qualitativen Sozialforschung: Erstens muss die *Nachvollziehbarkeit und Transparenz des Forschungsprozesses* gewährleistet und zweitens *Gegenstandsangemessenheit* produziert werden, ohne dass dies bedeutet, von einer korrespondenztheoretisch vorgestellten Abbildung sozialer Wirklichkeit in wissenschaftlichen Texten auszugehen. Vielmehr sind Nachvollziehbarkeit und Gegenstandsangemessenheit als Ergebnisse einer der informierten Rationalität der wissenschaftlichen Gemeinschaft plausiblen Form der Darstellung anzusehen (vgl. zur diesem Gedanken zugrunde liegende Konsensttheorie, Habermas 1973).

## 15.4 Von konventionell bis experimentell: Verschiedene Formen der Darstellung

### 15.4.1 Offenlegung der Datenbasis

Die Entwicklung und Kanonisierung von Darstellungsformen in der qualitativen Sozialforschung war stets von Forschungsinteressen, verfügbaren Techniken, theoretischen Entwicklungen, ethischen Ansprüchen und sozialen Praktiken in der Wissenschaft beeinflusst. Bis auf wenige Ausnahmen herrscht jedoch Einigkeit, dass es der Selbstan spruch einer Wissenschaft sein muss, in empirischen Texten und Vorträgen auch einen plausibilisierenden Bezug zum Forschungsprozess herzustellen, insbesondere, indem sie Dokumente sozialer Wirklichkeit, die vom Forscher produziert wurden, auch dem Leser bzw. Zuhörer zugänglich machen. Eine *Offenlegung der Datenbasis* erscheint erstens wichtig, weil diese Dokumente auch dem Forscher als Grundlage der Analyse und Interpretation dienen, so dass mit ihrer Offenlegung auch der Analyse- und Interpretationsprozess nachvollziehbar gemacht wird. Zweitens entzieht diese Praxis dem Schreibenden auch einen Teil seiner auktorialen Allmacht, indem sie den erforschten Subjekten (die ja selbst reflexive und kompetente soziale Akteure sind durch die direkte und unveränderte Zitation von Äußerungen zumindest teilweise eine Stimme) im sozialwissenschaftlichen Repräsentationsprozess zugesteht (Clifford 1993).

### 15.4.2 Forschungstradition, Datensorte und Darstellung

Allerdings differieren die einzelnen qualitativen Ansätze untereinander in Hinblick auf ihre Unterscheidung und Bewertung der einzelnen in der qualitativen Forschung gängigen *Datensorten*. Während für viele im Anschluss an die quantitative Sozialforschung arbeitende Forscher (z.B. die qualitative Inhaltsanalyse; Mayring/Fenzl, Kapitel 38 in diesem Band), aber auch etwa für die Biografieforschung (Rosenthal, Kapitel 35 in diesem Band) das *Interview* (als vertiefende Form der Befragung) die bevorzugte Sorte ist, bewertet die Diskursanalyse (Traue et al., Kapitel 34 in diesem Band) bei einer gewissen Präferenz für *Dokumente* (Salheiser, Ernst und Klein, Kapitel 62, 63 und 64 in diesem Band) unterschiedliche Datensorten ähnlich, da sie sämtliche unterliegende gesellschaftliche Strukturen („Diskurse“) widerspiegeln. Dies gilt auch für die Objektive Hermeneutik, die allerdings zusammen mit der Ethnomethodologie und der *Videoanalyse* (Schnettler/Tuma, Kapitel 68 in diesem Band) technisch unterstützte (Audio-/Video-) Aufzeichnungen für die Methode der Wahl hält, während die *Ethnographie* (Knoblauch, Kapitel 36 in diesem Band), die *Einzelfallanalyse* (Hering/Schmidt, Kapitel 37 in diesem Band) und die *Grounded Theory* (Strübing, Kapitel 32 in diesem Band) wiederum eine Tendenz zur *Datenpluralität* pflegen, indem sie Beobachtungsprotokolle, interviewartige Gespräche, technische Aufzeichnungen und Dokumente aus dem Feld in die Forschung einbeziehen. Die *sozialwissenschaftliche Hermeneutik* (Herbik/Kurt, Kapitel 33 in diesem Band) bezieht zudem

*bewegte Bilder* (Akremi und Schnettler/Tuma, Kapitel 69 und 68 in diesem Band) oder *stille Bilder* (Bohsack, Kapitel 67 in diesem Band) in ihr Datenspektrum ein.

Eine für die Bewertung qualitativer Datensorten zentrale Unterscheidung betrifft die Form, in der die Dokumente sozialer Wirklichkeit als qualitative Daten konserviert werden. Bergmann (2007) unterscheidet in diesem Zusammenhang zwischen einer registrierenden und einer rekonstruierenden *Form der Konservierung sozialer Phänomene*. Bei der *registrierenden Form* werden soziale Phänomene direkt in ihrer beobachtbaren Gestalt technisch aufgezeichnet (z.B. per Videoaufzeichnung), während sie bei der *rekonstruierenden Form* in einer Weise konserviert werden, die von vorne herein reduktiv und interpretierend ist (z.B. Beobachtungsprotokolle, Interviews).

Zwar werden bei *Audio- und Videoaufzeichnungen* sozialer Situationen unweigerlich Selektionen in Bezug auf Ort, Zeit und Perspektive bzw. Ausschnitt vorgenommen (Schnettler/Tuma, Kapitel 68 in diesem Band), doch sind die aufgezeichneten Ereignisse selbst in einer Form konserviert, die noch keine Interpretationen des Forschers beinhaltet. In sozialwissenschaftlichen Analysen werden diese Daten auf unterschiedliche Weise dargestellt (vgl. Büscher 2005 für vielfältige Möglichkeiten).

Bei *Beobachtungsprotokollen* und *Feldnotizen* (Thierbach/Petschick, Kapitel 66 in diesem Band) ist dies anders: Sie sind ebenso selektiv und darüber hinaus von Anfang an in einer Beschreibungssprache fixiert, die erstens notwendig reduktiv ist und zweitens dem beschriebenen sozialen Geschehen vorab Interpretationen zuordnet, die der Beschreibungssprache unvermeidlich eingeschrieben sind. Dieser vom Sozialforscher zugeschriebene Sinn kann sich vom gemeinten und erlebten Sinn der Akteure grundlegend unterscheiden. Audio- und Videoaufzeichnungen eröffnen zudem die Möglichkeit, eine Aufnahme immer wieder und in unterschiedlichen Geschwindigkeiten abzuspielen und so eine detailliertere sequenzielle Analyse der Vollzugspraktiken sozialer Akteure vorzunehmen. Dazu zählt auch das Identifizieren flüchtiger und subtiler Praktiken, die selbst einem geübten Beobachterauge entgehen können. Ein gelungenes Beispiel hierfür ist Scheffer (1997), der zeigt, wie Identität situativ hergestellt wird, indem Polizisten den Pass einer Person als temporäres Erkenntnisinstrument nutzen.

Im *Interview* (Helfferich, Küsters, Vogl, und Friedrichs/Schwinges, Kapitel 39, 40, 41 und 43 in diesem Band) werden von den Beteiligten per definitionem Aussagen generiert, die rückwirkende Deutungen sozialer Phänomene liefern und daher per se rekonstruierend sind. Selbst Formen wie das narrative Interview (Küsters, Kapitel 40 in diesem Band), bei denen subjektive Deutungsmuster von Individuen im Zentrum stehen, müssen sich auf sozial relevante Fragen beziehen – und es ist genau diese Beziehung zwischen der in der Interviewsituation verbalisierten subjektiven Perspektive und der gesellschaftlichen Ausagekraft, die problematisch, uneindeutig und spurenartig sein kann. In der Studie „Alltag in der Anstalt“ von Christa und Thomas Fengler (1980) wird z.B. das Thema Gleichbehandlung als Konzept des Anstaltpersonals mit Hilfe von Interviews und deren Kontextualisierungen untersucht.

*Dokumente* (Salheiser, Ernst und Klein, Kapitel 62, 63 und 64 in diesem Band) schließlich, die vom Untersuchungsfeld selbst produziert werden (z.B. Akten in Institutionen),

können als „natürliche Daten“ bezeichnet werden, da sie von den sozialen Akteuren in einem sozialen Handlungskontext selbst erzeugt werden. Harold Garfinkel (1967) z.B. untersucht in seiner Studie „‘Good’ Organizational Reasons for ‘Bad’ Clinic Records“ Krankenhausakten und deren praktische Einbettung durch das Personal im Krankenhaus (Produktions- und Aneignungskontexte).

Sowohl die videotragt als auch die ethnographische Sozialforschung nutzt die genuin reflexive Dimension des Sozialen, indem sie die selbstexplikative und selbstformulierende Qualität sozialer Praktiken ins Zentrum ihrer Methode stellen. Wartet z.B. jemand auf einen Bus, dann macht er dies durch den Vollzug spezifischer Praktiken des „Wartens-auf-den-Bus“ auch für andere verstehbar und von anderen Aktivitäten wie „kurz-Stehenbleiben-und-Überlegen“ unterscheidbar. Während die Videoforschung diese Praktiken, die den sozialen Akteuren verkörpert und präreflexiv, jedoch nicht unbedingt sprachlich zur Verfügung stehen, als sich schrittweise aufschichtende Interaktion im Detail nachvollzieht, expliziert sie die Ethnographie durch Techniken des Schreibens aus der Perspektive des teilnehmenden Beobachters (Amann/Hirschauer 1997: 23-24).

### 15.4.3 Transkription

Um Reflexivität und intersubjektive Nachvollziehbarkeit zu erreichen, erscheint es sinnvoll, den Lesern und Zuhörern wissenschaftlicher Texte die der Analyse zugrunde liegende Datengrundlage zugänglich zu machen. Doch wie kann das geschehen? Stichwortartige Zitate und Paraphrasen scheiden hier aus, da nur längere Passagen den Kriterien der Datenanreicherung und Kontextorientierung, die letztlich auch der exhaustiven Analyse eines generalisierbaren Einzelfalles dienen, gerecht werden können.

Längere Passagen aus bereits vorliegenden Dokumenten oder Beobachtungsprotokollen zu zitieren, erscheint wenig schwierig. Um hingegen Auszüge aus technisch aufgezeichneten Interviews, Gesprächen oder sozialen Situationen im Text zu repräsentieren, müssen diese transkribiert (Kuckartz/Rädiker, Kapitel 27 in diesem Band) werden. Für die *Transkription* haben sich mittlerweile eine ganze Reihe von Standards herausgebildet, die jedoch Teil des fortlaufenden Forschungsprozesses, situativ und durch die geltenden sozialwissenschaftlichen Theorien geprägt sind. Transkriptionen verändern zudem die strukturelle Organisation von empirischen Ereignissen: Visuelle Ereignisse (Mimik oder Gestik) z.B. lassen sich nicht einfach in eine schriftliche Form übertragen, da sie selbst nicht textlich verfasst sind. Eine Lösung hierfür bieten etwa Standbilder oder Glossen.

Die Form eines Transkripts hängt letztlich vom Einsatz im Forschungsprozess ab. Eine Transkription, die zum Zweck der Analyse angefertigt wurde, umfasst andere – oftmals weitaus mehr – Elemente als ein Transkript, das publiziert wird und daher leserfreundlich und in seinen Elementen allein auf die Analyse bezogen sein sollte. Daher wird das Transkript auch nicht als autonomer Text, sondern immer in seiner reflexiven Verbindung zur ursprünglichen Aufzeichnung verstanden, die um ein vielfaches komplexer als das situierte Transkript ist. Es wäre daher auch ein Fehler, anzunehmen, das Transkript könne

ein vergangenes soziales Geschehen *abbilden*. Was es jedoch dem Leser ermöglicht, ist die Nachvollziehbarkeit der Analyse des Autors auf der Basis der dafür relevanten Daten.

Audio- und Videoaufzeichnungen und ihre Transkripte fixieren und machen ein soziales Geschehen allerdings in einer Komplexität der Analyse zugänglich, das sich in seiner Flüchtigkeit nicht nur dem wissenschaftlichen Beobachter, sondern auch dem in der Situation partizipierenden sozialen Akteur entzieht (Hirschauer 2001: 434). Insofern handelt es sich dabei um artifizielle Produkte, die eine von den Akteuren nicht unbedingt auf diese Weise erlebte Realität widerspiegeln.

#### **15.4.4 Varianten wissenschaftlicher Darstellungen**

Wie also können Darstellungsformen gefunden werden, die dem Leser oder Zuhörer etwas über den gemeinten und erlebten Sinn des in den Blick genommenen sozialen Phänomens aus der emischen Sicht der Teilnehmer vermitteln? Wie können sie seine Präsenz und die dem Moment geschuldeten Selektivitäten und Subjektivitäten kommunizieren (Geertz 1974)? Die Frage, auf welche Weise die flüchtige und präsenzahe Dimension der sozialen Wirklichkeit dargestellt werden kann, wenn der Wissenschaftler nicht (mehr) von einer direkten Korrespondenz zwischen der Vorlage (der erforschten sozialen Wirklichkeit) und sprachlicher oder visueller Darstellung (Datum, wissenschaftlicher Text, bewegte Bilder) ausgeht, hat eine ganze Reihe von Wissenschaftlern beschäftigt. Tyler (1986: 137) nennt zwei Möglichkeiten der wissenschaftlichen Darstellung:

1. Die erste Form, „*Schreiben innerhalb der wissenschaftlichen Konventionsgrenzen*“, erfordert für seine Darstellung eine Präzision und Luzidität, die den Gegenstand so klar und deutlich hervortreten lässt, dass er für Leser und Zuhörer fassbar und anschaulich wird. Einen gelungenen Versuch für diese Form der Darstellung stellt Lynchs (1985) Studie über die alltägliche Arbeit des Produzierens, Dokumentierens und Interpretierens von Wissensobjekten in einem neurobiologischen Laboratorium dar. Sie muss allerdings selbst dem geübtesten und talentiertesten Schreiber versagt bleiben, wenn der zu beschreibende Gegenstand sich allen etablierten Kommunikationsmöglichkeiten entzieht. Dies ist nicht selten bei Gegenständen der Ethnologie der Fall, wenn die soziale Wirklichkeit der Beforschten so fremd ist, dass dem Wissenschaftler dazu jegliche Beschreibungssprache fehlt.
2. An dieser Stelle setzt die zweite Form an, das „*Schreiben am Limit*“. Es schreibt gegen die Grenzen des Schreibens „an“, indem es sprachliche Konventionen (Syntax, Semantik, Genre) herausfordert. Dazu arbeitet es mit dem Prinzip der „Evokation“, d.h. es nutzt alle möglichen Medien, um beim Leser oder Zuhörer einen Eindruck hervorzurufen, der dem nahekommt, was der teilnehmende Beobachter im Feld erlebt hat. Das evokative Darstellungsprinzip setzt mit seinen experimentellen Darstellungsformen auf Mimesis und Performanz („Erschreiben“) statt auf Informationen über Informationen („Beschreiben“). Beispiele hierfür sind das ineinanderweben verschiedener textlicher Formate (Dialoge, Tagebücher, Beobachtungen, einheimischer Narrative) zusammen

mit Zeichnungen, Fotografien und Filmen, um die Darstellung sowohl polyphon zu gestalten, als auch Perspektiven der ersten und dritten Person zu vermitteln (Lydall/Strecker 1979), gedichtartige ethnographische Darstellungen, die durch den Rhythmus und die Materialität der Sprache ein fremdes Ritual mimetisch nachempfinden lassen (Fichte 1976), oder körperliche Performanzen, die Exklusion und rassistische Anfeindungen miterlebbar machen (Reinelt 1996).

Für welche der unterschiedlichen Darstellungsoptionen sich der qualitative Forscher letztlich entscheidet, hängt sowohl vom Gegenstand, individuellen Präferenzen und wissenschaftsinstitutionellen Kontexten ab.

#### **15.4.5 Anonymisierung**

Ein ethischer Aspekt (Friedrichs und Mühlichen, Kapitel 3 und 4 in diesem Band) der Darstellungsfrage betrifft die Anonymisierung von präsentierten Daten. Der Schutz der Identität von Forschungsteilnehmern ist grundlegend und muss insbesondere dann vereinbart werden, wenn die Teilnehmer ihre Identität nicht preisgeben möchten oder bei einer Identifizierung gar einen Schaden befürchten müssen. In diesem Fall muss eine Anonymisierung des Datenmaterials im Rahmen eines Vortrags oder einer Veröffentlichung durchgeführt werden. Dabei können z.B. relevante Informationen, die zu einer Identifizierung eines Teilnehmers führen können, aus dem Datenmaterial „herausgenommen“ (z.B. geschwärzt) werden. Der Lesbarkeit dienlicher ist allerdings, wenn sie „umgeschrieben“ werden, indem z.B. reale Personennamen, Orts- und Organisationsbezeichnungen durch adäquate Pseudonyme ersetzt werden.

---

### **15.5 Neue Medien gleich neue Formen? Ein kurzer Ausblick**

Die neuen Medien eröffnen der qualitativen Sozialforschung neue Möglichkeiten der Darstellung, die dem gedruckten Text bislang versagt blieben. So wird z.B. zusätzliches Anschaulungsmaterial in Form von kurzen Videosequenzen im Internet begleitend zu Büchern oder Aufsätzen bereitgestellt (vgl. z.B. Büscher 2005). Dies vermittelt die Erlebnisperspektive auf die soziale Situation, kann allerdings nicht die Transkription ersetzen, da sich ansonsten die Probleme der Flüchtigkeit und Komplexität erneut stellen würden.

Aufgrund der steigenden Verfügbarkeit audio-visueller Medien im Alltag (Smartphones, intelligente Brillen und Armbanduhren etc.) ist auch eine weitere Veralltäglichung der audio-visuellen Selbstdokumentation zu erwarten. Deren öffentlich zugängliche Produkte können in Zukunft noch stärker als bisher als wissenschaftliche Daten verwendet werden.

## Literatur

- Amann, Klaus/Hirschauer, Stefan (1997): Die Befremdung der eigenen Kultur. In: Hirschauer, Stefan/Amann, Klaus (Hg.): *Die Befremdung der eigenen Kultur*. Frankfurt a. M.: Suhrkamp. 7-52
- Berg, Eberhard/Fuchs, Martin (Hg.) (1997): *Kultur, soziale Praxis, Text*. Frankfurt a. M.: Suhrkamp
- Bergmann, Jörg R. (2006): Qualitative Methoden der Medienforschung. In: Ayaß, Ruth/Bergmann, Jörg R. (Hg.): *Qualitative Methoden der Medienforschung*. Reinbek bei Hamburg: Rowohlt. 13-41
- Bergmann, Jörg R. (2007): Flüchtigkeit und methodische Fixierung sozialer Wirklichkeit. Aufzeichnung als Daten der interpretativen Soziologie. In: Hausendorf, Heiko (Hg.): *Gespräch als Prozess*. Tübingen: Gunter Narr Verlag. 33-66
- Büscher, Monika (2005): Social Life under the Microscope? In: *Sociological Research Online* 10 (1). URL: <http://www.socresonline.org.uk/10/1/buscher.html>, am 20.9.2013
- Cicourel, Aaron V. (1964): *Method and Measurement in Sociology*. New York: Free Press
- Clifford, James (1993): Über ethnographische Autorität. In: Berg, Eberhard/Fuchs, Martin (Hg.): *Kultur, soziale Praxis, Text*. Frankfurt a. M.: Suhrkamp. 109-157
- Crapanzano, Vincent (1996): Das Dilemma des Hermes. In: Bachmann-Medick, Doris (Hg.): *Kultur als Text*. Frankfurt a. M.: Fischer. 161-193
- Fengler, Christa/Fengler, Thomas (1980): *Alltag in der Anstalt*. Rehburg-Loccum: Psychiatrie-Verlag
- Fichte, Hubert (1976): *Xango*. Frankfurt a. M.: Fischer
- Garfinkel, Harold (1967): *Studies in Ethnomethodology*. Englewood Cliffs (NJ): Prentice Hall
- Geertz, Clifford (1972): Deep Play. In: *Daedalus* 101 (1): 1-37
- Geertz, Clifford (1974): "From the Native's Point of View". In: *Bulletin of the American Academy of Arts and Sciences* 28 (1): 26-45
- Geertz, Clifford (1980): Blurred Genres. In: *The American Scholar* 49 (2): 165-179
- Gitelman, Lisa (Hg.) (2013): "Raw Data" is an Oxymoron. Cambridge, MA: MIT Press
- Glaser, Barney G./Strauss, Anselm L. (1967): *The Discovery of Grounded Theory*. Chicago: Aldine
- Gluckman, Max (1961): Ethnographic Data in British Social Anthropology. In: *Sociological Review* 9 (1): 5-17
- Habermas, Jürgen (1973): *Wahrheitstheorien*. In: Fahrenbach, Helmut (Hg.): *Wirklichkeit und Reflexion*. Pfullingen: Neske. 211-265
- Hirschauer, Stefan (2001): Ethnographisches Schreiben und die Schweigsamkeit des Sozialen. In: *Zeitschrift für Soziologie* 30 (6): 429-451
- Lydall, Jean/Strecker, Ivo (1979): *The Hamar of Southern Ethiopia*. Hohenschäftlarn: Renner

- Lynch, Michael (1985): Art and Artifact in Laboratory Science. London: Routledge & Kegan Paul
- Rabinow, Paul (1993): Repräsentationen sind soziale Tatsachen. In: Berg, Eberhard/Fuchs, Martin (Hg.): Kultur, soziale Praxis, Text. Frankfurt a. M.: Suhrkamp. 158-199
- Reinelt, Janelle (1996): Performing Race. In: Modern Drama 39 (4): 609-617
- Rorty, Richard (1981): Der Spiegel der Natur. Frankfurt a. M.: Suhrkamp
- Scheffer, Thomas (1997): Der administrative Blick. In: Hirschauer, Stefan/Amann, Klaus (Hg.): Die Befremdung der eigenen Kultur. Frankfurt a. M.: Suhrkamp. 168-197
- Schütz, Alfred (1953): Common-Sense and Scientific Interpretation of Human Action. In: Philosophy and Phenomenological Research 14 (1): 1-38
- Shankman, Paul (1984): The Thick and the Thin. In: Current Anthropology 25 (3): 261-280
- Steinke, Ines (2000): Gütekriterien qualitativer Forschung. In: Flick, Uwe/von Kardorff, Ernst/Steinke, Ines (Hg.): Qualitative Forschung. Reinbek bei Hamburg: Rowohlt. 319-331
- Tyler, Stephen (1986): Post-Modern Ethnography. In: Clifford, James/Marcus, George E. (Hg.): Writing Culture. Berkeley: University of California Press. 122-140

Jürgen Friedrichs

Wissenschaftler/innen haben ein großes Interesse daran, ihre Forschungsergebnisse zu veröffentlichen. Sie tragen damit zum Erkenntnisfortschritt auf ihren jeweiligen Gebieten bei. Gleichzeitig dient die Veröffentlichung auch dazu, kritisiert und korrigiert zu werden – also die Ergebnisse und deren Interpretation durch die scientific community kontrollieren zu lassen.

Die Ergebnisse von Studien können mündlich vorgetragen oder schriftlich in Fachzeitschriften präsentiert werden. Die Darstellung kann in beiden Fällen einem inzwischen als Standard geltenden Aufbau folgen:

1. Problem
2. Theorie
3. Hypothesen
4. Methoden und Stichproben
5. Ergebnisse
6. Zusammenfassung / Folgerungen.

Ich stelle dies am Beispiel einer Studie über ein benachteiligtes Wohngebiet dar (Blasius et al. 2008). Unter (1) wird das Thema der Studie in einen weiteren (gesellschaftlichen) Zusammenhang gestellt: Wie kommt es zur Konzentration von Armut in wenigen Gebieten einer Stadt? Wie sind die Lebensbedingungen in einem solchen Gebiet? Macht ein armes Wohngebiet die Bewohner/innen noch ärmer?

In (2) sollten nicht alle Theorien angeführt und behandelt werden, die geeignet sein könnten, den untersuchten Sachverhalt zu erklären, sondern nur diejenige(n) Theorie(n), die tatsächlich verwendet wird/werden. In diesem Falle sind es Hypothesen über Kontexteffekte (des Gebiets auf die Bewohner) und Hypothesen über die positiven Effekte von sozialem Kapital.

In (3) wird man den Forschungsstand referieren, um damit am Ende die (eigenen) Hypothesen der Studie zu begründen. Die Teile (2) und (3) sind der methodologische

Teil des Berichts. Es folgt der methodische Teil (4), der darüber informiert, wie empirisch vorgegangen wurde: über die angewandte(n) Methode(n) und warum man diese gewählt hat. Bei den Stichproben sollten die Ausschöpfungsquoten, ein Vergleich mit der Brutstichprobe und der Grundgesamtheit vorgelegt werden. Ferner sind die Ausfallgründe anzuführen und vor allem ist zu erörtern, ob die Stichprobe eine systematische Verzerrung gegenüber der Grundgesamtheit aufweist, was u. U. dazu führt, sie gewichten zu müssen – was z.B. bei dem ALLBUS der Fall ist. Im Beispiel wurde eine Wahrscheinlichkeitsstichprobe der Bewohner/innen des Gebiets durch das Amt für Statistik der Stadt Köln gezogen. Die Befragung erfolgte face-to-face. Es wurden 707 Interviews mit deutschen und 239 mit türkischen Befragten erzielt. Die Ausschöpfungsquote lag bei 46 %. Eine wichtige Methode bestand darin, die Netzwerke der Befragten zu erheben.

Das umfangreichste Kapitel ist Teil (5), weil hier nun die Ergebnisse und statistischen Verfahren dargestellt werden. Bei empirischen Studien wird man meist zuerst die univariaten, dann die bivariaten und multivariaten Ergebnisse vorstellen. Wichtig ist, die Ergebnisse auf die im theoretischen Teil formulierten Hypothesen zu beziehen. Zwei wichtige Erkenntnisse unserer Studie waren: Je mehr soziales Kapital im Wohngebiet wahrgenommen wird, desto geringer sind die Fortzugsabsichten. Sozialhilfe-Empfänger hatten keine kleineren Netzwerke als Nicht-Empfänger. Türkische Befragte billigten signifikant weniger abweichendes Verhalten als deutsche Befragte. Zusätzlich haben sich drei aus der nordamerikanischen Literatur übernommene Skalen zur Messung von sozialem Kapital und des Ausmaßes an „Unordnung“ (disorder) bewährt.

Die Tabellen mit den einzelnen Berechnungen stellen die entscheidenden Belege dar. Zu beachten ist, dass die Tabellen aus sich heraus verständlich sein müssen, deshalb wichtige Erläuterungen nicht im Text sondern als Anmerkungen unter der Tabelle stehen sollten, z. B., ob horizontal oder vertikal prozentuiert wurde, um welche Regressionskoeffizienten es sich handelt, was die Abkürzungen der Variablen bedeuten. Gute Forscher/innen lesen nämlich zuerst die Tabellen und interpretieren die Ergebnisse, *bevor* sie die Interpretation der Autoren lesen. Es kann ja sein, dass die Autoren eine eigenwillige Interpretation vornehmen, der nicht alle Leser folgen wollen. Sehr technische statistische Ausführungen und spezielle Tabellen kann man in den Anhang zu dem Beitrag verlegen oder auf der eigenen Website zur Verfügung stellen.

Im Teil (6) werden die Ergebnisse zusammengefasst. Die Darstellung sollte sich sowohl auf das Problem als auch auf die Hypothesen als auch auf den Forschungsstand beziehen. Hier sollte diskutiert werden, warum einzelne Hypothesen sich vermutlich nicht bewährt haben (was zu modifizierten oder neuen Hypothesen führt), in unserem Beispiel: warum die Sozialhilfeempfänger/innen keine kleineren Netzwerke haben. Ferner, welche Folgerungen man für die Theorie zieht und welche Folgerungen sich für die künftige Forschung gezogen werden. Schließlich wird man anführen, was an der eigenen Studie unzulänglich war und vorschlagen, wie man diese Mängel oder Lücken in künftigen Studien beheben könnte. „Further research is needed“ ist die klassische Einsicht am Ende. In vielen nordamerikanischen Artikeln steht in dem Schlusskapitel auch noch, welche Bedeutung die Ergebnisse für politische Maßnahmen haben („policy implications“). In unserem Falle

war eine solcher Botschaften: Die türkischen Bewohner/innen in einem benachteiligten Wohngebiet haben konservativere Normen und stabilisieren das Gebiet. Zu solchen politischen Folgerungen gehört auch die Ergebnisse einer breiteren Öffentlichkeit, z.B. in Interviews, zu präsentieren.

Dem Beitrag wird eine kurze Zusammenfassung, ein Abstract, vorangestellt. Gemeinhin umfasst die Zusammenfassung etwa 100 bis 120 Wörter. Im besten Falle enthält sie zu jedem der fünf Punkte einen Satz. Die Leser können daraufhin entscheiden, ob sie den Aufsatz lesen wollen oder nicht. Je komplizierter diese Zusammenfassung geschrieben ist, desto weniger Wissenschaftler werden den Aufsatz lesen, weil zwischen der Qualität der Zusammenfassung und derjenigen des Artikels eine positive Korrelation besteht. Auch die Fachgutachter werden dann den Artikel selbst ungern lesen. Deshalb empfiehlt es sich, einfach und klar zu schreiben, auf Ausdrücke wie „fokussieren“, „Ausdifferenzierung“ oder „final“ zu verzichten. Auch sollte man bedenken, dass es doppelt so lange dauert, Texte mit vielen Substantivierungen zu lesen. Wird der Aufsatz bei einer Fachzeitschrift eingereicht, wird er bei fast allen Zeitschriften anonymisiert an zwei Gutachter/innen geschickt, die den Aufsatz anonym beurteilen (peer review). Die doppelte Anonymität soll Autoren wie Gutachter davor schützen, verzerrt zu urteilen. Im schlechtesten Falle wird er abgelehnt, in allen anderen Fällen wird er entsprechend den Kommentaren der Gutachter umgearbeitet und wieder eingereicht – und geht dann hoffentlich nicht an zwei andere Gutachter/innen.

## Literatur

Blasius, Jörg/Friedrichs, Jürgen/Klöckner, Jennifer (2008): Doppelt benachteiligt? Leben in einem deutsch-türkischen Stadtteil. Wiesbaden: VS Verlag.

---

## **Teil 2**

### **Stichproben, Datenaufbereitung und Güte**

Leila Akremi

---

## 17.1 Der Forschungsprozess und die drei Ebenen der Stichprobenentscheidung bei qualitativer Sozialforschung

Empirische Sozialforschung bedeutet, dass zur Gewinnung von Erkenntnissen über den jeweiligen Forschungsgegenstand immer Erfahrungen gesammelt werden müssen, die in Form von Datenmaterial nach wissenschaftlichen Gesichtspunkten ausgewertet und verallgemeinert werden können. Diese Daten werden aus einer unendlichen Fülle von möglichen Daten mehr oder weniger gezielt gewonnen. Das Thema „Stichprobenziehung“ oder „Sampling“ begleitet dabei den gesamten, in der qualitativen Forschung zumeist zyklisch verlaufenden, Forschungsprozess von der Forschungsfrage bis hin zur Präsentation von Forschungsergebnissen. *In jeder Phase des Forschungsprozesses stellen sich dem Forscher dabei unterschiedliche Fragen und es gilt auch jeweils unterschiedliche Stichprobenprobleme zu lösen (Tab. 17.1).* Merkens (2010: 286f.) spricht im Rahmen der qualitativen Stichprobenziehung von „Auswahlentscheidungen auf drei Ebenen“.

### 17.1.1 Datenerhebung: Auswahl der Erhebungseinheiten

Die erste Ebene bezieht sich auf die *Auswahl der Untersuchungseinheiten vor und während der Datenerhebungsphase im Allgemeinen*. Je nach Erhebungsmethode und Datentyp handelt es sich also um die Auswahl von Interviewpartnern (Helfferich, Küsters, Vogl, Häder/Häder und Friedrichs/Schwinges, Kapitel 39, 40, 41, 42 und 43 in diesem Band), Beobachtungssituationen (Thierbach/Petschick und Tuma/Schnettler, Kapitel 66 und 68 in diesem Band), Zeitungsartikeln (Klein, Kapitel 64 in diesem Band), Gebrauchsgegenständen oder technischen Artefakten (Schubert, Kapitel 70 in diesem Band) usw. Auf dieser Ebene müssen aber weitaus mehr Entscheidungen getroffen werden, als „nur“ die Auswahl der Erhebungseinheiten. Am Beginn des Forschungsprozesses steht eine mehr oder weniger konkrete Forschungsfrage oder ein spezifisches Erkenntnisinteresse. Zur Klärung

**Tab. 17.1** Zusammenhang zwischen Forschungsprozess und Stichprobenziehung

Ebene der Stichproben- ziehung	Phase des Forschungs- prozesses	Fragen	zu lösende Stichprobenprobleme
1. Ebene	Forschungs- frage		<ul style="list-style-type: none"> <li>- Fallkonstruktion</li> <li>- Absteckung der Grundgesamtheit</li> </ul>
	Entwicklung des Forschungs- designs	<ul style="list-style-type: none"> <li>- Was oder wen will ich wie untersuchen?</li> </ul>	<ul style="list-style-type: none"> <li>- Festlegung der Stichprobenstrategie im Einklang mit gewähltem Forschungsdesign und qualitativer Methode</li> </ul>
2. Ebene	Daten- erhebung	<ul style="list-style-type: none"> <li>- Wie wähle ich den Fall/ die Fälle konkret aus?</li> <li>- Wie komme ich an den Fall/die Fälle?</li> </ul>	<ul style="list-style-type: none"> <li>- Festlegung des Ausgangspunktes der Datenerhebung</li> <li>- Feldzugang</li> <li>- Sammlung des Datenmaterials/ Erstellung des Datenkorpus (=Stichprobe)</li> </ul>
3. Ebene	Daten- auswertung	<ul style="list-style-type: none"> <li>- Welche Fälle sind letztlich in meine Stichprobe gelangt?</li> <li>- Was wird aus den erhobenen Daten konkret ausgewertet?</li> <li>- Wie können die Ergebnisse verallgemeinert werden?</li> </ul>	<ul style="list-style-type: none"> <li>- Sichtung des erhobenen Datenmaterials</li> <li>- evtl. Einteilung in primäre, sekundäre usw. Daten</li> <li>- Sampling innerhalb der gesammelten Daten: <ul style="list-style-type: none"> <li>- Welche erhobenen Daten sollen auch ausgewertet werden?</li> <li>- Was soll innerhalb der ausgewählten Daten ausgewertet werden?</li> <li>- Was sind die Analyseeinheiten?</li> <li>- Nach welchen Kriterien werden sie festgelegt?</li> </ul> </li> </ul>
	Daten- präsentation	<ul style="list-style-type: none"> <li>- Was wird für die Veröffentlichung der Ergebnisse (Aufsatz, Monographie, Forschungsbericht etc.) genutzt?</li> </ul>	<ul style="list-style-type: none"> <li>- Entscheidung darüber, welche Teile der Daten für welchen Zweck präsentiert werden (Validierung, intersubjektive Nachvollziehbarkeit, Veranschaulichung, Illustration etc.).</li> <li>- Erläuterung der Generalisierbarkeit der Ergebnisse und Güte der Stichprobe</li> </ul>

was oder wen ich wie untersuchen möchte, gehören neben der Auswahl des Forschungsdesigns (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band), auch die Beschäftigung mit der Fallkonstruktion, Überlegungen zum Untersuchungsfeld sowie das Nachdenken über

die geplante Stichprobe bzw. die Festlegung des Startpunktes der Datenerhebung. Geht es an die Datenerhebung selbst, muss geklärt werden, *wie* die Fälle konkret ausgewählt werden und vor allem auch *wie* man an die Fälle herankommt.

### 17.1.2 Datenauswertung: Auswahl der Analyseeinheiten

Sobald Daten gesammelt worden sind, gilt es zu entscheiden, *was* davon genau ausgewertet wird, z.B. gibt es bei Interviews häufig Ausschweifungen und Exkurse, die zugelassen werden, um eine möglichst natürliche Gesprächssituation zu simulieren, aber für das Forschungsthema keine relevanten Informationen liefern. Zur Auswertung erhobener Daten gibt es daher häufig eine zweite Auswahl, nämlich die *Auswahl innerhalb des Materials*. Dazu muss sich der Forscher Gedanken darüber machen, was die konkreten Analyseeinheiten im Unterschied zu den Erhebungseinheiten sind und nach welchen Kriterien dies festgelegt wird. Die Erhebungs- und Analyseeinheiten können aber auch deckungsgleich sein.

### 17.1.3 Datenpräsentation: Auswahl des Materials für die Veröffentlichung

Als letztes wird im Rahmen der Präsentation der eigenen Forschungsergebnisse wiederum eine *Stichprobe aus den Daten zur Illustration, Veranschaulichung oder intersubjektiven Nachvollziehbarkeit* verwendet (Merkens 2010).

Was bei der Selektion innerhalb des Materials und der Entscheidung, welche Bestandteile in die Veröffentlichung der Ergebnisse integriert werden, zu beachten ist, ist eine methodologische Frage, die viel seltener als die Auswahl der Erhebungseinheiten auf der ersten Ebene im Kontext der Stichprobenproblematik betrachtet wird. Bei manchen qualitativen Ansätzen, wie z.B. der Grounded Theory (Strübing, Kapitel 32 in diesem Band), wird hingegen, wegen des zyklischen Forschungsprozesses, gar keine analytische Trennung zwischen Stichprobenziehung während der Datenerhebung und Stichprobenziehung zur Auswertung vorgenommen.

Dass es gerade für die Auswahl innerhalb des Datenmaterials und zur Präsentation von Ergebnissen wenig Handreichungen gibt, kann einerseits zu Unsicherheiten führen oder andererseits zur Pauschalkritik an der qualitativen Forschung, sie würde nach Belieben und Gutdünken einfach irgendwelche Ausschnitte aus dem Material interpretieren und schließlich präsentieren um Plausibilität zu suggerieren. Dem ist natürlich nicht so, denn methodisch vorzugehen bedeutet in der qualitativen Forschung – gemäß dem Wortsinne von „*methodus*“ –, dass es eine Systematik bzw. bestimmte Regeln gibt, an denen sich Forscher orientieren und diese zumindest implizit nachvollzogen werden können. „*Methode*“ ist nur eben kein Synonym für den Begriff der „*Standardisierung*“. Aus diesen Gründen kann das qualitative Vorgehen aufgrund anderer Forschungsinteressen und Forschungsstrategien weniger rezeptmäßig abgehandelt werden, als das in der quantitativen Forschung möglich ist.

## 17.2 Definition des Untersuchungsfeldes und der Erhebungseinheiten

Die erste Ebene der Stichprobenentscheidung bezieht sich auf die Frage, wie die interessierenden Erhebungseinheiten in Abhängigkeit vom Forschungsinteresse gesammelt werden können. Dazu muss zunächst der Forschungsgegenstand näher spezifiziert werden. Während es sich in der *quantitativen Forschung* dabei um die Definition der *objektorientierten Grundgesamtheit* handelt, die alle Elemente, Merkmalsträger oder Fälle einschließt, die für den Forschungsgegenstand relevant sind, wird in der *qualitativen Forschung* von der *Erschließung des Untersuchungsfeldes* (z.B. Flick 2007: 167f.) gesprochen. Es gibt zwei Strategien das Untersuchungsfeld abzugrenzen.

### 17.2.1 Abstrakt-methodologische Vorab-Definition vs. konkret-inhaltliche offene Herangehensweise

Das Untersuchungsfeld kann *theoriegeleitet vorab definiert* werden, wie dies auch in der quantitativen Forschung üblich ist. So haben Fuchs et al. (2009) in ihrer Studie zu „Gewalt an Schulen“ festgelegt, dass sich die Untersuchung auf alle Schüler bezieht, die eine allgemein- oder berufsbildende Schule in Bayern besuchen. Daraus ergeben sich mehrere Bedingungen für die Grenzen des Untersuchungsfeldes und für jede Erhebungseinheit muss geklärt werden, ob sie diese Bedingungen erfüllt: Ist bei jeder Schule, die es in Bayern gibt, eindeutig bestimmbar, ob sie in die Definition fällt oder nicht? Wie verhält es sich mit Schülern, die zwischenzeitlich die Schule, das Bundesland usw. wechseln, d.h. das Untersuchungsfeld sich ständig verändert?

Viele qualitative Ansätze, wie etwa die Grounded Theory (Strübing, Kapitel 32 in diesem Band), zeichnen sich aber nicht durch ein rein theoriegeleitetes Vorgehen aus, sondern sind *von der Anlage zunächst offen*. Es genügt zu Beginn eine vage Fragestellung, eine allgemeine soziologische Perspektive und das Alltagswissen, um das Untersuchungsfeld zu betreten und sich mit dem Gegenstandsbereich vertraut zu machen (Glaser/Strauss 2008: 45). In idealisierter Weise ergibt sich nach und nach der Fokus, die Begriffe definieren sich aus dem Feld, sozusagen von innen heraus.

Dies hat massive Auswirkung auf die Frage des Untersuchungsfeldes. Um im vorher eingeführten Beispiel „Gewalt an Schulen“ zu bleiben, würde es nach dieser Logik zunächst ausreichen, dass der Forscher aus der Alltagserfahrung weiß, dass es Schulen gibt, dort Schüler sind und diese nicht immer liebevoll miteinander umgehen. Für erste Erfahrungen mit dem Schulumfeld wären zunächst ethnographische Beobachtungen (Thierbach/Petschick, Kapitel 66 in diesem Band) oder Gespräche an jeder beliebigen Schule mit beliebigen Schülern hilfreich. Aus diesen gesammelten Informationen ergäbe sich dann ein erstes Bild. Möglicherweise ließen sich verschiedene Formen der Gewalt in unterschiedlichen Situationen beobachten, die dann kategorisiert werden können. Daraus könnten sich Fragen um Unterschiede in verschiedenen Schultypen anschließen, das Beobachtungsfeld

könnte sich auch auf außerschulische Situationen erweitern, man könnte überlegen, wie sich die Gewaltsituationen in der Schule von anderen vergleichbaren Situationen unterscheiden usw.

Bei diesem Vorgehen ist der Forschungsweg vorher nicht genau abgesteckt. Was konkret als Gewalt unter Schülern zählt und was die bedeutsamen Situationen sind, in denen sich Gewalt am besten beobachten lässt – in den Pausen, zwischen den Stundenwechseln –, ob es Unterschiede zwischen Schülern verschiedener Schultypen geben muss, ob es letztlich bei Gewalt an Schulen bleibt oder vielleicht um Gewalt unter Jugendlichen in unterschiedlichen Kontexten geht, wird vorab nicht definiert.

Das Ziel dieser offenen Herangehensweise ist vielmehr, die relevanten Aspekte des Forschungsfeldes zu entdecken und zu verhindern, dass eine Vorabdefinition zu Fehlwahrnehmungen führt. Damit soll der Blick stets frei sein für Erscheinungen, die nicht antizipiert werden können. Die gezielte Fokussierung und Zuspitzung erfolgt erst später, wenn man besser mit dem Untersuchungsgegenstand vertraut ist. Selbstverständlich wird das Forschungsprojekt eingebettet in den Stand der Forschung, nur steht das im zeitlichen Ablauf nicht an erster Stelle. Für die Finanzierung von Forschungsprojekten lässt sich das offen-induktive Vorgehen in idealisierter Weise aber nicht realisieren, weil für die Beantragung von Projekten die Aufarbeitung des Stands der Forschung sowie genaue Aufstellungen und Kostenkalkulationen zum Umfang der Datenerhebung angegeben werden müssen.

Die gegensätzlichen Vorgehensweisen, die von Flick (2007: 124ff.) als *abstrakt-methodologisch*, um die anfängliche Distanz zum Untersuchungsgegenstand zu verdeutlichen, und *konkret-inhaltlich*, als schrittweise Annäherung im offen-induktiven Paradigma, tituliert werden, finden sich auch bei der Definition der Erhebungseinheiten wieder.

### 17.2.2 Querschnitts- vs. Längsschnittdesign

Im Einklang mit dem Erkenntnisinteresse und den gegebenen Ressourcen wählt der qualitative Forscher ein geeignetes Forschungsdesign für seine Untersuchung aus (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) und entscheidet darüber, ob etwa narrative Interviews (Küsters, Kapitel 40 in diesem Band) über vergangene Erlebnisse oder Ereignisse, Momentaufnahmen in Form von Zustands- oder Prozessbeschreibungen (*Querschnittsdesign*) oder eher eine Analyse über einen längeren Zeitraum (*Längsschnittdesign*) geplant sind.

### 17.2.3 Einzelfallanalyse vs. Fallvergleich

Diese Entscheidungen haben bereits erste Auswirkungen auf die Anzahl der Erhebungseinheiten, die für die Untersuchung ausgewählt werden können und sollen. So ergibt es sich von selbst, dass für eine tiefergehende *Einzelfallanalyse* (Hering/Schmidt, Kapitel 37 in diesem Band) der eine interessierende und theoretisch als auch inhaltlich relevante

Fall gefunden werden muss. Um was es sich dabei konkret handelt, d.h., ob der einzelne Fall Person, Situation, Institution oder Phänomen ist (Flick 2010: 262) hängt von der inhaltlichen Frage ab. *Fälle liegen jedenfalls nicht objektiv vor, sie werden vom Forscher anhand seines Erkenntnisinteresses – wie auch das Untersuchungsfeld – konstruiert* und zwar, entweder durch theoriegeleitete Vorabfestlegung oder im Rahmen des Forschungsprozesses aus dem Feld heraus. Das können eben Personen, Organisationen usw. (Merkens 2010) sein.

„Einzelfallanalyse“ bedeutet allerdings nicht eine Stichprobengröße von n=1 im Sinne einer quantitativen Logik von Fällen. Denn um einen Einzelfall in all seinen Facetten beleuchten zu können, bedarf es einer Vielzahl von (unterschiedlichen) Daten, die in ihrer Gesamtheit den Datenkorpus für die Beschreibung des Einzelfalls bilden. Würde man z.B. untersuchen wollen, welche sozialen Mechanismen dazu führen können, dass ein zuvor als sehr aussichtsreich betrachteter Versuch der Mitarbeiter einer Firma sich gewerkschaftlich zu organisieren, letzten Endes doch scheitert, wie dies Krackhardt (1999) gemacht hat, dann könnte man dies als Einzelfallstudie definieren. Um aber zu verstehen, was diesen speziellen Einzelfall ausmacht, müssen viele Daten gesammelt werden. Dies können Interviews mit den Mitarbeitern (Helfferich, Kapitel 39 in diesem Band), Beobachtungen während der Arbeitszeit (Thierbach/Petschick, Kapitel 66 in diesem Band), Analysen der Beziehungsstrukturen (Baur, Kapitel 74 in diesem Band), Dokumente über die Firma (Salheiser, Kapitel 62 in diesem Band) sein.

Jede dieser einzelnen Erhebungseinheiten wird in der qualitativen Forschung schlicht als ein *Datum* bezeichnet. Je nach Methode und Datensorte kann es sich dabei um einen Zeitungsartikel, ein Beobachtungsprotokoll, eine Videoaufzeichnung oder ein Interviewtranskript handeln. Die Gesamtstichprobe aller erhobenen Daten wird als *Sample*, *Datenkorpus* oder *Datensammlung* (z.B. Strübing 2013: 171ff., Schnettler/Knoblauch 2007) bezeichnet.

Für *Fallvergleichsstudien* gilt es wiederum herauszufinden, welche Fälle überhaupt sinnvoll miteinander verglichen werden können und ob z.B. Ähnlichkeit (Konzentrationsprinzip) und/oder *Unähnlichkeit (Kontrastprinzip)* das Vergleichsmoment darstellen.

Je nach Forschungsstrategie kann die Absteckung von Untersuchungsfeld und Datenkorpus demnach zu Beginn oder in den überwiegenden Fällen während bzw. im Extremen erst gegen Ende des Forschungsprozesses geschehen.

---

### 17.3 Auswahl der Erhebungseinheiten

Einer der ersten Schritte im Sampling-Prozess ist die Klärung der Frage nach der Art der Erhebung. Diese wird im Zusammenhang mit Stichprobenziehung durch zwei Gegen- satzpaare gekennzeichnet. Die Unterscheidung zwischen Primär- und Sekundäranalysen bringt zum Ausdruck, dass bei *Primärerhebungen* der jeweilige Forscher auch selbst Daten produziert, z.B. narrative Interviews oder Beobachtungen durchführt, während bei einer *Sekundäranalyse* (Medjedović, Kapitel 13 in diesem Band) die Datenerhebungsphase und

damit auch die Stichprobenziehung „nur“ in der Beschaffung der erforderlichen Daten besteht. Sekundäranalysen sind in der qualitativen Forschung aufgrund von datenschutzrechtlichen Bestimmungen ein schwieriges, wenn nicht unmögliches Unterfangen, weil Originalaufzeichnungen von Interviews oder Videobeobachtungen nicht ohne weiteres anderen Forschern zur Verfügung gestellt werden können. Einfacher ist eine Sekundäranalyse daher häufiger bei von anderen Forschern zusammengestellten prozessproduzierten Daten (Salheiser, Kapitel 62 in diesem Band).

Auch die Unterscheidung zwischen Daten, die im Rahmen eines Forschungsvorhabens produziert werden und solchen, die unabhängig von der Wissenschaft entstehen, wie etwa Archivdokumente (Salheiser, Kapitel 62 in diesem Band) oder Zeitungsartikel (Klein in diesem Band), ist wichtig bei der Stichprobenziehung. Das zweite Gegensatzpaar lautet demnach *Forschungsdaten* versus *prozessproduzierte Daten*.

Um Missverständnissen vorzubeugen, sei hier noch anzumerken, dass die Unterscheidung zwischen primären und sekundären Daten in der qualitativen Sozialforschung noch für einen anderen wichtigen Sachverhalt gebraucht wird. Bei der Studie zu Powerpoint Präsentationen von Schnettler und Knoblauch (2007), die sich mit der Produktion von Wissen in Interaktionen zwischen Vortragenden, Powerpoint-Präsentation und Publikum beschäftigt, wird der Datenkorpus in Primär-, Sekundär- und Tertiärdaten unterteilt:

- Als *Primärdaten* gelten dabei die abgehaltenen Powerpoint-Vorträge verschiedener Referenten. Wurden die Vorträge im Rahmen größerer Veranstaltungen abgehalten, so gibt es um die Vorträge herum noch andere gewonnene Daten, die nicht zum Vortrag selbst, sondern zum Kontext zählen, in dem der Vortrag stattgefunden hat.
- Diese wichtigen Rahmeninformationen gehen als *Sekundärdaten* in den Datenkorpus mit ein. Zu den Sekundärdaten zählen außerdem auch Präsentationsdateien, Skizzen, Tagungs-Programme sowie Bearbeitungen der Primärdaten für die Analyse (Petschke 2007: 291).
- Eine dritte Art von Daten wird hier als *Tertiärdaten* bezeichnet. Es handelt sich dabei um Daten, die im Analyseprozess zur Dokumentation der Forschung und zur Weiterarbeit mit den Zwischenergebnissen entstanden sind und gesammelt als Projekttagebücher benannt werden (Petschke 2007: 291). Sie sind wesentlich für ein selbstreflexives Vorgehen im Laufe des Forschungsprozesses, welches insbesondere für die Beurteilung der Güte der Ergebnisse unerlässlich ist.

### 17.3.1 Zufallsstichproben vs. nicht-probabilistische Stichprobenstrategien

Bei der Auswahl einer Vielzahl von Zeitungsartikeln über einen langen Zeitraum hinweg oder bei der Entscheidung, welche Folgen einer Fernsehserie für eine Videoanalyse zu einem gegebenen Thema verwendet werden sollen, kann die Idee der Zufallsstichprobe, wie sie in der quantitativen Sozialforschung üblich ist (Häder/Häder, Kapitel 18 in diesem Band), durchaus nützlich sein, um die Gefahr der systematischen Verzerrung durch

willkürliche Auswahl zu umgehen. So könnten die zu interpretierenden Folgen einer Fernsehserie wie der Lindenstraße, bei der in 30 Jahren annähernd 1.500 Folgen ausgestrahlt wurden, per Zufallsmechanismus ausgewählt werden.

Von viel größerer Bedeutung für die qualitative Sozialforschung sind aber nicht-probabilistische Stichprobenstrategien, die sich in zwei Bereiche unterteilen lassen: Einerseits gibt es Stichprobenstrategien, die ohne vorher definierte Kriterien oder theoretische Erwägungen gezogen werden. Sie werden als *willkürliche Auswahl* oder auch als *Convenience-Sampling* (Flick 2007: 166) bezeichnet. Andererseits gibt es Datenerhebungsprozesse, denen *kriteriengeleitete bewusste Auswahlstrategien* zugrunde liegen.

### **17.3.2 Willkürliche Auswahl (Convenience-Sampling)**

Eine Art von willkürlicher Auswahl sind die sogenannten *Ad-hoc-* oder *Gelegenheitsstichproben*, bei denen z.B. Passanten auf der Straße als Fälle ausgewählt werden, die sich zu einem bestimmten Thema äußern sollen. Es kann sich auch um Beobachtungen bestimmter Situationen ohne konkrete Vorabbestimmungen handeln oder um eine „wahllose“ Sammlung von Dokumenten zu einem spezifischen Thema.

Auf sich *selbstgenerierende Stichproben*, eine weitere Form der willkürlichen Auswahl, stoßen wir in unserem Alltag in zunehmendem Maße, weil wir z.B. Kunden eines Unternehmens sind oder bestimmte Internetseiten aufrufen. Bei dieser Vorgehensweise erfolgt die Rekrutierung der Untersuchungseinheiten etwa für Gruppendiskussionen oder qualitative Interviews über einen Aushang am schwarzen Brett, einen Aufruf in einer Zeitungsannonce oder auf einer Internetseite. Personen entschließen sich dabei aus eigenem Antrieb an einer Studie teilzunehmen, ohne dass sie persönlich von den Forschern angeprochen wurden. Der Hauptvorteil dieser Vorgehensweise liegt darin, dass der Aufwand sowohl auf theoretischer Basis als auch in zeitlicher Hinsicht minimiert wird. Der Hauptnachteil ist der, dass der Forscher wenig Einfluss auf die Zusammensetzung der Stichprobe hat.

Als letztes ist unter der Rubrik der willkürlichen Auswahl das *Schneeballverfahren* zu nennen, bei dem der Forscher einen beliebigen Startpunkt für die Erhebung sucht und dann z.B. von bereits befragten Personen auf andere relevante Personen verwiesen wird, die wiederum andere relevante Personen kennen. Während zwar Überlegungen angestellt werden müssen, wer oder was als Startpunkt für die Erhebung dienen kann, entzieht sich der weitere Verlauf der Stichprobenziehung der Kontrolle des Forschers. Diese Stichprobenstrategie ist dann angebracht, wenn Forschungsfelder undurchsichtig sind, d.h. es vorab schwer ist, herauszufinden, welche Untersuchungseinheiten überhaupt theoretisch relevant sind. Eine andere Anwendungsmöglichkeit besteht, wenn Fragestellungen besonders heikel sind, wie etwa dem Thema des sexuellen Missbrauchs. In diesen Fällen kann davon ausgegangen werden, dass über andere Stichprobenstrategien nicht die theoretisch relevanten Fälle gefunden werden. In gewisser Weise könnte das Schneeballverfahren als spezifische Variante der sich selbstgenerierenden Stichproben bezeichnet werden. Beim

Schneeballverfahren gibt es jedoch zwischen einzelnen Fällen zwangsläufig Berührungspunkte, was bei allgemeinen, sich selbstgenerierenden Stichproben nicht der Fall sein muss. Diese Berührungs punkte können dazu führen, dass ein ganz spezifisches Netzwerk in die Untersuchung gelangt und eher randständige Personen oder gar Figuren mit Brückenfunktion zu anderen Netzwerken unter Umständen keine Berücksichtigung finden.

Willkürliche Stichproben tragen vor allem das Risiko der systematischen Verzerrung und folglich Verallgemeinerungsprobleme der Stichprobenergebnisse mit sich. Weder lässt sich beurteilen, ob wirklich die theoretisch relevanten Fälle ausgesucht wurden, noch in welchem Verhältnis die ausgewählten Elemente zum gesamten Untersuchungsfeld stehen. Sie sollten deshalb nicht die einzige Stichprobenstrategie sein. Als Teilstrategie im größeren Datenerhebungskonzept können sie jedoch nützliche Erkenntnisse liefern. Sollte es aufgrund der verfügbaren Ressourcen, z.B. im Rahmen einer Qualifikationsarbeit während des Studiums, nicht anders möglich sein, als eine willkürliche Stichprobe zu ziehen, so müssen die Einschränkungen in der theoretischen Generalisierbarkeit sehr genau reflektiert und widergespiegelt werden, da andernfalls kaum von wissenschaftlicher Forschung gesprochen werden kann.

### 17.3.3 Kriteriengeleitete bewusste Auswahl und theoretische Generalisierung

Bei kriteriengeleiteter bewusster Auswahl gehen theoretische Überlegungen der Datenerhebung immer voraus. Ziel ist die *theoretische Generalisierbarkeit* (Flick 2010: 260). Um von konkreten Fällen abstrahieren zu können, sind dabei inhaltliche Bandbreite und der Einsatz unterschiedlicher Methoden viel wichtiger als große Fallzahlen (Flick 2010: 260). Zwei Strategien nehmen eine besondere Stellung ein: die Quotenstichprobe und das theoretische Sampling, welches als Pendant zum Idealbild der Zufallsstichprobe angesehen werden kann.

Bei der *Quotenstichprobe* werden auf Basis theoretischer Überlegungen Merkmale oder Dimensionen zusammengestellt, nach deren Ausprägungen dann jeweils eine bestimmte Anzahl von Elementen für die Stichprobe gefunden werden müssen. Flick (2007: 167) verwendet wegen der Vorabfestlegung der Kriterien für die Stichprobenauswahl entgegen der gängigen Bezeichnung den Begriff „*A-priori-Determinierung*“. Wird die Quotenstichprobe als einzige Strategie verwendet, ist sie dann auch mit einer streng induktiven Forschungsstrategie unvereinbar. Die Ausprägungskombinationen finden sich in einem Quotenplan wieder.

*Quotenpläne* sind je nach Forschungsinteresse und der Kenntnis relevanter Merkmale mehr oder weniger differenziert und können sich auf unterschiedlichste Sachverhalte beziehen. Eine gängige Variante ist die Quotierung nach soziodemographischen Merkmalen wie Alter, Geschlecht, Bildungsgrad, Ethnizität usw., es gibt aber auch ganz andere Möglichkeiten. So sollten im Rahmen eines Methodenseminars im Wintersemester 2012/13 an der Technischen Universität Berlin mehrere Gruppendiskussionen zur Untersuchung

der Anforderungen von Wissenschaftlern an eine moderne Universitätsbibliothek durchgeführt werden. In der Konzeptionsphase dieser Studie ergaben sich Überlegungen, dass die Anforderungen an eine moderne Universitätsbibliothek möglicherweise mit den verschiedenen Fachdisziplinen als auch den unterschiedlichen Beschäftigungsverhältnissen variieren können. Folglich wurde zur Rekrutierung der Gruppendiskussionsteilnehmer ein Quotenplan erstellt, um sowohl eine bestimmte Anzahl an Vertretern verschiedener Fachdisziplinen als auch verschiedener Statusgruppen zu erhalten.

Gleichzeitig wurde sowohl mit dem Konzentrations- als auch dem Streuungsprinzip gearbeitet. Dies insofern, als dass die Teilnehmer einer einzelnen Gruppendiskussion nach Fächern und Status relativ homogen sein sollten (*Konzentrationsprinzip*), während die Teilnehmer verschiedener Gruppen möglichst heterogen (*Streuungsprinzip*) waren.

Das *theoretische Sampling* (ausführlich: Strübing, Kapitel 32 in diesem Band) durchzieht den gesamten Forschungsprozess und verknüpft dabei offen-induktives und theoriegeleitete-deduktives Vorgehen miteinander. Das Sampling ist zu Beginn des Forschungsprozesses offen (Strauss/Corbin 1996: 153) und wird im Verlauf durch vergleichende Analysen mit Hilfe von Konzentrations- und Kontrastprinzip sowie dem *Sampling extremer, kritischer, typischer oder sensibler Fälle* (ausführlich: Hering/Schmidt, Kapitel 37 in diesem Band) immer zielgerichteter und fokussierter. Die Stichprobenstrategie des theoretischen Samplings kann deshalb auch als Mischform zwischen willkürlicher und kriteriengeleiteter Auswahl bezeichnet werden.

---

## 17.4 Datenkorpus und Datenauswertung – Auswahl innerhalb des Materials

Sind die Fälle ausgewählt und ist das Material erhoben, besteht die zweite Ebene der Stichprobenentscheidung darin, im gesammelten Datenmaterial die „*Analyseeinheiten*“ (*datenorientierte Stichprobe*) abzustecken.

### 17.4.1 Theoriegeleitete Ableitung der Analyseeinheiten aus der Forschungsfrage

Bei *theoriegeleiteter qualitativer Forschung* können die Analyseeinheiten aus dem Fokus der Forschungsfragen abgeleitet werden. So haben sich Lenz et al. (2006) im Rahmen ihres Forschungsprojekts „Institutionalisierungsprozesse in Zweierbeziehungen“ unter anderem mit der Erforschung von Beziehungsanfängen im Wandel beschäftigt. Neben verschiedenen anderen qualitativen Daten wurden Spielfilme als „*Dokumente sozialer Wirklichkeit*“ (Lenz 2006: 123) analysiert. Auf der ersten Ebene der Stichprobenentscheidung musste dabei festgelegt werden, welche Filme nach welcher Stichprobenstrategie besonders relevant sind.

In Abhängigkeit von der theoretischen Einengung der Fragestellung wurde dann innerhalb der gesammelten Filme nochmals selektiert und zwar wurden die Teile der Filme herausgesucht, welche für eine Feinanalyse (Kurt/Herbrik, Kapitel 33 in diesem Band) benötigt werden. Lenz (2006: 118) definiert hierzu das Kennenlernen als „erste Begegnung von zwei einander bislang fremden Personen“. Damit ist der *Startpunkt der relevanten Sequenzen* festgelegt. Wann immer sich das potentielle Paar im Film zum allerersten Mal begegnet, *beginnt die Analyseeinheit*.

Für das *Ende der Analyseeinheit* gibt es unterschiedliche Möglichkeiten: Die *erste* Begegnung kann mit einer Verabschiedung der betreffenden Personen offiziell beendet werden oder aber es schließt sich per Kameraschnitt eine Sequenz mit anderen Personen oder an einem anderen Tag an, wobei das nächste Aufeinandertreffen des Paares dann eindeutig eine neue Situation darstellen muss.

Mit dieser Definition können alle Kennenlernszenen aus den gesammelten Filmen herausgesucht und einer Feinanalyse unterzogen werden. Die restlichen Bestandteile der Filme dienen dann als *Kontextinformationen*, in die die jeweiligen Situationen eingebettet sind. Diese theoriegeleitete Variante der Auswahl innerhalb des Materials kann auf andere Fragestellungen, Datensorten und Auswertungsmethoden übertragen werden.

#### 17.4.2 Offen-induktive Auswahl der Analyseeinheiten

Wesentlich schwieriger gestaltet sich dies beim *offen-induktiven Vorgehen*. In gewisser Weise soll der Forscher im Laufe des Forschungsprozesses an genau diesen Punkt kommen, d.h. aus dem offenen Sampling auf erster Ebene, *verengt sich Schritt für Schritt der Fokus*. Man würde mit einer relativ unkonkreten Fragestellung an solche Filme herantreten und diese im Gesamten auf sich wirken lassen.

Die intensivere Beschäftigung mit dem Material schließt eine *ständige Selbstbeobachtung des Forschers* mit ein. Wann immer dabei Sachverhalte unsere Aufmerksamkeit besonders binden oder wir uns daran reiben, sind dies Hinweise auf theoretische Relevanz. Diese Anhaltspunkte müssten weiterfolgt werden, um herauszufinden, was genau das theoretisch Relevante dabei ist. So lässt sich das Thema weiter fokussieren und auch eine gezieltere Auswahl innerhalb des Datenkorpus bzw. weiterer Filme erzielen. Strauss und Corbin (1996: 156) bezeichnen dies als „*Wachsamkeit für bedeutsame Daten*“.

Dieses Vorgehen verlangt dem Forscher hohe Kompetenzen beim Erkennen von theoretisch Relevantem, also hohe *theoretische Sensibilität*, und äußerste Selbstreflexivität ab. Zudem muss der Forscher die *Struktur* und den *Fokus*, die er *nicht* vorab gesetzt hat, in der Fülle der Daten finden und darf sich nicht darin verlieren. Für unerfahrene Forscher ist dieses Vorgehen daher nur bedingt ratsam.

### 17.4.3 Vergleichende Analysen von typischen, ähnlichen oder abweichenden Fällen

Ein weiterer wichtiger Aspekt, der die Auswahl innerhalb des Materials betrifft, sind die *vergleichenden Analysen von typischen, ähnlichen oder abweichenden Fällen*, die besonders bei der Grounded Theory von Bedeutung sind. Diese äußerst aufwendig erscheinenden Analysen müssen und sollen nicht nur auf der ersten Stichprobenebene stattfinden. Es ist explizit angedacht, dass dies auch innerhalb des Materials erfolgt. Je nach Ressourcenlage können auch nur wenige Interviews geführt werden oder Beobachtungen stattfinden, während die vergleichende Analyse intensiv auf Basis der Auswahl innerhalb des Materials erfolgt.

Nach dem offenen Kodieren des Materials (Strübing, Kapitel 32 in diesem Band) ergeben sich bestimmte Kategorien und Konzepte. Anschließend wird innerhalb des Materials ganz bewusst nach Bestätigungen und Widerlegungen oder Abweichungen von diesen Beobachtungen gesucht. Gemäß einer *falsifikationistischen Forschungsstrategie* lässt sich überprüfen (nicht beweisen!), ob die gezogenen Schlüsse auch Widerlegungsversuchen standhalten oder sich bei geringster Belastung auflösen. Obwohl ein idealtypisches Vorgehen im Sinne der Grounded Theory immer aufwendig bleibt, und nicht jede Forschungsarbeit auf Theorieentwicklung abzielt, ist es dennoch sinnvoll diese Kerngedanken im Rahmen der Analyse des Datenmaterials auch bei anderen qualitativen Ansätzen zu berücksichtigen, um theoretische Generalisierbarkeit zu erlangen.

---

## 17.5 Präsentation der Forschungsergebnisse

In allen Ausarbeitungen zu empirischen Studien, egal ob es sich dabei um Forschungsberichte, Monographien oder Aufsätze handelt, werden Daten präsentiert (Meyer/Meier zu Verl, Kapitel 15 in diesem Band). Die Einbeziehung der Originaldaten hat in der qualitativen Forschung unterschiedliche Ziele:

- Zunächst dienen die Ausschnitte zur *Stützung und Nachvollziehbarkeit der Argumentation* (Flick und Meyer/Meier zu Verl, Kapitel 29 und 15 in diesem Band).
- Gleichzeitig füllen Ausschnitte aus dem Material den Forschungstext mit „Leben“, weil die Daten für sich sprechen dürfen.

Die passenden Ausschnitte zu finden, ist häufig keine leichte Entscheidung. Während der Analyse entdeckt der Forscher aber Ausschnitte im Material, die für ihn bedeutsam sind, weil sie etwas auf den Punkt bringen oder einen speziellen Blick auf den Gegenstandsbereich liefern. Diese Ausschnitte sollten mit einer Erläuterung, z.B. in Form eines Memos (Strübing, Kapitel 32 in diesem Band), sofort festgehalten werden, denn ein strukturiertes Vorgehen bei der Analyse und Interpretation des Materials hilft einerseits, schon vorab zu entscheiden, welche Ausschnitte aus den Daten für die spätere Veröffentlichung wichtig sein könnten und verhindert andererseits eine unreflektierte Auswahl.

## 17.6 Stichprobengröße

Ein weiteres schwieriges Thema im Bereich der Stichprobenziehung ist die *Stichprobengröße bei qualitativer Sozialforschung*. Es gibt auch hier im Gegensatz zur quantitativen Sozialforschung keine allgemeinen Anhaltspunkte. Klar ist, dass die Auswertung nicht-standardisierter Daten zumeist aufwändiger ist, weshalb qualitative Stichproben häufig klein sind. Allerdings bedeutet das Ziel einer theoretischen Verallgemeinerung der Stichprobenergebnisse auch, dass es meist gar nicht erforderlich ist, große Stichproben zu ziehen. Ein breites und tiefgehendes Spektrum an Informationen über den Gegenstandsbereich lässt sich schon mit relativ wenigen gut gewählten Fällen erreichen und ab einem bestimmten Grad finden sich nur noch redundante Informationen. Nach der Idealvorstellung des *theoretischen Samplings* (Strübing, Kapitel 32 in diesem Band) wären prinzipiell Überlegungen zur Stichprobengröße zweitrangig, ergibt sie sich doch automatisch – man erhebt solange, bis sich *theoretische Sättigung* einstellt.

In der Forschungspraxis birgt dieses einfache Konzept jedoch Probleme. Zunächst kann sich theoretische Sättigung je nach Erfahrung des Forschers in ganz unterschiedlichen Phasen einstellen. Des Weiteren müssen, wie bereits erwähnt, im Rahmen von Forschungsanträgen bei Mittelgebern Kostenkalkulationen vorgelegt werden, die auch den Umfang der Datenerhebung beinhalten. Obwohl es keine eindeutigen Festlegungen für die Stichprobengrößen in der qualitativen Forschung geben kann, lassen sich zumindest *Empfehlungen für die Forschungspraxis* machen.

### 17.6.1 Verfügbare Ressourcen (Zeit, Geld, Personal)

Als erste Restriktion für die Stichprobengröße sind immer die zur Verfügung stehenden Ressourcen (Zeit, Geld, Personal) zu nennen. Dabei versteht es sich von selbst, dass im Rahmen von *Bachelor- oder Masterarbeiten* keine umfangreichen empirischen Analysen durchgeführt werden können, wenn der Zeitaufwand für Feldzugang, Datenerhebung, Datenauswertung, Aufarbeitung des Forschungsstandes usw. auch nur grob geschätzt wird. Wenn man z.B. von nur fünf einstündigen Interviews ausgeht, die knapp kalkuliert einen zeitlichen Umfang von insgesamt zehn Stunden für das Recherchieren der Interviewpartner, Terminabsprachen und die Durchführung beanspruchen, muss man zusätzlich noch die fünf- bis sechsfache Zeit, also etwa 30 Stunden, zum Transkribieren einberechnen. Daraus ergibt sich allein nur für die Datenerhebung und -aufbereitung eine volle Arbeitswoche. Gibt es zusätzlich noch Komplikationen im Forschungsprozess, wenn z.B. das Auffinden geeigneter Interviewpartner langwieriger ist als erwartet, vereinbarte Termine verschoben werden oder doch nicht stattfinden und erneut nach Interviewpartnern recherchiert werden muss, so kann dieser zeitliche Aufwand auch deutlich höher sein. Es sollten deshalb im Rahmen von empirischen qualitativen Bachelor- oder Masterarbeiten lieber weniger Daten erhoben werden, z.B. eben nur maximal fünf Interviews, für die dann auch Zeit ist, um sie ordentlich auszuwerten.

## 17.6.2 Inhaltliche Erwägungen

Ein zweiter Einflussfaktor, der direkt mit den Ressourcen zusammenhängt, ist die *Reichweite der Fragestellung*. Je umfassender das Forschungsvorhaben ist, desto mehr Datenmaterial muss tendenziell gesammelt werden. Je spezifischer die Fragestellung ist, desto eher kann theoretische Sättigung (Flick 2007) erzielt werden. Die Darstellung des Kennenlernens von Paaren in Spielfilmen ist z.B. ein relativ enges Thema. Wenn jedoch, wie bei Lenz, eine Längsschnittstudie geplant ist, um Wandlungsprozesse zu berücksichtigen oder wenn zusätzlich auch noch interkulturelle Vergleiche angestrebt werden, dann weitet sich das Thema aus und die Anforderungen an die Stichprobe steigen. Zumindest müssen Daten für jeden Vergleichszeitpunkt und für jede Vergleichskultur gesammelt werden, d.h. ein weiterer Faktor für die Stichprobengröße ist die *Anzahl der untersuchten Kriterien*, der eben mit der Reichweite der Fragestellung zusammenhängt.

Hinzu kommt noch der *Grad der Schichtung der Merkmale*. Denkt man hierbei etwa an einen Quotenplan, der nur nach drei Altersklassen und Geschlecht differenziert, wobei für jede Kombination nur zwei Interviews durchgeführt werden sollen, ergibt sich daraus bereits eine Stichprobengröße von n=12. Je mehr Merkmale also miteinbezogen werden und je größer die Anzahl der Ausprägungen ist, desto größer muss auch die Stichprobe sein, damit jede Zelle ausreichend mit Daten besetzt ist. Für die Fokussierung der Fragestellung und die Eingrenzung der Anforderungen an die Stichprobengröße spielt es demnach auch eine Rolle *wie die untersuchte Population beschaffen ist*, d.h. wie homogen oder heterogen sie ist und welche Besonderheiten sie aufweist (Mason 2010).

Ebenso beeinflussen die *Art der Datenerhebung* sowie die geplante *Auswertungsmethode* die Stichprobengröße. Dies hängt wiederum direkt mit den verfügbaren Ressourcen zusammen. Die Sammlung von Zeitungsartikeln (Klein, Kapitel 64 in diesem Band) oder Archivdokumenten (Salheiser, Kapitel 62 in diesem Band) kann (muss aber nicht) schneller vonstatten gehen als längere Beobachtungen (Thierbach/Petschick, Kapitel 66 in diesem Band) oder das Durchführen von Interviews (Helfferich, Küsters, Vogl, Häder/Häder und Friedrichs/Schwinges, Kapitel 39, 40, 41, 42 und 43 in diesem Band). Hermeneutische Interpretationen (Kurt/Herbrik, Kapitel 33 in diesem Band) oder Videoanalysen (Schnettler/Tuma, Kapitel 68 in diesem Band) sind sicherlich zeitintensiver als inhaltanalytische Bearbeitungen des Datenmaterials (Mayring/Fenzl, Kapitel 38 in diesem Band).

Als letztes soll noch die *Qualität der Daten* als Einflussfaktor für die Stichprobengröße genannt werden. Eine geringe Anzahl an gelungenen Interviews oder Beobachtungen kann mehr Erkenntnisgewinn liefern, als eine Vielzahl an schlecht ausgewählten Fällen.

Alle diese Faktoren müssen demnach gegeneinander abgewogen werden, um im Einklang mit den zur Verfügung stehenden Ressourcen das eigentliche Forschungsvorhaben realisieren zu können.

### 17.6.3 Konventionen der Forschungsliteratur und -praxis

Neben inhaltlichen Erwägungen sind ein weiteres Kriterium Konventionen der Forschungsliteratur und -praxis. Für qualitative Interviews zeigt sich dabei, dass die Vorschläge aus der Forschungsliteratur je nach Auswertungsmethode zwischen 5 und 50-60 Interviews schwanken (Mason 2010). Zu einem ähnlichen Intervall kam Mason auch in einer Metaanalyse von 560 Dissertationen in Großbritannien und Irland, in denen jeweils qualitative Interviews durchgeführt wurden. Die Arbeiten enthielten eine Fülle von Forschungsansätzen und -designs wie etwa Case Studies (Hering/Schmidt, Kapitel 37 in diesem Band), Grounded Theory (Strübing, Kapitel 32 in diesem Band), Inhaltsanalysen (Mayring/Fenzl, Kapitel 38 in diesem Band), Diskursanalysen (Traue et al., Kapitel 34 in diesem Band), Phänomenologie, Hermeneutik (Kurt/Herbrik, Kapitel 33 in diesem Band) usw. Guest et al. (2006) haben aber in einer systematischen Reanalyse einer ihrer eigenen Studien festgestellt, dass die meisten vergebenen Kategoriencodes bereits mit 6 von 60 Interviews abgedeckt waren, d.h. auch mit kleinen Stichprobengrößen z.B. weniger als zehn Interviews lassen sich schon wesentliche Themen des Untersuchungsgegenstandes herausarbeiten.

---

## 17.7 Verallgemeinerbarkeit qualitativer Stichproben

Nicht jedes Forschungsvorhaben muss das Ziel der größtmöglichen Verallgemeinerung der Ergebnisse über die Stichprobe hinaus verfolgen. Gerade im Bereich anwendungsorientierter Forschung z.B. den weitverbreiteten Evaluationsstudien (Döring, Kapitel 9 in diesem Band) ist es häufig ausreichend, den untersuchten Fall so breit wie möglich zu erfassen um beurteilen zu können, ob der Ist-Zustand auch dem Soll-Zustand entspricht. Ist dem nicht so, lassen sich aus den Ergebnissen konkrete Maßnahmen für den konkreten Einzelfall entwickeln. Wesentlich ist demnach nicht nur die Frage inwieweit die Stichprobenergebnisse verallgemeinerbar sind, sondern die Klärung, welchen *Grad der Verallgemeinerung* man überhaupt selbst anstrebt.

### 17.7.1 Generalisierungsziele

Mayring (2007) formuliert acht verschiedene *Generalisierungsziele*, die sich anhand ihres Verallgemeinerungsgrades in eine Rangfolge bringen lassen.

1. An oberster Stelle stehen die *universellen deterministischen Gesetzesaussagen*, die über Zeit und Raum hinweg Gültigkeit besitzen, aber für die sozialwissenschaftliche Forschung, selbst im quantitativen Bereich, wenig bedeutsam sind.
2. Da deterministische Gesetzesaussagen für soziales Handeln, menschliche Einstellungen usw. eher unzutreffend sind, versucht sich die quantitative Forschung zumindest im Aufstellen von *probabilistischen Aussagen*, die Abweichungen in einem gewissen Maße zulassen.

3. Für qualitative Forschung sind jedoch eher die anderen von Mayring genannten Generalisierungsziele wichtig. So kann versucht werden aus beobachteten Zusammenhängen *Regeln* herauszuarbeiten,
4. die gegebenenfalls nur unter bestimmten Bedingungen Gültigkeit besitzen, also *kontextspezifische Aussagen*.
5. Des Weiteren lassen sich Verallgemeinerungen auch innerhalb des Materials über das Aufdecken von *Gemeinsamkeiten und Unterschieden* vornehmen.
6. *Deskriptive Studien* und
7. *explorative Studien* können dazu genutzt werden, um eine Grundlage für Verallgemeinerungen in Folgestudien zu liefern.
8. Mayring (2007) führt schließlich auf der niedrigsten Stufe des Generalisierungsgrades die *Verallgemeinerung von Prozeduren zur Generierung von Ergebnissen* an.

### **17.7.2 Numerische Generalisierung (Statistische Inferenz)**

Für eine Verallgemeinerung der Stichprobenergebnisse gibt es unabhängig vom angestrebten Generalisierungsgrad im Wesentlichen drei unterschiedliche Herangehensweisen (Seale 2000: 106ff.). Im Rahmen der quantitativen Forschung wird mittels induktiver Statistik numerisch generalisiert, während in der qualitativen Forschung theoretische Generalisierung und dichte Beschreibung zwei Möglichkeiten der Verallgemeinerung darstellen.

### **17.7.3 Theoretische Generalisierung (Logische Inferenz)**

Die theoretische Generalisierung lässt sich anhand des theoretischen Samplings gut nachvollziehen. Beim offenen Kodieren werden allgemeine Codes für die beobachteten Sachverhalte gesucht. Dabei entstehen erste Vermutungen über *Zusammenhänge, Beziehungen, Typisches oder Untypisches*. Indem dann gezielt nach Bestätigungen und Widerlegungen gesucht wird, konkretisieren sich die Vermutungen und es lässt sich auf abstrakterer theoretischer Ebene formulieren, *unter welchen Bedingungen bestimmte Aussagen zutreffen und welche Einschränkungen bestehen*. Nach und nach entsteht so eine *gegenstandsorientierte Theorie*. Seale (2000: 109) spricht dabei in Anlehnung an Mitchell von *logischer Inferenz* im Gegensatz zu statistischer Inferenz.

### **17.7.4 Dichte Beschreibung (Idealtypen)**

Eine andere Variante der Generalisierung ist die dichte Beschreibung, die z.B. in deskriptiven oder explorativen Studien verwendet werden kann. Im Rahmen der dichten Beschreibung erfasst der Forscher nicht nur das, was konkret passiert, sondern er versucht auch interpretativ zu rekonstruieren, welche nichtbeobachtbaren Motive, Ziele

und Bedeutungen hinter bestimmten Handlungen stecken können. Dazu muss ein Einzelfall umfassend studiert werden. Die daraus gezogenen Erkenntnisse können, müssen aber nicht auf andere Einzelfälle übertragbar sein. So lässt sich mit Soeffners (2010: 172) Bemerkung zum Verhältnis von konkretem Fall und idealem Typus (Verallgemeinerung) im Rahmen der sozialwissenschaftlichen Hermeneutik abschließend zusammenfassen:

„Ziel der Analyse ist die verdichtende Rekonstruktion eines objektivierten Typus sozialen Handelns aus seinen konkreten, fallspezifischen Ausprägungen heraus. Dieser objektivierte Typus ist insofern ‚Idealtypus‘, als er mit dem Zweck konstruiert wird, einerseits gegenüber der Empirie insofern systematisch Unrecht zu haben, als er das Besondere im Einzelfall nur unzulänglich wiedergibt, andererseits gerade dadurch dem Einzelfall zu seinem Recht zu verhelfen, dass er das historisch Besondere vor dem Hintergrund struktureller Allgemeinheit sichtbar abhebt.“

## Literatur

- Flick, Uwe (2007): Qualitative Sozialforschung. Reinbek: Rowohlt
- Flick, Uwe (2010): Design und Prozess qualitativer Forschung. In: Flick, Uwe/Kardorff, Ernst von/Steinke, Ines (Hg.): Qualitative Forschung. Reinbek: Rowohlt. 252-264
- Fuchs, Marek/Lamnek, Siegfried/Luedtke, Jens/Baur, Nina (2009): Gewalt an Schulen. 1994 – 1999 – 2004. Wiesbaden: VS Verlag
- Glaser, Barney G./Strauss, Anselm L. (1999): The Discovery of Grounded Theory. New York: de Gruyter
- Guest, Greg/Bunce, Arwen/Johnson, Laura (2006): „How Many Interviews Are Enough? In: Field Methods 18 (1): 59-82
- Krackhardt, David (1999): The Ties that Torture. In: Bacharach, Samuel B./Andrews, S. B./Knoke, David (Hg.): Networks In and Around Organizations. Greenwich: JAI Press. 183-210
- Lenz, Karl (2006): Paare in Spielfilmen – Paare im Alltag. In: Mai, Manfred/Winter, Rainer (Hg.): Das Kino der Gesellschaft – die Gesellschaft des Kinos. Köln: Halem. 117-147
- Mason, Mark (2010): Sample Size and Saturation in PhD Studies Using Qualitative Interviews. In: FQS 11 (3): Art. 8
- Mayring, Philipp (2007): Generalisierung in qualitativer Forschung. In: FQS 8 (3): Art. 26
- Merkens, Hans (2010): Auswahlverfahren, Sampling, Fallkonstruktion. In: Flick, Uwe/Kardorff, Ernst von/Steinke, Ines (Hg.): Qualitative Forschung. Reinbek: Rowohlt. 286-298
- Petschke, Sabine (2007): Der Datenkorpus des DFG-Projektes „Die Performanz visuell unterstützter mündlicher Präsentationen“. In: Schnettler, Bernt/Knoblauch, Hubert (Hg.): Powerpoint-Präsentationen. Konstanz: UVK. 287-292
- Seale, Clive (2000): The Quality of Qualitative Research. London et al.: Sage
- Soeffner, Hans-Georg (2010): Sozialwissenschaftliche Hermeneutik. In: Flick, Uwe/Kardorff, Ernst von/Steinke, Ines (Hg.): Qualitative Forschung. Reinbek: Rowohlt. 164-174
- Strauss, Anselm L./Corbin, Juliet M. (1996): Grounded Theory. Weinheim: Beltz
- Strübing, Jörg (2013): Qualitative Sozialforschung. Eine komprimierte Einführung für Studierende. München: Oldenbourg

Michael Häder und Sabine Häder

---

## 18.1 Stichprobe oder Totalerhebung?

Eine wichtige Entscheidung im Rahmen der Designentwicklung bei einer quantitativen Erhebung betrifft die Frage, ob nur ein Teil der Elemente untersucht werden soll oder ob alle Elemente der Grundgesamtheit einzubeziehen sind. Bei der *Grundgesamtheit* handelt es sich um all jene Elemente, die aufgrund von bestimmten Eigenschaften für den For-scher von Interesse sind. So kann beispielsweise bei einer Wahlstudie die Grundgesamtheit bestimmt werden als die Menge an Personen, die in einem Land zu einem bestimmten Stichtag wahlberechtigt sind. Bei den Personen handelt es sich um die Elemente der Grundgesamtheit, deren gemeinsame Eigenschaft die Wahlberechtigung zu einem bestimmten Stichtag ist.

Werden alle Elemente der Grundgesamtheit in eine Untersuchung einbezogen, wie dies etwa bei einer Volkszählung erfolgt, so spricht man von einer *Totalerhebung*. Für eine Totalerhebung spricht, dass bestimmte Fehler, die bei einer Stichprobenziehung stets auftreten können, ausgeschlossen und die gewonnenen Befunde damit belastbarer und genauer sind. Gegen eine Totalerhebung ist der mit der Untersuchung einer sehr großen Zahl an Elementen verbundene finanzielle, personelle und zeitliche Aufwand ins Feld zu führen. So verursachte die Volkszählung in der Bundesrepublik Deutschland im Jahr 1987 finanzielle Kosten von umgerechnet mehr als 500 Mio. Euro.

Auch forschungspraktische Gründe sprechen oftmals gegen eine Totalerhebung. So sind in der Regel die für eine Untersuchung zur Verfügung stehenden Mittel begrenzt. Diese können effektiver bzw. konzentrierter eingesetzt werden, wenn nur eine relativ kleine Anzahl an Elementen untersucht werden soll. Mithilfe von beispielsweise Nachfassaktionen und Incentives können so bei einer Stichprobenerhebung eine bessere Ausschöpfung und bessere Resultate erzielt werden als bei einer Totalerhebung.

Angebracht und sinnvoll kann es in der empirischen Sozialforschung sein, beispielsweise Mitarbeiterbefragungen (Borg, Kapitel 57 in diesem Band) als Totalerhebungen zu

konzipieren. Hier ist eventuell die Grundgesamtheit nicht zu groß und es besteht außerdem eine Möglichkeit, alle Elemente der Grundgesamtheit zur Teilnahme zu motivieren. Schließlich ist es innerhalb von Unternehmen auch eher denkbar, dass die notwendigen Mittel für eine solche Erhebung vorhanden sind.

Bei einer *Stichprobe* handelt es sich um eine Auswahl von Elementen aus der Grundgesamtheit. Idealerweise soll diese Auswahl nach bestimmten mathematisch-statistisch begründeten Regeln erfolgen, um einen berechtigten Schluss auf die zugrunde liegende Gesamtheit zulassen zu können.

---

## 18.2 Zufallsauswahl

In der Praxis der quantitativen empirischen Sozialforschung kommen verschiedene Strategien zur Rekrutierung von Stichproben zum Einsatz. Neben zufallsgesteuerten Auswahlen handelt es sich dabei um bewusste und um willkürliche Auswahlen. Letztere eignen sich vor allem für die Schätzung von Modellparametern. Aufgrund der besonderen Bedeutung, die Zufallsauswahlen im Rahmen der quantitativen Sozialforschung besitzen, werden diese hier näher vorgestellt.

Zufallsauswahlen (auch *Wahrscheinlichkeitsauswahlen* oder engl. *random samples*) zeichnen sich dadurch aus, dass die Wahrscheinlichkeit für jedes Element der Grundgesamtheit, ein Element der Stichprobe zu werden, berechnet werden kann und größer als null ist. Nur bei Zufallsstichproben ist das Konzept des Stichprobenfehlers anwendbar, das Aussagen über die Präzision der Schätzung ermöglicht. Der Stichprobenfehler bezeichnet die Streuung der Stichprobenkennwerte um den Wert des gesuchten Parameters in der Gesamtheit. Je größer der Stichprobenfehler, desto breiter ist das Vertrauensintervall (auch *Konfidenzintervall*), das den gesuchten Parameter mit einer bestimmten Wahrscheinlichkeit überdeckt. Die Größe des Stichprobenfehlers hängt von der Streuung der Messwerte in der Grundgesamtheit und vom Stichprobenumfang ab.

### 18.2.1 Undercoverage, Overcoverage und fehlende Werte

Der Einsatz von Zufallsauswahlen setzt das Vorhandensein einer *Liste* (auch *Auswahlrahmen*, engl. *frame*) voraus, in der möglichst alle Elemente der Grundgesamtheit verzeichnet sind. In dem Fall, dass in einem Auswahlrahmen nicht alle Elemente der Grundgesamtheit vollständig enthalten sind, wird von *Undercoverage* gesprochen. Enthält der Auswahlrahmen jedoch Elemente der Grundgesamtheit mehrfach, so handelt es sich um *Overcoverage*. Under- und Overcoverage sind im Rahmen der Stichprobeneinschätzung ernste Probleme, da der systematische Ausfall von Elementen für die Ziehung zu verzerrten Schätzern führen kann. Ein Beispiel für einen derartigen Undercoverage ist, dass gegenwärtig das Telefonbuch nicht mehr als Auswahlrahmen verwendet werden kann, da sich Nichteingetragene und Eingetragene hinsichtlich soziodemographischer Merkmale unterscheiden.

Es existieren verschiedene Varianten von Zufallsauswahlen.

### 18.2.2 Einfache Zufallsauswahl

Zunächst soll die Aufmerksamkeit den *einfachen Zufallsauswahlen* (engl. *simple random samples*) gelten. Eine einfache Zufallsauswahl folgt dem Modell einer Urne bzw. einer Los-trommel, in der sich alle Elemente der Grundgesamtheit gut durchmischt befinden und aus der zufällig die Stichprobe, d.h. eine bestimmte Anzahl an Elementen, entnommen wird. Bei einfachen Zufallsauswahlen haben damit alle Elemente der Grundgesamtheit die gleiche Chance in die Stichprobe zu gelangen. Der Einsatz einfacher Zufallsauswahlen ist in der empirischen Sozialforschung beispielsweise bei Absolventen- oder Kundenbefragungen möglich, sofern hier eine geeignete Liste vorliegt. Auch wenn es sich bei der Grundgesamtheit um die Einwohner einer Gemeinde handelt, können einfache Zufallsauswahlen auf der Grundlage der Melderegister eingesetzt werden. Schließlich können auch standardisierte Inhaltsanalysen (Kuckartz/Rädiker, Kapitel 27 in diesem Band), bei denen beispielsweise die Ausgaben einer Tageszeitung (Klein, Kapitel 64 in diesem Band) analysiert werden sollen, auf dieses Auswahlverfahren zurückgreifen. Allerdings stehen in der Praxis nur relativ selten geeignete Auswahlrahmen für eine einfache Zufallsauswahl zur Verfügung, so dass häufig andere Designs zur Anwendung kommen.

### 18.2.3 Mehrstufige Auswahlen

Bei mehrstufigen Auswahlen werden zunächst Einheiten auf einer ersten Ebene zufällig ausgewählt, sogenannte *Primary Sampling Units* (PSUs), z.B. Gemeinden. Innerhalb dieser werden wiederum Elemente zufällig gezogen (*Seconday Sampling Units*, SSUs), z.B. Haushalte. Dieses Prinzip lässt sich auf weiteren Ebenen fortführen, z.B. in der Auswahl von Personen innerhalb der ausgewählten Haushalte in den gezogenen Gemeinden.

Bei Face-to-face-Umfragen in Deutschland, die an die Allgemeinbevölkerung der Bundesrepublik gerichtet sind, müssen mehrstufige Zufallsauswahlverfahren herangezogen werden, weil der empirischen Sozialforschung kein Auswahlrahmen für einfache Zufallsauswahlen zur Verfügung steht. Als qualitativ hochwertig gelten Einwohnermeldeamtstichproben, bei denen zunächst eine Anzahl von Gemeinden und innerhalb dieser eine konstante Anzahl von Personen pro Gemeinde ausgewählt werden. Bei diesem Verfahren haben alle Personen die gleiche Chance ausgewählt zu werden, insofern ist der Standardfehler nicht größer als bei einfachen Zufallsauswahlen. Nachteilig wirkt sich jedoch aus, dass eine größere Homogenität der Untersuchungsmerkmale aufgrund der Klumpung der Elemente in den Gemeinden erwartet werden muss. Dies führt zu einer geringeren Präzision der Schätzung (einem höheren Design-Effekt) als bei der einfachen Zufallsauswahl.

### 18.2.4 Geschichtete Auswahlen

Wenn Listen zur Verfügung stehen, die weitere Informationen über die Zielpersonen (wie z.B. deren Alter) enthalten, ist es günstig, diese Informationen in den Stichprobenplan aufzunehmen. Dieses Vorgehen bezeichnet man als Schichtung (engl. *Stratification*). Damit kann entweder sichergestellt werden, dass in der Zufallsstichprobe auch alle interessierenden Subpopulationen entsprechend ihrem Anteil in der Grundgesamtheit vertreten sind (proportionale Schichtung). Hierbei kann der Standardfehler u.U. sogar geringer ausfallen als bei der einfachen Zufallsauswahl. Es ist aber auch möglich, in der Gesamtheit geringer besetzte Schichten stärker zu berücksichtigen, um auch für diese Schätzungen mit höherer Präzision angeben zu können (disproportionale Schichtung). Hier müssen, sollen Aussagen über die Gesamtheit getroffen werden, die ungleichen Auswahlwahrscheinlichkeiten bei der Schätzung durch Gewichtung ausgeglichen werden.

### 18.2.5 Klumpenauswahlen

Die Umfrageforschung nutzt persönlich-mündliche Interviews, um von jenen Personen, die in die Stichprobe gelangt sind, Informationen zu gewinnen. Dazu werden diese Personen in der Regel von Interviewern in ihren Haushalten aufgesucht. Hierbei handelt es sich um eine kosten-, zeit- und personalaufwändige Vorgehensweise. Aus forschungspraktischen Gründen ist es deshalb angebracht, dass die zu interviewenden Personen räumlich möglichst eng beieinander wohnen (Klumpung). Dadurch wird der Aufwand, den die Interviewer für die Befragung aufzubringen haben, begrenzt. Ein bekanntes Beispiel, bei dem eine Klumpenauswahl zum Einsatz kam, ist die PISA-Studie, bei der die Befragungen von Schülern in Schulklassen erfolgten.

Zu beachten ist, dass die so gewonnenen Aussagen aufgrund des Designeffekts weniger belastbar sind als jene, die auf einer einfachen Zufallsauswahl beruhen. Es liegt nahe, dass mit steigender Anzahl an Klumpen, mit sinkender Klumpengröße und mit steigender Heterogenität innerhalb der Klumpen die Qualität der Klumpenstichprobe steigt.

---

## 18.3 Stichprobenumfang

Ein wichtiger Aspekt der Stichprobenziehung betrifft die Bestimmung des erforderlichen *Stichprobenumfangs*. Mit dem Stichprobenumfang (als Symbol  $n$ , engl. *sample size*) ist die Zahl der ausgewählten Elemente gemeint, die in die Untersuchung einbezogen werden sollen. Bei der Frage nach der erforderlichen Stichprobengröße wird zumeist der Repräsentativitätsbegriff bemüht. An dieser Stelle wird Repräsentativität als Synonym für eine Zufallsauswahl verstanden, da nur hierfür Verallgemeinerungen von Stichprobenkennwerten auf Parameter der Grundgesamtheit inferenzstatistisch fundiert erklärt sind. Es

bleibt allerdings festzuhalten, dass sich für den Begriff „Repräsentativität“ keine eigentliche mathematische Definition finden lässt (Gabler/Quatember 2012).

Zielführender ist es hingegen, wenn die Frage nach dem erforderlichen Stichprobenumfang mit der Frage nach der Größe des gewünschten *Vertrauensintervalls*, d.h. des Intervalls, in dem der gesuchte Parameter aus der Grundgesamtheit mit hoher Wahrscheinlichkeit liegt, verbunden wird: Wie präzise oder wie genau sollen die mit den Daten der zufällig gezogenen Stichprobe gefundenen Ergebnisse überhaupt sein? Um dies darzulegen, müssen zunächst einige Ausführungen zum *Stichprobenfehler* vorangestellt werden.

Wenn man sich vorstellt, dass sich in einer Urne 100 Kugeln befinden und davon 50 weiß und 50 schwarz sind und man nun aus dieser Urne eine zufällige Stichprobe von zehn Kugeln zieht, so kann man nicht unbedingt erwarten, dass mit Sicherheit fünf von den gezogenen Kugeln weiß und fünf schwarz sind. Den Unterschied zwischen der in der Grundgesamtheit vorhandenen und der in der Stichprobe ermittelten Verteilung nennt man *Stichprobenfehler*, er wird mit dem Buchstaben  $e$  abgekürzt. Würde man die Ziehung der Stichprobe häufiger wiederholen, so zeigte sich, dass dieser Stichprobenfehler um den tatsächlichen Wert normal verteilt ist. Dies ist eine wichtige Erkenntnis, die bei der Stichprobenziehung und konkret bei der Bestimmung des Vertrauensintervalls genutzt werden kann.

Um das Problem der Bestimmung des erforderlichen Stichprobenumfangs bzw. die Bedeutung des Vertrauensintervalls transparent zu machen, soll angenommen werden, dass im Rahmen einer Zufallsauswahl eine Anzahl an Personen ( $n = 1.000$ ) befragt wurde und dass bei diesen Personen eine bestimmte Ausprägung einer bestimmten Eigenschaft ermittelt wurde. Es könnte sich beispielsweise um die Absicht handeln, eine bestimmte Partei wählen zu wollen. Die größte Heterogenität (Streuung) in der Grundgesamtheit würde dann bestehen, wenn 50% diese Partei wählen wollten und 50% nicht. So stellt sich als nächstes die Frage, mit welchem Vertrauen man dem ermittelten Wert begegnen kann. Schließlich ist bekannt, dass bei Zufallsauswahlen Stichprobenfehler auftreten können. Um die Sicherheit bzw. das Vertrauensintervall anzugeben, wird eine Spanne ermittelt, in der – mit einer bestimmten Wahrscheinlichkeit – der wirkliche Wert in der Grundgesamtheit liegt. In einem ersten Schritt legt man den Wert für die Irrtumswahrscheinlichkeit  $\alpha$  fest. Eine für die Sozialwissenschaften übliche Größe beträgt 5%. Dies bedeutet, dass bei theoretisch angenommenen 100 unterschiedlichen Stichproben in nur fünf Stichproben ein Irrtum auftritt, d.h. der wahre Wert nicht vom ermittelten Vertrauensintervall überdeckt wird. Als Streuung für die Grundgesamtheit wählt man – da diese nicht bekannt ist – z.B. den Maximalwert 50% für einen Anteil, d.h.  $p=0,5$ . Danach kann das Intervall ermittelt werden, in dem – mit der gerade angegebenen Wahrscheinlichkeit – der tatsächliche Wert in der Grundgesamtheit liegt. Um einen Irrtum völlig auszuschließen, müssten in einer Totalerhebung alle Personen der Grundgesamtheit befragt werden. Aufgrund der Annahme der Normalverteilung des Stichprobenfehlers kann man aber das Vertrauensintervall berechnen.

Um nun im Vorfeld einer Untersuchung den notwendigen Stichprobenumfang zu bestimmen, muss man sich über folgende Aspekte Klarheit verschaffen. Der erste Aspekt

betrifft die Irrtumswahrscheinlichkeit, die akzeptiert werden soll. Zweitens spielen die Größe der Grundgesamtheit und drittens der erwartete Anteilswert eine Rolle. Schließlich muss der zulässige Stichprobenfehler festgelegt werden.

Es sei

$n_{srs}$  Stichprobenumfang unter SRS

$N$  Umfang der Grundgesamtheit

$z_{\alpha/2}$  tabellierter Wert aus der Standardnormalverteilung, für  $\alpha = 0,05$  ist  $= 1,96$

$p$  Anteil des interessierenden Merkmals in der Stichprobe, entweder aus Voruntersuchung bekannt oder ungünstigsten Fall  $p = 0,5$  annehmen

$e$  zulässiger absoluter Stichprobenfehler,  $2e$  entspricht der Länge des Konfidenzintervalls

1. Schätzung von Anteilswerten bei kleinem Auswahlsatz ( $n/N < 0,05$ )

$$n_{srs} \geq \left( \frac{z_{\alpha/2}}{e} \right)^2 \cdot p \cdot (1-p)$$

2. Schätzung von Anteilswerten bei großem Auswahlsatz

Berücksichtigung des Korrekturfaktors  $1 - \frac{n}{N}$  (Ziehen ohne Zurücklegen)

$$n_{srs} \geq \frac{N \cdot z_{\alpha/2}^2 \cdot p \cdot (1-p)}{z_{\alpha/2}^2 \cdot p \cdot (1-p) + N \cdot e^2}$$

Borg (2000:144) hat für die Bestimmung des erforderlichen Stichprobenumfangs eine Tabelle publiziert, die sehr hilfreich ist (vgl. Tab. 18.1).

Aus Tab. 18.1 kann man folgendes entnehmen: Zunächst ist die Größe der Grundgesamtheit ( $N$ ) in Betracht zu ziehen. Danach geht es um die Entscheidung für die zu tolerierende Fehlergröße  $e$ . Unterstellt man beispielsweise eine Größe der Grundgesamtheit zwischen  $N = 10.000$  und  $50.000$  Elementen und fordert man, dass die Schätzung auf  $\pm 3$  Prozentpunkte genau sein soll, so sind etwa  $n = 1.000$  zufällig ausgewählte Elemente der Grundgesamtheit erforderlich, um diese Maxime zu erfüllen. Es handelt sich dabei zudem um eine konservative Schätzung, d.h. um eine besonders sichere Variante einer Schätzung, da davon ausgegangen wird, dass die zu schätzenden Anteile bei 50% liegen.

Bei den bisherigen Überlegungen zu den Zufallsstichproben wurden die konkreten Realisierungsmöglichkeiten für die Ziehung von Stichproben nur kurz angedeutet. So ist es in der Praxis oft fraglich, ob es tatsächlich ein Verzeichnis gibt, das alle Elemente der Grundgesamtheit enthält und für die Sozialforschung zugänglich ist. Sollte das nicht der Fall sein, ist es nicht möglich, nach dem oben beschriebenen Modell uneingeschränkte Zufallsauswahlen für Bevölkerungsbefragungen zu ziehen. Deshalb sind zahlreiche weitere Strategien für die Ziehung von Stichproben entwickelt und implementiert worden.

**Tab. 18.1** Minimaler Stichprobenumfang  $n$  für gegebenen absoluten Stichprobenfehler  $e$  bei Irrtumswahrscheinlichkeit  $\alpha = 0,05$  für Anteile  $p = 0,5$  und  $p = 0,8$  (oder  $p = 0,2$ )

<b><math>p = 0,5</math></b>			<b><math>p = 0,8</math> (oder <math>p = 0,2</math>)</b>		
<b>N</b>	<b><math>e = 0,03</math></b>	<b><math>e = 0,05</math></b>	<b>N</b>	<b><math>e = 0,03</math></b>	<b><math>e = 0,05</math></b>
200	168	132	200	155	110
300	234	168	300	208	135
400	291	196	400	252	152
500	340	217	500	289	165
750	440	254	750	357	185
1.000	516	278	1.000	406	197
3.000	787	341	3.000	556	227
7.500	934	365	7.500	626	238
10.000	964	370	10.000	639	240
50.000	1.045	381	50.000	674	245
100.000	1.056	383	100.000	678	245

## 18.4 Stichprobendesigns für Bevölkerungsbefragungen in Deutschland

Wie festgestellt, ist es oft nicht möglich mit einfachen Zufallsauswahlen zu arbeiten. Neben dem inhaltlichen Ziel einer empirischen Studie und den sich daraus ergebenden Implikationen für die Auswahl des Erhebungsmodus sind auch die in der Praxis für die Stichprobenziehung vorhandenen Möglichkeiten bei der Entscheidung für einen Befragungsmodus und eine Auswahlstrategie entscheidend. So sind verschiedene Varianten von Zufallsauswahlen entwickelt worden, die sich an einer Umsetzung des Konzepts der Minimierung des Total Survey Error (Faulbaum, Kapitel 31 in diesem Band) orientieren. Weiterhin muss berücksichtigt werden, dass das zu wählende Stichprobenverfahren auch vom kulturellen Kontext (Braun, Kapitel 56 in diesem Band) abhängt. So wird man in Ländern mit einer niedrigen Telefondichte nicht auf die Idee kommen, telefonische Bevölkerungsbefragungen durchzuführen. Für Deutschland wurden einige Standardverfahren für bestimmte Befragungsmodi entwickelt, die im Folgenden kurz dargestellt werden.

### 18.4.1 Persönliche mündliche Befragung: Das ADM-Design

Beim Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (ADM) handelt es sich um eine Standesorganisation, zu der sich die bedeutendsten kommerziellen Umfrageinstitute der Bundesrepublik zusammengeschlossen haben. Ein Anliegen des ADM ist es, den Mitgliedsinstituten mit seinem ADM-Design einen Stichprobenplan für qualitativ

hochwertige Zufallsauswahlen für Befragungen in der Allgemeinbevölkerung der Bundesrepublik zur Verfügung zu stellen. Dieses Design gilt in der Bundesrepublik als „good practice“.

Beim ADM-Design (ADM 1999, Hoffmeyer-Zlotnik 1997: 33ff.) handelt es sich um eine mehrstufig geschichtete Klumpenauswahl. Diese kombiniert die oben beschriebenen Designmerkmale und stellt eine Vorgehensweise für die zufällige Rekrutierung von Personen für persönlich-mündliche Befragungen (Stocké, Kapitel 45 in diesem Band) dar. Es sieht die folgenden Schritte vor: In der ersten Stufe werden Gebiete zufällig ausgewählt, danach erfolgt auf der zweiten Stufe in diesen Gebieten die Auswahl der Haushalte und schließlich auf der dritten Stufe die Ermittlung der eigentlichen Zielpersonen in den zuvor ausgewählten Haushalten.

*Der erste Schritt:* Das bewohnte Gebiet der Bundesrepublik lässt sich auf der Basis der amtlichen Statistik in etwas mehr als 50.000 Flächen aufteilen. Hierbei handelt es sich um die feinste von der amtlichen Statistik mit Daten unterstützte Gliederung. Diese Flächen lassen sich bei Gemeinden mit über 10.000 Einwohnern zudem mithilfe von digitalisierten Straßenkarten abgrenzen. Aus diesen Flächen wird vom ADM nach dem Zufallsprinzip eine Stichprobe gezogen, wobei die dabei ermittelten Elemente als *Sample Points* bezeichnet werden. Während bis 2003 die etwa 80.000 Stimmbezirke der Bundesrepublik die Grundlage für die Festlegung der Sample Points bildeten, erfolgt seit 2003 eine verfeinerte Aufteilung der Sample Points in Straßenabschnitte.

Für die Bildung der Sample Points verwendet man die zur Verfügung stehenden kleinsten administrativen Flächeneinheiten bis auf Baublockebene. Dazu können inzwischen amtliche innerstädtische Gliederungen und Sachdaten systematisch herangezogen werden. Zu diesen Daten zählen Stadtteildaten, Einträge in Telefonverzeichnisse usw.

Einer Konvention folgend werden für das ADM-Design für 1.000 zu realisierende Interviews im westlichen Teil Deutschlands 210 und im östlichen (seit der Gebietsreform 1996) 48 solcher Sample Points ausgesucht. Diese sind so angelegt, dass sie durchschnittlich 600 bis 700 Haushalte umfassen. Als Schichtungskriterien dienen beim ADM-Design die Bundesländer, die Regierungsbezirke und die Kreise in Kombination mit einer Gemeindetypisierung.

„Die Zellen im Stichprobensystem werden aus den Landkreisen/kreisfreien Städten und den Gemeindegrößenklassen gebildet. Die Gemeindegrößenklassen werden hierbei in erweiterter Form in einer 10er-Einteilung herangezogen. [...] Die Gemeinden, die nicht zu einer Stadtregion gehören, werden innerhalb einer Schicht (Landkreise) entsprechend ihrer tatsächlichen (politischen) Größe zugeordnet.“ (ADM 1999: 86f.)

*Der zweite Schritt:* Hier besteht das Ziel darin, in den Sample Points die *Zielhaushalte zu bestimmen*, in denen später eine Befragung stattfinden soll. Dazu erfolgt, ausgehend von einer bestimmten Startadresse, eine zufällige Begehung (auch als Random Walk oder Random Route bezeichnet) des Sample Points. Für dieses Random Walk erstellen die Erhebungsinstitute Begehungsvorschriften, denen die Interviewer zu folgen haben.

Der Interviewer geht, beginnend bei der ihm vorgegebenen Startadresse, nach der festgelegten Vorschrift eine bestimmte Strecke ab und trägt in ein Adressenauflistungsblatt die ausgewählten Haushalte ein. Nun kann nach drei Varianten weiter vorgegangen werden: Die *erste* Möglichkeit besteht darin, dass das *Adressauflistungsblatt* an das Erhebungsinstitut geschickt wird und man dort dann (nochmals) eine zufällige Auswahl der zu befragenden Haushalte vornimmt (Häder/Häder 1997: 43ff.). Bei der zweiten und dritten Möglichkeit werden die Interviews von den Interviewern jeweils sofort geführt. Bei der *zweiten* Variante darf der Interviewer den Random Walk so lange weiter fortsetzen, bis es ihm gelungen ist, die ihm vorgegebene Anzahl an Interviews zu führen. Bei der *dritten* Variante wird den Interviewern eine Höchstzahl an zu kontaktierenden Haushalten vorgegeben. Nur bei diesen Haushalten darf der Interviewer dann versuchen, zu einem Interview zu gelangen. Für den Fall, dass er in den Haushalten niemanden erreicht beziehungsweise wenn er mit einer Verweigerung konfrontiert wird, reduziert sich die Zahl der von ihm realisierten Interviews.

Es ist für das Erhebungsinstitut sehr kompliziert, die genaue Einhaltung der – teilweise recht komplizierten – Begehungsvorschrift zu kontrollieren. Hinterhäuser, nicht existierende Hausnummern, Kreisverkehre, das Erreichen der Stadtgrenze, das Auffinden von Hausmeistern in Anstalten wie Krankenhäusern oder Schulkomplexen, Gemeinschaftsunterkünfte, Grundstücke auf denen vor dem Hunde gewarnt wird und eine Reihe ähnlicher Probleme stellen eine hohe Herausforderung an die Sorgfalt und die Ausdauer der Interviewer dar. Falls überhaupt möglich, so müsste die Kontrolle der Einhaltung der Regeln für den Random Walk mithilfe einer aufwändigen Nachbegehung durch einen anderen Interviewer erfolgen.

Der *dritte Schritt* des ADM-Stichprobendesigns gilt der *Auffindung der Zielperson* im ermittelten Haushalt. Hierfür werden in der Umfragepraxis vor allem zwei Strategien eingesetzt. Zunächst wird eine systematische Auflistung aller Haushaltmitglieder, die zur Grundgesamtheit gehören und damit für eine Befragung prinzipiell infrage kommen, vorgenommen. Diese hat so zu erfolgen, dass die Personen geordnet nach ihrem Alter in eine Liste eingetragen werden. Nun kann (Variante eins) jene Person um die Teilnahme an der Befragung gebeten werden, die als letzte Geburtstag hatte. Eine andere Möglichkeit (Variante zwei) besteht darin, auf dem Adressauflistungsblatt Zufallszahlen aufzudrucken und danach die im Haushalt zu befragende Zielperson zu ermitteln (sog. *Schwedenschlüssel*, engl. *Kish Grid*). Bei dieser Methode erscheint der Spielraum der Interviewer für mögliche Abweichungen von der Regel stärker eingeschränkt zu sein als bei der Geburtstagsmethode. Deshalb wird dieses Vorgehen bei persönlich-mündlichen Befragungen häufiger bevorzugt.

Dieser Auswahlschritt ist eine leicht einsehbare Notwendigkeit, da ja nicht jene Personen befragt werden sollen, die gerade in der Wohnung anzutreffen sind oder die sich für die Teilnahme an einer Befragung besonders interessieren. Auf der anderen Seite stellt es für den Interviewer eine gewisse Herausforderung dar, eine unter Umständen befragungswillige Person abzuweisen und stattdessen entweder den Haushalt noch einmal zu

kontaktieren (falls die eigentliche Zielperson nicht anwesend sein sollte) oder auch ganz auf ein Interview zu verzichten (falls die richtige Zielperson die Befragung verweigern sollte).

Während – zumindest theoretisch – bei der Kombination der beiden ersten Auswahlschritte alle Elemente der Grundgesamtheit die gleiche Chance hatten, zu Bestandteilen der Stichprobe zu werden, ist dies beim letzten Auswahlschritt nicht mehr der Fall. Hier haben Mitglieder kleinerer Haushalte eine größere Chance als jene aus größeren Haushalten. Um dies auszugleichen, müssen bei der Auswertung Gewichtungsprozeduren eingesetzt werden.

Es gilt, auf Grenzen und Fehlerquellen des ADM-Stichprobendesigns aufmerksam zu machen: Vor allem bei der Vorgabe einer zu befragenden Nettoanzahl an die Interviewer (Variante zwei) besteht die Gefahr, dass leicht erreichbare Personen in der Stichprobe überrepräsentiert sind und dadurch die Qualität der Stichprobe sinkt. Als ein Gegenmittel könnte die Zahl der von den Interviewern vorzunehmenden Kontaktversuche deutlich höher festgelegt werden. Auch durch die beiden anderen vorgestellten Varianten lässt sich diesem Fehler bis zu einem gewissen Grade begegnen. Zu beachten sind die höheren Kosten, die auf diese Weise entstehen.

Neben der schwer zu kontrollierenden Einhaltung der Vorschriften für den Random Walk stellt auch die korrekte Umsetzung des Schwedenschlüssels bzw. der Geburtstagsmethode eine nicht oder nur schwer zu überprüfende Fehlerquelle dar. So mag der eine oder andere Interviewer unter Umständen der Verlockung erliegen und eher eine befragungswillige Person im Haushalt interviewen, als unverrichteter Dinge aufzugeben (zu Fälschungen siehe Blasius, Kapitel 22 in diesem Band).

Ein weiteres Problem besteht darin, die bei einer Studie auftretenden *Ausfallgründe* (siehe Engel/Schmidt, Kapitel 23 in diesem Band) systematisch zu erfassen. Aus methodischer Sicht ist es erforderlich, zwischen stichprobenneutralen und systematischen Ausfällen zu unterscheiden. Auch bei der Führung der Statistik zum Nonresponse besitzt der Interviewer einen relativ großen und nur kompliziert kontrollierbaren Spielraum. So lässt sich im Nachhinein nur relativ schwer überprüfen, ob eine Zielperson etwa deshalb nicht an der Umfrage teilgenommen hat, weil sie nicht ausreichend deutsch spricht, oder ob hierfür ein anderer Grund, etwa eine explizite Verweigerung, vorlag. Hierzu sind Interviewerschulungen und –kontrollen (Glantz/Michael, Kapitel 21 in diesem Band) durch die Erhebungsinstitute notwendig.

Abschließend ist festzustellen, dass das ADM-Design vor allem Probleme bei der Kontrolle der Feldarbeit mit sich bringt. Üblich sind Überprüfungen der zumeist auf Honorarbasis tätigen Interviewer durch Telefonanrufe sowie mithilfe von Postkarten. Dabei werden die Teilnehmer einer Befragung danach gefragt, ob überhaupt ein Interviewer in ihrem (Ziel-)Haushalt war, wie er die zu befragende Person ermittelt hat, wie lange die Befragung gedauert hat, welche Themen angesprochen wurden u.ä.

### **18.4.2 Persönlich-mündliche und schriftlich-postalische Befragungen: Registerstichproben**

Angesichts der oben im Zusammenhang mit dem ADM-Design geschilderten Probleme gilt derzeit in der Bundesrepublik die Verwendung einer Registerstichprobe als die bestmögliche Vorgehensweise bei der Stichprobenziehung für Bevölkerungsbefragungen (Koch 1997: 99ff., Albers 1997: 117ff., ADM 1999: 119ff.). Als Register können in Deutschland die von den Einwohnermeldeämtern auf Gemeindeebene geführten Dateien benutzt werden. Diese stehen jedoch nur für wissenschaftliche Untersuchungen, die im öffentlichen Interesse liegen, zur Verfügung. Als Beispiel für die Nutzung von Registerstichproben lassen sich die gut dokumentierten ALLBUS-Studien der Jahre 2006, 2008 und 2010 sowie die Runden 1-5 des European Social Surveys heranziehen. Das Design dieser Studien war zweistufig:

*Die erste Stufe: Auswahl von Gemeinden.* Die Auswahl erfolgt – ähnlich wie beim ADM-Design – geschichtet nach Bundesländern, Regierungsbezirken und Kreisen. In der Regel sind die Gemeinden identisch mit den Sample Points. Sehr große Städte können aber durchaus mit mehreren Sample Points in die Auswahl gelangen. Für den ALLBUS wurden z.B. bei einer Studie für den Westen Deutschlands 111 Sample Points, die in 103 Gemeinden liegen, und für den östlichen Teil des Landes 51 in 45 Gemeinden gelegene Sample Points ausgewählt.

*Die zweite Stufe: zufällige Auswahl der Zielpersonen aus den Melderegistern,* z.B. 40 Personen-Adressen pro Sample Point. Die Melderegister verfügen – von Bundesland zu Bundesland unterschiedlich – neben den Adressen über einige zusätzliche Informationen zu ihren Einwohnern. Diese können für wissenschaftliche Studien zur Verfügung gestellt werden. Dabei handelt es sich vor allem um das Alter, das Geschlecht sowie um die Nationalität der in den jeweiligen Gemeinden lebenden Personen. Diese Informationen sind wichtig für die Kontrolle der Interviewer.

Der Einsatz von Registerstichproben ist zeit- und personalaufwändig und daher teuer. So muss vor allem die Zusammenarbeit mit jedem einzelnen Meldeamt organisiert werden. Hierbei liegen in der Praxis sehr unterschiedliche Erfahrungen vor. Auch hat sich gezeigt, dass die in den Meldeämtern geführten Dateien nicht immer aktuell sind.

Positiv hervorzuheben ist bei diesem Design vor allem, dass der Spielraum für die Interviewer vor Ort stark eingeschränkt wird. Es entfällt die Adressauflistung durch die Interviewer und damit die Unsicherheit, die mit diesem Schritt verbunden ist. Auch eine Auswahl der zu befragenden Zielperson im Zielhaushalt ist nicht mehr erforderlich. Stattdessen werden den Interviewern der Name und die Adresse der zu befragenden Person vorgegeben.

Eine Kontrolle der Interviewertätigkeit wird zusätzlich dadurch ermöglicht, dass die Altersangabe, die in den Interviews ermittelt wurde, mit derjenigen abgeglichen werden kann, die bei den Meldeämtern registriert ist. Zudem ist das Führen einer Nonresponse-Dokumentation wesentlich erleichtert. Dazu stehen ebenfalls die Informationen aus dem

Melderegister zur Verfügung. Damit kann die Feldarbeit deutlich transparenter beschrieben werden als beim ADM-Design.

Für schriftlich-postalische Befragungen (Reuband, Kapitel 47 in diesem Band) sind Registerstichproben die einzige Möglichkeit, bundesweite Bevölkerungsbefragungen auswahlstrategisch anspruchsvoll umzusetzen.

### **18.4.3 Telefonische Befragung**

Ähnlich wie z.B. auch in den USA stellen Telefonbücher in der Bundesrepublik seit der Aufhebung der Eintragungspflicht im Dezember 1991 (Heckel 2002: 13) keine vollständige Auflistung aller Teilnehmer mehr dar. Gegenwärtig steigt die Anzahl nicht gelisteter Anschlüsse weiter und wird – regional stark unterschiedlich – auf bis zu 50% geschätzt.

Die beste Praxis bei der Ziehung von Telefonstichproben stellt das sog. Gabler-Häder-Design dar (Gabler/Häder 1997, 2002, Heckel 2002), das bei ZUMA in Mannheim entwickelt wurde. Dabei werden zunächst aus der Menge aller im Telefonbuch registrierten Telefonnummern die offensichtlich nicht zur Grundgesamtheit der Wohnbevölkerung gehörenden Anschlüsse bzw. Nummernfolgen gelöscht. Dazu zählen bei allgemeinen Bevölkerungsumfragen vor allem die Geschäftsanschlüsse. Nun ordnet man die verbleibenden Nummern pro Ortsnetzbereich (davon gibt es in Deutschland 5.200) der Größe nach an. Bildet man daraufhin Blocks vom Umfang 100, z.B. von 0 bis 99, von 100 bis 199 usw., kann man die vorhandenen Telefonnummern in diese Blocks einordnen. Z.B. liegt die Nummer 06321 33703 im Block 33700-33799 des Ortsnetzbereiches 06321. Die grundlegende Vermutung ist nun, dass sich die nicht gelisteten Telefonnummern ebenfalls innerhalb der mit mindestens einer eingetragenen Nummer besetzten Blocks befinden. Deshalb werden im Weiteren alle Blocks, in denen keine Rufnummer im Verzeichnis gelistet ist, ausgesondert. In allen verbleibenden Blocks, in denen mindestens eine Rufnummer gelistet ist, werden alle 100 denkbaren Ziffernfolgen generiert, also zum Beispiel im Ortsnetzbereich 06321 die Folgen 33700, 33701, 33702, 33703, ..., 33799. Die Menge aller dieser Ziffernfolgen in allen Ortsnetzbereichen stellt dann den Auswahlrahmen dar, aus dem die Stichproben für Telefonumfragen (Hüfken, Kapitel 46 in diesem Band) zufällig gezogen werden können. Dieses Universum enthält eingetragene Telefonnummern (z.B. 33703), nichteingetragene Telefonnummern und Ziffernfolgen, die gegenwärtig nicht geschaltet sind. GESIS stellt für die akademische Sozialforschung Festnetzstichproben und mittlerweile auch Mobilfunkstichproben, die nach einem ähnlichen Prinzip generiert werden, kostenfrei zur Verfügung.

Bei der Auswahl der Zielperson im telefonisch kontaktierten Haushalt kommt schließlich zumeist die Geburtstagsmethode zum Einsatz. Das bedeutet, dass diejenige Person ermittelt wird, die im Haushalt als nächstes (oder alternativ zuletzt) Geburtstag hat(te) und die zugleich zu der zu untersuchenden Grundgesamtheit gehört. Diese ist dann die am Telefon zu befragende Zielperson. Möglich ist auch der Einsatz des Kish Grids für die Ermittlung der Zielperson.

Hilfreich ist es weiterhin, wenn zusätzlich zur Generierung der Telefonnummern eine Zuordnung zur Gemeinde, in der sich der jeweilige Anschluss befindet, erfolgt (Heckel 2002). Damit wird eine Schichtung der Stichprobe nach regionalen Merkmalen wie Kreisen oder Regierungsbezirken ermöglicht. Der ADM hat das Gabler-Häder-Design adaptiert und um eine solche regionale Schichtung ergänzt (von der Heyde 2002).

Indes bleiben einige andere Probleme bestehen:

- So muss eine Strategie gefunden werden, um mit angewählten Fax- und Geschäftsanschlüssen, mit Anrufbeantwortern und mit Mehrfachanschlüssen umzugehen. Haushalte mit Mehrfachanschlüssen sind deshalb ein statistisches Problem, da sie eine höhere Inklusionswahrscheinlichkeit besitzen als Haushalte mit nur einem Anschluss. Als adäquate Lösung bietet sich hier an, mit Hilfe einer entsprechende Frage die im Haushalt vorhandene Anzahl an Anschlüssen zu ermitteln und im Designgewicht zu berücksichtigen.
- Ein weiteres Problem stellen die Gate-Keeper dar. Hier handelt es sich um vorgeschaltete elektronische Filter, die bewirken, dass nur vom Anschlussinhaber zuvor ausgesuchte Anrufer durchgestellt werden. Eine Lösung für dieses Problem muss noch gefunden werden.
- Die Personenauswahl im Haushalt muss in einer aufgrund des hohen Abbruchrisikos kritischen Phase des Interviews erfolgen. Während einem Interviewer, der an der Wohnungstür um ein Gespräch bittet, verschiedene Strategien zur Verfügung stehen, um die Zielperson von der Notwendigkeit eines Interviews zu überzeugen, reduzieren sich diese Möglichkeiten am Telefon beträchtlich. Tests haben gezeigt (Friedrichs 2000: 171ff.), dass von Ankündigungsschreiben eine positive Wirkung auf die Teilnahmebereitschaft ausgehen kann. Das Versenden solcher Schreiben setzt jedoch voraus, dass die vollständigen Adressen der Zielpersonen dem Veranstalter vor Beginn der Befragung bekannt sind. Dies ist allerdings bei einem solchen Vorgehen nicht der Fall, so dass diese Technik nicht zum Einsatz kommen kann und die Haushalte in der Regel „kalt“ kontaktiert werden müssen.

Gegenwärtig verfügen ca. 11% der Haushalte in der Bundesrepublik über keinen Festnetzanschluss mehr (*Mobile Onlys*), sondern sind nur über ein Mobiltelefon zu erreichen (Heckel/Wiese 2012). Diese unterscheiden sich aber hinsichtlich soziodemographischer Merkmale von den Personen, die auch über einen Festnetzanschluss verfügen. Die Mobile Onlys sind eher jünger, niedriger gebildet, Singles und wohnen in Städten. Weil ein Verzicht auf diesen Personenkreis zu verzerrten Schätzern führen würde, musste ein Weg gefunden werden, um bei telefonischen Befragungen auch die Mobile Onlys zu erreichen. Die Lösung bietet ein Dual-Frame-Ansatz. Dabei werden die Befragungen parallel sowohl über das Festnetz als auch über den Mobilfunk geführt. Im Anschluss erfolgt dann die Synthese beider Substichproben mithilfe einer relativ aufwendigen Gewichtungsprozedur (Häder/Gabler/Heckel 2009: 21ff.).

Der Auswahlrahmen besteht hier aus der Menge an Mobilfunknummern, die durch die Bundesnetzagentur vergeben wird (zurzeit rund 306 Millionen Nummern). Die Zahl der Telefonbucheinträge für Mobilfunknummern beträgt aber nur 2,3 Millionen. Die Erstellung der Auswahlgrundlage erfolgt nun so:

Zunächst wird pro Vorwahl der tatsächlich gültige Rufnummernbereich ermittelt. Bewährt hat sich die Betrachtung von 10.000er Blöcken, über die man Informationen zusammenstellen kann. Da sind erstens die Einträge der Telefonbuch-CD. Zweitens sind auch Internetrecherchen erfolgsbringend. Zusätzlich kann man drittens auch spezielle Websites nutzen, die Nutzern Hinweise auf versteckte Kosten geben. Viertens gibt es auch Blöcke, die für technische Dienste der Anbieter oder Mailboxen reserviert sind und von daher auch für eine Befragung ausscheiden. Aus allen Informationsquellen wird pro Anbieter auf der 10.000er Blockebene oder, wenn die Informationen feiner vorliegen, auch auf einer 1.000er Block-Ebene ein Listenblatt erstellt, das Auskunft darüber gibt, welche Rufnummernblöcke innerhalb einer Vorwahl aktiv genutzt werden und welche noch inaktiv sind. Diese Listen bilden die Grundlage der Generierung der Rufnummern (Häder/Gabler/Heckel 2009: 43).

---

## **18.5 Stichprobenverfahren für andere Datentypen**

Stichproben spielen bei der quantitativen Sozialforschung nicht nur im Rahmen von Bevölkerungsbefragungen eine Rolle. So können im Rahmen der (computergestützten) Inhaltsanalyse z.B. auch Zeitungsartikel (Klein, Kapitel 64 in diesem Band) oder Dokumente (Salheiser, Kapitel 62 in diesem Band) ausgewählt werden. Die hierfür üblichen Auswahlstrategien werden in den entsprechenden Kapiteln vorgestellt. Stichprobenprobleme im Rahmen der amtlichen Statistik behandeln Hartmann/Lengerer (Kapitel 71 in diesem Band).

## Literatur

- ADM, Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. (1999) (Hg.): Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis. Opladen: Leske + Budrich
- Albers, Ines (1997): Einwohnermelderegister-Stichproben in der Praxis. Ein Erfahrungsbericht. In: Gabler, Siegfried / Hoffmeyer-Zlotnik, Jürgen H.P. (Hg.): 117-126
- Borg, Ingwer (2000): Führungsinstrument Mitarbeiterbefragung: Theorien, Tools und Praxiserfahrungen. Göttingen: Verlag für Angewandte Psychologie
- Friedrichs, Jürgen (2000): Effekte des Versands des Fragebogens auf die Antwortqualität bei einer telefonischen Befragung. In: Hüfken, Volker (Hg.): 171-182
- Gabler, Siegfried/Hoffmeyer-Zlotnik, Jürgen H.P. (1997) (Hg.): Stichproben in der Umfragepraxis. Opladen: Westdeutscher Verlag
- Gabler, Siegfried/Häder, Sabine (1997): Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland. In: ZUMA-Nachrichten 41: 7-18
- Gabler, Siegfried/Häder, Sabine (2002) (Hg.): Telefonstichproben. Methodische Innovationen und Anwendungen in Deutschland. Münster New York München: Waxmann
- Gabler, Siegfried/Quatember, Andreas (2012): Das Problem mit der Repräsentativität von Stichprobenerhebungen. VSMS-Jahrbuch 2012. 17-19
- Häder, Michael/Häder, Sabine (1997): Adressvorlaufverfahren: Möglichkeiten und Grenzen. Eine Untersuchung am Beispiel der Erhebung Leben Ostdeutschland. In: Gabler, Siegfried/Hoffmeyer-Zlotnik, Jürgen H.P. (Hg.): 33-42
- Heckel, Christiane/Wiese, Kathrin (2012): Sampling Frames for Telephone Surveys in Europe. In: Häder, Sabine/Häder, Michael/Kühne, Mike (Hg.): 103-119
- Heckel, Christiane (2002): Erstellung der ADM-Telefonauswahlgrundlage. In: Gabler, Siegfried/Häder, Sabine (2002) (Hg.): 11-31
- Häder, Sabine/Gabler, Siegfried/Heckel, Christiane (2009): Stichprobenziehung, Gewichtung und Realisierung. In: Häder, Michael/Häder, Sabine (Hg.): 21-49
- Häder, Michael/Häder, Sabine (2009) (Hg.): Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung. Wiesbaden: VS Verlag
- Hüfken, Volker (2000) (Hg.): Methoden der Telefonumfragen. Wiesbaden: Westdeutscher Verlag
- Häder, Sabine/Häder, Michael/Kühne, Mike (2012) (Hg.): Telephone Surveys in Europe: Research and Practice. Heidelberg: Springer Verlag
- Hoffmeyer-Zlotnik, Jürgen H.P. (1997): Random Route-Stichproben nach ADM. In: Gabler, Siegfried/Hoffmeyer-Zlotnik, Jürgen H.P. (Hg.): 33-42
- Koch, Achim (1997): ADM-Design und Einwohnermelderegister-Stichprobe. Stichprobenverfahren bei mündlichen Bevölkerungsumfragen. In: Gabler, Siegfried/Hoffmeyer-Zlotnik, Jürgen H.P. (Hg.): 99-116
- Von der Heyde, Christian (2002): Das ADM-Telefonstichproben-Modell. In: Gabler, Siegfried/Häder, Sabine (Hg.): 32-45

Martin Weichbold

---

## 19.1 Einleitung

Es ist in der empirischen Sozialforschung und insbesondere der Umfrageforschung heute Common Sense, vor der eigentlichen Datenerhebung Pretests durchzuführen. Ziel ist es, die Datenerhebung ex ante, d.h. vor ihrem eigentlichen Beginn zu optimieren. Aus heutiger Sicht versteht man Pretests nicht als punktuelle Verfahren oder einmalig durchzuführende Methode (typischerweise kurz bevor ein Fragebogen ins Feld geht), sondern als eine Menge von Verfahren zur Qualitätssicherung des Erhebungsinstruments, ja des gesamten Erhebungsdesigns, die im Zuge der Fragebogenerstellung und Erhebungsplanung Anwendung finden. Je nach angewendeten Verfahren wird dabei unterschiedlichen Qualitätsdimensionen der Befragung Aufmerksamkeit geschenkt. Pretests betreffen meist das Erhebungsinstrument (Formulierung der einzelnen Fragen bzw. Antwortmöglichkeiten sowie den Fragebogen als Ganzes), aber auch dessen technische Umsetzung (bei computergestützter Befragung), die Stichprobenziehung und die Durchführung der Interviews. Sie können aber auch eingesetzt werden, um mit Hilfe von offenen Fragen Antwortkategorien zu finden, um in der Hauptbefragung geschlossene Fragen verwenden zu können oder um die Anzahl von Items in Itembatterien für die Hauptbefragung reduzieren zu können.

Die Notwendigkeit von Pretests ergibt sich aus der Tatsache, dass bei der quantitativen Befragung (im Gegensatz zu qualitativen Interviews) Fragebogen und Befragungsdesign nach Beginn der Feldphase nicht mehr verändert werden können, um die Standardisierung der Erhebung nicht zu gefährden.

Die nachfolgende Darstellung der Pretestverfahren orientiert sich am Prozessverlauf empirischer Sozialforschung. Demnach können Pretests bereits im Zuge der Fragebogenentwicklung, unter Laborbedingungen zum Testen eines Fragebogenentwurfs sowie unter Feldbedingungen zum Testen des gesamten Befragungsablaufs zur Anwendung kommen (Campanelli 2008).

## 19.2 Pretests im Zuge der Fragebogenentwicklung

Ziel in dieser ersten Phase ist es, eine optimale Formulierung von Fragen und Antwortkategorien zu finden bzw. Entscheidungen zwischen Alternativen zu treffen sowie den Aufbau des Fragebogens zu optimieren. Hierzu gibt es mehrere Möglichkeiten:

- Die Durchführung der Fragebogenentwicklung oder die Diskussion von Fragealternativen in Fokusgruppen, die von Experten aus unterschiedlichen Bereichen zusammengesetzt sein sollen (Vogl, Kapitel 41 in diesem Band).
- Tiefeninterviews zur Exploration des Verständnisses von Fragen oder Begriffen (Helferich, Kapitel 39 in diesem Band).
- Die systematische Analyse von Fragen durch Checklisten. Für den deutschen Sprachraum haben Faulbaum et al. (2009) ein Frage-Bewertungs-System entwickelt, mit dessen Hilfe jede Frage hinsichtlich verschiedener Dimensionen (z.B. Textierung, Annahmen über Befragte, Erinnerungsleistungen) überprüft und eingeschätzt werden kann.
- Neben dieser standardisierten Überprüfung hat sich in dieser Phase auch die informelle Begutachtung von Kolleginnen und Kollegen bewährt, ebenso wie Pseudointerviews im Kollegenkreis, um Verständlichkeit und Flüssigkeit von Formulierungen abzuwägen („informal methods“, Campanelli 2008: 178).

Diese Verfahren werden von manchen Autoren nicht zu Pretests im engeren Sinn gezählt, sondern als Teil der Fragebogenentwicklung gesehen.

---

## 19.3 Test von einzelnen Fragen: Pretestverfahren im Labor

Liegt ein Fragebogenentwurf (oder auch nur Entwürfe zu einzelnen Fragen) vor, geht es darum, diesen daraufhin zu untersuchen, ob die Befragten die Fragen und Antwortmöglichkeiten verstehen und in der beabsichtigten Weise interpretieren. Dazu wurden ab den 1970er-Jahren unter dem Einfluss der Kognitionspsychologie mehrere Verfahren entwickelt, die ausgehend von Modellen der Informationsverarbeitung und des Antwortverhaltens versuchen, Erkenntnisse über den Prozess der Antwortgenerierung zu erhalten.

Ausgangspunkt dieser kognitiven Verfahren ist ein vierstufiges Modell des Antwortverhaltens (Faulbaum et al. 2009: 40): Eine Frage muss zunächst verstanden werden (comprehension; Fragetext und Antwortkategorien werden wahrgenommen und interpretiert), dann erfolgt die Informationsgewinnung (retrieval; die relevanten Fakten, Einschätzungen etc. werden ermittelt), daraus wird das Urteil gebildet (judgement; die entsprechenden Schlüsse im Sinne der Frage werden gezogen), das schließlich als Antwort geäußert wird (response; Wahl der entsprechenden Antwortkategorie oder Formulierung einer Antwort). Die Validität von Antworten kann auf jeder dieser Stufen gefährdet sein, etwa indem Frageformulierungen oder Antwortvorgaben unterschiedliche Interpretationen zulassen, weil die abgefragten Informationen nicht (mehr) im Gedächtnis sind oder nur mit großem Aufwand erinnert werden können, weil die Urteilsbildung durch vorangegangene Frageinhalte

in eine bestimmte Richtung gelenkt wird oder indem vorgegebene Antwortkategorien nur schwer mit dem gebildeten Urteil in Einklang zu bringen sind. Soziale Erwünschtheit kann zudem dazu führen, dass die gegebene Antwort vom zuvor gebildeten Urteil abweicht. Mitunter wird bei der Beantwortung von Fragen auch nicht der gesamte kognitive Prozess durchlaufen, sondern abgekürzt. In solchen Fällen wird die Frage zwar beantwortet, aber möglicherweise nicht auf die eigentlich zutreffende Weise. Durch einen guten Fragebogen und eine optimal gestaltete Befragungssituation sollen die Befragten dazu gebracht werden, alle Fragen adäquat zu beantworten. Dies soll durch diese Pretestverfahren überprüft und gesichert werden.

Derartige Tests finden nicht unter Feldbedingungen, sondern ‚im Labor‘, also einer künstlichen, aber dafür einfach handzuhabenden Umgebung statt und entsprechen einem fokussierten Interview mit dem Fragebogen als Interviewfokus. Bei der Auswahl der Befragten ist es nicht notwendig, die für die Befragung intendierte Zielgruppe repräsentativ zu erfassen, sie sollte aber zumindest ähnlich sein. Unter Umständen können sogar bewusst weniger gebildete Personen ausgewählt werden, wenn bei diesen eher Verständnisschwierigkeiten zu erwarten sind. Als erforderliche Größe der Stichprobe werden 10 bis 15 Personen pro Welle angegeben. Auch hier gibt es unterschiedliche Verfahren (Faulbaum et al. 2009: 98):

- Think-aloud (Technik des lauten Denkens): Die Befragten werden gebeten, ihre Gedanken im Zuge der Beantwortung der Fragen auszusprechen. Auf diese Weise sollen das Frageverständnis und der Antwortprozess offengelegt werden, um Hinweise auf Missverständnisse, irreführende Assoziationen oder Schwierigkeiten mit der Auswahl der Antwortkategorien zu erhalten.
- Probing (Nachfragetechnik): Nach Beantwortung der Frage werden Zusatzfragen gestellt. Diese können unterschiedliche Dimensionen betreffen, nämlich das Verständnis der Frage oder einzelner Begriffe, die Wahl der Antwortkategorie, den Besitz relevanter Informationen oder die Informationsbeschaffung. Das Nachfragen kann auch unspezifisch erfolgen („Ist Ihnen bei der Frage etwas aufgefallen?“). Ziel ist es, Schwierigkeiten mit Fragen sowie unterschiedliche Interpretationen von Frageinhalten oder Antwortkategorien offen zu legen.
- Paraphrasing (Paraphrasieren): Die Befragten werden gebeten, eine Frage mit eigenen Worten zu wiederholen. Auch damit kann das Verständnis von Fragen überprüft werden.
- Confidence Ratings (Einschätzung der Verlässlichkeit einer Antwort): Die Befragten bewerten, wie sicher sie sich bei der Beantwortung einer Frage gefühlt haben, um insbesondere bei Rückerinnerungsfragen die Sicherheit zu erheben, mit der Befragte eine Antwort wählen.
- Analyse der Latenzzeit: Die Zeit zwischen dem Stellen der Frage und ihrer Beantwortung wird gemessen, wobei die Dauer der Antwortfindung als Indikator für die Schwierigkeit der Frage interpretiert wird.

## 19.4 Test von Fragebogen und Befragungssetting: Pretestverfahren im Feld

In einem letzten Schritt wird ein Fragebogen unter Feldbedingungen, d.h. innerhalb der geplanten Zielgruppe und unter den geplanten Bedingungen (z.B. in der Wohnung der Befragten) getestet. Zu den Feldbedingungen gehört auch, dass die Stichprobe aus der intendierten Grundgesamtheit stammt, wobei hierzu eine einfacher zu ziehende Quotenstichprobe für ausreichend gehalten wird (Campanelli 2009: 179f.). Allerdings bietet sich an, auch das Stichprobenverfahren und die Kontaktierung der ausgewählten Personen einem Pretest zu unterziehen, um auf diesem Wege mögliche Schwierigkeiten zu identifizieren und eine erste Einschätzung der erreichbaren Ausschöpfungsquote zu erhalten. Die Empfehlungen zur Stichprobengröße bewegen sich meist zwischen 25 und 75 Personen. Neben dem Fragebogen als Erhebungsinstrument geht es beim Feld-Pretest auch um den Ablauf der Befragung. Die eingesetzten Interviewer sollten daher entsprechende Erfahrung haben bzw. geschult sein, um Auffälligkeiten erkennen und entsprechend einschätzen zu können. Auch beim Feld-Pretest stehen mehrere Verfahren zur Verfügung:

- Prüfung der Verteilung der Antworten: Trotz der geringen Stichprobengröße lassen sich erste Rückschlüsse auf die Verteilungen der Variablen ziehen. Eine Häufung von Antwortverweigerungen oder ein großer Anteil an „Weiß nicht“-Antworten deutet auf Probleme beim Verständnis oder der Beantwortung einer Frage hin. Auch extrem schiefe bzw. einseitige Antwortverteilungen sind problematisch.
- Split Ballot (Fragebogengabelung): Pretests im Feld werden dazu genutzt, unterschiedliche Frage(-bogen-)varianten gegeneinander zu testen. Dazu werden unterschiedliche Frageformulierungen oder Fragenreihenfolgen eingesetzt, um deren Effekte zu untersuchen.
- Erhebung von Zusatz-Informationen: Im Rahmen von Feld-Pretests können zusätzliche Informationen erhoben werden, die Hinweise auf potenzielle Schwierigkeiten zu verschiedenen Aspekten der Befragung geben:
  - Report der Interviewer: Wesentlicher Teil eines derartigen Pretests ist ein Bericht der Interviewer über Probleme und Auffälligkeiten während des Interviews. Dies können offensichtliche Schwierigkeiten der Befragten beim Verstehen von Frageinhalten, Probleme mit den vorgegebenen Antwortkategorien, aber auch im Handling des Fragebogens (Verzweigungen, Antwortkarten etc.) sein. Der Report kann unstrukturiert erfolgen, indem auftretende Auffälligkeiten notiert werden, oder strukturiert, indem Anmerkungen nach vorgegebenen Dimensionen bzw. Fragen erstellt werden. Eine einfache Form ist das sog. „problem coding“, bei dem vom Interviewer während des Interviews bei jeder Frage notiert wird, ob die Frage im Sinne der Fragestellung adäquat beantwortet wurde.
  - Behavior Coding: Dazu werden Interviews aufgezeichnet und im Nachhinein mittels eines Codesystems bewertet. Sowohl das Verhalten des Interviewers als auch jenes des Befragten wird danach klassifiziert, ob die Fragen korrekt vorgelesen wurden bzw. ob die Antwort adäquat erfolgte. Bitten um Wiederholung und nähere

Erklärung der Fragen sind beispielsweise Hinweise auf Qualitätsprobleme einer Frageformulierung, ebenso wie Antwortverweigerung. Auch vorzeitiges Antworten kann auf eine ungünstige Formulierung hindeuten.

- Registrierung der Befragungsdauer: Die Dauer einer Befragung kann als Indikator für die Schwierigkeit eines Interviews (oder Schwierigkeiten mit dem Interview) interpretiert werden. Die Gesamtdauer gibt wichtige Hinweise für die Planung der Feldphase. Bei computergestützten Befragungsverfahren können zudem Zeitmarken für jede einzelne Frage bzw. Antwort aufgezeichnet werden. Auf diese Weise können Fragen mit auffällig langer (oder auch kurzer) Beantwortungsdauer identifiziert werden.
- Probing: Die Nachfragetechnik, die oben unter den Labortechniken genannt wurde, kann auch im Feld-Pretest angewendet werden. Dabei werden zu einzelnen Fragen Zusatzfragen zum Verständnis oder zur Antwortfindung gestellt. Damit weicht man allerdings von der Vorgabe ab, die Probeinterviews möglichst ähnlich den (späteren) tatsächlichen Interviews zu gestalten. Um die Befragung nicht zu sehr in die Länge zu ziehen, werden meist nur einige Fragen ausgewählt, um Zusatzfragen zu stellen („Random probe“).
- Intensivinterviews: Nach der Durchführung des Interviews werden die Befragten zum Verständnis ausgewählter Fragen erneut befragt. Dazu werden die interessierenden Fragen noch einmal vorgelesen, die Befragten sollen beschreiben, wie die Antwort zustande kam bzw. welche Assoziationen und Überlegungen mit der Frage verknüpft wurden.

---

## 19.5 Einsatz von Pretests

Pretests sollten in allen drei beschriebenen Phasen bzw. Formen zum Einsatz kommen. Welche konkreten Verfahren dabei eingesetzt werden, hängt von der jeweiligen Befragung (Komplexität, Verwendung bereits getester Fragen, Thema etc.) ab, aber auch von den vorhandenen (zeitlichen und finanziellen) Ressourcen. Generell gilt jedoch, dass kein Fragebogen ohne Pretest ins Feld gehen soll. Selbst ein kleiner Pretest ist besser als kein Pretest.

Die Durchführung von (unter Umständen mehreren) Pretests ist also bereits in der Forschungsplanung zu berücksichtigen. Nimmt man Pretests als Instrument der Qualitätskontrolle und -verbesserung in der empirischen Sozialforschung ernst, müssen die Ergebnisse auch umgesetzt werden, also Fragen umformuliert oder Antwortkategorien verändert werden. Zu bedenken ist dabei, dass Fragen, die aufgrund eines Pretests neu in den Fragebogen aufgenommen wurden, selbst wiederum einem Pretest zu unterwerfen sind. Daher muss genügend Zeit vorgesehen werden, um unter Umständen mehrere Pretest-Runden durchzuführen.

## Literatur

- Campanelli, Pamela (2008): Testing Survey Questions. In: De Leeuw, Edith/Hox, Joop J./Dillman, Don A. (Hg.): 176-200
- De Leeuw, Edith/Hox, Joop J./Dillman, Don A. (2008) (Hg.): International Handbook of Survey Methodology. New York: Lawrence Erlbaum
- Faulbaum, Frank/Prüfer, Peter/Rexroth, Margrit (2009): Was ist eine gute Frage? Die systematische Evaluation der Fragenqualität. Wiesbaden: VS Verlag
- Presser, Stanley et al. (2004): Methods for Testing and Evaluating Survey Questions. In: Public Opinion Quarterly 68: 109-130
- Prüfer, Peter/Rexroth, Margrit (2000): Zwei-Phasen-Pretesting (ZUMA Arbeitsbericht 2000/8). Mannheim: ZUMA
- Prüfer, Peter/Rexroth, Margrit (Hg) (2004): QUEST 2003. Proceedings of the 4th Conference on Questionnaire Evaluation Standards (ZUMA Nachrichten Spezial Band 9). Mannheim: ZUMA
- Statistisches Bundesamt (1996): Pretest und Weiterentwicklung von Fragebogen (Band 9 der Schriftenreihe Spektrum Bundesstatistik). Stuttgart: Metzler-Poeschel

Anja Hlawatsch und Tino Krickl

---

## 20.1 Umfragebereitschaft und Datenqualität

Die Identifizierung von Einstellungen (Latcheva/Davidov, Kapitel 55 in diesem Band) zu bestimmten Inhalten oder Produkten steht im Fokus von Befragungen (Reinecke und Helfferich, Kapitel 44 und 30 in diesem Band). Doch hinter diesem vordergründigen Ziel einer Befragung steht immer auch die Einstellung zu Befragungen an sich. Der Einfluss dieser Einstellung auf die Datenqualität (Krebs/Menold, Engel/Schmidt und Faulbaum, Kapitel 30, 23 und 31 in diesem Band) wird dabei oft unterschätzt.

Im Rahmen umfragebasierter Forschung (Reinecke, Kapitel 44 in diesem Band) wird versucht, Meinungen und Einstellungen der Befragten möglichst unverzerrt zu erfassen. In der quantitativen Sozialforschung (Stein, Kapitel 7 in diesem Band) werden nach wissenschaftlichen Kriterien Fragebögen entwickelt und standardisierte, möglichst anonyme Befragungssituationen geschaffen. Doch selbst bei Studien, die streng nach wissenschaftlichen Kriterien durchgeführt werden, bleibt eine Quelle an möglichen Antwortverzerrungen, die vom Forscher nicht im Vorfeld durch die Einhaltung methodischer Standards ausgeschlossen werden kann: die *Einstellung der Befragten zu Umfragen*. Diese beeinflusst sowohl die Bereitschaft der Teilnahme an Befragungen, als auch das Antwortverhalten selbst. Damit sind die Einstellungen der Befragten zu Umfragen eine *mögliche Quelle systematischer Verzerrungen* und können somit die *Validität* (Krebs/Menold, Kapitel 30 in diesem Band) von Umfrageergebnissen *beeinträchtigen*.

### 20.1.1 Umfrageeinstellung und Teilnahmebereitschaft

Nach Goyder (1986) gibt es einen Zusammenhang zwischen Umfrageeinstellungen und der Teilnahmebereitschaft (Engel/Schmidt, Kapitel 23 in diesem Band): Personen mit positiver Einstellung zu Umfragen haben niedrigere *Verweigerungsquoten (Unit Nonresponse)*

als Personen mit einer negativen Einstellung. Rogelberg et al. (2001) zufolge hat die Einstellung zu Umfragen auch einen Einfluss auf die *Kooperationsbereitschaft* der Teilnehmer. Personen mit positiven Umfrageeinstellungen haben im Rahmen einer postalischen Befragung (Reuband, Kapitel 47 in diesem Band) schneller geantwortet, sich stärker an die Fragebogenanweisungen gehalten und weniger Fragen unbeantwortet (*Item Nonresponse*) gelassen. Mangelnde Kooperation der Befragten kann dazu führen, dass keine angemessenen Informationen gegeben werden. Sie wird daher von Hartmann (1991: 25ff.) als wichtige Ursache für systematische Verzerrungen gesehen.

### 20.1.2 Umfrageeinstellung und Antwortverhalten

Doch nicht nur die Wahrscheinlichkeit überhaupt eine Antwort zu erhalten wird durch die Einstellung zu Umfragen beeinflusst, auch die gegebenen Antworten selbst können hiervon beeinflusst sein. Stocké (2004) kommt zum Ergebnis, dass eine positive Einstellung zu Umfragen, wenn diese stark genug in der Identität der Befragten verankert ist, dazu führen kann, dass weniger *sozial erwünschte Antworten* gegeben werden. Nach Esser (1991) ist soziale Erwünschtheit die wichtigste Form der inhaltsbezogenen systematischen Verzerrung von Antworten. Hierbei passen Befragte in Abhängigkeit von Normen und Erwartungen ihre Antworten daran an, was sie in der Befragungssituation für sozial wünschenswert erachten. Dabei erhoffen Sie sich in der Regel, soziale Anerkennung zu erhalten oder negative Sanktionen zu vermeiden.

In einer Befragung können, je nach Befragtem, theoretisch alle Fragen unterschiedliche Anreize zu sozial erwünschtem Antwortverhalten bieten. Sudman und Bradburn (1983: 303) nennen hier beispielsweise „*bedrohliche Fragen*“. Unter solchen sind Fragen zu verstehen, die für den Befragten unangenehme oder peinliche Themen betreffen, oder auch Fragen, bei denen der Befragte Konsequenzen aufgrund seiner gegebenen Antworten fürchtet. Beispielsweise bei Fragen nach kriminellem Verhalten wird der Befragte ungern oder nicht antworten. In der Interviewsituation wird ihn etwa nicht die Befürchtung, seine bislang unentdeckten Straftaten könnten aufgedeckt werden, zum Nicht-Antworten bewegen, sondern die Furcht vor missbilligenden Reaktionen des Interviewers.

Nicht nur Antwortvorgaben können sozial (un-)erwünscht sein, sondern auch Fragen. Insbesondere bei *tabuisierten Themen*, die in der Öffentlichkeit nicht diskutiert werden, kann es bereits sozial unerwünscht sein, auf diese Fragen Antworten zu geben, egal in welche Richtung die Antworten gehen. Die Zuschreibung sozial unerwünschter Antwortkategorien kann als Indiz für andere sozial erwünschte Merkmale gesehen werden: Das Zugeben von kleineren oder größeren Schwächen kann unter Umständen positiv bewerteten Eigenschaften wie Aufrichtigkeit und Ehrlichkeit entsprechen. Die Zuschreibung eines sozial unerwünschten Merkmals kann also auch einer erwünschten Antwort entsprechen (vgl. Hartmann 1991: 48f.). Personen, die zu sozial erwünschten Antworten tendieren sind meist Personen, die über geringes Selbstbewusstsein verfügen oder aber generell sozial benachteiligt sind (Esser 1991).

## 20.2 Entstehung von Einstellungen zu Befragungen

Vor allem die *Teilnahme an früheren Umfragen* prägt die inhaltliche Meinung bzw. Einstellung der Befragten zu Umfragen und führt zu einer Zunahme der Einstellungsstärke (Stocké/Langfeldt 2003). Nach Porst (1998) fördert ein *persönliches thematisches Interesse und Spaß* an der Beantwortung der Fragen eine positive Einstellung zu Umfragen, da die Befragten somit einen „Nutzen“ aus der Umfrage gezogen haben. Negativ auf das Umfrageimage wirkt *Skepsis bezüglich der Anonymität* (Esser 1973: 96). *Altruistische Gründe*, wie der Glaube an die Nützlichkeit von Umfragen, können eine positive Einstellung zu Umfragen bewirken (Esser 1973). Darüber hinaus nehmen auch in der *medialen Berichterstattung* Umfrageergebnisse, wie beispielsweise Wahlprognosen, eine wichtige Rolle ein, die Einstellungen zu Umfragen beeinflussen bzw. prägen.

Für den deutschsprachigen Raum gibt es bislang nur wenige Studien, die sich damit beschäftigen, wodurch die Umfrageeinstellung geprägt wird. Im Folgenden wird auf die Ergebnisse von Stocké und Langfeldt (2003) eingegangen. Negative Einstellungen zu Befragungen bilden Befragte demnach häufig aus, wenn sie *in der Vergangenheit an langen Interviews und/oder an Interviews mit schwierigen Erhebungsinstrumenten teilgenommen* haben. Die empfundene Belastung durch das Interview wird von den Befragten in diesen Fällen als hoch eingeschätzt. Der Auftraggeber, sei es eine wissenschaftliche Einrichtung oder ein kommerzielles Markt- und Meinungsforschungsinstitut, hat hingegen keinen Einfluss auf die Einstellung zu Umfragen. Überraschenderweise haben sogar Befragte, bei denen ein Interview in ein *Verkaufsgespräch* mündete, ebenfalls keine negativeren Einstellungen zu Umfragen als solche, die nicht auf diese Weise getäuscht wurden. Problematisch an der Aussagekraft dieser Ergebnisse ist, dass potenzielle Befragte, die eine vergangene Interviewerfahrung negativ werten, eventuell gar nicht erst bereit sind, an einer Studie zur Messung von Einstellungen zu Befragungen teilzunehmen (Stocké/Langfeldt 2003: 29ff.).

---

## 20.3 Auswirkungen der Einstellungen zu Befragungen

Schnauber und Daschmann (2008: 113ff.) kommen zum Ergebnis, dass Personen mit Umfrageerfahrungen häufiger teilnehmen, wenn die letzte Interviewerfahrung positiv war. Den größten Einfluss hat hier der *persönliche Nutzen*; so hat das Item „Umfragen bringen Abwechslung und sind interessant“ den größten Einfluss auf die Teilnahmebereitschaft. *Angst vor mangelnder Anonymität* und *altruistische Gründe* („Umfragen sind für die Gesellschaft wichtig und sinnvoll“) haben zwar auch einen signifikanten Einfluss, dieser ist aber deutlich kleiner. Das wichtigste Problem für die Beeinträchtigung der Datenqualität ist diesem Studienergebnis zufolge somit eine mit *negativen Interviewerfahrungen* und einer vermutlich hieraus resultierenden negativen Einstellung zu Umfragen einhergehende, niedrigere Teilnahmequote. Vor diesem Hintergrund kann die Gefahr einer Verzerrung der Antworten durch soziale Erwünschtheit und Item Nonresponse aufgrund negativer Umfrageeinstellungen eher als nachrangig angesehen werden, da diese Personen vermutlich seltener an Befragungen teilnehmen.

Mit der Zunahme umfragebasierter Markt- und Sozialforschung ist davon auszugehen, dass immer mehr Menschen bereits persönliche Erfahrungen mit Umfragen gesammelt haben und damit stärker verankerte Einstellungen zu Umfragen bilden. So hat sich der Umsatz der im Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (ADM) zusammengeschlossenen privatwirtschaftlichen Institute von 2002 bis 2012 fast verdoppelt (ADM 2012: 15). Die Beeinflussung von Befragungsergebnissen durch Einstellungen zu Umfragen wird zukünftig somit eher steigen. Gemäß dem Rational Choice-Modell würden Befragte aus einer positiven Umfrageerfahrung einen Nutzen ziehen und daher sowohl motiviert werden, an weiteren Befragungen teilzunehmen, als auch positive Haltungen gegenüber Umfragen entwickeln (Porst 1998: 3).

---

## 20.4 Messung der Einstellung zu Befragungen

Die Einstellung zur Befragung kann auch die Datenvalidität (Krebs/Menold, Kapitel 30 in diesem Band) stark beeinflussen und sollte in der Befragung berücksichtigt werden. Insbesondere sollte es im Interesse wissenschaftlicher und kommerzieller Institute liegen, durch adäquate und seriöse Methoden ein gutes Bild der Umfrageforschung zu pflegen. Sollen mittels Umfragen belastbare und valide Ergebnisse generiert und die Gefahr von entsprechenden Verzerrungen, insbesondere durch niedrige Teilnahmequoten, minimiert werden, muss daher die Einstellung der Befragten zukünftig noch mehr in den Fokus der Forschung gerückt werden. Die Einstellung eines Individuums zu Befragungen ist jedoch keineswegs einfach strukturiert sondern ein komplexes Konstrukt, das nach Stocké/Langfeldt (2003: 5ff.) mehrere Dimensionen umfasst.

- Zunächst stellen die Befragten eine klare *Kosten-Nutzen-Kalkulation* auf: Was bringen mir Befragungen und wie viel Aufwand z.B. Zeit bedeutet es?
- Darüber hinaus spielt jedoch auch die wahrgenommene Zuverlässigkeit und damit die Qualität des Messinstrumentes und der Ergebnisse eine Rolle.

Auf Basis der drei Dimensionen Kosten, Nutzen und Zuverlässigkeit hat Stocké (2012) eine *Skala zur Messung von Einstellungen zu Befragungen* entworfen, die aus 12 (bzw. in der Kurzversion aus neun) Einzelitems besteht (siehe Tab. 20.1). Die Einzelitems sind jeweils mit einer 7-stufigen, an den Endpunkten verankerten Ratingskala mit den Antwortoptionen von 1 = stimme voll und ganz zu bis 7 = lehne voll und ganz ab verstehen. (siehe Tab. 20.1).

Dieses Instrument erfasst die *generalisierte Einstellung der Befragten zu Umfragen* und damit auch eventuell hervorgerufene systematische Verzerrungen. Durch die Bildung eines additiven Indexes (Blasius, Burzan, Latcheva/Davidov und Huinink, Kapitel 83, 81, 55 und 82 in diesem Band) aus den Skalenitems kann die generalisierte Einstellung zu Umfragen beispielsweise im Rahmen einer Drittvariablenkontrolle (Kühnel/Dingelstedt, Kapitel 80 in diesem Band) eingesetzt werden. Problematisch beim Einsatz einer solchen Skala ist jedoch, dass lediglich das Ausmaß einer systematischen Verzerrung im Nachhinein

**Tab. 20.1** Einzelitems der Skala „Einstellungen zu Befragungen“ nach Stocké (2012)**Wahrgeommener Nutzen**

01.	Umfragen sind sehr wichtig für Wissenschaft, Politik und Wirtschaft.
02.	Meinungsumfragen machen die Gesellschaft demokratischer.
*03.	Die Teilnahme an Umfragen ist in meinem eigenen Interesse.
04.	Umfragen bringen Abwechslung und sind interessant.

**Wahrgeommene Zuverlässigkeit**

05.	Die Ergebnisse von Umfragen sind in den meisten Fällen richtig.
*06.	Bei Umfragen werden normale und repräsentative Menschen befragt.
07.	Die Teilnehmer an Umfragen sagen meistens ihre wahre Meinung.
08.	Die Teilnehmer an Umfragen geben sich sehr viel Mühe, richtig zu antworten.

**Wahrgeommene Belastung**

09.	Durch Umfragen wird man nur von wichtigeren Dingen abgehalten.
10.	In Deutschland werden viel zu viele Umfragen durchgeführt.
*11.	Es ist anstrengend in Umfragen so viele Fragen zu beantworten.
12.	Interviewer sind unangenehme und aufdringliche Leute.

\* = Item ist nur in der 12 Item Version enthalten

geschätzt werden kann. Sie ist kein Instrument, welches im Vorfeld einer Studie zur besseren Gewährleistung der Datenqualität eingesetzt wird.

Zu beachten ist darüber hinaus, dass die Skala die generalisierte, also *allgemeine Einstellung zu Befragungen* fokussiert. Nicht erfasst werden demnach Faktoren wie *Interesse am Befragungsthema* oder die *Interviewsituation*, die sich direkt auf die konkrete Befragungserfahrung beziehen und selbstverständlich ebenfalls das Antwortverhalten beeinflussen. So ist davon auszugehen, dass potenzielle Befragte für die Entwicklung einer Einstellung zu Umfragen die für sie spezifischste Information nutzen. Folglich werden Einstellungen zu Umfragen am stärksten durch Erfahrungen geprägt, die sie bereits mit Umfragen gemacht haben. Diese können beispielsweise Themen früherer Befragungen, Verhaltensweisen des Interviewers oder der Umgang mit den gegebenen Antworten und die Transparenz der Studienergebnisse sein. Diese Einzelerfahrungen tragen auch zur Bildung einer generalisierten Einstellung zu Umfragen bei. Zu bedenken ist ebenfalls, ob nicht auch der Einsatz einer solchen Skala bei den Befragten zu einer gefestigteren Einstellung zu Umfragen führt. So werden die Befragten durch die Skalenitems unter Umständen erst dazu angeregt über Befragungen nachzudenken und sich eine Meinung darüber zu bilden, bzw. sich der Einstellungen hinsichtlich Befragungen bewusst zu werden. Vor allem bei negativen Einstellungen könnte die Skala dadurch selbst zu einer Verschlechterung der Datenqualität führen, sofern sie am Beginn des Fragebogens positioniert wird.

## 20.5 Teilnahmeanreize (Incentives)

Zu den Instrumenten, die die Kosten-Nutzen-Kalkulation der Befragten ansprechen, um die Motivation zur Teilnahme an Befragungen zu erhöhen, gehören die *Incentives*: Durch den Einsatz von *Gutscheinen* oder *monetären Anreizen* lässt sich die Rücklaufquote einer Befragung unter Umständen um mehr als zehn Prozentpunkte erhöhen (Stadtmüller/Porst 2005: 2). Ein Vorteil bei dieser Methode ist, dass sie im Vorfeld eingesetzt werden kann.

Der Einsatz von Incentives kann jedoch zugleich auch die Qualität der Daten beeinflussen. Die als „*Sponsorship Bias*“ bezeichnete systematische Verzerrung der Daten kann dadurch entstehen, dass die Befragten Vermutungen darüber anstellen, welche Interessen der Auftraggeber mit den Fragen verfolgt, und eventuell ihre Antworten diesbezüglich ausrichten (Hyman 1954: 185ff.). Dabei sind nicht nur die tatsächlichen Interessen des Auftraggebers, sondern insbesondere auch die Vorstellungen der Befragten hinsichtlich der erwarteten Interessen des Auftraggebers relevant.

## Literatur

- ADM (2012): Jahresbericht 2012. URL: [https://www.adm-ev.de/index.php?eID=tx\\_nawsecuredl&u=0&file=fileadmin/user\\_upload/PDFS/ADM\\_Jahresbericht\\_2012.pdf&t=1378929851&hash=faa9f7179a1120ac2696afcf61c90d86d3f19078](https://www.adm-ev.de/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/user_upload/PDFS/ADM_Jahresbericht_2012.pdf&t=1378929851&hash=faa9f7179a1120ac2696afcf61c90d86d3f19078) am: 10.09.2013
- Esser, Hartmut (1973): Kooperation und Verweigerung beim Interview. In: Erbslöh, Eberhard/Esser, Hartmut/Reschka, Willibald/Schöne, Dino (Hg.): Studien zum Interview. Meisenheim am Glan. 71-141
- Esser, Hartmut (1991): Die Erklärung systematischer Fehler in Interviews. In: Wittenberg, Rainhard von (Hg.): Person–Situation–Institution–Kultur. Berlin: Duncker & Humblot. 59-78
- Goyer, John (1986): Surveys on Surveys. In: Public Opinion Quarterly 50: 27-41
- Hartmann, Petra (1991): Wunsch und Wirklichkeit. Wiesbaden: Dt. Univ.-Verl
- Hyman, Herbert H. (1954): Interviewing in Social Research. Chicago: University of Chicago Press
- Porst, Rolf (1998): Erfahrung mit und Bewertung von Umfragen. ZUMA-Arbeitsbericht 98/03
- Rogelberg, Steven G./Fisher, Gwenith G./Maynard, Douglas C./Hakel, Milton D./Horvath, Michael (2001): Attitudes Toward Surveys. In: Organizational Research Methods 4: 3-25
- Schnauber, A./Daschmann, G. (2008): States oder Traits? In: Zeitschrift für empirische Sozialforschung 2: 97-123
- Stadt Müller, Sven/Porst, Rolf (2005): Zum Einsatz von Incentives bei postalischen Befragungen. In: ZUMA How-to-Reihe 14. URL: [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/howto/how-to14rp.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/how-to14rp.pdf) am: 10.09.2013
- Stocké, Volker/Langfeldt, Bettina (2003): Umfrageeinstellung und Umfrageerfahrung. In: Sonderforschungsbereich 504. Rationalitätskonzepte, Entscheidungsverhalten und ökonomische Modellierung. No. 03-34. URL: [https://ub-madoc.bib.uni-mannheim.de/2750/1/dp03\\_34.pdf](https://ub-madoc.bib.uni-mannheim.de/2750/1/dp03_34.pdf) am 10.09.2013
- Stocké, Volker (2004): Entstehungsbedingungen und Antwortverzerrungen durch soziale Erwünschtheit. In: Zeitschrift für Soziologie 33 (4): 303-320
- Stocké, Volker (2012): Deutsche Kurzskala zur Erfassung des Bedürfnisses nach sozialer Anerkennung. In: Glöckner-Rist, A. (Hg.): ZUMA Informationssystem. Elektronisches Handbuch sozialwissenschaftlicher Erhebungsinstrumente. ZIS Version 15.0. Bonn: GESIS
- Sudman, Seymour/Bradburn, Norman M. (1983): Asking Questions. San Francisco

Alexander Glantz und Tobias Michael

---

## 21.1 Was sind Interviewereffekte?

Interviewer nehmen eine zentrale, aber oft unterschätzte Rolle in der Umfrageforschung ein (Reinecke, Kapitel 44 in diesem Band). Blickt man jedoch genauer hin, dann übernehmen sie eine Reihe sehr wichtiger Aufgaben bei persönlichen (Stocké, Kapitel 45 in diesem Band) und telefonischen Umfragen (Hüfken, Kapitel 46 in diesem Band), die nachhaltige Auswirkungen auf die Qualität der erhobenen Umfragedaten haben können: Interviewer stellen meistens den ersten Kontakt mit den Befragten her und überzeugen sie, an der Umfrage teilzunehmen (Engel/Schmidt, Kapitel 23 in diesem Band). Während des Interviews administrieren sie oftmals komplexe Fragebögen, klären gegebenenfalls die Bedeutung von Fragen und zeichnen die Antworten der Befragten auf.

In den letzten Jahrzehnten wurde sowohl in der akademischen als auch der kommerziellen Umfrageforschung untersucht, wie Interviewer die Kooperationsbereitschaft und das Antwortverhalten der Befragten beeinflussen können, da Interviewereffekte eine wichtige Fehlerquelle (Krebs/Menold, Kapitel 30 in diesem Band) für nicht selbst-administrierte Formen der Befragung darstellen. Unter *Interviewereffekten* versteht man allgemein systematische Unterschiede zwischen den Befragten, die durch die wahrgenommenen Eigenschaften oder das Verhalten des Interviewers entstehen (Groves 1989).

Im Rahmen der wissenschaftlichen Diskussion um Interviewereffekte wird zunehmend auf ein allgemeines statistisches Modell Bezug genommen, das als Ansatz des totalen Umfragefehlers (*Total Survey Error, TSE*) bekannt ist (Faulbaum, Kapitel 31 in diesem Band). Das Ziel von Umfragen besteht darin, bestimmte Parameter der Grundgesamtheit (Häder/Häder, Kapitel 18 in diesem Band) zu schätzen, z.B. einen Mittel- oder Anteilswert. Dabei wird davon ausgegangen, dass Interviewer zwei Aspekte dieser Schätzer beeinflussen können:

- Der *Interviewer-Bias* besagt, dass bestimmte Populationswerte durch die Anwesenheit oder das Verhalten des Interviewers systematisch über- oder unterschätzt werden

können. Dies geschieht beispielsweise wenn Befragte ihr Einkommen gegenüber Interviewern systematisch nach oben korrigieren.

- Die *Interviewer-Varianz* hingegen hat zur Folge, dass die Streuung der Schätzer (Standardfehler) erhöht wird. Dies kommt daher, dass die systematischen Verzerrungen zwischen Interviewern variieren können, z.B. wenn das Ausmaß der Verzerrung im selbstberichteten Einkommen vom wahrgenommenen sozioökonomischen Status des Interviewers abhängt. Anders ausgedrückt, werden unterschiedliche Interviewer andere Mittelwerte für das Einkommen der gleichen Befragten erhalten, *die lediglich auf dem Erscheinen der Interviewer basieren, nicht aber auf dem tatsächlichen Einkommen des Haushaltes*. Dadurch wird insgesamt betrachtet die Präzision der Einkommensschätzung in der Umfrage verringert.

## 21.2 Maße zur Erfassung von Interviewereffekten

### 21.2.1 Intra-Interviewer-Korrelationskoeffizient

Der statistische Effekt des Interviewers auf Populationsschätzer wird durch den von Kish (1962) vorgeschlagenen Intra-Interviewer-Korrelationskoeffizienten  $\rho_{int}$  beschrieben. Kish (1962) definiert den Interviewereffekt als den Anteil der totalen Varianz der Antworten, der durch die Unterschiede zwischen den Interviewern erklärt werden kann. Je größer der Intra-Interviewer-Korrelationskoeffizient ist, umso mehr ähneln sich die Antworten der Befragten, die von dem gleichen Interviewer befragt wurden. Daher lässt sich  $\rho_{int}$  auch als *Maß für die Homogenität der Antworten innerhalb der Interviewer* interpretieren. Eine von Groves (1989) durchgeführte und durch mehrere Studien replizierte Analyse über verschiedene persönliche und telefonische Umfragen kommt zu einer mittleren Intra-Interviewer-Korrelation von 0.01, wobei die Werte zwischen 0.002 und 0.102 schwanken.

### 21.2.2 Interviewer-Designeffekt

Wie stark die *Präzision der Populationsschätzer* dadurch beeinträchtigt wird, hängt von der *Antworthomogenität* und der *durchschnittlichen Interviewerauslastung* ab. Dieser Zusammenhang wird durch den Interviewer-Designeffekt  $deff_{int}$  abgebildet (Biemer/Lyberg 2003: 163):  $deff_{int} = 1 + (m-1)\rho_{int}$ , wobei  $m$  die Anzahl der Interviews pro Interviewer bezeichnet. Die Wurzel aus dem Designeffekt – der sogenannte Designfaktor  $deft_{int}$  – gibt an, um wie viel sich der Standardfehler gegenüber einfachen Zufallsstichproben vergrößert, wenn der Interviewereffekt berücksichtigt wird. Gehen wir beispielsweise von einer Intra-Interviewer Korrelation von .01 und einer durchschnittlichen Anzahl von 25 Interviews pro Interviewer aus, dann erhöht sich der Standardfehler gegenüber einer einfachen Zufallsstichprobe um den Faktor 1.11 (oder 11 Prozent):

$$deft_{int} = \sqrt{1+(25-1)\times 0.01} = 1.11.$$

Dieser Verlust an Präzision lässt sich gleichsetzen mit einer *Verringerung des Stichprobenumfangs*, wobei sich die effektive Stichprobengröße  $n_{\text{eff}}$  aus der ursprünglichen Stichprobe  $n$  dividiert durch den Designeffekt  $\text{deff}_{\text{int}}$  ergibt. Anders ausgedrückt verringert sich  $n$  durch den Interviewerfehler um den Faktor  $1/(1/\text{deff}_{\text{int}})$ . In unserem Beispiel geht eine scheinbar geringe Antworthomogenität von 0.01 mit einer Verringerung der Stichprobe um nahezu 20 Prozent einher. Im Umkehrschluss bedeutet das, dass wir eine deutlich größere Stichprobe benötigen, um dasselbe Maß an Präzision zu erreichen.

---

## 21.3 Bestimmungsfaktoren von Interviewereffekten

In der Forschung wurden inzwischen zahlreiche Studien zu den Ursachen von Interviewereffekten durchgeführt, wobei folgende Faktoren im Mittelpunkt des Interesses standen: der Modus der Befragung, Merkmale von Fragen und spezifische Eigenschaften von Interviewern.

### 21.3.1 Erhebungsmodus

Vergleichende Untersuchungen zu den Auswirkungen des Erhebungsmodus zeigen, dass persönliche Befragungen (Stocké, Kapitel 45 in diesem Band) stärkere Interviewereffekte hervorrufen als telefonische Umfragen (Hüfken, Kapitel 46 in diesem Band). Groves (1989) berichtet durchschnittliche Intra-Interviewer-Korrelationen bei persönlichen Befragungen von 0.005 bis 0.102 mit einem Gesamtmittelwert von 0.031. Dagegen scheint es bei telefonischen Befragungen aufgrund der besseren Kontrollmöglichkeiten deutlich geringeren Raum für Interviewereinflüsse zu geben. Der durchschnittliche  $\rho_{\text{int}}$  reicht hier von 0.002 bis 0.018 mit einem Gesamtmittelwert über alle betrachteten Umfragen von 0.009 (Groves 1989). Somit ist die Antworthomogenität bei telefonischen Befragungen im Durchschnitt erheblich niedriger als bei persönlichen Interviews. Dabei gilt es aber zu berücksichtigen, dass die Interviewerauslastung in Telefonstudios oft deutlich höher ist als bei Face-to-Face-Umfragen. Das bedeutet, dass bei der Auswahl des Erhebungsmodus und der Bestimmung von Interviewer-Designeffekten sowohl die unterschiedliche Antworthomogenität als auch die Anzahl der vermutlich zu realisierenden Interviews berücksichtigt werden muss.

### 21.3.2 Merkmale von Fragen

Andere Untersuchungen haben sich damit befasst, inwieweit Interviewereinflüsse mit bestimmten Merkmalen von Fragen (Reinecke, Kapitel 44 in diesem Band) variieren. Dabei liegt die Hypothese nahe, dass insbesondere Fragen nach sensiblen Einstellungsthemen

(Hlawatsch/Krickl, Kapitel 20 in diesem Band) oder Fragen, bei denen die Befragten in ihren eigenen Worten antworten durch Interviewer beeinflusst werden. Die Ergebnisse einschlägiger Untersuchungen zeigen allerdings kein eindeutiges Bild. Während einige Studien belegen, dass sowohl Faktenfragen als auch geschlossene Fragen (Porst, Kapitel 50 in diesem Band) weniger anfällig für Interviewereffekte sind (Schnell/Kreuter 2005), fanden andere Studien keine signifikanten Unterschiede zwischen den Fragetypen (Groves 1989). Verschiedene Studien stützen hingegen die Hypothese, dass schwierige Fragen größere Interviewereffekte aufweisen, wobei der Schwierigkeitsgrad einer Frage davon abhängt, ob die Befragten größere Erinnerungsleistungen vollbringen oder Themen beurteilen müssen, über die sie nur schwach verankerte Einstellungen besitzen (Schaeffer et al. 2010).

### 21.3.3 Eigenschaften der Interviewer

Der Großteil der Forschung hat sich jedoch mit der Frage auseinandergesetzt, ob Interviewereffekte mit bestimmten Eigenschaften der Interviewer zusammenhängen. Nach Schaeffer et al. (2010) lassen sich diese Eigenschaften danach unterscheiden, ob sie für den Befragten sichtbar sind und ob sie eher persönlicher Natur sind oder zum Rollenverhalten des Interviewers gehören.

*Sichtbare Eigenschaften* werden von den Befragten direkt erfasst und können unmittelbare Auswirkungen auf ihr Verhalten haben. Dazu gehören persönliche Merkmale wie z.B. das Geschlecht und die Ethnie des Interviewers, aber auch Eigenschaften, die sich auf die Rolle des Interviewers beziehen. Das beinhaltet beispielsweise *Auftreten, Körpersprache* und das *verbale Verhalten des Interviewers*.

*Nicht-sichtbare Eigenschaften* treten meist erst im Laufe der Interaktion mit dem Interviewer zutage und beeinflussen die Befragten indirekt, indem sie das Verhalten des Interviewers prägen. Dazu zählen *stabile Persönlichkeitsmerkmale* wie etwa die Extrovertiertheit eines Interviewers, aber auch *rollenbezogene Eigenschaften*, wie die Einstellungen des Interviewers gegenüber dem Thema der Befragung und den Befragten.

Untersuchungen aus dem US-amerikanischen Kontext zeigen, dass die *ethnische Herkunft* zwar kaum Auswirkungen auf die Teilnahmebereitschaft, aber substantielle Effekte auf das Antwortverhalten der Befragten hat (Weisberg 2005, Schaeffer et al. 2010, El-Menouar, Kapitel 60 in diesem Band). Ähnliche Befunde bieten Studien zu *Geschlechtereffekten* in Interviews (Huddy et al. 1997). Dies legt die Vermutung nahe, dass Interviewer, die die gleiche ethnische Herkunft oder das gleiche Geschlecht wie die Befragten aufweisen, validere Antworten erhalten. Diese sogenannte *Matching-Hypothese* wird durch Ergebnisse aus mehreren Studien gestützt (Schaeffer et al. 2010). Allerdings kann ein Matching auch dazu führen, dass sich die Befragten veranlasst fühlen, ihre Antworten an die wahrgenommenen sozialen Erwartungen der Interviewer anzupassen. So demonstrieren Anderson et al. (1988), dass afroamerikanische Befragte ihre Beteiligung an US-Präsidentenwahlwahlen übertreiben, wenn sie von schwarzen Interviewern befragt werden.

Weiterhin gibt es Belege dafür, dass ältere Interviewer höhere Kooperationsraten erzielen und das Antwortverhalten positiv beeinflussen (Hox/de Leeuw 2002).

Obwohl es nur sehr wenige Untersuchungen zum *Stimmausdruck und nonverbalen Auftreten von Interviewern* gibt, existieren Hinweise dafür, dass subjektiv eingeschätzte Stimmerkmale wie die Feminität der Stimme sowie Stimmhöhe, Intonation und Sprachfluss positiv mit der Teilnahmebereitschaft zusammenhängen (Groves et al. 2007). Anderseits gibt es jedoch auch Studien, in denen keine Effekte der subjektiv wahrgenommenen Stimmerkmale auf die Erfolgsquote nachgewiesen werden konnten oder dies nur für die objektiv gemessene Stimmhöhe und Sprechgeschwindigkeit galt (Schaeffer et al. 2010).

Als nachgewiesen gilt, dass neben der Selbstsicherheit vor allem die *Erfahrung der Interviewer* eindeutig positive Zusammenhänge zur Teilnahmebereitschaft aufweist (Groves/Couper 1998, Hox/de Leeuw 2002). Ebenso sind *extrovertierte Interviewer* und *Interviewer, die davon überzeugt sind, Befragte zur Teilnahme bewegen zu können*, erfolgreicher (Jäckle et al. 2013, Weisberg 2005), wobei die Effekte dieser Persönlichkeitsmerkmale relativ schwach ausfallen.

Auf der anderen Seite gibt es kaum Untersuchungen dazu, inwiefern sich nicht-sichtbare Interviewereigenschaften auf die Antwortqualität auswirken. Die wenigen vorhandenen Studien legen nahe, dass *erfahrene Interviewer* mehr Antwortverweigerungen auf Fragen nach dem Einkommen und stärke Zustimmungstendenzen auf Einstellungsfragen (Olson/Bilgen 2011) hervorrufen. Eine mögliche Erklärung hierfür wäre, dass erfahrene Interviewer ihre Interviews schneller durchführen und die Produktivität auf Kosten der Interviewqualität erhöhen (Schaeffer et al. 2010).

Ein weiterer Faktor neben der Erfahrung sind die *Erwartungen der Interviewer*. Erwarteten Interviewer beispielsweise negative Reaktionen auf unangenehme Fragen, erzielen sie tatsächlich eine höhere Antwortverweigerung (Schaeffer et al. 2010).

Insgesamt legt der Forschungsstand nahe, dass die Ursachen von Interviewereffekten in der dynamischen Interaktion von Interviewern, Befragten und Befragungsthema liegen. So beeinflussen beispielsweise sichtbare Interviewereigenschaften am ehesten das Antwortverhalten, wenn die Gesprächssituation dieses Merkmal für den Befragten salient macht (Schaeffer et al. 2010). Neben methodischen Problemen lassen sich die teilweise uneinheitlichen Befunde auch damit erklären, dass in der Mehrzahl der betrachteten Studien jeweils einzelne Aspekte der Interviewsituation im Mittelpunkt standen, ohne deren Zusammenspiel zu berücksichtigen. Darüber hinaus sind aufgabenbezogene Merkmale der Interviewer, wie Erfahrung und professionelle Arbeitseinstellungen, wichtiger als persönliche Eigenschaften. Damit ist die Schlussfolgerung verbunden, dass es nicht den „typischen“ Interviewer mit einem festgelegten soziodemographischen Profil und Persönlichkeitssbild gibt, sondern Eigenschaften ausschlaggebend sind, die sich potentielle Interviewer aneignen können und die sich durch sorgfältiges Interviewertraining modifizieren lassen.

## 21.4 Minimierung von Interviewereffekten

Die meisten Forscher sind sich darin einig, dass die Schulung von Interviewern bedeutsamer für die Minimierung von Interviewereffekten ist als die Rekrutierung nach bestimmten Merkmalen, da persönliche Merkmale von Interviewern entweder keinen oder nur unter bestimmten Bedingungen einen Effekt auf das Befragtenverhalten haben (Weisberg 2005). Die *Interviewerschulung* sollte aus einer Basisschulung und jeweils speziellen Schulungen bestehen, sowie jeweils, die auf konkrete Aspekte einer Studie abgestimmt sind.

### 21.4.1 Basisschulung

Eine Basisschulung sollte neben einigen *Grundkenntnissen der Umfrageforschung* und dem *Umgang mit technischem Gerät* (z.B. Computer und Software für computerassistierte Befragungen) die *Kontaktaufnahme mit der Zielperson*, die *Behandlung von Einwänden* sowie *Grundlagen der standardisierten Gesprächsführung* beinhalten. Besondere Aufmerksamkeit sollte bei der Interviewerschulung dem Einstieg und der Behandlung von Einwänden gewidmet werden. In mehreren Studien wurde untersucht, wie sich der Inhalt und die Länge standardisierter Einleitungen auf die Kooperationsbereitschaft auswirken. Generell zeigt sich, dass Interviewer, die mehr Freiheit bei den *Einleitungen* haben und einige Kernpunkte als Leitfaden verwenden, höhere Ausschöpfungen erzielen, während Interviewer, die sich sehr stark an geskriptete Vorlagen halten, weniger erfolgreich sind (Groves et al. 2007; Morton-Williams 1993).

Die Ergebnisse verschiedener qualitativer Studien über das Verhalten von Interviewern und Aufzeichnungen realer Interaktionssituationen demonstrieren, dass Interviewer umso höhere Kooperationsraten erreichen, je besser sie in der Lage sind, sich auf ihr Gegenüber einzustellen (Couper/Groves 2002, Morton-Williams 1993). Das bedeutet, dass sie sehr schnell auf verbale und nonverbale Hinweise reagieren und ihre Überzeugungsversuche an die Bedenken des jeweiligen Befragten anpassen müssen. Dabei versuchen sie die Befragten in ein Gespräch zu verwickeln und die Interaktion aufrechtzuerhalten, um weitere Informationen zu erhalten und maßgeschneiderte Überzeugungstechniken einzusetzen. Die Erkenntnisse implizieren, dass standardisierte *Techniken der Einwandbehandlung* kaum zum Erfolg führen. Zwar muss ein Interviewer wesentliche Argumente kennen und überzeugend vermitteln, aber er sollte auch erkennen, welche Zielpersonen responsiv auf welche Argumente reagieren und eine Interaktion am Laufen halten, um mehr über die Beweggründe des Befragten zu erfahren. Diese Techniken stellen auch besondere Anforderungen an Interviewer, da sie schnell die Bedenken potentieller Befragter analysieren und diesen in ihren eigenen Worten begegnen müssen (Maynard/Schaeffer 2002).

Ein weiterer Kernpunkt der Basisschulung ist es, den Interviewern die *Ziele und Voraussetzungen standardisierter Gesprächsführung* zu vermitteln (Fowler/Mangione 1990). Das bedeutet, dass der Fragewortlaut, die Reihenfolge der Fragen und die Art und Weise, in der die Frage gestellt wird, genau festgelegt sind, um unabhängig vom Interviewer einen

identischen Fragestimulus zu generieren. Darüber hinaus sollten Antworten exakt protokolliert und nicht durch den Interviewer beurteilt werden. Insgesamt gilt es eine freundliche, aber aufgabenorientierte Gesprächssituation zu schaffen. Innerhalb der Forschung besteht ein breiter Konsens darüber, dass die standardisierte Gesprächsführung den Schlüssel zur Reduktion von Interviewereffekten darstellt; einen Überblick über die Kritik am Ansatz der standardisierten Interviewführung geben Maynard/Schaeffer (2006).

### 21.4.2 Studienspezifische Briefings

Das generelle Training wird ergänzt durch studienspezifische Briefings, in denen sichergestellt wird, dass Interviewer die Zielsetzung und den Hintergrund einer Studie verstehen und beschreiben können. Weiterhin sollte die Zielpersonenauswahl besprochen, der Fragebogen durchgearbeitet, auf besondere Fragen eingegangen und die Filterführung eingeübt werden. Diese Maßnahmen zielen zum einen darauf ab, die Motivation der Interviewer durch Hintergrundwissen über den wissenschaftlichen Zweck einer Studie zu steigern. Zum anderen helfen sie den Interviewern die Fragen der Respondenten zu beantworten und sie zur Teilnahme zu bewegen bzw. deren Kooperation im Interviewverlauf sicherzustellen.

### 21.4.3 Kontrolle der Interviewerperformanz

Ein weiterer wichtiger Aspekt, um die Qualität der Interviews zu gewährleisten, ist die Kontrolle der Interviewerperformanz (Biemer/Lyberg 2003). Interviewer sollten regelmäßig in folgenden Bereichen beobachtet und bewertet werden: Gestaltung des Intervieweinstiegs, Einwandbehandlung, Einhaltung der Regeln standardisierter Gesprächsführung, Umgang mit offenen Fragen und Weiß-nicht-Antworten, Exaktheit der Dateneingabe und administrative Aufgaben, wie das Ausfüllen von Kontaktprotokollen.

Weiterhin sollte insbesondere bei persönlichen Befragungen kontrolliert werden, ob Interviews tatsächlich stattgefunden haben, um eventuelle *Fälschungen* aufzudecken (Blasius, Kapitel 22 in diesem Band). Dies geschieht in der Regel durch telefonische oder postalische Nachkontakte. Einen Überblick über weitere Techniken der Interviewerkontrolle bieten Biemer/Lyberg (2003).

---

## 21.5 Herausforderungen für die Zukunft

Interviewer stellen einen zentralen Aspekt der Datenqualität dar. Wie die in diesem Kapitel besprochenen Untersuchungen verdeutlichen, können Interviewer unter bestimmten Bedingungen substantielle Effekte auf die Teilnahme und das Antwortverhalten der Befragten ausüben. Viele Studien belegen inzwischen, dass Design-Effekte aufgrund von

Interviewern mindestens ebenso bedeutsam sind wie der Effekt durch geographische Klumpung (Schnell/Kreuter 2005). Wie Weisberg (2005) betont, sind Interviewereffekte jedoch einer der wenigen Aspekte des totalen Umfragefehlers, der unter direkter Kontrolle der Forscher steht. Durch sorgfältiges Training und Monitoring können Interviewereinflüsse auf ein Minimum reduziert werden.

Angesichts ihrer Schlüsselfunktion für die Umfrageforschung ist es nicht verwunderlich, dass das Interesse an Interviewereffekten in den letzten Jahren stark zugenommen hat. Allerdings ist die Forschung in diesem Bereich mehr durch Daten als durch allgemeine Theorien geleitet. Die zukünftige Forschung sollte ihre Untersuchungen daher in eine Theorie des Interviewer- und Befragtenverhaltens einbetten, da erst eine solche Vorgehensweise einen systematischen Erkenntnisgewinn ermöglicht und das Potential hat, disparate Befunde zu erklären.

Zum Abschluss möchten wir noch auf einige methodische Herausforderungen der Forschung eingehen. Wie wir gesehen haben, entstehen Interviewereinflüsse aus dem komplexen Zusammenwirken von Befragten, Befragungsinstrument und Merkmalen der Interviewer. Auf der Suche nach den Ursachen von Interviewereffekten sollte sich die Forschung daher mehr auf die *Analyse von dynamischen Gesprächssituationen* konzentrieren, als auf die statische Betrachtung einzelner Aspekte. Dazu bedarf es auch neuer methodischer Ansätze wie der Konversations- (Maynard/Schaeffer 2006) oder Videoanalyse (Tuma/Schnettler, Kapitel 68 in diesem Band). Ein weiteres methodisches Problem für die empirische Prüfung von Interviewereffekten in persönlichen Befragungen ist die *Schwierigkeit, die Effekte von Interviewern und geographischen Stichprobeneinheiten zu trennen*. In der Praxis werden Interviewer den geographischen Stichprobeneinheiten (*sampling points*) (Häder/Häder, Kapitel 18 in diesem Band) meist aufgrund der räumlichen Nähe zugewiesen, um Feldzeit und Reisekosten zu minimieren. Um jedoch etwas über mögliche kausale Interviewereffekte zu erfahren sind sogenannte interpenetrierte Designs notwendig, in denen die Interviewer den Einheiten zufällig zugewiesen werden (Biemer/Lyberg 2003). Die beobachteten Effekte lassen sich somit eindeutig auf Unterschiede in der Wahrnehmung oder auf das Verhalten der Interviewer zurückführen.

Während es in den letzten Jahren ein steigendes Bewusstsein für stichprobenbedingte Designeffekte gibt, werden Interviewereffekte in der statistischen Datenanalyse weitgehend ignoriert. Ein weiteres methodisches Problem besteht darin, dass die meisten Datensätze nicht die für eine statistische Analyse von Interviewereffekten notwendigen Variablen enthalten. So fehlt beispielsweise meistens eine Identifikationsnummer des Interviewers. Die Aufnahme der Intervieweridentifikation und Interviewereigenschaften (Felderer/Birg/Kreuter, Kapitel 25 in diesem Band) sollten daher in Zukunft zur Standardausstattung eines Datensatzes gehören, um weitere Erkenntnisse über die sozialen Prozesse zu gewinnen, die Interviewereffekten zugrunde liegen.

## Literatur

- Anderson, Barbara A./Silver, Brian D./Abramson, Paul R. (1988): The Effects of the Race of the Interviewer on Race-Related Attitudes of Black Respondents in SRC/CPS National Election Studies. In: *Public Opinion Quarterly* 52: 289-324
- Biemer, Paul P./Lyberg, Lars E. (2003): *Introduction to Survey Quality*. Hoboken (NJ): Wiley
- Couper, Mick P./Groves, Robert M. (2002): Introductory Interactions in Telephone Surveys and Nonresponse. In: Maynard, Douglas W./Houtkoop-Steenstra Hanneke/Schaeffer, Nora C./van der Zouwen, Johannes (Hg.): *Standardization and Tacit Knowledge*. New York: Wiley. 161-177
- Fowler, Floyd J./Mangione, Thomas W. (1990): *Standardized Survey Interviewing*. Newbury Park (CA): Sage
- Groves, Robert M. (1989): *Survey Errors and Survey Costs*. New York: Wiley
- Groves, Robert M./Couper, Mick P. (1998): Nonresponse in Household Interview Surveys. New York: Wiley
- Groves, Robert M., O'Hare, Barbara C., Gould-Smith, Dottye, Benkí, José/Maher, Patty (2007): Telephone Interviewer Voice Characteristics and the Survey Participation Decision. In: Lepkowski, James M./Tucker, Clyde/Brick, J. Michael/de Leeuw, Edith/Japec, Lilli/Lavrakas, Paul J./Link, Michael W./Sangster, Rroberta L. (Hg.): *Advances in Telephone Survey Methodology*. Hoboken (NJ): Wiley. 385-400
- Hox, Joop J./de Leeuw, Edith D. (2002). The Influence of Interviewers' Attitude and Behavior on Household Survey Nonresponse. In: Groves, Robert M./Dillman, Don A./Eltinge, John L./Little, Roderick J. A. (Hg.): *Survey Nonresponse*. New York: Wiley. 103-120
- Huddy, Leonie/Billig, Joshua/Bracciodieta, John/Hoeffler, Lois/Moynihan, Patrick J./Puglianì, Patricia (1997): The Effect of Interviewer Gender on the Survey Response. In: *Political Behavior* 19: 197-220
- Jäckle, Annette/Lynn, Peter/Sinibaldi, Jennifer/Tipping, Sarah (2013): The Effect of Interviewer Experience, Attitudes, Personality and Skills on Respondent Co-operation with Face-to-Face Surveys. In: *Survey Research Methods* 7: 1-15
- Kish, Leslie (1962): Studies of Interviewer Variance for Attitudinal Variables. In: *Journal of the American Statistical Association* 57: 92-115
- Maynard, Douglas W./Schaeffer, Nora C. (2002): Refusal Conversion and Tailoring. In: Maynard, Douglas W./Houtkoop-Steenstra Hanneke/Schaeffer, Nora C./van der Zouwen, Johannes (Hg.): *Standardization and Tacit Knowledge*. New York: Wiley. 219-240
- Maynard, Douglas W./Schaeffer, Nora C. (2006): Standardization-in-Interaction. In: Drew, Paul/Raymond, Geoffrey/Weinberg, Darin (Hg.): *Talk and Interaction in Social Research Methods*. London: Sage. 9-27
- Morton-Williams, Jean (1993): *Interviewer Approaches*. Aldershot: Dartmouth Publishing

- Olson, Kristen/Bilgen, Ipek (2011): The Role Of Interviewer Experience on Acquiescence.  
In: *Public Opinion Quarterly* 75: 99-114
- Schaeffer, Nora C./Dykema, Jennifer/Maynard, Douglas W. (2010): Interviewers and Interviewing. In: Marsden, Peter V./Wright, James D. (Hg.): *Handbook of Survey Research*. Bingley: Emerald Group. 437-470
- Schnell, Rainer/Kreuter, Frauke (2005): Separating Interviewer and Sampling-Point Effects. In: *Journal of Official Statistics* 21: 389-410
- Weisberg, Herbert F. (2005): *The Total Survey Error Approach*. Chicago: University of Chicago Press

Jörg Blasius

---

## 22.1 Problemstellung

Es gibt viele Einführungen in die empirische Sozialforschung und eine Vielzahl von sozi-alwissenschaftlichen Fachzeitschriften mit Hunderten von Artikeln, in denen die Bedeutung von Interviewern in Face-to-Face-Situationen beschrieben wird (Stocké, Hlawatsch/Krickl, Glantz/Michael, Kapitel 45, 20 und 21 in diesem Band), aber es gibt nur wenige Studien, in denen das Problem von gefälschten Interviews diskutiert wird. Nach Empfehlung der American Association of Public Opinion Research (AAPOR 2003: 7) sollten 5 bis 15 Prozent der Interviews von den Markt- und Meinungsforschungsinstituten (Ziegler, Kapitel 10 in diesem Band) kontrolliert werden, um die Qualität der erhobenen Daten zu sichern und um zu gewährleisten, dass die Daten nicht gefälscht sind. Diese Art der Kontrolle scheint auch die Praxis zu sein, zumindest wenn man den Angaben der Datenerhebungsinstitute folgt. Die Frage ist, wie gründlich kann kontrolliert werden, und wie gründlich wird kontrolliert? Dabei ist es vermutlich sehr einfach herauszufinden, wenn ein Interviewer einen Haushalt oder eine Person überhaupt nicht kontaktiert hat, aber wie ist es, wenn nur Teile des Interviews gefälscht wurden?

Angesichts der geringen Anzahl von Publikationen, in denen das Problem „Fälschungen“ diskutiert wird, und in Anbetracht der Maßnahmen, die AAPOR und andere Institute für öffentliche Meinungsforschung zur Verhinderung von Fälschungen empfehlen (in Deutschland ist dies der ADM, der Arbeitskreis deutscher Marktforschungsinstitute, [www.adm-ev.de](http://www.adm-ev.de)), könnte man zu dem Schluss gelangen, dass dies kein zentrales Problem in der empirischen Forschung ist. Dem ist jedoch nicht so, wie einige aktuelle Studien zeigen (Blasius/Thiessen 2012, 2013, Blasius/Friedrichs 2012). Um dieses Problem zu diskutieren, beschränke ich mich im Folgenden auf Fälschungen durch die Interviewer, auch wenn es ebenfalls vorkommt, dass Mitarbeiter von Instituten ganze Interviews einfach kopieren, so dass diese zwei- und z.T. sogar mehrfach im Datensatz vorkommen (Blasius/Thiessen 2012, 2014).

## 22.2 Warum Interviewer fälschen

Die Diskussion über Fälschungen und Fälscher begann bereits Mitte der 1940er Jahre und wurde unter der Überschrift „Interviewermotivation“ geführt, und wie diese erhöht werden könnte. Dabei wurden Gründe gesucht, warum Interviewer Daten fälschen, die dann überwiegend bei der Art der Befragung und den zu stellenden Fragen gefunden wurden. Crespi (1945) war einer der ersten, der sich über das Fälschen von Interviews Gedanken gemacht hat. Er sieht das Fälschungsproblem als ein moralisches (Crespi 1945: 431), wobei er davon ausgeht, dass die Ursachen zum einen in den z.T. unrealistischen Bedingungen der Institute liegen und zum anderen in der persönlichen Integrität der Interviewer. So können eine zu lange Dauer des Interviews, zu viele Nachfragen (wieso, weshalb, warum, ...?), eine scheinbare Wiederholung von Fragen, langwierige Fragestellungen, zu komplexe und zu schwierige Fragen den Interviewer demoralisieren. Ein weiterer möglicher Grund ist der Druck durch den Supervisor oder einen noch höheren Vorgesetzten auf die Interviewer, z.B. in einem bestimmten Zeitraum eine bestimmte Anzahl von Interviews durchzuführen (Nelson/Kiecker 1996: 1110). Harrison/Krauss (2002: 325) sprechen davon, dass es vielleicht auch psychologische Belohnungen wie ein „Gefühl der Klugheit“ und „Aufregung bei einem riskanten Verhalten“ sind, die einen Interviewer zum Fälschen verleiten.

Obwohl es schon seit Mitte der 1940er-Jahre eine Diskussion über die Gründe gibt, warum Interviewer fälschen, der elementarste Grund wurde bislang kaum erwähnt: Fälscher sparen viel Zeit bzw. erhöhen ihren Stundenlohn deutlich, wenn sie (große) Teile ihrer Interviews am heimischen Schreibtisch ausfüllen bzw. wenn sie schon am Ort des Interviews mit einer oder wenigen Fragen gleich mehrere „beantworten“ – dies gilt zumindest wenn, wie üblicherweise praktiziert, die Bezahlung pro durchgeführtes Interview erfolgt (Blasius/Friedrichs 2012). Das Risiko bei derartigen Teilstückfälschungen entdeckt zu werden, ist relativ gering, da effektive Kontrollen sehr zeit- und damit auch kostenaufwändig sind. Des Weiteren sind die möglichen Folgen bei „Enttarnung“ vermutlich zu vernachlässigen, insbesondere dann, wenn die Anweisungen aus dem Institut unrealistisch sind (Nelson/Kiecker 1996). Zudem dürfte es im Interesse der betroffenen Institute liegen zu verhindern, dass die Öffentlichkeit erfährt, dass ihre Mitarbeiter (Teile von) Interviews gefälscht haben (Harrison/Kraus 2002).

---

## 22.3 Ausmaß der Fälschungen

In einem Projekt des US Bureau of Census zur Erkennung von Fälschungen, die zwischen September 1982 und August 1987 stattfanden, berichten Biemer/Stokes (1989: 25), dass zwischen drei und fünf Prozent aller befragten Interviewer angaben, in irgendeiner Form schon einmal Interviews gefälscht zu haben. Schäfer et al. (2005: 10) berichteten von bis zu 2,4 Prozent an Fälschungen, die sie bei der ersten Welle des deutschen Sozio-ökonomischen Panels (SOEP) entdeckt haben. Nelson/Kiecker (1996) befragten 173 Telefoninterviewer danach, wie genau sie die Anweisungen des Instituts befolgt hätten. Obwohl sie im

Telefonlabor überwacht wurden, berichteten 10 Interviewer, dass sie zumindest manchmal ein vollständiges Interview gefälscht hätten, weitere 22 gaben an, dass sie zumindest manchmal die eine oder andere fehlende Antwort ersetzen. Harrison/Kraus (2002: 319) kamen aufgrund ihrer Beobachtungen zu dem Schluss, dass Fälschungen durch die Interviewer wesentlich weiter verbreitet sind als es allgemein angenommen wird.

Blasius/Thiessen (2012, 2013) vermuten, dass vollständige Fälschungen nur selten vorkommen, da dies von den Instituten, welche die Daten erheben, zu einfach per Kontrollanruf nachgewiesen werden kann. Vielmehr nehmen sie an, dass sämtliche zentralen und leicht nachprüfbarer Angaben wie das Alter, das Geschlecht und die Bildung korrekt erhoben werden, dass aber bei langen Itembatterien (Porst, Franzen, Kapitel 50 und 51 in diesem Band) geschummelt wird. Da werden von vielleicht zehn Fragen zum Vertrauen in Institutionen (Bundesverfassungsgericht, politische Parteien, Polizei, ...) zwei oder drei gestellt, die verbleibenden Antworten werden einfach „entsprechend“ eingetragen. Diese Vorgehensweise resultiert in Interviewer-typischen Antwortstrukturen, die statistisch nachgewiesen werden können (Blasius/Thiessen 2012, 2013).

---

## 22.4 Wie können Interviews gefälscht werden?

Es gibt verschiedene Arten von gefälschten Interviews. In einigen Fällen ist es sogar fraglich, ob ein bestimmtes Interview überhaupt als „gefälscht“ bezeichnet werden sollte. AAPOR (2003: 2) zufolge ist ein Interview bereits gefälscht, wenn eine der folgenden Aussagen zutrifft: „a) fabricating all or part of an interview – the recording of data that are not provided by a designated survey respondent and reporting them as answers of that respondent; b) deliberately misreporting disposition codes and falsifying process data (e.g., the recording of a refusal case as ineligible for the sample; reporting a fictitious contact attempt); c) deliberately miscoding the answer to a question in order to avoid follow-up questions; d) deliberately interviewing a non-sampled person in order to reduce effort required to complete an interview; or e) otherwise, intentionally misrepresenting the data collection process to the survey management.“

Dieser Definition zur Folge müsste ein Interview bereits als gefälscht eingestuft werden, wenn ein Interviewer bei einer „Random-Walk-Anweisung“ (Häder/Häder, Kapitel 18 in diesem Band) von der „richtigen Route“ abgewichen ist bzw. wenn er die Anweisung bei der Auswahl der Zielperson (z.B. die Person zu befragen, die als nächstes Geburtstag hat) nicht eingehalten hat. Doch wie soll ein Institut eine derartige Vorgabe (möglichst kostengünstig) prüfen? Beim Random-Walk könnte zwar ein zweiter Interviewer den Weg nachgehen, aber wie hat der Interviewer den „fünften“ Haushalt „abgezählt“ – und wenn er sich „verzählt“ hat, sollte da Absicht unterstellt werden? Und zum Beginn des Interviews als erstes zu fragen: „wer hat als nächstes Geburtstag?“, um die richtige Zielperson zu ermitteln, dürfte kaum jemanden zur Teilnahme motivieren, dafür ist diese Frage zum einen viel zu sensibel und zum anderen ist es sehr leicht zu sagen, dass eine andere Person „als nächstes Geburtstag“ hat, die „bedauerlicherweise“ gerade nicht erreichbar ist. Das ist

weder im Interesse des Interviewers, der viel Zeit verliert, um den Kontakt zur richtigen Zielperson herzustellen (die dann vielleicht auch noch ihre Teilnahme verweigert), noch des Erhebungsinstituts, welches in möglichst kurzer Zeit die vertraglich zu erbringenden Interviews bei einer möglichst hohen Ausschöpfungsquote (Engel/Schmidt, Kapitel 23 in diesem Band) erbringen möchte.

Wird davon ausgegangen, dass ein Interviewer sich über mögliche Konsequenzen beim Aufdecken des Betruges keine Gedanken gemacht hat bzw. dass er diese in Kauf nimmt, vielleicht unter der wohlwissenden Annahme, dass es zu keiner Anzeige kommt, dann begünstigt dies eine mögliche Absicht zum Fälschen. AAPOR (2003: 7) schlägt Fragen zu folgenden Themenbereichen vor, um zu prüfen, ob ein Interview stattgefunden hat – und diese Art der Kontrolle kann dem Interviewer durchaus bekannt sein, z.B. aus dem Internet:

1. die Haushaltzzusammensetzung,
2. die Art der Datenerhebung,
3. Länge des Interviews,
4. die Weitergabe von „Vergütungen“ an die Befragungsperson (sofern es welche gab),
5. die Nutzung eines Computers während der Datenerhebung,
6. ob zentrale Themen diskutiert wurden,
7. ob bestimmte Fragebatterien den Befragten vorgelegt wurden, insbesondere die zeitintensiven.

Wird ein Prüfverfahren vorausgesetzt, welches diesen Empfehlungen folgt, können folgende „Ratschläge für den erfolgreichen Fälscher“ gegeben werden (angelehnt an Blasius/Friedrichs 2012):

- Notieren Sie die Adresse der Interviewten genau so wie vom Institut gefordert. Stellen Sie dabei sicher, dass der/die Befragte sich bewusst ist, dass das Forschungsinstitut das Interview telefonisch oder postalisch überprüfen könnte. Dabei ist es nicht wichtig, wie viel Zeit das Interview tatsächlich in Anspruch nahm (Befragte können sich nicht erinnern, im Zweifel verging die „Zeit anscheinend wie im Flug“), es ist nur wichtig, dass Sie in seiner/ihrer Wohnung waren und Fragen gestellt haben.
- Reduzieren Sie den Fragebogen auf die zentralen Fragen, insbesondere Alter, Geschlecht und Bildung sind wichtig, da sie leicht zu überprüfen sind. Zentrale Themenblöcke sollten Sie erwähnen, aber es müssen längst nicht alle Fragen gestellt werden.
- Bereiten Sie eine Liste von möglichen Befragten vor, die Sie wieder verwenden können (Freunde, Freunde von Freunden, ...), insbesondere in Quotenstichproben (Häder/Häder, Kapitel 18 in diesem Band) können Sie diese immer wieder verwenden.
- Achten Sie auf logische und plausible Antworten, wenn Sie den Fragebogen ausfüllen, das ist per Computer leicht zu überprüfen.
- Arbeiten Sie bei mehreren Instituten gleichzeitig, so können Sie u.U. die gleichen Adressen zweimal oder öfter verwenden (kaum jemand wird sich an den Namen des Instituts erinnern).

- Arbeiten Sie mit anderen „erfahrenen“ Interviewern zusammen; beachten Sie dabei aber, dass Sie in etwa gleichen Alters und möglichst auch gleichen Geschlechts sind; Adressen können ausgetauscht und doppelt verwendet werden.
- Beachten Sie, dass die Kunden u.U. bestimmte Ergebnisse von dem Institut erwarten, achten Sie darauf, dass die entsprechenden Erwartungen erfüllt werden.
- Achten Sie auf Filterfragen, füllen Sie die Fragen auch dann regelmäßig aus, wenn diese auch „übersprungen“ werden können (zu wenig ausgefüllte Fragenblöcke sind verdächtig und leicht zu überprüfen); achten Sie sich auch grundsätzlich auf eine „durchschnittliche Stichprobe“, wenn z.B. 30 Prozent Sozialdemokraten in der Bevölkerung (der Grundgesamtheit) vorhanden sind, dann sollten drei von zehn „Befragten“ Sozialdemokraten sein. Aber Achtung, der Anteil von Sozialdemokraten ist in Arbeitervierteln höher.
- Den Befragten müssen zwar keine Listen gezeigt werden, wohl aber Fotos und Karten, denn dies kann von Instituten leicht abgefragt werden und diese speziellen Fragen sind relativ leicht zu erinnern.
- Nicht gefälscht werden sollten größere Teile von Panel-Daten, da ist die Gefahr zu groß, entdeckt zu werden.
- Wenn es Vergütungen gibt, dann müssen diese grundsätzlich an die Befragten überreicht werden; die nehmen die „angemessene“ Bezahlung für (relativ) wenig eigenen Aufwand gerne an.
- Manchmal kann Betrug auch harte Arbeit sein – man muss etwas über die Produkte wissen, zu denen Fragen gestellt werden. Im Zweifelsfall schauen Sie im Lexikon nach, fragen Sie einen Freund oder verwenden Sie Google, und wenn auch dies nicht ausreicht, dann führen Sie einige Interviews korrekt durch.
- Achten Sie darauf, dass Sie Fragebatterien nicht zu konsistent und vor allem nicht zu stereotyp beantworten, dies können Blasius/Thiessen (2012, 2013) mit Hilfe ihrer Screening-Prozeduren entdecken.

---

## 22.5 Entdecken und Vermeidung von Fälschungen

Es gibt verschiedene Möglichkeiten, mit denen Institute prüfen können, ob die Interviews korrekt oder nicht korrekt durchgeführt wurden. In einem *Telefonlabor*, welches inzwischen zum Standard aller großen Sozial- und Meinungsforschungsinstitute gehört, kann der Supervisor die einzelnen Interviewer zu jeder Zeit kontrollieren, ohne dass diese es merken (Hüfken, Kapitel 46 in diesem Band). Von daher sollte es hier eigentlich keine Probleme geben, es sei denn, das Institut (ein verantwortlicher Mitarbeiter des Instituts) nimmt es mit der Kontrolle selbst nicht so genau.

Bei *Face-to-Face-Interviews* (Stocké, Kapitel 45 in diesem Band) können mit Hilfe von *Computer Assisted Personal Interviews (CAPI)* Datum und Zeit der Interviews sowie die Dauer für die Beantwortung einzelner Frageblöcke aufgezeichnet werden (sogenannte Paradata, Felderer/Birg/Kreuter, Kapitel 25 in diesem Band), anhand derer die

Befragungssituation genau analysiert werden kann (Lück/Landrock, Kapitel 28 in diesem Band). Allerdings ist die Verwendung von CAPI noch keine Garantie dafür, dass diese z.T. recht aufwendigen Berechnungen zur Überprüfung des zeitlichen Ablaufs der Interviews auch tatsächlich vorgenommen werden. So hat auch der Einsatz von CAPI beim ALLBUS 2008, eine der wichtigsten sozialwissenschaftlichen Studien in Deutschland (Mochmann, Kapitel 14 in diesem Band), längst nicht alle Interviewer davon abgehalten, Teile des Fragebogens zu fälschen (Blasius/Thiessen 2013).

Zusätzlich zu den im Internet publizierten Empfehlungen (AAPOR 2003) geben zumindest die großen Institute an, ihre eigenen Routinen entwickelt zu haben, um Fälschungen zu entdecken; diese werden allerdings als „Betriebsgeheimnisse“ nicht öffentlich gemacht – und dass sie zumindest für den ALLBUS 2008 nicht funktioniert haben (nicht angewendet wurden?), belegen Blasius/Thiessen (2013). Eine weitere Möglichkeit Fälschungen schon zu Beginn der Feldphase zu entdecken, besteht darin, schon beim Eingang der ersten 100 oder 200 Fragebögen einer Studie, diese nach „verdächtigen“ Antwortmustern zu screenen (Blasius/Thiessen 2012, 2013).

## Literatur

- AAPOR (2003): Interviewer Falsification in Survey Research: Current Best Methods for Prevention, Detection and Repair of Its Effects. [http://www.aapor.org/AM/Template.cfm?Section=Resources\\_for\\_Researchers&Template=/CM/ContentDisplay.cfm&ContentID=1462](http://www.aapor.org/AM/Template.cfm?Section=Resources_for_Researchers&Template=/CM/ContentDisplay.cfm&ContentID=1462)
- Biemer, Paul P./Stokes, S. Lynne (1989): The Optimal Design of Quality Control Samples to Detect Interviewer Cheating. In: *Journal of Official Statistics* 5: 23-39
- Blasius, Jörg/Friedrichs, Jürgen (2012): Faked Interviews. In: Salzborn et al. (Hg.): 49-56
- Blasius, Jörg/Thiessen, Victor (2012): Assessing the Quality of Survey Data. London: Sage
- Blasius, Jörg/Thiessen, Victor (2013) Detecting Poorly Conducted Interviews. In: Winker et al. (Hg.): 67-88
- Blasius, Jörg/Thiessen, Victor (2014): Should we Trust Survey Data? Response Simplification and Data Fabrication in the PISA Schools Principals' Survey. Eingereicht
- Crespi, Leo P. (1945): The Cheater Problem in Polling. In: *Public Opinion Quarterly* 9: 431-445
- Harrison, David E./Krauss, Stefanie L. (2002): Interviewer Cheating: Implications for Research on Entrepreneurship in Africa. In: *Journal of Developmental Entrepreneurship* 7: 319-330
- Merton, Robert K. (1957): Social Theory and Social Structure. Glencoe, Ill.: The Free Press.
- Nelson, James E./Kiecker, Pamela L. (1996): Marketing Research Interviewers and Their Perceived Necessity of Moral Compromise. In: *Journal of Business Ethics* 15: 1107-1117
- Salzborn, Samuel/Davidov, Eldad/Reinecke, Jost (Hg.) (2012): Methods, Theories, and Empirical Applications in the Social Sciences. *Festschrift für Peter Schmidt*. Wiesbaden: Springer VS
- Schäfer, Christian/Schräpler, Jörg-Peter/Müller, Klaus-Robert/Wagner, Gert G. (2005): Automatic Identification of Faked and Fraudulent Interviews in the German SOEP. In: Schmollers Jahrbuch 125: 183-193
- Winker, Peter/Menold, Natalja/Porst, Rolf (Hg.) (2013): Interviewers' Deviations in Surveys – Impact, Reasons, Detection and Prevention. Frankfurt: Peter Lang

Uwe Engel und Björn Oliver Schmidt

---

## 23.1 Unit-Nonresponse

Wenn Stichproben nach dem Zufallsprinzip (Häder/Häder, Kapitel 18 in diesem Band) generiert werden, kommt es regelmäßig vor, dass in die Stichprobe gezogene Personen von vornherein nicht teilnehmen, den Fragebogen nur teilweise beantworten oder aus unterschiedlichen Gründen nicht erreicht werden können (Engel et al. 2004: 43). Von *Unit-Nonresponse* (*Totalausfall*, *Komplettausfall*) ist dabei die Rede, wenn in die Stichprobe gezogene Personen die Teilnahme an einer Befragung verweigern oder aus Gründen der Nichterreichbarkeit oder z.B. bedingt durch Sprach- oder Verständigungsprobleme keine gültigen Aussagen vorliegen (Bartsch 2012: 48).

Das Phänomen sinkender Ausschöpfungsquoten beschäftigt die Umfrageforschung schon seit Langem. Die *Ausschöpfungsquote* oder auch *Response-Rate* beschreibt, wie viele Personen der Ausgangsstichprobe anteilig an einer Umfrage teilgenommen haben. Sie ist u.a. für die Beurteilung der Qualität einer Stichprobe von Bedeutung, da der Spielraum für ausfallbedingte Verzerrungen einer realisierten Stichprobe steigt, wenn die Ausschöpfung sinkt. Studien deuteten zwischenzeitlich auf einen allgemeinen, wenngleich im Ländervergleich variierenden Trend hin. So haben de Leeuw/de Heer (2002) in einer ländervergleichenden Analyse vorliegender Studien festgestellt, dass die Zahl an Verweigerern pro Jahr um 0,3 Prozentpunkte zunimmt. Zweifel daran, ob dies für einzelne Umfragemodi in gleicher Weise anzunehmen ist, wecken jedoch neuere Studien. Liegen beispielsweise die Antwortraten im Rahmen von face-to-face Befragungen (Stocké, Kapitel 45 in diesem Band) in der dritten Runde des European Social Survey (ESS) über alle einbezogenen Länder hinweg innerhalb eines Spektrums von ca. 46 bis 73% (Stoop et al. 2010: 93), so werden für neuere telefonische Befragungen (Hüfken, Kapitel 46 in diesem Band) Antwortraten von 20% und weniger berichtet (Häder et al. 2012: 124; Engel et al. 2012: 66f.). Zudem können Stoop et al. (2010) für face-to-face Befragungen im Vergleich der ersten drei Runden des ESS kein klares Bild in den Antwortraten feststellen und sehen den Trend

sinkender Antwortraten unter Verweis auf neuere Studien gestoppt bzw. sogar umgedreht. Zugleich finden die Autoren Hinweise darauf, dass Veränderungen in den Antwortraten auch von Veränderungen in der Feldarbeit abhängen (Stoop et al. 2010: 106ff.). Zuvor hatte auch eine Metaanalyse (Wagner/Weiß, Kapitel 88 in diesem Band) über Studien aus den 1970er-, 1980er- und 1990er-Jahren ergeben, dass nach statistischer Kontrolle einer Reihe von Studienbedingungen keine Unterschiede in den Antwortraten festzustellen waren, die diesen drei Perioden zuzuschreiben gewesen wären (Engel/Schnabel 2004: 23f.).

Kompletausfälle werden für die Datenqualität problematisch, wenn sich teilnehmende Personen in relevanten Merkmalen (z.B. in Form von soziodemografischen Charakteristika, Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band) systematisch von Verweigerern und Nicht-Erreichbaren unterscheiden. Nonresponse wird nie völlig zu vermeiden sein und muss deswegen schon bei der Untersuchungsplanung Berücksichtigung finden. Strebt man einen bestimmten Umfang der Nettostichprobe (Umfang der Bruttostichprobe abzüglich aller Ausfälle) an, so muss eine entsprechend große Bruttostichprobe (Umfang aller Personen in der Stichprobe) gezogen werden, um auf eine ausreichende Anzahl an Personen zu kommen, die den Fragebogen auch tatsächlich beantworten. Da sämtliche in die Stichprobe gezogenen Personen kontaktiert werden, ist zu beachten, dass hierdurch ein Mehr an Kosten entstehen kann. Bei schriftlich-postalischen Befragungen beispielsweise durch den Druck und den Versand der Fragebögen.

*Verweigerungen* liegen vor, wenn Kontakt zu einem potentiellen Befragten hergestellt werden konnte, eine Teilnahme aber abgelehnt wird. Von *Nicht-Erreichbarkeit* spricht man, wenn ausgewählte Personen aus unterschiedlichen Gründen während einer Feldphase nicht erreicht werden können. Werden Populationsparameter infolgedessen allein auf Basis der Angaben der Teilnehmer geschätzt, so müssen diese Stichprobenschätzungen die zu schätzenden Werte in der Population zwangsläufig verfehlten. Die Rede ist dann von „*bias*“ und *verzerrten Schätzungen* (Groves et al. 2002: 4). So kann es hierdurch beispielsweise zu einer Unter- oder Überschätzung interessierender Maßzahlen kommen.

### **23.1.1 Ursachen für Unit-Nonresponse**

Eine Ursache für Unit-Nonresponse ist der *Befragungsmodus*. So variieren die Antwortraten (Engel et al. 2012: 77) beispielsweise im Vergleich von telefonischen Befragungen (Hüfken, Kapitel 46 in diesem Band) und webbasierten Befragungen (Wagner/Hering, Kapitel 48 in diesem Band). Vergleichsweise hohe Antwortraten sind in Abhängigkeit vom jeweiligen Design der Umfrage in der Regel im Rahmen persönlich-mündlicher Befragungen zu erwarten. Beispielsweise haben sich in der dritten Runde des European Social Survey (ESS) in Deutschland 53% persönlich-mündlich befragen lassen (Stoop et al. 2010: 93). Damit liegt Deutschland im europäischen Vergleich im unteren Drittel. Postalische Befragungen treffen häufig auf ähnliche Resonanz. So folgert Schnell (2012), dass bei postalischen Befragungen unter der „allgemeinen Bevölkerung“, bei professioneller Umsetzung, mit einer Antwortrate von 50% und höher gerechnet werden kann. Wird telefonisch

befragt, werden mehrheitlich geringere Antwortraten berichtet. In der in den Jahren 2007 und 2008 via Festnetz durchgeführten CELLA-(Haupt-)Studie haben beispielsweise 20% der ausgewählten Personen ein Interview gegeben (Häder et al. 2009: 72). Webbasierte Umfragen erzielen im Vergleich zu „traditionellen“ Befragungsmodi, wie z.B. postalisch, einen geringeren Rücklauf (Fan/Yan 2010). So konnten z.B. Converse et al. (2008) nachweisen, dass die Antwortrate in einer Onlineumfrage, im Vergleich zu einer gleichzeitig durchgeführten postalischen Umfrage, signifikant niedriger ausfällt.

Es gibt allerdings eine Gruppe von Faktoren, die unabhängig vom Modus der Befragung für die Höhe der Ausschöpfungsquote verantwortlich ist (Couper/Groves 1996: 66). Aus theoretischer Perspektive wird der *Entscheidungsprozess eines potentiellen Befragten für oder gegen eine Teilnahme* zumeist handlungsorientiert erklärt. Ein verbreiteter Ansatz ist, den Entscheidungsprozess als mehrstufige, rationale Kosten-Nutzen-Abwägung („Rational Choice Ansatz“) zu beschreiben. Dillman et al. (2009) haben hierzu angemerkt, dass die Kosten einer Teilnahme für den Befragten gering zu halten und der Nutzen zu maximieren seien, sodass sich möglichst viele Personen zu einer Teilnahme bereit erklären. Nutzen ist hier in dem Sinne zu verstehen, dass ein potentieller Befragter für sich den Eindruck gewinnt, dass er mit seiner Teilnahme einen *wichtigen Beitrag zur öffentlichen Meinungsbildung sowie Wissenschaft und Forschung* leistet. Ein praktischer Nutzen liegt beispielsweise dann vor, wenn *Teilnahmeincentives* eingesetzt und so im Vorhinein einer Befragung ein geringer Geldbetrag als „kleines Dankeschön“ vergeben wird. Dieser sollte seitens des Befragten allerdings nicht als Bezahlung wahrgenommen werden. Denn wird ein Teilnahmeincentiv im ökonomischen Sinne als eine Art Gegenleistung für die Mühe und Aufwendung von Zeit zur Beantwortung eines Fragebogens wahrgenommen und dieser ggf. als inadäquat empfunden, fällt es dem potentiellen Befragten relativ leicht, ein Anliegen abzulehnen (Engel 2006: 6). Dies zeigt exemplarisch, dass das Kosten-Nutzen-Verhältnis durch *Designelemente einer Umfrage* teilweise direkt beeinflusst werden kann. Wird zudem beachtet, dass die Bereitschaft, sich befragen zu lassen, nachweislich auch davon beeinflusst wird, ob bzw. welche Erfahrungen jemand zuvor bereits mit Umfragen gemacht hat (Engel 2013), so wird erkennbar, dass auch ein professionelles und im Vorhinein durch einen sogenannten „Pre-Test“ ausreichend erprobtes Studiendesign (Weichbold, Kapitel 19 in diesem Band) sowie insbesondere ein guter Fragebogen (Reinecke, Klöckner/Friedrichs, Kapitel 44 und 49 in diesem Band) zentral für die Höhe der Ausschöpfungsquote sind. Wenn die Bilanz der rationalen Kosten-Nutzen-Abwägung trotz aller Bemühungen negativ ausfällt, z.B. durch die verlorene Zeit oder Desinteresse am Thema der Umfrage, wird ein potentieller Befragter eine Teilnahme in der Regel verweigern (Engel et al. 2004: 55).

Die zentralen Ursachen für Totalausfälle sind primär in *Verweigerungen und Nichterreichbarkeit* zu sehen (de Leeuw/de Heer 2002: 44). So gibt es Personen, zu denen zwar ein Kontakt hergestellt werden kann, die sich aber nicht bereit erklären, teilzunehmen. Bei interviewervermittelten Umfragen kann der Versuch eines Interviews aber bereits daran scheitern, dass trotz aller Versuche kein Kontakt zur potentiellen Befragungsperson hergestellt werden kann. Insofern ist es nötig, mehrere Versuche zu unternehmen, um die

Wahrscheinlichkeit eines Kontaktes zu erhöhen, respektive sogenannte „Reminder“ (bei selbstdurchführten Befragungsformen) zu versenden, um an die noch ausstehende Teilnahme an einer Umfrage zu erinnern. Im Rahmen telefonischer Umfragen sollten mindestens fünf Kontaktversuche unternommen werden, da die Wahrscheinlichkeit, eine Person zu erreichen, bei den ersten Versuchen am größten ist (Groves/Couper 1998; Engel et al. 2012: 120-123). Eine vorherige Kontaktaufnahme zur Ankündigung des Telefonanrufs kann in der Regel nicht erfolgen. Denn Telefonnummern werden häufig per Zufallsprinzip (Häder/Häder, Kapitel 18 in diesem Band) generiert, sodass keine weiteren Informationen für eine erste Kontaktaufnahme vorliegen.

Bei mündlich-persönlichen Befragungen ist darauf zu achten, dass ein Kontaktversuch im Vergleich zu anderen Befragungsformen mit höheren Kosten verbunden ist. Hier muss also ein Kompromiss zwischen dem bei einer Befragung zur Verfügung stehenden Budget und der relevanten Anzahl an Versuchen, eine Person in einem Haushalt anzutreffen, gefunden werden. In der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) werden beispielsweise vier Kontaktversuche unternommen, bevor eine Adresse als „nicht erreichbar“ und somit als Ausfall definiert wird (Haarmann et al. 2006).

Für postalische Befragungen werden von Dillman et al. (2009: 243) insgesamt fünf Kontaktierungsschritte vorgeschlagen:

1. Ankündigungsschreiben
2. Fragebogen inklusive Anschreiben
3. Erinnerungskarte
4. Ersatzfragebogen
5. Ggf. Erinnerungsanruf

Die ersten drei Kontaktierungsschritte sollten dabei in wöchentlichen sowie die letzten beiden in zweiwöchigen Abständen erfolgen (Dillman et al. 2009). Für eine Onlinebefragung werden häufig zunächst Einladungen per E-Mail an die potentiellen Teilnehmer versandt (Fan/Yan 2010). Da bei Bevölkerungsumfragen in der Regel kein geeigneter Auswahlrahmen für die zufällige Ziehung von E-Mail Adressen existiert, werden Personen häufig zunächst auf anderem Wege kontaktiert (z.B. telefonisch oder postalisch), um die benötigten Kontaktinformationen zu eruieren. Nachdem eine Umfrageeinladung per E-Mail versandt wurde, können in Abhängigkeit der Teilnehmerzahlen in regelmäßigen Abständen Erinnerungs-E-Mails („Reminder“) an diejenigen verschickt werden, die den Fragebogen noch nicht beantwortet haben. Die Häufigkeit der Erinnerungen kann dabei von der Höhe der angestrebten Antwortrate abhängig gemacht werden.

Trotz häufiger Kontaktversuche werden wissenschaftliche Umfragen stets dem Problem ausgesetzt sein, dass zu ausgewählten Personen kein Kontakt hergestellt werden kann. Zu fragen ist also zunächst nach den Faktoren, die die *Kontaktwahrscheinlichkeit* beeinflussen. Durch gesellschaftliche Veränderungen sind bestimmte Bevölkerungsgruppen heutzutage immer schlechter zu erreichen, sodass wissenschaftliche Umfragen häufig mit dem Problem konfrontiert werden, dass zu ausgewählten Personen kein Kontakt hergestellt werden kann (de Leeuw/de Heer 2002: 52; Engel et al. 2004: 45):

- Bei Telefonbefragungen (Hüfken, Kapitel 46 in diesem Band) liegt eine Ursache darin, dass Haushalte vermehrt über einen *Anrufbeantworter* oder eine *Anruferkennung* verfügen (Groves et al. 2002: 8). So können Personen bei einer ihnen unbekannten Telefonnummer den Hörer entweder erst gar nicht abnehmen oder den Anrufbeantworter anspringen lassen.
- Ebenso gibt es immer mehr Singlehaushalte, wodurch bei *berufsbedingter Abwesenheit* keine Person im Haushalt erreichbar ist (Engel et al. 2004). Insgesamt sollte demnach das Survey Design so angelegt sein, dass einzelne Kontaktversuche zu möglichst variierenden Wochentagen sowie Tageszeiten stattfinden. So besteht nach Groves et al. (2002) eine höhere Chance eines eventuellen Kontaktes, wenn vor allem abends und am Wochenende angerufen wird. Engel (2013) konnte diesbezüglich zeigen, dass im Rahmen von telefonischen Befragungen die Interviewwahrscheinlichkeit zur Mittagszeit sinkt, um am frühen Abend wieder zu steigen.
- Vor allem jüngere Menschen gehören immer häufiger zu der Gruppe der „*Mobile Onlys*“, die keine Festnetznummer besitzen, sondern ausschließlich über das Handy erreichbar sind. Hierdurch wird eine telefonische Kontaktaufnahme zusätzlich erschwert.
- Bei face-to-face Befragungen (Stocké, Kapitel 45 in diesem Band) wird die Kontaktaufnahme oft durch *physikalische Hindernisse* erschwert, wie z.B. verschlossene zentrale Eingänge.
- Im Rahmen postalischer (Reuband, Kapitel 47 in diesem Band) sowie webbasierter Umfragen (Wagner/Hering, Kapitel 48 in diesem Band) führen in der Regel *falsche Kontaktinformationen* (Postanschriften oder E-Mail-Adressen) oder E-Mail-Spamfilter dazu, dass Personen nicht erreicht werden können (Groves et al. 2002).

Kommt ein Kontakt zustande, so ist die Wahrscheinlichkeit der Realisierung eines Interviews von einer Reihe weiterer Merkmale abhängig. Ob ein Befragter zu einer Teilnahme bereit ist, liegt einerseits an Aspekten des Umfrage-Designs, der Kontakt- und Interviewsituation selbst sowie andererseits an den Eigenschaften des Befragten und dessen Lebensumständen (Engel et al. 2012: 14). Couper/Groves (1996) teilen die *Determinanten der Umfragebereitschaft* in vier Blöcke ein:

1. Soziale Umwelt (Umfrageklima, Urbanisierungsgrad),
2. Eigenschaften des Haushaltes bzw. des Befragten selbst (Haushaltsstruktur, soziodemografische Charakteristika, psychologische Dispositionen),

Tabelle 23.1 gibt einen Überblick darüber, welche Gruppen im Rahmen wissenschaftlicher Umfragen in der Regel unterrepräsentiert sind.

**Tab. 23.1** Unterrepräsentierte Bevölkerungsgruppen innerhalb wissenschaftlicher Umfragen

Merkmal	Befragungsmodus
Lebt in urbaner Umgebung (vs. Land)	Face-to-face, telefonisch, web, postalisch
Zugehörigkeit zu ethnischer Minderheit	Face-to-face, telefonisch, web, postalisch
Zugehörigkeit zu oberer/älterer Altersgruppe	Face-to-face, telefonisch, gilt vor allem für mobil und web
Niedriger Bildungsabschluss	Face-to-face, telefonisch, web, postalisch

Allgemein sind Personen unterrepräsentiert, die als älter und schlechter gebildet zu charakterisieren sind (Bartsch 2012; Engel et al. 2012). Hier wird häufig von einem sogenannten „Bildungsbias“ gesprochen, was meint, dass es eher die besser im Gegensatz zu den schlechter gebildeten Personen sind, die sich an Umfragen beteiligen. Kaum vertreten sind ältere Personen vor allem bei online durchgeführten Befragungen sowie bei Befragungen über das Mobilfunknetz (Bartsch 2012). Der Grund dafür liegt zumeist darin, dass ältere Personen seltener über einen Internetanschluss bzw. ein Handy verfügen. Nicht zuletzt sind es in Deutschland eher die Stadt- als die Landbewohner, die sich seltener an wissenschaftlichen Umfragen beteiligen.

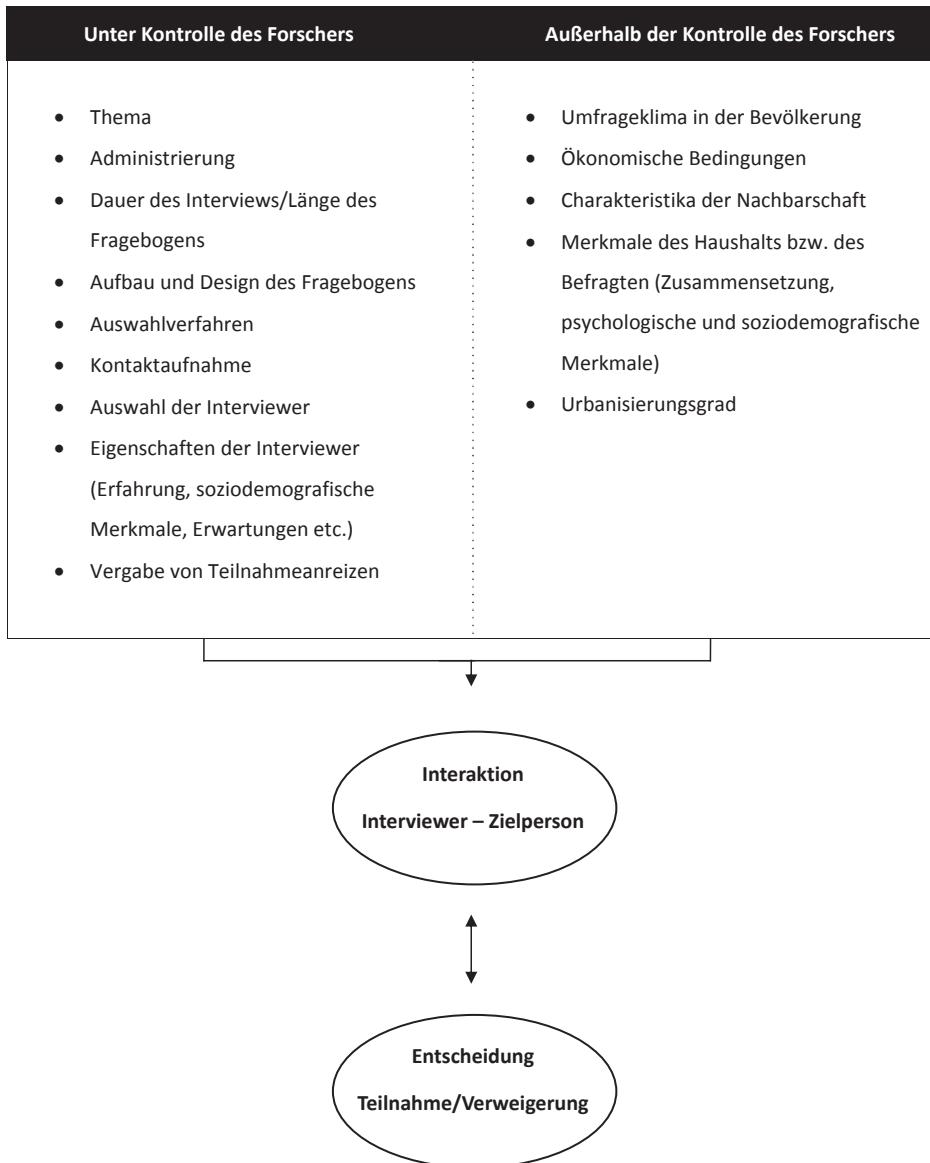
3. Survey Design (Thema, Auswahl der Befragten, Länge des Fragebogens etc.).  
Die Länge des Fragebogens und die damit von Seiten des Befragten zu investierende Zeit zur Beantwortung ist ein ausschlaggebender Faktor dafür, ob sich jemand bereit erklärt, einem Umfragegesuch zuzustimmen oder abzulehnen. Insofern ist für ein professionelles Survey Design ein Gleichgewicht zwischen der Anzahl an Fragen (Informationsvielfalt) und der Länge des Fragebogens zu finden. Orientierungshilfe zur Planung eines Surveys und der damit verbundenen Dauer eines Interviews gibt Schnell (2012: 309):
4. Interviewer (soziodemografische Charakteristika, Erfahrung, Erwartungen).  
Neben einer guten Schulung (Glantz/Michael, Kapitel 21 in diesem Band) ist es von Vorteil, Interviewer einzusetzen, die bereits über Erfahrung verfügen. Diese können beispielsweise flexibler auf mögliche Bedenken eines Befragten reagieren und erhöhen so die Wahrscheinlichkeit einer Teilnahme (Engel et al. 2012: 195f.). Darüber hinaus sollte ein Interviewer, besonders bei face-to-face Interviews, ein möglichst selbstsicheres Auftreten haben. Als förderlich hat es sich zudem erwiesen, wenn Interviewer davon überzeugt sind, dass tendenziell jede Person zu einer Teilnahme motiviert werden kann (Engel et al. 2012; Hox/de Leeuw 2002).

**Tab. 23.2** Maximale Dauer eines Interviews

Face-to-Face	Telefonisch	Online	Postalisch
60 Minuten	20 Minuten	15 Minuten	15 Minuten

Quelle: Angaben (leicht adaptiert) aus Tabelle 12.1 in Schnell 2012: 309 übernommen.

Punkt 4 fällt als Einflussmöglichkeit bei den selbst administrierten Befragungsformen „online“ (Wagner/Hering, Kapitel 48 in diesem Band) und „postalisch“ (Reuband, Kapitel 47 in diesem Band) weg, da hierfür keine Interviewer eingesetzt werden. Abbildung 23.1 gibt in Anlehnung an Couper/Groves (1996) sowie Engel et al. (2004) einen Überblick über Determinanten der Umfrageteilnahme.



**Abb. 23.1** Determinanten der Umfrageteilnahme in Anlehnung an Couper/Groves (1996) sowie Engel et al. (2004)

Die *Faktoren der sozialen Umwelt*, wie das Umfrageklima oder der Urbanisierungsgrad sowie Eigenschaften des Befragten selbst, entziehen sich dabei in der Regel der Kontrolle des Forschers (Abb. 23.1). Das Umfrageklima in der Bevölkerung leidet dabei zunehmend darunter, dass Verkaufsanrufe unter dem Deckmantel von Befragungen in den letzten Jahren zugenommen haben (Engel et al. 2012: 69). Ferner wird häufig versucht, die Umfragetriebereitschaft mittels soziodemografischer Charakteristika der Befragten zu erklären. Insgesamt sind die Ergebnisse hierzu nicht eindeutig und durch das unterschiedliche Design von Umfragen nur bedingt miteinander vergleichbar.

Engel et al. (2012: 279ff.) haben beispielsweise herausgefunden, dass sich die Soziodemografie nicht nennenswert auf die Wiederbefragungsbereitschaft innerhalb eines Panels auswirkt. Als relevant erweist sich die Soziodemografie (insbesondere der Bildungsgrad) jedoch, wenn es um die Wahrscheinlichkeit einer ggf. ersten Umfrageteilnahme geht. Denn in realisierten Stichproben sind höhere Bildungsgruppen typischerweise überrepräsentiert („Bildungsbias“). Zufallsstichproben können aber auch dadurch selektiv werden, dass sich zur Teilnahme an einer Umfrage eingeladene Personen verstärkt beteiligen, wenn diese schon zuvor zu Befragungen eingeladen worden waren und teilgenommen haben (Engel 2013: 14). Die Umfrageteilnahme erklärt sich hier endogen aus sich selbst heraus: Man nimmt teil, weil man zuvor bereits teilgenommen hatte.

Insgesamt scheint es so zu sein, dass soziodemografische Charakteristika, je nach Modus der Befragung, einen unterschiedlichen Einfluss auf die Teilnahmewahrscheinlichkeit ausüben. So gibt es empirische Evidenz dahingehend, dass in telefonischen, face-to-face und Online-Befragungen jüngere Personen überrepräsentiert sind (Bartsch 2012; Engel et al. 2004: 105). Die Teilnahmebereitschaft variiert offensichtlich mit dem Alter und dem jeweils eingesetzten Erhebungsverfahren.

### **23.1.2 Umgang mit Unit-Nonresponse**

Zum Umgang mit Nonresponse führt Schnell (2012: 180f.) treffend aus: „Die einzige mögliche Antwort auf Nonresponse kann nur in der Berücksichtigung des Nonresponse bei der Planung eines Surveys, sorgfältiger Feldarbeit, deren Dokumentation und einer statistischen Analyse, die auf Nonresponse Rücksicht nimmt, bestehen.“ Dabei sollte vor allem eine Kombination von *Maßnahmen zur Reduktion von Nonresponse* führen, zu denen Schnell (2012: 181) die folgenden zählt:

- Schriftliche Ankündigung der Befragung,
- Erhöhung der Anzahl der Kontaktversuche,
- Wechsel der Art der Kontaktaufnahme,
- erneute Kontaktierung und der Versuch der Konvertierung von Verweigerern sowie die
- Verwendung von Befragungsanreizen.

Über die Gestaltung einer Umfrage ist es in der Tat möglich, die Kontakt- und Kooperationswahrscheinlichkeit – und darüber vermittelt präventiv auch Unit-Nonresponse – positiv zu beeinflussen. So stellen *Interviewer* (Glantz/Michael, Kapitel 21 in diesem Band) eine relevante Einflussgröße der Teilnahmewahrscheinlichkeit dar, weshalb eine adäquate Schulung im Vorhinein unumgänglich ist (Hox/de Leeuw 2002). Es gilt als empirisch bestätigt, dass Interviewer im Rahmen telefonischer Befragungen durch ihre stimmlichen Merkmale, die Art und Weise der Gesprächseröffnung oder ihre Einstellungen gegenüber dem Umgang mit zögerlichen Personen einen direkten Einfluss auf die Interviewwahrscheinlichkeit ausüben (Engel et al. 2012: 170f.). Erfahrene Interviewer sind darüber hinaus eher in der Lage, flexibel auf schwierige Situationen zu reagieren und so beispielsweise das jeweils passende Argument zu finden, um zögerliche Personen von einer Teilnahme zu überzeugen (Hox/de Leeuw 2002).

Über *adaptive Umfragedesigns* (Bethlehem et al. 2011: 395ff.) ist es möglich, einzelne Elemente flexibel an Erfordernisse der Kontaktsituation anzupassen, um die Wahrscheinlichkeit von Verweigerungen zu minimieren. So ist es im Rahmen von *Mixed-Mode Designs* möglich, es den Befragten zu überlassen, in welchem Umfragemodus das erbetene Interview stattfinden soll. Als vorteilhaft hat es sich zudem erwiesen, wenn Interviewer neben der Vollfassung auch kürzere Interviewversionen anbieten können, um zögerliche Personen von einer Umfrageteilnahme zu überzeugen; „basic question“-Ansatz ist dafür eine aktuelle Bezeichnung (Bethlehem et al. 2011: 295ff.). Man gewinnt durch diesen Ansatz Personen für eine Umfrage, die ansonsten nicht teilnehmen würden. Dies wiederum erhöht nicht nur die Ausschöpfung der Stichprobe, sondern liefert zugleich auch relevante Informationen zur Modellierung des Ausfallprozesses (Engel et al. 2012: Kap. 10). Ein Umfragedesign kann zudem vorsehen, dass Interviewer in der Kontaktsituation flexibel auf Bedenken und Argumente zögerlicher Personen eingehen, die um ein Interview gebeten werden (Groves et al. 2000). Zwar erweisen sich entsprechende Versuche in der Tat als wirksam, jedoch darf nicht übersehen werden, dass ihr Spielraum umso stärker begrenzt wird, je höher in einer Studie der Anteil von Spontanverweigerern ausfällt. Bei telefonischen Befragungen sind dies Personen, die den Telefonhörer gleich wieder auflegen, ohne dass zuvor ein Gespräch zustande kommen konnte, oder Personen, die sofort verweigern, ohne mit sich reden zu lassen. Im Rahmen der Rekrutierungsphase des PPSM Access Panels (für weitere Informationen hierzu siehe Engel et al. 2012; Engel 2013) waren beispielsweise 44% der 56.848 Fälle, in denen telefonisch Kontakt mit dem Haushalt hergestellt werden konnte, Personen, die den Hörer spontan gleich wieder aufgelegt oder die Bitte nach dem Interview ebenso spontan wie bestimmt zurückgewiesen haben. Darüber hinaus lag der Anteil der Personen, die überhaupt Bedenken vorgetragen haben und somit per definitionem als „zögerliche Personen“ angesehen werden konnten, „lediglich“ bei 28% (Engel 2013: 14, 51). Nur ein Teil der Personen ließ also überhaupt mit sich reden. Entsprechend begrenzt ist die Anwendbarkeit von Techniken, die darauf abzielen, zögerliche Personen auf argumentativem Wege zu einer Interviewteilnahme zu bewegen.

Bekannt ist, dass *Ankündigungsschreiben* (Klöckner/Friedrichs, Kapitel 49 in diesem Band) im Vorfeld der tatsächlichen Kontaktaufnahme oder Erinnerungsschreiben bei selbst administrierten Befragungsformen eine Teilnahme wahrscheinlicher werden lassen (Groves et al. 2002: 11, Reuband, Kapitel 47 in diesem Band). Als außerordentlich ausschöpfungserhöhend erweisen sich zudem vorab gewährte monetäre *Befragungsanreize* (Engel/Schnabel 2004; Engel et al. 2012: 128f.) und die Kombination solcher Incentives mit dem Einsatz von Ankündigungsschreiben und einem Mixed-Mode Design, welches die Kontakt-/Erhebungsmodi „Telefon“ und „Online“ kombiniert (Engel/Schmidt 2013).

Über präventive Maßnahmen kann die Anzahl von Ausfällen reduziert werden, Ausfälle werden aber nach aller Erfahrung weiterhin noch in erheblichem Umfang auftreten. Von essentieller Bedeutung ist daher, erkennen zu können, ob bzw. in welcher Weise es sich um *systematische Ausfälle* handelt. Diese liegen in der Regel dann vor, wenn Ausfälle nicht zufällig erfolgen. Dieser Fall tritt ein, wenn die Wahrscheinlichkeit einer Umfrageteilnahme mit Eigenschaften des Befragten, der Umfrage selbst (z.B. mit dem Umfragethema) oder mit der einem Interview oder einer Verweigerung unmittelbar vorausgehenden Kontakt situation zwischen Interviewer und angefragter Person zusammenhängt. Eine Möglichkeit festzustellen, ob es sich um systematische Ausfälle handelt, bietet eine *Analyse der Antwortneigung*, also eine Analyse von Determinanten der Teilnahmewahrscheinlichkeit, um diese dann z.B. als Basis für eine kompensatorische Stichprobengewichtung zu nutzen (Engel et al. 2004: 117ff.). Dafür sind allerdings auch Informationen über diejenigen in eine Stichprobe (Häder/Häder, Kapitel 18 in diesem Band) gezogenen Personen erforderlich, die nicht an der Befragung teilnehmen.

Ein Vorteil von *registergestützten Auswahlen* liegt darin, genau solche Informationen zu bieten. Melderegisterstichproben wären dafür ein Beispiel. Zwar ist die Erklärungskraft der im Melderegister verfügbaren Informationen stark begrenzt, über die Wohnadressen besteht aber die Möglichkeit, die Wohnumgebungen durch *Aggregat- bzw. Gebietsdaten* zu charakterisieren und diese Informationen in die Analyse der Antwortneigung einzubeziehen. Die Möglichkeit, Gebietsdaten einzubeziehen, besteht über die sogenannten Gemeindekennziffern z.B. auch bei telefonischen Befragungen.

Verfügt man über solche Informationen, ist es möglich, *individuelle Response-Wahrscheinlichkeiten zu schätzen*, um darüber vermittelt z.B. ausfallbedingte Verzerrungen in Stichprobenschätzungen identifizieren und quantifizieren zu können. Gestützt auf das „random response model“ hat z.B. Bethlehem eine Formel entwickelt, in die diese individuellen Antwortneigungen eingehen (Bethlehem et al. 2011: 43-45).

Auch besteht die Möglichkeit, *Stichprobenschätzungen so zu korrigieren*, dass differenzielle Antwortneigungen z.B. qua Berechnung modifizierter Horvitz-Thompson-Schätzer berücksichtigt werden. Bethlehem et al. (2011: 330f.) beschreiben die damit verbundene Idee sehr schön: Horvitz und Thomson hatten gezeigt, dass es stets möglich sei, den Mittelwert der Häufigkeitsverteilung einer Variable in der Population (Grundgesamtheit) unverzerrt zu schätzen, wenn die individuellen Auswahlwahrscheinlichkeiten bekannt sind. Der „modifizierte“ Horvitz-Thompson-Schätzer resultiert nun, wenn (neben bekannten Auswahlwahrscheinlichkeiten) auch die oben schon angesprochenen, individuell variierenden

Response-Wahrscheinlichkeiten berücksichtigt werden. Diese sind zwar unbekannt, können aber auf der Basis von Modellen geschätzt werden. Interindividuelle Unterschiede in diesen geschätzten Response-Wahrscheinlichkeiten können dann bei der Schätzung des Populationsmittelwerts berücksichtigt werden.

Zudem kann systematischen Unterschieden zwischen den Gruppen derer, die der Einladung zu einer Umfrage folgen vs. nicht folgen (Verweigerungen und Nicht-Erreichte), durch den Einsatz von *Gewichtungsfaktoren* (sogenannte *Redressment-Gewichte*, siehe Kiesl, Kapitel 24 in diesem Band) begegnet werden. Würde man bei Vorliegen solcher Unterschiede Schätzungen nur auf die Angaben der Teilnehmenden einer Umfrage stützen, so könnten diese Schätzungen die zu schätzenden Populationswerte nur verfehlten.

In dem zu einer Studie gehörenden Methodenbericht sollte immer die jeweilige Ausschöpfungsquote angegeben sein. Die entsprechend benötigten Informationen hierfür müssen bereits während des Datenerhebungsprozesses erfasst werden. Im Anschluss an eine Befragung muss zwischen Teilnehmern, Verweigerern und Personen differenziert werden können, die aus unterschiedlichen Gründen nicht erreicht werden konnten. Die American Association for Public Opinion Research (AAPOR) hat für die unterschiedlichen Befragungsmodi sogenannte „Standard Definitions“ herausgegeben, die regeln, wie Ausfälle zu klassifizieren und in die Berechnung der Antwortrate einzubeziehen sind (AAPOR 2011). Die Beachtung dieses online verfügbaren Regelwerks ist von größter Bedeutung für die korrekte Berechnung von Antwortraten. Dabei geht es vor allem um die korrekte Unterscheidung von relevanten und neutralen Ausfällen und darum, wie diesbezüglich unklare Fälle in die Berechnung der Antwortrate einzubeziehen sind.

---

## 23.2 Item-Nonresponse

Item-Nonresponse liegt vor, wenn aus unterschiedlichen Gründen nur ein Teil eines Fragebogens beantwortet wird (de Leeuw et al. 2003; 153, Shoemaker et al. 2000: 193). Eine Person erklärt sich zunächst bereit, an einer Befragung teilzunehmen, beantwortet dabei allerdings eine oder mehrere Fragen mit den Kategorien „keine Angabe“ oder „weiß nicht“ oder bricht die Befragung während des Interviews ab. Somit liegen nur für einen Teil der Fragen gültige Antworten vor. In diesem Fall ist von *fehlenden Werten* („missing values“) die Rede.

### 23.2.1 Ursachen für Item-Nonresponse

Mögliche Ursachen für Item-Nonresponse können im *kognitiven Prozess der Antwortfindung* sowie in der sozialen Interaktion zwischen Befragtem und Interviewer liegen (de Leeuw et al. 2003; Essig/Winter 2003; Shoemaker et al. 2000). Eine Frage muss zunächst verstanden und es müssen ggf. relevante Informationen aus dem Gedächtnis abgerufen werden, bevor im Hinblick auf zumeist vorgegebene Kategorien eine Antwort geformt

werden kann (Essig/Winter 2003). Dabei kann auftreten, dass es einer Person zu schwer fällt, eine Frage zu beantworten, beispielsweise weil sie sich an relevante Informationen nicht erinnern kann. Eine einfache Möglichkeit, sich einer als (zu) schwer oder mühselig empfundenen kognitiven Aufgabe zu entledigen, liegt darin, sich für die Antwortoptionen „keine Angabe“ oder „weiß nicht“ zu entscheiden.

Eine Entscheidung für die Kategorie „weiß nicht“ kann allerdings auch ein Indikator für *mangelnde Motivation* sein. Befragte unterscheiden sich in dem Grad an Energie, den sie bereit sind, in die Findung einer passenden Antwort zu investieren. So kann es vor allem bei Fragen mit vielen Antwortoptionen vorkommen, dass Personen mangels Motivation nicht die Antwortoptionen wählen, die ihre Meinung am ehesten wiederspiegeln, sondern dass es von der Reihenfolge der Optionen abhängt, welche sie wählen (Krosnick/Alwin 1987: 203). Engel/Köster (2014) konnten zeigen, dass Personen, die erst von einer Teilnahme an einer Befragung überzeugt werden mussten, in der Folge eher zu kognitiv *genügsamen Antwortverhalten (Satisficing)* neigen. Zudem konnten sie nachweisen, dass der Einfluss von Responseeffekten systematisch von Satisficing und der perzipierten Sensitivität einer Frage beeinflusst wird.

Eine weitere Quelle von Item-Nonresponse ist im perzipierten *Sensitivitätsgrad einer Frage* zu sehen (Hlawatsch/Krickl, Kapitel 20 in diesem Band). So konnten Shoemaker et al. (2000) zeigen, dass mit dem kognitiven Aufwand und dem Grad an Sensitivität auch der Umfang an Nonresponse zunimmt. Sensitive Fragen zielen in der Regel auf persönliche Informationen ab, die unter Umständen soziale Normen betreffen können. Befragte sind aber eher nicht dazu bereit, sich abweichendes Verhalten bzw. negativ bewertete Eigenschaften einzugehen und dies einem „fremden“ Interviewer gegenüber auch noch preiszugeben. Eine sensitive Frage kann aber einfach auch nur eine Frage sein, die von den Befragten als zu persönlich empfunden wird. Die Einkommensfrage ist dafür sicherlich ein gutes Beispiel. Auch andere Fragen zur Soziodemografie, wie z.B. die Frage nach dem jeweiligen Bildungsstatus oder dem Beruf, fallen durchaus in diese Kategorie (Engel 2013: 68f.).

Im Rahmen selbstadministrierter Befragungsformen (Reuband, Wagner/Hering, Kapitel 47 und 48 in diesem Band) verursachen sensitive Fragen weniger fehlende Werte, da es zu keiner sozialen Interaktion zwischen Befragtem und Interviewer kommt und die Befragungssituation somit als vertraulicher empfunden wird (de Leeuw et al. 2003: 161). Interviewer können andererseits einen durchaus positiven Effekt auf die Reduzierung von Item-Nonresponse haben, da sie z.B. einzelne Fragen erklären können, wenn der Befragte diese nicht versteht.

Nicht zuletzt ist ein *gut durchdachter* (Reinecke, Porst, Franzen, Klöckner/Friedrichs, Kapitel 44, 50, 51 und 49 in diesem Band) und im Vorhinein *getesteter Fragebogen* (Weichbold, Kapitel 19 in diesem Band) von Bedeutung. Durch einen logischen und übersichtlichen Aufbau kann die Wahrscheinlichkeit für Item-Nonresponse verringert werden (Klöckner/Friedrichs, Kapitel 49 in diesem Band). Eine Filterführung ist hierbei unumgänglich, sodass Teilnehmern nur die Fragen gestellt werden, die für sie relevant sind.

### 23.2.2 Umgang mit fehlenden Werten

Den in der Praxis der Sozialforschung eingesetzten Techniken zur statistischen Analyse von Datenmatrizen, in denen fehlende Werte auftreten, liegt in aller Regel die Annahme zugrunde, dass die Werte, die fehlen, zufällig fehlen. Mit fehlenden Werten kann dann im Rahmen von Datenanalysen auf unterschiedliche Weise umgegangen werden (Lück/Landrock, Kapitel 28 in diesem Band). Grundsätzlich bestehen dafür zwei Möglichkeiten:

- Fälle mit fehlenden Werten in einzelnen Variablen werden von statistischen Berechnungen ausgeschlossen oder
- fehlende Werte werden durch inhaltlich plausible Werte ersetzt.

Beim *Ausschluss von Fällen mit fehlenden Werten* werden wiederum zwei Varianten unterschieden: der listenweise und der paarweise Ausschluss (Allison 2002; Little/Rubin 2002). Relevant wird diese Unterscheidung, wenn in einer Analyse die gemeinsame Häufigkeitsverteilung von drei oder mehr als drei Variablen untersucht wird. Bei drei Variablen besteht diese „Liste“ aus drei Variablen, bei fünf simultan zu analysierenden Variablen aus fünf etc.

Der *listenweise Ausschluss von Fällen mit fehlenden Werten* („*complete case analysis*“) bezieht sich auf solche Listen und meint, dass von den Berechnungen alle Fälle ausgeschlossen werden, die in irgendeiner dieser Variablen fehlende Werte aufweisen. Bei einer aus fünf Variablen bestehenden Liste würden dann z.B. von der Berechnung der zwischen den Variablen 1 und 2 bestehenden Korrelation auch die Fälle ausgeschlossen werden, die zwar in diesen beiden Variablen gültige Werte haben, allerdings fehlende Werte in einer oder mehreren der anderen Variablen aufweisen.

Genau hierin liegt der Unterschied zum *paarweisen Ausschluss* („*available case analysis*“), da bei diesem Verfahren immer alle Fälle mit gültigen Werten in den einer Berechnung zugrundeliegenden Variablenpaaren berücksichtigt werden.

Beim listenweisen Ausschluss ist von Vorteil, dass sich statistische Berechnungen einheitlich auf dieselbe Zahl von Fällen beziehen und somit die Vergleichbarkeit univariater Statistiken ermöglicht ist (Little/Rubin 2002: 41). Nachteilig ist hingegen, dass sich beim listenweisen Ausschluss die Zahl der auszuschließenden Fälle sehr schnell erhöhen kann. Man verliert ggf. viele Fälle für eine Analyse und muss auch damit rechnen, dass sich die resultierende Teilmenge von Fällen mit validen Werten systematisch von der ursprünglichen Gesamtmenge von Fällen unterscheiden wird. Davon ist auszugehen, wenn sich innerhalb der Liste eine oder mehrere Variablen befinden, bei denen fehlende Werte nicht zufällig fehlen (Little/Rubin 2002: 41, Cielebak/Rässler, Kapitel 26 in diesem Band), sondern die Wahrscheinlichkeit, dass ein Wert fehlt, z.B. von eben diesem Wert abhängt. Das wäre z.B. der Fall, wenn es von der Einkommenshöhe abhängen würde, ob Umfrageteilnehmer auf Befragen ihr Einkommen angeben oder nicht. Von Vorteil ist daher, sich für statistische Berechnungen auf das jeweilige Maximum an verfügbaren empirischen Informationen stützen zu können. Dies ist z.B. unter Rückgriff auf moderne Statistiksoftware wie Mplus möglich, sofern wie dort Maximum-Likelihood-Schätzungen für unterschiedliche Muster fehlender Werte realisiert werden können (Muthén/Muthén 1998-2010: 7f.).

Eine grundsätzliche Alternative zu diesen beiden Ausschlussvarianten liegt darin, fehlende Werte durch inhaltlich plausible Werte zu ersetzen. Dieser Vorgang wird als *Imputation* bezeichnet (Cielebak/Rässler, Kapitel 26 in diesem Band). Da solche „inhaltlich plausiblen“ Werte aber nur auf der Basis von Modellen gefunden werden können und infolgedessen unterschiedliche Modelle auch zu unterschiedlichen plausiblen Werten führen können, kommt einer sorgfältigen Modellbildung große Bedeutung zu. Auch bei Imputationsverfahren existieren einige grundsätzliche Unterscheidungsmöglichkeiten.

Eine erste besteht darin, dass fehlende Werte durch *Schätzwerte oder durch beobachtete Werte ähnlich antwortender Fälle* ersetzt werden können. Wird ein fehlender Wert durch einen Schätzwert ersetzt, so könnte es sich bei diesem Wert im einfachsten Fall um das *arithmetische Mittel der Häufigkeitsverteilung derjenigen Variable handeln*, bei der die fehlenden Werte auftreten. Nehmen wir zur Illustration dieser Methode an, dass in einer Umfrage die Frage nach dem Nettoeinkommen (Y) von 30 Prozent des Samples nicht beantwortet wurde. Auf Basis der 70 Prozent des Samples, bei denen die individuellen Einkommensangaben vorliegen, könnte dann das Durchschnittseinkommen berechnet und dieser Wert dazu verwendet werden, bei den 30 Prozent der Befragten mit fehlenden Einkommensangaben diese fehlenden Werte zu ersetzen. Eine solche einfache Mittelwertersetzung ist allerdings problematisch, da sie individuell varierende (Einkommens-)Werte durch ein und denselben, konstanten (Mittel-) Wert ersetzt und dadurch die durchschnittliche Streuung der Häufigkeitsverteilung künstlich reduziert. Die 30 Prozent der Befragten würden im Beispiel alle denselben Einkommenswert, beispielsweise 2.400 Euro, zugewiesen bekommen. Ist man dann z.B. daran interessiert, über diese Variable eine Analyse der Einkommensungleichheit in der Gesellschaft durchzuführen, so würde diese Ungleichheit künstlich geringer ausfallen als sie es tatsächlich ist.

Abgemildert werden kann dieser Effekt dadurch, dass fehlende Werte nicht durch den unbedingten Mittelwert, sondern durch *bedingte* Mittelwerte ersetzt werden. Dies ist z.B. auf der Basis von Regressionsanalysen möglich. Erweitern wir unser Beispiel dazu um drei weitere Variablen. Nehmen wir also an, dass in besagter Umfrage die Frage nach dem Nettoeinkommen (Y) von 30 Prozent des Samples nicht beantwortet wurde, dass zugleich aber alle Befragten die Fragen nach der Bildung (X1), der beruflichen Stellung (X2) und der Geschlechtszugehörigkeit (X3) beantwortet haben. Für die 70% des Samples mit kompletter Information könnten dann z.B. über eine multiple Regression Einkommenswerte auf der Basis von Bildung, beruflicher Stellung und Geschlecht *geschätzt* werden (das sind genau die angesprochenen bedingten mittleren Einkommenswerte). Unter Rückgriff auf die geschätzten Effekte (Regressionskoeffizienten) dieser drei Variablen könnte dann bei den 30 Prozent des Samples mit fehlenden Einkommenswerten für jede individuelle Konstellation von Bildung, beruflicher Stellung und Geschlecht ein individueller Einkommenswert ermittelt und dieser individuelle Schätzwert dazu verwendet werden, den fehlenden Einkommenswert zu ersetzen. Auch hierbei würde die Streuung der Y-Werte noch künstlich reduziert werden, wenngleich nicht mehr so stark.

Um diesem Effekt entgegenzuwirken, besteht die Möglichkeit, zum Y-Schätzwert eine zufällige Komponente zu addieren und dann diese Summe aus jeweiligem Schätzwert und

Zufallskomponente zur Ersetzung eines fehlenden Wertes zu nutzen. Dazu kann beispielsweise jedem Y-Schätzwert per Zufallsprinzip ein Residuum aus der Verteilung der Residuen zugeordnet werden. Betrachten wir dazu wieder unser Einkommensbeispiel. Da bei einer Regressionsanalyse die geschätzten Y-Werte im Allgemeinen von den beobachteten Y-Werten abweichen werden, entstehen sogenannte Residuen. Diese ergeben sich jeweils als Differenz zwischen beobachtetem und geschätztem Y-Wert. Ein solches Residuum fällt bei jedem Fall an, so dass wir über alle 70% unseres Beispiel-Samples mit kompletter Information hinweg eine Verteilung individueller Residualwerte erhalten, aus der per Zufall Residualwerte gezogen und zu den jeweiligen geschätzten Einkommenswerten addiert werden können. Jede Person mit zuvor fehlenden Einkommenswerten erhält so zu ihrem geschätzten Einkommenswert wieder ein Residuum hinzu, nur gewissermaßen nicht ihr eigenes. Gleichwohl wird dadurch näherungsweise simuliert, was bei einer Regressionsanalyse strikt gelten würde: Und zwar, dass die Summe aus Y-Schätzwert und Residuum einen beobachteten Y-Wert ergibt. Ein zuvor fehlender Einkommenswert würde bei diesem Verfahren also durch einen Einkommenswert ersetzt werden, der einem beobachteten Einkommenswert zumindest nahekommen wird. Werden bei diesem Verfahren Residualwerte aus der empirischen Verteilung der Residuen gezogen, besteht alternativ aber auch die Möglichkeit, Residualwerte aus einer Wahrscheinlichkeitsverteilung, etwa einer Normalverteilung, zu ziehen. Bei Verwendung einer Statistiksoftware wie SPSS würden diese beiden Optionen standardmäßig angeboten werden.

Eine weitere Verfahrensmöglichkeit besteht darin, fehlende Werte gleich durch *beobachtete* Werte von ähnlich antwortenden Fällen zu ersetzen. Auch hierbei gibt es Varianten. Betrachten wir wieder unser Beispiel. Möglich wäre etwa, zwischen Fällen mit kompletter Information einerseits und Fällen mit fehlenden Einkommensangaben andererseits jeweils den Grad der Ähnlichkeit in Bildung und beruflicher Stellung über ein Distanzmaß zu ermitteln, um aus der Einkommensverteilung der Fälle mit jeweils geringster Distanz per Zufall Einkommenswerte zu ziehen, um damit fehlende Einkommenswerte zu ersetzen. Möglich wäre aber auch, Ähnlichkeit über den geschätzten Y-Wert einer Regressionsanalyse zu ermitteln, wie wir dies oben bei unserem Einkommensbeispiel beschrieben haben. Fälle wären dann per definitionem ähnlich, wenn sie ähnliche geschätzte Einkommenswerte aufweisen. Anstelle nun aber an die jeweils geschätzten Einkommenswerte Zufallskomponenten anzuhängen, um beobachteten Werten möglichst nahe zu kommen, kann auch gleich auf die zugehörigen beobachteten Einkommenswerte zurückgegriffen werden. Die Einkommenswerte können ja nur über eine Regression geschätzt werden, wenn zu jedem Fall auch ein beobachteter Einkommenswert vorgelegen hat. Also kann man nun für jeden Fall und seinen geschätzten Einkommenswert schauen, welcher beobachtete Einkommenswert vorgelegen hat. Über die Fälle mit jeweils gleichem geschätzten Einkommenswert hinweg betrachtet wird so die Verteilung der beobachteten Einkommenswerte sichtbar, aus der per Zufall Einkommenswerte gezogen und zur Ersetzung fehlender Einkommenswerte genutzt werden können. Dieses als „Predictive Mean Matching“ bezeichnete Verfahren ist in SPSS z.B. eine von zwei Standardoptionen in der Routine zur multiplen Imputation.

Schließlich wird bei Imputationsverfahren zwischen einfacher und multipler Imputation unterschieden (Cielebak/Rässler, Kapitel 26 in diesem Band). *Multiple Imputation* besagt dabei, dass ein fehlender Wert nicht nur einmal, sondern parallel mehrfach, z.B. fünfmal, durch inhaltlich plausible Werte ersetzt wird. Es entstehen dann fünf parallele Datensätze, die im Zuge von statistischen Auswertungen wieder zusammengeführt werden. Dafür kann in Statistikprogrammen wie SPSS, Stata, R oder Mplus auf eigens dafür entwickelte Routinen zurückgegriffen werden. Für einen weitergehenden Überblick über Missing-Data Techniken sowie einen vertieften Einblick in einzelne Techniken zur Analyse von Daten mit fehlenden Werten verweisen wir auf Spieß (2008).

## Literatur

- Allison, Paul D. (2002): Missing Data. Thousand Oaks: Sage
- American Association for Public Opinion Research (2011): Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition: AAPOR
- Bartsch, Simone (2012): ... würden Sie mir dazu Ihre E-Mail-Adresse verraten? Internetnutzung und Nonresponse beim Aufbau eines Online Access Panels. Baden Baden: Nomos
- Bethlehem, Jelke/Cobben, Fannie/Schouten, Barry (2011): Handbook of Nonresponse in Household Surveys. Hoboken: John Wiley
- Converse, Patrick D./Wolfe, Edward W./Huang, Xiaoting/Oswald, Frederick L. (2008): Response Rates for Mixed-Mode Surveys Using Mail and E-Mail/Web. In: American Journal of Evaluation 29: 99-107
- Couper, Mick P./Groves, Robert M. (1996): Household-Level Determinants of Survey Nonresponse. In: New Directions in Evaluation 70: 63-79
- De Leeuw, Edith D./de Heer, Wim (2002): Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In: Groves et al. (Hg.): 41-54
- De Leeuw, Edith D./Hox, Joop/Huisman, Mark (2003): Prevention and Treatment of Item Nonresponse. In: Journal of Official Statistics 19: 153-176
- Dillman, Don A./D. Smyth, Jolene /Leah, M. Christian (2009): The Tailored Design Method. New York: John Wiley
- Engel, Uwe/Schnabel, Christiane (2004): Markt- und Sozialforschung. Metaanalyse zum Ausschöpfungsgrad. <https://www.adm-ev.de/index.php?id=forschungsprojekte>
- Engel, Uwe/Pötschke, Manuela/Schnabel, Christiane/Simonson, Julia (2004): Nonresponse und Stichprobenqualität. Frankfurt am Main: Verlagsgruppe Deutscher Fachverlag
- Engel, Uwe (2006): Anreizeffekte in Studien der Markt- und Sozialforschung. [http://www.sozialforschung.uni-bremen.de/Engel\\_Anreizeffekte.pdf](http://www.sozialforschung.uni-bremen.de/Engel_Anreizeffekte.pdf)
- Engel, Uwe/Bartsch, Simone/Schnabel, Christiane/Vehre, Helen (2012): Wissenschaftliche Umfragen. Methoden und Fehlerquellen. Frankfurt/New York: Campus
- Engel, Uwe (2013): Access Panel and Mixed-Mode Internet Survey. PPSM Panel Report. [http://www.sozialforschung.uni-bremen.de/PPSM\\_Panel\\_Report\\_Online.pdf](http://www.sozialforschung.uni-bremen.de/PPSM_Panel_Report_Online.pdf)
- Engel, Uwe/Schmidt, Björn Oliver (2013): Ein Feldexperiment zur Wirkung von Befragungsanreizen innerhalb eines Mixed-Mode Designs. <http://www.sozialforschung.uni-bremen.de/html/downloads.html>
- Engel, Uwe/Köster, Britta (2014): Response Effects and Cognitive Involvement in Answering Survey Questions. Forthcoming in: Engel et al. (Hg.)
- Engel, Uwe/Jann, Ben/Lynn, Peter/Scherpenzeel, Annette/Sturgis, Patrick (Hg) (2014): Improving Survey Methods. New York: Routledge
- Essig, Lothar/Winter, Joachim (2003): Item Nonresponse to Financial Questions in Household Surveys: An Experimental Study of Interviewer and Mode Effects. University of Mannheim: Mannheim Research Institute for the Economics of Aging (MEA)

- Fan, Weimiao/Yan, Zheng (2010): Factors Affecting Response Rates of Web Survey: A Systematic Review. In: Computers in Human Behavior 26: 132-140
- Groves, Robert M./Couper, Mick P. (1998): Nonresponse in Household Interview Surveys. New York: John Wiley
- Groves, Robert M./Singer, Eleanor/Corning, Amy (2000): Leverage – Saliency Theory of Survey Participation. Description and an Illustration. In: Public Opinion Quarterly 64: 299-308
- Groves, Robert M./Dillman, Don A./Eltinge, John L./Little, Roderick J.A. (Hg.) (2002): Survey Nonresponse. New York: John Wiley
- Haarmann, Alexander/Scholz, Evi/Wasmer, Martina/Blohm, Michael/Harkness, Janet (2006): Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 2004. ZUMA Methodenbericht 06
- Häder, Sabine/Gabler, Siegfried/Heckel, Christiane (2009): Stichprobenziehung für die CELLA-Studie. In: Häder/Häder (Hg.): 21-49
- Häder, Michael/Häder, Sabine (Hg.) (2009): Telefonbefragungen über das Mobilfunknetz. Wiesbaden: VS Verlag
- Häder, Sabine/Häder, Michael/Kühne, Mike (Hg.) (2012): Telephone Surveys in Europe. Research and Practice. Heidelberg: Springer
- Hox, Joop/de Leeuw, Edith (2002): The Influence of Interviewers' Attitude and Behavior on Household Survey Nonresponse: An International Comparison. In: Groves et al. (Hg.): 103-120
- Krosnick, Jon A./Alwin, Duane F. (1987): An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurements. In: Public Opinion Quarterly 51: 201-219
- Little, Roderick J.A./Rubin, Donald B. (2002): Statistical Analysis with Missing Data. 2nd ed. Hoboken: John Wiley
- Muthén, Linda K./Muthén, Bengt O. (1998-2010). Mplus User's Guide. 6th ed. Los Angeles, CA: Muthén & Muthén
- Schnell, Rainer (2012): Survey-Interviews. Methoden standardisierter Befragungen. Wiesbaden: VS Verlag
- Shoemaker, Pamela J./Eichholz, Martin/Skewes, Elizabeth A. (2000): Item Nonresponse: Distinguishing Between don't Know and Refuse. In: International Journal of Public Opinion Research 14: 193-201
- Spieß, Martin (2008): Missing-Data Techniken. Analyse von Daten mit fehlenden Werten. Hamburg: LIT Verlag
- Stoop, Ineke/Billiet, Jaak/Koch, Achim/Fitzgerald, Rory (2010): Improving Survey Response. Lessons Learned from the European Social Survey. Chichester: John Wiley

Hans Kiesl

## 24.1 Einführung

Ziel einer Analyse quantitativer Daten ist die Verallgemeinerung der Stichprobenergebnisse auf die interessierende Grundgesamtheit (Häder/Häder, Kapitel 18 in diesem Band). Tatsächlich unterscheiden sich Stichproben in bestimmter Hinsicht aber fast immer von der Grundgesamtheit; sei es durch ein geplantes „Oversampling“ einer bestimmten Teilstichprobe (die Genauigkeit einer Schätzung wird im Wesentlichen von der Fallzahl in der Stichprobe bestimmt, weshalb seltene Teilpopulationen, für die valide Schätzungen möglich sein sollen, mit einem größeren Auswahlsatz in die Erhebung aufgenommen werden) oder durch selektiven Nonresponse (Engel/Schmidt, Kapitel 23 in diesem Band). Viele Befragungen weisen etwa einen so genannten „Mittelschichtsbias“ auf; Personen mit mittlerem bis gehobenem Bildungsniveau (gemessen durch den höchsten Schulabschluss) zeigen sich am öftesten bereit, an Umfragen teilzunehmen, sie sind in den Erhebungsdaten daher überrepräsentiert.

Ob man mit den Stichprobendaten deskriptive Berechnungen durchführen will (z.B. Schätzungen von Durchschnittseinkommen oder Armutsrisikoquoten) oder Modellparameter schätzt (z.B. Regressionskoeffizienten): die üblichen Berechnungen setzen meistens eine *einfache Zufallsstichprobe* (Häder/Häder, Kapitel 18 in diesem Band) voraus und liefern unter dieser Annahme unverzerrte Punktschätzungen, unverzerrt geschätzte Standardfehler und korrekte p-Werte. Wenn keine einfache Stichprobe vorliegt, müssen die Verfahren ggf. modifiziert werden. *Gewichtung ist eine Möglichkeit, auch bei komplexen Auswahlverfahren und selektivem Nonresponse unverzerrte Schätzungen zu erhalten.* Vereinfacht gesagt, werden Stichprobeneinheiten aus unter- bzw. überrepräsentierten Gruppen in den Berechnungen höher bzw. niedriger gewichtet. Im Folgenden wird diskutiert, welche Arten von *Gewichten* (auch: *Hochrechnungsfaktoren*) es gibt, wie sie berechnet werden und wann und wie sie bei Schätzungen verwendet werden sollten.

## 24.2 Designgewichte

Das gewählte Zufallsstichprobenverfahren (Häder/Häder, Kapitel 18 in diesem Band) determiniert für jede Einheit der Grundgesamtheit eine Wahrscheinlichkeit, in die Stichprobe zu gelangen, die so genannte *Auswahlwahrscheinlichkeit*. Bei einer einfachen Zufallsauswahl besitzt jede Einheit der Grundgesamtheit die gleiche Wahrscheinlichkeit, in die Stichprobe aufgenommen zu werden, bei komplizierteren Designs können verschiedene Einheiten unterschiedliche Auswahlwahrscheinlichkeiten besitzen. So haben Haushalte, die unter vielen Telefonnummern erreichbar sind, bei telefonischen Befragungen durch Random Digit Dialling (Hüfken; Kapitel 46 in diesem Band) eine größere Auswahlwahrscheinlichkeit; wird in jedem erreichten Haushalt genau eine Person ausgewählt, haben Personen in großen Haushalten eine kleinere Auswahlwahrscheinlichkeit. Im Rahmen der Panelbefragung (Schupp, Kapitel 73 in diesem Band) PASS („Panel Arbeitsmarkt und soziale Sicherung“; Trappmann et al. 2010) wurde die Hälfte der befragten Haushalte aus dem Register der Arbeitslosengeld-II-Empfänger gezogen (*geplantes Oversampling*); diese hatten daher im Vergleich zu anderen Haushalten eine größere Auswahlwahrscheinlichkeit.

Einheiten mit größerer Auswahlwahrscheinlichkeit sind in der Stichprobe tendenziell überrepräsentiert. Die Kehrwerte der Auswahlwahrscheinlichkeiten heißen *Designgewichte* und werden dazu verwendet, diese Überrepräsentation bei Schätzungen auszugleichen.

Zunächst sei das Problem betrachtet, einen Mittelwert (z.B. das Durchschnittseinkommen oder den Wähleranteil einer Partei) aus Stichprobendaten zu schätzen.  $N$  bezeichne die Größe der Population,  $n$  den Umfang der Stichprobe, und  $m_x$  sei der Mittelwert eines Merkmals  $X$  in der Gesamtheit. Für  $i = 1, \dots, n$  sei  $x_i$  der Wert des Merkmals  $X$  bei der  $i$ -ten Stichprobeneinheit,  $\pi_i$  die zugehörige Auswahlwahrscheinlichkeit und  $d_i = 1/\pi_i$  das Designgewicht. Die *ungewichtete Schätzung* für  $m_x$  ist das arithmetische Mittel der Stichprobenwerte:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i .$$

Bei unterschiedlichen Auswahlwahrscheinlichkeiten ist diese Schätzung aber im Allgemeinen nicht erwartungstreue. (Ein Schätzer heißt *erwartungstreu* oder *unverzerrt*, wenn der Erwartungswert der Schätzung über alle möglichen Stichproben, die gezogen werden können, dem zu schätzenden Wert entspricht; andernfalls heißt der Schätzer *verzerrt*.) Mit Hilfe der Designgewichte lässt sich aber ein *gewichteter Schätzer* konstruieren, der immer erwartungstreue ist:

$$\bar{x}^g = \frac{1}{N} \sum_{i=1}^n d_i x_i .$$

Offensichtlich werden hier die Stichprobenwerte umso höher gewichtet, je größer die Designgewichte sind. Während die gewichtete Schätzung immer unverzerrt ist, ist die ungewichtete Schätzung nur dann erwartungstreue, wenn das zu schätzende Merkmal und die Auswahlwahrscheinlichkeiten unkorreliert sind, also (etwas vereinfacht) wenn sich

bestimmte Populationsgruppen, die per Design in der Stichprobe über- oder unterrepräsentiert sind, hinsichtlich des Merkmals X nicht vom Rest der Population unterscheiden. Diese Annahme ist im Allgemeinen nicht überprüfbar, weshalb gewichtete Schätzungen von Mittelwerten (und auch von anderen Verteilungskennziffern wie Quantilen oder Varianzen) den ungewichteten vorzuziehen sind.

Wenn die Designgewichte sehr unterschiedlich sind, können die Standardfehler der gewichteten Schätzungen groß werden; man wird daher bei der Stichprobenziehung versuchen, große Unterschiede bei den Auswahlwahrscheinlichkeiten zu vermeiden.

Sind alle Auswahlwahrscheinlichkeiten (also auch alle Designgewichte) gleich groß, nennt man die Stichprobe *selbstgewichtet*; dies ist z.B. bei einer einfachen Stichprobe oder bei einer geschichteten Stichprobe mit proportionaler Aufteilung (Häder/Häder, Kapitel 18 in diesem Band) der Fall. Von standardisierten Gewichten wird gesprochen, wenn die Gewichte so reskaliert werden, dass ihre Summe dem Stichprobenumfang entspricht.

### 24.3 Nonresponse-Gewichtung

Jede sozialwissenschaftliche Erhebung leidet unter Unit Nonresponse, d.h. Personen, die Elemente der Stichprobe sind, können nicht befragt werden, weil sie nicht erreichbar sind oder sich einer Teilnahme verweigern (Engel/Schmidt, Kapitel 23 in diesem Band). Wenn sich Stichprobeneinheiten, die nicht an der Erhebung teilnehmen, von den Teilnehmern unterscheiden, sind Schätzer, die den Nonresponse nicht berücksichtigen, verzerrt. Nonresponse-Gewichtung ist eine Möglichkeit, diese Verzerrungen auszugleichen.

Zu diesem Zweck stellt man sich den Nonresponse als einen Zufallsprozess vor; man nimmt also an, dass alle Stichprobeneinheiten eine (von Einheit zu Einheit unterschiedliche) Wahrscheinlichkeit haben, erreicht und zur Teilnahme überredet zu werden, falls sie in die Stichprobe gelangen. Diese Sichtweise entspricht einer zweistufigen Stichprobenprozedur: in der ersten Stufe werden die Stichprobeneinheiten ausgewählt (Bruttostichprobe), in der zweiten Stufe nimmt nur ein Teil der Stichprobeneinheiten tatsächlich an der Erhebung teil (Nettostichprobe).

Sei  $n_r$  der Umfang der Nettostichprobe, und sei  $\pi_i$  wieder die Auswahlwahrscheinlichkeit der  $i$ -ten Stichprobeneinheit. Wenn mit  $\pi_i^f$  die Wahrscheinlichkeit bezeichnet wird, dass Einheit  $i$  auch an der Erhebung teilnimmt, falls sie ausgewählt wird (Antwortwahrscheinlichkeit), ist das Produkt  $\pi_i \cdot \pi_i^f$  die Wahrscheinlichkeit, dass eine Einheit  $i$  in der Nettostichprobe enthalten ist. Sei  $w_i$  der Kehrwert dieser Wahrscheinlichkeit, dann kann analog zu Abschnitt 24.2 folgender unverzerrter, gewichteter Schätzer für den Mittelwert definiert werden:

$$\bar{x}^{g,nr} = \frac{1}{N} \sum_{i=1}^{n_r} w_i x_i .$$

Das Problem in der Praxis besteht darin, dass die Antwortwahrscheinlichkeiten  $\pi_i^r$  nicht bekannt sind und aus der Stichprobe geschätzt werden müssen; einen guten Überblick über geeignete Schätztechniken gibt Brick (2013).

Im Wesentlichen kann man zwei Verfahren der Nonresponse-Gewichtung unterscheiden:

- das Bilden von Gruppen, innerhalb derer man gleiche Antwortwahrscheinlichkeiten für alle Einheiten vermutet, und
- die Schätzung eines expliziten Modells für die Antwortwahrscheinlichkeiten (z.B. eines Logit- oder Probit-Modells; Little 1986).

In beiden Fällen müssen für alle Einheiten der Bruttostichprobe, also auch für die Nichtteilnehmer, alle Variablen bekannt sein, die die Antwortwahrscheinlichkeit (also Erreichbarkeit bzw. Teilnahmeentscheidung) beeinflussen (im ersten Fall die Variablen, die die Gruppenzugehörigkeit bestimmen, im zweiten Fall die Kovariablen des Modells). Diese starke Annahme ist in der Praxis allenfalls näherungsweise erfüllt, zumal oft wenig Informationen über Nichtteilnehmer vorliegen: Bei Einwohnermeldestichproben sind z.B. für die Nichtteilnehmer nur die Informationen aus dem Register verfügbar (u.a. Geschlecht, Alter und Adresse), bei Telefonbefragungen im Festnetz gibt es lediglich Regionalinformationen (Vorwahl). Besser sieht es bei Panelbefragungen aus; hier können alle Informationen aus dem Interview der jeweiligen Vorwelle verwendet werden. In allen Fällen gilt: Nonresponse-korrigierte Gewichte können Verzerrungen, die durch selektiven Nonresponse (Engel/Schmidt, Kapitel 23 in diesem Band) verursacht sind, nur dann vollständig ausgleichen, wenn alle Faktoren bekannt sind, die die Teilnahmeentscheidung beeinflussen. In der Praxis hofft man darauf, dass die Nonresponse-Korrektur etwaige Verzerrungen zumindest verringert.

---

## 24.4 Kalibrierung

Nach Design- und Nonresponse-Gewichtung schließt sich häufig noch ein dritter Gewichtungsschritt an, die so genannte Kalibrierung (auch Redressment oder Benchmarking genannt). Kenngrößen bestimmter Variablen der Zielpopulation sind oft aus externen Quellen bekannt, z.B. soziodemographische Merkmale wie Geschlecht, Alter, Erwerbsstatus und höchster Schulabschluss, deren Verteilung aus dem Mikrozensus bekannt ist. Mit Hilfe der Design- und/oder Nonresponse-Gewichte können die Verteilungen dieser Variablen auch aus der Stichprobe geschätzt werden. Wahre und geschätzte Verteilungen werden sich im Allgemeinen unterscheiden, entweder allein durch Zufallseinfluss oder (realistischer) durch selektiven Nonresponse, den die explizite Nonresponse-Gewichtung nicht ausgleichen konnte. Eine naheliegende Idee ist nun, die Gewichte der Stichprobeneinheiten etwas (aber möglichst wenig) so zu verändern, dass die Schätzungen für diese Variablen (in diesem Zusammenhang auch Eckwerte oder Benchmarks genannt) mit den

**Tab. 24.1** Designgewichte vs. kalibrierte Gewichte am Beispiel fiktiver Umfragewerte

Person	Geschlecht (0 = m, 1 = w)	persönliches Einkommen	Designgewicht	an Geschlecht kalibriertes Gewicht
1	0	2600	125	167
2	1	700	125	100
3	1	1800	125	100
4	0	1500	125	167
5	1	1000	125	100
6	1	2000	125	100
7	0	1800	125	167
8	1	600	125	100

veränderten Gewichten exakt mit den aus externen Quellen bekannten Werten übereinstimmen; die neuen Gewichte nennt man dann kalibriert.

Das folgende fiktive Beispiel soll die Idee und das Vorgehen illustrieren. Aus einer Bevölkerungsgruppe (Umfang 1.000 Personen) sei eine einfache Stichprobe vom Umfang 20 gezogen worden; nur acht Personen waren aber bereit, an der Befragung teilzunehmen.

Ohne weitere Information über die Nicht-Teilnehmer wird jede Person das Gewicht  $1.000/8 = 125$  erhalten. Das Durchschnittseinkommen in der Stichprobe beträgt dann  $12.000/8 = 1.500$  €. Nehmen wir nun an, wir wissen aus externer Quelle, dass es in dieser Bevölkerungsgruppe genauso viele Männer wie Frauen gibt. Die Stichprobe enthält aber einen höheren Frauenanteil (62,5%), d.h. Frauen sind in der Stichprobe überrepräsentiert. Das könnte man ausgleichen, indem man die Designgewichte der Frauen mit dem Verhältnis  $50%/62,5\% (\times 125)$  und die Designgewichte der Männer mit dem Verhältnis  $50%/37,5\% (\times 125)$  multipliziert (siehe kalibrierte Gewichte in der Tabelle). Der mit den kalibrierten Gewichten geschätzte Frauenanteil beträgt nun 50% (entspricht also dem wahren Anteil), die gewichtete Schätzung des Durchschnittseinkommens hat sich auf knapp 1.600 € ( $(167 \times 2.600 + 100 \times 700 + \dots + 100 \times 600)/1.000 = 1.595,30$ ) erhöht. Man sieht den Daten an, dass es einen Zusammenhang zwischen Einkommen und Geschlecht gibt, also ist die höhere gewichtete Schätzung sicher plausibel. Ganz ähnlich könnte man Gewichte erzeugen, die an zusätzliche Variablen kalibriert sind, also z.B. an bekannte sozioökonomische Verteilungen wie Schulabschluss, Erwerbsstatus oder Alter.

Welchen Sinn macht eine Kalibrierung? Wenn es *keinen Nonresponse* gibt, gleicht die Kalibrierung lediglich Zufallsschwankungen bei den betroffenen Merkmalen aus. Dies führt zu kleineren Schätzfehlern auch bei anderen Merkmalen, falls diese mit den Benchmarks korreliert sind, ändert aber nichts an der Unverzerrtheit der Schätzungen.

Für den realistischeren Fall, dass die Erhebung unter *Nonresponse* leidet, der unter anderem mit den Eckwert-Variablen (im obigen Beispiel das Geschlecht) zusammenhängt, gleicht eine Kalibrierung Verzerrungen hinsichtlich der Eckwert-Variablen und (teilweise)

hinsichtlich anderer Variablen aus, die mit den Eckwert-Variablen korreliert sind (im obigen Beispiel das Einkommen). Somit ist die Kalibrierung eine Ergänzung oder Alternative zur expliziten Nonresponse-Gewichtung (im Gegensatz dazu müssen bei der Kalibrierung keine Informationen über die Nichtteilnehmer vorliegen; ausführlich siehe: Särndal/Lundström 2005). Ein vollständiger Ausgleich einer durch Nonresponse verursachten Verzerrung ist aber durch Kalibrierung nur gewährleistet, wenn die Antwortwahrscheinlichkeit ausschließlich von den Benchmarks abhängt, eine Annahme, die nicht überprüft werden kann. Werden sehr viele Benchmarks verwendet, erhöht sich tendenziell wieder die Varianz der Gewichte, was zu größeren Standardfehlern bei Schätzungen führt. Eine Kalibrierung an Eckwerten, die den Nonresponse nur wenig erklären, kann also trügerisch sein, wenn der Anwender fälschlich glaubt, damit das Nonresponse-Problem gelöst zu haben, während in Wahrheit nur die Varianz vergrößert wird. Kalibrierung an allzu vielen Variablen wird in der Literatur daher auch kritisch gesehen (Schnell 1997: 245f). Als *Faustregel für die Kalibrierung* gilt: so viele Eckwerte wie nötig, um Verzerrungen durch Nonresponse auszugleichen, aber so wenige wie möglich, um die Standardfehler zu begrenzen.

---

## 24.5 Braucht man Gewichte auch bei Schätzung von Modellparametern?

Die Begründung für eine Gewichtung wurde in den vorangegangenen Abschnitten nur für eine Schätzung von Total- und Mittelwerten gegeben. Grundsätzlich gilt aber für jede Schätzung aus Stichprobendaten, dass unterschiedliche Inklusionswahrscheinlichkeiten und/oder Unit Nonresponse (wie auch das gesamte Stichprobendesign) berücksichtigt werden sollten, weil sonst verzerrte Schätzungen und/oder (bei statistischen Tests) falsche p-Werte resultieren können. Inwieweit die Gewichte dabei am effizientesten berücksichtigt werden, hängt vom statistischen Verfahren ab; bei Verwendung geeigneter Software muss sich der Anwender mit den statistischen Details nicht beschäftigen.

Die verbreitete Meinung, dass bei multivariaten Modellschätzungen – wie z.B. Regressions (Blasius/Baur, Kühnel, Kapitel 79 und 80 in diesem Band) oder Mehrebenenmodellen (Pötschke, Kapitel 87 in diesem Band) – die Berücksichtigung von Gewichten unnötig ist, ist in dieser Allgemeinheit falsch. Wenn alle Variablen, die für unterschiedliche Auswahlwahrscheinlichkeiten und/oder für selektiven Nonresponse verantwortlich sind, als Kovariablen im Modell aufgenommen sind, ist eine Berücksichtigung von Gewichten tatsächlich unnötig, in anderen Fällen können ungewichtete Schätzungen aber verzerrt sein (eine Simulationsstudie hierzu findet sich in Arzheimer 2009).

Typischerweise führt eine Verwendung von Gewichten bei der Schätzung von Modellparametern zu größeren Standardfehlern, weshalb Anwender oft vor einer gewichteten Schätzung zurückgeschrecken. Das allein sollte aber keine Begründung für eine ungewichtete Schätzung sein. Es ist im Einzelfall zu entscheiden, ob die Gewichte Informationen zum Nonresponse berücksichtigen, die nicht in den Kovariablen des Modells enthalten sind. Als Faustregel darf gelten, dass bei annähernd gleichen Schätzergebnissen die

ungewichtete Schätzung (wegen kleinerer Standardfehler) vorzuziehen ist, während bei Unterschieden in den Schätzungen diese genauer analysiert werden sollten (und bei plausiblem Nonresponse-Modell gewichtete Schätzungen normalerweise weniger verzerrt sind). Eine umfassende Diskussion pro und contra Gewichte bei Schätzung von Modellparametern findet sich bei Gelman (2007). Die (nicht triviale) Verwendung von Gewichten bei Mehrebenenmodellen diskutieren Rabe-Hesketh/Skrondal (2006).

---

## 24.6 Software

Mittlerweile stehen unterschiedliche Softwarepakete zur Verfügung, die es Anwendern erlauben, Gewichte in die Analyse von Stichprobendaten zu integrieren, ohne sich mit den entsprechenden Formeln befassen zu müssen, darunter Stata, R und SPSS:

- *Stata* bietet hierfür die svy-Kommandos an (svy für „survey“). Neben dem Stichprobendesign (Schichtung, Klumpung etc.) ist die Gewichtungsvariable als so genanntes „pweight“ (für „probability weight“) zu spezifizieren.
- Die Statistik-Software *R* stellt verschiedene Zusatzpakete zur Auswertung von Stichprobendaten unter Verwendung von Gewichtungsfaktoren zur Verfügung, darunter das Paket „survey“.
- *SPSS* erlaubt die korrekte Berücksichtigung von Gewichten nur in dem Zusatzpaket „Complex Samples“.

Zu beachten ist, dass manche Softwarepakete auch andere Arten von Gewichten kennen. In SPSS gibt es z.B. den Befehl „WEIGHT BY“, in Stata korrespondiert dazu „fweight“ (kurz für „frequency weight“). Bei Verwendung dieser Syntax werden die Gewichte als Häufigkeiten interpretiert, d.h. es wird angenommen, dass die jeweilige Beobachtung so oft im Datensatz vorkommt, wie das Gewicht anzeigt (ein Gewicht von 10 würde hier so interpretiert, dass es zehn identische Einheiten in der Stichprobe gibt). Dass dies eine falsche Verarbeitung von Stichprobendaten ist, zeigt sich z.B. in viel zu kleinen Standardfehlern. Als Ausweg wird bisweilen die Verwendung von standardisierten Gewichten empfohlen; die Standardfehler werden dann realistischer, berücksichtigen aber das Stichprobendesign nicht vollständig. Schätzungen mit Stichprobendaten sollten daher immer mit spezialisierter Software erfolgen, z.B. den oben genannten Programmen.

## Literatur

- Arzheimer, Kai (2009): Gewichtungsvariation. In: Schoen et al. (Hg.): 361-388
- Brick, J. Michael (2013): Unit Nonresponse and Weighting Adjustments: A Critical Review. In: Journal of Official Statistics 29: 329-374 (mit Diskussion)
- Gelman, Andrew (2007): Struggles with Survey Weighting and Regression Modelling. In: Statistical Science 22: 153-164
- Little, Roderick J. A. (1986): Survey Nonresponse Adjustments for Estimates of Means. In: International Statistical Review 54: 139-157
- Rabe-Hesketh, Sophia/Skrondal, Anders (2006): Multilevel Modeling of Complex Survey Data. In: Journal of the Royal Statistical Society A 169: 805-827
- Särndal, Carl.-E./Lundström, Sixten (2005): Estimation in Surveys with Nonresponse. Hoboken: Wiley
- Schnell, Rainer (1997): Nonresponse in Bevölkerungsumfragen. Opladen: Leske + Budrich
- Schoen, Harald/Rattinger, Hans/Gabriel, Oscar (Hg.) (2009): Vom Interview zur Analyse. Baden-Baden: Nomos
- Trappmann, Mark/Gundert, Stefanie/Wenzig, Claudia/Gebhardt, Daniel (2010): PASS: a Household Panel Survey for Research on Unemployment and Poverty. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften 130: 609-622

Barbara Felderer, Alexandra Birg und Frauke Kreuter

---

## 25.1 Was sind Paradaten?

Bei allen Umfragen (Reinecke, Kapitel 44 in diesem Band) entstehen neben den eigentlichen Umfragedaten auch Daten durch den Erhebungsprozess. Dazu gehören z.B. bei computergestützten Erhebungen Zeitmarken bei der Eingabe von Antworten, mit der die Dauer für die Beantwortung einer Frage bestimmt werden kann, oder automatisch aufgenommene Maus-Bewegungen in einem Websurvey. In einem Vortrag auf dem Joint Statistical Meeting der Amerikanischen Statistischen Gesellschaft 1998 in Dallas prägte Mick Couper für diese Art von Daten den Begriff *Paradaten* (Couper 1998). Solche Paradaten erlauben eine Prozess- und Qualitätskontrolle für die Durchführung von Befragungen (Weichbold, Krebs/Menold, Faulbaum, Engel/Schmidt, Kapitel 19, 30, 25 und 23 in diesem Band).

Nicht zu verwechseln sind Paradaten mit Metadaten, die als Daten über Daten (z.B. Informationen über Variablen-Bezeichnungen oder Werte-Bezeichnungen) beschrieben werden (Kreuter 2013a).

---

## 25.2 Die Erhebung von Paradaten

Paradaten können in allen Phasen des Erhebungsprozesses entstehen. Beim Rekrutieren von Befragten werden Interviewer häufig damit beauftragt *Kontaktprotokolle* zu erstellen, also Notizen über jeden Rekrutierungsversuch (Glantz/Michael, Kapitel 21 in diesem Band). Diese beinhalten z.B. Angaben über den Zeitpunkt des Kontaktversuchs und dessen Ergebnis, ob die zu befragende Person angetroffen wurde, ob sie am Interview teilgenommen hat, ob ein Termin vereinbart wurde, oder ob die Person ein Interview generell verweigert (Schnell, Hill, Esser 2013).

Zusätzlich werden Interviewer mitunter gebeten eigene *Beobachtungen der Nachbarschaft oder des zu befragenden Haushalts* durchzuführen, zum Beispiel über das Vorhandensein von Müll in der Umgebung, Einschätzungen des sozialen Status des Haushalts oder darüber, ob in dem Haushalt Kinder leben (West/Sinibaldi 2013).

Bei der Durchführung der Befragung können je nach Befragungsmodus unterschiedliche Paradaten anfallen (Olson/Parkhurst 2013 für einen Überblick). Bei persönlichen und telefonischen Befragungen können Paradaten über die Interaktion von Interviewern und Befragten erhoben werden sowie Informationen über die Interviewsituation, wie die Anwesenheit Dritter oder die Einschätzung des Interviewers über Verständnisschwierigkeiten des Befragten. Außerdem entstehen Paradaten über den Frage-Antwortprozess, beispielsweise Zeitmarken, Tastenanschläge und Zurückgehen im Fragebogen. Diese können auch in Websurveys (Wagner/Hering, Kapitel 48 in diesem Band) erhoben werden, hier werden sie allerdings vom Befragten selbst und nicht vom Interviewer erzeugt.

Auch wenn Paradaten automatische Nebenprodukte von computergestützten Datenerhebungs-Systemen sein können, müssen Forscher schon vor Beginn der Studie entscheiden, ob und welche dieser Informationen sie überhaupt sammeln und auswerten wollen. Es ist dabei wichtig herauszufiltern welche Daten für die eigene Forschung relevant sind und wie diese analysiert werden sollen. Nicht alle Befragungssysteme speichern Tastenanschläge oder Antwortzeiten automatisch, einige müssen vorher entsprechend programmiert werden. Eine Herausforderung dabei ist, dass nicht alle Softwareprogramme Paradaten „gebrauchsfertig“ erfassen können, und somit Benutzer-generierte Programme entwickelt werden müssen, um die Aufzeichnung von Paradaten zu unterstützen. Ein Beispiel hierfür ist das Paradatenskript von Lars Kaczmirek zur Aufdeckung von Antwortverhalten (Kaczmirek 2013, Callegaro 2013 für einen Überblick). Will man mit Hilfe von Eye-Tracking Studien festhalten, worauf sich die Augen des Befragten auf einer Bildschirmseite fixieren und für wie lange sind spezielle Geräte und entsprechende Software notwendig. Paradaten, wie z.B. Interviewer-Beobachtungen über ausgewählte Merkmale und Fälle müssen ebenfalls vor Beginn der Feldphase exakt bestimmt werden, und Interviewer müssen für die Erhebung dieser Daten geschult werden.

---

### **25.3 Möglicher Erkenntnisgewinn durch Paradaten**

Um zuverlässige Analysen auf Basis von Umfragebedaten durchführen zu können ist es wichtig, dass diese eine hohe Qualität aufweisen. Allerdings ist es nicht immer einfach sicherzustellen, dass die Befragten die Gesamtpopulation repräsentativ vertreten oder dass die Variablen fehlerfrei erhoben werden. So kann eine selektive Teilnahme an der Befragung zu verzerrten Schätzungen führen. Auch Messfehler bei der Datenerhebung können die Güte der Schätzungen stark beeinträchtigen. Ineffiziente Datenerhebungsprozesse können zudem unnötige Kosten verursachen und die gewünschte Fallzahl kompromittieren.

Ziel der Survey-Forschung ist es deshalb, alle Elemente einer Umfrage zu optimieren. Dazu gehören u.a. die Stichprobenziehung, die Rekrutierung von Befragten, die Fragen

selbst (Häder/Häder, Hlawatsch/Krickl, Porst, Franzen, Kapitel 18, 20, 50 und 51 in diesem Band), Daten-Eingabemasken bei Online-Befragungen (Wagner/Hering, Kapitel 48 in diesem Band) bzw. Fragebogenaufbau (Klöckner/Friedrichs, Kapitel 49 in diesem Band), Erhebungs-Prozesse und Gewicht-Anpassungs-Prozesse (Kiesl, Kapitel 24 in diesem Band).

Zur Identifikation möglicher Fehler in den verschiedenen Stufen des Umfrage-Produktionsprozesses werden Experimente (Eifler, Kapitel 11 in diesem Band), Simulations-techniken (Jann/Hevenstone, Kapitel 12 in diesem Band) und mitunter auch qualitative Methoden verwendet (Akremi, Kapitel 17 in diesem Band). Paradaten sind ein weiteres Instrument in dieser Reihe. Werden Paradaten während der Datenerhebung prozessbegleitend erfasst und analysiert, können Probleme direkt aufgedeckt werden, und im Idealfall kann der Datenerhebungsprozess direkt verbessert werden. Werden Paradaten im Anschluss an eine Datenerhebung analysiert, dient dies in der Regel der Verbesserung zukünftiger Prozesse oder der statistischen Korrektur der Schätzungen.

### 25.3.1 Paradaten zur Coverage-Forschung

Die Coverage-Forschung untersucht die Qualität der Auswahlgrundlagen (Häder/Häder, Kapitel 18 in diesem Band) aus denen die Samples gezogen werden, und den Einfluss von Fehlern in den Auswahlgrundlagen auf die Umfragedaten. Es gibt viele verschiedene Typen von Auswahlgrundlagen, z.B. Haushalts-Frames, Telefon-Frames, Haushaltmitglieder-Listen, Populations-Register oder Frames für Websurveys, dementsprechend unterschiedlich sind die Paradaten, die bei der Erstellung der Frames erzeugt werden (Eckman 2013 für einen Überblick).

Wenn Interviewer an der Erstellung der Auswahlgrundlagen beteiligt sind, wie beispielsweise bei einem Random-Route-Verfahren, bei dem die Adressen während einer Begehung aufgelistet werden (Häder/Häder, Kapitel 18 in diesem Band), kann zum Beispiel mit Hilfe von GPS-Signalen der von den Interviewern verwendeten Mobilgeräte geprüft werden ob die Begehungsanweisungen korrekt umgesetzt wurden. Paradaten in Form von Kontaktprotokollen wurden bereits in einer Studie des Deutschen Jugendinstituts von 1988 verwendet und lieferten Hinweise, dass Interviewer beim Random-Route dazu neigen, die Personen auszuwählen, die am stärksten motiviert sind, an der Befragung teilzunehmen. Im Vergleich zu einer Stichprobe aus dem Einwohnermelderegister waren bei diesen Personen weniger Kontakte notwendig bis ein Interview realisiert werden konnte (Alt et al. 1991).

### 25.3.2 Paradaten zur Untersuchung von Nicht-Teilnahme

Die Nicht-Teilnahme (Nonresponse) an der Befragung ist ein großes Problem vieler Umfragen (Engel/Schmidt, Kapitel 23 in diesem Band). Daher suchen Umfrageforscher nach Strategien um Nonresponse-Raten und -Verzerrungen zu reduzieren, oder wenigstens das

Ausmaß der Verzerrung durch Nonresponse abschätzen und entsprechend korrigieren zu können. Es gibt derzeit drei gängige Typen von Paradaten um Nonresponse in Umfragen zu untersuchen: *Kontaktprotokolle*, die überwiegend für die Feldkontrolle und Feldsteuerung verwendet werden. Mit ihrer Hilfe kann kontrolliert werden, ob alle zufällig gezogenen Fälle auch tatsächlich eine ausreichende Chance hatten befragt zu werden. Insbesondere wird hier darauf geachtet, dass Kontaktversuche an verschiedenen Tagen und in verschiedenen Zeitfenstern stattfinden und dass ausreichend Zeit zwischen den Kontaktversuchen liegt, so dass kurzfristig verreiste Befragte ebenfalls eine Chance haben, an der Befragung teilzunehmen. *Interviewer-Beobachtungen* der Wohnumgebung oder der gezogenen Einheit werden zum Teil in der Feldarbeit eingesetzt, um Informationen an nachfolgende Interviewer weiter zu geben („Vorsicht bissiger Hund“), spielen aber auch in der Nonresponse-Forschung eine Rolle, wenn Merkmale beobachtet werden, die stark mit den interessierenden Zielgrößen korreliert sind. Im U.S. National Survey of Family Growth gaben Interviewer zum Beispiel ihre Einschätzung darüber ab, ob Kinder zu den Zielhaushalten gehörten oder ob sich die Zielperson in einer aktiven gegengeschlechtlichen Partnerschaft befand (Kirgis/Lepkowski 2013). Liegen diese Beobachtungen sowohl für Befragte als auch für Nicht-Befragte vor, und sind sie ausreichend mit den Zielgrößen und dem Ausfallmechanismus korreliert, können sie unter Umständen für die Nonresponsegewichtung verwendet werden.

Aufzeichnungen der *Interviewer-Haushalt-Interaktionen* können Aufschluss über erfolgreiche Rekrutierungsstrategien geben. Erste Analysen von stimmlichen Eigenschaften von Telefoninterviewern zeigen, dass Interviewer in der Rekrutierung erfolgreicher sind, wenn sie ein moderates Maß an sprachlichen Füllwörtern verwenden (Conrad et al. 2013).

### **25.3.3 Paradaten zur Messfehler-Untersuchung**

Eine weitere Fehlerquelle, die mit Hilfe von Paradaten untersucht werden kann, sind Messfehler bei Befragungen (Krebs/Menold, Faulbaum, Kapitel 30 und 31 in diesem Band). Diese treten immer auf, wenn Antworten von den „wahren“ Werten eines Befragten abweichen (Latcheva/Davidov, Kapitel 55 in diesem Band), was in jeder Stufe des Frage-Antwort-Prozesses der Fall sein kann. Über Messfehler informieren sprachliche Ausdrücke von Zweifel und Unsicherheit, paralinguistische und nichtverbale Hinweise (zum Beispiel zu schnelles oder zu langsames Antworten) bzw. Antwort-Änderungen. Einige dieser Hinweise können mit Hilfe von Paradaten gespeichert und quantifiziert werden. Insbesondere Tastenanschläge können dazu verwendet werden die Benutzerfreundlichkeit und Interaktionen mit einem Computer-gestützten Erhebungsinstrument zu untersuchen. Sie helfen beim Identifizieren von Problemen, die Interviewer und Befragte mit einem Befragungsinstrument haben (Caspar/Couper 1997; Couper 2000). Paradaten, die Messfehler indizieren, werden bisher meistens dazu verwendet problematische Fragen zu entdecken und zu verbessern. Zur Korrektur von Messfehlern mithilfe von Paradaten gibt es bisher sehr wenige Studien.

## 25.4 Welche Paradaten werden in Deutschen Studien erhoben und wofür werden sie verwendet?

Auch in Befragungen, die in Deutschland durchgeführt werden, werden Paradaten erhoben, so z.B. im „German Internet Panel“ (GIP), im Erwachsenenpanel des „Nationalen Bildungspanels“ (NEPS), im „Panel Arbeitsmarkt und soziale Sicherung“ (PASS), im „Sozio-oekonomischen Panel“ (SOEP), in der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS), im „Beziehungs- und Familienpanel“ (pairfam) im „Programme for the International Assessment of Adult Competencies“ (PIAAC), im „Survey of Health, Ageing and Retirement in Europe“ (SHARE) und im „European Social Survey“ (ESS).

Die *datenschutzrechtliche Lage* für die Veröffentlichung von Paradaten ist weitgehend ungeklärt, daher werden Paradaten von den meisten Studien nur intern bzw. von Kooperationspartnern verwendet. Die folgenden Angaben zur Veröffentlichung beziehen sich jeweils auf die letzte Welle. Die Angaben über das Vorhandensein und die Veröffentlichung der Paradaten stammen – wenn nicht anders gekennzeichnet – aus persönlichen Gesprächen mit den Verantwortlichen der Befragungen<sup>1</sup>.

Alle genannten Studien bekommen Kontaktdata (mit sog. *Call Record Data* und *Outcome Codes*) von den jeweiligen Erhebungsinstituten geliefert. Den Studien SHARE, GIP, PASS, ESS, ALLBUS und PIAAC liegen diese für jeden einzelnen Kontaktversuch vor. Diese Daten beinhalten den Tag, die Uhrzeit und die Art des Kontaktversuchs sowie das Ergebnis. SOEP, pairfam und NEPS erhalten nur aggregierte Kontaktinformationen, wie die Gesamtzahl der Kontaktversuche und den Final Disposition Code, wobei NEPS ab der nächsten Welle diese Daten auch auf Kontaktebene erhält. Im ESS werden zusätzlich für jeden Kontakt das Geschlecht und das geschätzte Alter der Kontaktperson erfasst, sowie bei Nicht-Teilnahme der Verweigerungsgrund und die Einschätzung des Interviewers, ob die Person zu einem späteren Zeitpunkt teilnehmen wird. Die Kontaktdata werden beispielsweise von SHARE, GIP und PASS zur Überwachung der Feldphase verwendet, und in PASS gehen diese auch in die Nonresponse-Gewichte ein. In ALLBUS, pairfam, SOEP und NEPS werden die Kontaktdata aggregiert als Anzahl der Kontaktversuche, getrennt nach Kontaktart und verbunden mit dem Datum des Interviews, veröffentlicht. Im PASS ist nur das Interviewdatum und in SHARE der Interviewmonat und das Interviewjahr enthalten. Auf der Ebene der einzelnen Kontakte werden die Daten mit Ausnahme des ESS nur zum internen Gebrauch oder für Kooperationsprojekte freigegeben. Sie werden zum

<sup>1</sup> Persönliche Information durch Johanne Bristle (SHARE), Max-Planck-Institut für Sozialrecht und Sozialpolitik, München; Ulrich Krieger (GIP), Universität Mannheim; Corinna Kleinert (NEPS), Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg; Gerrit Müller (PASS), Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg; Eva-Maria Trüdinger (ESS), Institut für Sozialwissenschaften, Universität Stuttgart; Simone Bartsch (SOEP), Deutsches Institut für Wirtschaftsforschung, Berlin; Michael Blohm (ALLBUS), GESIS-Leibniz-Institut für Sozialwissenschaften, Mannheim; Laura Castiglioni (pairfam), Universität München; Beatrice Rammstedt (PIAAC), GESIS-Leibniz-Institut für Sozialwissenschaften, Mannheim.

Beispiel vom PASS dazu verwendet herauszufinden, ob sich ein idealer Kontaktzeitpunkt für die nächste Welle vorhersagen lässt.

Auch *Zeitmarken* werden in allen Umfragen erhoben, auf Frageebene (SHARE, GIP, PIAAC) oder auf Ebene inhaltlicher Blocks (NEPS, PASS, ESS, pairfam, ALLBUS). Das SOEP erhebt bislang nur die Gesamtlänge des Interviews, plant aber für zukünftige Wellen die Erhebung der Zeitmarken auf Frageebene. Diese werden häufig zur Fragebogenkonstruktion verwendet, d.h. zur Abschätzung, wie lange der Fragebogen oder ein Frageblock sein wird, oder zur Qualitätskontrolle der Interviewer, wobei besonders lange oder kurze Antwortzeiten als Hinweis auf mögliche Fälschungen dienen (Blasius, Kapitel 22 in diesem Band). Zeitmarken werden mit Ausnahme des ESS nicht detailliert, sondern höchstens als Gesamtdauer des Interviews veröffentlicht.

Den meisten Umfrageorganisationen liegen außerdem vom Erhebungsinstitut gelieferte oder am Ende der Befragung erhobene *Informationen über die Interviewer* vor. Dies sind meist Alter, Geschlecht und Bildungsstand des Interviewers sowie dessen Erfahrung als Interviewer. Alter und Geschlecht werden im ESS und pairfam veröffentlicht, im ALLBUS und NEPS zusätzlich auch Bildungsstand und Erfahrung als Interviewer. Im SHARE ist die Interviewer-Demographie bisher nur für die erste Welle veröffentlicht.

Zum Teil stammen die Informationen über die Interviewer von speziell durchgeführten *Interviewer-Befragungen* (siehe z.B. PASS, SHARE, GIP, PIAAC, pairfam, ALLBUS und SOEP; NEPS plant eine Interviewerbefragung für die nächste Welle), die über die soziodemografischen Merkmale hinaus auch Einstellungsfragen beinhalten. Untersucht werden mit ihnen der Einfluss von Interviewer-Eigenschaften oder Einstellungen und Erwartungen im Hinblick auf den Verlauf und die Ergebnisse der Befragung. Auch die Daten der Interviewer-Befragungen sind in der Regel nur intern verfügbar.

SHARE, ESS, GIP (Rekrutierung), SOEP, ALLBUS und PIAAC erheben regelmäßig *Interviewerbeobachtungen*. Diese Beobachtungen beziehen sich z.B. darauf, in welcher Art von Wohngebäude sich ein Haushalt befindet, ob es Anzeichen für die Anwesenheit von Kindern oder Behinderten gibt oder ob der Haushalt über eine Gegensprechchanlage verfügt. Häufig wird auch eine Einschätzung der Wohngegend und der sozialen Schicht des zu Befragenden erhoben. Im NEPS wurde nur in der zweiten Welle eine Information zur Einschätzung der Wohngegend erhoben und auch PASS erhebt sie nicht standardmäßig (bislang nur in der fünften Welle). Für die sechste Welle des ESS in Deutschland lauten die Fragen beispielsweise (ESS6 Kontaktformular):

1. In welcher Art von Gebäude wohnt die Befragungsperson?
2. Gibt es eine Gegensprechchanlage oder ein verschlossenes Tor, so dass Sie die Befragungsperson nicht direkt an ihrer Haustüre kontaktieren können?
3. Wie beurteilen Sie den Zustand des Hauses / des Gebäudes, in dem die Befragungsperson wohnt?
4. In welchem Umfang haben Sie Müll bzw. Abfall in der unmittelbaren Umgebung des Hauses der Befragungsperson wahrgenommen?

- 
5. In welchem Umfang haben Sie beschmierte Hauswände (Graffiti), mutwillig zerstörtes Eigentum oder Vandalismus (z.B. eingeschlagene Fenster) in der unmittelbaren Umgebung des Hauses der Befragungsperson wahrgenommen?

Interviewerbeobachtungen werden im ALLBUS, ESS und SOEP veröffentlicht. ALLBUS und ESS veröffentlichen darüber hinaus auch Informationen zur *Einschätzung der Interview-Situation* durch den Interviewer. In pairfam liegen diese Informationen für die Ankerperson vor. In der sechsten Welle des ESS lauten die Fragen beispielsweise (ESS6 Interviewerfragebogen): Hat der/die Befragte bei irgendwelchen Fragen um Erläuterungen gebeten? Hat der/die Befragte Ihrem Eindruck nach bestimmte Fragen nur mit Widerwillen beantwortet? Hat sich der/die Befragte Ihrem Eindruck nach bemüht, die Fragen zu beantworten so gut er/sie konnte? Hatten Sie alles in allem den Eindruck, dass der/die Befragte die Fragen verstanden hat? War beim Interview jemand anderes anwesend, der/die sich in die Befragung eingemischt hat? Wer war das? Systematische Auswertungen dieser Beobachtungen gibt es bisher eher selten.

---

## 25.5 Zukünftige Entwicklung

Obwohl Paradaten immer weiter verbreitet werden und immer einfacher zu erheben sind, wird ihr Potential oft nicht ausgeschöpft. Selbst die Verwendung von Paradaten zur Feldsteuerung und Feldkontrolle gehört noch nicht zum Alltag aller Umfragedesigner und -manager. Aus diesem Grund bemühen sich einige Datenerhebungsorganisationen derzeit darum „Paradaten Dashboards“ zu entwickeln, um die Nutzung zu vereinfachen (Kirgis/Lepkowski 2013).

Auch die Veröffentlichung von Paradaten wie Kontaktprotokolle oder Antwortzeiten zusammen mit den Umfragedaten ist noch nicht Standard. Ein breiterer Zugriff auf solche Daten, wäre wünschenswert, damit Forscher bei der Datenanalyse beispielsweise selbst Nonresponse-Gewichte für ihre speziellen Zwecke erstellen oder schwierige Fragen identifizieren können.

Generell ist noch nicht geklärt, inwieweit Befragte über die Erhebung und Verwendung von Paradaten informiert werden sollten bzw. ob deren Einverständnis dazu notwendig ist. Während man für Befragte argumentieren kann, dass Paradaten, die im Interview erhoben werden, nicht über die Interviewsituation hinaus gehen, und daher das Einverständnis interviewt zu werden das Einverständnis Paradaten zu erheben mit einschließt, ist dies für Paradaten der Nicht-Teilnehmer nicht so leicht zu entscheiden. Die Entwicklung einheitlicher Richtlinien ist daher der nächste nötige Schritt (Couper/Singer 2013).

So wie die Qualität der Befragungsdaten entscheidend für die Güte der auf ihnen basierenden Schätzungen ist, so hängt auch der Nutzen der Paradaten stark von deren Qualität ab. Zur Sicherung der Qualität der Paradaten ist es wichtig, mögliche Fehlerquellen bei der Erhebung von Paradaten zu identifizieren und den Erhebungsprozess zu optimieren.

Paradaten bieten großes Potential und große Herausforderungen. Die genannten Beispiele zeigen, dass die Erhebung von Paradaten in Deutschland mittlerweile durchaus üblich ist und auch deren Analyse immer mehr Interesse findet. Zukünftige Forschung wird zeigen, ob der Nutzen den Aufwand rechtfertigt.

## Literatur

- Alt, Christian ; Bien, Walter ; Krebs, Dagmar: Wie zuverlässig ist die Verwirklichung von Stichprobenverfahren? Random route versus Einwohnermeldeamtsstichprobe. In: ZUMA Nachrichten 15 (1991), 28: 65-72.
- Callegaro, Mario (2013): Paradata in Web Surveys. In Kreuter (Hg.): 261-276
- Caspar, Raphael A./Couper, Mick P. (1997): Using Keystroke Files to assess Respondent Difficulties with an ACASI Instrument. In: Proceedings of the Section on Survey Research Methods, American Statistical Association: 239-244
- Conrad, Frederick G./Broome, Jessica S./Benkí, José R./Kreuter, Frauke/Groves, Robert M./Vannette, David & McClain, Colleen (2013): Interviewer Speech and the Success of Survey Invitations. In: Journal of the Royal Statistical Society Series A 176(1): 191-210
- Couper, Mick P. (1998): Measuring Survey Quality in a CASIC Environment. In: Proceedings of the Survey Research Methods Section, American Statistical Association: 41-49
- Couper, Mick P. (2000). Usability Evaluation of Computer-Assisted Survey Instruments. In: Social Science Computer Review 18(4): 384-396
- Couper, Mick P./Singer, Eleanor (2013): Informed Consent for Web Paradata Use. In: Survey Research Methods 7(1): 57-67
- Eckman, Stephanie (2013): Paradata for Coverage Research. In Kreuter (Hg.): 97-116
- ESS6 Kontaktformular: [http://www.europeansocialsurvey.org/docs/round6/fieldwork/germany/ESS6\\_contact\\_form\\_DE.pdf](http://www.europeansocialsurvey.org/docs/round6/fieldwork/germany/ESS6_contact_form_DE.pdf) (03.07.2014)
- ESS6 Interviewerfragebogen: [http://www.europeansocialsurvey.org/docs/round6/fieldwork/germany/ESS6\\_interviewer\\_questionnaire\\_DE.pdf](http://www.europeansocialsurvey.org/docs/round6/fieldwork/germany/ESS6_interviewer_questionnaire_DE.pdf) (03.07.2014)
- Kaczmarek, Lars (2013): A Survey of the Uses of Paradata. Präsentiert beim MESS Workshop, 30-31 August 2013, Den Haag
- Kirgis, Nicole G./Lepkowski, James M. (2013): Design and Management Strategies for Paradata-driven Responsive Design: Illustrations from the 2006-2010 National Survey of Family Growth. In: Kreuter (Hg.): 123-143
- Kreuter, Frauke (Hg.) (2013): Improving Surveys with Paradata: Analytic Uses of Process Information. Hoboken, NJ: Wiley
- Kreuter, Frauke (2013a): Improving Surveys with Paradata: Introduction. In: Kreuter (Hg.): 1-9
- Mitchell, Susan/Carley-Baxter, Lisa/Day, Orin/Peytchev, Andy/Sadler-Redmond, Susan (2011): Developing Standardized Paradata and Dashboards for Use Across Multiple Surveys. Präsentiert auf dem 2011 Federal CASIC Workshops, Washington, DC, März 2011
- Olson, Kristen/Parkhurst, Bryan (2013). Collecting Paradata for Measurement Error Evaluations. In: Kreuter (Hg.): 43-65
- Schnell, Hill, Esser (2013). Methoden der Empirischen Sozialforschung. Oldenbourg.
- West, Brady T./Sinibaldi, Jeniffer (2013): The Quality of Paradata: A Literature Review. In: Kreuter (Hg.): 339-356

Julia Cielebak und Susanne Rässler

---

## 26.1 Einleitung

Besteht Interesse an der Analyse der gemeinsamen Verteilung mehrerer Variablen, z.B. bezüglich deren Assoziation oder Interaktionseffekten, werden üblicherweise alle Merkmale bei jedem Element einer Stichprobe (Häder/Häder, Kapitel 18 in diesem Band) erhoben. Prinzipiell lassen sich solche multivariaten Verteilungen (Blasius/Baur, Kapitel 79 in diesem Band) untersuchen, wenn die Ausprägungen für sämtliche Variablen bei allen Untersuchungseinheiten vollständig vorliegen. Begrenzte Ressourcen schränken jedoch in der Forschungspraxis oft sowohl die Stichprobenumfänge, als auch die Anzahl der Variablen in Umfragen ein. Ein weiterer Gesichtspunkt sind die Probleme bei Sekundäranalysen bereits vorliegender Daten, die ursprünglich auf die Ansprüche der Primärforscher zugeschnitten wurden und die interessierenden Variablen möglicherweise nicht vollständig enthalten (Mochmann, Schupp, Kapitel 14 und 73 in diesem Band).

Die *Zusammenführung mehrerer Datenquellen* ist in der Praxis daher weithin die einzige Möglichkeit an Datenmaterial zu gelangen, das bestimmte relevante Variablen gemeinsam enthält und erlaubt, kosten- und zeitintensive Primärerhebungen zu umgehen, wobei es zwei Varianten gibt:

- Bei einer *Datenintegration* (auch *Record Linkage*) werden dabei für jeden Fall (z.B. für jede Person) Variablenwerte aus unterschiedlichen Erhebungen zusammengespielt, z.B. Antworten aus unterschiedlichen Befragungen (Reinecke, Kapitel 44 in diesem Band) oder Befragungsdaten und Sozialversicherungsdaten (Hartmann/Lengerer, Kapitel 71 in diesem Band).
- Wenn bestimmte Informationen über einen Fall komplett fehlen und sich auch nicht aus unterschiedlichen Datenquellen rekonstruieren lassen, kann man diesem Fall im Rahmen einer *Datenfusion* (auch *Statistical Matching*, *Data Fusion*) auch alternativ Variablenwerte zuspielen, die von einem anderen Fall stammen, der diesem Fall statistisch ähnlich ist.

Typische Einsatzgebiete sind beispielsweise die Fusion und Analyse von Konsumverhalten und Mediennutzung in der Marktforschung oder die Verknüpfung von Datenmaterial aus unterschiedlichen amtlichen Registern. Umfangreiche *Single-Source-Untersuchungen*, bei denen alle Informationen aus einem Fragebogen stammen, bergen aufgrund der Länge der Befragung ein hohes Risiko für Interviewabbrüche. In der hohen Antwortbelastung der untersuchten Einheiten und der damit einhergehenden Antwortverweigerung findet sich ein weiterer Grund für die Expansion von Datenfusionen. Eine Datenfusion lässt sich den *Missing-Data Techniken* (Engel/Schmidt, Kapitel 23 in diesem Band) zuordnen und als Ergänzungs- bzw. Imputationsproblem verstehen. Der vorliegende Beitrag definiert unterschiedliche Ausfallmuster und Ausfallmechanismen für fehlende Daten, grenzt die Begriffe Datenfusion und Datenintegration ab und gibt einen Einblick in die Grundidee, Voraussetzungen und Annahmen von Imputationsverfahren.

---

## 26.2 Datenintegration, Datenfusion und statistisches Matching

### 26.2.1 Datenintegration

Datenintegration, auch als *Record Linkage* bezeichnet, ist die Zusammenführung von Daten aus unterschiedlichen Quellen, die zu denselben Einheiten (Personen, Haushalte, Unternehmen) gehören, anhand eindeutig identifizierender Schlüssel (wie Name, Adresse, Sozialversicherungsnummer, SteuerID).

Liegen die Daten vollständig und fehlerfrei vor, ist dieser Prozess trivial und kann anhand eines „Ist-Gleich-Abgleichs“ durchgeführt werden. Oft fehlen jedoch in den Schlüsselvariablen einzelne Werte oder diese unterscheiden sich in den Datenquellen aufgrund von Übertragungsfehlern bzw. abweichenden Schreibweisen (z.B. Maximilian Müller, M. Mueller, Max Müller). In diesem Fall muss eine Ähnlichkeitssuche in den Datensätzen durchgeführt werden. Der Datenintegrationsprozess kann in folgende Schritte unterteilt werden: Bereitstellung der Daten, Standardisierung der Schlüsselvariablen, Berechnung der Ähnlichkeiten, manuelle Zusammenführung „schwieriger Fälle“ und tatsächliche Verknüpfung der Dateien.

Da die Datenintegration in der Praxis relativ einfach zu handhaben ist, konzentrieren wir uns in der folgenden Darstellung auf Probleme der Datenfusion.

### 26.2.2 Datenfusion

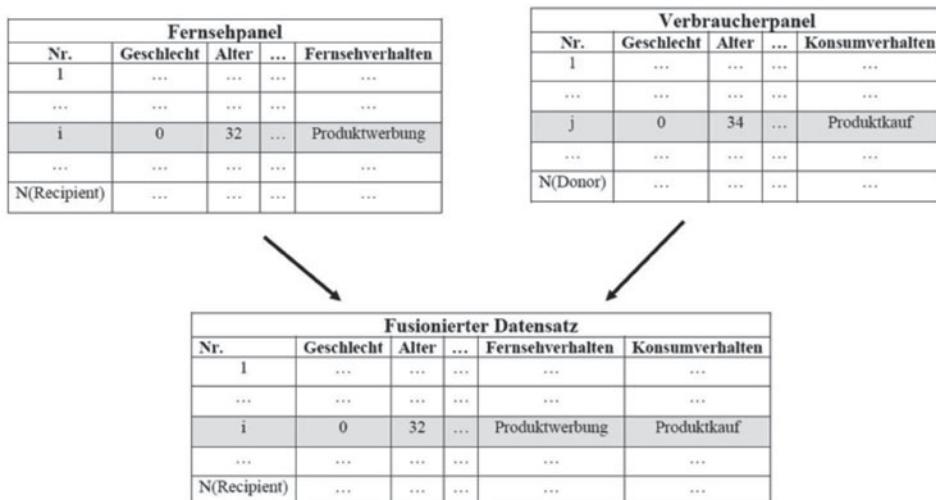
Bei einer *Datenfusion* werden (mindestens) zwei Datensätze miteinander verknüpft, die teilweise nicht-identische Variablen beinhalten, aber im Unterschied zur Datenintegration bzw. zum Record-Linkage werden hierbei *nicht dieselben* Einheiten gesucht, sondern Einheiten, welche in bestimmten *gemeinsamen* Merkmalen (z.B. Geschlecht, Alter, Beruf) gleich oder zumindest ähnlich sind. Im Falle der Datenfusion geht man davon aus, dass in

den beiden Datensätzen jeweils unterschiedliche Einheiten vorliegen. Von diesen ähnlichen Einheiten werden sodann bestimmte *spezifische* Variablen übertragen, die jeweils nur in einem Datensatz vorliegen und der Einheit im anderen Datensatz fehlen. Der entscheidende inhaltliche Unterschied in den formal ähnlichen Vorgehensweisen von Record-Linkage und Datenfusion wird von Rodgers (1984: 102) sehr treffend beschrieben:

“It is a relatively small step, computationally, from such procedures for exact matching of identical individuals to statistical matching of similar individuals. A small step for the computer is in this case a giant step for the statistician – a step that should only be taken with full awareness of the importance of the implicit assumptions and the potential consequences of the incorrectness of those assumptions.”

In traditionellen Fusionsalgorithmen werden in der Regel so genannte *Nearest-Neighbor-Verfahren* eingesetzt, d.h. man sucht nach *statistischen Zwillingen*. Dabei wird zu jedem Befragten im *Empfängerdatensatz* eine Einheit im *Spenderdatensatz* gesucht, die in den gemeinsamen Merkmalen übereinstimmt. Werden mehrere passende Spender (*Donoren*) gefunden, wird zufällig einer ausgesucht. Insbesondere dann, wenn viele gemeinsame Variablen für die Suche verwendet werden, findet sich häufig kein Spender, der dem Empfänger (*Rezipient*) in allen Ausprägungen gleicht. In diesem Fall werden Donoren gesucht, die dem Rezipienten im Sinne eines bestimmten Distanzmaßes möglichst ähnlich sind (Kiesl/Rässler 2005).

Abbildung 26.1 stellt ein stark vereinfachtes Beispiel für ein Nearest-Neighbor-Matching in der Datenfusion dar. Ziel ist es aus einem Verbraucherpanel, in dem das Konsumverhalten aufgezeichnet wurde und einem Datensatz aus dem Fernsehpanel mit Informationen zur Mediennutzung und Fernsehverhalten einen einzelnen Datensatz zu erzeugen.



**Abb. 26.1** Prinzip des Nearest-Neighbor Matching, angelehnt an Kiesl/Rässler (2005)

Dieser soll Charakteristiken zum Konsum- und Fernsehverhalten zusammen enthalten. Anhand der gemeinsamen Merkmale, Geschlecht und Alter wird für jede Beobachtung der Empfängerdatei (Fernsehpanel) eine Beobachtung in der Spenderdatei (Verbraucherpanel) gesucht, die dieser in den gemeinsamen Variablen möglichst ähnlich ist. Von dem gefundenen Donor wird die Ausprägung der spezifischen Variable (Produktkauf) zum Rezipienten übertragen. Nach der Fusion sind nun Analysen zu beiden Zielvariablen (Produktwerbung und Produktkauf) möglich.

Es gibt eine ganze Reihe von *Algorithmen*, die sich zwar im Detail unterscheiden, dieses Grundprinzip jedoch gemein haben. Sie können anhand bestimmter Unterscheidungsmerkmale wie z.B. einfacher oder mehrfacher Verwendung von Beobachtungen aus dem Spenderdatensatz oder Verwendung von Zusatzinformationen und logischen Einschränkungen kategorisiert werden. Eine umfassende Übersicht traditioneller und alternativer Fusionsalgorithmen findet sich in Rässler (2002).

Zu beachten ist, dass unterschiedliche Definitionen des Begriffs „Datenfusion“ existieren. Vor allem in der Markt- und Medienforschung wird er häufig gleichgesetzt mit dem Verschmelzen zweier (oder mehrerer) Datenquellen anhand von *statistischen Zwillingen*. Unserer Ansicht nach ist diese Definition zu eng gefasst. Hintergrund ist, dass sie sich lediglich auf eine bestimmte Klasse von Methoden bezieht (Nearest-Neighbor-Matchingverfahren) und gleichzeitig nichts über die Analyseziele besagt. Wir folgen daher im Folgenden der in der statistischen Literatur üblichen Definition, gemäß der Datenfusion nicht nur den Verknüpfungsprozess, sondern ein spezifisches Datenausfallmuster, also einen Spezialfall der Missing Data-Problematik bezeichnet (Rässler 2002: 6ff.).

### 26.2.3 Statistisches Matching

„Datenfusion“ und „statistisches Matching“ meinen nicht notwendigerweise dasselbe: Die unterschiedlichen Definitionen von „Datenfusion“ sind u.a. auf den englischen Begriff *Statistical Matching* zurückzuführen, der in den USA und in Kanada auch für *Datenfusion (Data Fusion)* mit dem typischen Identifikationsproblem verwendet wird (Rässler 2002: 1f.). Dagegen steht der Begriff „statistisches Matching“ in Europa meist für das Auffinden von statistischen Zwillingen. Bacher (2002) zeigt dabei folgende Anwendungsgebiete für statistisches Matching auf, verstanden als Suche nach statistischen Zwillingen:

- *Datenfusion*: Zwei Datensätze sollen über eine Menge gemeinsamer Merkmale fusioniert werden.
- *Bestimmung einer Kontrollgruppe*: Zu einer Untersuchungsgruppe wird zur Effektschätzung eine Kontrollgruppe aus anderen Daten gezogen, die sich hinsichtlich einer Menge an Kovariablen nicht unterscheidet.
- *Item Nonresponse*: Es gilt fehlende Informationen in einem Datensatz zu ergänzen, z.B. fehlende Einkommensangaben bei einigen Personen.

- *Berechnung des empirischen Re-Identifikationsrisikos:* Bei anonymisierten Registerdaten wird zur Abschätzung des Re-Identifikationsrisikos überprüft, welcher Anteil statistischer Zwillinge den realen re-identifizierten Zwillingen entspricht.

Datenfusion ist in dieser Definition folglich nur *eine* mögliche Variante des statistischen Matching.

---

## 26.3 Datenfusion als Spezialfall des Missing-Data-Problems

### 26.3.1 Ausfallmuster für Missing Data

Eine Datenfusion kann aufgrund der fehlenden spezifischen Variablen in jeweils einem der verknüpften Datensätze grundsätzlich als ein Problem fehlender Werte, also als Missing-Data-Problem (Engel/Schmidt, Lück/Landrock, Kapitel 23 und 28 in diesem Band), verstanden werden. In Abb. 26.2 sind vier unterschiedliche Ausfallmuster für Missing Data dargestellt. Die grauen Felder stehen für vorhandene, die weißen für fehlende Werte.

- Dabei zeigt Abb. 26.1 den allgemeinen Fall einer *unregelmäßigen Ausfallstruktur*. Die Angaben fehlen bei mehreren Variablen (X, Y, V) und bei unterschiedlichen Einheiten, z.B. durch Antwortverweigerung. Es existiert jedoch mindestens eine gemeinsame Variable (Z), die bei allen Einheiten vollständig vorliegt.
- Abbildung 26.2 zeigt eine Situation, *bei der für einen Teil der Beobachtungen bestimmte Merkmale (X) komplett fehlen*, während bei einer Unterstichprobe alle Variablen (Z und X) beobachtet sind. Dies könnte z.B. bei einer Zensusumfrage der Fall sein, bei der bestimmte Sachverhalte nur bei einer Unterstichprobe abgefragt werden, während bestimmte soziodemographische Angaben für alle Einheiten vorliegen.
- In Abb. 26.4 sehen wir ein mögliches Design für ein Fragebogensplitting, bei dem der Fragenkatalog bewusst in *disjunkte Fragenblöcke* aufgeteilt wird, so dass Gruppen von Befragten jeweils nur einen Teil aller Fragen beantworten müssen, um die Antwortbelastung und somit die Wahrscheinlichkeit der Nichtbeantwortung zu minimieren. Bei sehr umfangreichen Erhebungen neigen die Teilnehmer dazu, die Beantwortung z.B. aufgrund von Ermüdungserscheinungen abzubrechen oder die Teilnahme komplett zu verweigern. Beim Fragebogensplitting muss einer der Blöcke (Z) allen Befragten vorgelegt werden, die anderen Blöcke ( $X_1$  bis  $X_4$ ) werden jeweils nur von einem Teil der Einheiten beantwortet. Es wird dabei darauf geachtet, dass alle möglichen Kombinationen von Blockpaaren jeweils bei einer Befragtengruppe gemeinsam erhoben werden.
- Abbildung 26.3 verdeutlicht schließlich die Situation einer *Datenfusion*. Es soll ein Datensatz, der die Variablen Z und X enthält mit einem anderen Datensatz verknüpft werden, der die Variablen Z und Y beinhaltet. Hängt man diese beiden aneinander, entsteht das abgebildete Ausfallmuster, auch *File Concatenation* genannt. Die gemeinsamen Merkmale Z liegen für alle Teilnehmer vor, während die Ausprägungen der spezifischen Zielvariablen X bzw. Y jeweils nur für einen Teil von ihnen zur Verfügung

### 1. Allgemeines Ausfallmuster, unregelmäßige Struktur fehlender Werte

gemeinsam $Z$	spezifisch $X$	spezifisch $Y$	spezifisch $V$

### 2. Database Enrichment

gemeinsam $Z$	spezifisch $X$

### 3. Datenfusion

gemeinsam $Z$	spezifisch $X$	spezifisch $Y$

### 4. SQS: Split Questionnaire Survey Design (Fragebogensplitting)

gemeinsam $Z$	spezifisch $X_1$	spezifisch $X_2$	spezifisch $X_3$	spezifisch $X_4$

beobachtet

fehlend

**Abb. 26.2** Unterschiedliche Ausfallmuster, angelehnt an Rässler (2004)

stehen. Dieses Ausfallmuster hat die bezeichnende Eigenschaft, dass die spezifischen Variablen  $X$  und  $Y$  bei keiner Einheit gemeinsam beobachtet werden und somit keine direkten Informationen zu ihrem Zusammenhang vorliegen (Rubin 1986, Rässler 2002:

72). Dieses *Identifikationsproblem* der Datenfusion liegt bei den drei anderen Ausfallmustern nicht vor, da hier mindestens bivariate Zusammenhänge zwischen allen Variablenpaaren aus den vorhandenen Daten geschätzt werden können.

### 26.3.2 Ursachen für Missing Data

Vor der Behandlung eines Missing-Data-Problems, sollte der Mechanismus, der zu einem Datenausfall führt, näher betrachtet werden, da die Art des Ausfalls beeinflusst, wie man mit ihm umgehen kann oder muss. Mögliche Ausfallmechanismen wurden von Rubin (1976) in die folgenden drei Kategorien eingeteilt:

- *Missing Completely at Random (MCAR)*: Der Datenausfall ist rein zufällig. Er hängt weder von den vorhandenen, noch von den fehlenden Variablenwerten ab. Man spricht von MCAR, wenn die Antwort des Befragten beispielsweise durch äußere Umstände, wie Kaffeeflecken auf dem Fragebogen, unleserlich geworden ist.
- *Missing at Random (MAR)*: Dieser bedingt zufällige Datenausfall ist gegeben, wenn das Fehlen der Angabe nicht von den Variablenwerten selbst abhängt, sondern über die Ausprägungen der beobachteten Variablen erklärt werden kann. Demnach liegt MAR vor, wenn beispielsweise Angaben zum Einkommen nicht aufgrund der Einkommenshöhe, sondern von Personen im höheren Alter überdurchschnittlich oft verweigert werden.
- *Missing Not at Random (MNAR)*: Ein systematischer, verzerrnder Datenausfall liegt vor, wenn dieser direkt von den Werten der fehlenden Variable abhängig ist, z.B. überdurchschnittlich viele fehlende Angaben zum Einkommen bei Hochverdienern. Hieraus würde eine systematische Unterschätzung des Einkommens resultieren.

Ist der Datenausfall MCAR oder MAR und kann angenommen werden, dass die Parameter des datengenerierenden Prozesses von den Parametern des Ausfallprozesses a priori unabhängig sind, kann von einem so genannten *ignorierbaren Ausfallmechanismus* ausgegangen werden. Der Ausfallmechanismus bei einer Datenfusion (sowie beim Split Questionnaire Design) wird auch *Missing by Design* genannt. Hier fehlen die Angaben nicht aufgrund der Ausprägungen der fehlenden oder der beobachteten Variablen, sondern aufgrund der Tatsache, dass sie gar nicht erst erfragt wurden.

In der Praxis ist der Ausfallmechanismus nicht immer einfach zu bestimmen. Demzufolge ist die Entscheidung für die richtige Behandlung der fehlenden Werte nicht trivial. Zur Beantwortung der Frage, ob es sich um einen systematischen oder unsystematischen Datenausfall handelt, gilt es anhand des vorliegenden Datenmaterials festzustellen, inwieweit der datengenerierende Prozess von den beobachteten oder fehlenden Variablenwerten abhängt. Erfüllt der Datenausfall weder das MCAR noch das MAR-Kriterium, liegt ein systematischer Ausfallmechanismus vor, der nicht ignoriert werden darf – dieses ist die größte Herausforderung. Im Rahmen der Datenfusion ist es sinnvoll als Ausfallmechanismus MAR oder sogar MNAR anzunehmen (Rässler 2002: 7).

## 26.4 Varianten des Umgangs mit Missing Data

### 26.4.1 Complete-Case Analyse

Allgemein ist die einfachste Herangehensweise an Analysen unvollständiger Daten das Löschen aller Objekte, bei denen mindestens eine Beobachtung fehlt, d.h. eine Complete-Case Analyse (CC). Außer im seltenen Fall des MCAR führt dieser Ansatz jedoch meistens zu verzerrten Ergebnissen (Little/Rubin 1987), mindestens jedoch zu Effizienzverlusten. Das Ausmaß der Verzerrung hängt neben dem Anteil der fehlenden Werte von den angewandten Analysemethoden sowie dem Grad der Abweichung von der MCAR-Annahme ab.

### 26.4.2 Available-Case Analyse

Eine Alternative zum Complete-Case Ansatz ist die Available-Case Analyse (AC), bei der diejenigen Objekte in die Untersuchung einfließen, die in den zu untersuchenden Variablen vollständige Angaben aufweisen (*listwise deletion*; Engel/Schmidt, Kapitel 23 in diesem Band). Diese Technik kann allerdings vor allem dann zu Problemen führen, wenn die MCAR-Annahme verletzt ist und mehrere Schätzer verglichen oder zu einem kombiniert werden sollen (Little/Rubin 2002, Rässler et al. 2008). Zusätzlich nimmt man bei CC und AC oft einen extremen Datenverlust in Kauf, da u.U. auch Fälle unberücksichtigt bleiben, bei denen lediglich eine oder wenige Variablen fehlen und ansonsten auswertbare Informationen vorhanden wären. CC und AC sind in der Standardsoftware implementiert und somit einfach anzuwenden. Dennoch ist angesichts der Nachteile davon abzuraten. Für die Datenfusion sind diese Ansätze aufgrund des speziellen Ausfallmusters im Regelfall ohnehin nicht geeignet.

### 26.4.3 Gewichtung durch Propensity-Scores

Liegt eine Teilnahmeverweigerung, also *Unit Nonresponse* vor (Engel/Schmidt, Kapitel 23 in diesem Band), kann eine Gewichtung des Befragten durch Propensity-Scores helfen (Kiesl, Kapitel 24 in diesem Band) die Verzerrung zu korrigieren, welche durch die Complete-Case Analyse entstehen würde, sofern MCAR nicht vorliegt. Derartige Gewichtungen basieren auf zusätzlichen Informationen, die für alle Befragten vorliegen. Die Grundidee einer Gewichtung für Unit Nonresponse besteht darin, die Teilnahmeverweigerer als eine eigene Stichprobenschicht zu betrachten und jeden Teilnehmer mit der Inversen seiner geschätzten Auswahl- und Antwortwahrscheinlichkeit zu gewichten. Als Ausfallmechanismus wird implizit MAR unterstellt.

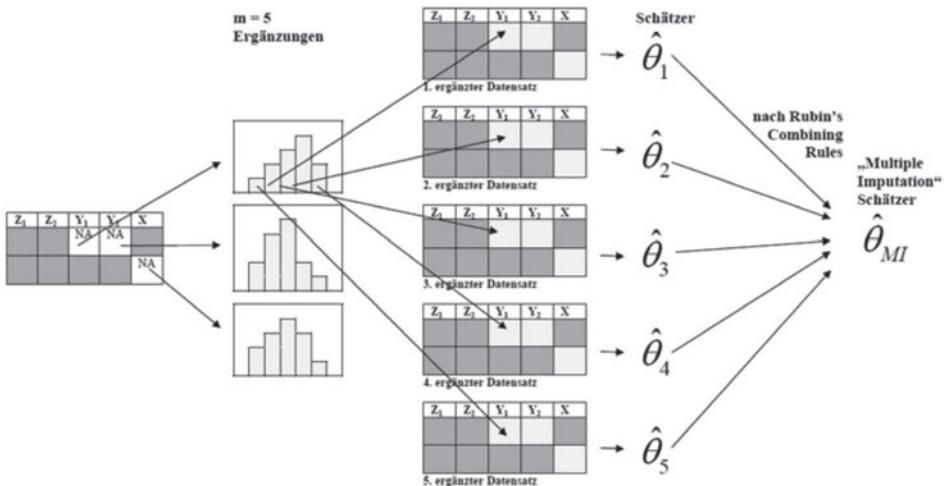
### 26.4.4 Einfache Imputation

Für den Fall des *Item Nonresponse*, d.h. sofern für einen Befragten nicht der komplette Fragebogen fehlt, sondern nur einzelne Werte (Engel/Schmidt, Kapitel 23 in diesem Band), sind Gewichtungsverfahren weniger geeignet. Das Verwerfen ganzer Fälle, selbst wenn nur wenige Fragen unbeantwortet blieben, ohne dass die in den Daten vorhandene Information für die Gewichterstellung genutzt wird, ist als kritisch anzusehen. Aus diesem Grund werden in der Regel Ergänzungsmethoden (*Imputation*) bei Item Nonresponse bevorzugt.

*Einfache Imputation* (*Single Imputation, SI*) umfasst Verfahren, bei denen für jeden fehlenden Wert ein plausibler Wert eingesetzt wird, z.B. ein (bedingter) Mittelwert. Das Ergebnis ist *ein vervollständigter Datensatz*. Hierbei wird als Ausfallmechanismus MAR unterstellt. Ein grundlegender Vorteil der Single Imputation ist, dass die datenerhebende Person ihr zusätzliches Wissen um die Gründe für den Datenausfall in die Imputation einfließen lassen kann. Zusätzlich kann der vervollständigte Datensatz grundsätzlich mit Standardmethoden analysiert werden. Problematisch ist allerdings, dass bei der Verwendung von Standardanalysen und Standardsoftware bei einfach imputierten Daten, diese behandelt werden, als seien sie tatsächlich so erhoben worden. Selbst wenn der Schätzer im vervollständigten Datensatz unverzerrt ist, wird die zusätzliche Variabilität durch das Nichtkennen der fehlenden Werte vernachlässigt, also die Streuung des Parameterschätzers unterschätzt. Fehlt beispielsweise bei einigen Untersuchungseinheiten die Einkommensangabe, und wird im vereinfachten Fall bei allen Einheiten das Durchschnittseinkommen imputiert, führen Analysen des so vervollständigten Datensatzes nicht nur zu einer unterschätzten Streuung, sondern auch zu engeren Konfidenzintervallen und verfälschten Testergebnissen. Folglich sollte für eine statistisch valide Inferenz bei einer einfachen Imputation eine Korrektur des Varianzschätzers erfolgen (Rubin 1987: 11ff., Rässler et al. 2008).

### 26.4.5 Mehrfache Ergänzung

Der mehrfachen Ergänzung (*Multiple Imputation, MI*) fehlender Werte liegt die Idee zu Grunde, dass jeder fehlende Wert nicht einmal, sondern  $m > 1$  Mal ergänzt wird. Abbildung 26.3 veranschaulicht das Prinzip der MI angewendet auf das Ausfallmuster der Datenfusion. Demnach bekommt jeder fehlende Wert (not available oder NA)  $m > 1$ , hier  $m=5$  plausible Werte zugeordnet, die aus einer Wahrscheinlichkeitsverteilung gezogen werden. Somit entstehen  $m=5$  vervollständigte Dateien, die jeweils mit Hilfe von Standardanalysen die interessierenderen Schätzer liefern. Diese werden anschließend, Rubins Combining Rules (Rubin 1987: 76ff.) folgend, gemittelt und ergeben den MI Schätzer. Die zusätzliche Unsicherheit aufgrund des Fehlens der Werte spiegelt sich im höheren Varianzschätzer wider, der aus der „Within-Imputation“-Varianz und der „Between-Imputation“-Varianz (Rubin 1987: 76ff.) kombiniert wird. Somit sorgt die mehrfache Ergänzung für eine valide Inferenz.



**Abb. 26.3** Konzept der Multiplen Imputation, angewendet auf Datenfusion, angelehnt an Rässler (2002: Abb. 4.2)

Verfahren der Multiplen Imputation haben neben den positiven Eigenschaften der Single Imputation noch weitere Vorteile: Das zufällige Ziehen der imputierten Werte aus einer Verteilung erhöht die Effizienz der Schätzer. Ein weiterer Vorteil ergibt sich aus der Tatsache, dass die  $m$  ergänzten Datensätze Inferenzanalysen zur Sensitivität auf unterschiedliche Modelle erlauben (Rubin 1987: 15f.).

Im Allgemeinen existieren drei wesentliche Nachteile der Multiplen Imputation (MI) (Rubin 1987: 17f.):

- der größere Aufwand bei der Erstellung mehrerer ergänzter Datensätze,
- der größere Rechnerspeicherbedarf sowie
- der Mehraufwand bei der Analyse.

Alle drei Nachteile relativieren sich angesichts der mittlerweile hohen verfügbaren Rechnerleistung und der für jedermann zugänglichen Softwarepakete, z.B. in R oder SPSS. Das grundlegende Ziel der MI ist es, dem Analysten mehrere vervollständigte Datensätze zur Verfügung zu stellen, die jeweils mit Standardsoftware und -analysemethoden ausgewertet werden können.

## 26.5 Konsequenzen impliziter Verteilungsannahmen der Datenfusion

Prinzipiell ist MI für eine Datenfusion sehr gut geeignet. Ein entscheidender Unterschied zu anderen Missing-Data Problemen liegt im *Identifikationsproblem*. Da die spezifischen Variablen aus den beiden zusammengeführten Datensätzen nie gemeinsam beobachtet werden, ist deren Zusammenhang bzw. deren gemeinsame Verteilung nicht aus dem vorhandenen Datenmaterial direkt schätzbar. Standard-MI Verfahren stellen bei Datenfusion die bedingte Unabhängigkeit ein. Man kann jedoch zulässige Verteilungen durch die vorhandenen Randverteilungen und Korrelationen etwas eingrenzen.

Für den gemeinhin schwierigen Fusionsprozess sind *implizite, nicht überprüfbare Annahmen über Zusammenhänge zwischen den spezifischen Variablen X und Y* zu treffen. Diese werden bei einer Datenfusion nie gemeinsam beobachtet, weshalb deren wahre gemeinsame Verteilung nicht aus den vorliegenden Daten eindeutig abzuleiten ist. Die (künstliche) gemeinsame Verteilung nach der Fusion kann aus den Verteilungen zwischen X und Z sowie Y und Z in den einzelnen Ausgangsdatensätzen konstruiert werden. Diese ist genau dann gleich der wahren (aber unbekannten) gemeinsamen Verteilung, wenn die spezifischen Variablen X und Y bedingt unabhängig gegeben die gemeinsame Variable Z sind (Sims 1972, Kiesl/Rässler 2005). Es ist bekannt, dass traditionelle Fusionsalgorithmen die „bedingte Unabhängigkeit“ herstellen. Demnach wird diese Annahme (*Conditional Independence Assumption* oder CIA) bei traditionellen Verfahren implizit getroffen.

Bedenkt man, dass eine Fusion meist durchgeführt wird, um Zusammenhangsstrukturen zwischen den nicht gemeinsam erhobenen spezifischen Variablen zu entdecken und Handlungsempfehlungen daraus abzuleiten, ist die Richtigkeit der CIA zumindest kritisch zu hinterfragen. Bemüht man erneut das Beispiel aus Abb. 26.1, bei dem Daten aus dem Fernsehpanel mit Daten aus dem Verbraucherpanel fusioniert werden, würde das bedeuten: Im fusionierten Datensatz wären in den einzelnen Teilgruppen, der weiblichen Befragten eines bestimmten Alters bzw. der männlichen Befragten eines bestimmten Alters, also bedingt auf die gemeinsamen Variablen Alter und Geschlecht, die beiden spezifischen Variablen Fernsehverhalten und Konsumverhalten stochastisch unabhängig. Nun liegt bei weiteren Analysen der Fokus meist gerade auf dem Zusammenhang dieser spezifischen Variablen, der in der künstlich gebildeten Fusionsstichprobe u.U. nicht richtig abgebildet wird. Dies wiederum birgt die *Gefahr des statistischen Fehlschlusses* (Graeff, Kapitel 72 in diesem Band). Zudem vertraut man allein auf die Erklärungskraft der gemeinsamen soziodemographischen Variablen für die Zusammenführung der Einzeldatensätze. Waren die spezifischen Variablen perfekt mit den gemeinsamen korreliert, wäre die Verteilung nach der Fusion perfekt wiedergegeben. In der Praxis fällt diese Korrelation aber meist nur gering aus. Es hat sich deshalb gerade in der Markt- und Medienforschung bewährt, zusätzlich zu den gemeinsamen soziodemographischen Merkmalen „spezifische gemeinsame“ Eigenschaften abzufragen, z.B. eine Zusatzfrage zum Kaufverhalten der Teilnehmer des Fernsehpanels (Rässler 2002: 33ff.), welche die Fusion „stützen“ soll.

## 26.6 Validitätsebenen der Datenfusion

Um die Frage zu beantworten, wann eine Fusion erfolgreich durchgeführt wurde, lassen sich deren unterschiedliche wünschenswerte Eigenschaften und Ziele betrachten. Rässler (2002: 29-33) diskutiert die folgenden vier Validitätsebenen einer Datenfusion:

### 26.6.1 Erstens: Reproduktion der einzelnen unbekannten Werte

Fälschlicherweise wird die exakte Reproduktion der einzelnen unbekannten wahren Werte häufig als oberstes Ziel der Datenfusion gesehen. Man könnte versuchen, die Anzahl solcher exakt geschätzten Werte (Treffer) zu maximieren. Jedoch ist es aus mehreren Gründen nicht sinnvoll, die Qualität der Fusion am Anteil der Treffer zu messen. Für stetige Variablen ist die Wahrscheinlichkeit für einen Treffer Null, für diskrete Variablen mit vielen Ausprägungen ist diese nur sehr gering und für einen multivariaten Fall mit mehreren spezifischen Variablen nahe Null. Ein weitaus gewichtigeres Argument ist, dass mit einem höheren Trefferanteil nicht unbedingt die Reproduktion der Verteilung gewährleistet ist. Kiesl/Rässler (2005: 25) diskutieren hierfür das folgende Beispiel:

„Zu ergänzen sei im Empfängerdatensatz eine binäre Variable Y, die im Spenderdatensatz kaum Korrelation mit den gemeinsamen Variablen Z zeigt. Die Häufigkeit von Y=1 im Spenderdatensatz sei 60%. Eine Fusionsstrategie könnte nun darin bestehen, im Empfängerdatensatz für Y generell den Wert 1 zu setzen; man erreicht damit eine Trefferrate von ungefähr 60%, hat aber die Randverteilung von Y völlig verfälscht. Setzt man dagegen Y in jeder Empfängerbeobachtung zufällig mit einer Wahrscheinlichkeit von 60% auf den Wert 1, wird man am Ende die Randverteilung (und wohl auch die gemeinsame Verteilung) sehr gut widerspiegeln; die zu erwartende Trefferrate ist aber nur  $0,6 \cdot 0,6 + 0,4 \cdot 0,4 = 52\%$ .“

Eine hohe Trefferrate ist offenbar kein relevantes Validitätskriterium für eine erfolgreiche Datenfusion und entspricht nicht den eigentlichen Zielen, die damit verfolgt werden. Datenfusionen werden durchgeführt, um im weiteren Schritt statistische Analysen zu ermöglichen, nicht um die wahren individuellen Werte für jedes Objekt zu „erraten“.

### 26.6.2 Zweitens: Erhalt der gemeinsamen Verteilung

Der Erhalt der gemeinsamen Verteilung ist das eigentliche erstrebenswerte Ziel der Datenfusion. Würde man es schaffen, die wahre unbekannte gemeinsame Verteilung nach der Fusion einzustellen, wären alle Analyseergebnisse, die auf dem verknüpften Datensatz basieren, valide. Das offensichtliche Problem besteht darin, dass das Erreichen dieser Validitätsstufe nicht ohne Zusatzinformation, mit dem vorliegenden Datenmaterial, überprüft werden kann. Es kann lediglich im Rahmen von Simulationsstudien untersucht werden, inwiefern ein bestimmter Fusionsalgorithmus Ergebnisse liefert, welche die wahre gemeinsame Verteilung gut widerspiegeln.

### 26.6.3 Drittens: Erhalt der Korrelationsstruktur

Da ein häufiges Untersuchungsziel bei fusionierten Daten weniger Aussagen zur gemeinsamen Verteilung, als zur Korrelation zwischen den spezifischen Variablen sind, ist es sinnvoll den Erhalt der Korrelationsstruktur nach der Fusion anzustreben. Das kann allerdings mit den vorliegenden Daten nicht direkt überprüft werden. Die wahre Korrelationsstruktur im fusionierten Datensatz wird korrekt wiedergegeben, wenn die spezifischen Merkmale X und Y im Durchschnitt bedingt unabhängig sind, und damit bedingt unkorreliert gegen (?) die gemeinsamen Variablen Z. Diese Eigenschaft lässt sich wiederum nicht überprüfen. Das verfügbare Datenmaterial erlaubt es jedoch, einen Bereich möglicher Korrelationskoeffizienten zwischen X und Y anzugeben (Rässler 2002: 115ff., Rässler 2004, Kiesl/Rässler 2005).

### 26.6.4 Viertens: Erhalt der Randverteilungen

Die Mindestanforderung an einen fusionierten Datensatz ist der Erhalt der Randverteilungen, die bereits in den einzelnen Datensätzen vor der Fusion bestanden. So sollten statistische Analysen bezüglich der Parameter der Randverteilungen nach einer validen Fusion mit den Ergebnissen aus den Einzeldatensätzen vor der Fusion übereinstimmen (Rässler 2002: 31f.). In der Praxis kommen hierzu oft Signifikanztests für die Verteilungsparameter, z.B.  $\chi^2$ - oder t-Tests zum Einsatz, deren Ziel die Beibehaltung der Nullhypothese ist. Bei der Ableitung belastbarer Aussagen aus dem Beibehalten der Nullhypothese solcher Tests ist im Gegensatz zum Ablehnen Vorsicht geboten. Zudem ist selbst bei einer Reproduktion der Randverteilungen nicht unbedingt gewährleistet, dass auch die gemeinsame Verteilung tatsächlich reproduziert wurde (Kiesl/Rässler 2005). Meist sind mehrere unterschiedliche gemeinsame Verteilungen vorstellbar, die zu den vorliegenden Randverteilungen passen.

---

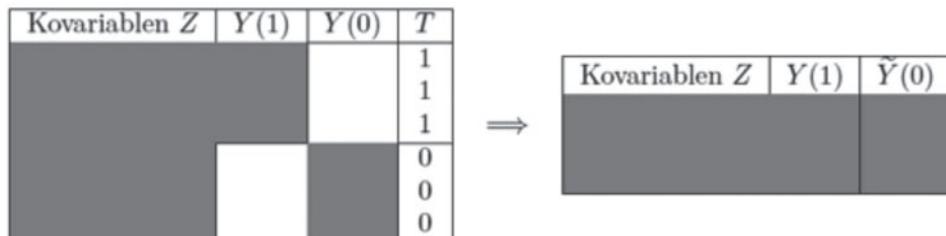
## 26.7 Statistisches Matching vs. Propensity-Score Matching

Abschnitt 4 hat gezeigt, dass das Statistisches Matching, verstanden als Suche nach statistischen Zwillingen, bei unterschiedlichen Missing-Data Problemen Anwendung findet. Zum Zweck der Bestimmung einer Kontrollgruppe zur Effektschätzung einer Maßnahme ist Statistisches Matching, etwa Propensity-Score Matching nach Rosenbaum/Rubin (1983), weit verbreitet. Der zugrunde liegende Ansatz hierzu wird als *Rubins Kausalmodell* bezeichnet (Rubin 1974 sowie viele weitere Veröffentlichungen in den folgenden Jahren). Hierbei werden zwei Personengruppen betrachtet, von denen eine ein bestimmtes Treatment erhält, z.B. eine Maßnahme der aktiven Arbeitsmarktpolitik, die Zuweisung zu dem Treatment aber nicht zufallsgesteuert ist. Zur Messung kausaler Effekte gilt das Zufallsexperiment nach wie vor als „Königsweg“ (Kühnel/Dingelstedt, Kapitel 80 in diesem Band),

ist aber häufig nicht durchführbar. Sei  $T=1$ , falls eine Person  $i$  das Treatment  $T$  erhält,  $T=0$  sonst, und  $Y$  eine interessierende Ergebnisvariable, beispielsweise das Einkommen nach fünf Jahren oder die Dauer der Arbeitslosigkeit. Die beiden potentiellen Ergebnisse für Person  $i$  sind  $Y_i(1)$  bei Maßnahmeteilnahme, und  $Y_i(0)$  sonst. Zusätzlich gibt es für jede Person weitere Variablen  $Z$ , welche die Zuweisung zur Maßnahme beeinflussen können wie Alter, Bildung oder Berufserfahrung. Im Idealfall wären beide Ergebnisse für jedes Individuum bekannt, so dass jeweils der individuelle Treatment-Effekt  $Y_i(1) - Y_i(0)$ , die Differenz in der Ergebnisvariable mit und ohne Maßnahmeteilnahme bekannt wäre. Es können jedoch nur die beiden Fälle  $Y(1)|T=1$ , Ergebnisvariable nach Teilnahme bei den Teilnehmern, und  $Y(0)|T=0$ , Ergebnisvariable ohne Teilnahme bei den Nicht-Teilnehmern, tatsächlich beobachtet werden. Die kontrafaktischen Situationen  $Y(1)|T=0$ , Ergebnis bei Teilnahme bei den Nicht-Teilnehmern, und  $Y(0)|T=1$ , Ergebnis bei Nicht-Teilnahme bei den Teilnehmern, hingegen nicht. Der durchschnittliche Effekt der Teilnahme auf die Teilnehmer, der sogenannte Average Treatment Effect on the Treated (ATET), ist gegeben durch  $ATET = E(Y(1)|T=1) - E(Y(0)|T=1)$ . Das nicht beobachtbare Ergebnis von Teilnehmern im Falle einer Nicht-Teilnahme  $E(Y(0)|T=1)$  lässt sich mit Hilfe von Matching-Verfahren schätzen, indem eine Vergleichsgruppe mit gleichen oder ähnlichen relevanten Kovariablen  $Z$  gebildet wird. Relevante Merkmale sind solche, die sowohl die Maßnahmeteilnahme als auch den Erfolg der Maßnahme, d.h.  $Y(1)$  beeinflussen.

Die Schätzung des ATET ist über die statistischen Zwillinge möglich, wenn die Annahme bedingter Unabhängigkeit (CIA) erfüllt ist, d.h. für gegebene  $Z$  das Ergebnis nach der Teilnahme  $Y(1)$  und ohne Teilnahme  $Y(0)$  unabhängig von der Zuweisung zum Treatment ist. Die Suche nach statistischen Zwillingen für die Bildung der Vergleichsgruppe anhand aller relevanten Variablen  $Z$  kann allerdings rechnerisch sehr aufwendig sein. Eine wesentliche Vereinfachung bildet das Propensity-Score Matching, welches die Wahrscheinlichkeit für die Zuweisung zum Treatment gegeben die relevanten Variablen  $p(Z) = P(T=1|Z)$  verwendet (Rosenbaum/Rubin 1983).

Die Analogie zwischen der Ausgangssituation einer Datenfusion und einer Effektschätzung nach Rubins Kausalmodell ist offensichtlich, wenn man Abb. 26.4 und Abb. 26.2 (Bild 3) miteinander vergleicht. Infolgedessen liegt der Gedanke nahe, Propensity-Score Matching auch für die Datenfusion zu verwenden. Warum ist dieses Verfahren für die



**Abb. 26.4** Statistisches Matching zur Schätzung kausaler Effekte

Datenfusion ungeeignet? Es wurde entwickelt, um Stichproben mit gleichen Strukturen zu erzeugen. Matching auf die relevanten Kovariablen Z oder den Propensity-Score  $p(Z)$  balanciert die Verteilungen der Personenmerkmale zwischen Maßnahmeteilnehmern und der Vergleichsgruppe und bildet faktisch eine zufällige Zuweisung nach. Bei einer Datenfusion handelt es sich bereits um Zufallsstichproben aus der gleichen Grundgesamtheit. Das Propensity-Score-Modell hat hier keinerlei Aussagekraft. Für eine Effektschätzung im Kausalmmodell spielt die quasi-zufällige Zuweisung keine Rolle, bei einer Datenfusion ist sie dagegen nicht erwünscht, denn Ziel einer Datenfusion ist gerade die Auswertung von Zusammenhangsstrukturen zwischen den spezifischen Merkmalen.

Rässler (2002: 36-43) zeigt in einer Reihe von Simulationsstudien, dass die gemeinsame Verteilung der spezifischen Variablen im Kontext der Datenfusion nach einem Matching via Propensity-Scores verfälscht wird und lediglich bei bedingter Unabhängigkeit funktioniert. Obwohl mehrere Artikel diese Methode vorschlagen, ist von der Verwendung des Propensity-Score Matching für die Datenfusion abzusehen.

---

## 26.8 Zusammenfassung und Ausblick

Datenfusionen werden bereits seit vielen Jahren durchgeführt, wobei die Bekanntheit und Beliebtheit stetig zunimmt. Nicht zuletzt dank der gesteigerten Rechnerleistung und der mittlerweile verfügbaren Software. Wie gezeigt werden konnte, sind die Begrifflichkeiten rund um Datenfusion und die statistischen Methoden, die hierfür genutzt werden, nicht einheitlich. Folglich sollten diese vor der Durchführung einer Datenfusion neben den Untersuchungszielen genau definiert werden. Angesichts der Vielzahl unterschiedlicher Analyseziele und Verfahren erweist es sich als schwierig, ein universell bestes Ergänzungsverfahren zu benennen. Parametrische MI-Verfahren eignen sich jedoch sehr gut dafür und führen zu valider Inferenz, wohingegen Propensity-Score Matching eher als ein Standardverfahren für die Effektschätzung über Kontrollgruppen gilt. Das Identifikationsproblem und die damit einhergehenden Annahmen stellen die entscheidende Herausforderung bei einer Datenfusion dar. Versuche, diese zu überwinden und zu validen Fusionen zu gelangen, sind aktueller Forschungsgegenstand.

## Literatur

- Bacher, Johann (2002): Statistisches Matching. In: ZA-Informationen 51: 3-66
- Kiesl, Hans/Rässler, Susanne (2005): Techniken und Einsatzgebiete von Datenintegration und Datenfusion. In: König, Christian/Stahl, Matthias/Wiegand, Erich (Hg.): Datenfusion und Datenintegration. Bonn: Informationszentrum Sozialwissenschaften. 17-32
- Little, Roderick J.A./Rubin, Donald B. (1987): Statistical Analysis with Missing Data. New York: John Wiley
- Little, Roderick J.A./Rubin, Donald B. (2002): Statistical Analysis with Missing Data. Hoboken: Wiley
- Rässler, Susanne (2002): Statistical Matching. New York: Springer
- Rässler, Susanne (2004): Data Fusion. In: Austrian Journal of Statistics 33: 153-171
- Rässler, Susanne/Rubin, Donald B./Zell, Elizabeth R. (2008): Incomplete Data in Epidemiology and Medical Statistics. In Rao, Calyampudi Radhakrishna/Miller, J. Philip/Rao, Dabeeru C. (Hg.): Handbook of Statistics 27: Epidemiology and Medical Statistics. Amsterdam: Elsevier Science & Technology Books. 569-601
- Rodgers, Willard L. (1984): An Evaluation of Statistical Matching. In: Journal of Business and Economic Statistics 2 (1): 91-102
- Rosenbaum, Paul R./Rubin, Donald B. (1983): The Central Role of the Propensity Score in Observational Studies for Causal Effects. In: Biometrika 70 (1): 41-55
- Rubin, Donald B. (1974): Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. In: Journal of Educational Psychology 66: 688-701
- Rubin, Donald B. (1976): Inference and Missing Data. In: Biometrika 63: 581-592
- Rubin, Donald B. (1986): Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations. In: Journal of Business and Econometric Statistics 4: 87-95
- Rubin, Donald B. (1987): Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Son
- Sims, Christopher A. (1972): Comments. In: Annals of Economic and Social Measurement 1: 343-345

Udo Kuckartz und Stefan Rädiker

---

## 27.1 Einleitung

Charakteristisch für qualitative Daten ist ihre große Vielfalt, die von Antworten auf offene Fragen in Bevölkerungsumfragen (Züll/Menold, Kapitel 52 in diesem Band) über transkribierte narrative Interviews (Küstrers, Kapitel 40 in diesem Band) und Gruppendiskussionen (Kapitel 41 in diesem Band) bis hin zu visuellen Daten wie Photographien (Bohnsack, Kapitel 67 in diesem Band) und Videos (Tuma/Schettler, Kapitel 68 in diesem Band) reicht. Dabei spielen qualitative Daten nicht nur im Rahmen qualitativer Forschung eine Rolle, auch in der quantitativ orientierten Forschung, etwa im Bereich der Surveyforschung (Stein, Reinecke, Kapitel 7 und 44 in diesem Band) sind qualitative Daten sehr häufig anzutreffen. Dies gilt erst recht für Mixed-Methods Ansätze (Kapitel 8 in diesem Band), die zunehmend auf Interesse stoßen und die in unterschiedlicher Art und Weise qualitative und quantitative Daten in ihren Designs kombinieren und integrieren.

Der Sinn und Zweck qualitativer Daten im Rahmen von quantitativer Forschung ist ebenso vielfältig wie die Bandbreite ihrer Erscheinungsformen. Think-Aloud Protokolle werden bei der Instrumentenentwicklung eingesetzt und Befragte können ihre Angaben in Online-Befragungen (Wagner/Hering, Kapitel 48 in diesem Band) durch erklärende Freitextantworten ergänzen. Im Kontext von Mixed-Methods können qualitative Daten je nach gewähltem Design der Exploration des Forschungsfeldes und der Explanation der quantitativen Forschungsergebnisse dienen (Creswell/Plano Clark 2010). Im Konzept der Triangulation (Flick 2004) übernehmen die qualitativen Daten die Aufgabe, die quantitativen Ergebnisse zu validieren – und vice versa.

Heutiger Standard ist, für die Auswertung qualitativer Daten spezielle Software, genannt QDA-Software, einzusetzen. Die weltweit bekanntesten Programme sind Atlas.ti, MAXQDA und NVivo, welche in zahlreichen Disziplinen Verwendung finden. Diese Art der Software eignet sich als Hilfsmittel in allen Phasen des Forschungsprozesses von der Literaturanalyse (Salheiser, Ernst, Klein, Kapitel 62, 63 und 64 in diesem Band) zur

Entwicklung einer Fragestellung über die Datenaufbereitung bis zur Analyse und Ergebnispräsentation (Meyer/Meier zu Verl, Friedrichs, Kapitel 15 und 16 in diesem Band).

---

## 27.2 Formen qualitativer Daten

Die Vielfalt qualitativer Daten lässt sich grob in zwei Kategorien aufgliedern: in *textförmige Daten* und *Multimedia-Daten*. Unter „Daten“ verstehen wir bei dieser Aufteilung die Rohdaten, welche die Forscher und Forscherinnen aus dem Feld mitbringen, die in der Regel für eine spätere Analyse erst entsprechend aufbereitet werden müssen.

### 27.2.1 Textförmige Daten

Zu den sehr häufig vorkommenden textförmigen Daten zählen alle Mitteilungen, die in Schriftsprache verfasst wurden (Salheiser, Kapitel 62 in diesem Band). Dies können einerseits Texte sein, die primär für Forschungszwecke entstanden sind, wie es bei Freitextantworten in Befragungen oder schriftlichen Online-Gruppendiskussionen der Fall ist. Andererseits zählen hierzu auch nicht-reaktive Daten, die außerhalb eines Forschungskontextes entstanden sind, der Forschung aber potenziell zur Verfügung stehen, wie beispielsweise Zeitungsbeiträge oder Abstracts von Zeitschriftartikeln (Klein, Kapitel 64 in diesem Band). Hierzu zählen ebenfalls Daten, die originär aus dem Internet stammen: Forendiskussionen, Twitternachrichten, Blogeingänge, ganze Webseiten von Unternehmen und Einrichtungen sowie zahlreiche PDF-Dokumente, die zum Download bereitstehen, wie Parteiprogramme, Regierungsreden oder Hausarbeiten von Studierenden (Schmitz/Yanenko, Kapitel 65 in diesem Band). Durch die Verbreitung des Internets und den zunehmend einfacheren Verfahren, um im Internet als Privatperson Texte etwa in Webblogs zu veröffentlichen, hat die Vielfalt und vor allem die Verfügbarkeit derartiger Daten in der letzten Dekade erheblich zugenommen. Für diese Daten steht weniger die Erhebung im klassischen Sinne als das Auffinden und ihre sinnvolle, mit technischen Fallstricken verbundene Aufbereitung im Vordergrund.

Es existieren zahlreiche, teilweise sich überlappende Merkmale, die sich den einzelnen Textsorten zuschreiben lassen und anhand derer die textförmigen Daten weiter differenziert werden können. Neben dem bereits genannten *Entstehungskontext* und dem damit verbundenen *Ausmaß möglicher Reaktivität* unterscheiden sich Texte insbesondere durch ihren *Umfang* und die *Anzahl der zu differenzierenden Fälle*. Während bei schriftlichen Befragungen (Reuband, Kapitel 47) und Twitterbeiträgen zumeist sehr viele, aber relativ kurze Antworten vorliegen, verhält es sich bei Parteiprogrammen genau umgekehrt und es liegen deutlich weniger, aber dafür längere Texte vor. Diese beiden Beispiele unterscheiden sich zumeist auch bezüglich ihrer *Elaboriertheit* und ihrer *Verständlichkeit*, denn es ist unterschiedlich viel Kontextwissen und Fachwissen notwendig, um die Texte verstehen zu können. Zunehmend spielt auch das Kriterium der *orthographischen Korrektheit*

eine große Rolle, was ein großes Problem für die automatisierte Verarbeitung (lexikalische Suche, automatische Vercodung etc.) darstellen kann. So weisen zahlreiche Internet-Texte wie Twitterbeiträge und Antworten auf Online-Surveys eine deutlich schlechtere Orthographie auf als etwa Parteiprogramme oder andere PDF-Dokumente. Ein weiterer Aspekt betrifft schließlich den *Interaktionsgrad*, den ein Text abbildet: Handelt es sich gewissermaßen um einen Monolog einer Einzelperson, wie es in einer schriftlichen Befragung der Fall ist, oder spiegeln die Texte einen Dialog, wie eine Forendiskussion, an der zahlreiche Teilnehmende beteiligt sind, die gegenseitig Bezug aufeinander nehmen?

## 27.2.2 Multimedia-Daten

Im Bereich der Multimedia-Daten erscheinen uns drei unterschiedliche Datenformen als zentral.

### 27.2.2.1 Bilder und Fotos

Die Bedeutung von Bildern und Fotografien für die Sozialwissenschaft hat in den letzten Jahren aufgrund technischer Errungenschaften wie Digitalkameras, Farb-Drucker und Bildbearbeitungs-Software für den Heimgebrauch einerseits und der Etablierung einer „Bildwissenschaft“ (Schulze 2013: 530, Sachs-Hombach 2005, Bohnsack, Tuma/Schnettler, Kapitel 67 und 68 in diesem Band) andererseits an erheblicher Bedeutung gewonnen. Unter Bildern verstehen wir Visualisierungen, die „mit der Hand“ gezeichnet wurden. Sie entstehen häufig während eines Face-to-face-Interviews (Stocké, Kapitel 45 in diesem Band), bei dem Genogramme, narrative Stegreifskizzen (Behnken/Zinnecker 2013), Team- und Organisationsstrukturen oder auch Netzwerkbilder von den Befragten erstellt werden. Im Gegensatz zu Bildern werden Fotos mithilfe einer Kamera produziert. Forschende nehmen Fotos entweder selbst auf, wenn sie beispielsweise eine Vor-Ort-Visitation durchführen oder eine Fotodokumentation einer Tagung für Evaluationszwecke erstellen, oder sie greifen auf Fotos zurück, die in Werbeanzeigen, Broschüren oder auf Webseiten abgebildet sind. Während selbst erstellte Bilder zumeist nur auf Papier vorliegen, stehen die meisten Fotos als digitale Bilddateien bereit.

### 27.2.2.2 Audio

Audioaufnahmen werden heutzutage mithilfe qualitativ hochwertiger digitaler Geräte erzeugt, zunehmend greifen Forschende aber auch auf Handys und Tablets zurück, deren Aufnahmequalität stetig zunimmt und für die inzwischen auch spezielle Apps zur Datenerhebung verfügbar sind, etwa die MAXApp ([www.maxqda.de](http://www.maxqda.de)). Eine Übersicht über Apps für die mobile und cloud-basierte Datenerhebung und -auswertung findet sich auf der Webseite der wöchentlichen Online-Zeitschrift für qualitative Forschung „The Qualitative Report“ ([www.nova.edu/ssss/QR/apps.html](http://www.nova.edu/ssss/QR/apps.html)). Am häufigsten werden in der qualitativen Sozialforschung Einzelinterviews, mit steigender Tendenz auch Gruppeninterviews aufgenommen. Mithilfe von Adaptern ist es möglich, Telefongespräche mitzuschneiden. Die

Aufnahmen liegen als digitale Audiodateien im MP3-, WMA- oder WAV-Format vor, problematisch sind proprietäre Formate der Aufnahmegeräte, wie etwa DSS, aufgrund ihrer geringen Qualität und der Schwierigkeit, derartige Dateien weiterzuverarbeiten.

### 27.2.2.3 Video

Auch bezüglich Videodaten ist aufgrund der fortgeschrittenen Digitaltechnik und großer Speichermedien ein Aufschwung zu beobachten, der sich unter anderem in neu erschienenen Publikationen niederschlägt (Corsten et al. 2010, Reichertz/Englert 2011). Videoaufnahmen zeigen häufig Gruppeninteraktionen, da sie den Vorteil bieten, neben der Sprache auch die Gestik, Mimik, Körperhaltung und die Platzierung im Raum festzuhalten. Während man früher üblicherweise auf selbst aufgenommene Videodaten angewiesen war, besteht seit wenigen Jahren auch die Möglichkeit, auf Videodaten im Internet zurückzugreifen (Schmitz/Yanenko, Kapitel 65 in diesem Band). Portale wie YouTube hosten Millionen an Videoclips, die für nicht-wissenschaftliche Zwecke produziert wurden, aber für die Forschung relevant sein können. Zudem finden sich im Kontext von Forschung entstandene Videoclips im Netz, die für Sekundäranalysen bereitgestellt werden. Unter [www.healthtalkonline.org](http://www.healthtalkonline.org) finden sich sortiert nach Krankheitsbereichen mehr als 2.000 Videoclips und Audiobeiträge (einschließlich der zugehörigen Transkripte) von Menschen, die über ihre Erfahrungen und ihren Umgang mit den jeweiligen Krankheiten berichten (Seale 2013). Problematisch ist insbesondere bei Videoaufnahmen, zum Teil auch bei Fotos, die Einhaltung des Datenschutzes (Zweidler 2013, Mühlichen, Kapitel 4 in diesem Band), denn häufig besteht der Wunsch Fotos von Szenen in Ergebnisberichte zu übernehmen.

### 27.2.3 Kombinierte Daten

Wenngleich wir hier zwischen textförmigen und Multimedia-Daten unterscheiden, muss angemerkt werden, dass diese beiden Datenarten häufig auch in kombinierter Form auftreten. So finden sich in vielen Dokumenten wie Prospekten, Volkshochschulprogrammen, Werbeanzeigen oder Gesundheitsbroschüren sowohl Texte als auch Fotos. Diese können zwar beide getrennt voneinander analysiert werden, ihr Informationspotenzial wird man aber erst bei der gemeinsamen Analyse beider Datenformen ausschöpfen können.

---

## 27.3 Aufbereitung von Daten für die Analyse mit QDA-Software

Die Fortschritte der Computertechnik und die damit zusammenhängende Möglichkeit, auch riesige Datenmengen zu speichern und der Analyse zugänglich zu machen, haben den für die Aufbereitung der Daten notwendigen Aufwand erheblich reduziert. Dennoch gibt es auch heute noch Bereiche, in denen man gespannt auf weitere Fortschritte wartet: So ist nach wie vor die automatische Transkription von offenen Interviews ein Desiderat. Art und Ausmaß der Datenvorbereitung differieren je nach Form der Daten und der

angestrebten Art der Analyse. Soweit Daten bereits in digitalisierter Form vorliegen, ist der Aufbereitungsaufwand eher gering. Wenn jedoch für die angestrebte Analyseform eine Transkription erforderlich ist, beträgt der Aufwand ein Vielfaches.

### 27.3.1 Texte und Dokumente

Textförmige Daten wie beispielsweise Parlamentsreden, Parteiprogramme, Schulinspektionsberichte, Selbstreporte im Rahmen von Evaluationen, Feldprotokolle, Methodenberichte und Ähnliches liegen meist in Form von PDF-Dokumenten vor (Salheiser, Kapitel 62 in diesem Band). In der Regel ist bei solchen Dateien keine weitere Vorarbeit erforderlich, sondern die PDF-Dateien können einfach so wie sie sind in das Datenset übernommen werden. Das PDF-Format, dessen Abkürzung für Portable Document Format steht, ist ein Dokumentenformat, das es erlaubt ein Dokument immer in genau dem Format anzusehen und auszudrucken, wie der Autor oder die Autorin es festgelegt hat. Änderungen eines solchen Dokuments durch den Leser oder die Leserin sind deshalb auch faktisch nicht möglich. Eine PDF-Datei kann ein sehr komplexes Layout besitzen, bspw. aus mehreren Spalten bestehen, Bilder, Grafiken und Tabellen beinhalten.

Anders verhält es sich mit Dateien im DOC- bzw. DOCX-Format, einem von Microsoft für Dokumente entwickelten Format, das standardmäßig von Textverarbeitungsprogrammen wie WORD verwendet wird. Solche Dateien lassen sich editieren. Die Wahl von Schriftart, Schriftgröße, Absatzformatierung und Seitengestaltung erfolgt durch den Benutzer. Bei solchen Dokumenten kann es sich empfehlen eine gewisse Vereinheitlichung und Standardisierung aller Rohdaten eines Forschungsprojekts durchzuführen. Analysiert man bspw. Schulaufsätze, die als DOC-Dateien vorliegen, so ist eine Vereinheitlichung in Bezug auf Schriftart, Schriftgröße etc. vor Beginn der Auswertung durchaus sinnvoll. Auch die Überprüfung auf orthografische Fehler und ggf. ihre Verbesserung sollte bereits vor Beginn der eigentlichen Auswertung erfolgen.

Wann immer man die Wahl hat, sollte man Dateien im DOC-Format solchen im PDF-Format vorziehen: DOC-Dateien eröffnen bessere Möglichkeiten zur lexikalischen Suche und automatischen Codierung, denn sie besitzen eine offen liegende und leicht veränderbare Absatzstruktur, die es erlaubt, gezielt nach dem gemeinsamen Vorkommen von Wörtern im gleichen Absatz – oder in definierter Absatz-Entfernung – zu suchen.

### 27.3.2 Vorstrukturierte textförmige Daten

Viele der in der empirischen Sozialforschung verwendeten textförmigen Daten weisen bereits eine innere Strukturierung auf. Führt man eine aus offenen und geschlossenen Fragen bestehende Online-Befragung durch (Wagner/Hering, Kapitel 48 in diesem Band), so erzeugen die entsprechenden Softwaretools eine rechteckige Datenmatrix wie die in Abb. 27.1 dargestellte, die von LimeSurvey.org, einer Open Source Software zur Datenerhebung mittels Onlineumfragen, erstellt wurde.

id	ErsteTN	Gliederung	FreundlDoz	KommentarLehrmaterialien	GutGefallen	
1	1	4	4	Im Vergleich zwischen Powerpoint-Folien und Lehrbuch Kombination aus Vorlesung und Übung, dienstags die I		
2		4	4	Zu viele Fehler in der aktuellen Auflage, aber ansonsten Hinweise auf die Praxis Lösungen der Aufgaben wurden		
3	1	4	4	Für mich war es nicht nachvollziehbar, wie sich bei grÜbungen		
4	1	4	4	Die Powerpoint Präsentationen habe ich als sehr hilfreich. Das es zu jeder Vorlesung Übungen gibt.		
5	1	4	3	Die Lehrmaterialien sind verständlich geschrieben.	Die Übungsstunden dienstags.	
6	1	4	4	Gute Struktur der Veranstaltung, da in jeder Sitzung eDie vielen Angebote der Unterstützung in Form von T		
7	1	4	4	Tutorium: Ich selber wäre sehr gerne zum Tutorium gegang. Die klar geglieederter Vorlesungen die sehr gut überreir		
8		4	4	Die Veranstaltung am Dienstag ist hilfreicher als die VDie Prof./ Dozenten/ Mitarbeiter sind sehr bemüht, de		
9	1	3	3	Die Übungszettel helfen mir sehr das Theoretische beDas Dienstagsseminar mit Tests und anschließender		
10	1	4	3	Das Lehrangebote ist entsprechend der Bedürfnisse uDie Übungszettel, da man sich und seinen Lernerfolg		

**Abb. 27.1** Auszug einer mithilfe von LimeSurvey erstellten Datenmatrix

Man kann in Abb. 27.1 erkennen, dass die Datei neben Spalten mit numerischen Variablen auch Spalten mit textförmigen Informationen enthält. Hierbei handelt es sich um Antworten auf die offen gestellten Fragen des Online-Surveys, welche die Befragten selbst in die entsprechenden Freitextfelder der Umfrage eingetippt haben. Diese Vorab-Strukturierung der Daten kann man sich für die spätere Auswertung zu Nutzen machen. Dort, wo eine solche Strukturierung von Daten in Form von Fragen und dazu gehörenden Antworten zwar besteht, aber noch nicht in digitaler Form abgebildet ist, lässt sich eine solche Struktur bei der Aufbereitung der Daten erzeugen. Dies geschieht am besten in Form einer Excel-Tabelle, die wie im obigen Beispiel die entsprechenden Fragen in der ersten Zeile und in der jeweiligen Spalte alle Antworten aller Befragten auf eben diese Frage enthält. QDA-Software wie bspw. MAXQDA ist in der Lage, alle zu einer Frage gehörenden Antworten automatisch zu codieren und ggf. in einem mehrstufigen Prozess weiter manuell oder automatisch zu bearbeiten. In MAXQDA wird eine so vorbereitete Excel-Datei dann automatisch aufgeteilt: Numerische Daten werden als Variable in die Datei der Fall-Variablen eingefügt und entsprechend gekennzeichnete Spalten mit textförmigen Daten werden als qualitative Daten eingefügt und vorcodiert, wobei der in der Kopfzeile enthaltene Text – zumeist ein Kürzel der gestellten Frage – als Kategoriename verwendet wird.

Im Falle von Blogs oder Foren ist die Vorbereitungsarbeit in der Regel umfangreicher als bei aus Surveys stammenden Antworttexten. Hier ist zunächst zu entscheiden, ob jeder Forenbeitrag als Fall definiert werden soll. Dies ist dann sinnvoll und notwendig, wenn die einzelnen Beiträge codiert und anschließend nach Codes ausgezählt werden sollen. Wenn es lediglich um die Identifizierung von Denkfiguren, Argumenten und Diskursen geht, ohne dass eine fallbezogene quantitative Auswertung stattfindet, ist das Splitten der Forenbeiträge nicht erforderlich, sondern man kann mit dem Gesamtkorpus arbeiten. Mit dem Text-Preprocessor von MAXQDA lässt sich eine aus dem Web kopierte Datei von Foren- oder Blogbeiträgen für eine fallbezogene Auswertung vorbereiten. Hierzu ist eine vorgesetzte Syntax einzuhalten, wobei zu Beginn jedes Beitrags das Schlüsselwort „#TEXT“ einzufügen ist. In Abb. 27.2 lässt sich erkennen, dass hinter diesem Schlüsselwort der Name des Verfassers des Blog- bzw. Foreneintrags gefolgt vom Datum steht. Wie

```
#TEXTantares56
```

20.06.2013 10:25 UHR

Wir haben schon gemerkt, wie Angela vor Obama gekrochen ist. Wir haben aber auch nichts anderes von ihr erwartet!

```
#TEXTkarmen
```

20.06.2013 10:17 UHR

Ja, die „Netzgemeinde“ ... immer cool und schön selbstgenügsam ... Sie definiert sich über ein Medium, statt es mit Inhalten zu füllen hält sie es für DEN Inhalt schlechthin. Ganz hohl aber immer cool dabei.

**Abb. 27.2** Auszug aus einer Datei mit mehreren Texten für den automatisierten Import

viel Arbeitsaufwand bei der Vorbereitung tatsächlich nötig wird, hängt natürlich von den technischen Gegebenheiten des jeweiligen Forums ab, das man analysieren möchte. In jedem Fall sollte direkt hinter dem Schlüsselwort „#TEXT“ sofort der Name der Verfasser\_in des Beitrags stehen, denn dieses wird als Textname interpretiert. In diesem Beispiel werden also ein Text namens „antares 56 ...“ und ein Text namens „karmen“ erzeugt.

Funktionen zur Vorab-Strukturierung von qualitativen Daten können auch genutzt werden, um Teile eines Textes bereits beim Einlesen automatisch einem Code bzw. einer Kategorie zuordnen. Dies ist bei der Aufbereitung von herkömmlichen offenen Fragen im Papierfragebogen sinnvoll. Abb. 27.3 zeigt einen Auszug einer entsprechend vorbereiteten Datei mit Antworttexten aus einem Fragebogen. Das erste eingefügte Schlüsselwort „#CODE“ bewirkt, dass der folgende Textabschnitt bis zum Schlüsselwort „#ENDCODE“ mit dem Code „Positives Feedback“ codiert wird und der folgende Abschnitt hinter „#CODE“ mit dem Code „Verbesserungsvorschläge“.

### 27.3.3 Visuelle Daten: Bilder, Audio und Videos

Sofern diese Daten ohne vorherige Transkription analysiert werden sollen, sind keine weiteren Vorbereitungsarbeiten erforderlich. Man muss sich allerdings über den Umfang der Daten, deren Speicherung und die Dateiformate Gedanken machen. Auch die *Dropbox* bietet kostenfrei derzeit nur 6 GB Speicherplatz, ein zehnminütiges Video in HD-Qualität kann jedoch schon mehrere 100 MB ausmachen. Bei Bildern und Grafiken besteht die Möglichkeit, durch Veränderung der Auflösung die Größe der Datei erheblich zu verkleinern. Hiervon sollte man normalerweise Gebrauch machen.

#TEXTFragebogen 1

#CODEPositives Feedback

Die Veranstaltung war durchgängig gut strukturiert und klar aufgebaut. Der Reader war sehr hilfreich.

#ENDCODE

#CODEVerbesserungsvorschläge

Die Übungszettel sollten frühzeitiger im Netz verfügbar sein.

#ENDCODE

#TEXTFragebogen 2

#CODEPositives Feedback

Die Verbindung der zahlreichen verwendeten Medien sorgt für ein besseres Verständnis der Thematik

#ENDCODE

#CODEVerbesserungsvorschläge

Ich würde es begrüßen, wenn zu den Übungsaufgaben detailliertere Lösungen mit Querverweisen auf das Buch vorliegen würden.

#ENDCODE

**Abb. 27.3** Auszug aus einer Datei mit vorcodierten Textabschnitten

#### 27.3.4 Transkription von Audio- und Videodaten

Qualitative Interviews und Gruppendiskussionen bzw. Fokusgruppen werden üblicherweise als Audio- oder Videoaufzeichnung festgehalten. Es sind allerdings nicht diese Rohdaten, die ausgewertet werden, sondern deren Verschriftlichung, d.h. die Audio- und Videodaten stellen nur eine Art Zwischenprodukt dar, auf das ggf. noch einmal zurückgegriffen werden kann, wenn man es bei der Analyse für sinnvoll hält, auf die O-Töne zurückzugreifen.

Audio- oder Videoaufnahmen zu transkribieren ist eine zeitaufwändige Angelegenheit, bei der zunächst zu entscheiden ist, wie detailliert und nach welchen Regeln transkribiert werden soll. Es existieren zahlreiche, mehr oder weniger komplexe Transkriptionssysteme, die an anderer Stelle ausführlich beschrieben sind (Kuckartz 2012). Der Hauptunterschied zwischen den Transkriptionssystemen besteht in der Art und Weise, wie verschiedene Textmerkmale (bspw. Betonungen, Lautstärke, Sprechpausen, Überlappungen zwischen den Äußerungen verschiedener Sprecher, Dialektfärbungen und paraverbale Äußerungen) in der Transkription berücksichtigt werden.

1. Es wird wörtlich transkribiert, also nicht lautsprachlich oder zusammenfassend. Vorhandene Dialekte werden nicht mit transkribiert, sondern möglichst genau in Hochdeutsch übersetzt.
2. Sprache und Interpunktionszeichen werden leicht geglättet, d.h. an das Schriftdeutsch angelehnt. Zum Beispiel wird aus „Er hatte noch so'n Buch genannt“ -> „Er hatte noch so ein Buch genannt“. Die Satzform, bestimmte und unbestimmte Artikel etc. werden auch dann beibehalten, wenn sie Fehler enthalten.
3. Deutliche, längere Pausen werden durch in Klammern gesetzte Auslassungspunkte (...) markiert. Entsprechend der Länge der Pause in Sekunden werden ein, zwei oder drei Punkte gesetzt, bei längeren Pausen wird eine Zahl entsprechend der Dauer in Sekunden angegeben.
4. Besonders betonte Begriffe werden durch Unterstreichungen gekennzeichnet.
5. Sehr lautes Sprechen wird durch Schreiben in Großschrift kenntlich gemacht.
6. Zustimmende bzw. bestätigende Lautäußerungen der Interviewer (mhmm, aha etc.) werden nicht mit transkribiert, sofern sie den Redefluss der befragten Person nicht unterbrechen.
7. Einwürfe der jeweils anderen Person werden in Klammern gesetzt.
8. Lautäußerungen der befragten Person, die die Aussage unterstützen oder verdeutlichen (etwa Lachen oder Seufzen), werden in Klammern notiert.
9. Absätze der interviewenden Person werden durch ein „I:“, die der befragten Person(en) durch ein eindeutiges Kürzel, z.B. „B4;“, gekennzeichnet.
10. Jeder Sprechbeitrag wird als eigener Absatz transkribiert. Sprecherwechsel wird durch zweimaliges Drücken der Enter-Taste, also einer Leerzeile zwischen den Sprechern deutlich gemacht, um so die Lesbarkeit zu erhöhen.
11. Störungen werden unter Angabe der Ursache in Klammern notiert, z.B. (Handy klingelt).
12. Nonverbale Aktivitäten und Äußerungen der befragten wie auch der interviewenden Person werden in Doppelklammern notiert, z.B. ((lacht)), ((stöhnt)) und Ähnliches.
13. Unverständliche Wörter werden durch (unv.) kenntlich gemacht
14. Alle Angaben, die einen Rückschluss auf eine befragte Person erlauben, werden anonymisiert.

**Abb. 27.4** Transkriptionsregeln für die computerunterstützte Auswertung (Kuckartz 2012: 136)

Wie detailliert transkribiert wird, hängt nicht zuletzt von finanziellen Gesichtspunkten ab, denn die Transkription ist zeitaufwändig und verursacht erhebliche Kosten. Selbst für einfache Transkriptionen benötigt man etwa das fünf- bis zehnfache der Interviewzeit. Für die meisten qualitativen Forschungsprojekte reichen relativ einfache Transkriptionsysteme wie das in Abb. 27.4 dargestellte völlig aus. Dieses wurde ursprünglich im Rahmen eines Evaluationsprojektes entwickelt, ist einfach und schnell erlernbar und wurde inzwischen in vielen Projekten mit Erfolg verwendet.

Sehr hilfreich ist die heute durch QDA-Software gebotene Möglichkeit zur synchronen Betrachtung von Transkription und zugrundeliegenden Rohdaten, also Ton bzw. Bild. Während der Originalton läuft, besteht die Möglichkeit ähnlich wie bei einem Teleprompter

die verschriftlichte Fassung zu lesen. Auf diese Weise lässt sich die Güte der Transkription quasi „on the fly“ überprüfen. Darüber hinaus lassen sich während der Transkription Zeitmarken einfügen, die es später erlauben, an interessierenden Textstellen auf die Originalaufnahme zurückzugreifen.

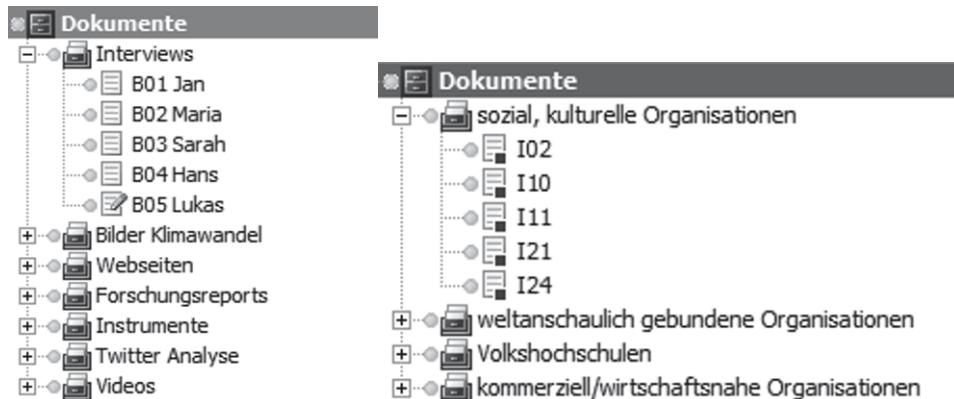
---

## 27.4 Datenimport in die QDA-Software, Datenbereinigung und Datenmanagement

Nachdem die qualitativen Daten aufbereitet und überprüft wurden, können sie im nächsten Schritt in die QDA-Software importiert werden und von diesem Zeitpunkt an mit Hilfe der QDA-Software verwaltet werden. Beim Importprozess sind zahlreiche Entscheidungen zu treffen, die unter anderem die Speicherorte und Speicherverfahren sowie die Untergliederung der Daten betreffen. QDA-Software bietet hier zahlreiche Möglichkeiten und Flexibilität in der Gestaltung, doch lassen sich einige Entscheidungen im Nachhinein meist nur bedingt revidieren und sollten daher bewusst und mit Blick auf die Konsequenzen getroffen werden. Dies gilt vor allem, wenn mehrere Personen eines Forschungsteams die Daten gemeinsam analysieren. Mit QDA-Software lassen sich unterschiedliche Datenformate auswerten: Texte im weit verbreiteten Word-Format, Tabellen im Excel-Format, PDF-Dateien sowie Bilder in den üblichen Formaten JPG, GIF, PNG oder TIF. Die Fülle der heutzutage verwendeten Audio- und Videoformate ist sehr groß, wobei man auf Standardformate wie AVI, MPG oder WMV setzen sollte, die von QDA-Software meist problemlos gemanagt werden können.

In der QDA-Software wird zunächst ein neues Projekt angelegt, das als Container für die zu analysierenden Daten dient. Textförmige Daten, die im DOC/X- oder einem vergleichbaren Format für Textverarbeitungsprogramme vorliegen, nehmen meist nur einen geringen Speicherplatz in Anspruch und können direkt in die Projektdatei der QDA-Software integriert werden, so dass diese leicht gesichert oder per Mail verschickt werden kann. PDF-Dateien sind zwar speicherintensiver als DOC/X-Dateien, können bei kleiner Anzahl jedoch ebenfalls problemlos in die Projektdatei integriert werden. Eine sehr große Anzahl an PDF-Dateien, aber vor allem auch Multimedia-Daten erfordern hingegen ein anderes Speicherprinzip, weil sie unhandlich große Projektdateien erzeugen würden. In diesem Fall werden die großen Datenmengen an einem zentralen Ort auf der Festplatte oder im Netzwerk gespeichert und in die Projektdatei wird nur ein Verweis auf die extern gespeicherten Dateien übernommen. Der Import von Dateien oder auch ganzen Ordner von Dateien in die QDA-Software ist denkbar einfach. Aus dem Windows-Explorer oder dem Mac-Finder werden diese mit der Maus in die QDA-Software gezogen und dort abgelegt.

Für die weitere Verarbeitung der Daten ist es notwendig, eine sinnvolle Gliederung vorzusehen, um einen leichten Zugriff auf die Daten zu gewährleisten und um Differenzierungen bei Analysen vorsehen zu können. Diese Gliederung kann sich an unterschiedlichen Datenarten orientieren oder sie kann die Subgruppen widerspiegeln, die im Rahmen eines Forschungsprojekts befragt wurden (Abb. 27.5). Alternativ lassen sich die Daten



**Abb. 27.5** Untergliederung von Daten nach verschiedenen Formen (links) und nach Subgruppen (rechts) in der QDA-Software MAXQDA

nach Fällen, für die jeweils mehrere Dokumente vorliegen, oder nach unterschiedlichen Erhebungszeitpunkten gruppieren. Die QDA-Software erlaubt es, verschiedene Gliederungsoptionen für die Daten zu speichern und nach diesen Gruppierungen differenzierende Analysen vorzunehmen.

Es kommt häufig vor, dass die Daten bereits vorstrukturiert vorliegen. Für diesen Fall hält QDA-Software Import-Prozeduren bereit, welche den Forschenden erhebliche Arbeitserleichterung bringen und es erlauben, mit wenigen Klicks 1.000 Dokumente und mehr zu importieren und automatisch zu vercoden. Bibliographische Daten mit Literaturangaben und Abstracts lassen sich aus einer RIS-Datei einlesen, wobei jede Literaturangabe automatisch in ein Dokument umgewandelt wird. Antworten auf offene Fragen eines Fragebogens liegen meist in einer Datenmatrix vor, in der die Fälle die Zeilen bilden und die Antworten auf die offenen Fragen zusammen mit weiteren Variablenwerten in den Zeilen abgetragen sind. Bei dem Import solch strukturierter Daten in die QDA-Software wird für jeden Fall ein eigenes Dokument angelegt, die Freitextantworten werden automatisch mit der gestellten Frage codiert und die Variablenwerte zu jedem Fall in einer Datenmatrix gespeichert.

Neben den Primärdaten sind häufig auch ergänzende Informationen und Kontextinformationen interessant für die Analyse. Diese Zusatzinformationen können entweder als eigene Dokumente (z.B. Selbstbericht einer Organisation als PDF), als Memos zu einzelnen Fällen (z.B. Interviewprotokoll), als Variablen (z.B. forschungsrelevante Skalenwerte der Befragten) oder als Hyperlinks (z.B. auf die Homepage einer Organisation) in das QDA-Projekt integriert werden.

Sollen Interviews analysiert werden, empfiehlt es sich, nicht nur deren Verschriftlichungen in die QDA-Software zu importieren, sondern ergänzend auch die zugehörige Audio- oder Videoaufnahme in der Projektdatei bereitzustellen. Text und Media-Datei lassen sich mithilfe von Zeitmarken verknüpfen, so dass mit einem Klick der Originalton

aufgerufen werden kann. Diese Funktionalität ermöglicht es unter anderem Transkripte zu überprüfen und die genauen Betonungen von Wörtern und Sätzen überprüfen zu können.

QDA-Software ist primär als Analysesoftware und nicht als Programm zur Textverarbeitung konzipiert und stellt weniger Möglichkeiten der Texteditierung bereit. Deshalb sollten die Daten vor dem Import bereits bereinigt sein, sprich die Texte sollten im Hinblick auf Rechtschreibung und Zeichensetzung korrigiert und ggf. anonymisiert sein. Dies ist leistbar für eine überschaubare Anzahl an Texten, bei mehreren tausend Freitextantworten einer Online-Befragung jedoch kaum möglich. Bei diesen Daten ist es daher unter Umständen notwendig, Rechtschreibfehler zu korrigieren, wenn sie in der Software auffallen und für die weitere automatische Verarbeitung der Daten Fehlerfreiheit notwendig ist. Bei Audio- und Videodaten, aber auch bei Bildern und Fotos ist meist keine Bereinigung, höchstens eine Anonymisierung, notwendig.

Nach dem Import in die QDA-Software haben die Forschenden leichten Zugriff auf die Daten. Per Doppelklick lassen sich die importierten Text- und Tabellendokumente, Bilder, PDF- und Media-Dateien öffnen und betrachten, sie können ggf. editiert und durchsucht werden.

---

## 27.5 Mögliche Analyseformen

Die Aufbereitung der Daten sollte immer schon mit Blick auf die spätere Auswertung geschehen. Würde man beispielsweise eine Auswertung planen, die sich an der klassischen quantitativen Inhaltsanalyse (Rössler 2010) orientiert, so wären die Daten so aufzubereiten, dass normalerweise eine Untergliederung in Texteinheiten vorzusehen ist. Solche quantitativen Inhaltsanalysen, bei denen die Vercodung auf der Basis eines Diktionärs automatisch vorgenommen wird, sind allerdings heute in der Forschungspraxis nicht sehr häufig anzutreffen.

Unter den Verfahren zur Analyse qualitativer Daten dominieren die *qualitative Inhaltsanalyse* und die *Grounded Theory*. Die Grounded Theory (Mey/Mruck 2011, Grounded Theory Reader, Strübing, Kapitel 32 in diesem Band) ist allerdings eher ein Forschungsstil als ein Auswertungsverfahren im engeren Sinne. Charakteristisch für die Grounded Theory ist das Ziel der Theoriegenerierung, wobei in einer mehrstufigen Vorgehensweise von der offenen Codierung des Materials bis hin zur Identifikation von Schlüsselkategorien vorangeschritten wird. Eine systematische Codierung des gesamten Materials ist für die Theoriegenerierung nicht erforderlich – stattdessen gilt das Kriterium der „theoretischen Sättigung“ (Strauss/Corbin 1996: 159), welche dann erreicht ist, wenn die Erhebung zusätzlicher Daten und die weitere Auswertung von Kategorien keinen weiteren Erkenntniszuwachs erwarten lassen, d.h. die Kategorie ist genügend ausgearbeitet, die Beispiele wiederholen sich, Neues und Differenzierteres wird nicht mehr entdeckt.

Anders verhält es sich bei der *qualitativen Inhaltsanalyse* (Kuckartz 2012, Mayring 2010, Schreier 2012, Mayring/Fenzl, Kapitel 38 in diesem Band), die nicht per se mit dem Ziel der Theoriegenerierung verknüpft ist. Als Auswertungsverfahren lässt sich die qualitative

Inhaltsanalyse sowohl zur Theoriegenerierung als auch zur Theorieüberprüfung verwenden – zudem kann sie auch bei rein deskriptiver Zielsetzung eingesetzt werden. Im Mittelpunkt der qualitativen Inhaltsanalyse stehen wie bei der klassischen Content Analysis (vgl. Krippendorff 2004) die Kategorien, die sowohl deduktiv als auch induktiv oder in kombiniert deduktiv-induktiver Vorgehensweise gebildet werden können. Inhaltlich relevante Teile des Textes (oder des Bildes, Videos, Audios bei anderen Datenarten) werden Kategorien zugeordnet, wobei das gesamte Datenmaterial systematisch durchgearbeitet und codiert wird. Sehr weit verbreitet sind Formen inhaltlich strukturierender Inhaltsanalyse (Kuckartz 2012, Mayring 2010, Mayring/Fenzl, Kapitel 38 in diesem Band), wobei in den meisten Fällen mit thematischen Kategorien und Subkategorien gearbeitet wird. Moderne QDA-Software erlaubt eine Vielzahl von Auswertungsformen, mit denen sich etwa der Zusammenhang von Kategorien qualitativ und quantitativ darstellen lässt. Eine besondere Bedeutung besitzen hierbei visuelle Darstellungen, die von Diagrammen des gleichzeitigen Vorkommens von Kategorien bis hin zu Konzept-Maps zur Visualisierung von Wirkungszusammenhängen reichen.

Interessante Analysemöglichkeiten ergeben sich auch bei Mixed Methods-Projekten (Kelle, Kapitel 8 in diesem Band), bei denen sowohl qualitative als auch quantitative Daten erhoben wurden. Hier können – so die Daten entsprechend aufbereitet wurden – sogenannte „Joint Displays“ (Creswell/Plano-Clark 2010: 226) erstellt werden, bei denen die Resultate beider Analysestränge in einer gemeinsamen Tabelle oder Visualisierung dargestellt werden.

## Literatur

- Behnken, Imbke/Zinnecker, Jürgen (2013): Narrative Landkarten. In: Friebertshäuser, Barbara/Langer, Antje/Prengel, Annedore (Hg.): Handbuch qualitative Methoden in der Erziehungswissenschaft. Weinheim/München: Juventa. 547-562
- Corsten, Michael/Krug, Melanie/Moritz, Christine (2010): Videographie praktizieren. Wiesbaden: VS Verlag
- Creswell, John W./Plano Clark, Vicki L. (2010): Designing and Conducting Mixed Methods Research. Thousand Oaks (CA.): Sage Publications
- Flick, Uwe (2004): Triangulation. Wiesbaden: VS Verlag
- Krippendorff, Klaus (2004): Content Analysis. Thousand Oaks (CA.): Sage Publications
- Kuckartz, Udo (2012): Qualitative Inhaltsanalyse. Weinheim: Beltz
- Mayring, Philipp (2010): Qualitative Inhaltsanalyse. Weinheim: Beltz
- Mey, Günter/Mruck, Katja (Hg.) (2011): Grounded Theory Reader. Wiesbaden: VS Verlag
- Reichert, Jo/Englert, Carina (2011): Einführung in die qualitative Videoanalyse. Wiesbaden: VS Verlag
- Rössler, Patrick (2010): Inhaltsanalyse. Konstanz: UVK
- Sachs-Hombach, Klaus (2005): Bildwissenschaft. Frankfurt/Main: Suhrkamp
- Schreier, Margrit (2012): Qualitative Content Analysis in Practice. London: Sage
- Schulze, Theodor (2013): Bildinterpretation in der Erziehungswissenschaft. In: Friebertshäuser, Barbara/Langer, Antje/Prengel, Annedore (Hg.): Handbuch qualitative Methoden in der Erziehungswissenschaft. Weinheim/München: Juventa. 529-546
- Seale, Clive (2013): Secondary analysis of qualitative data. Vortrag auf der Tagung für Computergestützte Analyse Qualitativer Daten zum Thema Methoden, Innovationen, Praxis. Marburg. Zugriff am 07.03.2013. [http://caqd.de/images/stories/caqd13/Folien/Seale\\_CAQD2013\\_Keynote.pdf](http://caqd.de/images/stories/caqd13/Folien/Seale_CAQD2013_Keynote.pdf) am 22.08.13
- Strauss, Anselm L./Corbin, Juliet M. (1996): Grounded Theory. Weinheim: Beltz
- Zweidler, Reinhard (2013): Evaluationsstandards und Datenschutz, Nutzung von Bildmaterial. Vortrag auf der Frühjahrstagung des Arbeitskreises Aus- und Weiterbildung in der Evaluation in der Gesellschaft für Evaluation zum Thema „Datenschutz in der Evaluation.“ Bonn 03.05.2013. [http://www.degeval.de/images/stories/Arbeitskreise/AK\\_AUWE/AKAUWE\\_FJT\\_2013/Zweidler\\_Vortrag\\_DeGEval\\_3.5.2013.pdf](http://www.degeval.de/images/stories/Arbeitskreise/AK_AUWE/AKAUWE_FJT_2013/Zweidler_Vortrag_DeGEval_3.5.2013.pdf), Zugriff am 11.06.13

Detlev Lück und Uta Landrock

Eine erfolgreiche Analyse von Daten setzt neben einer intelligenten und gewissenhaften Datenerhebung auch die sorgfältige und gründliche Bereinigung und Aufbereitung der Daten voraus. Dieser Schritt mag trivial erscheinen. Doch er ist technisch durchaus anspruchsvoll und hält Fehlerquellen bereit. Er sollte in seiner Bedeutung und in seinem Arbeitsaufwand schon deswegen nicht unterschätzt werden, weil Fehler in der Datenaufbereitung und Datenbereinigung in der Lage sind, alle Befunde zu verfälschen, die je mit den entsprechenden Daten generiert werden.

Im Folgenden erläutern wir zunächst, in welcher Form Daten vorliegen müssen, um mittels Statistiksoftware ausgewertet werden zu können. Wir erläutern, auf welchem Wege die im Zuge einer standardisierten Erhebung erfassten Daten in diese Form gelangen. Wir stellen die Arbeitsschritte vor, die notwendig sind, um die Daten – insbesondere auch für Sekundärnutzer – nutzbar zu machen. Und wir stellen mögliche Strategien dar, mit Hilfe derer eventuelle Fehler, die sich möglicherweise im Zuge des Datenerhebungsprozesses eingeschlichen haben, in den Daten identifiziert und korrigiert werden können. Wir erläutern die Vorgehensweisen grundsätzlich exemplarisch für die Statistiksoftware SPSS und verweisen vereinzelt auf Gemeinsamkeiten und Unterschiede in anderen Programmen.

---

## 28.1 Aufbau einer Datenmatrix

Für die statistische Analyse werden Daten in Form einer *Datenmatrix* organisiert. Diese ist in den meisten Statistikprogrammen (u.a. Stata, SPSS, SAS, Excel) auch als solche sichtbar. Darin werden typischerweise Merkmalsträger oder auch Fälle in Zeilen und Merkmale bzw. Variablen in Spalten aufgelistet. In den Zellen sind dann die Merkmalsausprägungen erfasst: die Ausprägung, die einen bestimmten Merkmalsträger hinsichtlich eines bestimmten Merkmals kennzeichnet (zur Darstellung von Netzwerkdaten siehe Baur, Kapitel 74 in diesem Band).

1 : v1		4615						
	v1	v2	v3	v4	v5	v6	v7	v6
1	4615	1	2	1	2	1	10	
2	4615	2	1	1	1	1	10	
3	4615	3	2	1	1	1	4	
4	4615	4	2	2	2	1	6	
5	4615	5	1	1	1	1	7	
6	4615	6	2	1	1	1	6	
7	4615	7	2	2	1	1	0	

**Abb. 28.1** Allbus 2012 im SPSS-Datenfenster

Als Beispiel dient uns der SPSS-Daten-Editor, der den Allbus 2012 zeigt (Abb. 28.1). Aus Platzgründen sind nur die ersten sechs Fälle und die ersten sieben Variablen dargestellt. Merkmalsträger sind hier einzelne Personen, die im Zuge der Datenerhebung befragt wurden. Jede der Zeilen steht für eine solche Person. Liest man die Werte einer Zeile, erhält man die Informationen, die über eine bestimmte Person vorliegen – also im Wesentlichen das, was sie im Interview geantwortet hat. Aus Gründen des Datenschutzes sind die Befragten allerdings nicht namentlich erfasst. Stattdessen erhält jeder von ihnen eine anonyme Identifikationsnummer. Diese Nummer ist in Form der Variablen V2 erfasst: Der erste Befragte erhält die Nummer 1, der zweite die Nummer 2 usw.

Jede Variable entspricht einer Spalte. Liest man die Werte einer Spalte, erhält man beispielsweise alle vergebenen Identifikationsnummern oder alle Antworten, die die unterschiedlichen Personen auf eine bestimmte Frage gegeben haben. Neben der Identifikationsnummer und einigen anderen technischen Informationen (z.B. der Studiennummer V1) sind als Variablen alle Merkmale erfasst, die im Zuge des Interviews erfragt wurden. Die Variable V6 etwa entspricht der Frage, ob der Befragte die deutsche Staatsangehörigkeit hat oder nicht. Der Wert 1 signalisiert *Ja*, der Wert 2 signalisiert *Nein*. Die vom Interviewer selbst auf einer Skala von 1 bis 11 einzuschätzende Attraktivität des Befragten ist in Variable V7 gespeichert. Demnach sind die ersten beiden im Allbus 2012 befragten Personen sehr attraktiv gewesen. Welcher Variablenname welcher Frage entspricht und welcher Wert für welche Antwort steht, ist im Codeplan oder Codebuch definiert, auf das wir später noch eingehen werden (vgl. den Abschnitt „Datendokumentation“).

Die Merkmalsträger müssen keine Individuen sein. Es können auch für Paare, Haushalte, Organisationen oder Länder jeweils bestimmte Merkmale erfasst und in dieser Form organisiert werden. Dann entspricht eine Zeile im Datensatz beispielsweise einem Haushalt. Die Variablen wiederum müssen nicht (genau) einer Frage in einem Fragebogen entsprechen. Beispielsweise können auch technische Daten eines Interviews (z.B. Interviewdauer), Einschätzungen durch den Interviewer (z.B. Attraktivität des Befragten),

biologische Merkmale einer Person (z.B. Körpergröße), technische Merkmale eines Fahrzeugs, volkswirtschaftliche Daten eines Landes oder die Anzahl der Nennungen des Begriffs „Krise“ in den verschiedenen Ausgaben einer Zeitung jeweils als Variablen gespeichert werden. Auch wird häufig von einer Frage mit Mehrfachantwortmöglichkeit in einem Interview jede mögliche Antwort jeweils als eine eigene binäre Variable (mit den Ausprägungen *trifft zu* und *trifft nicht zu*) erfasst.

---

## 28.2 Variablencodierung, Variablenformate und Datenformate

Die Ausprägung einer Variablen, die in eine Zelle der Datenmatrix geschrieben wird, kann in absoluten Zahlen, Dezimalzahlen, Prozentwerten, Buchstaben, Wörtern oder auch beliebigen Zeichenfolgen bestehen. Man denke nur an das Merkmal *Sozialversicherungsnummer*. Eine Grenze stellen grundsätzlich nur die Tastatur, der auf dem Computer installierte Zeichensatz sowie die Software dar, die möglicherweise bestimmte Sonderzeichen nicht erkennt.

Allerdings sehen die gängigen Statistikprogramme vor, dass für jede Variable ein bestimmtes *Variablenformat* definiert werden muss. Diese – von Software zu Software leicht unterschiedlich vordefinierten – Variablenformate erkennen jeweils nur bestimmte Arten von Ausprägungen. In SPSS gibt es unter anderem das Format *String*, das für Text gedacht ist und relativ beliebige Aneinanderreihungen von Buchstaben und Ziffern akzeptiert. Weitaus gängiger ist jedoch das Format *Numerisch*, das ausschließlich Zahlenwerte zulässt und in dessen Rahmen auch eine bestimmte maximale Anzahl von Stellen und von Nachkommastellen definiert werden muss.

Die Frage, welcher Eintrag in der Datenmatrix welche Information repräsentiert, ist die der *Variablencodierung*. Dass eine Merkmalsausprägung als Zahlenwert erfasst, also numerisch codiert wird, drängt sich in einem üblichen standardisierten Interview nur bei wenigen Merkmalen auf, und zwar bei metrischen Variablen, wie etwa die Antwort auf die Frage, wie viele Kinder man habe. Will man auf Basis dieser Angaben statistische Auswertungen vornehmen, beispielsweise eine durchschnittliche Kinderzahl berechnen, so ist dies technisch nur möglich, wenn die Angaben als Zahlenwerte vorliegen (und nicht etwa in Form der Wörter „eins“, „zwei“, „drei“ etc.). Sehr häufig werden von Befragten in Interviews Einschätzungen auf einer Skala (Franzen, Kapitel 51 in diesem Band) von 1 („stimme voll und ganz zu“) bis 5 („stimme überhaupt nicht zu“) gefordert, etwa die Zustimmung zu der Aussage „Die Aufgabe des Mannes ist es, Geld zu verdienen, die der Frau, sich um Haushalt und Familie zu kümmern“. Diese Skalen werden von vielen Forschern als ordinal angesehen, so dass von ihnen ohnehin nur Anteilswerte ausgewiesen werden (zur Unterscheidung von binären, nominalen, ordinalen und metrischen Variablen siehe: Blasius/Baur, Kapitel 79 in diesem Band). Doch sie werden auch von vielen Forschern als metrisch interpretiert und für entsprechende multivariate Auswertungsverfahren verwendet, so dass sich auch hier eine numerische Codierung anbietet.

Das gleiche gilt für binäre Variablen wie das Geschlecht. Binäre Variablen eignen sich für viele statistische Verfahren ebenso wie metrische Variablen, werden von der Statistiksoftware aber meist nur akzeptiert, wenn ihre Ausprägungen als Zahlenwerte erfasst sind. Für nominale Merkmale mit mehr als zwei Ausprägungen entfällt dieses Argument. Sie können – ohne zuvor in binäre Variablen umcodiert worden zu sein – in den meisten multivariaten Analysen nicht berücksichtigt werden. So könnten beispielsweise Antworten auf die Frage nach dem Familienstand in Form der Wörter „ledig“, „verheiratet“, „eingetragene Lebenspartnerschaft“, „geschieden“ und „verwitwet“ erfasst werden, ohne dass sich die Möglichkeiten zu ihrer Analyse nennenswert einschränken würden. Dennoch wird in aller Regel auch eine solche Information numerisch codiert – schon deshalb, weil so das Erfassen der Antworten schneller und weniger fehleranfällig ist. Im Zweifel werden dazu die vorgegebenen Antwortkategorien durchnummeriert.

Da verschiedene Statistikprogramme neben den Variablennamen und Merkmalsausprägungen noch andere Informationen vorsehen, speichern und verarbeiten, legt jedes Programm Daten in einem bestimmten *Datenformat* an. SPSS zum Beispiel speichert seinen Daten-Editor im Format *sav*, Stata im Format *dta*. Grundsätzlich ist aber jedes gängige Programm in der Lage, die Datenformate anderer gängiger Programme in das eigene Format zu konvertieren.

---

## 28.3 Wie kommen die Daten in den Datensatz?

Um Daten, also beispielsweise die Antworten im Zuge einer standardisierten Befragung, in die Form einer Datenmatrix in einem bestimmten Datenformat zu bekommen, gibt es mehrere Wege. Üblich ist es heute, dass bereits der Interviewer während des Interviews die Antworten des Befragten *an einem Computer digital erfasst*. Dabei führt er entweder telefonische Interviews (Hüfken, Kapitel 46 in diesem Band) in einem Telefonstudio (computer-assisted telephone interview, CATI) oder er führt ein persönliches Interview (Stocké, Kapitel 45 in diesem Band) vor Ort mit einem tragbaren Computer (computer-assisted personal interview, CAPI). Bereits der Fragebogen ist in diesen Verfahren üblicherweise durch eine Software programmiert und wird lediglich digital am Bildschirm angezeigt. Gängige Programme dafür sind beispielsweise *Odin* von NIPO, *Quancept*, *Bellview* und *CI2* von Sawtooth. Sie sind in der Lage, die so erfassten Daten entweder in den Formaten gängiger Statistikprogramme abzuspeichern oder zumindest in Formaten, die von diesen Statistikprogrammen importiert und konvertiert werden können.

Ähnliches gilt für Onlinebefragungen (Wagner/Hering, Kapitel 48 in diesem Band), die naturgemäß ebenfalls mit digital *programmierten Fragebögen* arbeiten. Hier erfasst der Befragte seine Antworten per Mausklick oder Tastendruck selbst in digitaler Form.

Wenn eine Befragung auf die herkömmlichen *gedruckten Fragebögen* zurückgreift, in denen Antworten von Hand angekreuzt werden (paper and pencil interview, PAPI), entweder durch den Interviewer oder, in einer schriftlichen Befragung (Reuband, Kapitel 47 in diesem Band), durch den Befragten selbst, dann werden diese Fragebögen üblicherweise

*eingescannt*. Auch hier wird spezielle Software eingesetzt. Sie wird so programmiert, dass dunkle und helle Flächen an bestimmten Stellen zum Beispiel als 1 und 0 gelesen werden. Typischerweise wird bereits der Fragebogen mit der gleichen Software gestaltet. Das Programm sorgt dann für ein Fragebogendesign (z.B. Größe und Form der anzukreuzenden Kästchen), das es selbst beim Scannen verarbeiten kann, und setzt zum Beispiel in die vier Ecken jeder Seite jeweils einen Markierungspunkt, durch den später auch eine leicht schräg eingescannte Fragebogenseite korrekt eingelesen werden kann.

Natürlich ist es auch in den Datenfenstern der Statistikprogramme so wie SPSS, Stata und SAS möglich, Daten manuell einzugeben. Dazu wählt man per Mausklick oder Kurstorsteuerung jeweils eine bestimmte Zelle in der Datenmatrix aus und gibt dann über die Tastatur den (Zahlen-) Wert ein, der in dieser Zelle stehen soll. Dies ist allerdings ein vergleichsweise aufwändiges Verfahren und wird typischerweise nur bei sehr kleinen Datensätzen eingesetzt. Für große Studien mit drei-, vier- oder fünfstelligen Fallzahlen und ein paar Hundert Variablen wäre es zu umständlich und zeitaufwändig.

---

## 28.4 Datenaufbereitung

Wie schon angeklungen ist, genügt es nicht, die im Zuge der Datenerhebung erfassten Merkmalsausprägungen in das gewünschte Datenformat zu bringen. Bevor die Auswertung beginnen kann, ist eine *Datenaufbereitung* nötig. Sie beinhaltet vor allem das bereits erwähnte Formatieren der Variablen, die Vergabe von Variablen- und von Wertelabels sowie das Definieren von fehlenden Werten.

Als *Variablenformat* kommen bei SPSS, wie bereits erwähnt, unter anderem *Numerisch* und *String* in Frage. Diese gibt es so oder so ähnlich auch in anderen Statistikprogrammen. Daneben gibt es bei SPSS noch sechs andere Formate (u.a. ein Datumsformat), die aber selten verwendet werden. Das Programm R kennt außerdem *logische Variablen* (logical vector). Variablen, die so formatiert sind, können die Ausprägungen *true* (wahr), *false* (falsch) und *not available* (nicht verfügbar) annehmen, was beispielsweise für Filterbedingungen sinnvoll ist.

*Labels* sind Beschriftungen, die es demjenigen, der mit den Daten arbeitet, erleichtern sollen, zu erkennen, welche Informationen in einer Variablen gespeichert sind. Beschriften lässt sich zum einen die Variable selbst (*Variablenlabel*). Typischerweise werden die Variablennamen so gewählt, dass sie die Nummerierung der Fragen im Fragebogen aufgreifen (z.B. V1, V2, V3, ...), um ggf. anhand des Variablenamens schnell den exakten Fragewortlaut nachschlagen zu können. Dann bietet es sich an, der Variablen zusätzlich einen Label zu geben, der den Inhalt der Frage andeutet (z.B. „Kinderzahl“ oder „deutsche Staatsbürgerschaft“), so dass das Nachschlagen im Fragebogen nicht zur Dauerbeschäftigung wird.

Beschriften lassen sich außerdem die einzelnen Ausprägungen einer Variablen (*Wertelabels*). Das ist immer dann sinnvoll, wenn die eigentlich erhobene Information keine Zahl ist (wie im Falle der Frage nach der Kinderzahl), sondern wenn Antwortvorgaben oder andere Informationen lediglich durch Zahlenwerte repräsentiert werden. So könnte

beispielsweise das Geschlecht eines Befragten durch die Werte 0 und 1 erfasst werden. Da die Zuordnung von weiblich und männlich zu den Zahlenwerten 0 und 1 aber willkürlich ist, ist es entscheidend, diese Zuordnung zu dokumentieren. Das geschieht u.a. in Form von Wertelabels.

Beim *Definieren von fehlenden Werten* („missing values“) (Engel/Schmidt, Kapitel 23 in diesem Band) geht es darum, tatsächlich vorhandene Informationen im Sinne der Datenerhebung von Informationen über die Gründe für das Fehlen von „echten“ Informationen zu unterscheiden. Für die Auswertung macht es einen Unterschied, ob eine Antwort deshalb nicht gegeben wurde, weil der Befragte sie nicht geben wollte, oder deshalb, weil er sie nicht geben konnte. Im einen Fall ist mit einer Verzerrung der Ergebnisse zu rechnen. Beispielsweise verweigern besonders gut und besonders schlecht Verdienende häufiger die Auskunft über ihr Einkommen als Menschen, deren Einkommen in der Nähe des Durchschnitts liegt (siehe Schütte/Schmies, Kapitel 61 in diesem Band). Im anderen Falle ist eine Verzerrung nicht unbedingt zu erwarten. (Dass beispielsweise ein Rentner kein Erwerbseinkommen angibt, spiegelt die empirische Realität adäquat wider.) Aus diesem Grund werden in Variablen üblicherweise nicht nur die eigentlichen Merkmalsausprägungen erfasst, sondern ggf. auch unterschiedliche Gründe dafür, dass keine Merkmalsausprägung vorliegt. Diese *Residualkategorien* müssen als solche gekennzeichnet werden, um zu vermeiden, dass sie in statistische Analysen eingehen. So wäre es zum Beispiel fatal, wenn beim Erfassen des monatlichen Nettoeinkommens der Ausfallgrund „will ich nicht beantworten“ mit 999.999 codiert würde und dieser Wert gemeinsam mit allen anderen Werten zur Ermittlung des Einkommensdurchschnitts herangezogen werden würde.

Die Datenaufbereitung kann, je nach Ausgangssituation, noch zahlreiche andere Arbeitsschritte beinhalten, die zum Teil sehr komplex und zeitaufwändig sind und die hier nicht im Detail erläutert werden können. So müssen mitunter bestimmte Angaben *nachcodiert* werden. Beispielsweise werden Angaben zur beruflichen Tätigkeit in Interviews häufig offen erfragt und als Text erfasst und nach der Erhebung in den vierstelligen numerischen ISCO-Code übertragen (ILO 2012). Wenn Erhebungen den Anspruch haben, repräsentativ zu sein, wird meist ein *Gewichtungsfaktor* (Kiesl, Kapitel 24 in diesem Band) berechnet. Zuweilen werden fehlende Werte durch *Imputation* (Cieleback/Rässler, Kapitel 26 in diesem Band) geschätzt. Wenn beispielsweise sehr viele Befragte die Auskunft zu ihrem Einkommen verweigert haben, dieses aber für eine Analyse oder eine Gewichtung von großer Bedeutung ist, kann es sinnvoll sein, die fehlenden Angaben zu schätzen. Die Schätzung erfolgt z.B. mit Hilfe einer Regression aufgrund von vorhandenen Informationen, die mit dem Einkommen nachweislich statistisch stark zusammenhängen (z.B. höchster Bildungsabschluss, Beschäftigungsumfang, ...) (Gartner 2005; Rässler et al. 2008).

In der Regel sind für bestimmte Analyseschritte jeweils weitere Vorarbeiten nötig, die ebenfalls zur Datenaufbereitung gezählt werden können. Vor allem werden neue Variablen gebildet, die die Informationen des ursprünglichen Datensatzes in eine besser „verwertbare“ Form bringen. Um beispielsweise den Zusammenhang von Einkommen und Lebensalter in einer Kreuztabelle darzustellen, müssen beide Informationen vorher klassiert werden. Um Bildung als unabhängige Variable in einer Regressionsanalyse verwenden zu können, muss

die eine ordinal skalierte Variable in mehrere dichotome *Dummy-Variablen* recodiert werden, also in Variablen, die für jeweils eine Ausprägung (z.B. Hauptschulabschluss) ausweisen, ob sie zutrifft (1) oder nicht (0). Möglicherweise muss eine Variable erst neu berechnet werden, beispielsweise das Alter des Befragten bei der Geburt seines ältesten Kindes – als Differenz aus dem Geburtsjahr des ältesten Kindes und dem Geburtsjahr des Befragten. Bestimmte multivariate Verfahren wie beispielsweise die Clusteranalyse (Blasius/Baur, Kapitel 79 in diesem Band) oder die Faktorenanalyse (Blasius, Kapitel 83 in diesem Band) können zur Typen- bzw. Skalenbildung verwendet werden, also ebenfalls neue Variablen generieren, und im Rahmen der Datenaufbereitung der eigentlichen Analyse vorgeschaltet werden.

Aufwändig ist ggf. der *Wechsel der Analyseebene* (Lück 2011a): Zuweilen sollen Analysen nicht nur auf der Ebene durchgeführt werden, in der die Daten organisiert sind (meist die Individualebene), sondern z.B. auch auf Paar- oder Haushaltsebene. Dann muss die gesamte Struktur aus Fällen und Variablen umorganisiert werden. Ähnliches gilt für Ereignisanalysen (Mika/Stegemann und Pötter/Prein, Kapitel 85 und 86 in diesem Band), die bestimmte zeitlich aufeinander folgende und typischerweise im Lebenslauf angeordnete Ereignisse (z.B. Berufseintritt, Heirat, Geburt) miteinander in Beziehung setzen. Auch sie benötigen eine eigene Analyseebene und eine Reorganisation der Datenstruktur (Blossfeld/Rohwer 2002). Für diese konkrete Form der Datenreorganisation hält das Programm Stata einen einfachen Befehl bereit, mit dessen Hilfe vom „wide“- ins „long“-Format gewechselt werden kann (Kuhn/Ruf 2006: 20f.).

---

## 28.5 Fehlersuche und Datenbereinigung

Auch wenn es das hebre Ziel einer Studie ist, kann nicht davon ausgegangen werden, dass Datenerhebung und Datenaufbereitung fehlerfrei ablaufen. Gerade bei großen Fallzahlen wird es quasi immer vorkommen, dass sich in die Daten unmögliche, unwahrscheinliche und widersprüchliche Werte einschleichen. Diese gilt es zu identifizieren und zu dokumentieren, ggf. aus der Analyse auszuschließen oder – sofern möglich – zu korrigieren (Lück 2011b).

Zunächst macht es Sinn, sich eine theoretische Vorstellung davon zu verschaffen, wie falsche Werte in die Daten gelangen können. In Anlehnung an Lück (2011b) können folgende *Fehlerquellen* unterschieden werden:

1. *Fehler im Erhebungsinstrument*: Eine erste Fehlerquelle kann das Erhebungsinstrument selbst sein. Wenn bei der Konstruktion z.B. eines Fragebogens Antwortkategorien (Porst und Franzen, Kapitel 50 und 51 in diesem Band) vergessen werden oder sich überschneiden oder Filterführungen (Klöckner/Friedrichs, Kapitel 59 in diesem Band) falsch gelegt werden, besteht die Gefahr, dass falsche oder zumindest unvollständige Informationen erhoben werden. Fehler im Erhebungsinstrument können leicht identifiziert werden, indem man den Fragebogen gegenliest oder im Falle der computerunterstützten Befragungsmodi am Computer „durchklickt“ und prüft. Vermieden werden können diese Fehler durch eine sorgfältige Konstruktion des Erhebungsinstruments und systematische Pretests.

2. *Reaktivität des Befragten:* Eine häufige Fehlerquelle resultiert aus der Reaktivität des Befragten (Hlawatsch/Krickl, Kapitel 20 in diesem Band). Diese besteht darin, dass der Befragte unzutreffende Angaben macht, weil er nicht zugeben möchte, dass er etwas nicht weiß oder dass er keine Meinung hat (z.B. bei der Beurteilung eines Politikers). Auch durch Tendenzen der sozialen Erwünschtheit können unwahre Daten entstehen (z.B. bei der Beantwortung von Verhaltensfragen wie zu Mülltrennung oder Fernsehkonsum).
3. *Irrtum durch den Befragten:* Unzutreffende Angaben können darauf beruhen, dass der Befragte sich irrt oder Ereignisse verwechselt bzw. sich falsch erinnert (z.B. wann er in welchem Ort welche Beschäftigung aufgenommen hat). Diese fehlerhaften Informationen entstehen versehentlich und sind schwer zu identifizieren, da sie – offenkundig auch für den Befragten – plausibel sind. Fehler, die auf Irrtümern beruhen, können unter Umständen durch Kontrollfragen identifizierbar gemacht werden.
4. *Interpretation des Fragebogens:* Wenn der Interviewer oder der Befragte eine Frage nicht richtig versteht, werden unter Umständen fehlerhafte Daten erzeugt. Das Auftreten von Missverständnissen (wie etwa die Verwechslung von Brutto- und Nettoeinkommen) kann wahrscheinlich nicht vollständig verhindert werden, das Risiko kann aber beispielsweise durch eindeutige Formulierungen, übersichtlich gestaltete Fragebögen und/oder Interviewerschulungen minimiert werden.
5. *Kommunikation zwischen Interviewer und Befragtem:* Wenn ein Interviewer (Glantz/Michael, Kapitel 21 in diesem Band) beteiligt ist, können in der Kommunikation zwischen Interviewer und Befragtem eine weitere Fehlerquelle liegen und Missverständnisse auftreten, selbst wenn Befragter und Interviewer den Fragebogen inhaltlich richtig interpretieren. Das Risiko vergrößert sich, wenn einer der Gesprächspartner mit Dialekt bzw. Akzent spricht oder Jargons benutzt. Aus solchen Missverständnissen entstandene fehlerhafte Angaben sind in aller Regel kaum zu korrigieren. Dennoch kann im Falle eines entsprechenden Verdachts eine Rücksprache mit dem Interviewer sinnvoll sein, um zumindest diese Fehlerquelle ggf. identifizieren zu können.
6. *Bewusst falsche Auskunft durch den Befragten:* Wenn Angaben eindeutig falsch sind und andere Fehlerquellen nicht identifiziert werden können, besteht die Möglichkeit, dass der Befragte bewusst falsche und unrealistische Angaben gemacht hat. Begründet ist dieses Antwortverhalten z.B. durch Vorbehalte gegen die Sozialforschung oder gegen Befragungen allgemein. Bewusst falsche Auskünfte sind letztendlich als Verweigerungen zu betrachten.
7. *Fälschung von Interviews durch den Interviewer:* Dieser Verdacht drängt sich dann auf, wenn sich falsche Angaben bei einem bestimmten Interviewer über mehrere Interviews hinweg häufen und Missverständnisse schwer vorstellbar sind – etwa weil die Formulierungen eindeutig sind und die eingetragenen Werte auch bei unterschiedlicher Auslegung der Frage nicht zutreffen können. Bei diesem Verdacht bietet es sich an, alle Fälle des betreffenden Interviewers in Augenschein zu nehmen. Bestätigt sich die Vermutung, dass sich die Fehler auf einen bestimmten Interviewer konzentrieren, sollte dieser kontaktiert werden, um den Fälschungsverdacht zu bestätigen oder

auszuräumen. Gefälschte Interviews (Blasius, Kapitel 22 in diesem Band) können naturgemäß nicht korrigiert werden, da keine (wahre) Information über die Befragten vorliegt. Die einzige Möglichkeit wäre eine Nacherhebung.

8. *Technischer Fehler bei der Datenerfassung:* Ein typischer technischer Fehler liegt z.B. vor, wenn der ausgefüllte Fragebogen einer PAPI-Erhebung beim Scannen nicht korrekt eingelesen werden kann. Das tritt u.a. dann auf, wenn ein Kreuz in der falschen Farbe oder mit dem falschen Stift oder an einer falschen Stelle gesetzt wird, wenn das Papier nicht richtig eingezogen wird oder der Scanner aus sonstigen Gründen nicht richtig funktioniert. Fehler dieser Art sind mit einem Blick in den Originalfragebogen schnell zu identifizieren und zu korrigieren.
9. *Tippfehler bei der Datenerfassung:* Wenn die Angaben eines Fragebogens per Hand erfasst werden, können Tippfehler auftreten. Diese haben in der Regel typische Muster (z.B. Zahlendreher) und können nach Sichtung des Originalfragebogens bereinigt werden.

Um Fehler im Datensatz zu identifizieren, bieten sich verschiedene Strategien an. Sie reichen von sehr einfachen Verfahren, die jeder Primärforscher anwenden sollte, bis hin zu komplexeren, die möglicherweise nur bei großen, fehleranfälligen Datensätzen sinnvoll sind. Diesen Verfahren vorangestellt kann ein Blick in das Datenfenster der verwendeten Statistiksoftware zunächst helfen, Auffälligkeiten festzustellen. So können Häufungen vieler leerer Zellen ein Hinweis auf Filterfragen, aber auch auf technische Probleme oder auf Fehler im Erhebungsinstrument sein.

Ein grundlegender erster Schritt zur Identifikation von Fehlern im Datensatz besteht darin, sich die *Häufigkeitsverteilungen aller Variablen* eines Datensatzes anzusehen und einige Fragen zu beantworten:

- *Ist jede Identifikationsnummer (ID, Befragtennummer) nur einmal vergeben?* Eine Mehrfachvergabe darf nicht vorkommen und würde dem Zweck einer eindeutigen ID entgegenstehen.
- *Ist der Anteil fehlender Werte bei einer Variablen außergewöhnlich hoch?* Bei Filterfragen wird nur ein Teil der Befragten antworten und auch bei heiklen Fragen sind höhere Ausfälle zu erwarten. Wenn beides nicht der Fall ist und auch die Filterführung richtig funktioniert, spricht einiges dafür, dass der Fehler im Zuge der Datenaufbereitung entstanden ist.
- *Liegen die Werte innerhalb des gültigen Bereichs?* Wenn für die Variable Geschlecht beispielsweise der Wert 3 ausgewiesen wird, obwohl 1 für weiblich und 2 für männlich codiert wurde, liegt offensichtlich ein Fehler vor.
- *Ist die höchste oder niedrigste Ausprägung einer Variablen plausibel?* Gerade bei Variablen mit vielen Ausprägungen ist es einfacher, sich auf diese Ausprägungen zu konzentrieren; unrealistisch hohe oder niedrige Werte sind ein Hinweis darauf, dass Werte falsch erfasst wurden.
- *Weisen alle Ausprägungen Wertelabels auf?* Bei Variablen, denen bei der Datenaufbereitung Wertelabel zugewiesen wurden, sind fehlende Wertelabel ein Hinweis darauf, dass es sich um einen unzulässigen Wert handelt.

Neben diesen Fehlern, die bei der Prüfung lediglich einer Variablen erkennbar sind, gibt es viele Unstimmigkeiten, die erst bei der Betrachtung mehrerer Variablen gleichzeitig aufgedeckt werden können. Dabei hilft die Berechnung von *Kreuztabellen*. Auf diese Art kann z.B. geprüft werden, ob die *Filterführung* richtig angelegt und programmiert wurde. So sollten Befragte, die angegeben haben kinderlos zu sein, bei der darauf folgenden – gefilterten – Frage zum Alter der Kinder keine Angaben gemacht haben. Das entsprechende Feld in der Kreuztabelle sollte keine Fälle enthalten. Nicht nur fehlerhafte Filterführungen, sondern auch inhaltliche Auffälligkeiten können so identifiziert werden, etwa wenn die angegebene Größe der Wohnung nicht mit deren Miethöhe korrespondiert. Bei SPSS ist zu beachten, dass Werte, die als *missing* definiert sind, in Kreuztabellen standardmäßig nicht angezeigt werden.

Tabelle 28.1 fasst die häufigsten Fehlerarten, ihre Manifestierung im Datensatz und Möglichkeiten ihrer Identifizierung zusammen. Die letzte Tabellenspalte zeigt einige Möglichkeiten der Fehlerkorrektur.

Ist ein Fehler einmal identifiziert, stellt sich die vielleicht schwierigste und wichtigste Frage: *Wie soll mit ihm umgegangen werden?* Hierzu stellt die Tab. 28.1 mögliche Strategien vor: Bei einigen Fehlerarten wie Werten außerhalb des gültigen bzw. realistischen Bereichs, ungültig fehlenden Werten oder Inkonsistenzen innerhalb des Fragebogens (Zeilen 1 bis 4 in Tabelle 1), sollte, wenn das Erhebungsverfahren es zulässt (PAPI-Befragungen), immer der Originalfragebogen zu Rate gezogen werden. Mit den Informationen aus dem Originalfragebogen können Zweifel an der Richtigkeit der Daten zerstreut oder bestätigt und ggfs. Korrekturen durchgeführt werden. Bei computerunterstützten Erhebungen (CATI, CAPI) gibt es die Möglichkeit der (telefonischen) Nachbefragung, wobei hier das Verhältnis von Aufwand und Nutzen abzuwegen ist. Wenn in diesen und anderen Fällen eine Korrektur der fehlerhaften Werte nicht möglich ist, sollten diese gekennzeichnet oder ggfs. gelöscht werden. Es muss aber in jedem Fall eine Einzelfallentscheidung vorgenommen werden. *Grundsätze* sollten sein:

- Kann nicht zweifelsfrei nachgewiesen werden, dass ein Wert fehlerhaft ist, sollte er *nicht* gelöscht werden; es sollte lediglich auf ihn hingewiesen werden. Dafür bieten sich sogenannte Flag-Variablen an. Diese Flag-Variablen erhalten bei Vorliegen einer Auffälligkeit den Wert 1 und anderenfalls den Wert 0.
- Kann zweifelsfrei nachgewiesen werden, dass ein Wert fehlerhaft ist, aber nicht, welches der wahre Wert wäre, sollte der Wert gelöscht werden.
- Kann zweifelsfrei nachgewiesen werden, dass ein Wert fehlerhaft ist, und mit sehr hoher Wahrscheinlichkeit geklärt werden, welches der wahre Wert ist, sollte der Wert ersetzt werden.

Jede Datenkorrektur muss für den Forschenden selbst und für Außenstehende dokumentiert werden.

Insgesamt gilt: Wenn ein Fall wiederholt zweifelhafte Merkmalsausprägungen aufweist, sollte man überlegen, den gesamten Fall aus den Analysen auszuschließen. Analoges gilt, wenn eine Variable über den gesamten Datensatz hinweg Unstimmigkeiten aufweist, z.B.

**Tab. 28.1** Fehlerarten, Identifikation und mögliche Korrekturen

Fehlerart (Manifestation)	Beispiel	Prüfung	Korrekturmöglichkeiten
Wert außerhalb des gültigen Bereichs	Geschlecht = 3	Häufigkeitsverteilung sichten	Suche nach Information im Originalfragebogen (PAPI) oder (telefonische) Nacherhebung (CATI, CAPI), sonst: <i>Missing</i>
Wert außerhalb des realistischen Bereichs	Anzahl Zimmer = 14	Häufigkeitsverteilung sichten	Suche nach Information im Originalfragebogen (PAPI) oder (telefonische) Nacherhebung (CATI, CAPI), sonst: Kennzeichnung des Falles mit einer Flagvariablen
Ungültig fehlender Wert	Geburtsjahr = (leeres Feld)	Häufigkeitsverteilung sichten	Suche nach Information im Originalfragebogen (PAPI) oder (telefonische) Nacherhebung (CATI, CAPI), sonst: <i>Missing</i>
Inkonsistenz innerhalb des Fragebogens	pers. Einkommen = „2500 €“ Haushaltseink. = „1800 €“	Kreuztabelle sichten	Suche nach Information im Originalfragebogen (PAPI) oder (telefonische) Nacherhebung (CATI, CAPI), sonst: Kennzeichnung des Falles mit einer Flagvariablen
Ungültiger Wert trotz Filterführung (Spezialfall einer Inkonsistenz innerhalb des Fragebogens)	Filterfrage = „kinderlos“ Alter des 1. Kindes = „9“	Kreuztabelle sichten	Variable über alle Fälle recodieren, sonst: <i>Missing</i>
Widerspruch zu bestimmten anderen Fällen (z.B. bei Befragung von Paaren)	Mann: „Frau nimmt Pille.“ Frau: „Wir verhüten nicht.“	Fälle matchen und Kreuztabelle sichten	Kennzeichnung des Falles mit einer Flagvariablen
Widerspruch gegenüber der Gesamtheit oder Informationen außerhalb der Datenerhebung	Pers. X: Kindergeld = „60 €“ andere Pers./Recherche: „184 €“	Häufigkeitsverteilung sichten	falls möglich: Korrektur anhand anderer Informationen, sonst: <i>Missing</i>

durch einen Fehler im Erhebungsinstrument. Hier kann es nötig werden, auf die Auswertung der betroffenen Variable gänzlich zu verzichten; diese sollte dann in weiteren Analysen weder als unabhängige noch als abhängige Variable verwendet werden.

## 28.6 Datendokumentation

Das Ziel der Datendokumentation ist es, den vollständigen Prozess der Datenerhebung, der Datenaufbereitung und -bereinigung sowie im Falle einer Publikation empirischer Ergebnisse ggf. auch die durchgeführten Analysen für sich und andere nachvollziehbar und rekonstruierbar zu machen.

Zur Datendokumentation zählt zunächst die *Erstellung des Codebooks*. Darin werden die Fragen und Antwortmöglichkeiten des Fragebogens im Wortlaut den Variablen und Ausprägungen im Datensatz zugeordnet und die Variablen- und Wertelabels sowie die fehlenden Werte dokumentiert.

Typischerweise wird das Codebook ergänzt durch einen *Methodenbericht*, also eine Beschreibung des vollständigen Prozesses der Datenerhebung, inklusive Erstellung des Erhebungsinstrumentes, Grundgesamtheit, Stichprobenauswahl, Anzahl der Ausfälle aus jeweils bestimmten Ausfallgründen, Erhebungstechnik, Erhebungsdauer, Anfang und Ende der Feldphase, Interviewer- und Datenkontrolle, um nur die wichtigsten Stichworte zu nennen. Des Weiteren ist die Dokumentation aller Bearbeitungsschritte und Datenmanipulationen (z.B. durchgeführte Prozeduren wie Recodierungen oder Berechnungen) erforderlich, die sowohl bei der Fehlerkorrektur als auch bei der Bildung neuer Variablen (z.B. klassiertes Alter oder Flag-Variablen) anfallen.

Für die *interne Dokumentation*, d.h. für die eigene Verwendung, kann das z.B. in Form von ausdokumentierten und *kommentierten Syntax- bzw. Do-Files der entsprechenden Statistiksoftware* erfolgen. Als alternative bzw. ergänzende Speicherung ist ein *einfaches Textdokument* (Dateiendung txt) zu empfehlen. Dieses Dateiformat bietet den Vorteil, plattformübergreifend und unabhängig vom verwendeten Statistikprogramm und dessen jeweiligen Versionen (Auf- und Abwärtskompatibilität) lesbar zu sein. Dokumentationen für externe Zwecke werden in der Regel als pdf-Dokumente zur Verfügung gestellt.

Die Datendokumentation ist eine Voraussetzung dafür, dass die Daten *archiviert* und für die weitere Forschung verwendbar gemacht werden können: Die erhobenen und dokumentierten Daten werden gesichert und z.B. als scientific use file für Sekundäranalysen zur Verfügung gestellt. Einen Überblick über die Archive geben Mochmann und Schupp (Kapitel 14 und 73 in diesem Band).

## Literatur

- Blossfeld, Hans-Peter/Rohwer, Götz (2002): Techniques of Event History Modeling. Mahwah: Lawrence Erlbaum Associates
- Gartner, Hermann (2005): The Imputation of Wages Above the Contribution Limit of the German IAB Employment Sample. FDZ-Methodenreport 2/2005
- International Labour Office ILO (2012): International Standard Classification of Occupations 2008 (ISCO-08). Volume I: Structure, Group Definitions and Correspondence Tables. Genf: ILO. verfügbar als PDF-Dokument „wcms\_172572.pdf“ unter [www.ilo.org](http://www.ilo.org)
- Kuhn, Andreas/Ruf, Oliver (2006): Einführung in die Statistiksoftware STATA. Institute for Empirical Research in Economics, University of Zurich, Working Paper No. 277. Zürich. URL: [www.econ.uzh.ch/static/workingpapers.php?id=277](http://www.econ.uzh.ch/static/workingpapers.php?id=277)
- Lück, Detlev (2011a): Mängel im Datensatz beseitigen. In: Akremi, Leila/Baur, Nina/Fromm, Sabine (Hg.): Datenanalyse mit SPSS für Fortgeschrittene 1: Datenaufbereitung und uni- und bivariate Statistik. Wiesbaden: VS Verlag. 66-80
- Lück, Detlev (2011b): Zusammenführen von Datensätzen und Wechsel der Analyseebene. In: Akremi, Leila/Baur, Nina/Fromm, Sabine (Hg.): Datenanalyse mit SPSS für Fortgeschrittene 1: Datenaufbereitung und uni- und bivariate Statistik. Wiesbaden: VS Verlag. 81-108
- Rässler, Susanne/Rubin, Donald B./Schenker, Nathaniel (2008): Incomplete Data: Diagnosis, Imputation, and Estimation. In: de Leeuw, Edith/Hox, Joop J./Dillman, Don A. (Hg.): International Handbook of Survey Methodology. Hillsdale: Lawrence Erlbaum. 370-386

Uwe Flick

---

## 29.1 Qualitätssicherung in der qualitativen Sozialforschung

Die Frage, wie sich die Qualität qualitativer Sozialforschung bestimmen lässt, ist seit der Wiederentdeckung der qualitativen Forschung in den 1960er-Jahren virulent. In diesem Beitrag kann die relative diversifizierte Diskussion um die Qualität(-skriterien) qualitativer Forschung nur in Auszügen wiedergegeben werden. Das liegt auch daran, dass es – anders als in der quantitativen Forschung (Krebs/Menold, Kapitel 30 in diesem Band) – keine einheitliche Diskussion über einen allgemein akzeptierten Kriteriensatz gibt. Vielmehr wird die Diskussion in unterschiedlichen Bereichen einerseits und unterschiedlich für verschiedene Forschungsansätze in der qualitativen Forschung andererseits geführt. Dabei lassen sich verschiedene *Diskussionsstränge* unterscheiden:

1. Manche Autoren verfolgen ein eher eng *fokussiertes, gleichzeitig aber generalistisches Konzept*: Eng fokussiert deshalb, weil die (weiter gefasste) Frage nach der Qualität reduziert wird auf die Formulierung und Anwendung von Kriterien (bspw. für die Validität von Interviewaussagen). Generalistisch deshalb, weil dahinter – häufig zumindest – die Hoffnung steht, dass die Kriterien dieselben oder ähnliche sind wie in der quantitativen Forschung und dass sie zumindest dieselbe Funktion hier wie dort erfüllen.
2. Andere Autoren vertreten ein eher *weit gefasstes, gleichzeitig aber spezifisches Konzept*: Weit gefasst deshalb, weil die Qualitätsfrage nicht auf die Erfüllung methodischer Kriterien reduziert wird. Spezifisch deshalb, weil dabei davon ausgegangen wird, dass für qualitative Forschung andere Wege der Qualitätsbestimmung verfolgt werden müssen als in der quantitativen Forschung und ggf. ansatzspezifische Wege eingeschlagen werden sollten.

In diesem Spannungsfeld wird im Folgenden die Problematik entfaltet.

## 29.2 „Klassische“ Gütekriterien im Kontext qualitativer Sozialforschung

Wie lassen sich subjektive Sichtweisen, Alltagswissen (oder andere Gegenstände qualitativer Forschung) verlässlich ermitteln? Und wie lassen sich darüber Aussagen mit einer ausreichenden Gültigkeit zum Untersuchungsthema treffen? Sind erhobene Daten und gezogene Schlussfolgerungen ausreichend unabhängig von der konkreten Person, die sie erhoben bzw. gezogen hat?

Diese Fragen, mit denen sich die Grundbedeutung von Reliabilität, Validität und Objektivität (Krebs/Menold, Kapitel 30 in diesem Band) umschreiben lässt, stellen sich für jede Untersuchung. Die generellen Ansprüche, die hinter solchen Fragen stehen, sollten auch für qualitative Forschung erfüllt sein. Allerdings stellt sich als Problem, inwieweit die in der quantitativen Forschung zur Bestimmung der Gültigkeit verwendeten Kriterien der Reliabilität, Validität und Objektivität („klassische Kriterien“) sich mit den Besonderheiten bzw. Eigenschaften qualitativer Forschung vereinbaren bzw. sich auf diese anwenden lassen (Steinke 1999).

So zeigen Steinke (1999) sowie Kirk und Miller (1986), dass die *Reliabilität* von Daten und Verfahren im traditionellen Sinne – als die Stabilität von Daten und Ergebnissen bei mehreren Erhebungen – für die Bewertung qualitativer Daten eher ungeeignet ist: Die identische Wiederholung einer Erzählung bei wiederholten narrativen Interviews (Küsters, Kapitel 40 in diesem Band) – im Sinne von Replikationsstudien und des Kriteriums der Reliabilität – liefert eher Hinweise auf eine „zurecht gelegte“ Version als auf die Verlässlichkeit des Erzählten.

Kirk und Miller (1986: 21) fassen die Frage der *Validität* darin zusammen, ob „der Forscher sieht, was er (...) zu sehen meint“. Hier ergeben sich ebenfalls Probleme bei der unmittelbaren Anwendung klassischer Validitätskonzeptionen.

*Interne Validität* (Eifler, Kapitel 11 in diesem Band) wird etwa erhöht bzw. sichergestellt, indem ausgeschlossen werden soll, dass andere als die in der Untersuchungshypothese enthaltenen Variablen den beobachteten Zusammenhang bestimmen (Bortz/Döring 2006: 53). In diesem Verständnis liegen bereits die Probleme bei der Übertragung auf qualitative Forschung begründet: Interne Validität meint in der quantitativen Forschung, wie eindeutig ein gemessener Zusammenhang bestimmt werden kann, d.h. inwieweit Störeinflüsse externer Variablen ausgeschlossen werden können. Sie soll durch eine möglichst umfassende Kontrolle der Kontextbedingungen in der Untersuchung erhöht werden. Zu diesem Zweck wird in der quantitativen Forschung die weitgehende Standardisierung der Erhebungs- bzw. Auswertungssituation (Stein, Kapitel 7 in diesem Band) angestrebt. Eine solche Standardisierung des Vorgehens ist jedoch mit dem größten Teil der qualitativen Methoden nicht kompatibel bzw. stellt ihre eigentlichen Stärken in Frage.

*Objektivität* gibt im Kontext quantitativer Forschung an, in wie weit ein Messwert unabhängig ist von der Person, die eine Messung durchführt. Im Kontext der qualitativen Forschung wird Objektivität in zwei Hinsichten zum Thema. Einerseits wird sie zur Charakterisierung von bestimmten Analyseverfahren verwendet. Objektive oder

sozialwissenschaftliche Hermeneutik (Kurt/Hebrik, Kapitel 33 in diesem Band) sind in Abgrenzung zu Verfahren, die auf die Analyse subjektiven Sinns abzielen, an objektivierbarem Sinn aufgrund der Analyse von Texten interessiert. Dabei wird „Objektivität“ allerdings nicht zu einem Überprüfungskriterium vorgenommener Analysen, sondern zu einem Wesensmerkmal der konzipierten und angewendeten Methode. Als Gütekriterium für qualitative Forschung wird Objektivität insgesamt seltener (als bspw. Validität) diskutiert. Eine Ausnahme stellt der Artikel von Madill et al. (2000) dar. Die Autoren machen die Objektivität qualitativer Forschung an der Analyse qualitativer Daten fest. Sie stellen die Frage in den Vordergrund, ob zwei Forscher zu gleichen Ergebnissen bei der Analyse vorliegender qualitativer Daten kommen. Dabei setzen sie Objektivität mit der „Konsistenz der Bedeutung“ durch die Verknüpfung und Gegenüberstellung der „Ergebnisse zweier unabhängiger Forscher“ (Madill et al. 2000: 17) gleich.

Gelegentlich wird zwar der Anspruch erhoben, qualitative Forschung müsse sich zumindest den Fragen stellen, die mit Konzepten wie Reliabilität und Validität oder Objektivität verknüpft sind. Die Anwendung der entsprechenden Kriterien der quantitativen Forschung auf qualitative Forschung wird jedoch eher abgelehnt, da „das ‚Wirklichkeitsverständnis‘, von qualitativer und quantitativer Forschung dafür „zu unterschiedlich“ (Lüders/Reichert 1986: 97) sei. Schon Glaser und Strauss (1979: 92) „bezwifeln, ob der Kanon quantitativer Sozialforschung als Kriterium (...) auf qualitative Forschung (...) anwendbar ist. Die Beurteilungskriterien sollten vielmehr auf einer Einschätzung der allgemeinen Merkmale qualitativer Sozialforschung beruhen – der Art der Datensammlung (...), der Analyse und Darstellung und der (...) Weise, in der qualitative Analysen gelesen werden.“

Als *Konsequenz* solcher Skepsis wird die Auseinandersetzung mit Gütekriterien in der qualitativen Forschung in zwei Richtungen geführt:

1. Wie lassen sich Konzepte wie Reliabilität und Validität für die qualitative Forschung reformulieren bzw. modifizieren?
2. Welche methodenangemessenen Kriterien können entwickelt und an die Stelle von Kriterien wie Objektivität, Validität und Reliabilität gesetzt werden?

Als Beispiel für beide Diskussionsstrände sollen im Folgenden die Reformulierung der Validierung und die daraus resultierenden alternativen Ansätze behandelt werden.

---

## 29.3 Validierung durch Kommunikation

### 29.3.1 Validierung mit den Untersuchungsteilnehmern

Seit Längerem wird in der qualitativen Forschung diskutiert, wie kommunikative Strategien zur Validierung von Daten, Ergebnissen oder allgemeiner von Vorgehensweisen genutzt werden können. Zunächst einmal finden sich verschiedene Bezeichnungen für *Validierung unter Einbeziehung der Interviewten, oder Beobachteten etc.* Lincoln und

Guba (1985) sprechen von „*Member Checks*“, Scheele und Groeben (1988) verwenden den Begriff „*kommunikativer Validierung*“ und Bryman (2004) bezieht sich auf „*Member Validation*“. Gemeint ist dabei jeweils, dass die Forschenden (Zwischen-)Ergebnisse an Interviewpartner (Helfferich, Kapitel 39 in diesem Band) an einem Punkt im Forschungsprozess (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) zurückmelden. Die Zustimmung der Interviewten zu den Ergebnissen oder ihre Ablehnung wird dann als Ansatz der Validierung der Resultate gesehen.

Unterschiede im Vorgehen und Stellenwert ergeben sich einerseits daraus, wann diese Form der Validierung eingesetzt wird:

1. Scheele und Groeben (1988) führen diese kommunikative Validierung ca. zwei Wochen nach dem Interview durch. Sie begrenzen die Validierung auf die Aussagen im *Interview* und damit auf die Erhebung der Daten.
2. Davon unterscheiden sich Ansätze, in denen die *Interpretationen der Daten* den Interviewten vorgelegt werden. Wertz et al. (2011) haben ein Interview in einer Einzelfallstudie mit verschiedenen methodischen Zugängen (z.B. Diskursanalyse, psychoanalytische und narrative Analysen) vorgenommen und deren Ergebnisse verglichen. Dann haben sie die Befragte um ihre Einschätzung der jeweiligen Analysen gebeten. In Ihrem Artikel macht McSpadden (2011) deutlich, wie sie sich als die Befragte mit den Kategorien der Analyse, vor allem aber auch mit den als Belegen herangezogenen Zitaten aus ihrem Interview und den darauf bezogenen Interpretationen auseinandersetzt, welche Schlüsse sie akzeptiert und welche sie zunächst zurückweist. In den weiteren Prozess der Analyse wurde die Reaktion der Befragten berücksichtigt. Obwohl die Forschenden hier nicht eine Validierung der Analysen durch die Befragte anstreben (2011: 334) und die Befragte die Analysen nicht wirklich in Frage stellen will (2011: 341), zeigt dieses Beispiel, welche Irritation die Vorlage von Interpretationen ihrer eigenen Aussagen bei untersuchten Personen auslösen kann.
3. Demgegenüber setzen *member checks* häufig erst nach Vorliegen der *vergleichenden Ergebnisse der Analyse mehrerer Fälle (oder des Forschungsprojekts insgesamt)* an.

Neben technischen Fragen – in welcher Form werden Daten, Analysen oder Ergebnisse den Beteiligten präsentiert – ergeben sich methodologische und ethische Fragen solcher Validierungsstrategien. *Methodologisch* ist zu fragen: *Geht es dabei um Rückmeldung von Ergebnissen oder tatsächlich um deren Validierung?*

Zur ersten Alternative (Rückmeldung von Ergebnissen) haben wir in einem Projekt, in dem Interviews mit Ärzten und Pflegekräften zu ihren Gesundheits- und Alterskonzepten (Flick et al. 2004) geführt wurden, gegen Ende der Studie die Interviewten noch einmal zu focus groups (Vogl, Kapitel 41 in diesem Band) eingeladen, in denen wir ausgewählte Ergebnisse aus der Analyse der Interviews präsentiert haben. Einerseits wurden die Teilnehmenden gebeten, die Ergebnisse zu bewerten und zu kommentieren und ihre Wichtigkeit zu beurteilen. Andererseits wurden sie gebeten, die vorgestellten Ergebnisse hinsichtlich ihrer Relevanz für die eigene berufliche Praxis zu diskutieren und im letzten Schritt mögliche bzw. notwendige Veränderung dieser Praxis zu erörtern. Im ersten Schritt

handelte es sich dabei um einen Ansatz der kommunikativen Validierung, der sich von anderen Ansätzen (etwa bei Scheele/Groeben 1988) in zwei Hinsichten unterscheidet:

1. Nicht die Aussagen der einzelnen Teilnehmenden an der Studie wurden zur Diskussion gestellt, sondern die Trends, die sich aus der Analyse der Daten insgesamt ergeben hatten.
2. An diese Validierung wurde nicht die Weiterverwendung von Aussagen (oder Ergebnissen) geknüpft, wie dies etwa bei Scheele und Groeben praktiziert wird: Bei Scheele/Groeben wird eine Aussage, die Interviewte im Rahmen der kommunikativen Validierung ihrer Interviews ablehnen und damit nicht validieren, aus dem jeweiligen Interview gestrichen und im weiteren Prozess der Datenanalyse nicht mehr verwendet. In unserem Vorgehen war die Analyse bereits abgeschlossen und die kommunikative Validierung bezog sich auf die Plausibilität der Trends, die sich aus der Analyse ergeben hatten.

An dieser Unterscheidung lässt sich auch die allgemeine Frage nach dem *Stellenwert der kommunikativen Validierung (bzw. des Member Checks) in der Bewertung von Forschung und ihren Ergebnissen* festmachen: Müssen bei der Validierung von Ergebnissen durch Zustimmung alle Untersuchten diesen zustimmen, oder die meisten oder wie viele von ihnen, damit dies als relevant für die Validität der Ergebnisse erachtet werden kann? Wie verbindlich ist die Zustimmung der einzelnen Untersuchten für die Forschenden insgesamt?

*Ethische Fragen* ergeben sich daraus, was die Konfrontation einer Interviewpartnerin mit den Interpretationen ihrer Aussagen durch die Forschenden möglicherweise bei dieser auslöst. Das weiter oben behandelte Beispiel von McSpadden (2011) zeigt die Irritationen, die etwa die psychoanalytische oder diskursanalytische Interpretation (Traue et al., Kapitel 34 in diesem Band) des Interviews, das sie gegeben hatte und ihrer Aussagen darin bei der Interviewten bewirken können.

### 29.3.2 Expertenvalidierung

Weniger problematisch erscheint in diesem Zusammenhang die Validierung von Ergebnissen durch die Einbeziehung von ExpertInnen. Hier lassen sich wieder zwei *Vorgehensweisen* unterscheiden:

1. die *Validierung von Ergebnissen durch andere Forschende*, denen sie mit Bitte um Kommentierung und Bewertung vorgelegt werden. Lincoln und Guba (1985) sprechen hier von „*peer debriefing*“.
2. die *Validierung von Ergebnissen durch ExpertInnen aus dem Feld*. Nach Interviews mit einer Zielgruppe der Versorgung (z.B. obdachlosen Jugendlichen, vgl. Flick/Röhnsch 2008) werden etwa die Aussagen und Trends, die sich aus deren Analyse ergeben haben, Professionellen aus verschiedenen Bereichen der Versorgung vorgestellt mit der Bitte, die Plausibilität und Angemessenheit der Erkenntnisse zu prüfen und bewerten.

Bspw. werden bei der Analyse der Interviews mit den Betroffenen bestimmte Barrieren im Zugang zu Versorgungsangeboten identifiziert wie ihre fehlende Krankenversicherung oder Erfahrungen der Zurückweisung durch Ärzte. Diese Barrieren werden dann Fachkräften aus verschiedenen Institutionen, an die sich die Jugendlichen wenden könnten, vorgestellt. Ziel ist dabei zu prüfen, wie diese Hindernisse aus ExpertInnen-sicht einzuschätzen sind – als angemessen oder nicht zutreffend. Das kann verknüpft werden mit der Frage, wie die Barrieren abgebaut werden können.

Verschiedene methodologische Fragen stellen sich auch hier: Wenn dieser Schritt der Validierung der Erkenntnisse dienen soll, wie umfassend muss die Zustimmung der ExpertInnen dann sein und was ergibt sich aus widersprüchlichen Einschätzungen? Wie lässt sich der Differenz der Perspektiven – zwischen der KlientInnen- und der Versor-gungsperspektive – in diesem Fall Rechnung tragen?

Insgesamt betrachtet können die gerade skizzierten Vorgehensweisen wesentliche Auf-schlüsse über die Einschätzung von Aussagen, Ergebnissen und Erkenntnissen liefern, entsprechend der Feststellung von Kvale (2007): „To validate is to communicate“. Sie liefern aber gleichzeitig Hinweise darauf, dass auch seine anderen Feststellungen in diesem Zusammenhang – „to validate is to theorize“ und „to validate is to question“ – zutreffen. Damit auf diesen Wegen eine Validitätsprüfung im klassischen Sinne realisiert werden kann und die Gültigkeit von Ergebnissen bestätigt oder zurückgewiesen werden kann, müsste zunächst das „Benchmarkproblem“ in diesem Zusammenhang gelöst werden: Wie viel Zustimmung bzw. Zustimmung von wie vielen Beteiligten ist notwendig, um sagen zu können „eine Aussage ist valide (oder nicht)“?

Die zuletzt diskutierten Probleme treffen mehr oder weniger stark auf alle Versuche zu, neue Kriterien für die qualitative Forschung zu formulieren, bei denen keine Unterschei-dungspunkte festgelegt sind, ab wann sie erfüllt sind (Flick 2007a, b).

---

## 29.4 Standards nicht standardisierter Forschung

Als Alternative zu (traditionellen, re-formulierten oder neuen) Kriterien diskutiert Bohn-sack (2005) Standards nicht-standardisierter Forschung. Diese Standards sollten nicht „am grünen Tisch“ entwickelt, sondern im Nachgang aus der Rekonstruktion qualitativer For-schungspraxis abgeleitet und expliziert werden: „Die Methoden und Standards qualitativer Forschung werden auf der Grundlage einer empirischen Rekonstruktion der Forschungs-praxis entwickelt“ (2005: 65).

Da sich mittlerweile eine ganze Reihe von mehr oder minder kanonisierten Methoden in der qualitativen Forschung entwickelt und etabliert haben, stehen Forschende heute häufig erstens vor der Frage, welche davon sie anwenden wollen für die Beantwortung ihrer Forschungsfragen.

Hier stellt sich dann zweitens die Frage, worin sich gute von weniger guten Anwendung bestimmter Methoden unterscheiden lässt. Bohnsack schlägt vor, Standards qualitativer Forschung aus der Analyse alltäglicher Standards der Kommunikation zu entwickeln und darüber die Kriterien Gültigkeit und Zuverlässigkeit in der qualitativen Forschung zu rekonstruieren (2005: 76). Dabei wird argumentiert, dass die wesentliche Bezugsebene für die Formulierung von Standards die methodologische und theoretische Begründung des jeweiligen Vorgehens ist. Bohnsack unterscheidet bei qualitativen Methoden zwischen offenen und rekonstruktiven Verfahren. Erstere sind bspw. Leitfadeninterviews (Helfferich, Kapitel 39 in diesem Band) und die Qualitative Inhaltsanalyse (Mayring/Fenzl, Kapitel 38 in diesem Band), Letztere sind etwa das narrative Interview (Küsters, Kapitel 40 in diesem Band), die objektive Hermeneutik (Kurt/Herbrik, Kapitel 33 in diesem Band) und die von Bohnsack entwickelte dokumentarische Methode (Bohnack, Kapitel 67 in diesem Band). Eine – seinen Ansprüchen entsprechende – methodologische und theoretische Begründung weisen nur die rekonstruktiven Verfahren, auf, weshalb auch nur diese Methoden den von Bohnsack entwickelten Qualitätsstandards entsprachen (2005: 74).

Bohnack liefert hier theoretische und methodologische Vorschläge für eine meta-theoretische Fundierung der Diskussion über die Qualität qualitativer Forschung, lässt jedoch verschiedene Fragen unbeantwortet:

1. Lässt sich die Formulierung von Standards in dem heterogenen Feld der qualitativen Forschung (überhaupt bzw. schon zum gegenwärtigen Zeitpunkt) realisieren?
2. Ist mit der Formulierung von Standards nicht die Gefahr der Standardisierung (von Vorgehensweisen und Prozeduren) verknüpft – was den Ansatz qualitativer (häufig auch als nicht-standardisiert bezeichneter) Forschung zumindest in einen Widerspruch verwickelt?
3. Wird nicht über die Formulierung von Standards, wie sie Bohnack vorschlägt, die Frage der Geltungsbegründung von der Ebene der Qualitätsbestimmung praktischer Vorgehensweisen im Feld auf die Ebene der Angemessenheit ganzer Forschungsprogramme verlagert?

Wendet man Bohnacks Vorschlag an, weiß man zwar, dass bestimmte Ansätze – rekonstruktive Verfahren (z.B. das narrative Interview) – den Standards qualitativer Forschung entsprechen, andere – offene Verfahren (z.B. Leitfaden-Interviews) – dagegen nicht. Seine Vorschläge beantworten jedoch nicht, wie die konkreten Anwendungen und Vorgehensweisen – ob nun offener oder rekonstruktiver Methoden – in einem Forschungsprojekt oder Artikel zu bewerten sind.

---

## 29.5 Strategien der Geltungsbegründung

Als dritte Alternative neben Kriterien und Standards in der Auseinandersetzung mit der Frage der Bestimmung der Qualität qualitativer Forschung werden Strategien der Geltungsbegründung diskutiert, von denen hier einige exemplarisch behandelt werden sollen.

## 29.5.1 Suche nach Gemeinsamkeiten und Widersprüchen

Die erste Strategie bezieht sich darauf, welche zusätzlichen Schritte explizit im Forschungsprozess enthalten sind, um Gemeinsamkeiten und Widersprüche in den Ergebnissen zu identifizieren.

### 29.5.1.1 Triangulation

Das Konzept der „Triangulation“ wurde von Denzin (1970) in die qualitative Forschung eingeführt. Triangulation heißt allgemein, dass der Ansatz der Forschung (ein Forschender untersucht den Gegenstand ausgehend von einer Theorie mit einer Methode und einer Datenform) erweitert wird und ein Gegenstand mit mehreren Methoden oder von mehreren Forschenden untersucht wird. Denzin (1970) unterscheidet vier *Formen*:

- Denzins zentrales Konzept ist die *methodologische Triangulation* (vgl. auch Kelle, Kapitel 8 in diesem Band). Sie lässt sich innerhalb einer Methode (*within-method*) anwenden, indem z.B. verschiedene Zugänge in einer Methode kombiniert werden. Das episodische Interview (Flick 2007) enthält einerseits Erzählaufrüfferungen („Könnten Sie bitte erzählen, wie Ihr erster Schultag verlief?“) und andererseits Fragen zu allgemeineren Zusammenhängen („Könnten Sie bitte erläutern, was für Sie einen guten Lehrer kennzeichnet?“). Durch diese Zugänge sollen verschiedene Aspekte eines Gegenstandes („Subjektive Erfahrungen von Schule“) beleuchtet werden. Methodologische Triangulation kann auch durch die Verwendung verschiedener Methoden (*between-method*) realisiert werden. In der erwähnten Studie zur Versorgung obdachloser Jugendlicher (Flick/Röhrnsch 2008) wurden (1) teilnehmende Beobachtungen in verschiedenen Kontexten durchgeführt. Sie wurden trianguliert mit (2) episodischen Interviews mit den Jugendlichen und (3) mit ExpertInneninterviews mit Professionellen der Medizin und Sozialarbeit, die mit dieser Zielgruppe arbeiten.
- Als zweite Form schlägt Denzin *Daten-Triangulation (data triangulation)* vor. Diese kombiniert Daten, die verschiedenen Quellen entstammen und zu verschiedenen Zeitpunkten, an unterschiedlichen Orten oder bei verschiedenen Personen erhoben werden. Im gerade genannten Beispiel ergeben sich diese Daten aus der Verwendung der o.g. Methoden – es resultieren Interviewdaten und Beobachtungsdaten. Zusätzlich können existierende Daten (z.B. Dokumente wie Akten, Leitbilder von Versorgungsinstitutionen) einbezogen werden.
- *Investigator Triangulation* kennzeichnet den Einsatz mehrerer Forschender bei der Durchführung von Beobachtungen bzw. Interviews, um subjektive Einflüsse Einzelner auszugleichen. Dies ist vor allem aufschlussreich, wenn diese unterschiedliche methodische oder inhaltliche Hintergründe in die Forschung einbringen.
- *Theorien-Triangulation* meint die Annäherung an den Forschungsgegenstand ausgehend von verschiedenen theoretischen Perspektiven und etwa die Analyse von Daten mit unterschiedlichen Ansätzen. In unserem Beispiel wurde eine Theorie, die sich auf die Unterschiedlichkeit von Wissen bei verschiedenen Gruppen (Betroffene vs. Professionelle oder Ärzte vs. Sozialarbeiter) bezieht kombiniert mit einer Praxistheorie, die Kommunikation von Professionellen mit KlientInnen fokussiert.

Diese Formen der Triangulation wurden ursprünglich als Ansätze der Überprüfung der Qualität qualitativer Forschung diskutiert. Zunächst ging es auch Denzin um die Validierung von Ergebnissen durch die Verwendung eines zweiten methodischen Zugangs. Denzin nahm zunächst an, dass mit verschiedenen Methoden ein und derselbe Gegenstand – in Denzins Formulierung „das selbe Phänomen“ – abgebildet werde und dass die resultierenden Teilansichten nur noch zusammengesetzt werden müssten. Dieses Verständnis wurde in den 1980er Jahren vor allem im Rahmen ethnografischer Forschung (Knoblauch, Kapitel 36 in diesem Band) kritisiert (ein Überblick über die Diskussion findet sich in Flick 2011) und in der Folge von Denzin auch erweitert. Die oben genannten Formen der Triangulation werden auch in der aktuellen Diskussion und Forschung angewendet. Der Beitrag der Triangulation zur Geltungsbegründung qualitativer Forschung ist, dass der methodisch-theoretische Zugang zum Untersuchungsgegenstand erweitert wird – durch die Verwendung einer zweiten Methode, verschiedener Daten, unterschiedlicher Theorien etc. Zur aktuellen Verwendung des Konzepts sind zwei Anmerkungen notwendig:

1. Triangulation beschränkt sich nicht auf die Überprüfung (oder Geltungsbegründung) von Forschung, sondern wird zunehmend als breiter angelegte Erkenntnisstrategie eingesetzt – als zentraler Bestandteil eines Projekts und nicht nur als Ansatz seiner Überprüfung oder als Kriterium.
2. Triangulation beschränkt sich nicht auf die Bestätigung der Ergebnisse einer (bspw. qualitativen) Methode durch die Ergebnisse einer anderen (bspw. quantitativen) Methode, sondern versucht der Vielschichtigkeit des Untersuchten durch die Erweiterung der methodischen und theoretischen Perspektiven umfassender gerecht zu werden als dies mit einem Zugang möglich wäre. Das Missverständnis, dass Triangulation sich auf die Bestätigung von Ergebnissen beschränke wird v.a. in der Diskussion um Mixed Methods propagiert (Bryman 2004, Kelle, Kapitel 8 in diesem Band).

Beide Anmerkungen zielen darauf ab, dass Triangulation eher der Berücksichtigung und Erfassung von Diversität und Widersprüchlichkeit im Untersuchungsgegenstand dient. Ihre Verwendung ist besonders aufschlussreich, wenn etwa die Beobachtung des Handelns der untersuchten Personen Widersprüche zu ihren Aussagen in Interviews aufdeckt, und weniger, wenn das eine das andere nur bestätigt.

### 29.5.1.2 Analytische Induktion

Die Analytische Induktion prüft nach der *Entwicklung einer vorläufigen Theorie (bzw. eines Musters, Modells etc.) bzw. eines Ergebnisses das Gefundene durch die Analyse von abweichenden bzw. kritischen Fällen (oder gar Gruppen)* (Hering/Schmidt, Kapitel 37 in diesem Band). Darunter ist nach Bühler-Niederberger (1985, S. 476) zu verstehen: „Analytische Induktion ist eine Methode systematisierter Ereignisinterpretation, die sowohl den Prozess der Genese wie auch der Prüfung von Hypothesen umfasst. Ihr entscheidendes Instrument ist die Analyse der Ausnahme, des von der Hypothese abweichenden Falls.“

Als Strategie der Geltungsbegründung fokussiert die analytische Induktion explizit auf die *Suche nach Gegenbeispielen*. Analytische Induktion ist an der Absicherung von gewonnenen Theorien und Erkenntnissen durch die Analyse bzw. Integration abweichender Fälle orientiert. Das Vorgehen der analytischen Induktion beinhaltet nach Bühler-Niederberger (1985: 478) folgende Schritte:

1. Eine grobe Definition des zu erklärenden Phänomens wird formuliert.
2. Eine hypothetische Erklärung des Phänomens wird formuliert.
3. Ein Fall wird im Lichte dieser Hypothese studiert, um festzustellen, ob die Hypothese den Tatbeständen in diesem Fall entspricht.
4. Trifft die Hypothese nicht zu, so wird sie umformuliert oder das zu erklärende Phänomen so umdefiniert, dass dieser Fall ausgeschlossen wird.
5. Praktische Sicherheit kann erreicht werden, nachdem eine kleine Zahl von Fällen untersucht wurde, aber die Entdeckung jedes einzelnen negativen Falls durch den Forscher oder einen anderen Forscher widerlegt die Erklärung und verlangt eine Umformulierung.
6. Es werden solange Fälle studiert, das Phänomen umdefiniert und die Hypothesen umformuliert, bis eine universelle Beziehung etabliert wird; jeder negative Fall ruft nach einer Umdefinition oder Umformulierung.

Die Überprüfung entwickelter Theorien (Muster, Typologien etc.) basiert hier darauf, dass transparent (gemacht) wird, wie mit den Abweichungen und Widersprüchen in den Daten umgegangen wurde und wie bzw. inwieweit diese in die Ergebnisse integriert wurden.

### **29.5.2 Geltungsbegründung durch Transparenz der Vorgehensweisen**

Die zweite Strategie der Geltungsbegründung, auf die hier kurz eingegangen wird, bezieht sich auf die Transparenz der Vorgehensweisen – wie wird nachvollziehbar gemacht, wie die Forschung vorgegangen ist, welche methodischen Entscheidungen getroffen wurden und wie diese die Ergebnisse beeinflusst haben? Diese Strategie konkretisiert sich auf zwei Ebenen:

Einerseits ist bei qualitativer Forschung eine genaue und möglichst vollständige *Forschungsdokumentation* zu empfehlen, in der detailliert festgehalten wird, warum welche Methode ausgewählt wurde, welche Entscheidungen im Forschungsprozess gefallen sind, wie stimmig sie aufeinander bezogen sind etc. Als methodologischer Rahmen für solche Dokumentationen lässt sich der Vorschlag von Lincoln und Guba (1985) nutzen, die zur Überprüfung der Verlässlichkeit qualitativer Daten einen Prozess des „auditing“ skizzieren, der am Vorgang der Buchprüfung im Finanzwesen orientiert ist. Er enthält einen „Überprüfungspfad“ (*auditing trail*) der Folgendes erfassen sollte (Lincoln/Guba 1985: 320f.):

- die Rohdaten, ihre Erhebung und Aufzeichnung;

- Datenreduktion und Ergebnisse von Synthesen durch Zusammenfassung, theoretische Notizen, Memos etc., Summaries, Kurzdarstellungen von Fällen etc.;
- Datenrekonstruktionen und Ergebnisse von Synthesen anhand der Struktur entwickelter und verwendeter Kategorien (Themen, Definitionen, Beziehungen), Erkenntnisse (Interpretationen und Schlüsse) sowie die erstellten Berichte mit ihren Integrationen von Konzepten und den Bezügen zu existierender Literatur;
- Prozessnotizen, d.h. methodologische Notizen und Entscheidungen auch hinsichtlich der Herstellung von Vertrauens- und Glaubwürdigkeit der Erkenntnisse;
- Materialien in Bezug auf Absichten und Anordnungen wie die Forschungskonzeption, persönliche Aufzeichnungen und Erwartungen der Beteiligten;
- Informationen über die Entwicklung der Instrumente einschließlich der Pilotversio-nen und vorläufigen Plänen.

Die Verwendung von Programmen wie Atlas/ti oder MAXqda (Kuckartz/Rädiker, Kapitel 27 in diesem Band) unterstützt solche Auditingprozesse, da sie zumindest für den Prozess der Datenanalyse automatisch mitdokumentieren und die Erstellung entsprechender Dokumentationen unterstützen.

Transparenz wird darüber andererseits nur dann geschaffen, wenn Forschung im Sinne der größtmöglichen Nachvollziehbarkeit (Steinke 1999) auch so dargestellt wird, dass Leser die methodischen Entscheidungen und Vorgehensweisen einer Studie möglichst detailliert beurteilen können (vgl. auch Meyer/Meier zu Verl, Kapitel 15 in diesem Band).

---

## 29.6 Strategien als Kriterien?

Es lassen sich immer wieder Versuche festhalten, die hier zuletzt skizzierten Strategien in Sets von Kriterien zu integrieren (z.B. Lincoln/Guba 1985). Relativ umfassend ist der Vorschlag von Tracy (2010), die acht „Big Tent“ Kriterien vorschlägt. Diese Bezeichnung verwendet sie, da die Kriterien sich nicht auf einen einzelnen Schritt im Forschungsprozess beziehen – wie etwa in einer Validitätsprüfung in der quantitativen Forschung die Gültigkeit der Messung geprüft wird, andere Aspekte, eher außer Acht gelassen werden, z.B. ob bei der Studie überhaupt ein relevantes Problem untersucht wird. Tracy (2010: 839) definiert ihre Kriterien wie folgt: „.... high quality qualitative methodological research is marked by (a) worthy topic, (b) rich rigor, (c) sincerity, (d) credibility, (e) resonance, (f) significant contribution, (g) ethics, and (h) meaningful coherence.“

Dabei werden alle Kriterien detaillierter beschrieben.

1. „Worthy topic“ z.B. meint: „The topic of the research is relevant; timely; significant; interesting.“
2. „Rich rigor“ bezieht sich auf Folgendes: „The study uses sufficient, abundant, appropriate, and complex theoretical constructs; data and time in the field; sample(s); context(s); data collection and analysis processes“ (2010: 840f.).

3. Im Kriterium „*credibility*“ sind Strategien wie Triangulation, member checks und der Umgang mit abweichenden Fällen (hier unter dem Stichwort „*multivocality*“ diskutiert) zusammengefasst (2010: 844).

Insgesamt betrachtet wird hier der Forschungsansatz als Ganzes bewertet – von der Relevanz der Fragestellung zur Bedeutsamkeit der Ergebnisse (auf theoretischer und praktischer Ebene) und zur Ethik (im Feld, in den Beziehungen und nach Abschluss des Projekts). Dadurch unterscheidet Tracys Vorgehen sich vom Kriterienverständnis in der quantitativen Forschung und legt nach der eingangs formulierten Unterscheidung ein weit gefasstes, aber spezifisches Konzept vor.

Tracys Vorschläge sind aber mit dem selben Problem konfrontiert wie diejenigen von Lincoln und Guba (1985): Es lassen sich keine Grenzen (oder Grenzwerte) definieren, wie viel „*worth*“, „*rigor*“, „*credibility*“ oder „*sincerity*“ gegeben sein sollten, damit eine Studie diese Kriterien erfüllt. Darin unterscheiden sich solche Kriterievorschläge von den „klassischen“ Kriterien, bei denen Grenzwerte definiert werden können, ab wann eine Aussage reliabel oder valide ist.

---

## **29.7 Zum Stand der Dinge und Konsequenzen für den Forschungsprozess**

In der qualitativen Forschungslandschaft lässt sich noch nicht ausmachen, ob es in absehbarer Zeit zu einer allgemeinen Verständigung auf Kriterien oder Standards kommen wird. Auch ist nicht klar, ob das überhaupt wünschenswert ist angesichts der Diversifizierung qualitativer Forschung und angesichts dessen, dass sie im Wesentlichen vom Verzicht auf Standardisierung von Vorgehensweisen geprägt ist und lebt. Von daher sind Replikationsstudien als Gütekriterium in der qualitativen Forschung nicht zielführend. allerdings kann als Anspruch an qualitative Forschung formuliert werden, dass

- die Wahl der Methoden begründet dargestellt wird,
- die konkreten Vorgehensweisen expliziert werden,
- die dem Projekt zu Grunde liegenden Ziel- und Qualitätsansprüche benannt werden und
- die Vorgehensweisen so transparent dargestellt werden, dass Leser sich ein eigenes Bild über Anspruch und Wirklichkeit des Projektes machen können.

Ob die hier exemplarisch diskutierten Ansätze der Geltungsbegründung – wie kommunikative Validierung, Triangulation, Analytische Induktion – für die eigene Forschung geeignet sind, hängt auch von der konkreten Fragestellung und Methodik ab.

## Literatur

- Bohnsack, Ralf (2005): Standards nicht-standardisierter Forschung in den Erziehungs- und Sozialwissenschaften. In: Zeitschrift für Erziehungswissenschaft 8 (4): 63-81
- Bortz, Jürgen/Döring, Nicola (2006): Forschungsmethoden und Evaluation für Sozialwissenschaftler. Berlin: Springer
- Bryman, Alan (2004): Social Research Methods. Oxford: Oxford University Press
- Bühler-Niederberger, Doris (1985): Analytische Induktion als Verfahren qualitativer Methodologie. In: Zeitschrift für Soziologie 14 (6): 475-485
- Denzin, Norman K. (1970): The Research Act. Chicago/Englewood Cliffs (NJ): Aldine/Prentice Hall
- Flick, Uwe/Walter, Ulla/Fischer, Claudia/Neuber, Anke/Schwartz, Friedrich W. (2004): Gesundheit als Leitidee? Bern: Huber
- Flick, Uwe (2007a): Managing the Quality of Qualitative Research. London: Sage
- Flick, Uwe (2007b): Qualitative Sozialforschung. Reinbek: Rowohlt
- Flick, Uwe/Röhnsch, Gundula (2008): Gesundheit auf der Straße. Weinheim/München: Juventa
- Flick, Uwe (2011): Triangulation. Wiesbaden: VS Verlag
- Glaser, Barney G./Strauss, Anselm L. (1979): Die Entdeckung gegenstandsgrundbegründeter Theorie: Eine Grundstrategie qualitativer Forschung. In: Hopf, Christel/Weingarten, Elmar (Hg.): Qualitative Sozialforschung. Stuttgart: Klett. 91-112
- Kirk, Jerome/Miller, Marc L. (1986): Reliability and Validity in Qualitative Research. Beverley Hills: Sage
- Kvale, Steinar (2007): Doing Interviews. London: Sage
- Lincoln, Yvonna S./Guba, Egon G. (1985): Naturalistic Inquiry. London: Sage
- Lüders, Christian/Reichert, Jo (1986): Wissenschaftliche Praxis ist, wenn alles funktioniert und keiner weiß warum. In: Sozialwissenschaftliche Literaturreviews 12 (1). 90-102
- Madill, Anna/Jordan, Abbie/Shirley, Caroline (2000): Objectivity and reliability in qualitative analysis. In: British Journal of Psychology 91 (1): 1-20
- McSpadden, Emalinda (2011): The participant's response. In: Wertz, Frederick J./Charmaz, Kathy/McMullen, Linda M./Josselson, Ruthellen/Anderson Rosemarie (Hg.): Five ways of doing qualitative analysis. New York: Guilford. 334-352
- Scheele, Brigitte/Groeben, Norbert (1988): Dialog-Konsens-Methoden zur Rekonstruktion Subjektiver Theorien. Tübingen: Francke
- Steinke, Ines (1999): Kriterien qualitativer Forschung. München: Juventa
- Tracy, Sarah J. (2010): Qualitative Quality: Eight „Big-Tent“ Criteria for Excellent Qualitative Research. In: Qualitative Inquiry 16 (10): 837-851

Dagmar Krebs und Natalja Menold

---

## 30.1 Einführung

Für alle Schritte der Datenerhebung und -auswertung in der quantitativen Sozialforschung (Stein, Kapitel 7 in diesem Band) gibt es Gütekriterien, die es zu beachten gilt, damit die Daten möglichst fehlerfrei erhoben und die erzielten Resultate angemessen interpretiert werden können. Dabei wird zwischen *Gütekriterien für Messinstrumente* (deren Zuverlässigkeit und Gültigkeit) und *Gütekriterien für das gesamte Forschungsdesign* (die Generalisierbarkeit und Eindeutigkeit der Ergebnisse) unterschieden. Qualitätskriterien empirischer (Sozial-)Forschung sind dem Prinzip der Wertfreiheit verpflichtet. Das Postulat der Wertfreiheit kann – bezogen auf die einzelnen Phasen des Forschungsprozesses (Stein, Kapitel 7 in diesem Band) – nur im sog. Begründungszusammenhang (Reichert, Kapitel 2 in diesem Band) eingelöst werden, denn sowohl der Entstehungszusammenhang (weshalb entscheidet sich ein Forscher für eine bestimmte Thematik; Reichertz, Kapitel 2 in diesem Band) wie auch der Verwertungszusammenhang (wie werden erzielte Ergebnisse interpretiert und wozu werden sie verwendet; Friedrichs, Kapitel 16 in diesem Band) unterliegen subjektiven Entscheidungen und enthalten somit implizit immer eine Wertung.

*Gütekriterien* in der (quantitativen) empirischen Sozialforschung im Allgemeinen und der Umfrageforschung (Reinecke, Kapitel 44 in diesem Band) im Besonderen können sich also nur auf die *Phase des Begründungszusammenhangs* beziehen. In diese Phase fallen alle Schritte, die bei der Umsetzung einer Forschungsidee in ein Forschungsdesign zu durchlaufen sind: Instrumentenentwicklung, Stichprobenziehung und Datenerhebung bis zur Bereitstellung von Ergebnissen. Jeder einzelne dieser Schritte ist sachlich begründ- und dokumentierbar und somit interindividuell nachvollziehbar. Für jeden dieser Schritte sind Qualitätskriterien erarbeitet worden, die ausführlich in, Groves et al. (2004), Schnell et al. (2011) sowie De Leeuw et al. (2008) beschrieben werden. Alle im Begründungszusammenhang der Forschung erarbeiteten Gütekriterien liefern die Basis für eine angemessene Bewertung von Forschungsergebnissen.

In diesem Kapitel werden die Gütekriterien von Messungen (Objektivität, Reliabilität, Validität), die Gütekriterien von Forschungsdesigns (externe und interne Validität) und die Gütekriterien von Ergebnissen statistischer Analysen (statistische Validität und praktische Bedeutsamkeit) beschrieben.

---

## 30.2 2. Gütekriterien von Messinstrumenten

Als Messinstrumente werden in der quantitativen Sozialforschung standardisierte Fragebögen bezeichnet (Reinecke, Kapitel 44 in diesem Band), mittels derer Informationen bei Befragungspersonen erhoben werden. Erfragt werden neben demographischen Angaben (z.B. Alter, Bildung, Geschlecht) (Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band)) vor allem Einstellungen (Latcheva/Davidov, Kapitel 55 in diesem Band) zu einer Vielfalt gesellschaftlich interessierender Themen.

Einstellungen, z.B. die Einstellung zu „in Deutschland lebenden Ausländern“, sind nicht direkt beobachtbare theoretische Konstrukte, die auch als *latente Variablen* (Blasius, Kapitel 83 in diesem Band) bezeichnet werden. Beobachtbar (oder messbar) werden Einstellungen durch zustimmende oder ablehnende Antworten auf Fragen oder Feststellungen, die Aussagen über ein Einstellungsobjekt, z.B. „die in Deutschland lebenden Ausländer“ (Allgemeine Bevölkerumfrage, ALLBUS, 2010), enthalten. Dabei werden zur Messung einer Einstellung mehrere Aussagen (Items) verwendet, die vom Forscher als *Indikatoren (manifeste Variablen)*; Burzan, Kapitel 81 in diesem Band) für die latente Variable Einstellung bestimmt (postuliert) werden. Dieser methodische Schritt wird als *Operationalisierung* bezeichnet. Der Zusammenhang zwischen einem Indikator und einer latenten Variable ist also zunächst eine vom Forscher (mehr oder weniger explizit) formulierte Hypothese, die empirisch überprüfbar ist. Hypothesen über den Zusammenhang zwischen mehreren Indikatoren und einer latenten Variable „Einstellung“ sind Gegenstand einer sog. *Mess- oder Hilfstheorie*.

### 30.2.1 Objektivität

Während das Gütekriterium der *Objektivität* – bestehend aus dem Dreiklang von Durchführungs-, Auswertungs- und Interpretationsobjektivität – in der psychologischen Forschung als weitgehend unproblematisch angesehen wird (Bortz/Döring 2006: 195), kann die quantitative Sozialforschung dieses Kriterium nur eingeschränkt einlösen:

Objektivität bei der Interpretation sozialwissenschaftlicher Forschungsergebnisse (*Interpretationsobjektivität*) kann es nicht geben, da Interpretationen subjektiven Bewertungen (Werturteilen) unterliegen (können).

*Durchführungsobjektivität* kann weitgehend durch die *Standardisierung* von Fragebögen (alle Befragten erhalten identisch formulierte Fragen, also identische Stimuli) realisiert werden (Reinecke und Porst, Kapitel 44 und 50 in diesem Band). Allerdings können

Merkmale des Fragebogens, beispielsweise Effekte der Reihenfolge von Fragen (Klöckner/Friedrichs, Kapitel 49 in diesem Band) und/oder Antwortvorgaben (Franzen, Kapitel 51 in diesem Band), oder Merkmale der Befragungssituation und/oder der Interviewer (Glantz/Michael, Kapitel 21 in diesem Band) die Durchführungsobjektivität einschränken. Diese Effekte können jedoch durch geeignete *a priori-Maßnahmen* vor der Befragung, z.B. durch Variation der Fragenreihenfolge sowie durch *ex post-Analyse* der Antworten in Abhängigkeit von Interviewermerkmalen, wenn nicht ausgeschaltet, so doch kontrolliert werden (Sudman et al. 1996; Tourangeau et al. 2000).

Die *Auswertungsobjektivität* (das ist die intersubjektive Nachvollziehbarkeit der – auf der Basis des vorliegenden Datenmaterials – durchgeföhrten Analysen) wird durch eine sorgfältige *Dokumentation der Datenaufbereitung* gewährleistet, bei der die Daten auf fehlerfreie Eingabe und Vollständigkeit überprüft werden. Außerdem ist zu dokumentieren, wie die Daten für *statistische Analysen* (Blasius/Baur, Kapitel 79 in diesem Band) *modifiziert* (Lück/Landrock, Kiesl, Cielebak/Rässler, Kapitel 28, 24 und 26 in diesem Band) wurden. So können Werte einer Variablen zusammengefasst oder in andere Werte umkodiert werden. Sorgfältig zu dokumentieren ist auch der *Umgang mit fehlenden Werten* (Engel/Schmidt, Kapitel 23 in diesem Band).

Die Dokumentation der angewandten statistischen Verfahren sowie der Methoden von Skalierung (Blasius, Kapitel 83 in diesem Band) und Indexbildung (Huinink, Kapitel 82 in diesem Band) tragen ebenso zur Sicherstellung der Auswertungsobjektivität bei. So werden zur Erfassung einer latenten Variable (einer Einstellung) oft nicht einzelne sondern mehrere Indikatoren verwendet, die (additiv oder multiplikativ) miteinander verknüpft, d.h. zu einem Index zusammengefasst werden.

### 30.2.2 Reliabilität

Reliabilität (Zuverlässigkeit) einer Messung wird definiert als das *Ausmaß, in dem wiederholte Messungen eines Einstellungsobjekts zu gleichen Werten führen*. Ist das Einstellungsobjekt „die in Deutschland lebenden Ausländer“, dann sind mehrere Aussagen (Einstellungssitems bzw. -indikatoren) wiederholte Messungen der latenten Variablen „Einstellung zu Ausländern“. So wird etwa im ALLBUS 2010 die Einstellung zu Ausländern mit folgenden vier Items erhoben:

- A => Anpassung: „Die in Deutschland lebenden Ausländer sollten ihren Lebensstil ein bisschen besser an den der Deutschen anpassen“;
- B => Rotation: „Wenn Arbeitsplätze knapp werden, sollte man die in Deutschland lebenden Ausländer wieder in ihre Heimat zurückschicken“;
- C => Politische Rechte: „Man sollte den in Deutschland lebenden Ausländern jede politische Betätigung in Deutschland untersagen“;
- D => Interethnische Ehen: „Die in Deutschland lebenden Ausländer sollten sich ihre Ehepartner unter ihren eigenen Landsleuten auswählen“.

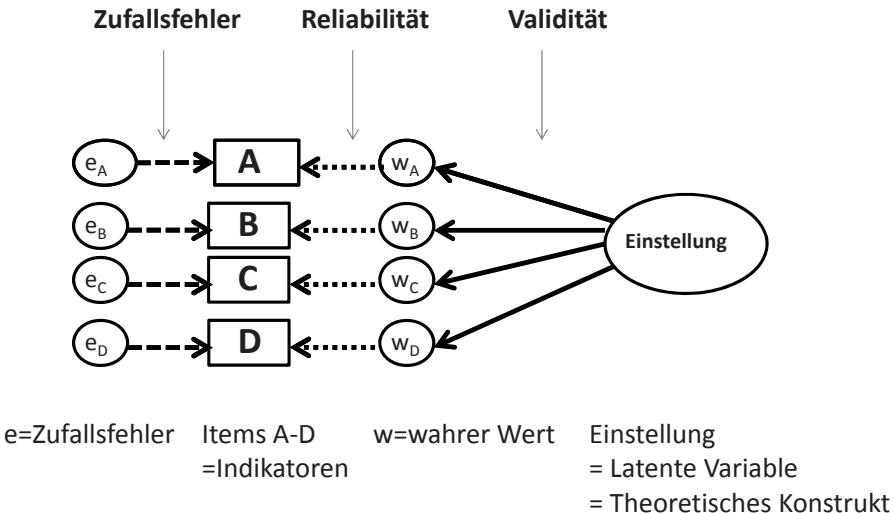
Dabei müssen die *Messwiederholungen* (hier: die vier Items oder Indikatoren) *unabhängig voneinander sein*, d.h. dass die Beantwortung jedes einzelnen Items unabhängig von der Beantwortung der übrigen Items ist. Reliabilität bezieht sich also auf die *Replizierbarkeit von Messungen* und beschreibt die *Genauigkeit (Präzision, Zuverlässigkeit)*, mit der ein (Einstellungs-)Objekt erfasst werden kann.

Allerdings ist nicht davon auszugehen, dass derartige Messungen fehlerfrei sind. Für den Umgang mit Messfehlern bietet die *klassische Testtheorie* ein (mathematisches) Modell, in dem ein beobachteter Messwert – in der Umfrageforschung also die Antwort auf eine Frage bzw. ein Einstellungssitem – als Summe aus „wahrem Wert“ und „Fehler“ definiert wird (siehe Abb. 30.1). Dabei „wirkt“ die Einstellung (die latente Variable) auf die Ausprägung des „wahren Werts“ (w) und dieser bedingt das Ausmaß der Zustimmung zu einem Item (A oder B oder C oder D), wobei der Messwert des Indikators (des Items) nicht mit dem wahren Wert identisch ist, sondern durch einen Messfehler (e) „abgefälscht“ wird. Grundlage dieses Modells sind folgende Annahmen:

1. Der „wahre Wert“ einer Messung ist als *Erwartungswert (Mittelwert)* einer (unendlich großen) Anzahl von unabhängigen Messungen aufzufassen. Fehler bei diesen Messungen (Messfehler) treten nur zufallsbedingt auf.
2. *Zufällige Messfehler* weichen vom „wahren Wert“ mal nach oben, mal nach unten ab und heben sich gegenseitig auf, so dass der Erwartungswert der zufälligen Messfehler (kurz: des *Zufallsfehlers*) bei einer großen Zahl von wiederholten Messungen Null ist. Bei *systematischen Messfehlern* (siehe Abschnitt 30.2.3) würde die Abweichung stets in die gleiche Richtung gehen, so dass der Erwartungswert eines systematischen Messfehlers niemals Null sein kann.
3. *Unabhängigkeit der Messungen* impliziert, dass (zufällige) Messfehler verschiedener (wiederholter) Messungen desselben (Einstellungs-)Objekts nicht miteinander und auch nicht mit dem „wahren Wert“ korrelieren.

Weiterhin basiert die Feststellung (Schätzung) der Reliabilität auf der Annahme der Stabilität von Messungen über die Zeit sowie auf der Annahme der Äquivalenz von Messungen. Die *zeitliche Stabilität von Messungen* kann entweder mit der Test-Retest-Methode oder der Paralleltestmethode erfasst werden.

- Bei der *Test-Retest-Methode* wird eine Einstellung (z.B. die Einstellung zu Ausländern) zu zwei aufeinanderfolgenden Messzeitpunkten bei denselben Personen mit identischen Items gemessen. Die Korrelation zwischen den beiden Messungen gilt hier als Reliabilitätsschätzung, wobei angenommen werden muss, dass sich die „wahren Werte“ zwischen den beiden Messzeitpunkten nicht verändern.
- Bei der *Paralleltestmethode* werden zwei unterschiedliche Messinstrumente verwendet, die dieselbe Dimension messen und identische Messeigenschaften haben (sollen). Die Korrelation zwischen den „parallelen“ Messungen wird wiederum als Reliabilitätsschätzung angesehen. Letztere Methode wird in der Praxis der quantitativen Sozialforschung allerdings kaum angewendet, weil die Entwicklung paralleler Items sehr schwierig ist.

**A: Anpassung; B: Rotation; C: politische Rechte; D: Interethnische Ehen**


**Abb. 30.1** Messmodell zur Beschreibung von Reliabilität und Validität als Beziehungen zwischen latenter Variable, wahrem Wert, Indikatoren und Messfehlern

Die am häufigsten verwendete Methode der Reliabilitätsschätzung basiert auf der *Annahme der Äquivalenz (Gleichwertigkeit) von (wiederholten) Messungen*. Werden z.B. die vier im ALLBUS 2010 zur Messung der Einstellung zu Ausländern verwendeten Items als Stichprobe aus einer Gesamtheit von äquivalenten (gleichwertigen) Indikatoren zur Messung der latenten Variable „Einstellung zu Ausländern“ und damit untereinander als Messwiederholungen verstanden, dann bietet die Eindimensionalität (alle Indikatoren messen die Dimension „Einstellung zu Ausländern“) einen Anhaltspunkt für die Äquivalenz (ähnlich hohe Korrelationen zwischen „wahrem Wert“ und Indikator, siehe Abb. 30.1) der verwendeten Items. Die Hypothese der Eindimensionalität wird dabei mit Hilfe der *konfirmatorischen Faktorenanalyse* (Blasius, Kapitel 83 in diesem Band) überprüft.

Neben der *Reliabilität einzelner Items* interessiert in der Einstellungsmessung vor allem die *Reliabilität der* zur Messung einer Einstellung verwendeten *Einstellungsskala* (Latcheva/Davidov und Blasius, Kapitel 55 und 83 in diesem Band). Es sei angemerkt, dass der Begriff „Einstellungsskala“ missverständlich sein kann, da die „Messlatte“, mit der das Ausmaß der Zustimmung zu den Indikatoren bzw. Einstellungselementen erfasst wird, ebenfalls als „Skala“ (Franzen, Kapitel 51 in diesem Band) bezeichnet wird. Zu unterscheiden ist daher zwischen der „Einstellungsskala“ (dem Instrument, also z.B. der Gesamtheit der vier Items A-D zur Messung der „Einstellung zu Ausländern“) und der „Antwortskala“ (den Antwortvorgaben, die die Intensität der Einstellung als Ausmaß der Zustimmung (1=„stimme gar nicht zu“ bis 7=„stimme voll und ganz zu“) zu den Items erfassen). Welche der beiden

„Skalen“ gemeint ist, dürfte aus dem Kontext erkennbar sein – hier geht es im Folgenden um die Reliabilität von Einstellungsskalen.

*Eindimensionalität* und *Äquivalenz der verwendeten Items* sind die Basis für die Reliabilitätsschätzung eines Messinstruments. Diese kann mit der „*split-half*“-Methode festgestellt werden: Hierbei wird die Gesamtzahl der Items halbiert, so dass zwei Instrumentenhälften entstehen. Die Korrelation zwischen den Messungen der beiden Hälften ist dann die Reliabilitätsschätzung. Je nach Aufteilung der Items können hieraus jedoch unterschiedliche Reliabilitätsschätzungen resultieren. So gibt es z.B. für die vier Items zur Messung der Einstellung zu Ausländern sechs Paare (Hälften) des Instruments (die Items 1&2, 1&3, 1&4, 2&3, 2&4 und 3&4) und damit drei mögliche Korrelationen (1&2 mit 3&4, 1&3 mit 2&4 und 1&4 mit 2&3) zwischen den Instrumentenhälften bzw. den split-half-Versionen.

Dieses Problem wird durch den Koeffizienten „Cronbachs Alpha“ gelöst, der als Mittelwert aller möglichen „*split-half*“-Korrelationen definiert ist (Cronbach 1951). Auch hier muss einem möglichen Missverständnis vorgebeugt werden: Da der Buchstabe „Alpha“ in vielfacher Bedeutung verwendet wird (etwa auch als Ausdruck für die statistische Signifikanz), ist die Benennung für den Reliabilitätskoeffizienten stets „Cronbachs Alpha“.

Der *Reliabilitätskoeffizient Cronbachs Alpha* ist ein Maß für die interne Konsistenz (die Übereinstimmung) der Antworten auf die zur Einstellungsmessung verwendeten Items. Der Reliabilitätskoeffizient kann Werte zwischen Null (keine Konsistenz) und Eins (vollständige Übereinstimmung) annehmen, wobei Werte  $> 0,80$  erwünscht sind, aber oft schon Werte  $\geq 0,70$  akzeptiert werden. Die Werte von Cronbachs Alpha hängen von der Anzahl, der *Homogenität* und der *Trennschärfe* der verwendeten Items ab:

Während *Homogenität* (gemessen durch die mittlere Korrelation zwischen den Items einer Einstellungsskala) die Ähnlichkeit der Indikatoren untereinander angibt, informiert die *Trennschärfe* eines Indikators darüber, wie gut das jeweilige Item die zu messende Einstellung repräsentiert, d.h. wie gut die Einstellung anhand der Antwort(en) auf ein Item prognostizierbar ist. Trennscharf ist ein Item dann, wenn ein hoher Itemwert mit einer starken Ausprägung der zu messenden Einstellung einhergeht. Im Beispiel der Einstellung zu Ausländern ist ein Item dann trennscharf, wenn es mit extremer Zustimmung beantwortet wird und wenn extreme Zustimmung bei allen Items ein hohes Maß an Ausländerfeindlichkeit bedeutet. Werden die Werte der Antworten auf alle Items einer Einstellungsskala addiert, dann resultiert daraus ein Index, der die Stärke der Einstellung zum Ausdruck bringt. Als Maß für die Trennschärfe eines Items in einer Einstellungsskala gilt die *Korrelation eines Itemwerts mit dem Wert des Index* (gebildet aus den übrigen n-1 Items der Einstellungsskala).

### **30.2.3 Validität**

Während die Reliabilität angibt, wie genau ein Messinstrument misst (ungeachtet welcher Inhalt zu messen ist), beschreibt das Gütekriterium der *Validität (Gültigkeit)* das *Ausmaß, in dem ein Messinstrument das Phänomen misst, das gemessen werden soll* (Bühner 2004: 36).

Zur Sicherung der Validität eines Messinstruments ist eine *Hilfstheorie* (siehe Abschnitt 30.2 und Abb. 30.1) notwendig, die das *Einstellungsobjekt* und die *Dimensionen (Facetten)* der *Einstellung* möglichst genau und vollständig beschreibt. Im Beispiel der Ausländeritems des ALLBUS kann eine Theorie über Vorurteile gegenüber fremden Gruppen als Basis für die Entwicklung der Hilfstheorie dienen, die zur Operationalisierung der „Einstellung zu Ausländern“ herangezogen wird. Durch die Verwendung einer solchen Hilfstheorie wird sichergestellt, dass z.B. die Items der ALLBUS-Skala „Einstellung zu Ausländern“ nur die Haltung gegenüber den „in Deutschland lebenden Ausländern“ und nicht eine andere ähnliche Einstellung (z.B. die Einstellung zur EU-Erweiterung oder den Nationalismus) erfassen.

Einstellungen sind oftmals komplex, d.h. sie enthalten unterschiedliche Aspekte (Facetten) der Haltung gegenüber dem Einstellungsobjekt. So kann die Einstellung zu Ausländern Ansichten über die Überlegenheit bzw. Unterlegenheit anderer Nationen, über Kontakte zu Angehörigen anderer Nationalitäten oder über Diskriminierung erfassen (Multifactorial Racial Attitudes Inventory, MRAI, Woodmansee/Cook 1967).

Neben den unterschiedlichen *Einstellungsfacetten* beschreibt eine Hilfstheorie auch den (*vermuteten*) *Zusammenhang (Korrelation)* zwischen den *Facetten*. Durch die Verwendung einer Hilfstheorie wird also gewährleistet, dass die Messinstrumente (Items) diese Komplexität widerspiegeln und die interessierende Einstellung nicht zu eng fassen.

Zur Feststellung der Validität einzelner Items oder (additiv oder multiplikativ zusammengesetzter) Indizes gibt es verschiedene Möglichkeiten:

1. Zu nennen ist zunächst die *Inhaltsvalidität* (auch „*face validity*“ genannt), für deren Beurteilung allerdings keine objektiven Kriterien existieren. Mit „*face validity*“ wird quasi ein „*measurement per fiat*“ (so soll es sein) erreicht, d.h. die Gültigkeit der Messung wird durch Augenschein festgestellt. Genau genommen ist mit dieser Validitätsform die Forderung nach einer *möglichst umfassenden und intersubjektiv nachvollziehbaren Operationalisierung eines theoretischen Konstrukt*s (im Beispiel: der Einstellung zu Ausländern) verbunden. Eine objektive Überprüfung von Operationalisierungen durch Augenschein ist allerdings kaum möglich. Daraus folgt, dass die Inhaltsvalidität in der sozialwissenschaftlichen Forschung kaum zur Bestimmung der Validität von Messungen geeignet ist.
2. Eine weitere Form ist die *Kriteriumsvalidität*. Hierbei geht es um die Übereinstimmung zwischen einer Messung, z.B. einer Einstellung, und einem externen, von der Einstellungsmessung unabhängigen (möglichst objektiven) Kriterium. In der Praxis existiert allerdings oft kein geeignetes (hinreichend genau definiertes und gemessenes) Kriterium, um eine Einstellungsmessung zu validieren, d.h. deren Gültigkeit festzustellen. Ein Beispiel für die Kriteriumsvalidierung bei der Erfassung der Einstellung zu verschiedenen politischen Themen könnte die Wahlentscheidung (objektives Außenkriterium, das von der Einstellungsmessung unabhängig ist) für diejenige Partei sein, deren Themen bei den Befragten am meisten Zustimmung (die positivste Einstellung) erfahren haben. Zur Feststellung der Kriteriumsvalidität werden korrelative Verfahren verwendet. Eine von Null verschiedene Korrelation zwischen den Einstellungswerten,

die mit dem zu validierenden Messinstrument erfasst werden, und dem Wert des Kriteriums dient als eine Kennzahl für die Kriteriumsvalidität. Ein entsprechend objektives Kriterium für die Einstellung zu Ausländern ist dagegen kaum vorstellbar.

3. Schließlich gibt es die *Konstruktvalidität*. Dabei werden aus der oben beschriebenen Hilfsthéorie *Hypothesen über den Zusammenhang zwischen zwei oder mehreren Konstrukten* (z.B. zwischen der Einstellung zu Ausländern und dem Grad des Nationalismus) abgeleitet und überprüft. Wenn es zur Messung des Nationalismus ein bereits erprobtes Messinstrument gibt, dann kann dieses zur Validierung des Konstrukts „Einstellung zu Ausländern“ verwendet werden. Validiert ist die Einstellungsskala „Ausländerfeindlichkeit“ dann, wenn der Zusammenhang zwischen den Konstrukten „Nationalismus“ und „Ausländerfeindlichkeit“ in der postulierten Form beobachtbar, also empirisch nachweisbar ist. Diese Form der Konstruktvalidierung ist die am häufigsten verwendete Methode zur Feststellung der Validität eines Messinstruments. Unterschieden wird dabei zwischen konvergenter und divergenter Validität. *Konvergente oder konvergierende Validität* (Konvergenz = Vergleichbarkeit/Austauschbarkeit von Indikatoren) liegt dann vor, wenn Ähnlichkeit zwischen zwei Messinstrumenten (einem bewährten und einem neu entwickelten) behauptet wird und auch beobachtbar ist (positive Korrelation). *Divergente oder diskriminante Validität* (Diskriminanz = Verschiedenheit von Indikatoren) liegt dementsprechend dann vor, wenn in theoretisch fundierten Hypothesen behauptet wird, dass sich die Werte zweier Messinstrumente gegenläufig verhalten (negative Korrelation) oder dass kein Zusammenhang zwischen den Konstrukten besteht.

Gibt es kein empirisch bewährtes Messinstrument, das zur Validierung eines neu entwickelten Messinstruments verwendet werden kann, dann kann Konstruktvalidität als *faktorielle Validität* (auch *Strukturvalidität* genannt) festgestellt werden. Die faktorielle Validität bezieht sich auf die Überprüfung der theoretisch angenommen Dimensionalität eines Messinstruments. Sinnvoll ist dieses Vorgehen nur bei Messinstrumenten, die aus mehreren Items bestehen (also bei Einstellungsskalen). Wird z.B. vermutet, dass die Items zur Messung der Einstellung zu Ausländern nur *eine* Facette erfassen, so gilt der empirische Nachweis der Eindimensionalität als faktorielle Validität. Wird theoriebasiert angenommen, dass ein Einstellungsobjekt unterschiedliche Facetten hat, dann liegt faktorielle Validität vor, wenn sich die postulierte Mehrdimensionalität der Items zur Messung des Einstellungsobjekts empirisch nachweisen lässt. Zur Überprüfung der faktoriellen Validität wird die konfirmatorische Faktorenanalyse verwendet.

Eine weitere Methode der Validitätsbestimmung ist die *Kreuzvalidierung*. Dabei werden Messergebnisse aus zwei unterschiedlichen Stichproben (einer Grundgesamtheit) miteinander verglichen. Bei übereinstimmenden Resultaten kann ein so überprüftes Messinstrument als kreuzvalidiert gelten.

Die Validität von Messungen wird durch *systematische Messfehler* (*Verzerrungen*) beeinträchtigt. Systematische Messfehler hängen entweder mit Antworttendenzen (z.B. Zustimmungstendenz oder soziale Erwünschtheit) zusammen (Hlawatsch/Krickl, Kapitel

20 in diesem Band) oder können Hinweise auf einen Stichprobenfehler (Häder/Häder und Engel/Schmidt, Kapitel 18 und 23 in diesem Band) geben. Wird z.B. bei der Messung der Einstellung zu Ausländern eine Stichprobe von Personen verwendet, die dem politisch rechten Spektrum der Gesellschaft zuneigen, dann ist die Einstellungsmessung aufgrund eines Stichprobenfehlers nicht valide (weil tendenziös). Die Gefahr eines systematischen Fehlers durch die Zustimmungstendenz besteht dann, wenn bei der Überprüfung der Hilfstheorie (der Validierung) nur Personen mit geringer formaler Bildung berücksichtigt werden, die unabhängig vom „wahren Wert“ ihrer Einstellung zu positiven (zustimmenden) Antworten neigen.

Systematische Messfehler „verzerren“ eine Messung stets in eine Richtung, so dass letztlich bei einer Vielzahl von Messwiederholungen eine Abweichung vom „wahren Wert“ resultiert. Bei zufälligen Messfehlern (siehe Abschnitt 30.2.2) gleichen sich die Abweichungen über eine Vielzahl von Messwiederholungen aus, so dass im Endeffekt der Mittelwert (Erwartungswert) des zufälligen Messfehlers Null ist, was für den systematischen Messfehler nicht gilt.

Messungen werden nachhaltiger durch systematische als durch zufällige Messfehler beeinträchtigt. Das zeigt sich auch in der Beziehung zwischen den Gütekriterien Reliabilität und Validität: *Ein valides Messinstrument misst immer auch reliabel, ein reliables Messinstrument ist dagegen nicht gleichzeitig valide.* So kann z.B. eine schlecht geeichte Waage stets dasselbe Gewicht anzeigen, die Angaben sind aber wegen der fehlerhaften Eichung nicht valide.

### 30.2.4 Kritische Anmerkungen zur Bestimmung von Reliabilität und Validität

Bei den Ergebnissen der beschriebenen Methoden zur Bestimmung von Reliabilität und Validität ist eine *konservative (vorsichtige) Interpretation zu empfehlen*. So ist die *Test-Retest-Korrelation* als Maßzahl für die Zuverlässigkeit von Messungen insofern problematisch, als eine hohe Korrelation nicht nur die substantielle Stabilität eines Merkmals sondern auch die Stabilität von Verzerrungen (bspw. durch Antworttendenzen) anzeigen kann.

Für den Koeffizienten *Cronbachs Alpha* gilt: Nur unter der Bedingung, dass die verwendeten Items das Kriterium der Eindimensionalität erfüllen und äquivalente Indikatoren einer latenten Variablen sind, gibt der Wert von Cronbachs Alpha eine Schätzung der Reliabilität, mit der die Indikatoren die zu messende Eigenschaft erfassen. Äquivalenzkriterien sind dabei:

1. *Parallelität der Items (parallele Items)*: Die latente Variable wirkt in gleicher Weise (identische Faktorenladungen) auf die Indikatoren und auch die Messfehler sind bei allen Items identisch.
2. *Tau-Äquivalenz der Items*: Die Faktorenladungen der Items sind identisch, Faktorenladungen der Items identisch, aber die Messfehler der Items sind unterschiedlich groß.

3. *Kongenerische Items:* Sowohl die Faktorenladungen als auch die Messfehler der Items haben unterschiedliche Werte.

Deshalb wird in der Praxis *vor* der Berechnung des Reliabilitätskoeffizienten die Ein-dimensionalität mittels einer (konfirmatorischen) Faktorenanalyse (Blasius, Kapitel 83 in diesem Band) überprüft. Da die mit Cronbachs Alpha gemessene Reliabilität von der Anzahl der Items in einer Einstellungsskala sowie deren Homogenität und Trennschärfe (siehe Abschnitt 30.2.2) abhängt, besteht die *Gefahr einer „Überhomogenisierung“ von Messinstrumenten* durch

1. geringe sprachliche Variation bei der Itemformulierung,
2. rein mechanisches (ex post) Eliminieren von Items mit geringer Trennschärfe und
3. Erhöhung der Itemzahl.

Für die Bestimmung der Validität mit herkömmlichen Methoden gilt: *Konstruktvalidität* als Korrelation zwischen Messinstrumenten (üblicherweise als additive Indizes aus den Indikatoren gebildet) ist kein hinreichendes Kriterium für die Gültigkeit von Messungen, denn hohe Korrelationen implizieren nicht nur die Ähnlichkeit von Messungen, sondern können ebenso durch die Existenz systematischer Messfehler, bedingt durch z.B. eine Erhebungsmethode, hervorgerufen werden.

Ahilfe können hier die *konfirmatorische Faktorenanalyse bzw. Strukturgleichungsmodelle* schaffen. Bei einem Strukturgleichungsmodell wird das Modell in Abb. 30.1 um eine weitere latente Variable und deren Indikatoren sowie die Messfehler dieser Indikatoren erweitert. Zwischen den beiden latenten Variablen wird eine korrelative Beziehung (symbolisiert durch einen Doppelpfeil) postuliert. Strukturgleichungsmodelle stellen eine Verbesserung der Methoden zur Feststellung von Reliabilität und Validität dar,

1. weil mit Hilfe von Messmodellen statistische Tests der in der Instrumenten- bzw. Hilfstheorie (siehe Abschnitt 30.2.3) formulierten Hypothesen über den Zusammenhang zwischen latenten Variablen und Indikatoren möglich sind und
2. weil durch explizite Berücksichtigung systematischer Messfehler deren Einfluss auf die Validität festgestellt werden kann.
3. Weiterhin werden in Strukturgleichungsmodellen die Hypothesen über konvergente und diskriminante Validität sowie
4. die Übereinstimmung aller im Modell postulierten Beziehungen mit den Daten statistisch geprüft.

Die beschriebenen herkömmlichen Methoden zur Bestimmung der Reliabilität und Validität von Messungen bieten diese Möglichkeiten nicht. Mit Hilfe von Strukturgleichungsmodellen wird die bivariate Korrelation zwischen den verwendeten Items – nach Berücksichtigung der Messfehler – neu berechnet, was zu einer Messfehlerkorrektur bei nachfolgenden Analysen genutzt werden kann (Saris/Gallhofer 2007: Kap 9 und 10).

## 30.3 Gütekriterien von Forschungsdesigns

Die Validität von Forschungsdesigns – in Kontrast zu der Validität der Erhebungsinstrumente – gewährleistet die angemessene Interpretation des Forschungsergebnisses und der daraus abgeleiteten Schlussfolgerungen bezüglich des Forschungsziels.

### 30.3.1 Externe Validität

Ergänzend zur Validität von Erhebungsinstrumenten gewährleistet die Validität von Forschungsdesigns eine angemessene Interpretation des Forschungsergebnisses und der daraus abgeleiteten Schlussfolgerungen bezüglich des Forschungsziels. Zum einen wird betrachtet, ob die Ergebnisse *auf die vorab definierte Grundgesamtheit (Population) übertragbar* sind (z.B. Bortz/Döring 2006). Während die Validität eines Messinstrumentes seine Eigenschaft beschreibt, den „wahren Wert“ einer latenten Variablen (z.B. „Einstellungen zu Ausländern“) erfassen zu können, beschreibt die *externe Validität* (auch äußere *Validität* oder ökologische *Validität* genannt), in welchem Ausmaß das realisierte Ergebnis eine Schätzung des korrespondierenden Populationswertes erlaubt. Die Validität des Messinstrumentes gilt hier als vorausgesetzt.

Die externe Validität wird zum einen durch die *Abweichungen der realisierten Stichprobe von der Grundgesamtheit* und zum anderen durch die *Abweichung der Erhebungssituation von der natürlichen Situation* gemindert (Schnell et al. 2011; Bortz/Döring 2006). So werden bspw. ALLBUS-Daten zur Überprüfung der Hypothese genutzt, dass die Einstellung zu Ausländern (im ALLBUS gemessen als Ausländerfeindlichkeit) in Ostdeutschland stärker ausgeprägt ist als in Westdeutschland. Bei diesem Beispiel wäre zu beurteilen, inwieweit aufgrund der ALLBUS-Stichprobe eine Generalisierung auf die Bundesbevölkerung möglich ist und inwieweit die Situation der Datenerhebung im ALLBUS von einer natürlichen Situation abweicht. Zur Beurteilung des Stichprobendesigns und der Realisierung der Stichprobe im Feld existieren nationale und internationale Standards (z.B. ADM 1999; AAPOR 2011). Die Natürlichkeit der Erhebungssituation in einer face-to-face-Umfrage (Reuband, Kapitel 47 in diesem Band) wie dem ALLBUS ist höher als in einem Laborexperiment (Eifler, Kapitel 11 in diesem Band). So gilt die externe Validität von Laborexperimenten *a priori* methodenbedingt als besonders stark eingeschränkt.

### 30.3.2 Interne Validität

Bei der Ergebnisbeurteilung stellt sich auch die Frage, inwieweit die *Eindeutigkeit der Ergebnisinterpretation* gegeben ist. Auf diesen Aspekt bezieht sich die interne Validität. Das Gütekriterium „*interne Validität*“ wird vorrangig im Zusammenhang mit experimentellen Forschungsdesigns (Eifler, Kapitel 11 in diesem Band) thematisiert. Betrachtet man jedoch

jede Untersuchungsanordnung zur Hypothesenüberprüfung als Experiment (Schnell et al., 2011), dann erscheint die interne Validität als Kriterium zur Beurteilung der Eindeutigkeit des Ergebnisses der Hypothesenprüfung auch in nicht experimentellen Studien als sinnvoll. Generell gilt: *Je besser ein Design dazu geeignet ist, die aufgestellten Hypothesen zu prüfen und Alternativerklärungen auszuschließen, umso höher ist die interne Validität.*

Die interne Validität kann durch eine Reihe von – dem Forscher bekannten oder auch nicht bekannten – Faktoren gefährdet sein. Solche Faktoren werden als *Störfaktoren* bezeichnet. Dazu zählen:

1. *externe Wirkungen* wie zwischenzeitliches Geschehen und Reifungsprozesse,
2. *interventionsbedingte Wirkungen* z.B. des Messvorgangs und der Erhebungssituation und zusätzlich
3. *die Unterschiedlichkeit der untersuchten Personen.*

*Experimentelle randomisierte Designs mit Zufallsverteilung der Personen* auf experimentelle Gruppen sind insbesondere dazu geeignet, unerwünschte Effekte unterschiedlicher Störfaktoren zu kontrollieren. Allerdings muss auch in Bezug auf die nicht experimentellen Untersuchungen kritisch hinterfragt werden, welche Alternativerklärungen für ein Ergebnis auszuschließen sind. Angenommen, man findet bei der Indexvariablen „Einstellung zu Ausländern“ mit ALLBUS-Daten einen signifikanten Mittelwertunterschied zwischen Ost- und Westdeutschland. Hier wäre beispielsweise auszuschließen, dass sich die Erhebungsinstrumente und die Erhebungssituation im Osten systematisch von derjenigen im Westen unterscheiden (z.B. durch Unterschiede in Instruktionen für die Interviewer). Als nächstes wäre zu prüfen, inwiefern sich die Teilstichproben der Ost- und Westdeutschen im Hinblick auf Merkmale, die mit der Einstellung zu Ausländern zusammenhängen können (z.B. Bildung oder Einkommen, Bevölkerungsdichte), systematisch voneinander unterscheiden.

### **30.3.3 Statistische Signifikanz**

Schließlich sollte gesichert sein, dass ein Ergebnis nicht nur statistisch (statistische Validität), sondern auch praktisch relevant ist (praktische Bedeutsamkeit). Da mit steigendem Umfang von Stichproben auch sehr geringe Kennwerte – ohne jegliche inhaltliche Bedeutung – signifikant werden, wird die praktische Bedeutsamkeit zur Beurteilung von inferenzstatistischen Ergebnissen herangezogen.

Die *statistische Validität* (auch: *statistische Signifikanz*) ist gewährleistet, wenn die empirisch ermittelte Fehlerwahrscheinlichkeit ( $p$ ) bei der Ablehnung der Nullhypothese kleiner ist als die a-priori festgelegte Irrtumswahrscheinlichkeit ( $\alpha$ ).

### 30.3.4 Praktische Relevanz

Während die statistische Signifikanz auch dann angegeben wird, wenn sie, wie z.B. bei großen Stichproben, trivial ist, wird die *praktische Bedeutsamkeit* (auch: *praktische Relevanz*) in der empirischen Sozialforschung oft vernachlässigt.

Die Maße für die praktische Bedeutsamkeit sind sog. *Effektgrößen*. Sie beziehen sich unabhängig von der Stichprobengröße auf die Größe der analysierten Zusammenhänge oder Unterschiede. So kann ein in einer großen Stichprobe erzieltes signifikantes Ergebnis (z.B. Korrelation oder Mittelwertunterschied zwischen zwei Variablen) so gering sein, dass es sich kaum von Null unterscheidet und vernachlässigbar ist.

Ein Beispiel soll dies veranschaulichen: In einer Evaluationsstudie mit insgesamt 2000 Schülern brachte die Anwendung einer neuen Unterrichtsmethode gegenüber einer traditionellen Unterrichtsform eine Leistungssteigerung von 0,2 Punkten bei einer maximalen Punktzahl von 13 Punkten (Menold 2007). Der gefundene Unterschied ist zwar statistisch signifikant (aufgrund der Stichprobengröße), praktisch jedoch wegen einer geringen Wertdifferenz vollkommen unbedeutend.

Es wird empfohlen, Effektgrößen als Werte für praktische Bedeutsamkeit zusammen mit den Ergebnissen der Signifikanztests zu berichten. Je nach statistischem Test können das Varianzanteile, standardisierte Mittelwertdifferenzen oder standardisierte Zusammenhangsmaße sein. So wäre ein Mittelwertunterschied bei Einstellungen zu Ausländern zwischen Osten und Westen mit der standardisierten Mittelwertdifferenz Cohen's  $d = 0,1$  von geringer praktischer Bedeutsamkeit. Bei dem Maß „Cohen's  $d$ “ wird die Mittelwertdifferenz an der Streuung relativiert; beim Wert  $d \leq 0,2$  spricht man von kleinen Effekten. Eine Berücksichtigung der Effektgröße ermöglicht nicht nur eine Bewertung des Ergebnisses unabhängig von der Stichprobengröße, sondern auch eine ökonomische Stichprobenplanung (Cohen 1988).

## Literatur

- AAPOR (2011): Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys. URL: <http://aapor.org/Content/NavigationMenu/AboutAAPOR/StandardsampEthics/StandardDefinitions/StandardDefinitions2011.pdf>
- ADM (Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.)/AG.MA (Arbeitsgemeinschaft Media-Analyse e.V.) (Hg.) (1999): Stichproben-Verfahren in der Umfrageforschung. Opladen: Leske + Budrich
- ALLBUS (2010): Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. gesis Datenarchiv für Sozialwissenschaften. <http://www.gesis.org>
- Bortz, Jürgen/Döring, Nicola (2006): Forschungsmethoden und -evaluation für Human- und Sozialwissenschaftler. Heidelberg: Springer
- Bühner, Markus (2004): Einführung in die Test- und Fragebogenkonstruktion. München: Pearson Studium
- Cohen, Jacob (1988): Statistical Power Analysis for the Behavioral Sciences. Hillsdale: Lawrence Erlbaum
- Cronbach, Lee J. (1951): Coefficient Alpha and the Internal Structure of Tests. In: *Psychometrika* 16 (3): 297-334
- Groves, Robert M./Floyd, Floyd J., Jr./Couper, Mick P./Lepowski, James M./Singer, Eleanor/Tourangeau, Roger (2004): Survey Methodology. Hoboken (NJ): John Wiley & Sons
- De Leeuw, Edith/Hox, Joop J./Dillman, Don A. (2008): International Handbook of Survey Methodology. New York: Lawrence Erlbaum Associates
- Menold, Natalja (2007): Methodische und methodologische Aspekte der Wirkungsmessung von (sozial)pädagogischen Interventionen. In: Sommerfeld, Peter/Hüttemann, Matthias (Hg.): Evidenzbasierte Soziale Arbeit. Baltsmannweiler: Schneider Verlag Hohengehren. 26-39
- Saris, Willem E./Gallhofer, Irmtraud, N. (2007): Design, Evaluation and Analysis of Questionnaires in Survey Research. New York: John Wiley & Sons
- Schnell, Rainer/Hill, Paul B./Esser, Elke (2011): Methoden der Empirischen Sozialforschung. München: R. Oldenburg Verlag
- Sudman, Seymour/Bradburn, Norman M./Schwarz, Norbert (1996): Thinking about Answers. San Francisco: Jossey-Bass Publishers
- Tourangeau, Roger/Rips, Lance J./Rasinski, Kenneth (2000): The Psychology of Survey Response. Cambridge: Cambridge University Press
- Woodmansee, John J./Cook, Stuart W. (1967): Dimensions of Racial Attitudes: Their Identification and Measurement. In: *Journal of Personality and Social Psychology* 7: 240-250

Frank Faulbaum

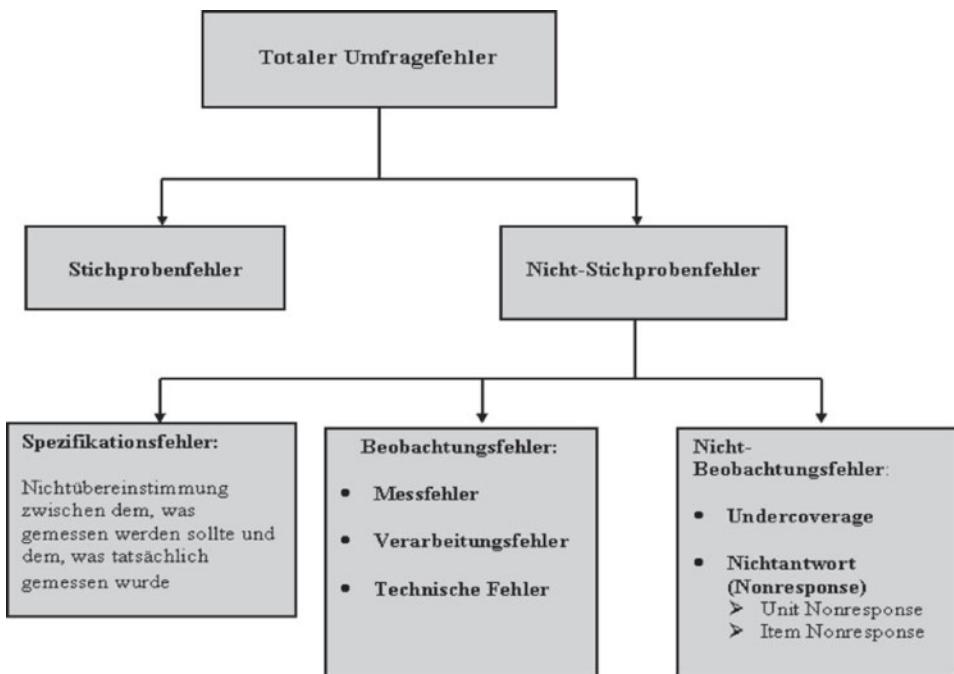
---

## 31.1 Die Komponenten des totalen Umfragefehlers

Von zentraler Bedeutung für die Durchführung von Umfragen (Reinecke, Kapitel 44 in diesem Band) ist die *Umfragequalität* (z.B. Biemer/Lyberg, 2003) und deren Sicherung. Die Qualität von Umfragen steht auch im Mittelpunkt des Konzepts des *totalen Umfragefehlers* (*Total Survey Error, TSE*). Thematisiert werden in diesem Konzept die Fehlerquellen, die während der Datenerhebung von der Stichprobenauswahl (Häder/Häder, Kapitel 18 in diesem Band) bis zum vorliegenden, bereinigten Datensatz die Schätzung der Populationsparameter verzerren könnten. Die mit dem Konzept verbundene Absicht ist also, die Abweichung der Parameterschätzung (Schätzung von Populationsparametern wie Mittelwert, Varianz etc.) auf der Basis des (bereinigten) Datensatzes (Lück/Landrock, Kapitel 28 in diesem Band) auf verschiedene Fehlerquellen zurückzuführen, die auf unterschiedlichen Stufen des Erhebungsprozesses in unterschiedlichem Ausmaß zur Gesamtabweichung der Parameterschätzung vom wahren Populationswert beitragen. (Groves, 2004; Weisberg, 2005). Die Fehlerquellen des TSE sind in Abbildung 1 dargestellt.

Der TSE bezieht also in erster Linie diejenigen Aspekte der Umfragequalität mit ein, die sich negativ auf die Schätzgenauigkeit der Populationsparameter auswirken. Das Konzept stellt dabei vor allem die quantitativen Aspekte der Umfragequalität in den Mittelpunkt und konzentriert sich auf deren Messbarkeit.

Ein ähnliches Konzept wie das des TSE definiert die Qualität der Schätzungen von Populationswerten über den Vergleich der tatsächlichen Schätzungen mit denen, die man erhalten würde, wenn die Umfrage unter idealen Bedingungen hätte durchgeführt werden können, d.h., wenn die Ziele der Messung optimal ohne „störende“ Einflüsse hätten umgesetzt werden können (Dippo, 1997); d.h. je näher die Realität am Ideal, desto höher die Schätzqualität.



**Abb. 31.1** Zerlegung des totalen Umfragefehlers

### 31.2 Stichprobenfehler und Schätzqualität

Die bei der Durchführung von Umfragen auftretenden Fehler lassen sich zunächst in die Klasse der Stichprobenfehler (Häder/Häder, Kapitel 18 in diesem Band) und die Klasse der Nicht-Stichprobenfehler zerlegen. *Stichprobenfehler* entstehen einfach dadurch, dass statt der vollständigen Grundgesamtheit nur eine Stichprobe von Elementen erhoben wird. Dieser Fehler wird gelegentlich noch einmal zerlegt in einen *Schätzfehler* (*estimation error*) und einen *Auswahlfehler* (*selection error*) (Bethlehem 2009: 181).

Der *Schätzfehler* (*estimation error*) betrifft den Sachverhalt, dass als Folge einer Zufallsauswahl bei jeder Ziehung eine andere Schätzung resultiert, die vom Populationswert mehr oder weniger stark abweicht. Der *Auswahlfehler* (*selection error*) entsteht dadurch, dass die in der Formel für den Schätzer des Populationswerts eingetragenen Wahrscheinlichkeiten für die Auswahl einer Stichprobe nicht korrekt sind. In einigen Befragungsarten wie etwa in Webumfragen (Wagner/Hering, Kapitel 48 in diesem Band) sind die Auswahlwahrscheinlichkeiten oft unbekannt. In diesem Fall kann auch nicht mehr von einer Zufallsauswahl im statistischen Sinn gesprochen werden, da von einer Zufallsauswahl die Kenntnis bzw. die Berechenbarkeit der Auswahlwahrscheinlichkeiten gefordert wird.

Wichtige Kriterien für die Schätzqualität sind der Bias des Schätzers, die Varianz des Schätzers, der mittlere quadratische Fehler des Schätzers sowie der Designeffekt. Diese

Größen beziehen sich auf den Fall vollständiger Messungen aller Stichprobenelemente, d.h. auf eine Stichprobe, die weder durch *Unit-Nonresponse* noch durch *Item-Nonresponse* beeinflusst wurde (Engel/Schmidt, Kapitel 23 in diesem Band). Ihre Berechnung ergibt sich bereits aus den mathematisch/statistischen Merkmalen des Stichprobendesigns.

### 31.2.1 Bias

Der Begriff des „*Bias*“ bzw. der *Verzerrung eines Schätzers* spielt an verschiedenen Stellen des Umfrageprozesses eine Rolle, so etwa im Fall der Stichprobenverzerrung durch den Ausfall ganzer Einheiten aus der Bruttostichprobe (siehe unten). Auf der Ebene der Stichprobenfehler geht es speziell um den Bias, mit dem auf der Grundlage des gewählten Stichprobendesigns (z.B. einfache Zufallsauswahl) gerechnet werden muss. Ein *Stichprobendesign* gibt für jede Stichprobe des gewählten Auswahlschemas die Wahrscheinlichkeit an, mit der die Stichprobe ausgewählt wird (Särndal et al. 1992: 27). Diese muss bei einer Zufallsauswahl bekannt sein. Der Bias ist eine Größe, die unabhängig vom Erhebungsprozess und seinen Unwägbarkeiten auf der Basis mathematisch/statistischer Ableitungen berechnet werden kann.

Der (statistische) Bias gibt an, wie genau ein Populationsparameter  $\theta$  im Durchschnitt aller möglichen Stichproben des gleichen Umfangs aus der Zielpopulation  $U$  von einem Schätzer (z.B. dem Median, allgemein: einem statistischen Kennwert oder einer Statistik) geschätzt wird. Er betrifft die Abweichung bzw. den Abstand des Durchschnitts der statistischen Kennwerte aller möglichen Stichproben  $s$  aus der Grundgesamtheit  $U$  vom wahren Populationswert. Wenn wir etwa als Beispiel für einen Schätzer des wahren Populationsmittelwerts das arithmetische Mittel nehmen und für alle Stichproben des gleichen Umfangs das arithmetische Mittel berechnen würden, dann wäre der Bias die Differenz zwischen dem Durchschnitt aller arithmetischen Mittel über alle Stichproben vom Mittelwert in der Population. Der Durchschnitt aller arithmetischen Mittelwerte ist der Erwartungswert der Wahrscheinlichkeitsverteilung der arithmetischen Mittel. Der Bias lässt sich somit als *durchschnittliche Performanz* des Schätzers in Bezug auf den Populationswert auffassen. Formal ist der Bias folgendermaßen definiert:

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta,$$

wobei  $\theta$ : Populationsparameter  $\hat{\theta}$ : Parameterschätzer;  $E(\hat{\theta})$ : Erwartungswert bzw. Durchschnitt des Schätzers. Ist der Bias 0, so heißt der Schätzer *erwartungstreu (unbiased)*.

*In Worten:* Bias  $\stackrel{\text{def}}{=}$  Erwartungswert der Schätzer (Mittelwert der Schätzer über alle Stichproben) minus Populationsparameter.

Der *Erwartungswert (Mittelwert)*  $E(\hat{\theta})$  eines Schätzers  $\theta$  ist formal definiert durch

$$E(\hat{\theta}) = \sum_s p(s) \hat{\theta}(s)$$

wobei  $p(s)$  die bekannte Wahrscheinlichkeit für die Auswahl der Stichprobe  $s$  und  $\hat{\theta}(s)$  der Wert des Schätzers in der Stichprobe  $s$  ist. Die Formel summiert also über alle Stichproben das Produkt aus dem Wert des Schätzers der Stichprobe und der Auswahlwahrscheinlichkeit der Stichprobe.

### 31.2.2 Varianz eines Schätzers

Unterschiedliche Stichprobendesigns können zu unterschiedlichen Schätzern führen, die sich durch unterschiedlichen Bias und unterschiedliche Varianzen auszeichnen. Die *Varianz eines Schätzers*,  $V(\hat{\theta})$  ist ein Maß für die *Präzision eines Schätzers* und gibt an, wie stark die Schätzungen von Stichprobe zu Stichprobe variieren; d.h. die Varianz des Schätzers ist gleich dem Mittelwert der quadrierten Abweichungen der Schätzer aller Stichproben vom Mittelwert der Schätzer. Der mathematische Ausdruck für die Varianz ist

$$V(\hat{\theta}) = \sum_s p(s)(\hat{\theta}(s) - E(\hat{\theta}))^2.$$

Angestrebt wird ein erwartungstreuer Schätzer mit möglichst geringer Varianz.

### 31.2.3 Mittlerer quadratischer Fehler

Der mittlere quadratische Fehler (*Mean Square Error*)  $MSE$  ist die mittlere quadratische Abweichung der Schätzer vom Populationsparameter. Sie gibt an, wie hoch der Durchschnitt aller quadrierten Abweichungen der Stichproben-Schätzer vom Populationsparameter ist:

$$MSE(\hat{\theta}) = E[\hat{\theta} - \theta]^2 = \sum_s p(s)[\hat{\theta}(s) - \theta]^2.$$

Man kann formal zeigen, dass der mittlere quadratische Fehler gleich der Summe aus dem quadrierten Bias und der Varianz des Schätzers ist (Särndal et al. 1992: 40):

$$MSE = \text{Varianz des Schätzers} + \text{Bias}$$

Die oben vorgestellten Konzepte des Bias, der Varianz und der mittleren quadratischen Abweichung eines Schätzers sollen im Folgenden an einem fiktiven Beispiel erläutert werden: Angenommen, eine Population  $U$  würde aus  $N = 5$  Personen  $\{U = 1, 2, 3, 4, 5\}$  bestehen, die folgende Altersstufen aufweisen: Person 1: 32 Jahre; Person 2: 35 Jahre; Person 3: 33 Jahre; Person 4: 29 Jahre; Person 5: 31 Jahre. Der Populationsmittelwert hat dann den Wert  $\mu = 32+35+33+29+31/5=160/5=32$  Jahre. Gezogen werden soll eine Stichprobe von drei Personen nach dem Verfahren der einfachen Zufallsauswahl. Die Anzahl aller möglichen Stichproben mit einem Umfang von  $n = 3$ , die aus  $U$  gezogen werden können, ist

$$\frac{N}{n} = \frac{5}{3} = 5 \times 4 \times 3 / 3 \times 2 \times 1 = 10.$$

Für die zehn Stichproben ergeben sich folgende Werte für das arithmetische Mittel:  
 $s_1 = \{1,2,3\}: 33; s_2 = \{1,2,4\}: 32; s_3 = \{1,2,5\}: 32,7; s_4 = \{1,3,4\}: 31,3; s_5 = \{1,3,5\}: 32; s_6 = \{1,4,5\}: 30,7;$   
 $s_7 = \{2,3,4\}: 32,3; s_8 = \{2,3,5\}: 33; s_9 = \{2,4,5\}: 31,7; s_{10} = \{3,4,5\}: 31$ . Bei einer einfachen Zufallsauswahl werden alle Stichproben mit der gleichen Wahrscheinlichkeit

$$p(s) = 1 / \binom{n}{k} = 1 / \binom{5}{3} = 1 / 10$$

gezogen. Der Erwartungswert ergibt sich dann wie folgt:

$$E(\hat{\mu}) = 0,1 \times (33 + 32 + 32,7 + 31,3 + 32 + 30,7 + 32,3 + 33 + 31,7 + 31) = 0,1 \times 319,7 = 31,97.$$

Der Bias beträgt also  $E(\hat{\theta}) - \hat{\theta} = 31,97 - 32 = -0,03$  Jahre. Ob dieser sehr gering erscheinende Bias statistisch signifikant ist oder nicht, kann nur nach Konstruktion eines Konfidenzintervalls entschieden werden. Die Varianz des Schätzers  $\hat{\mu}$  berechnet sich wie folgt:

$$\begin{aligned} V(\hat{\mu}) &= 0,1 \times (1,03^2 + 0,03^2 + 0,73^2 + (-0,67)^2 + 0,03^2 \cdot (-1,27)^2 + 0,39^2 + 1,03^2 + (-0,27)^2 + (-0,97)^2) \\ &= 0,1 \times (1,061 + 0,001 + 0,533 + 0,449 + 0,001 + 1,613 + 0,152 + 1,061 + 0,072 + 0,941) \\ &= 0,1 \times 5,884 = 0,588. \end{aligned}$$

Als mittlere quadratische Abweichung des Schätzers  $\hat{\mu}$  erhält man den folgenden Wert:

$$\begin{aligned} \text{MSE}(\hat{\mu}) &= 0,1 \times (1,0^2 + 0 + 0,7^2 + (-0,7)^2 + 0 + (-1,3)^2 + 0,3^2 + 1,0^2 + (-0,3)^2 + (-1,0)^2) \\ &= 0,1 \times (1,000 + 0 + 0,490 + 0,490 + 0 + 1,690 + 0,090 + 1,000 + 0,090 + 1,000) = 0,1 \times 5,850 \\ &= 0,585. \end{aligned}$$

Das Beispiel liefert eine empirische Bestätigung für die additive Zusammensetzung von  $\text{MSE}(\hat{\mu}) = \text{Varianz}(\hat{\mu}) + \text{Bias}(\hat{\mu}) = 0,588 - 0,030 = 0,585$ .

### 31.2.4 Effekt des Stichprobendesigns (Designeffekt)

Der *Designeffekt (deff)* basiert auf einem Vergleich der Varianz des Schätzers für eine Statistik unter irgendeinem Stichprobendesign, z.B. unter dem Design einer Klumpenauswahl (Häder/Häder, Kapitel 18 in diesem Band) mit der Varianz des Schätzers unter der Annahme einer einfachen Zufallsauswahl:

$$deff = \frac{\text{Varianz des Schätzers unter einem Design } p}{\text{Varianz des Schätzers unter dem Design einer einfachen Zufallsauswahl}}$$

Er misst den Präzisionsverlust oder den Präzisionsgewinn, der entsteht, wenn man statt einer einfachen Zufallsauswahl ein anderes Design wählt. Ein Verlust tritt ein, wenn

der Designeffekt größer als 1 ist. Ist er kleiner als 1, wird Präzision gewonnen. So kann man z.B. zeigen, dass bei einer 1-stufigen Klumpenauswahl, bei der aus einer Anzahl von Klumpen eine Stichprobe nach dem Verfahren der einfachen Zufallsauswahl gezogen wird (*simple random cluster sampling*), der Designeffekt des Mittelwertschätzers mit wachsender Homogenität innerhalb der Klumpen (Intraklassenkorrelation) steigt (vgl. Särndal/Swensson/Wretman, 1992: 132). Ein Präzisionsgewinn tritt nur ein, wenn es eine ausreichende Variation innerhalb der Cluster gibt. Die Berechnung des Designeffekts kann auf verschiedene Arten von Klumpung angewandt werden wie z.B. Klumpungen durch Interviewer und durch Sampling Points.

---

### 31.3 Nicht-Stichprobenfehler (nonsampling errors)

Neben den Verzerrungen der Parameterschätzung durch die Stichprobenauswahl einschließlich des Stichprobendesigns kann es weitere Quellen für Verzerrungen geben, die nicht auf das Stichprobendesign zurückzuführen sind. Diese betreffen

1. *Beobachtungsfehler (observation errors)*: Beobachtungsfehler beziehen sich auf Verzerrungen der Schätzung, die auf die Beobachtungsdaten zurückzuführen sind.
2. *Nichtbeobachtungsfehler (nonobservation errors)*: Nichtbeobachtungsfehler sind Fehler, die sich weder auf die Stichprobenauswahl noch auf die beobachteten Daten beziehen.
3. *Spezifikationsfehler (specification errors)*

Der Spezifikationsfehler betrifft den Grad der Übereinstimmung zwischen dem, was gemessen werden soll und dem, was tatsächlich gemessen wird, also die sog. *Validität bzw. Gültigkeit einer Messung*.

#### 31.3.1 Nichtbeobachtungsfehler

Zu den *Nichtbeobachtungsfehlern* werden gerechnet:

- *Deckungs- bzw. Coverage-Fehler*: Verzerrungen durch Unterdeckung (coverage error bzw. coverage bias) (Häder/Häder, Kapitel 18 in diesem Band)
- *Nonresponse-Fehler*: Verzerrungen durch Nichtantwort (nonresponse error bzw. non-response bias) (Engel/Schmidt, Kapitel 23 in diesem Band)
  - Verzerrungen durch Unit-Nonresponse
  - Verzerrungen durch Item-Nonresponse

##### 31.3.1.1 Coverage-Fehler

Verzerrungen der Parameterschätzung durch Unterdeckung betreffen die Qualität der *Auswahlgrundlage (sampling frame)*, aus der die Stichprobe der Zielpopulation gezogen wird. Der in der Auswahlgrundlage enthaltene Teil der Zielpopulation wird auch als

*Auswahlgesamtheit* bezeichnet. Das Verhältnis von Zielpopulation und Auswahlgesamtheit kann durch folgende Eigenschaften gekennzeichnet sein:

- Die Auswahlgrundlage enthält alle Elemente der Zielpopulation (*vollständige Deckung bzw. coverage*);
- Es gibt Elemente der Zielpopulation, die durch die Auswahlgrundlage nicht identifiziert werden können (*Unterdeckung bzw. undercoverage*). Ein Beispiel für Unterdeckung ist die Unvollständigkeit des Telefonbuchs als Auswahlgrundlage für Telefonstichproben;
- Es gibt Elemente in der Auswahlgrundlage, die nicht der Zielpopulation angehören (*Überdeckung bzw. overcoverage*). Diese Elemente heißen *nicht auswählbar (ineligible)*, weil sie nicht zur Zielpopulation gehören. Ein einfaches Beispiel ist eine Meldeamtsstichprobe, in der Personen aller Altersstufen aufgeführt sind, die Zielpopulation aber nur aus Personen im Alter über 18 Jahren besteht. Alle Personen im Alter unter 18 Jahren wären nicht auswählbar.
- Es gibt Elemente der Zielpopulation, die in der Auswahlgrundlage mehrfach aufgeführt sind (*doppelte Auflistungen bzw. Duplizierungen; duplicate listings, duplications*). Beispiele sind das Auftreten einer Person unter mehreren Adressen oder unter mehreren Telefonnummern.

Die in der Auswahlgrundlage vorhandenen Elemente der Zielpopulation bilden den durch die Auswahlgrundlage *abgedeckten Teil der Zielpopulation* (*covered population*; Groves et al., 2009: 55). Durch eine unvollständige Abdeckung kann die Parameterschätzung zusätzlich verzerrt werden. So weiß man etwa, dass die sich im Telefonbuch Eingetragenen von den Nichteingetragenen oder die Alleinnutzer von Mobiltelefonen von den Nutzern des Festnetzes systematisch unterscheiden (Hüfken, Kapitel 46 in diesem Band). Dies bedeutet, dass mit einer Verzerrung der Parameterschätzer der betroffenen Variablen und der mit ihnen korrelierten Variablen zu rechnen ist.

Der *Coverage-Fehler* bzw. der *Coverage-Bias* lässt sich formal wie folgt definieren (Groves et al., 2009: 55):

$$\bar{Y}_C - \bar{Y} = \frac{U}{N}(\bar{Y}_C - \bar{Y}_U)$$

wobei:

$\bar{Y}$ : Mittelwert einer Variablen Y in der gesamten Zielpopulation

$\bar{Y}_C$ : Mittelwert der Population, die in der Auswahlgrundlage enthalten ist

$\bar{Y}_U$ : Mittelwert der Zielpopulation, die nicht in der Auswahlgrundlage enthalten ist

N: Gesamtanzahl der Elemente in der Zielpopulation

U: Gesamtanzahl der auswählbaren Elemente, die nicht in der Auswahlgrundlage enthalten sind (Anzahl der nicht in der Auswahlgrundlage enthaltenen Elemente der Zielpopulation)

Das Verhältnis  $\frac{U}{N}$  heißt auch *Unterdeckungsrate (noncoverage rate)*.

Der Unterdeckungsfehler bezieht sich auf die Differenz zwischen dem Mittelwert des durch die Auswahlgrundlage abgedeckten Teils der Zielpopulation und dem Mittelwert der Zielpopulation. Diese Differenz steigt bei konstanter Unterdeckungsrate mit größer werdender Differenz zwischen dem Mittelwert des durch die Auswahlgrundlage abgedeckten und dem Mittelwert des nicht abgedeckten Teils der Population. Sie steigt bei konstanter Differenz zwischen dem abgedeckten und dem nichtabgedeckten Teil auch, wenn die Unterdeckungsrate steigt.

Auswirkungen der Unterdeckung lassen sich für unterschiedliche Befragungsarten nachweisen. So ergab sich etwa in einer Studie zum Wahlverhalten, dass sich Mobile Onlys (Personen, die nur über das Mobilfunknetz erreichbar sind) durch spezifische soziodemographische Merkmale auszeichnen (Keeter, 2006). Diese waren jünger, häufiger alleinstehend und kinderlos. Keeter et al. (2007) fanden, dass die Schätzer vieler Variablen des National Health Interview Survey (NHIS in der Teilgruppe der Jüngeren verzerrt waren. Der Coverage-Bias durch das Weglassen von Mobilfunknutzern wird auch durch andere Analysen bestätigt (Peytchev et al. 2012).

### 31.3.1.2 Nonresponse-Fehler

Verzerrungen durch Nonresponse betreffen einerseits den vollständigen Ausfall aller Messungen eines Stichprobenelements (Unit-Nonresponse) oder den Ausfall einzelner Messungen auf bestimmten Variablen, so dass einige Variablen im Datensatz fehlende Werte aufweisen (Item-Nonresponse). Zusammenfassend können folgende Arten von Unit-Nonresponse unterschieden werden:

- *Nichtkontakt (noncontact)*: Die Bitte um Teilnahme an der Umfrage konnte nicht vorgetragen werden (z.B. bei Nichterreichbarkeit, Nichtauffindbarkeit, Nichtzustellbarkeit des Fragebogens, etc.);
- *Verweigerung (refusal)*: Untersuchungseinheit verweigert die Teilnahme;
- *Unvermögen, an der Umfrage teilzunehmen* (z.B. wegen Krankheit, mangelnder Sprachkompetenz);
- *Ausfälle durch gefälschte oder aus anderen Gründen ungültige Interviews*.

Die Verzerrung der Ausgangsstichprobe durch Nonresponse betrifft die Differenz zwischen dem Mittelwert einer Zielvariablen in der Ausgangsstichprobe und dem Mittelwert dieser Variablen in der Teilstichprobe der Respondenten. Sie kann als Funktion der Differenz zwischen den Mittelwerten der Respondenten und der Nichtrespondenten beschrieben werden:

$$\bar{y}_r - \bar{y}_s = \frac{m_s}{n_s} (\bar{y}_r - \bar{y}_m),$$

wobei:

$\bar{y}_s$  : Mittelwert von  $y$  in der vollständigen Stichprobe s

$\bar{y}_r$  : Mittelwert der Respondenten in der Stichprobe s

$\bar{y}_m$  : Mittelwert der Nichtrespondenten in der Stichprobe s

- $n_s$  : Gesamtanzahl der Elemente in der Stichprobe s  
 $m_s$  : Gesamtanzahl der Nichtrespondenten in der Stichprobe s

Das Verhältnis  $\frac{m_s}{n_s}$  heißt auch *Nonresponse-Rate (nonresponse rate)*.

Allerdings schwankt der Bias von Ausgangsstichprobe zu Ausgangsstichprobe. Groves (2004; vgl. auch Groves et al., 2009: 59) definiert den *Nonresponse-Fehler* daher als Mittelwert des obigen Ausdrucks über alle Stichproben s. Diese Definition nimmt allerdings keinen Bezug auf die konkrete Modellierung der Antwortwahrscheinlichkeiten (Särndal et al. 1992: 575ff.). Die Definition des Nonresponse-Bias lässt sich auch auf die Zielpopulation beziehen und unter der Annahme formulieren, dass es in der Zielpopulation eine feste Teilmenge von Respondenten und Nichtrespondenten gibt. In diesem Fall beschreibt die Formel einen deterministischen Zusammenhang.

Aus der Formel für den Nonresponse-Bias ist ersichtlich, dass der Nonresponse-Fehler den Wert Null annimmt, wenn es keinen Mittelwertunterschied zwischen den Respondenten und den Nichtrespondenten gibt. Er wird umso größer, je größer diese Differenz ist. Ähnlich wie beim Unterdeckungsfehler steigt er bei konstanter Mittelwertdifferenz zwischen Respondenten und Nichtrespondenten aber auch, wenn die Nonresponse-Rate steigt.

In der obigen Form bezieht der Nonresponse-Fehler sowohl Unit-Nonresponse als auch Item-Nonresponse ein. Nichtrespondenten in Bezug auf eine Variable sind alle Fälle, die auf dieser Variablen keinen Wert besitzen einschließlich solcher, die auf keiner Variablen einen Wert haben.

Der Nonresponse-Fehler tritt in Bezug auf eine bestimmte Variable auf, wenn die Antwortwahrscheinlichkeit mit dieser Variablen korreliert ist. Wie bereits die Formel für den Nonresponse-Bias zeigt, ist die Nonresponse-Rate für den Bias nicht allein ausschlaggebend (Groves, 2006; Groves/Peytcheva, 2008). Ist die Teilmenge der Respondenten eine zufällige Teilstichprobe der Ausgangsstichprobe, so ist kein Bias des Schätzers zu erwarten, allerdings ein Einfluss auf dessen Varianz.

Groves/Peytcheva (2008: 183) kommen in ihrer Metaanalyse (Wagner/Weiß, Kapitel 88 in diesem Band), deren Beschränkungen sie explizit erwähnen, u.a. zu der Schlussfolgerung, dass hohe Responseraten das Risiko von Verzerrungen reduzieren können; dies ist aber umso weniger der Fall, je stärker die Ausfallursache mit einer Variablen korreliert ist. Es gibt Umfragen mit einer niedrigen Nonresponse-Rate, aber einem großen Bias.

Da Umfragen unterschiedlichen Designs unterschiedliche Nonresponse-Raten produzieren können, empfiehlt es sich, bei der Berechnung des Nonresponse-Fehlers von den individuellen Antwortwahrscheinlichkeiten  $p_k$  auszugehen. Eine Verzerrung durch Nonresponse entsteht, wenn die Mechanismen, welche die Teilnahme beeinflussen, in Beziehung zu den Variablen stehen, die in den Schätzer eingehen. Groves (2006) empfiehlt daher, die von Bethlehem (2002) vorgeschlagene Formel

$$\text{Bias}(\bar{y}) = \frac{\sigma_{yp}}{\bar{p}}$$

zu verwenden. In diesem Ausdruck ist  $\sigma_{yp}$  die Kovarianz zwischen der Variablen  $y$  und der individuellen Antwortwahrscheinlichkeit (propensity).  $\bar{p}$  ist die mittlere Antwortwahrscheinlichkeit. Nach dieser Formel steigt der Bias mit wachsender Kovarianz zwischen der betrachteten Untersuchungsvariablen  $y$  und den individuellen Antwortwahrscheinlichkeiten bei gleich bleibender durchschnittlicher Antwortwahrscheinlichkeit und sinkt mit zunehmender durchschnittlicher Wahrscheinlichkeit der Nichtantwort. Obige Formel trägt der Erkenntnis Rechnung, dass der Bias von der Korrelation der Ziel- bzw. Umfragevariablen und der Antwortwahrscheinlichkeit abhängt.

Da Unit-Nonresponse auf unterschiedliche Weise entstehen kann, wäre es wünschenswert, die verschiedenen Anteile von Nichterreichbarkeit, Verweigerung etc. abzuschätzen, was auf Grund der Datenlage aber nicht immer möglich ist. Ein Beispiel für eine solche Analyse geben Sakshaug et al. (2010).

Nichtantworten bei spezifischen Fragen sind Folge individueller Antwortentscheidungen (Beatty/Herrmann, 2002). Die Häufigkeit ihres Auftretens kann ebenfalls durch zahlreiche Faktoren wie Befragungsart, Interviewer-Verhalten etc. beeinflusst werden (West/Olson, 2010; Groves/Couper, 1998; die Beiträge in Groves et al., 2002; Fricker et al. 2005).

### 31.3.2 Beobachtungsfehler

*Beobachtungsfehler* werden im Rahmen des TSE-Ansatzes noch einmal unterteilt in *Messfehler (measurement errors)*, *Verarbeitungsfehler (processing errors)* und *technische Fehler (technical errors)*.

#### 31.3.2.1 Messfehler

Unter den Beobachtungsfehlern kommt dem Messfehler besondere Bedeutung zu. Er geht auf die psychometrische Unterscheidung zwischen einer wahren Messung und der beobachteten Messung zurück. Elaboriert wird diese Konzeption in der klassischen Testtheorie (Lewis, 2007; Lord/Novick, 1968). Die *klassische Testtheorie* (Latechva/Davidov, Kapitel 55 in diesem Band) bezieht sich auf quantitative Variablen und geht dabei von einer additiven Zerlegung des beobachteten Werts einer Variablen  $x_i$  in einen *wahren Wert (true score)*  $\tau_i$  und einen *Messfehler (measurement error)*  $\varepsilon_i$  aus; d.h., sie nimmt zwischen der beobachteten Variablen und der wahren Variablen einen *linearen Zusammenhang* an:

$$x_i = \tau_i + \varepsilon_i.$$

Eine beobachtete Variable wird also als Summe einer wahren Variablen (*true variable*) und einer Fehlervariablen (*error variable*) betrachtet:

$$x = \tau + \varepsilon.$$

Der Fehler ist gemäß dieser Zerlegung definiert als Differenz zwischen dem gemessenen Wert und dem wahren Wert:

$$\varepsilon_i = x_i - \tau_i \text{ bzw } \varepsilon = x - \tau.$$

In der klassischen Testtheorie wird der wahre Wert als Erwartungswert der Messung in einer gegen unendlich gehenden Anzahl von Wiederholungsmessungen unter den gleichen Bedingungen verstanden. Der Messfehler wird als normalverteilte Zufallsvariable mit einem Erwartungswert von Null aufgefasst. Sowohl die wahre Variable als auch die Fehlervariable stellen unbeobachtete Variablen dar, so dass die Zerlegung einer beobachteten Variablen in eine wahre Variable und eine Fehlervariable stets zu einem statistischen Modell mit latenten Variablen führen muss.

Für die Umfrageforschung ist von Bedeutung, dass der Messfehler durch verschiedene Fehlerquellen beeinflusst werden und daher in verschiedene Komponenten zerlegt werden kann (hierzu auch Biemer/Trewin, 1997). Quellen für den Messfehler können sein:

- Sichtbare und unsichtbare Merkmale der Interviewer und des Interviewer-Verhaltens;
- Befragungssituation (z.B. Anwesenheit Dritter);
- Frageformulierung;
- Fragebogengestaltung;
- Befragungsart (Administrationsform, Kommunikationskanal, Befragungstechnologie);
- Merkmale der Befragten (Antworttendenzen).

Auch Mängel des Antwortprozesses, d.h. des zwischen Frage und Antwort intervenierenden mentalen Prozesses im Befragten, in der Literatur wird von *Satisficing* gesprochen, kann zu Messfehlern führen (Krosnick, 1991).

### 31.3.2.2 Verarbeitungsfehler

Beispiele für Verarbeitungsfehler sind Editierfehler, Eingabefehler, Tabellierungsfehler und Gewichtungsfehler (Biemer/Lyberg, 2003: 39). Letztere bestehen in der falschen Berechnung von Gewichten.

### 31.3.2.3 Technische Fehler

Technische Fehler werden nicht immer separat aufgeführt. Sie bekommen aber durch neue technologische Innovationen in der Datenerhebung stärkeres Gewicht. Geräteausfälle, fehlerhafte Software, etc. können zu technischen Fehlern führen, die das Antwortverhalten der Befragten beeinflussen können.

### 31.3.3 Spezifikationsfehler

Der *Spezifikationsfehler* schließlich betrifft den Grad der Übereinstimmung zwischen dem, was gemessen werden soll und dem, was tatsächlich gemessen wird, also die *Validität* bzw. *Gültigkeit einer Messung* (Krebs/Menold, Kapitel 30 in diesem Band). Je größer der Spezifikationsfehler, desto gefährdeter ist die inhaltliche Interpretierbarkeit der Daten. Theoretisch wäre es möglich, dass sich die Parameterschätzung einer im Übrigen fehlerfreien

Umfrage auf eine falsch spezifizierte Variable bezieht, was bedeuten würde, dass die Interpretation aller statistischen Ergebnisse, die sich auf diese Variable beziehen, in Zweifel zu ziehen wäre. Der Spezifikationsfehler lässt sich im Rahmen spezifischer Umfrage- und/oder Fragebogendesigns quantifizieren wie etwa durch Korrelation eines empirischen Indikators mit dem durch ihn gemessenen Konstrukt (theoretische Validität). Die Validität einer Messung hängt sehr davon ab, wie gut die Frage im Sinne des Forschers in ihrer Bedeutung verstanden wurde und daher von einer eingehenden Prüfung der Fragequalität (Faulbaum et al. 2009), für die verschiedene Verfahren zur Verfügung stehen (Presser et al., 2004).

---

### 31.4 Kritische Bestandsaufnahme des Konzepts

Im Rahmen ihrer kritischen Betrachtung des Konzepts des TSE und seiner Rolle als Qualitätsstandard in der Umfrageforschung zählen Groves/Lyberg (2010) einige Unzulänglichkeiten des in Abbildung 1 dargestellten Rahmenkonzepts vor, das dennoch nach wie vor eine akzeptable Hintergrundfolie für die Beurteilung von Verzerrungen darstellt. Die Autoren kritisieren insbesondere das Fehlen nicht-statistischer Qualitätsindikatoren wie Transparenz, Glaubwürdigkeit, etc. sowie den Aufwand für die Messung bestimmter Fehlerkomponenten in konkreten Umfragen, der die Messung in praktischen Anwendungen erschwert. In der Tat sind die notwendigen Informationen über die Population oder die Bruttostichprobe oft nicht bekannt. Zeit- und Kostenüberlegungen können die Anwendung komplexerer Analyseverfahren verhindern. In der Regel bleibt die Messung von Verzerrungen also die Aufgabe wissenschaftlicher Untersuchungen und nicht, wie es eigentlich sein sollte, auch Aufgabe eines konkreten Anwenders. Die Autoren fordern außerdem im Rahmen von Kausalanalysen (Kühnel/Dingelstedt, Kapitel 80 in diesem Band) eine verstärkte Suche nach den Fehlerursachen sowie eine Betrachtung der Korrelationen zwischen den Fehlerkomponenten, da die Komponenten des Umfragefehlers nicht unbedingt unabhängig voneinander sein müssen. Ähnlich wie Groves/Lyberg argumentiert auch Smith (2011), der eine weitere Vervollständigung der Fehlerkomponenten und eine Anpassung an die Erfordernisse des interkulturellen Vergleichs vorschlägt.

Der TSE wie er in Abbildung 1 dargestellt ist, thematisiert im Wesentlichen die Qualität der *Schätzung*, d.h. die Abweichung des Schätzers vom wahren Populationswert. Die Qualität einer Umfrage muss aber als Qualität eines aus mehreren Schritten bestehenden Prozesses verstanden werden. Daher spricht man in diesem Zusammenhang auch von *Prozessqualität* (Lyberg et al., 1997). Die Art und Weise, wie die einzelnen Schritte des Umfrageprozesses realisiert werden, ist maßgeblich für deren Qualität und damit für die Qualität der gesamten Umfrage einschließlich der aus ihr resultierenden Daten; d.h. die Optimierung einer Umfrage beinhaltet die Optimierung aller Schritte und daher die Optimierung der Prozessqualität (Lyberg et al., 1997; Biemer/Lyberg, 2003). Qualitätseinbußen in einzelnen Stufen lassen sich nach Abschluss einer Umfrage in der Regel nicht mehr oder nur mit einem methodisch-statistischen Zusatzaufwand korrigieren.

Mit Recht weist die einschlägige Denkschrift der Deutschen Forschungsgemeinschaft (Kaase, 1999: 96) darauf hin, dass sich Methodenstudien in der Umfrageforschung in der Regel nur mit einzelnen Aspekten und Teilfragen befassen, dass eine Methodologie der Qualitätsbewertung von Umfragen aber eine ganzheitliche Perspektive einnehmen müsse. Sie schlägt den Begriff des *Verfahrensmodells* als Beschreibung eines *Umfragetypus* vor. Ein Verfahrensmodell besteht aus der Spezifikation verschiedener Komponenten oder Dimensionen des Gesamtablaufs einer Umfrage, wobei folgende Komponenten unterschieden werden:

- Ausgangsstichprobe (Adäquation, coverage) (Häder/Häder, Kapitel 18 in diesem Band)
- Ausfälle in der Befragung (nonresponse) (Engel/Schmidt, Kapitel 23 in diesem Band)
- Interviewmethode (mode effects)
- Datenerfassung und -verarbeitung (Fehlermöglichkeiten/Fehlervermeidung) (Lück/Landrock, Kapitel 28 in diesem Band)
- Zeitbedarf (Schnelligkeit)
- Dokumentation (Transparenz) (Friedrichs, Kapitel 16 in diesem Band)
- In Bezug auf die Qualitätsbewertung wird die Trennung von zwei Ebenen vorgeschlagen:
  - die *erreichbare* Qualität der jeweiligen Verfahrensmodelle;
  - die tatsächlich erreichte Qualität der Verfahrensmodelle.

Bei der Übernahme dieses Vorschlags wird man sich der Gefahr einer vorschnellen Relativierung auf niedrige Qualitätsnormen bewusst sein. Jeder einzelne Schritt des Umfrageprozesses besitzt eigene Qualitätsmerkmale, die bei seiner Realisierung beachtet werden sollten. Dies beginnt bereits bei der Präzisierung der Forschungsfragestellung und der in ihr verwendeten Begriffe sowie der Operationalisierung, in der die theoretischen Begriffe durch Zuordnung empirischer Indikatoren messbar gemacht werden müssen und setzt sich allen Schritten fort. Die Optimierung aller Schritte ist auch der Weg zur Minimierung des TSE.

## Literatur

- Beatty, Paul/Herrmann, Douglas (2002): To Answer or Not to Answer: Decision Processes Related to Survey Item Nonresponse. In: Groves et al. (Hg.): 71-85
- Bethlhem, Jelke (2002): Weighting Nonresponse Adjustments Based on Auxiliary Information. In: Groves et al. (Hg.): 275-287
- Bethlehem, Jelke (2009): Applied Survey Methods. Hoboken, NJ: Wiley
- Biemer, Paul P./Lyberg, Lars E. (2003): Introduction to Survey Quality. Hoboken, NJ: Wiley
- Biemer, Paul P./Trewin, Dennis (1997): A Review of Measurement Error Effects on the Analysis of Survey Data. In: Lyberg et al. (Hg.): 603-632
- Dippo, Cathryn S. (1997): Survey Measurement and Process Improvement. In: Lyberg et al. (Hg.): 457-474
- Faulbaum, Frank/Prüfer, Peter/Rexroth, Margit (2009): Was ist eine gute Frage? Wiesbaden: VS Verlag
- Fricker, Scott/Galesic, Mirta/Tourangeau, Roger/Yan, Ting (2005): An Experimental Comparison of Web and Telephone Surveys. In: Public Opinion Quarterly 69: 370-392
- Groves, Robert M. (2004): Survey Errors and Survey Costs (2nd edition). Hoboken, NJ: Wiley
- Groves, Robert M. (2006): Nonresponse Rates and Nonresponse Bias in Household Surveys. In: Public Opinion Quarterly 70: 646-675
- Groves, Robert M./Couper, Mick P. (1998): Nonresponse in Household Interview Surveys. Hoboken, NJ: Wiley
- Groves, Robert M./Lyberg, Lars E. (2010): Total Survey Error. In: Public Opinion Quarterly 74: 849-879
- Groves, Robert M./Peytcheva, Emilia (2008): The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. In: Public Opinion Quarterly 72: 167-189
- Groves, Robert M./Dillman, Don A./Eltinge, John L./Little, Roderick J. A. (Hg.) (2002): Survey Nonresponse. Hoboken, NJ: Wiley
- Groves, R.M./Fowler, Floyd J./Couper, Mick P./Lepkowski, James M./Singer, Eleanor/Tourangeau, Roger (2009): Survey Methodology (2nd edition). Hoboken, NJ: Wiley
- Kaase, Max (Hg.) (1999): Qualitätskriterien in der Umfrageforschung. Berlin: Akademie Verlag
- Keeter, Scott (2006): The Impact of Cell Phone Noncoverage Bias on Polling in the 2004 Presidential Election. In: Public Opinion Quarterly 70: 88-98
- Keeter, Scott/Kennedy, Courtney/Clark, April/Tompson, Trevor/Mokrzycki, Mike (2007): What's Missing from National Landline RDD Surveys. The Impact of the Growing Cell-Only Population. In: Public Opinion Quarterly 71: 772-792
- Krosnick, Jon A. (1991): Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. In: Applied Cognitive Psychology 5: 213-236
- Lewis, Charles (2007): Selected Topics in Classical Test Theory. In: Rao/Sinharay (Hg.): 29-43

- Lord, Frederic M./Novick, Melvin R. (1968): Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley
- Lyberg, Lars E./Biemer, Paul/Collins, Martin/De Leeuw, Edith D./Dippo, Cathryn S./Schwarz, Norbert/Trewin, Dennis (Hg.) (1997). Survey Measurement and Process Quality. New York: Wiley
- Peytchev, Andy/Carley-Baxter, Lisa R./Black, Michele C. (2012): Coverage Bias in Variances, Associations, and Total Error from Exclusion of the Cell Phone-Only Population in the United States. In: Social Science Computer Review 28: 287-302
- Presser, Stanley/Rothgeb, Jeniffer M./Couper, Mick P./Lessler, Judith T./Martin, Elizabeth/Martin, Jean/Singer, Eleanor (Hrsg.) (2004): Methods for Testing and Evaluating Survey Questions. Hoboken, NJ: Wiley
- Rao, C. Radhakrishna/Sinharay, Sandip (Hg.) (2007): Handbook of Statistics 26: Psychometrics. New York: Springer
- Särndal, Carl-Erik/ Swensson, Bengt/Wretman, Jan (1992): Model Assisted Survey Sampling. New York: Springer
- Sakshaug, Joseph W./Yan, Ting/Tourangeau, Roger (2010): Nonresponse Error, Measurement Error, and Mode of Data Collection. In: Public Opinion Quarterly 74: 907-933
- Smith, Tom W. (2011): Refining the Total Error Perspective. In: International Journal of Public Opinion Research 23: 464-484
- Weisberg, Herbert F. (2005): The Total Survey Error Approach. Chicago: University of Chicago Press
- West, Brady T./Olson, Kristen (2010): How Much of Interviewer Variance is Really Nonresponse Error Variance? In: Public Opinion Quarterly 74: 1004-1026

---

## **Teil 3**

### **Forschungsparadigmen in der qualitativen Sozialforschung**

Jörg Strübing

---

## 32.1 Zur Entstehung der Grounded Theory

Der Forschungsstil der Grounded Theory (im folgenden: GT) wurde in den 1960er Jahren von Barney Glaser und Anselm Strauss entwickelt und 1967 in „The Discovery of Grounded Theory“ (1967/1998) erstmals publiziert. Veröffentlicht in einer Zeit des intellektuellen und politischen Aufbruchs und gezielt an die junge Generation revoltierender, nach neuen Wegen suchender Sozialforscher adressiert, wurde das Buch zu einem Klassiker der empirischen Sozialforschung. Obwohl es bis heute gerne als Lehrbuch für die Methode der Grounded Theory verstanden und genutzt wird, handelt es sich tatsächlich eher um ein Skizzenbuch, in dem zentrale Verfahren der GT grob umrissen und in ihrer Funktion bestimmt, nicht aber systematisch und in sich konsistent dargelegt werden.

Die Bezeichnung des Forschungsstils als „Grounded Theory“ bietet gelegentlich Anlass zu Missverständnissen. Der Grund dafür liegt in der substantivischen und damit objektivierenden Form des Labels „Grounded Theory“, das seine Doppeldeutigkeit daraus bezieht, dass es die zentrale Qualität der mit dem Verfahren zu erarbeitenden Theorien zugleich zum Namen für das Verfahren selbst erhebt. Damit verweist die Bezeichnung „Grounded Theory“ gleichermaßen auf *Prozess und Ergebnis*, auf *problemloses Forschungshandeln* und auf die dabei *hervorgebrachten gegenstandsbezogenen Theorien*, gerade weil das Ergebnis angemessen nur aus dem Arbeitsprozess heraus zu verstehen ist, in dem es produziert wurde. Dieses Verständnis von Theorie als Prozess deckt sich auch mit den großen argumentativen Linien im sozialtheoretischen Werk von Strauss (Strübing 2007a).

Wenn daher verschiedentlich in der Sekundärliteratur zur GT vorgeschlagen wird, die scheinbare Missverständlichkeit des Labels durch das Suffix „-Methodologie“ (so z.B. Mey/Mruck 2009) oder „-Methodik“ (Breuer 2009) zu heilen und statt von GT von GTM zu sprechen, dann ist dadurch nichts gewonnen. Im Gegenteil: Gerade mit dem Label GT wird ausdrücklich darauf verwiesen, dass es sich hier eben nicht um eine Methode, aber auch nicht um eine Methodologie handelt, sondern um einen *Forschungsstil*.

Das „Discovery“-Buch konnte auch deshalb eine übermäßige Bedeutung als Pseudo-Lehrbuch erlangen, weil sowohl Strauss, als auch Glaser zum Forschungsstil der Grounded Theory eine Dekade lang nicht weiter publizierten. Erst 1978 veröffentlichte Barney Glaser unter dem Titel „Theoretical Sensitivity“ ein Buch, in dem er sein Verständnis von Grounded Theory systematisch und in nachvollziehbarer Form darlegt (Glaser 1978). Es sollten noch einmal fast zehn Jahre vergehen, bis Anselm Strauss ein eigenes Methodenlehrbuch veröffentlichte (Strauss 1987/1991). Obwohl Strauss in diesem Buch ganze Abschnitte aus Glasers Werk zitiert, werden doch zwischen beiden Verfahrensdarstellungen zunehmend Unterschiede sichtbar, die schließlich in einem offenen Dissens münden, als Strauss gemeinsam mit Juliet Corbin unter dem Titel „Basics of Qualitative Research“ (Strauss/Corbin 1990/1996) ein weiteres einführendes Lehrbuch zum Verfahren der GT publiziert und Glaser eine polemische Replik darauf verfasst (Glaser 1992; vgl. dazu auch: Strübing 2011).

Von den beiden Gründern der GT lebt heute nur noch Glaser, Strauss starb bereits 1996. Im vorliegenden Text jedoch beziehe ich mich ausdrücklich auf die GT-Version von Anselm Strauss, die heute von einigen seiner Schülerinnen sowie von seiner ehemaligen Weggefährtin Juliet Corbin fortgeführt und weiterentwickelt wird. Die Gründe für diese Beschränkung liegen in den m.E. erkenntnistheoretisch nicht haltbaren Positionen Glasers, der einen in der Wissenschaftstheorie längst überwundenen, naiven Induktivismus vertritt. Diese Festlegung geht einher mit der These, dass es erforderlich ist, sich zu entscheiden, welche der beiden Varianten von GT wir jeweils meinen, wenn wir uns für empirische Studien auf diesen Forschungsstil berufen.

---

## 32.2 Theoretische Grundlagen

Das in Lehrtexten verbreitete instrumentelle Verständnis von Forschungsmethoden über sieht, dass empirische Methoden aus bestimmten Denkschulen und Wissenschaftstraditionen heraus entstehen und daraus ihre Funktionsbestimmung und Legitimation beziehen. Dies wird an zwei Punkten im Forschungsverlauf bedeutsam: Zum einen bei der Interpretation von Verfahrensregeln in der empirischen Arbeit, zum anderen aber in der Wissenschaftskommunikation, wenn also die geleistete Forschungsarbeit präsentiert (Meyer/Meier zu Verl, Kapitel 15 in diesem Band), legitimiert und für ihre Relevanz argumentiert wird. Theoretisch und erkenntnislogisch informierte Darstellungen entgehen leichter der Gefahr, von unangemessenen wissenschaftstheoretischen Positionen vereinnahmt, zu Unrecht mit deren Kriterien evaluiert und schließlich mit falschen Argumenten kritisiert zu werden.

Anselm Strauss entstammt der interaktionistischen Theorietradition. Als Student bei Herbert Blumer an der University of Chicago lernt er die klassische symbolisch-interaktionistische Theorieperspektive kennen, die George Herbert Mead entwickelt und Blumer weiter expliziert und schließlich benannt hat. Zudem wird Strauss von dem kanadischen Soziologen Everett C. Hughes beeinflusst, der stärker durch die ethnographische

Forschungstradition (Knoblauch, Kapitel 36 in diesem Band) des Robert E. Park geprägt ist. An dieser Schnittstelle zwischen Interaktionismus und humanökologischer Feldforschungstradition, geschult durch Blumers scharfsinnige methodologische Argumentation, aber auch desillusioniert, was dessen empirisch-methodische Kompetenz betrifft, beginnt Strauss ab Mitte der 1950er Jahre eine qualitativ-empirische Praxis der Ethnographie in Organisationen zu entwickeln. Für Strauss besteht das Problem vor allem darin, von einer rein deskriptiven Repräsentation empirischer Phänomene zu einer systematischen Theoriebildung zu kommen. Damals gibt es noch kaum gründlich durchgearbeitete Vorschläge zu einer empirisch begründeten Theoriegenese. Schon vor der gemeinsamen Arbeit mit Glaser beginnt Strauss daher in der Studie „Boys in White“ (Becker et al. 1961) eine konsequent komparative Strategie der Datenanalyse zu praktizieren und damit den Grundstein für die spätere GT zu legen.

Die interaktionistische Sozialtheorie basiert auf den wissenschafts- und erkenntnistheoretischen Postulaten des amerikanischen Pragmatismus (Shalin 1986). Das wichtigste verbindende Element zwischen GT und Pragmatismus ist die Auffassung von Handeln und Forschen als Problemlösungsprozess (Dewey 2002).

Die Grundidee ist der Spannungswechsel zwischen Gewissheit und (praktischem) Zweifel: Im vorreflexiven Strom routinierten Handelns besteht solange Gewissheit, bis das gewohnheitsmäßige Handeln an Grenzen stößt, gehemmt wird und Zweifel aufkommt. Die dann einsetzende Serie von Aktivitäten lässt sich am besten als eine iterativ-zyklische Sequenz von Problembestimmungs- und -lösungsprozessen beschreiben (Strübing 2013, Abb. 2). Entscheidend für die Problemlösung ist, dass diese immer abduktive Erkenntnisprozesse erfordert (Reichert, Kapitel 2 in diesem Band). Abduktiv gewonnene mögliche Lösungen werden gedankenexperimentell auf ihre Plausibilität geprüft und schließlich praktisch erprobt. Soweit das problematisch gewordene Handeln sich auf dieser Basis erfolgreich weiterführen lässt, geht die so gefundene Problemlösung in das nun erweiterte Handlungsrepertoire ein und aus Zweifel wird erneut Gewissheit. Im anderen Fall wird ein erneuter, nun zusätzlich informierter Problemlösungsprozess initiiert. Daraus ergibt sich ein iterativ-zyklischer Verlauf des Problemlösens, in dem induktive, abduktive und deduktive Schlussformen ineinander greifen.

Vom pragmatistischen Modell lässt sich zunächst lernen, dass auch die Wissenschaft nicht allein auf der Basis formallogischer Schlüsse zu neuen Erkenntnissen kommt. Der Wissensfortschritt liegt vielmehr gerade in jenem kreativen Moment der Abduktion, in dem spontan, wenngleich nicht zufällig, mögliche Lösungen zur Integration zuvor disperater Wahrnehmungen und Wissensbestände aufscheinen. Dies hat allerdings den Preis, dass die so gewonnene Erkenntnis eben nicht formallogisch abgesichert und legitimiert ist, sich also immer erst im praktischen Handeln bewähren muss, um als Gewissheit zu gelten.

Die Denkfigur einer Spannung von Gewissheit und Zweifel verweist zugleich auf die axiomatische Annahme, dass sich die Bedeutung aller Objekte der physischen wie der sozialen Umwelt allein handlungspraktisch bestimmen lässt, den Dingen mithin nicht „an sich“ Bedeutung innewohnt. Handeln ist diesem Grundmotiv pragmatistischen Denkens zufolge nicht einfach das „Fenster“ durch das hindurch wir einen Blick auf die Realität

erhaschen können, es ist selbst der Modus von Realität. Für Pragmatisten ist die „Welt da draußen“ eher eine undefinierte Offenheit, die erst im Handeln als je spezifische Art von Widerständigkeit erfahren wird, mit der umgegangen werden muss und die so, in praktischen Problemlösungsprozessen, Bedeutung erlangt.

Dieser Realitätsbegriff hat gravierende Konsequenzen für zentrale Konzepte der empirischen Forschung, insbesondere für das Verständnis von Daten. Im Unterschied zur Vorstellung einer sozialen Konstruktion von Daten betont etwa Mead mit dem von ihm verwendeten Bild des „Herausmeißelns“ der Daten aus der empirischen Welt (Mead 1938: 660) sehr deutlich die Idee, dass sich auch die Aktivität der Herstellung von Daten an einer materiellen oder sozialen Widerständigkeit abarbeiten muss: Was ein Datum ist, wird nicht in freier Vereinbarung unter Forschenden ausgehandelt, sondern fortwährend durch den substantiellen Gehalt der empirischen Welt gerahmt. Daten sind demnach das prozesshafte Produkt der Interaktion von Forschenden und Feld, die durch die sich ebenfalls entwickelnde Forschungsfrage als „Problem“ strukturiert wird. Das pragmatistische Realitätsverständnis konterkariert die Idee einer Universalität von Realität: Weil das Handeln als Modus der Realitätshervorbringung an die je unterschiedlichen Perspektiven der Akteure gebunden ist, muss grundsätzlich von Realität im Plural ausgegangen werden – auch wenn die Widerständigkeit der „Welt da draußen“ im Verbund mit der sozialen Integration der Handelnden (Sozialisation, Enkulturation) bereits ein hohes Maß an Kongruenz erzeugt. Wenn aber schon Realität nicht universell gedacht wird, dann kann auch das Wissen über Realität nicht universell sein, womit das konventionelle, für die abendländische Wissenschaft so zentrale Verständnis objektiven Wissens in Frage gestellt wird (Strübing 2007b).

Wie aber kann empirische Forschung ohne diese Form objektiven Wissens funktionieren und legitimiert sein? In pragmatistischer Perspektive wird dieses Problem dahingehend aufgelöst, dass „wahres“ im Sinne von richtigem Wissen solches ist, das eine „Steigerung der Handlungsmacht gegenüber einer Umwelt“ (Joas 1992: 29) hervorbringt – Handelnde also in die Lage versetzt, sich kompetenter in ihrer Umwelt zu bewegen. Insofern zielt die GT auf gegenstandsbezogene Theorien mit praktischem Erklärungspotential, die auch für die Akteure im untersuchten Handlungsfeld zu einem verbesserten Verständnis ihrer Praxis beitragen – ganz im Sinne jenes pragmatistischen Problemlösungszyklus nach Dewey.

Ein weiteres zentrales Element pragmatistischen Denkens, das seinen Niederschlag auch in der GT gefunden hat, ist die Kritik gegenüber dichotomen Konstruktionen, wie Leib/Seele, Akteur/Umwelt, Mensch/Natur etc., die mehr Erkenntnismöglichkeiten verstellen als sie eröffnen. Im Pragmatismus dominiert dagegen in genetischer und prozessorientierter Perspektive die Vorstellung der Ko-Konstitution. Das Konzept dichotom-existentieller Getrenntheit wird durch das einer relationalen Verbundenheit ersetzt: Ohne physiologische Prozesse wäre etwa ein Phänomen wie Geist nicht existent, ohne unseren Geist könnten wir die Physis nicht denken. Ähnliches gilt für Akteure und ihre Umwelt: Für die Untersuchung sozialer Prozesse ist nicht die Getrenntheit, sondern die Verbundenheit von Handelnden mit ihrer Umwelt zentral. Die relational zu bestimmenden Prozesse zwischen den nur scheinbar getrennten Entitäten sind die entscheidenden Objekte sozialwissenschaftlichen Erkenntnisgewinns: Eine Gruppe ist eine Gruppe dadurch, dass sie im

---

Interaktionsprozess sich selbst so konstituiert, aber eben auch von anderen so behandelt wird.

Perspektivisch gebundene Wirklichkeitskonstitution und eine anti-dualistische, auf Genese und Prozesshaftigkeit orientierende Analyseeinstellung sind auch die Grundlagen sowohl des soziologisch-theoretischen wie des empirisch-methodischen Denkens von Anselm Strauss. So wenig Theorien von der empirischen Welt unabhängig sind, so wenig sind empirische Methoden theorie- und gegenstandsneutrale ‚Instrumente‘ des Forschens. Jedes strukturierte Denken über methodische Wirklichkeitszugänge impliziert neben epistemologischen und wissenschaftstheoretischen immer auch sozialtheoretische Grundannahmen – die mehr oder weniger plausibel sein mögen, am Ende aber immer axiomatische Setzungen bleiben: riskant und ohne die Möglichkeit einer logisch zwingenden Letztbegründung, da sie empirisch allenfalls irritierbar, aber weder falsifizierbar noch verifizierbar sind.

Die Perspektive, die Strauss auf Gesellschaft entwickelt, ist die eines „Universums von enormer Fluidität“ (Strauss 1978: 123), in dem gesellschaftliche Phänomene immer wieder entstehen, vergehen und neu und modifiziert hervorgebracht werden. All dies geschieht im Handeln, also nicht z.B. durch soziale oder materielle Strukturen, die sozusagen durch die Akteure hindurch im Handeln wirksam werden. Ebenso wenig wie Daten die Theorie erzeugen, erzeugen Strukturen neue Strukturen. Es bedarf jeweils des interpretativ-creativen, problemlösenden Handelns der unterschiedlichen Akteure, um situative Gegebenheiten wahrzunehmen, mit ihnen umzugehen und zu neuen Gegebenheiten zu gelangen. Auch in diesem Punkt erweisen sich Alltagshandeln und Forschung sowie Sozialtheorie und empirische Methoden als Kontinuen.

---

## 32.3 Das praktische Vorgehen

### 32.3.1 Prinzipien

Gerade weil GT nicht als Methode im Sinne eines präskriptiven Sets von Verfahrensregeln zu verstehen ist, sondern als ein an die konkrete Forschungspraxis flexibel anzupassendes Gerüst von Verfahrensvorschlägen (Strauss 1991: 33), kommt den Arbeitsprinzipien, die Strauss zu Beginn seines Einführungsbuches formuliert, besondere Bedeutung zu. Sie können die Forscherin über den Sinn der im Einzelnen vorgeschlagenen Verfahrensschritte orientieren und deren situative Interpretation anleiten.

1. Strauss fasst *Forschen als Arbeiten* auf, also als eine Reihe von Tätigkeiten, die zu leisten sind, um die Aufgabe erfolgreich bewältigen zu können (Strauss 1991: 25). Damit rückt die praktische Forschungsorganisation ins Zentrum der methodologischen Betrachtung. In dieser Perspektive werden zwei Vorschläge gemacht, die auf den ersten Blick und verglichen mit der Darstellung anderer Methoden überraschen: Die *Parallelisierung der Arbeitsschritte* und die *Sequenzierung des Samplings*. Anstatt Datengewinnung, Datenanalyse und Theoriebildung als nacheinander zu bewältigende Arbeitsschritte

anzulegen, besteht die Forschungslogik der GT gerade darin, dass alle drei Tätigkeiten parallel betriebene Modi des Forschens sind, die sich gegenseitig produktiv beeinflussen. Analytische Ideen bei der Auswertung der Daten sollen nicht nur die Entwicklung der gegenstandsbezogenen Theorie beeinflussen, sie wirken auch, ebenso wie die Theoriegenese selbst, auf den Prozess der Datengewinnung zurück, indem etwa andere Datentypen oder modifizierte Formen der Datengenese eingesetzt werden, aber auch indem die Auswahl der Fälle selbst, also das Sampling, durch die entstehende Theorie gesteuert wird (*Theoretisches Sampling*, s.u.).

2. Parallelität der Arbeitsschritte bedeutet zugleich, dass die *Analyse bereits mit dem ersten Fall beginnt* – und dass mit der Analyse dieses Falles bereits theoretische Aussagen gemacht werden können. Es müssen also nicht erst ganze Sets von Fällen akkumuliert werden, bevor die Untersuchung Ergebnisse erbringt. Umgekehrt aber muss der Fall als ganzer analytisch erschlossen und verstanden werden. Die GT ist somit, wie eine Reihe anderer qualitativ-interpretativer Verfahren auch, ein *einzelfallanalytisches Verfahren*. Als komparative Methode bleibt es in der GT jedoch nicht bei der Analyse eines einzelnen Falles, sie stellt vielmehr einen für die Theoriegenese sehr wichtigen, aber nicht allein entscheidenden Arbeitsschritt dar. Es ist sinnvoll, den *ersten zu analysierenden Fall* mit Bedacht auszuwählen, da er zumindest zu Beginn großen Einfluss auf den Theoriebildungsprozess hat. Hier liegt allerdings ein Problem: Bei einer theoriegenerierenden Vorgehensweise ist gerade zu Beginn einer Studie schwer zu bestimmen, welcher Fall und welches Material als erstes bearbeitet werden sollten. Drei Überlegungen sind hier hilfreich:
  - Erstens wird die Wahl des ersten Falles entdramatisiert, wenn man sich vergegenwärtigt, dass jeder Fall, wenn er dem fraglichen Untersuchungsfeld entstammt, etwas zur gegenstandsbezogenen Theorie beizusteuern hat.
  - Zweitens ist evident, dass, wenn wir den Kern unserer zu entwickelnden Theorie noch nicht kennen, jede Auswahl provisorisch und tentativ sein muss.
  - Drittens schließlich hindert uns nichts daran, bereits Material zu zwei oder drei Fällen erarbeitet zu haben, bevor wir dann einen ersten davon analytisch zu bearbeiten beginnen. Im Prozess der Datengewinnung, also beim Interviewen, Transkribieren, Beobachten, Protokollieren, gewinnen wir bereits einen ersten Eindruck davon, ob das jeweilige Material für unsere Fragestellung gut geeignet ist. Und wenn wir mehr als einen Fall erheben, haben wir die Möglichkeit, denjenigen auszuwählen, der am ergiebigsten zu sein verspricht.
3. In der GT wird die *Kreativität* betont, die für die analytische Arbeit erforderlich ist (z.B. Strauss/Corbin 1996: 11f). Gerade weil Theorie nicht aus Daten emergiert, ist eine kreative Eigenleistung der Forschenden erforderlich. Damit ist nicht künstlerische Freiheit gemeint, sondern die unabdingbare subjektive Leistung der Forschenden im zielorientierten, kontrollierten Prozess der empirischen Untersuchung (Strauss/Corbin 1996: 27ff.). Die Idee kreativer Eigenleistungen birgt offensichtlich die Gefahr idiosynkratischer Abirrungen in sich.

4. Das produktive Moment von Kreativität kann leicht durch subjektive „Überschüsse“ konterkariert werden. Nicht zuletzt aus diesem Grund wird vorgeschlagen, *Forschung als kollektiven Prozess* zu organisieren (Strauss 1991: 68ff). Damit ist nicht arbeitsteilige Projektbearbeitung gemeint, wie sie in größeren Projekten unvermeidlich ist, es geht vielmehr um die gemeinsame analytische Arbeit am Material. Gerade das offene und das axiale Kodieren sind Arbeitsschritte, in denen mehrere Forscherinnen einander ideal ergänzen, aber auch wechselseitig ‚in der Spur halten‘ können.

### 32.3.2 Theoretisches Sampling und Heuristiken des Vergleichens

Für die praktische Analysearbeit der GT sind zwei Verfahrensmodi von zentraler Bedeutung, die hier integriert dargestellt werden, weil sie im Forschungsverlauf eng miteinander verwoben sind: Das „*Theoretische Sampling*“ als sukzessive Konstruktion der Auswahlgesamtheit und die fortgesetzte Iteration zweier einander abwechselnder Vergleichsmodi. Diese „*Methode des ständigen Vergleichens*“ setzt an einer basalen Alltagsheuristik an: Beobachtete Übereinstimmungen und Differenzen sind es, mit denen wir unseren Alltag ordnen und uns verfügbar machen. Wie im Alltag so gilt auch in wissenschaftlichen Vergleichsoperationen, dass das Unterscheidungsvermögen in der handelnden Person angelegt ist. Die zu vergleichenden Objekte sind nicht ‚an sich‘ unterschiedlich oder ähnlich, sondern es bedarf eines Akteurs, der diesen Vergleich auf Basis seiner Erfahrung und seines Wissens vornimmt und anhand seiner Relevanzstrukturen die Vergleichskriterien bestimmt.

In der analytischen Arbeit im Rahmen der GT geht es nicht um alltagspraktische Orientierung, sondern darum, an empirischem Material relevante theoretische Konzepte und Aussagen zu generieren. Als unabdingbare Voraussetzung für diesen Prozess braucht die Forscherin eine Forschungsfrage, im Lichte derer ein empirisches Phänomen überhaupt erst als solches wahrnehmbar und in seiner Bedeutung für das Forschungsproblem zu bestimmen ist. Es ist zunächst die durch Forschungsproblem, gegebenes Vorwissen, thematische Interessen und forschungspraktische Erfahrungen gerichtete Perspektive der Forschenden, die darüber entscheidet, was wir als Phänomen wahrnehmen und was genau seine Relevanz ausmacht.

Dieser Hinweis ist insofern wichtig als in der GT-Literatur mit Blick auf die Methoden des ständigen Vergleichens gelegentlich ein Konzept-Indikator-Modell zur Erläuterung der Entwicklung theoretischer Konzepte aus empirischen Indikatoren verwendet wird, bei dem eine Kette von Indikatoren in einem theoretischen Konzept zu resultieren scheint. Richtig ist daran, dass die fortgesetzte, systematisch-vergleichende Einbeziehung immer weiterer Indikatoren die notwendige Grundlage jeder Konzeptentwicklung bildet: Ein einzelner empirischer Indikator mag den Anstoß für die Entwicklung erster Konturen eines theoretischen Konzeptes geben, und eine fortgesetzte Kette von Indikatoren ist zur Verfeinerung und Spezifizierung des Konzeptes unabdingbar. Doch nur wenn forschendes

Problemlösen diese Indikatoren als relevant für seinen Gegenstand erachtet, wird daraus die Grundlage der Theoriebildung.

Heuristiken des Vergleichens kommen in der GT bei gleichem Grundmuster auf unterschiedlichen Ebenen und in verschiedenen Kodierschritten zum Tragen: Gleichviel, ob im traditionellen Sinne Fälle miteinander verglichen werden oder Vorkommisse von relevanten Phänomenen, immer werden zunächst Vergleichsmaterialien/Fälle herangezogen, die dem ersten Materialstück in den jeweils als relevant bestimmten inhaltlichen Dimensionen möglichst ähnlich sind. In dieser *minimalen Kontrastierung* soll der Kern eines Falltypus oder eines theoretischen Konzeptes erarbeitet werden. Die Grundfrage lautet hier: Was bleibt bei aller unvermeidlichen Variation in den Fällen konstant in Bezug auf das Phänomen, um dessen theoretische Konzeptionalisierung es jeweils geht? Wir finden also in einem kleineren Set gleichartiger Fälle bestimmte Eigenschaften, die über alle Fälle hinweg weitgehend stabile Ausprägungen aufweisen (Dimensionen) und die damit den Kern unseres theoretischen Konzeptes bilden. Zugleich aber fallen meist auch kleinere oder größere Variationen auf, bei denen sich bei genauerer Analyse erweist, dass sie mit bestimmten Kontexten kovariieren. Wenn diese Variationen konzeptionell ausgearbeitet werden, ergeben sich Subkonzepte.

Zugleich wird der stabile Kern des zentralen Konzeptes zunehmend gefestigt und verdichtet, indem weitere homogene Fälle herangezogen und verglichen werden. Dabei stellt sich die Frage des Abbruchkriteriums: Wie lange soll dieser Vergleichsmodus beibehalten werden? Die GT schlägt hier das Kriterium der *theoretischen Sättigung* vor (Glaser/Strauss 1998: 68ff.; Strauss 1991: 49): Wenn das fortgesetzte Einbeziehen weiterer homogener Fälle keine zusätzlichen Eigenschaften mehr erbringt, gilt dieser Theoretisierungsschritt als gesättigt und ein gleichartiges Fortfahren würde keinen zusätzlichen Erkenntnisgewinn erbringen.

Basierend auf den im ersten Vergleichsschritt zentralen Kriterien werden nun im nächsten Schritt gezielt abweichende Fälle aufgesucht: Was ändert sich, aber auch: was bleibt stabil, wenn wir den Kontext der Fälle variieren lassen? Diesen Vergleichsschritt kann man als Test der vorläufigen Theoretisierungsleistung verstehen: Wie weit trägt das bisher ausgearbeitete Konzept, welche Konstellationen erklärt es, und ab wann wird eine Modifikation welcher Art erforderlich? Hier wird also die „konzeptuelle Repräsentativität“ (Strübing 2014: 31) eines Konzeptes getestet. Im Ergebnis erhalten wir nicht nur Auskunft über die Reichweite des jeweiligen Konzeptes, sondern auch Hinweise auf für einen angebaren Geltungsbereich valide alternative Konzepte.

In diese multiplen Vergleichsaktivitäten passt sich auch die Samplingstrategie der GT ein, das *theoretische Sampling* (Glaser/Strauss 1998: 53ff.; Strauss 1991: 70f.; Strübing 2014: 29ff.): Wenn ein erster Fall mit weiteren homogenen Fällen verglichen werden soll, dann muss erstens sichergestellt sein, dass weitere homogene Fälle im Material vorhanden sind. Zweitens müssen die Kriterien für Homogenität und Heterogenität dem laufenden Theoriebildungsprozess entstammen, damit die Relevanz der Vergleiche und Kontrastierungen gewährleistet ist. Dies schließt eine Vorab-Festlegung des Samples aus, denn eine solche müsste sich auf Kriterien beziehen, deren Angemessenheit und Relevanz noch völlig offen

ist. Daher wird in der GT wird das Sampling sukzessive im Projektverlauf durchgeführt und an Auswahlkriterien aus der entstehenden Theorie orientiert.

Das Wechselspiel von minimaler und maximaler Kontrastierung findet im Projektverlauf fortgesetzt und auf verschiedenen analytischen Ebenen auch parallel statt, also etwa indem Befragte als Fälle miteinander verglichen werden, aber auch z.B. bestimmte Situationsdefinitionen einer Person unter wechselnden Umständen oder die Bewältigung bestimmter Ereignisse über verschiedene Personen-Fälle hinweg. Die Unterscheidung von Fällen und Phänomenen erweist sich hier als im Grund obsolet, denn es geht beim Begriff des Falls immer um eine *relationale* Bestimmung: Wofür ist etwas ein Fall? Welche im Material gefundenen oder neu erhobenen Daten sind als Fälle für welches Phänomen und welches theoretische Konzept des Phänomens zu betrachten? Was ein Fall ist, kann immer wieder variieren. Das bedeutet für die analytische Arbeit, dass wir fortwährend auf unterschiedlichen Ebenen im Material Auswählen treffen, um Konzepte und deren Variationen und Reichweiten zu erarbeiten.

Forschungspraktisch ist diese Samplingstrategie nicht immer strikt durchzuhalten. Häufig sind die am Fall erarbeiteten Auswahlkriterien so beschaffen, dass sie den Fällen, von ‚außen‘ nicht anzusehen sind, vor der Datengewinnung also gar nicht mit Sicherheit festzustellen ist, ob ein Fall tatsächlich den Kriterien entsprechen wird. In anderen Projektkonstellationen wiederum ist der Feldzugang zeitlich so eng limitiert, dass eine gründliche Analyse der Fälle parallel zu ihrer Gewinnung nicht zu leisten ist. In diesen Fällen kann es im Sinne einer flexiblen Anpassung der Verfahrensregeln z.B. erforderlich sein, zu Beginn eines Projektes in einer komprimierten Feldphase einen größeren Materialumfang zu erarbeiten und die Regeln des theoretischen Samplings dann erst im Verlauf der analytischen Bearbeitung des Materialkonvoluts anzuwenden – und die sich dabei notwendig ergebenden Einschränkungen und Risiken (z.B. das Fehlen bestimmter wünschenswerter Kontrastierungsfälle) in Kauf zu nehmen.

### 32.3.3 Dateninterpretation und -analyse als Kodieren

Strauss unterscheidet drei Modi des Kodierens: das offene, das axiale und das selektive Kodieren (Strauss 1991: 57ff.). Glaser beschränkt sich auf zwei Kodierschritte, das offene und das theoretische Kodieren, wobei Leistungen des axialen und selektiven Kodierens bei Glaser im Arbeitsschritt des theoretischen Kodierens erbracht werden (Glaser 1978).

Im *offenen Kodieren* werden thematische Zugänge zum Material geschaffen sowie relevante Materialstücke ausgewählt und im Detail analysiert. Insbesondere für den Beginn der analytischen Arbeit und wenn es um das Erschließen neuer thematischer Dimensionen, weiterer Falldomänen oder besonders problematischer Materialabschnitte geht, kommt dabei die „Line-by-line-Analyse“ zum Einsatz: Ein Textstück wird kleinschrittig-sequentiell (Kurt/Hebrik, Kapitel 33 in diesem Band) betrachtet, um hinter der leicht für selbstverständlich und vertraut genommenen Oberfläche des manifesten Textes weitere Sinndimensionen zu erschließen, den Text also „aufzubrechen“ (Strauss/Corbin 1996:

45). Diese Metapher verweist auf die Anstrengung, die gegenüber dem Text aufzubringen ist, um ihn in theoriegenetischer Perspektive verfügbar zu machen: Sätze werden nicht als geschlossene ganze gelesen, sondern zunächst einmal nur einzelne Worte und Satzabschnitte. Diese werden nun aktiv befragt: Was wird hier thematisiert? Was ist für die Forschungsfrage relevant? Welche Situationsdefinition, welches Handlungsproblem zeigt der Sprecher durch seine Art der Präsentation des Themas an? Was hätte hier auch stehen können? Die dabei erforderliche aktive Leistung der Forscherin wird von Strauss ausdrücklich betont: „Der Punkt ist tatsächlich der, dass das Potential an Aspekten nicht so sehr im Dokument selbst liegt als vielmehr in der Qualität der Beziehung zwischen Datum und forschendem Geist sowie in der Übung, die der Forscher hat, wenn er die Arbeit des offenen Kodierens tatkräftig und phantasievoll angeht.“ (Strauss 1991: 58)

Die hier gemeinte Kompetenz der Forscher bezeichnen Glaser und Strauss als *theoretische Sensibilität* (1998: 54). Damit ist, wie Strauss und Corbin später formulieren, „ein Bewusstsein für die Feinheiten in der Bedeutung von Daten (gemeint), das abhängt „vom vorausgehenden Literaturstudium und von Erfahrungen, die man entweder im interessierenden Phänomenbereich selbst gemacht hat oder die für diesen Bereich relevant sind“ (Strauss/Corbin 1996: 25). Theoretische Sensibilität ist die Ausgangsbasis für Vergleichsheuristiken, wie sie bereits in der Line-by-line-Analyse zum Einsatz kommen. Dies geschieht etwa mit „*weithergeholt Vergleichen*“ (Strauss/Corbin 1996: 69f.), bei denen die fraglichen Phänomene mit auf den ersten Blick weit entfernt liegenden Phänomenen konfrontiert und auf Ähnlichkeiten und Unterschiede befragt werden. Vergleiche mit Phänomenen außerhalb der konkreten Studie sind hilfreich und zielführend, weil sich an ihnen das Typische des eigentlich interessierenden Phänomens offenbart. Der weithergeholt Vergleich erschließt damit sukzessive weitere Eigenschaften und Dimensionen von Konzepten.

Die Line-by-line-Analyse ist nicht die einzige Form offenen Kodierens und wird auch nicht auf alles zu analysierende Material angewandt. Das wäre auch praktisch kaum zu leisten, denn der Aufwand für die Bearbeitung schon kurzer Textpassagen ist beträchtlich. Strauss notiert zu den alternierenden Modi des offenen Kodierens: „Wenn ein Kode eine relative Sättigung erreicht hat (...), dann wird der Forscher die Daten automatisch schneller durchgehen, in der Zeile-für-Zeile Analyse Wiederholungen finden und folglich die Daten überfliegen, bis etwas Neues seine Aufmerksamkeit erregt. Dann beginnt wieder die minutiöse Untersuchung.“ (Strauss 1991: 61)

Insbesondere die Integration weiterer aus dem theoretischen Sampling gewonnener Daten ist häufig ein Anlass zu erneuter Line-by-line-Analyse.

### **32.3.4 Exkurs: Was ist ein Konzept?**

Die Frage, was im Sinne der GT ein theoretisches Konzept ist, berührt gerade den prekären Übergang von der Empirie zur Theorie, an ihr entscheidet sich, ob die Auswertung der Daten in einer Paraphrase des Materials endet oder in eine theorieförmige Gestalt mündet.

Die Entwicklung theoretischer Konzepte ist jener Schritt im Forschungsprozess der GT, bei dem die Heuristik ständigen Vergleichens vorrangig Verwendung findet. Beginnend mit einer ersten empirischen Instanz, dem Auftauchen eines empirischen Phänomens in einem ersten Fall, wird sukzessive nach weiteren Instanzen des gleichen Phänomens gesucht. Die nacheinander in Betracht gezogenen Instanzen werden im Blick auf die interessierende Untersuchungsfrage miteinander verglichen und Ähnlichkeiten und Unterschiede herausgearbeitet. Ziel ist dabei, jene Eigenschaften herauszuarbeiten, die für das sich abzeichnende Konzept „wesentlich“ sind, die also für Existenz und Funktionieren der damit in der jeweils rekonstruierten Perspektive bezeichneten Phänomene konstitutiv sind – und das Konzept damit von anderen zu unterscheiden erlauben. Diese Eigenschaften sind zu trennen von je konkreten, mehr oder weniger zufälligen Begleiterscheinungen. Allerdings: Was wesentlich und was eher zufällig und beliebig ist, das gilt es in der vergleichenden Analyse erst herauszufinden.

Um sicherzustellen, dass wir die relevanten Eigenschaften und Dimensionen eines Konzeptes sorgfältig und angemessen herausgearbeitet haben, kommt das theoretische Sampling zum Einsatz: Weitere Vorkommnisse eines am ersten Fall erarbeiteten Konzeptes im Material werden sukzessive in die Analyse einbezogen. Durch solch fortgesetzte Vergleiche mit weiteren, ähnlichen Phänomenen sind wir dann nicht nur in der Lage das Konzept genauer zu bestimmen, wir werden auch zu Differenzierungen gelangen und Subkonzepte entwickeln. Diese Differenzierungen in der Konzeptentwicklung sind besonders wichtig, weil es im Fortgang der Analyse und Theorieentwicklung von Interesse sein wird zu bestimmen, welche Variante des Konzeptes in welchem Typ von Situation anzutreffen ist (und letztlich natürlich auch: warum). In anschließenden *kontrastiven Vergleichen* werden zusätzlich Vorkommnisse des Ausgangskonzeptes untersucht, die sich gerade durch eine Variation der den Ausgangsfall prägenden Konzeptdimensionen auszeichnen.

### 32.3.5 Axiales Kodieren

Indem wir weitere Vorkommnisse unterschiedlicher Varianten des Ausgangskonzeptes aufsuchen, interessiert zunehmend die Frage nach Ursachen, Umständen und Konsequenzen dieser unterschiedlichen Varianten. Darauf orientiert sich das *axiale Kodieren*: Es geht um das Kodieren „um die Achse“ einer zentralen Kategorie. Erst so kann aus der Entwicklung theoretischer Konzepte schließlich eine Theorie erwachsen, denn eine reine Taxonomie von Sub-Konzepten, Konzepten und Kategorien ist noch nicht erklärend. Axiales Kodieren zielt also auf erklärende Bedeutungsnetzwerke, die die jeweils fokussierte Kategorie möglichst umfassend erklären. Dabei werden nicht alle im Material identifizierten Phänomene systematisch vergleichend untersucht, sondern nur diejenigen, von denen – nach dem vorläufigen Stand der Analyse – angenommen werden kann, dass sie für die Klärung der Forschungsfrage relevant sind. Damit wird implizit bereits eine Reihe zunächst vager Hypothesen entwickelt, die im weiteren Gang der Analyse überprüft werden.

Diese mutmaßlich relevanten Konzepte werden im axialen Kodieren nacheinander eingehend befragt. Strauss hat dazu eine Reihe von Aspekten benannt, deren Klärung für die theoretische Einbettung eines Konzeptes zentral ist. Er schlägt vor, „dass Daten nach der Relevanz für die Phänomene, auf die durch eine gegebene Kategorie verwiesen wird, kodiert werden, und zwar nach:

- den Bedingungen
- den Interaktionen zwischen den Akteuren
- den Strategien und Taktiken
- den Konsequenzen“ (Strauss 1991: 57).

Später erweitert er dieses Fragenschema, das er als „*Kodierparadigma*“ bezeichnet, noch um die Aspekte „Kontext“ und „intervenierende Bedingungen“ (Strauss/Corbin 1996: 78ff.). Insgesamt sind die im Kodierparadigma benannten Fragen nur Systematisierungen all jener *Wer-Wie-Wo-Was-Warum-Fragen*, mit denen wir auch im Alltag den Sinn von Ereignissen zu erschließen versuchen, indem wir nach erklärenden Zusammenhängen forschen.

Im Unterschied zum selektiven Kodieren (s.u.) wendet sich das axiale Kodieren explizit einzelnen empirischen Vorkommnissen sowie deren Abstraktionen zu. Es geht nicht um die Beantwortung der umfassenden Forschungsfrage, sondern um die Erklärung des Zustandekommens und der Konsequenzen eines bestimmten Ereignisses bzw. eines bestimmten Ereignistyps. Damit entstehen im axialen Kodieren nach und nach Theorie-Miniaturen, von denen jede in sich den Kern einer Erklärung aufweist – indem sie das jeweils fokussierte Phänomen erklärt und in seinen Konsequenzen bestimmen kann –, die aber noch nicht so weit integriert sind, dass eine befriedigende Antwort auf die Forschungsfrage zu erkennen wäre. Auch für das Kodierparadigma gilt, dass es variabel und anpassungsbedürftig ist, also keine präskriptive Funktion hat. Mindestens für die Soziologie und verwandte Wissenschaften ist der Wert dieser Heuristik allerdings nicht zu unterschätzen – nicht weil hier auf Ursachen und Wirkungen abgezielt wird, sondern weil über die Dimensionen *Kontext* und *intervenierende Bedingungen* ein Angebot zur Verknüpfung situativ-interaktiver Momente mit Prozessen auf entfernteren und abstrakteren Ebenen von Gesellschaftlichkeit gemacht wird.

### **32.3.6 Selektives Kodieren**

Mit dem axialen Kodieren steht eine Antwort auf die zentrale Forschungsfrage immer noch aus. Es fehlt so etwas wie der *rote Faden*, der die vielen kleinen Zusammenhänge, die bislang ausgearbeitet wurden, zu einem kohärenten Theorieentwurf zusammenfasst. Zu einem solchen gelangen wir weder durch fortgesetztes Befolgen von Kodierregeln, noch durch den Rückzug auf logische Schlussverfahren. Es bedarf vielmehr a) einer guten Idee, d.h. der Zusammenhang muss von uns aktiv entdeckt werden, und b) einer Entscheidung: Wir müssen unter verschiedenen, sich anbietenden zentralen Konzepten dasjenige

auswählen, mit dem wir unser Forschungsproblem am besten gelöst sehen. Mit dieser Entscheidung beginnt das sogenannte *selektive Kodieren*, das man auch als ein Re-Kodieren verstehen kann: Nun wird die gesamte bisher erarbeitete Landschaft von Konzepten und Beziehungen zwischen Konzepten noch einmal mit Blick auf die nun fokussierte *Schlüssel- oder Kernkategorie* überdacht, teilweise – durch die nun veränderte Perspektive – umkodiert und so die gesamte analytische Struktur in einen Theorieentwurf integriert: „Die Schlüsselkategorie wird jetzt zur Richtschnur für Theoretisches Sampling und Datenerhebung. Der Forscher sucht nach Bedingungen, Konsequenzen usw., die in Bezug zur Schlüsselkategorie stehen, indem er nach dieser kodiert“ (Strauss 1991: 63).

Die damit implizierte Überarbeitung der bisherigen Kodierungen stellt allerdings nicht eine Korrektur im Sinne der Verbesserung fehlerhafter Kodierungen dar, sondern eine Neujustierung der analytischen Perspektive: Was bislang in Bezug auf eine Reihe unterschiedlicher, im Projektverlauf immer wieder modifizierter, tentativer Sichtweisen kodiert wurde, soll im selektiven Kodieren nun insgesamt auf eine einheitliche Analyseperspektive hin überarbeitet werden. Das Ergebnis ist weder in einem höheren Maße richtig, noch kommt ihm eine erhöhte Gültigkeit zu. Es sollte lediglich ein höheres Maß an Konsistenz aufweisen.

### 32.3.7 Von den Daten zur Theorie – und zurück!

Die GT zielt explizit auf die Generierung von Theorie und setzt sich damit von rein deskriptiv, „explorativ“ orientierten Ansätzen qualitativer Sozialforschung ab. Allerdings ist der jeweilige Theoriebegriff klärungsbedürftig, und muss an axiomatische Setzungen im Bereich der Epistemologie und Wissenschaftstheorie anschlussfähig sein. Für die GT sind es vor allem zwei Merkmale, mit denen ihr *Theorieverständnis* an den pragmatistischen Ursprung anknüpft:

1. Strauss wie auch Glaser verstehen die zu generierenden Theorien primär als *Wissen zur Bewältigung praktischer Probleme im Handlungsfeld*. So zielt die von ihnen durchgeführte ethnographische Studie zu Tod und Sterben in Krankenhäusern darauf, Pflegekräften und Medizinern Wissen für einen reflektierteren Umgang mit dem Problem des Sterbens an die Hand zu geben. Mit der handlungspraktischen Relevanz und ‚Passung‘ der empirisch basierten Theorien ist in pragmatistischer Perspektive zugleich auch ihr zentrales Gütekriterium verbunden: Sie sind so gut, wie die Handlungs- und Problemlösungsfähigkeit im jeweiligen Feld durch sie gesteigert werden.
2. Neben der praktischen Relevanz ist das *prozessuale Verständnis von Theorie* prägend für die GT. Für die praktische Forschungsarbeit bedeutet dies Theorien als grundsätzlich unabgeschlossen zu betrachten (Glaser/Strauss 1967/1998: 50) und den Prozess fortgesetzten Theoretisierens in das Zentrum der analytischen Aktivitäten zu rücken: Die Systematik von theoretischem Sampling, aufeinander aufbauenden Vergleichs- heuristiken, die Moduswechsel im Kodierprozess sowie das fortgesetzte generative Fragen dienen dem Ziel, einen Theoriebildungsprozess auf empirischer Grundlage zu

gestalten, der zugleich kreativ und begründbar ist. Entscheidend ist dabei die Integration der Theoriebildung in den Prozess der Datengewinnung und -analyse. Statt eine Theorie am Ende der Untersuchung „zu schreiben“, wird sie in der GT in einzelnen begründbaren und verifizierbaren Schritten entwickelt.

Das Theorie-Empirie-Verhältnis ist in Bezug auf die GT immer wieder kritisch reflektiert worden, nicht zuletzt in der Glaser-Strauss-Kontroverse (Strübing 2011). Erwecken einige Formulierungen in „The Discovery“ den Eindruck, es würde dort ein Tabula rasa-Modell vertreten (obwohl genau dies explizit ausgeschlossen wird), so betont Strauss in späteren Arbeiten deutlich den heuristischen Wert von praktisch-gegenstandsbezogenem wie von theoretischem Vorwissen. Der Unterschied liegt im Modus des Bezugs auf Wissen, ob also Wissen *ex ante* zu einem die Studie präformierenden theoretischen Rahmen verdichtet oder aber zur Sensibilisierung für generative Fragen genutzt wird.

Abgesehen von erkenntnistheoretischen gibt es auch ganz pragmatische Gründe, sich Vorwissen über den Gegenstand zunutze zu machen: Forscherinnen als ausgebildete Spezialistinnen ihres Forschungsgebietes verfügen per se über einschlägiges Wissen in ihrem Forschungsgebiet. Überdies bietet dieses Wissen ein großes Anregungspotential für die Forschung – vorausgesetzt, wie auch Strauss und Corbin (1994: 277) betonen, die Forscher beziehen sich darauf nicht in der Suche nach Antworten, sondern nach Fragen.

---

## **32.4 Anwendbarkeit, Qualitätssicherung und weiterführende Literatur**

Die GT hat ihren Ursprung in der Erforschung organisationaler Praktiken und hat von dort ausgehend in vielen Feldern der empirischen Sozialforschung Anwendung gefunden. Die große Bandbreite der Anwendungsfelder hat zunächst damit zu tun, dass es sich um einen Forschungsstil handelt, der nicht an bestimmte eingeschränkte Formen der Datengewinnung und entsprechende Datensorten gebunden ist. Reine Interviewstudien kommen hier ebenso zum Einsatz wie Dokumentenanalysen, Webnographien oder ethnographische Feldforschungen. Auch führt die Offenheit und Adaptierbarkeit der Verfahrensvorschläge dazu, dass GT in unterschiedlichen Forschungszusammenhängen mit teilweise spezialisierten Erkenntnisinteressen fruchtbar kombiniert werden können (z.B. Biographieforschung, Diskursanalyse). Grenzen ergeben sich für die GT vor allem aus der Abwägung von Aufwand und Ertrag: In stabilen, bereits gut erforschten Feldern ist der hohe Aufwand einer GT-Studie ebenso wenig vertretbar, wie dort, wo es vor allem darum geht, Makrophänomene neu zu ‚vermessen‘ und entsprechende Theorien zu testen.

Methodologische Prinzipien und methodische Regeln können keine Garantie für „gute“ Forschung und sachangemessene Ergebnisse sein. Dabei liegt das Problem allerdings weniger in der Produktion von Qualität als vielmehr in der Möglichkeit einer verlässlichen externen Überprüfung dieser Qualität durch die wissenschaftliche Öffentlichkeit. Für die GT lassen sich dabei zwei Ebenen unterscheiden: Zum einen integriert sie

eine Vielzahl von qualitätssichernden Maßnahmen, wie ein am Theoriebildungsprozess orientiertes Sampling, die iterative Zyklik vieler kleiner und wiederholter Hypothesentests oder die kontrastiven Vergleichsheuristiken zur Bestimmung der Reichweite theoretischer Konzepte. Sie befördern, sinnvoll praktiziert, eine hohe Qualität der Ergebnisse. Um eine externe Gütebeurteilung zu ermöglichen, sind neben der praktischen Verwendung der Forschungsergebnisse umfassende Informationen über die konkrete Umsetzung der qualitätssichernden Maßnahmen erforderlich. Hier liegt eine zentrale Aufgabe der Forschungspublikationen, die nicht nur Ergebnisse darlegen, sondern auch den Weg zu ihrer Erarbeitung dokumentieren sollen (Strauss/Corbin 1996).

Unter den inzwischen entwickelten Verfahrensvarianten ist zunächst Kathy Charmaz zu nennen (2006), die sowohl Strauss als auch (und insbesondere) Glaser vorwirft, eine objektivistische Methode zu propagieren und die soziale Konstruiertheit der Daten im Forschungsprozess methodisch nicht angemessen zu reflektieren. Charmaz' Vorschlag zu einer Weiterentwicklung der GT fügt indes den hier skizzierten Verfahren keine neuen hinzu und schlägt auch keine wesentlichen Veränderungen dieser Verfahren vor. Es handelt sich eher um einen Appell für einen Perspektivwechsel weg von der von Glaser bemühten Emergenz-Metapher hin zur Vorstellung einer im Forschungsprozess sozial konstruierten Theorie des Gegenstandes.

Einen anderen Ansatz verfolgt Adele Clarke (2004/2012), die wie Charmaz von Strauss ausgebildet wurde. Rhetorisch an den Postmodernismus anknüpfend, entwickelt sie die GT weiter in Richtung auf eine umfassende Situationsanalyse, die expliziter als Strauss auch die diskursiven und historischen Ebenen von Sozialität einbezieht. Ihr Mittel dazu ist eine „kartographische Situationsanalyse“, d.h. eine Technik des „Mapping“, bei der auf unterschiedlichen Ebenen die zur Analyse der Situation relevanten Elemente zueinander in Beziehung gesetzt werden. Ansätze dazu gab es schon bei Strauss und Corbin mit ihrem Vorschlag der Erweiterung des Kodierparadigma zu einer „conditional matrix“ (Strauss/Corbin 1996), doch Clarke differenziert diese Perspektive auch methodenpraktisch stärker aus und verbindet sie mit sozialtheoretischen Konzepten von E.C. Hughes (Ecology Approach), Anselm Strauss (Theorie sozialer Welten) sowie Michel Foucault (Diskurs).

Insgesamt ist allen Versuchen einer Weiterentwicklung der GT gemein, dass sie das Element des Reflexiven und Perspektivischen von Forschungsgegenstand und Forschungsprozess stärker akzentuieren und damit Momente stärken, die der pragmatistischen Ontologie bereits inhärent sind, in den methodologischen Arbeiten von Strauss jedoch wenig betont werden – und bei Glaser völlig fehlen. Zugleich haben viele methodische Elemente der GT Eingang in die Methodenpraxis auch anderer Ansätze gefunden. Insbesondere das theoretische Sampling und das Kriterium der theoretischen Sättigung, aber auch die systematische Variation von Fallvergleichen mit der Heuristik des kontrastiven Vergleichs finden zunehmend Verbreitung in der qualitativen Sozialforschung.

## Literatur

- Becker, Howard S./Geer, Blanche/Hughes, Everett C/Strauss, Anselm L. (1961): Boys in White. Chicago: University of Chicago Press
- Breuer, Franz (2009): Reflexive Grounded Theory. Wiesbaden: VS Verlag
- Charmaz, Kathy (2006): Constructing Grounded Theory. London: Sage
- Clarke, Adele (2004): Situational Analysis. Thousand Oaks: Sage (dt. Übers.: 2012. Situationsanalyse. Wiesbaden: VS Verlag)
- Corbin, Juliet/Strauss, Anselm L. (1990): Grounded Theory Research. In: Zeitschrift für Soziologie 19 (6): 418-427
- Dewey, John (2002): Logik. Frankfurt am Main: Suhrkamp (amerik. Orig.: Logic. New York: Holt, Rinehart and Winston. 1938)
- Glaser, Barney G. (1978): Theoretical Sensitivity. Mill Valley, CA: Sociology Press
- Ders. (1992): Emergence vs Forcing. Mill Valley (CA): Sociology Press.
- Glaser, Barney G./Strauss, Anselm L. (1998/1967): Grounded Theory. Göttingen: H. Huber (amerik. Orig.: The Discovery of Grounded Theory. Chicago: Aldine, 1967).
- Joas, Hans (1992): Von der Philosophie des Pragmatismus zu einer soziologischen Forschungstradition. In: ibidem (Hg.): Pragmatismus und Gesellschaftstheorie. Frankfurt a. M.: Suhrkamp. 23-65
- Mead, George Herbert (1938). The Philosophy of the Act. Chicago: University of Chicago Press
- Mey, Günter/Mruck, Katja (2009): Methodologie und Methodik der Grounded Theory. In: Kempf, Wilhelm/Kiefer, Markus (Hg.): Forschungsmethoden der Psychologie. Bd. III. Berlin: Regener. 100-152
- Shalin, Dmitri N (1986): Pragmatism and Social Interactionism. In: American Sociological Review 51: 9-29
- Strauss, Anselm L. (1978): A Social World Perspective. In: Studies in Symbolic Interaction 1: 119-128
- Strauss, Anselm L. (1991). Grundlagen qualitativer Sozialforschung. München: Fink (amerik. Orig.: Qualitative Analysis for Social Scientists. New York: Cambridge University Press, 1987)
- Strauss, Anselm L./Corbin, Juliet (1990/1996): Grounded Theory. Weinheim: Beltz/ Psychologie Verlagsunion (amerik. Orig.: Basics of Qualitative Research, Newbury Park: Sage 1990)
- Strübing, Jörg (2007a): Anselm Strauss. Konstanz: UVK
- Strübing, Jörg (2007b): Pragmatistisch-interaktionistische Wissenssoziologie. In: Schützeichel, Rainer (Hg.): Handbuch Wissenssoziologie. Konstanz: UVK. 127-138
- Strübing, Jörg (2011): Zwei Varianten von Grounded Theory. In: Mruck, Katja/Mey, Günter (Hg.): Grounded Theory Reader, Wiesbaden: VS Verlag. 261-277
- Strübing, Jörg (2013): Qualitative Sozialforschung. München: Oldenbourg
- Strübing, Jörg (2014): Grounded Theory. Zur sozialtheoretischen und epistemologischen Fundierung eines pragmatistischen Forschungsstils. Wiesbaden: VS Verlag

Ronald Kurt und Regine Herbrik

---

## 33.1 Was ist sozialwissenschaftliche Hermeneutik?

*Sozialwissenschaftliche Hermeneutik* (heute auch häufig *Hermeneutische Wissenssoziologie* genannt) bezeichnet eine von Hans-Georg Soeffner und seiner Arbeitsgruppe in den 1980er Jahren entwickelte Methodologie und ein Set methodischer Verfahrensweisen zur Interpretation textförmiger Daten (Soeffner 1989). Mittlerweile wurden ihre Grundsätze auch auf die Interpretation von Bild- und Videodaten übertragen (Raab 2008, Kurt 2002, 2008, Herbrik 2011, Tuma/Schnettler, Kapitel 68 in diesem Band). Ziel des Verfahrens ist es, die gesellschaftliche Bedeutung von sozialen Handlungen, Interaktionen und Interaktionsprodukten in Form von Typenbildungen rekonstruktiv zu verstehen (Kurt 2004, Reichertz 2013). Als angewandte verstehende Soziologie ist die sozialwissenschaftliche Hermeneutik ein Mittel zur *Ex-Post-Analyse empirischer Einzelfälle*.

Für hermeneutische Verfahrensweisen sind gleich zu Beginn drei *Besonderheiten* zu erwähnen:

1. Die Hermeneutik ist kein starres Verfahren, das immer wieder auf genau die gleiche Art und Weise angewandt werden könnte. Sie ist im Gegenteil *flexibel* und kann bzw. muss dem jeweiligen Untersuchungsgegenstand immer wieder neu angepasst werden.
2. Hermeneutisches Arbeiten beruht auf einer *Haltung, die praktisch eingeübt werden muss*. Die Lektüre eines Artikels wie diesem, der die sozialwissenschaftliche Hermeneutik erläutert, ist daher lediglich der erste Schritt auf dem Weg zum Erlernen und Anwenden der Methodik. Sie kann das praktische Ausprobieren der hermeneutischen Techniken am Material nicht ersetzen.
3. Kennzeichnend für die hermeneutische Praxis ist die Bereitschaft, sich immer wieder von den Daten und den eigenen Deutungen irritieren zu lassen. Mit dieser Haltung ist es möglich, *Neues zu entdecken, Felder zu explorieren, Perspektiven aufzuschließen, ein reicheres Verstehen zu ermöglichen und wissenschaftliche Kreativität zu fördern*.

### 33.2 Was ist Verstehen?

Menschen sind soziale Lebewesen. Sie orientieren sich in ihrem Denken, Fühlen, Wollen und Handeln an anderen Menschen. Es gehört zu den Bedingungen dieses Aufeinanderbezogenseins, dass kein Mensch Zugang zum *Bewusstsein* eines anderen Menschen hat. Um inneren Sinn zum sinnlichen Ausdruck zu bringen, ist es Menschen aber möglich, *Zeichen* (wie z.B. Worte, Gesten, Texte oder Bilder) zu verwenden, mit denen sie sich wechselseitig etwas zu verstehen geben. Das *Zeichensetzen* und *Zeichenverstehen* bildet die *Basiskompetenz sozialer Orientierung*. Menschen müssen sich ausdrücken und sie müssen verstehen (Kurt 2009a: 10ff.), um die für sie lebensnotwendigen sozialen Beziehungen herstellen und aufrechterhalten zu können.

*Verstehen* bedeutet, menschlichen Lebensäußerungen Sinn zu geben. Mit Wilhelm Diltheys Worten (1957: 318): „Wir nennen den Vorgang, in welchem wir aus Zeichen, die von außen sinnlich gegeben sind, ein Inneres erkennen: Verstehen.“

In diesem Sinne handelt es sich genau dann um Verstehen, wenn Menschen meinen, dass sie es mit den Zeichensetzungen anderer Menschen zu tun haben. Fehlt diese für das Verstehen notwendige Sinn- und Subjektivitätsunterstellung, dann liegt auch kein Verstehen vor.

#### 33.2.1 Grundmerkmale des Verstehens

Bereits die grundlegendsten Formen des Verstehens weisen gemeinsame Charakteristika auf, die im Weiteren auch für das wissenschaftliche Verstehen relevant sind und im Folgenden kurz vorgestellt werden.

- *Universalität des Verstehens*: Um sich in der Welt orientieren zu können, sind Menschen ständig darauf angewiesen zu verstehen, indem sie sinnlich Gegebenes (Laute, Gerüche, Gegenstände, menschliches Verhalten etc.) als Ausdruck von subjektivem Sinn auffassen.
- *Perspektivität des Verstehens*: Das Verstehen ist immer an den subjektiven Standpunkt des Verstehenden gebunden; es perspektiviert sich in seinem Hier, Jetzt und So.
- *Soziohistorisches Apriori des Verstehens*: Die Seinsgebundenheit des Verstehens ist ein Produkt der Prägung durch Geschichte, Gesellschaft und Lebenslauf. Hineingeboren in eine von anderen bereits gedeutete Welt lernt der Mensch im Verlauf seiner Sozialisation die soziale Wirklichkeit auf eine bestimmte Weise zu verstehen.
- *Geschichtlichkeit des Verstehens*: Verstehen bezieht sich auf Ereignisse in der Geschichte und ist selbst ein Ereignis in der Geschichte – weshalb derselbe Mensch das gleiche Ereignis zu einem späteren Zeitpunkt möglicherweise ganz anders versteht.
- *Vorstruktur des Verstehens*: Jedes Verstehen beruht auf Vorverständnissen. Das Verstehen beginnt nicht bei Null, weil sich jedes Verstehen im Sinnhorizont des bereits Verstandenen vollzieht: als Einordnung des zu Verstehenden in pragmatisch bewährte

Wissensbestände. So steht beispielsweise für den Mann mit dem weißen Kittel und dem Stethoskop um den Hals die 'Schublade' Arzt parat.

- *Selektivität des Verstehens:* Menschen können nicht alles auf einmal verstehen. Das Verstehen bedarf deshalb der Ausrichtung und Fokussierung. Verstehensgegenstände konstituieren sich in (bewussten oder auch nichtbewussten) Auswahlprozessen.
- *Motiviertheit des Verstehens:* Menschen sind pragmatisch und verstehen deshalb interessensorientiert. So hören wir aus den Aussagen anderer gerne das uns bzw. unsere Meinung Bestätigende heraus.
- *Unabschließbarkeit des Verstehens:* Aufgrund seiner Perspektivität und Geschichtlichkeit kann kein Verstehen letztgültig sein. Andere Zeiten, andere Interpretationen ... (vgl. Kurt 2009b: 10).

### 33.2.2 Grenzen des Verstehens

Aus den Grundmerkmalen lassen sich gleichzeitig auch die folgenden Grenzen des Verstehens ableiten:

1. *Man kann nie ganz verstehen.* Gegen hundertprozentiges Verstehen sprechen nicht nur die Perspektivität und Geschichtlichkeit des Verstehens, sondern auch die Unzugänglichkeit und Unzulänglichkeit des Bewusstseins: Es ist für sich selbst nicht transparent, und es kann inneren Sinn nicht 1:1 in Ausdruck übersetzen, geschweige denn vollständig versprachlichen. „Spricht die Seele so spricht ach! schon die Seele nicht mehr“ (Schiller 1983: 297).
2. *Voraussetzungsfreies Verstehen ist nicht möglich.* Jedes Verstehen beruht auf Vorverständnissen, auf einem Immer-schon-verstanden-Haben. Die Vorstruktur des Verstehens (Heidegger 1993: 151) hat eine Doppelfunktion: Sie ermöglicht Verstehen und präfiguriert es zugleich.
3. *Das Verhältnis zwischen Zeichen und Bezeichnetem ist nicht eindeutig.* Die Mehrdeutigkeit des Zeichensinns macht das Verstehen zu einer unendlichen Aufgabe, die sich immer wieder neu und anders stellt (Schleiermacher 1959: 31).

Aus der Unzugänglichkeit des Bewusstseins, der Verwurzelung des Verstehens in Vorverständnissen und der Uneindeutigkeit des Zeichensinns folgt, dass die Voraussetzungen für angemessenes Verstehen heikel sind. Die Hermeneutik ist eine Antwort auf die Frage, wie die Operation „Verstehen“ gleichwohl gelingen kann – als methodisch kontrollierte wissenschaftliche Tätigkeit.

### 33.3 Varianten hermeneutischen Denkens

#### 33.3.1 Hermeneutik

Ursprünglich eine lose Sammlung von Verfahren, mit denen sich (z.B. in der antiken Bibellexegese) unverständliche Texte in verständliche verwandeln ließen, wurde das hermeneutische Denken von Friedrich Daniel Ernst Schleiermacher (1768-1834) Anfang des 19. Jahrhunderts in die Form einer „*Kunstlehre des Verstehens*“ gebracht (Schleiermacher 1995). In der „Kunst, die Rede eines anderen richtig zu verstehen“ (Schleiermacher 1995: 75) geht es um zwei Auslegungsaufgaben: „die Rede zu verstehen als herausgenommen aus der Sprache, und sie zu verstehen als Tatsache im Denkenden“ (Schleiermacher 1995: 77). Das Ziel des sich am Allgemeinen wie am Individuellen ausrichtenden Interpretierens besteht darin, in der „Umkehrung eines Aktes des Redens“ zu rekonstruieren, „welches Denken der Rede zum Grunde gelegen“ hat (Schleiermacher 1995: 76). Indem er „sich gleichsam in den anderen verwandelt“ (Schleiermacher 1995: 169), nähert sich der Verstehende dem zu Verstehenden in einer Rückwärtsbewegung an. Wie kam er dazu, sich sprachlich so und nicht anders auszudrücken?

In der Suche nach Antworten vollzieht das hermeneutische Denken Bewegungen, die vom Einzelnen zum Ganzen und von diesem wieder zu jenem führen. Diese heute als *hermeneutischer Zirkel* bekannte Denkfigur beruht auf der Annahme, „daß alles Einzelne nur verstanden werden kann vermittelst des Ganzen und also jedes Erklären des Einzelnen schon das Verstehen des Ganzen voraussetzt“ (Schleiermacher 1995: 328). Als paradoxe Verstehensforderung formuliert: „Man muß den Menschen schon kennen, um die Rede zu verstehen, und doch soll man ihn erst aus der Rede kennen lernen“ (Schleiermacher 1959: 44).

Wilhelm Dilthey (1833-1911) emanzipierte das Verstehen als die Methode der Geisteswissenschaften vor dem naturwissenschaftlichen Erklären (Eifler und Kühnel/Dingelstedt, Kapitel 11 uns 80 in diesem Band) – „Die Natur erklären wir, das Seelenleben verstehen wir“ (Dilthey 1957: 144) – und erweiterte den Anwendungsbereich der Hermeneutik, indem er nicht nur sprachliche, sondern alle menschlichen Ausdrucksformen zu möglichen Gegenständen des wissenschaftlichen Verstehens erklärte; unter der Voraussetzung, dass sie in der Form von Dokumenten, (Salheiser, Kapitel 62 in diesem Band), Zeitungsartikeln (Klein, Kapitel 64 in diesem Band), literarischen Texten (Ernst, Kapitel 63 in diesem Band), Bildern (Bohsack, Kapitel 67 in diesem Band), Videodaten (Tuma/Schnettler, Kapitel 68 in diesem Band), Filmen (Akremi, Kapitel 69 in diesem Band), Kleidern, Gärten und anderen Artefakten jedweder Art (Schubert, Kapitel 70 in diesem Band) als objektive Daten methodisch kontrolliert und für andere nachvollziehbar interpretiert werden können. Der Fluchtpunkt der Interpretation von „dauernd fixierten Lebensäußerungen“ (Dilthey 1957: 319) besteht für Dilthey darin, die individuellen Sinnzusammenhänge des Seelenlebens im Kontext von Geschichte und Gemeinschaft zu verstehen. Dabei stellt sich Dilthey (1958: 147) Deutende wie Gedeutete als Mitglieder von Wir-Sphären vor: „der einzelne erlebt, denkt und handelt stets in einer Sphäre der Gemeinsamkeit, und nur in

einer solchen versteht er. ... Wir leben in dieser Atmosphäre, sie umgibt uns ständig. Wir sind eingetaucht in sie“.

Die *Hauptaufgabe der Hermeneutik* sahen Schleiermacher und Dilthey in der *Sinnrekonstruktion*, also in der *hypothetischen Nachbildung des Sinnzusammenhangs*, der verstehbar macht, warum sich ein Mensch so ausgedrückt hat, wie er sich ausgedrückt hat.

### 33.3.2 Sozialwissenschaftliche Hermeneutik

Die von Hans-Georg Soeffner entwickelte *sozialwissenschaftliche Hermeneutik* knüpft an dieses Verstehensverständnis an. Sie fragt nach den (allgemeinen) Problemen, die Menschen in ihrem Handeln (auf besondere Art und Weise) zu lösen versuchen, sie fragt nach den Deutungsroutinen, Handlungsmustern, Ritualen und Symbolen, mit denen Menschen im sozialen Aufeinanderbezogensein ihrem Alltagsleben Sinn und Struktur verleihen, sie fragt nach den biographischen, soziohistorischen und kulturellen Hintergründen individueller Lebenswelten, und sie fragt nach Freiheit: nach Möglichkeiten des Andersdenken-, Andersdeuten- und Andershandelnkönnens.

Im Vergleich zu der auf die Entdeckung *latenter Strukturen* ausgerichteten *objektiven Hermeneutik* Ulrich Oevermanns (Oevermann et al. 1979) geht es in der *sozialwissenschaftlichen Hermeneutik* als Methode einer (im Anschluss an Weber, Schütz und Berger/Luckmann) wissenssoziologisch forschenden verstehenden Soziologie primär um die *Rekonstruktion individueller und gesellschaftlicher Sinnkonstruktionen*. Die sinnrekonstruierende Zeicheninterpretation zielt dabei nicht auf das vollständige Verstehen von Individuen, sondern auf ein typenbildendes Verstehen sozialer Sinnzusammenhänge. Die sozialwissenschaftliche Hermeneutik knüpft hier an Max Webers Konzept des methodologischen Individualismus an: Der Weg zum Sozialen führt in der Verstehenden Soziologie über das Individuum. Auf diesem Weg bewegt sich die sozialwissenschaftliche Hermeneutik im Zwischen zweier *Prämissen*:

1. *dass Verstehen durch die Gleichheit bzw. Ähnlichkeit menschlichen Denkens und Deutens gewährleistet ist.* In diesem Sinn konzipiert Dilthey das Verstehen als „ein Wiederfinden des Ich im Du“ (Dilthey 1958: 191). Hier garantieren die Wesensgleichheit und die soziale Seelenverwandtschaft, dass Verstehen zwischen ego und alter möglich ist.
2. *dass der Mensch dem Menschen ein Fremder ist.* Durch radikale Differenz voneinander getrennt, sind sich die Menschen am nächsten im Nichtverstehen; in dem Paradox, dass die Präsenz des Anderen in seiner Absenz erfahrbar ist. „Die Abwesenheit des anderen ist gerade seine Anwesenheit als des anderen“ (Lévinas 1984: 65).

In der sozialwissenschaftlichen Hermeneutik können diese gegensätzlichen Grundannahmen gleichberechtigte Orientierungsalternativen sein. In die Form einer Forschungshaltung gebracht: Betrachte das zu Verstehende *nie nur* unter den Vorzeichen von Identität/Gleichheit/Ähnlichkeit, sondern *immer auch* unter den Vorzeichen von Differenz.

### 33.4 Alltagsweltliches Verstehen – Wissenschaftliches Verstehen

Für gewöhnlich verstehen wir, ohne das uns bewusst wäre, dass und wie wir verstehen. Die Als-Struktur des Verstehens (Heidegger 1993: 151), etwas – zum Beispiel dieses Schwarz auf Weiß – als etwas – einen Text – aufzufassen, bleibt im Alltag unterschwellig. Solange das Verstehen wie ein Automatismus abläuft, vollzieht es sich weitgehend nichtbewusst. Erlernte Deutungsroutinen tragen dafür Sorge, dass sich fast immer fast alles wie von selbst versteht. Man erkennt Bekanntes, versteht wie üblich und tut, was man tut, so wie man es schon immer tat – und arbeitet pragmatisch die Forderungen des Tages ab. Das ist die (von Alfred Schütz so genannte) *natürliche, alltagsweltliche Einstellung*, in der uns das Verstehen aufgrund seiner Selbstverständlichkeit verborgen bleibt. In der *Alltagspraxis* geschieht das Verstehen also weitgehend reflexartig: als blitzschnelles Einordnen gegebener Zeichen in bereit stehende Wissensbestände.

In der *wissenschaftlichen Praxis* darf das Verstehen aber kein Reflex sein. Es muss eine Reflexion sein: ein bewusster, sich selbst beobachtender und steuernder, den Sinn gegebener Zeichen aus der Perspektive des Zeichen Setzenden rekonstruierender Interpretationsprozess. Die sozialwissenschaftliche Hermeneutik ist somit nicht ausschließlich Kunst, wie ihr oft unterstellt wird, denn ihre Durchführung folgt Regeln, die angegeben, gelernt und eingebütt werden können. Sie ist jedoch genauso wenig ausschließlich eine Technik, die bei ordnungsgemäßer Anwendung aus jedem Datum quasi automatisch eine ‚richtige‘ Interpretation presst. Die sozialwissenschaftliche Hermeneutik ähnelt vielmehr einem *Spiel*, denn innerhalb ihres Regelwerks eröffnen sich Freiräume für Ungeregeltes und nicht Vorhersehbare: für Improvisation. So wie zum Spiel (z.B. bei Caillois 1982) sowohl seine Regelhaftigkeit als auch seine Wildheit und Ungeordnetheit gehören, und so wie die Sprache selbst im Spannungsverhältnis von relativ geregelter Grammatik und relativ freier situativer Aktualisierung steht, so zeichnet sich das hermeneutische Interpretieren durch ein stetes Hin-und-Her-Wechseln zwischen *Wirklichkeitssinn* und *Möglichkeitssinn* (Musil 1997) aus (siehe auch unten). Ein Grundzug dieses Wechselspiels besteht darin, am Datenmaterial generierte Möglichkeitsannahmen an der Wirklichkeit zu überprüfen (bzw. Lesarten am Text zu testen – wie in der in 6. und 7. beschriebenen Sequenzanalyse).

#### 33.4.1 Hermeneutische Haltung

Im Wechsel vom Modus „Alltag“ in den Modus „Wissenschaft“ wird das alltägliche Verstehen, die Konstruktionen ersten Grades, zum Gegenstand des wissenschaftlichen Verstehens, den Konstruktionen zweiten Grades (Schütz 2010: 334). In diesem *Verstehen des Verstehens* wird das Verstehen *selbstreflexiv*.

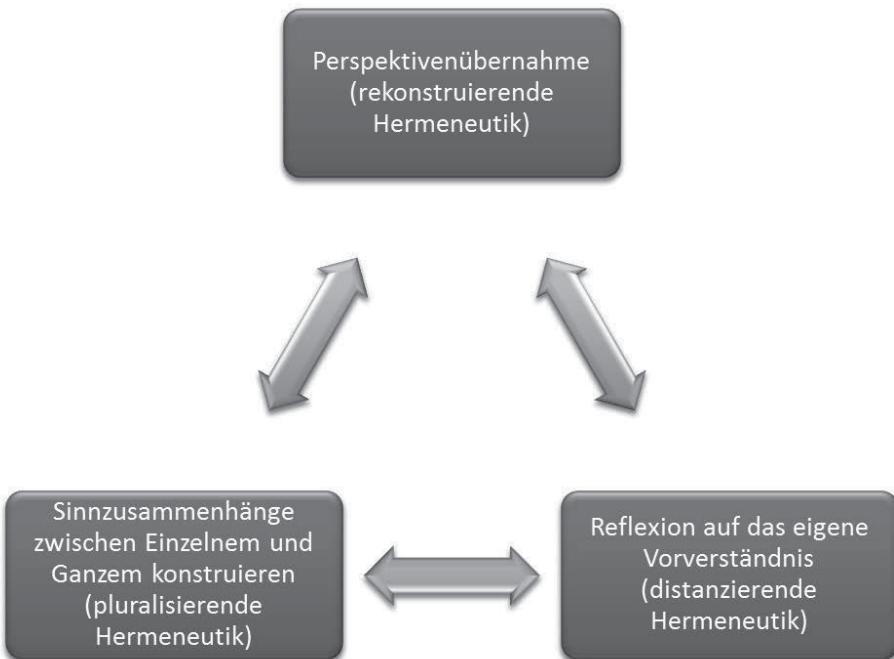
Zu diesem Zweck gilt es, eine hermeneutische Haltung einzunehmen. Der Wechsel in eine solche Haltung setzt voraus, dass die Mechanismen des Alltagsverständens systematisch außer Kraft gesetzt werden; durch ein umfassendes Negationsprogramm (Kurt 2009b: 82ff., Soeffner 1989: 10-50), das die folgenden Punkte umfasst:

- *Nicht schnell verstehen, sondern in aller Ruhe:* Die Entlastung vom Handlungs- und Zeitdruck des Alltagsleben ist eine notwendige Bedingung für wissenschaftliches Verstehen.
- *Nicht pragmatisch verstehen, sondern distanziert und interesselos:* Im Gegensatz zum pragmatischen alltagsweltlichen Verstehen basiert wissenschaftliches Verstehen darauf, zu den eigenen Gedanken und Gefühlen ein unpersönliches Verhältnis zu haben.
- *Nichts als selbstverständlich hinnehmen, sondern alles in Frage stellen:* Der systematische Zweifel ist das A und O der Wissenschaft. Während sich im Alltag alles wie von selbst versteht, darf sich im wissenschaftlichen Verstehen nichts von selbst verstehen. Hermeneutik ist methodisierte Skepsis.
- *Nicht wie üblich verstehen, sondern anders verstehen:* Im alltagsweltlichen Verstehen wird Unbekanntes auf Bekanntes zurückgeführt; d.h. alles Fremde und Neue wird pragmatisch bewährten Schemata subsumiert. Wissenschaftliches Verstehen funktioniert umgekehrt: Hier wird das scheinbar Selbstverständliche mittels Befremdung in Nicht-Selbstverständliches transformiert, um aus diesem methodisch erzwungenen Nicht-Verstehen neue Verstehensmöglichkeiten zu entwickeln.
- *Nicht nach Eindeutigkeit streben, sondern nach Mehrdeutigkeit:* Das alltägliche Verstehen drängt nach Eindeutigkeit, das wissenschaftliche nach Mehrdeutigkeit. Erst die offene Frage, ob etwas so oder so oder so zu verstehen ist, lässt neue Wege des Verstehens entstehen.

### 33.4.2 Richtungen der Interpretation

Aus dieser hermeneutischen Haltung heraus kann die Sinn rekonstruierende Zeicheninterpretation in drei Richtungen gehen, wobei es in der konkreten Interpretation darauf ankommt, diese Verstehensbewegungen so miteinander zu verknüpfen, dass sie sich gegenseitig widersprechen und konstruktiv ergänzen können (vgl. auch Abb. 33.1). Die drei Richtungen hermeneutischen Interpretierens wenden sich:

- *zum Anderen (Perspektivenübernahme, rekonstruierende Hermeneutik):* Bei dem Versuch, sich hypothetisch in andere hineinzuversetzen, können die folgenden Fragen hilfreich sein:
  - *Sinnlich:* Was nimmt der Andere von seinem Standpunkt aus wahr?
  - *Subjektiv:* Auf welche Motive und Relevanzen sind die Lebensäußerungen anderer zurückzuführen?
  - *Sozial:* Welche Sozialisation hat der Andere erfahren? Wie definiert er die Situation und welche Rolle spielt er in ihr? Was erwartet er von sich und anderen?
  - *Symbolisch:* In welche Symbolsysteme und Wertüberzeugungen ist das Denken und Handeln anderer eingebettet?
- *zum Eigenen (Reflexion auf das eigene Vorverständnis, distanzierende Hermeneutik):* Bei der Selbstbesinnung auf das eigene Vorverständnis ist immer zu fragen:



**Abb. 33.1** Die drei Richtungen hermeneutischen Interpretierens

- Legen wir aus, oder legen wir hinein? Das heißt: Kann das eigene Vorverständnis vom Anderen unterschieden werden oder wird das Eigene unreflektiert in Andere hineinprojiziert?
- zu den Zeichen (*Sinnzusammenhänge zwischen Einzelnen und Ganzem konstruieren, pluralisierende Hermeneutik*) (Marquard 1991): Bei der (Re-)Konstruktion von Sinnzusammenhängen ist zu fragen, in welchem Kontext welchen Zeichen welcher Sinn zukommt. Wie konstituiert sich das Ganze im Verhältnis der Teile? Wie lautet die Frage, auf die das Auszulegende die Antwort ist?

### 33.5 Verwandtschaft zu anderen Verfahren

Besonders ertragreich lässt sich sozialwissenschaftliche Hermeneutik innerhalb eines Forschungsdesigns (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) realisieren, das sich auf ein mit ihren Grundlagen kompatibles Verständnis von Sinn, sozialem Handeln und Lebenswelt gründet. Sie kommt daher insbesondere dort zur Anwendung, wo mithilfe *ethnografischer Zugänge* (Knoblauch, Kapitel 36 in diesem Band) soziale Felder, „kleine soziale Lebenswelten“ oder soziale Milieus erschlossen werden. Um die Perspektive nicht bereits zu Beginn des Forschungsprozesses zu verengen, ist die Fragestellung möglichst

offen zu wählen. Statt anhand bestehender Theorien Hypothesen zu entwickeln und diese dann am Material zu testen, bietet es sich an, sich bei der Datenerhebung von den im Feld vorfindlichen Gegebenheiten leiten zu lassen. Das bedeutet jedoch nicht, dass auf Vorwissen nicht reflektiert werden sollte. Auch Forschungsfragen können dabei hilfreich sein, einen ersten Zugang zum Feld zu erarbeiten.

Aufgrund ihrer methodologischen Verwandtschaft mit der „*Grounded Theory*“ (Strübing, Kapitel 32 in diesem Band) lässt sich die sozialwissenschaftliche Hermeneutik gut mit diesem Forschungsstil kombinieren. Das komparative Vorgehen hat für beide Verfahren eine ähnliche Relevanz. Während im Falle der Grounded Theory der Vergleich zwischen unterschiedlichem Vorfindlichen im Vordergrund steht, ist für die sozialwissenschaftliche Hermeneutik darüber hinaus der Vergleich zwischen dem Vorfindlichen und dem Möglichen relevant. Dies bedeutet auch, dass unterschiedliche Möglichkeitsentwürfe miteinander verglichen werden können. Die Hermeneutik nutzt dabei die menschliche Fähigkeit, über das unmittelbar Wahrnehmbare hinauszugehen, also zu imaginieren, und das Offensichtliche zu überschreiten. Dadurch eröffnet sie auch dort die Möglichkeit zu vergleichen, wo dies ohne das Transzendieren des tatsächlich Gegebenen nicht möglich wäre.

---

### 33.6 Sequenzanalyse

Ein zentrales Verfahren des Verstehens stellt die Sequenzanalyse dar. Die *Sequenzanalyse* ist ein *Interpretationsverfahren*, das den Sinn jeder Art menschlichen Handelns Sequenz für Sequenz, also Sinneinheit für Sinneinheit, in der Linie des ursprünglichen Geschehens zu rekonstruieren versucht. Geleitet wird die Sequenzanalyse von der Annahme, dass sich im Nacheinander des Handelns – also auch des Sprechens, Schreibens und Filmens – Sinnzusammenhänge realisieren. Der hermeneutische Ansatz, das Einzelne als Teil eines Ganzen zu denken, überspannt in der Sequenzanalyse das konkrete menschliche Handeln hypothetisch mit einer alle Handlungsschritte umfassenden Sinngestalt. Ohne diese Sinninvestition – als „Vorgriff der Vollkommenheit“ (Gadamer 1990: 299) – wären Sequenzanalysen nicht durchführbar.

Die Sequenzanalyse ist nicht dazu gedacht, große Datenmengen in möglichst kurzer Zeit zu sichten und zu organisieren. Vielmehr ist sie ein zeitaufwendiges und personalintensives Verfahren, dessen Anwendung sich insbesondere dort lohnt, wo ein erster (oder neuer) Zugang zu einer Fragestellung erarbeitet werden soll, wo bedeutungsmäßig besonders „dichte“ Schlüsselstellen intensiv ausgedeutet werden sollen und wo es herauszufinden gilt, welche Fragestellung für die Erforschung eines bestimmten Feldes überhaupt angemessen sein könnte.

Die grundlegenden Verfahrensschritte (Kurt 2004: 240ff.) der Sequenzanalyse, die weiter unten noch detailliert erklärt werden, sind in Tab. 33.1 dargestellt.

**Tab. 33.1** Verfahrensschritte der Sequenzanalyse

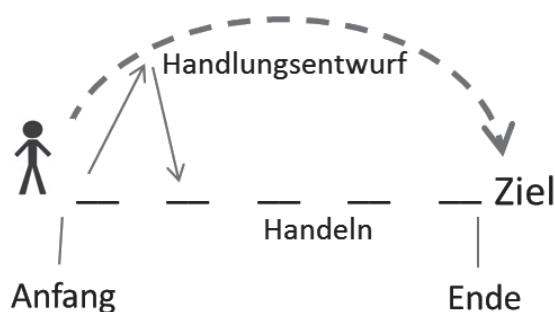
Verfahrensschritte	Beispiel
1. <i>Sequenzbestimmung</i> : Eine Sequenz isolieren und das Wissen über den Kontext ausblenden.	Erste Sequenz: „Es war einmal“
2. <i>Hypothesenentwicklung</i> : Lesarten entwickeln. Lesarten sind Handlungszusammenhänge, in denen die in Frage stehende Sequenz als Teil des Ganzen Sinn ergibt (Reichertz 1986: 250).	<p><i>Erste Lesart</i>: Die Sequenz könnte der Anfang eines Märchens sein. Wenn es sich bei dem Text um ein Märchen handelt, könnte der Text so weitergehen: „ein Müller“ oder: „eine Prinzessin“ etc.</p> <p><i>Zweite Lesart</i>: Die Sequenz könnte der Beginn eines Kinderliedes sein. Ein möglicher Anschluss wäre dann: „... ein Apfel“</p>
<p>3. <i>Hypothesenprüfung</i>: Überprüfung der fantasiierten Lesarten</p> <p>(a) anhand der Folgesequenzen (<i>innerer Kontext</i>) und/oder</p> <p>(b) anhand des Handlungskontextes (<i>äußerer Kontext</i>).</p>	Die nächste Sequenz lautet: „ein Politiker“. Diese Erwartungsenttäuschung im Hinblick auf die erste und zweite Lesart führt zur Formulierung neuer Hypothesen, die dann wieder empirisch am (Kon-)Text zu prüfen sind.
4. <i>Verallgemeinerung</i> : Typisierung des Einzelfalls als besondere Lösung eines allgemeinen Problems.	In diesem fiktiven Beispiel ist der märchenhafte Beginn eine einzelfallspezifische Lösung des Problems, einen Politiker als Lügner darzustellen.

### 33.7 Arbeitsschritte und Vorgehensweisen der Sequenzanalyse

Für die Durchführung einer Sequenzanalyse ist es hilfreich, sich an einem Ablaufschema zu orientieren. Insbesondere während des Einübens der Methode kann es der *Interpretationsgruppe* ein hilfreicher Leitfaden sein.

Abbildung 33.2 soll die *handlungstheoretischen Hintergrundannahmen* und die *Verfahrensschritte* der Sequenzanalyse veranschaulichen. Sie stellt das Handeln als schrittweise Realisierung eines Handlungsentwurfs dar (Schütz 2004). Der Pfeil, der von der ersten Sequenz des Handelns auf das Ganze des Handlungsentwurfs zielt, markiert die hermeneutische Denkbewegung vom Einzelnen zum Ganzen. Ausgehend von einer einzelnen Sequenz („Es war einmal“), wird fantasiiert, um welches Ganze (Lesart 1: Märchen, Lesart 2: Kinderlied) es sich handeln könnte. Der vom Entwurf zurück in Richtung Handeln verlaufende Pfeil stellt dar, wie im Rückgang vom Ganzen zum Einzelnen die nächste Sequenz antizipiert wird („eine Prinzessin“ bzw. „ein Apfel“). In diesem Verfahrensschritt geht es also darum, zu fragen, was unter der Prämisse der entworfenen Lesart sinnvoller Weise auf die erste Sequenz folgen könnte.

**Abb. 33.2** Der Handlungsentwurf und seine schrittweise Realisierung im Handeln



### 33.7.1 Alles auf Anfang

Weil im Anfang schon das Ende mitgedacht ist, der Handlungsentwurf als vorweg gedachtes Ganzes also den Beginn (und auch den weiteren Handlungsprozess) bestimmt (Mead 1973: 50), fängt die Sequenzanalyse vorzugsweise am Anfang eines Handlungsablaufs an. Das bedeutet jedoch nicht, dass grundsätzlich am Beginn eines Textes bzw. Datums (Akremi, Kapitel 17 in diesem Band) angefangen werden muss.

### 33.7.2 Isolierung der Anfangssequenz

Um das Bedeutungspotential der Anfangssequenz zu ergründen, wird die erste Sequenz isoliert betrachtet. Wo Sequenzen beginnen und enden, bestimmen die Interpretierenden. Je nach Datenmaterial kann eine Sequenz beispielsweise ein Wort, ein Halbsatz oder ein ganzer Satz sein. Häufig ist es sinnvoll, zu Beginn einer Interpretation sehr kurze Sequenzen auszuwählen und erst im Fortgang des hermeneutischen Prozesses etwas längere Sequenzen zu definieren. Eine Sequenz zu isolieren, bedeutet praktisch, alle folgenden Sequenzen abzudecken (zum Beispiel mit Papier) oder sie gedanklich nicht zu berücksichtigen.

### 33.7.3 Ausblendung des Kontextwissens

Der gesamte bekannte Kontext des Handlungszusammenhangs wird ausgeblendet. Man tut so, als ob man nicht wüsste, in welchen Kontext die Sequenz gehört und in welchem Zusammenhang das Handeln geschah. Man stellt sich unwissend (und ist es unter Umständen auch). Dieser Schritt erfordert von den Interpretierenden eine befremdende Denkweise: sich auf etwas (möglicherweise sogar gut) Bekanntes so einzulassen, als ob es vollkommen neu und unbekannt sei.

### 33.7.4 Lesartenbildung

Die vorübergehend aus ihrem Kontext herausgelöste Sequenz wird interpretiert, um eine Vielzahl von wahrscheinlichen und unwahrscheinlichen Lesarten zu entwickeln. Hier heißt die Leitfrage: In welchem Kontext könnte diese Sequenz Sinn ergeben? Je widersprüchlicher und unwahrscheinlicher die (in sich stimmigen) Lesarten und entworfenen Kontexte ausfallen, desto besser für den Interpretationsprozess.

Der Stoff für die Lesartenbildung entstammt dem Wissensrepertoire der Interpretationsgruppe. Eine wissensspezifisch heterogene Gruppenzusammensetzung ist einer homogenen deshalb vorzuziehen. Bei der Lesartenbildung darf auf keinen Fall auf Folgesequenzen vorgegriffen werden! Grundsätzlich gilt: Von Früherem auf Späteres schließen – und niemals umgekehrt, denn andernfalls bräche die Logik des sequenziellen Vorgehens in sich zusammen.

Bei der Lesartenbildung kann es hilfreich sein, *Interpretationsschwerpunkte* zu setzen:

- *Grammatische Interpretation*: Wortbedeutung (etymologischer, buchstäblicher, metaphorischer Sinn), Satzbau, Textart etc. analysieren und Paraphrasierungen formulieren: Wie hätte das Gleiche alternativ ausgedrückt werden können?
- *Sachliche Interpretation*: Worum geht es?
- *Soziologische Interpretation*: Wer sagt das wann, wo, wie, zu wem, warum in welcher Situation? Was könnte vorangegangen sein? Was könnte nun folgen? Auf welches soziale Problem wird wie reagiert?
- *Psychologische Interpretation*: Was denkt, fühlt, will der Handelnde? Welche Weil-Motive und welche Umzu-Motive liegen seinem Verhalten zugrunde (Schütz 2004) und was hätte der Handelnde stattdessen wollen oder tun (bzw. lassen) können?
- *Kulturelle Interpretation*: Wie ist die Handlungssituation kulturell codiert? (Hier ist nach Werten, Ideen, Symbolen und Weltanschauungen zu fragen.)

Diese Liste kann einzelfallspezifisch reduziert oder auch erweitert werden, beispielsweise durch eine juristisch oder historisch ansetzende Interpretation.

### 33.7.5 Hypothesenbildung

Anhand der im vorherigen Arbeitsschritt konstruierten Lesarten und möglichen Kontexten wird nun in Form von Hypothesen antizipiert, was als nächstes geschehen könnte. Lesart für Lesart wird konkretisiert, wie es im Text weitergehen und wie die Handlung enden könnte.

Dieses Phantasieren von Handlungsabläufen stützt sich auf das Wissen der Mitglieder der Interpretationsgruppe und ihre Möglichkeiten zum Beispiel durch Experten oder anhand eigener Recherchen (z.B. mittels Internet) andere Wissensquellen in den Prozess der Deutung einzubringen.

### 33.7.6 Hypothesenüberprüfung

Die Hypothesen werden im nächsten Schritt an der Folgesequenz (und gegebenenfalls am tatsächlichen Kontext – z. B. einer Interviewsituation als Ursprungskontext eines Transkripts) überprüft. Lesarten, die mit der Folgesequenz unvereinbar sind, scheiden aus dem Interpretationsverfahren aus. (Es empfiehlt sich, die falsifizierten Lesarten nicht durchzustreichen, sondern einzuklammern, weil sie im weiteren Verlauf der Interpretation noch eine wichtige Rolle spielen können.) Die nicht falsifizierten Lesarten werden in die nächste Phase der Auslegung übernommen. Der Text dient damit als objektive Korrekturinstanz für die Lesarten.

Wenn das tatsächliche Handeln durch eine Veränderung der Situation oder des Handlungsplans bzw. -ziels nicht wie antizipiert, sondern anders verläuft, dann bleibt die Verfahrenslogik davon unberührt. In Fällen wie diesen gilt es aufzudecken, warum das Handeln an welcher Stelle mit welcher Motivation die Richtung wechselte –, es ist ein Grundsatz der Sequenzanalyse, dass jede bewusste oder auch nichtbewusste Zeichensetzung motiviert geschieht (Oevermann et al. 1979: 394).

### 33.7.7 Möglichkeitssinn

Im Zuge dieses Arbeitsschrittes wird der Text nicht als Wirklichkeit, sondern als Möglichkeit betrachtet. In diesem Sinne fragen die Interpretierenden, was der Handelnde stattdessen hätte tun können. Welche Handlungsweisen wären aus Sicht des Subjekts objektiv möglich gewesen?

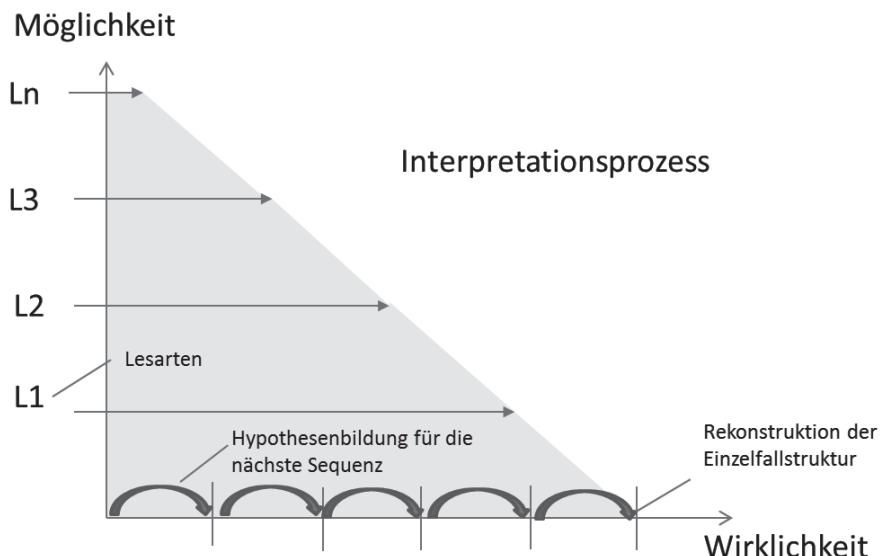
### 33.7.8 Wirklichkeitssinn

Das Gegenstück zu der Frage nach Alternativen bildet die Frage nach den objektiven Rahmenbedingungen des Handelns. Wie wurde das Handeln durch datums-externe Faktoren wie Lebenslauf und Lebenslage mitbestimmt? Auf diese Weise wird Kontextwissen methodisch kontrolliert in den Interpretationsprozess integriert. Wissenschaftliches Wissen, das zum besseren Verstehen des Textes beitragen kann, wie zum Beispiel Wissen aus den Geschichtswissenschaften und der Psychologie, sollte ebenfalls Berücksichtigung finden. (Fand das im Text Beschriebene tatsächlich statt? Und: Wie erklären Psychologen, dass Menschen meinen etwas erlebt zu haben, was sie gar nicht erlebt haben können?) Entscheidend für die Referenz auf Textexternes ist, dass sie Fragen folgt, die in der Analyse der Sequenzen entstanden sind. Um ein Verstehen-wie-Üblich zu verhindern, folgt die Bezugnahme auf Kontextwissen und wissenschaftliches Wissen nicht nur (zeitlich) *auf*, sondern auch (logisch) *aus* der Sequenzanalyse.

### 33.7.9 Reduktion der Lesarten

Die Schritte 7.4 bis 7.8 werden nun von *Sequenz zu Sequenz* wiederholt. Dabei reduziert sich die Anzahl der Lesarten durch die Überprüfung des Möglichen am Wirklichen. Einerseits; andererseits bringt der Prozess der Interpretation unweigerlich immer wieder neue Lesarten hervor – die dann ihrerseits wieder am Text zu testen sind. Ist schließlich eine Lesart und mit dieser eine Strukturhypothese gefunden, die dem Text standhält, dann wird das Verfahren der Sequenzanalyse umgestellt. Nun geht es nicht mehr um die (induktive und abduktive) Generierung neuer Deutungen, sondern um die (deduktive) Suche nach Textstellen, mit denen die favorisierte Deutung falsifiziert werden kann. Gelingt die Falsifizierung nicht, dann gilt die Interpretation im Sinne Poppers als bis auf weiteres gültig. Die Sequenzanalyse folgt hier der Popperschen Forderung, 1. falsifizierbare Hypothesen aufzustellen, 2. ihre empirische Widerlegung anzustreben und sie 3., sofern eine Falsifizierung nicht gelingt, als vorläufig gültig zu betrachten.

Abbildung 33.3 zeigt, wie im Prozess der Interpretation durch das kontinuierliche Kontrastieren von Möglichkeit und Wirklichkeit und der damit einhergehenden Reduktion der Lesarten die Typik des Einzelfalls rekonstruiert wird. Im Idealfall verläuft der Falsifikationsprozess so, dass sich die Anzahl der zu Beginn aufgestellten Lesarten von Sequenz zu Sequenz verringert, bis schließlich nur noch eine Lesart übrig bleibt. Der Praxis der Interpretation dient diese Verlaufsform als regulative Idee. Schließlich ist das Bilden und Prüfen von Lesarten kein Selbstzweck, sondern ein Mittel, um beim Querdenken zwischen



**Abb. 33.3** Zwischen Möglichkeit und Wirklichkeit

den Zeilen auf gute Ideen zu kommen. In diesem Sinn basiert das sequenzanalytische Verfahren auf einem produktiven Paradox: Das strenge methodische Vorgehen befähigt die Einbildungskraft und führt so zum Take-Off der Theorie-Fantasie.

### 33.7.10 Die Struktur des Einzelfalls

Im Verlauf der Sequenzanalyse wird das So-und-nicht-anders-Sein des interpretierten Handelns vor dem Hintergrund alternativer Deutungs- und Handlungsmöglichkeiten in seiner spezifischen Einzelfallstruktur erkennbar. Die Interpretation endet, wenn die handlungsleitenden Deutungs- und Verhaltensmuster typisiert sind und der Einzelfall als besondere Lösung eines allgemeinen Problems rekonstruiert werden kann.

Um die Nachvollziehbarkeit und Überprüfbarkeit der Sequenzanalyse zu gewährleisten, ist es notwendig, dass ein Gruppenmitglied (und/oder ein digitaler Datenträger) den gesamten Interpretationsprozess dokumentiert. Die Protokollierung kann wie folgt geschehen: Zunächst wird die zu interpretierende Sequenz auf das Protokollblatt geschrieben (S1). Darunter werden sinngemäß die Lesarten festgehalten (L1, L2, ..., Ln). Für jede Lesart sind dann Hypothesen für die Folgesequenz zu notieren (H1, H2, ..., Hn). Die falsifizierten Lesarten können dann bei der Interpretation der nächsten Sequenzen eingeklammert werden. Um blitzartig aufscheinenden Einfällen, Fragen, Fallbesonderheiten, theoretischen Assoziationen, Strukturhypothesen und Grafiken Raum zu geben, empfiehlt es sich, mindestens 1/3 des Protokollpapiers auf der rechten Seite frei zu lassen. Die Teilung des Protokollblatts systematisiert und dynamisiert den dialektischen Deutungsprozess: Links wird in Form der Lesarten und Hypothesen das Allgemeine des Falls und seine Konkretion dokumentiert, rechts werden die Fallbesonderheiten notiert. Im kontinuierlichen Hin-und-her-Denken zwischen den beiden Seiten kristallisiert sich das Allgemeine des Besonderen und das Besondere des Allgemeinen heraus.

---

## 33.8 Sequenzanalyse als Improvisation und Spiel

In der sozialwissenschaftlichen Hermeneutik nimmt die Sequenzanalyse aus den folgenden Gründen eine zentrale Rolle ein: Die Sequenzanalyse ist *objektiv*,

1. weil ihr (beispielsweise in Form eines Textes) objektiv Gegebenes zugrunde liegt;
2. weil sich mit ihrer Hilfe das Sinnpotential menschlicher Zeichensetzungen entfalten und im Modus des objektiv Möglichen (im Rahmen des Wissenshorizontes der Interpretierenden) rekonstruieren lässt;
3. weil ihr ein Testverfahren eigen ist, mit dem sich die im Prozess der Interpretation gebildeten Hypothesen objektiv (d.h. intersubjektiv nachvollziehbar) überprüfen lassen.

Die Sequenzanalyse ist nicht nur ein *hypothesengenerierendes*, sondern zugleich auch ein *hypothesentestendes Verfahren* – d.h. abduktive, deduktive und induktive Denkprozesse werden in der Sequenzanalyse methodisch ineinander verzahnt (Reichertz, Kapitel 2 in diesem Band).

Ein zentrales *Qualitätsmerkmal* (Flick, Kapitel 29 in diesem Band) der Sequenzanalyse besteht darin, dass sie am besten in der Gruppe funktioniert. Das Entscheidende ist, dass es hier nicht der Genius Einzelner ist, der das Neue gleichsam aus dem Nichts erzeugt. Die Antithese zum Mythos des genialen Einzelnen ist die Genialität improvisatorischer Teamarbeit. Es ist der *Genius der Gruppe*, der das Neue erschafft – aus einer kollektiven Improvisation heraus, welche die Ich-Grenzen der interpretierenden Individuen in einem gemeinsamen Flow-Erlebnis zum vollständigen Verschwinden bringen kann. Das Ergebnis einer Sequenzanalyse ist demnach nicht einem einzelnen Autor, sondern dem Interpretationskollektiv bzw. dem spiel-ähnlichen Prozess zuzuschreiben, in dem die Gruppenmitglieder dialogisch Deutungen generieren. Als Gruppeninterpretationsverfahren ist die Sequenzanalyse eine Methode, mit der sich nicht nur soziale Blockaden lösen lassen. Es ist auch eine Methode, um sich von der allzu schematischen Anwendung wissenschaftlicher Methoden abzubringen, ohne in die Problematik einer rein subjektiven und für andere nicht nachvollziehbaren Interpretation zu geraten. Die Erfahrung zeigt, dass die Praxis dieses Theorieideal, unter anderem aufgrund der hierarchischen Struktur akademischer Kommunikation, oft verfehlt. Reflexionen auf die eigene Kommunikationskultur gehören deshalb zur Verstehensarbeit der Interpretationsgruppe mit hinzu.

Die Sequenzanalyse fordert die Interpretationsgruppe zwangsläufig zum Improvisieren heraus und gibt ihr mithilfe einiger Spielregeln gleichzeitig den Rahmen vor, der die *Improvisation* begrenzt und zugleich ermöglicht (zur Improvisation als Methode der qualitativen Sozialforschung, Kurt 2012). Die Gruppenmitglieder versetzen sich mittels Perspektivenübernahmen gedankenexperimentell in die Situation handelnder Menschen und antizipieren anhand einer Sequenz aus dem Stegreif, zu welchem Ende ein gegebener Anfang führen könnte. Zu welcher Ausgangsaktion passen welche Anschlussaktionen? Aus der so erzeugten Vervollständigungsnot – man will ja wissen wie es weitergeht – können sich die Gruppenmitglieder nicht anders als durch das spontane Kreieren von Lesarten befreien. „Ein halbes Wort, und meine Seele eilt hinein in seinen Gedanken“ (Kierkegaard 1984:67).

Dieser gruppendynamische Prozess lässt dann quasi wie von selbst nicht vorhergesehene Lesarten entstehen. Die Perspektive des einen fordert einen zweiten zur Formulierung einer alternativen Lesart heraus, was wiederum einen dritten dazu anstachelt, seine Sicht der Dinge kundzutun. Auf nichts ist mehr Verlass als darauf, dass sich die Gruppenmitglieder gegenseitig aus der Denkspur hebeln (wozu sie ohne die Hilfe der Mitinterpretanten kaum in der Lage wären, weil jeder dazu neigt, sich viel zu schnell auf eine Lieblingslesart festzulegen). So bleibt die Offenheit der Improvisation gewährleistet, während gleichzeitig eine – für Gruppen-Spiele typische – Wettbewerbsdimension hinzugewonnen wird, die zur Motivation der Interpretierenden beiträgt.

Im spontanen Mit- und Gegeneinander-Denken können auch Zufälle eine wichtige Rolle spielen. Die Gruppe muss nur erkennen können, dass sie per Zufall etwas

wissenschaftlich Relevantes fand (Abduktion, vgl. auch Reichertz. Kapitel 2 in diesem Band), nach dem sie ursprünglich aber unter Umständen gar nicht auf der Suche war. Das ist in etwa das, was Merton mit dem Begriff „Serendipity“ umschreibt: „ein unerwartetes, von der Regel abweichendes, forschungsrelevantes Datum, das die Gelegenheit bietet, eine neue Theorie zu entwickeln oder eine bereits bestehende Theorie zu erweitern“ (Merton/Barber 2004: 293, Übersetzung d.V.).

In die Suche nach Auslegungsmöglichkeiten mischen sich zudem Reflexionen über die Angemessenheit dieser Auslegungen ein. Sozialen Sinn erhält eine Interpretationsaktion ja im Grunde erst durch die Reaktion auf sie. Und diese Reaktionen können nicht nicht bewertend sein. Das heißt, die Gruppe muss auf die Entäußerung von Deutungen zwangsläufig mit Bewertungen reagieren. Zum Beispiel mit einem zustimmenden Kopfnicken, das dem Interpreten signalisiert, seine Lesart weiter zu entwickeln. Die Zurückweisung einer Lesart als „zu weit her geholt“ oder „nicht überprüfbar“ kann den Interpreten wiederum dazu herausfordern, seine Lesart an dieser Kritik zu schärfen. Irrige Annahmen, die ertragreich scheitern, können für den Interpretationsprozess durchaus bereichernd sein (Oevermann et al. 1979: 393). So oder so, es geht darum, dass sich der spontan Interpretierende von der Gruppe getragen fühlt. Von demotivierenden, Denkblockaden auslösenden Stellungnahmen ist in diesem Zusammenhang daher abzusehen. Im *kontinuierlichen Oszillieren zwischen dynamisierender Lesartenproduktion und steuernder Lesartenevaluation* kann sich die Gruppe durch den tendenziell chaotischen Interpretationsprozess hindurch navigieren und am Ende der Sinn rekonstruierenden Textinterpretation eine Antwort auf die Frage geben, auf welche Frage der Einzelfall die Antwort ist.

## Literatur

- Caillois, Roger (1982): Die Spiele und die Menschen. Frankfurt a.M.: Ullstein
- Dilthey, Wilhelm (1957): Die geistige Welt. Stuttgart/Göttingen: Teubner/Vandenhoeck & Ruprecht
- Dilthey, Wilhelm (1958): Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften. Stuttgart/Göttingen: Teubner/Vandenhoeck & Ruprecht
- Gadamer, Hans-Georg (1990): Wahrheit und Methode. Tübingen: Mohr
- Heidegger, Martin (1993): Sein und Zeit. Tübingen: Niemeyer
- Herbrik, Regine (2011): Die kommunikative Konstruktion imaginärer Welten. Wiesbaden: VS Verlag
- Kierkegaard, Søren (1984): Die Wiederholung. Frankfurt a.M.: Syndikat
- Kurt, Ronald (2002): Menschenbild und Methode der Sozialphänomenologie. Konstanz: UVK
- Kurt, Ronald (2004): Hermeneutik. Konstanz: UVK/UTB
- Kurt, Ronald (2008): Vom Sinn des Sehens. Phänomenologie und Hermeneutik als Methoden visueller Erkenntnis. In: Raab, Jürgen/Pfadenhauer, Michaela/Stegmaier, Peter/Dreher, Jochen/Schnettler, Bernt (Hg.): Phänomenologie und Soziologie. Wiesbaden: VS Verlag. 369-378
- Kurt, Ronald (2009a): Indien und Europa. Bielefeld: transcript
- Kurt, Ronald (2009b): Hermeneutik. Die Kunstlehre des (Nicht-)Verstehens. In: Rehbein, Boike/Saalmann, Gernot (Hg.): Verstehen. Konstanz: UVK. 71-91
- Kurt, Ronald (2012): Improvisation als Grundbegriff, Gegenstand und Methode der Soziologie. In: Göttlich, Udo/Kurt, Ronald (Hg.): Kreativität und Improvisation. Soziologische Positionen. Wiesbaden: Springer VS Verlag. 166-186
- Lévinas, Emmanuel (1984): Die Zeit und der Andere. Hamburg: Meiner
- Marquard, Odo (1991): Frage nach der Frage, auf die die Hermeneutik die Antwort ist. In: ibidem (Hg.): Abschied vom Prinzipiellen. Stuttgart: Reclam. 117-146
- Mead, George Herbert (1973): Geist, Identität und Gesellschaft aus der Sicht des Sozialbehaviorismus. Frankfurt a.M.: Suhrkamp
- Merton, Robert/Barber, Elinor (2004): The Travels and Adventures of Serendipity. Princeton: University Press
- Musil, Robert (1997): Der Mann ohne Eigenschaften. Reinbek: Rowohlt
- Oevermann, Ulrich/Allert, Tilman/Kronau, Elisabeth/Krambeck, Jürgen (1979): Die Methodologie einer objektiven Hermeneutik und ihre allgemein forschungslogische Bedeutung in den Sozialwissenschaften. In: Soeffner, Hans-Georg (Hg.): Interpretative Verfahren in den Sozial- und Textwissenschaften. Stuttgart: Metzler. 352-433
- Raab, Jürgen (2008): Visuelle Wissenssoziologie. Konstanz: UVK
- Reichertz, Jo (1986): Probleme qualitativer Sozialforschung. Frankfurt a.M./New York: Campus
- Reichertz, Jo (2013): Gemeinsam interpretieren. Wiesbaden: VS Verlag

- Schiller, Friedrich (1983): Werke in vier Bänden. Band 1. Salzburg: Das Bergland-Buch
- Schleiermacher, Friedrich (1959): Hermeneutik. Heidelberg: Carl Winter Universitätsverlag
- Schleiermacher, Friedrich (1995): Hermeneutik und Kritik. Frankfurt a.M.: Suhrkamp
- Schütz, Alfred (2004): Der sinnhafte Aufbau der sozialen Welt. Konstanz: UVK
- Schütz, Alfred (2010): Zur Methodologie der Sozialwissenschaften. Konstanz: UVK
- Soeffner, Hans-Georg (1989): Auslegung des Alltags – Der Alltag der Auslegung. Frankfurt a.M.: Suhrkamp

Boris Traue, Lisa Pfahl und Lena Schürmann

---

## 34.1 Was ist ein Diskurs?

Der Begriff „Diskurs“ bezeichnet in seiner allgemeinsten Bedeutung die Produktion sozialen Sinns, verstanden als die symbolische (d.h. sprachliche, visuelle, auditive) Darstellung, Vermittlung und Konstitution von Gegenständen in kommunikativen Prozessen.

In der sozialwissenschaftlichen Diskursforschung (Keller et al. 2001) wird heute einerseits die großflächige Produktion von Sinn bzw. die Genese und Zirkulation von materiell oder symbolisch fixiertem Wissen in Diskursgemeinschaften untersucht. Andererseits setzt sie einen zweiten Schwerpunkt auf die interaktive Aushandlung von Sinn zwischen mehreren Kommunikationspartner/-innen. Damit stellt die empirische Erforschung des Verhältnisses von Wissen und Sozialstruktur bzw. von Wissen und Macht den zentralen Bezugspunkt der Diskursforschung dar – im Unterschied zu stärker sprachwissenschaftlich geprägten Perspektiven wie der ethnomethodologischen Konversationsanalyse (vgl. Bergmann 1981). Zwei unterschiedliche Modelle der Sinnkonstitution bilden in Diskursanalysen sich ergänzende Bezugspunkte: Während im *strukturalistischen Modell* davon ausgegangen werden kann, dass Sinn durch Differenzbildungen zwischen diskursiven Einheiten entsteht, kann im Anschluss an den *philosophischen Pragmatismus* und die *phänomenologisch orientierte Soziologie* davon ausgegangen werden, dass Sinn aus der sequentiellen Verknüpfung dieser Einheiten im Handeln spezifischer Akteure hervorgeht.

Aus welchen *diskursiven Einheiten* sich Diskurse jeweils zusammensetzen, ist durch die Methodologie nicht vorentschieden. Auch die Frage, wann von einem Diskurs gesprochen werden kann, der sich meist aus unterschiedlichen Diskurssträngen zusammensetzt, muss jeweils am Gegenstand bestimmt werden. Entscheidendes Kriterium für die Bestimmung eines Diskurses ist, ob sich unterschiedliche Praktiken auf einen übersituativen, überindividuellen und zumindest vorläufig stabilen Sinn beziehen bzw. diesen herstellen. So kann

---

<sup>1</sup> Diese Publikation wurde von der Deutschen Forschungsgemeinschaft im Rahmen des Forschungsprojektes “Audiovisuelle Kulturen der Selbstthematisierung” gefördert.

beispielsweise von Diskursen der Sexualität, der Medizin, der Behinderung, der Kunst, der Weiblichkeit und ebenso vom Schmutz-Diskurs oder vom neoliberalen Diskurs die Rede sein.

---

### 34.2 Was ist Diskursanalyse?

Die *Diskursanalyse* ist eine Forschungsperspektive, die es sich zum Ziel setzt, Diskurse, d.h. übersituative und überindividuelle Praktiken der Wissens- bzw. Sinnproduktion zu analysieren. Sie hat seit den 1970er-Jahren ausgehend von Frankreich in den Sozial- und Kulturwissenschaften weltweite Verbreitung gefunden. Die Diskursanalyse beruht auf einer historisch-vergleichenden Vorgehensweise im Anschluss an die wissenschaftshistorischen Studien von Gaston Bachelard und Georges Canguilhem. Sie hat damit starke Ähnlichkeiten mit historisch-soziologischen Arbeiten, wie sie etwa Max Weber und Marcel Mauss in ihren wirtschafts- und religionssoziologischen Studien entwickelt haben. Vor allem in ihren strukturtheoretischen Aspekten weist sie Ähnlichkeiten mit funktionalistischen Perspektiven in der Soziologie auf, insofern es in der Diskursanalyse – wie etwa auch in der Soziologie Talcott Parsons – um eine *Analyse der Genese und der Bestandserhaltungsbedingungen von Normen, institutionellen Gefügen, sozialen Rollen sowie Persönlichkeitsstrukturen* („*Subjektivierung*“) geht. Von den normativistischen und funktionalistischen Perspektiven unterscheidet sich die Diskursanalyse allerdings durch eine handlungstheoretische Grundannahme: Diskurse bestehen immer aus *diskursiven Praktiken* – also aus lokal situierten Handlungsvollzügen und Materialitäten – und sind dadurch Gegenstand und Anlass von Transformationsprozessen. Vor allem aufgrund dieses Fokus auf diskursive „Brüche“ wird die Diskursanalyse zu den poststrukturalistischen Ansätzen gezählt. Gegenüber hermeneutischen Zugängen (Kurt/Herbrik, Kapitel 33 in diesem Band) zur sozialen Konstruktion der Wirklichkeit wird hier die *Heterogenität (Uneinheitlichkeit)*, *Ambivalenz (Vieldeutigkeit)* und *Brüchigkeit (zeitliche Begrenzung)* von *Sinnphänomenen* betont.

Die Diskursanalyse gehört mit ihrer Analyse längerer Zeiträume zu den *historischen Methoden*. Im Allgemeinen zielt sie darauf ab, Gegenwartsphänomene in ihrer historischen Gewordenheit zu erschließen. In dieser zugleich historischen und gegenwartsbezogenen Ausrichtung liegt ein zweiter wichtiger Unterschied zu hermeneutischen Forschungsmethoden wie der sozialwissenschaftlichen Hermeneutik (Kurt/Herbrik, Kapitel 33 in diesem Band) oder der dokumentarischen Methode (Bohnsack, Kapitel 67 in diesem Band). Bei Diskursanalysen handelt es sich somit immer um historische Längsschnittuntersuchungen. Wenn die Untersuchung der Genese eines Diskurses die Konstitution der Subjektivität der Diskursproduzent/-innen oder -adressat/-innen einschließt, wird – im Anschluss an Foucault's Nietzsche-Rezeption – von einer *Genealogie* gesprochen (Saar 2007).

Im Mittelpunkt der sozialwissenschaftlichen Diskursforschung steht also die *Analyse von Wissensformen* in ihrem Verwendungszusammenhang über eine gewisse Zeitdauer. Dabei geht es nicht um die Untersuchung konkreter Sprechsituationen, d.h. um die

einzelne „Außerung“ (enunciation), sondern um eine Untersuchung der Wechselwirkung zwischen der *Aussage* (enoncé) und den gesellschaftlichen Bedingungen und Folgen dieser Aussage. Während z.B. in der interaktionstheoretisch orientierten Konversationsanalyse die situative und flüchtige Produktion von Sinnbezügen im Vordergrund steht, geht es in der Diskursanalyse um die *Erschließung von übersituativen Simfiguren und deren situativen und strukturellen Bestandsbedingungen*, die insbesondere in Machtverhältnissen gesehen werden. *Machtverhältnisse* können dabei Bedingungen von Diskursen sein – etwa beim Zugang und der Besetzung von Sprecherpositionen („Wer darf sprechen?“), sie können aber auch als Effekt von Diskursen verstanden werden – etwa wenn diese normative, normalisierende oder legitimatorische Deutungen produzieren („Was kann/darf gesagt werden?“). Während bis in die 1990er-Jahre beinahe ausschließlich sprachliche Muster Gegenstand von Diskursanalysen waren, werden heute auch Fotografien (Bohnack, Kapitel 67 in diesem Band), Video (Tuma/Schnettler, Kapitel 68 in diesem Band), Filme (Akremi, Kapitel 69 in diesem Band) und andere nicht-zeichenhafte Symbolformen einbezogen (Traue 2013).

---

### 34.3 Varianten der sozialwissenschaftlichen Diskursanalyse

Die Ende der 1960er Jahre durch den französischen Sozialphilosophen Michel Foucault (1926–1984) als Methode eingeführte Diskursanalyse (Foucault 1981) knüpft an verschiedene kulturhistorische und sozialwissenschaftliche Verfahrensweisen an. Sie bietet die Möglichkeit, insbesondere solche gesellschaftlichen Konstruktionen von Wirklichkeit empirisch zu untersuchen, die über individuelle Intentionen und lokale Interaktionen hinausgehen, dabei aber dennoch einen strategischen, d.h. machtvollen Charakter besitzen (können). Die Macht der Diskurse wird dabei nicht als einschränkendes, sondern als produktives Moment gedacht, etwas als „Redeauflorderung“ (Foucault 1989: 49) gedacht, das verschiedene Deutungen und Handlungsformen stützt, während es andere erschwert oder verhindert.

Die verschiedenen Varianten der sozialwissenschaftlichen Diskursanalyse gehen dementsprechend mit unterschiedlichen Akzentsetzungen auf die Frage ein, wie das *Verhältnis von Wissen, Macht, Akteurskonstellationen und sozialen Konstruktionsprozessen* zu fassen ist. In Frankreich entwickelte Michel Pêcheux parallel zu Foucault eine quantitativ ausgerichtete Diskursanalyse, die sich verändernde Häufigkeiten von bestimmten Worten oder Wortfeldern in großen Korpora untersucht, die Aufschluss über massenmediale Bedeutungsproduktion geben (zur Übersicht vgl. Diaz-Bone 2003). Einen Überblick über weitere Ausdifferenzierungen im Forschungsfeld gibt das interdisziplinäre Wörterbuch der Diskursforschung (Reisigl et al. 2014). Im Folgenden werden die im deutschsprachigen Raum verbreiteten, unterschiedlichen Varianten der Diskursanalyse kurz vorgestellt. Dabei wurden die Varianten ausgewählt, die für sozialwissenschaftlichen Fragestellungen bedeutsam sind: Kritische Diskursanalysen, Gouvernementalitätsstudien, postmarxistische Diskurstheorien, wissenssoziologische Diskursanalysen sowie Diskursanalysen in den Gender Studies und Queer Studies.

### **34.3.1 Kritische Diskursanalyse**

Die „kritische Diskursanalyse“ (Jäger 2004) ist eine sprachwissenschaftlich orientierte Variante der Diskursforschung, die sich explizit als ideologiekritisch begreift und der Verschränkung von Sprache und Herrschaftszusammenhängen nachgeht. Sie wird in Großbritannien von Norman Fairclough, im deutschsprachigen Bereich insbesondere von Siegfried Jäger, Ruth Wodack und Marianne Bublitz vertreten (zur Übersicht vgl. Reisigl 2011). Analysiert werden einzelne Sprechereignisse ebenso wie institutionelle oder massenmediale Diskurse, um nachzuvollziehen, wie die Artikulationsmöglichkeiten Einzelner oder sozialer Gruppen beschränkt bzw. in spezifische Richtungen kanalisiert werden und wie es zur Verfestigung gesellschaftlicher Herrschaftsverhältnisse kommt (Fairclough/Wodack 1997). Eine eigenständige Variante (die hier aufgrund ihrer sprachwissenschaftlichen Ausrichtung in die kritische Diskursanalyse eingeordnet wird) bildet die von Johannes Angermüller entwickelte „Aussagenanalyse“ (Angermüller 2007).

### **34.3.2 Gouvernementalitätsstudien**

In der Gouvernementalitätsforschung wird die Regierung von Individuen und Bevölkerungen als Verhältnis von Fremdführung und Selbstführung bestimmt (Foucault 1987). Im Zentrum stehen Formen der Regierung des Sozialen in der Moderne, die diskursanalytisch untersucht werden. Dies schließt eine Analyse der diskursiv angeleiteten Sozial- und Selbsttechniken im Rahmen von Dispositiven (Schneider/Bührmann 2012) ein. In Großbritannien sind die Arbeiten von Nikolas Rose und Monica Greco sowie das critical accounting („kritische Buchhaltung“) von zentraler Bedeutung für die dortige sozialwissenschaftliche Rezeption der Diskursanalyse. Diese Forschungen werden in den deutschsprachigen Gouvernementalitätsstudien fortgesetzt (Bröckling et al. 2000).

### **34.3.3 Postmarxistische und psychoanalytische Diskurstheorie**

Die postmarxistische Diskurstheorie von Ernesto Laclau und Chantal Mouffe (Laclau/Mouffe 2001) verbindet Fragen nach der Subjektkonstitution und der Konstruktion kollektiver Identitäten mit demokratietheoretischen Themen und sucht dabei Anschluss an die Ideologietheorie von Louis Althusser und die Subjekttheorie Jacques Lacans. In psychoanalytischer Perspektive geht Lacan von einem begehrenden Subjekt aus, das einen grundsätzlichen Mangel erlebt. Aufgrund dieses Mangels sind Akteure anfällig für Versprechungen, die durch Diskurse kommuniziert werden, die keinen eindeutigen Gegenstand haben („leerer Signifikant“), wodurch Hegemonien, also Vorherrschaften von bestimmten Theorien, Wissenschaften oder Selbst- und Weltbezügen verankert werden.

### 34.3.4 Wissenssoziologische Diskursanalyse

Mit der wissenssoziologischen Diskursanalyse (WDA) (Keller 2005) liegt eine Integration der französischen Diskursanalyse Foucaultscher Prägung in das interpretative Paradigma und insbesondere in die Wissenssoziologie (Kurt/Herbrik, Kapitel 33 in diesem Band) vor. Die „diskursive Konstruktion gesellschaftlicher Wirklichkeit“ vollzieht sich demnach in Prozessen intersubjektiv und medial vermittelter Kommunikation. Die WDA erweitert die Analyse situationsübergreifender diskursiver Praktiken um Analysen der interaktiv-kommunikativen und medialen Konstruktion (Keller et al. 2012) des diskursiven Wissens. Eine solche Diskursanalyse versteht soziale Akteure sowohl als Produzent/-innen als auch als Rezipient/-innen von Diskursen. Sie sind damit diskursiv konstituierte und regelinterpretierende Handelnde, deren Mitwirkung am diskursiven Geschehen jedoch für die Existenz von Diskursen unumgänglich ist, denn Diskurse existieren nur, soweit sie handelnd artikuliert werden. Gegenüber anderen diskursanalytischen Ansätzen werden die pragmatischen Aspekte der Konstitution von diskursivem Sinn im Handeln betont sowie der Prozess der Institutionalisierung von Wissen re- und dekonstruiert.

### 34.3.5 Diskursanalyse in den Gender Studies und Queer Studies

Innerhalb der Gender und Queer Studies führte die Rezeption des diskursanalytischen Ansatzes zu einer grundlegenden Diskussion der wissenschaftlichen Praxis und zu einer Thematisierung des Verhältnisses von Gegenstand und Erkenntnisinteresse. Hinterfragt wurde die Konstruktion eines vorgeblich neutralen Beobachtungsstandpunkts in wissenschaftlichen Untersuchungen und damit der Rolle der Wissenschaftler/-innen selbst, die immer gleichzeitig Adressat/-innen, Teilnehmer/-innen und Produzent/-innen gesellschaftlicher Wissensproduktion von Geschlecht(-erdifferenzen) sind (Engel 2009, Hark 2011). Daran knüpft sich eine systematische Reflexion der eigenen Position als Forschende/-r und der Wissens- und Wahrheitstechniken des wissenschaftlichen Feldes an (Bublitz et al. 2000).

### 34.3.6 Visuelle Diskursanalysen

Um der zunehmenden visuellen und audiovisuellen Sinnproduktion Rechnung zu tragen, wurde die diskursanalytische Methodologie auf die Analyse von Bildern und Bildregimen ausgeweitet (Maasen et al. 2006; Engel 2009). Fotografien (Bohnsack, Kapitel 67 in diesem Band), Videos (Tuma/Schnettler, Kapitel 68 in diesem Band), Filme (Akremi, Kapitel 69 in diesem Band), Sounds und andere Modalitäten von Sinn werden dabei als konstitutiver Bestandteil von Diskursen aufgefasst. Die Untersuchung von Bilddiskursen als ästhetische Praktiken öffnet die Diskursanalyse für Fragen nach der performativen Umgestaltung von Wissensordnungen in der (Gegenwarts-)Gesellschaft. Felder des Sagbaren werden dabei in ihrer Wechselwirkung mit Feldern des Sichtbaren analysiert (Traue 2013).

## 34.4 Zur Vorgehensweise sozialwissenschaftlicher Diskursanalysen

Im Folgenden werden die grundlegenden Verfahrensschritte für sozialwissenschaftliche Diskursanalysen dargelegt. Dabei werden zunächst die methodischen Grundbegriffe erläutert, anschließend wird auf die Korpusbildung bzw. das Sampling sowie auf einzelne Auswertungsschritte im Forschungsprozess (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) eingegangen.

Da es – wie die obige Darstellung zeigt – unterschiedliche Varianten der Diskursanalyse gibt und deren theoretischen Annahmen zumindest in Teilen zu verschiedenen Verfahrensweisen führen, greifen wir bei der Darstellung der Begrifflichkeiten (Abschnitt 34.4.1) auf die von Foucault beschriebenen grundlegenden Konzepte zurück (Gegenstand, Äußerungsformen, Begriffe und Strategien des Diskurses sowie die Einheit des Diskurses und diskursive Formation). Anschließend werden Korpusbildung (Abschnitt 34.4.2) und einzelne Auswertungsschritte (Abschnitt 34.4.3) beschrieben.

### 34.4.1 Grundbegriffe

Das Konzept des Diskurses entstand im Zuge der Bemühungen Michel Foucaults, die Vorgehensweise seiner frühen Untersuchungen zu erklären, insbesondere im Hinblick auf die Auswahl von Daten und der Ebenen ihrer Interpretation. In der „Archäologie des Wissens“ (Foucault 1981), einem zentralen frühen Grundlagentext der Diskursforschung, wird vorgeschlagen, vier Aspekte – Begriffe, Gegenstände, Äußerungsmodalitäten, Strategien – zu unterscheiden und diese in ihren historisch spezifischen und gesellschaftlich verorteten Wissensordnungen (Foucault 1981) zu rekonstruieren.

Was ist im Einzelnen mit Gegenständen, Äußerungsformen, Begriffen und Strategien gemeint?

Als *Gegenstand eines Diskurses* wird dasjenige Phänomen bezeichnet, worauf dieser verweist, bzw. welches durch den Diskurs als außerhalb des Diskurses stehend konstruiert wird: Auch „Unsichtbares“ kann im Diskurs zur Sprache kommen und zum Gegenstand des Diskurses werden, wie z.B. Kompetenz und Leistung. Um das Erscheinen der Gegenstände zu erklären, ist es nötig, die Oberfläche ihres Auftauchens zu untersuchen und die Regeln zu benennen, mit der die zugehörigen diskursiven Bezeichnungen eingeführt und aufrechterhalten werden. Foucault zufolge werden in Diskursen vielfältige Gegenstände „genannt, umschrieben, analysiert, dann berichtet, erneut definiert, in Frage gestellt, gelöscht“ (Foucault 1981: 62). Worum es geht, ist die Regeln festzustellen, die ihr Erscheinen erklären können.

Auf diese Gegenstände, also auf außerdiskursive Wirklichkeiten, verweist ein Diskurs nicht einfach, sondern sie werden in ihm selbst durch die symbolischen Formen ihrer Thematisierung hervorgebracht. Deshalb wird es notwendig, die *Äußerungsformen*, also die *Art und Weise der Gegenstandskonstruktion*, durch Beschreibungen, Narrationen,

Klassifikationen, Abbildungen, Illustrationen, Modelle etc. sowie die Verbindung zwischen diesen zu rekonstruieren.

Begriffe erfüllen im Diskurs, vereinfacht gesagt, die Aufgabe, *Gegenstände miteinander zu verknüpfen und ins Verhältnis zu setzen*. Wie kann die Formation der Begriffe untersucht werden? In Diskursanalysen muss die einfache ideengeschichtliche Rekonstruktion einer Wissensform, z.B. eines wissenschaftlichen Theoriegebäudes, streng vermieden werden. Stattdessen wird untersucht, in welchem thematischen und strategischen Zusammenhang ein Begriff auftaucht, wie sich sein Gebrauch wandelt oder welche Anschrift von Themen er erlaubt oder forciert. „Anstatt die Begriffe in einem virtuellen deduktiven Gebäude erneut anordnen zu wollen, müsste man die Organisation des Feldes der Aussagen beschreiben, in dem sie auftauchen und zirkulieren“ (Foucault 1981: 83). Damit wird einerseits die Frage nach dem Inhalt der Begriffe gestellt: Mit welchen anderen Begriffen, Redeweisen und Themen unterhält ein Begriff Korrespondenzen? In welchen Relationen steht er zu anderen? In diesem Zusammenhang ermöglicht die empirische Analyse der Diskurse bzw. der Diskursstränge eine Abgrenzung der Anwendungsfelder von Begriffen und erlaubt damit auch deren Systematisierung.

Die Formation der Strategien bezeichnet die Absichten, die Akteure mit dem Einsatz einer Aussage verfolgen. Dabei wird der Diskurs auf seine institutionellen, ökonomischen und technischen Möglichkeits- und Erhaltungsbedingungen bezogen. Strategien umfassen im Diskurs all das, was Akteure und Akteursgruppen absichtlich oder unabsichtlich anstreben. Die Bestimmung der Diskontinuitäten und Wendepunkte eines Diskurses lässt Rückschlüsse auf die Absichten der Diskursakteure zu. Die Bruchpunkte des Diskurses zeigen damit immer auch die (wechselnden) disziplinären Strategien der Expertenschaft auf (Foucault 1981: 96).

In welche sozialstrukturellen, wirtschaftlichen und professionellen Kräfteverhältnisse ist der Diskurs eingebunden? Um diese Fragen zu beantworten, muss untersucht werden, welche Funktion der Diskurs in sozialen Feldern ausübt (Foucault 1981: 99). Eine Funktion von Diskursen kann etwa darin bestehen, die Existenz von Institutionen zu legitimieren. Wissenschaftliche Diskurse oder die Diskurse bestimmter Expertengruppen können Legitimationstheorien für Institutionen bereitstellen und damit eine „sekundäre Objektivation von Sinn“ (Schütz/Luckmann 1984: 98) leisten. Wie Institutionen kategoriale gesellschaftliche Grenzziehungen und damit institutionelle Wahrheit (re-)produzieren, lässt sich dann durch eine Analyse der wissenschaftlichen und professionellen Diskurse rekonstruieren.

Eine zentrale machtvolle Wirkung des Diskurses, die einen Teil seiner Strategie ausmacht, liegt in seinem *Wahrheitseffekt*. Die „Ordnung des Diskurses“ (Foucault 1991) gibt vor, welche Aussagen in seinem Rahmen akzeptabel sind, d.h. Wahrheitswert haben. Aussagen können damit innerhalb oder außerhalb des Wahren des Diskurses liegen und damit als verrückt, radikal, verschroben, altmodisch und letztlich irrelevant gelten. Diskurse sind also, wie bereits eingangs dargestellt, als überindividuelle Wissensbestände zu bezeichnen, die die Deutungsprozesse Einzelner anleiten. Diskurse beinhalten Problematisierungen: Sie bieten interpretative Rahmen zur Deutung von Handlungsproblemen, legen sie nahe

oder zwingen sie Akteuren auf. Ihre Entstehung ist nicht an subjektiv-biographische Relevanzen gebunden, obwohl solche Relevanzen in sie eingehen.

Woraus setzen sich nun diese *diskursiven Wissensvorräte* zusammen, was sind also die Einheiten des Diskurses?

Das zentrale methodische Problem einer Diskursanalyse besteht darin, die jeweils zu untersuchenden *Einheiten des Diskurses* zu bestimmen, also eine nicht-reduktionistische Frageweise zu entwickeln, die Diskurse weder auf Ideen, noch auf sprachliche Codes, noch auf Akteursstrategien oder auf ihre Referenzen beschränkt. Die Einheit des Diskurses ist dabei kein Forschungskonstrukt, sondern als diskursive Formation ein Effekt gesellschaftlicher Kräfteverhältnisse. Diese können nachträglich von Forschenden mit einem Interesse an der Beschreibung und Kritik solcher Einheitsbildungen dekonstruiert werden.

Doch auf welche Weise fügen sich die Einheiten, also die kleinsten und zugleich formgebenden Elemente eines Diskurses zu größeren Wissenskomplexen zusammen, die in der Diskursanalyse *diskursive Formation* genannt werden?

Um eine diskursive Formation zu rekonstruieren, muss das Verhältnis zwischen dem Diskurs als Wissensform und den sozialen Regeln und Umständen seiner Herstellung geklärt werden. Dazu sollte zwischen der Form der Äußerungen, den Begriffen, den Gegenständen und den Akteursstrategien unterschieden werden, um Regelmäßigkeiten zu identifizieren (Foucault 1981: 58). *Formationsregeln* sind die Bedingungen, unter denen die Gegenstände der Diskurse angeordnet sind. Sie beschreiben das Auftauchen und Verschwinden von Begriffen und Äußerungen, begründen die Etablierung eines Gegenstandes und geben Auskunft über die strategische Orientierung der Diskursgestalter/-innen: „Die Formationsregeln sind Existenzbedingungen (aber auch Bedingungen der Koexistenz, der Aufrechterhaltung, der Modifizierung und des Verschwindens) in einer gegebenen diskursiven Verteilung“ (Foucault 1981: 58).

### **34.4.2 Korpusbildung und Sampling-Strategie**

Im Rahmen der sozialwissenschaftlichen Diskursforschung müssen zunächst ähnliche Entscheidungen getroffen werden, wie sie in anders gelagerten qualitativen und sogar quantitativen Forschungen auch notwendig sind: dies betrifft die Frage nach einer geeigneten Auswahl von Daten, die eine Erforschung des Gegenstands ermöglichen und zugleich die Generierung falsifizierbarer Aussagen über den gewählten Gegenstand erlauben.

In Diskursanalysen werden prinzipiell „natürliche Daten“ (Salheiser, Kapitel 62 in diesem Band) interpretiert. Die Daten sind „prozessgeneriert“, d.h. sie sind nicht durch die Forschenden hergestellt (wie bei einer „Erhebung“ von Interviews), sondern werden im Feld vorgefunden (wie beim Besuch eines „Archivs“). Die Zusammenstellung von Daten zu einem sogenannten *Korpus* (Akremi, Kapitel 17 in diesem Band) ist deshalb von großer Bedeutung: Mit der Auswahl der zentralen Texte, Dokumente (Salheiser, Kapitel 62 in diesem Band), Zeitungsartikel (Klein, Kapitel 64 in diesem Band), literarischen Quellen und persönlichen Dokumenten (Ernst, Kapitel 63 in diesem Band), Bilder (Bohsack, Kapitel 67

in diesem Band), Filme (Akremi, Kapitel 69 in diesem Band), Artefakte (Schubert, Kapitel 70 in diesem Band) etc. werden der Gegenstand, der inhaltliche Zuschnitt und die Reichweite einer Diskursanalyse festgelegt. Auf die Festlegung des Korpus einer Diskursanalyse muss also besondere Aufmerksamkeit verwendet werden, weil Diskursforscher/-innen im Forschungsprozess immer an der Konstituierung dessen beteiligt sind, was als Diskurs beschrieben wird.

Die *Sampling-Strategie* der Diskursanalyse kann als eine Spielart des theoretischen Samplings (Strübing, Kapitel 32 in diesem Band) angesehen werden. Eine explorative Untersuchung des Forschungsfeldes durch Literaturstudium (Ohly/Weber, Kapitel 5 in diesem Band), Sichtung von Dokumenten und, wenn möglich, Gewinnung eigener Erfahrungen mit dem Forschungsgegenstand durch teilnehmende Beobachtungen (Thierbach/Petschick, Kapitel 66 in diesem Band) leiten die Zusammenstellung eines Korpus an. Der Korpus wird dabei in Grundzügen zu Beginn der Untersuchung festgelegt, kann aber im Verlauf der Forschung erweitert oder eingeschränkt werden. Diese Einschränkungen oder Erweiterungen erfolgen nicht beliebig: Jede Diskursanalyse muss nachweisen, dass eine oder mehrere diskursive Formationen, d.h. Verdichtungen von Begriffen, Strategien, Gegenständen und – man könnte ergänzen – Medialitäten (Traue 2013) vorliegen.

Denkbar ist etwa die Zusammenstellung eines Korpus aus allen Ausgaben einer bestimmten Zeitschrift, die zugleich Verlautbarungen eines Verbandes über einen gewissen Zeitraums darstellen (Pfahl 2011; Schürmann 2013); Äußerungen mächtiger und marginalisierter Diskursproduzenten (Freitag 2005); Darstellungen bzw. Nennungen eines Themas in bestimmten Massenmedien (Küçük 2008); Ratgeber und Managementliteratur (Bröckling 2007), Praktikerliteratur von Berufsgruppen (Traue 2010) sowie Werbefotografie (Engel 2009) oder Videoblogs (Traue 2013).

Zu vermeiden ist eine „armchair discourse analysis“, also eine Diskursanalyse, die allein vom Schreibtisch aus betrieben wird. Diskursanalysen erfordern im Allgemeinen eine Vertrautheit der Forschenden mit dem untersuchten sozialen Feld. Methodische Fehler in Diskursanalysen (etwa arbiträre Zusammenstellungen von Themen) sind meist das Ergebnis fehlender Vertrautheit mit dem Forschungsfeld und der Organisation der Wissensproduktion im jeweiligen Feld. Erkennbar werden solche Qualitätsmängel, wenn vorgebliebene Diskursanalysen zurückfallen in Begriffs- oder Ideengeschichte, reine Institutionenbeschreibungen oder gar Inhaltsanalysen.

### 34.4.3 Auswertungsschritte und Analyseebenen

Die Diskursanalyse interessiert sich also „für die Formationsmechanismen von Diskursen, die Beziehung zwischen Diskursen und Praktiken sowie strategisch-taktische Diskurs-Performanz sozialer Akteure“ (Keller 2005: 182). Diskurse stellen überindividuelle Wissensformen dar, die einerseits durch Deutungspraktiken fortgeschrieben werden und sich andererseits ‚hinter dem Rücken‘ der Akteure reproduzieren und transformieren. Gegenüber dem hermeneutisch-wissenssoziologischen Ansatz (Kurt/Herbrik, Kapitel 33

in diesem Band), in dem Deutungsleistungen der Handelnden nicht systematisch in Bezug zu kollektiven Wissensvorräten gestellt werden, steht dieser Zusammenhang in der sozialwissenschaftlichen Diskursanalyse im Mittelpunkt. Ziel der Auswertung ist daher „nicht nur die Deskription der empirischen Vielfalt von subjektiven, typisierbaren Wissensvorräten, sondern auch die Analyse der kollektiven und institutionellen Prozesse, in denen spezifisches Wissen zur gesellschaftlichen Wirklichkeit wird“ (Keller 2005: 185).

Die Diskursanalyse ist ein interpretatives Verfahren, jedoch geht es in ihr anders als bei hermeneutischen Verfahren nicht um die Entschlüsselung latenter impliziter oder tiefenstruktureller Sinnsschichten. Die interpretative Leistung – und ein Großteil der Untersuchungsarbeit – liegt in der vergleichenden Anordnung des Materials und der Zuordnung seiner Elemente zu Aussageformationen, die sich empirisch bewähren. Worin die Einheit eines Diskurses jeweils besteht, ist nicht *a priori* bekannt – anders als bei den hermeneutischen Verfahren, bei denen im Vorhinein entschieden wird, welches Material interpretiert werden soll. Der Erkenntnisfortschritt in den hermeneutischen Verfahren erfolgt demgemäß typischerweise schrittweise-zirkulär, während in Diskursanalysen die vorgefundenen diskursiven Elemente solange angeordnet werden, bis sich eine Passung von Begriffssystemen, Institutionendaten, sozialen Ordnungen und diskursiven Praktiken anzeigt. Dadurch ergibt sich in der Darstellung der Ergebnisse der Diskursforschung oft der Eindruck einer Evidenz bzw. Einfachheit der Befunde; es scheint, als zeigten sich die Befunde „von selbst“ oder als seien sie dem Forschenden von vornherein bekannt gewesen. Dies ist selbstverständlich nicht der Fall, scheint aber dem ungeübten Lesenden leicht so – analog dem Vorurteil, die Hermeneutikerin hätte etwas in das Material „hineininterpretiert“.

Wie wird praktisch vorgegangen?

Zunächst muss festgestellt werden, dass es in Diskursanalysen *kein feststehendes Schema von notwendigen Analyse- und Beschreibungsschritten* gibt. Die analytischen Einheiten und Beschreibungsebenen müssen streng am jeweiligen Gegenstand entlang entwickelt werden. Formale und inhaltliche Elemente des betreffenden (hypothetischen) Diskurses werden benannt und in ihrer „spezifischen Gestalt“ bzw. „Phänomenstruktur“ rekonstruiert (Keller 2004: 99ff.). Über die verschiedenen Untersuchungszeitpunkte hinweg wird die Gegenstands- bzw. Phänomenstruktur des jeweiligen Diskurses erarbeitet, indem die Gegenstände, Themen, Begriffe des Diskurses herausgearbeitet werden.

„Die tatsächlichen Bausteine einer solchen Problemstruktur sind nicht vor der Datenanalyse bekannt, sondern sie müssen aus den empirischen Daten – und dort aussageübergreifend – erschlossen werden. Einzelne Daten enthalten dazu in der Regel nur partielle Elemente.“ (Keller 2004: 99). Die in den Texten vorkommenden Begriffe, Gegenstandsbeschreibungen und Problematisierungen, Argumentationen, Deutungsfiguren und Metaphern werden zunächst *kodiert* (Mayring/Fenzl, Kapitel 38 in diesem Band). In einem nächsten Schritt werden typische Äußerungen einer *Feinanalyse* (Kurt/Herbrik, Kapitel 33 in diesem Band) unterzogen. Dabei können *Aspekte der Analyse*, die in der hermeneutischen Wissenssoziologie eingeführt sind, benutzt werden (Deutungsmuster, Klassifikationen, narrative Strukturen, Visualisierungen etc.):

- In Diskursanalysen kann z.B. untersucht werden, welche *Deutungsmuster* im Diskurs produziert werden, wie sich Deutungsmuster im Zeitverlauf verändern oder was ihre jeweiligen Entstehungsbedingungen sind. Keller definiert Deutungsmuster als „grundlegende bedeutungsgenerierende Schemata, die durch Diskurse verbreitet werden und nahe legen, worum es sich bei einem Phänomen handelt. Diskurse verknüpfen verschiedene Deutungsmuster zu einem Deutungsarrangement“ (Keller 2005: 238). Deutungsmuster vermitteln gesellschaftliche Problemdeutungen in das Alltagswissen. Auf Seiten der Subjekte entfalten sie eine handlungsentlastende, weil routinstiftende Funktion.
- *Klassifikationen* können als institutionalisierte Form sozialer Typisierung begriffen werden, als kontingente Modelle der Wirklichkeitskonstruktion durch Gruppenbildungen. In Klassifikationsprozessen werden disparate Dinge zu Klassen zusammengefasst und mit moralischem und politischem Gehalt versehen (Douglas 1991: 107f.). Die Untersuchung der diskursiven Klassifikationsprozesse und -praktiken ermöglicht Aussagen darüber, welche Klassifikationen innerhalb von Diskursen und durch sie vorgenommen werden. Aber wie werden soziale Phänomene durch Klassifikationen konstituiert? Gerade für machtanalytisch angelegte Fragestellungen ist die Auseinandersetzung mit durch einen Diskurs produzierten Klassifikationsschemata wichtig, da institutionelle Klassifikationsschemata Bewertungen und Zuschreibungen enthalten (Pfahl/Traue 2012). Für Untersuchungen, die sich mit den Wirkungen von Diskursen auf Handelnde befassen (z.B. Subjektivierungsanalysen), kommt der Frage zentrale Bedeutung zu, wie Klassifikationen in den Selbstbeschreibungen aufgegriffen werden und welche identitätsstiftende Wirkung diese haben.
- Einen weiteren Aspekt stellen die *narrativen Strukturen* eines Diskurses dar. Viehöver (2011) betrachtet Diskurse als Narrationen (Küsters, Kapitel 40 in diesem Band) und nutzt narrationsanalytische Verfahrensschritte der Textanalyse. Im Zentrum steht hierbei die Rekonstruktion der Erzählstrukturen von Diskursen: Wie werden die Elemente eines Diskurses, also die Deutungsmuster oder Klassifikationen, durch einen dramaturgischen Plot miteinander in Beziehung gesetzt und zu einer Geschichte mit zeitlichen, räumlichen Strukturen und mit einem Personal versehen? Kommt es zu Moralisierungen und der Adaption moralischer Schemata innerhalb der Narration?
- Die Analyse der *Sprecherpositionen* bietet Aufschluss über die institutionelle Verfasstheit eines Diskurses. Dementsprechend werden die Zugangs- und Teilnahmebedingungen am diskursiven Geschehen untersucht und bestimmt, welche legitimen Sprecherpositionen ein Diskurs bereithält (Hark 2011). Wer wird im Diskurs als Experte gehört und wer wird ausgeschlossen? Welche Zuschreibungen werden an unterschiedliche Sprecherpositionen innerhalb eines Diskurses gemacht?
- Nicht zuletzt bildet der Aspekt der *nicht-diskursiven Praktiken* einen wichtigen Bestandteil der Diskursanalyse. Zu fragen ist in diesem Zusammenhang, welche nicht-sprachlichen Objektivationen ein Diskurs hat und wie seine institutionelle und technische Infrastruktur beschaffen sind (Traue 2013). Diese Fragen verweisen stärker als die zuvor genannten auf die machtanalytische Ausrichtung von Diskursanalysen.

## 34.5 Diskursanalyse als Machtanalyse

Abschließend gehen wir auf die Art und Weise ein, wie im Rahmen der Diskursanalyse Machtverhältnisse analysiert werden. Hier bestehen in der Forschungspraxis die meisten Missverständnisse und Qualitätsprobleme, da die verschiedenen Analyseebenen miteinander verknüpft werden müssen, um die Funktionsweise eines Diskurses (Abschnitt 34.5.1) beschreiben zu können. Wie wir im Folgenden erläutern, müssen die Machtwirkungen jeweils mühsam *empirisch* herausgearbeitet werden, sie sind integraler Bestandteil von Diskursanalysen, und können nicht – wie es oft geschieht – unterstellt werden; Diskursanalysen, in denen eine Kritik materieller Verhältnisse als absoluter Ausgangspunkt gesetzt wird, verletzen dieses Prinzip. Anschließend gilt es, die rekonstruierte innerdiskursive Ordnung mit der Untersuchung des Außen des Diskurses in Zusammenhang zu setzen (Abschnitt 34.5.2).

Die sozialwissenschaftliche Diskursanalyse geht im Anschluss an Foucault von einer wechselseitigen Beziehung zwischen gesellschaftlichen Herrschaftsverhältnissen und den Wissens- und Wirklichkeitskonstruktionen einer Gesellschaft aus. In Diskursanalysen wird also weder ein („materialistisches“) Primat der Macht noch ein („idealstisches“) Primat des Wissens vorausgesetzt, sondern die *Beziehung zwischen Wissen und Macht* wird als eine zirkuläre Beziehung verstanden. Die sozialen Verhältnisse, innerhalb derer Wissensproduktionen situiert sind, bleiben dem Wissen (und der Wissenschaft) und seinen Gegenstandskonstruktionen nicht äußerlich – wie etwa noch Scheler (1980) mit seiner Unterscheidung in Realfaktoren und Idealfaktoren annahm –, sondern sie formen die Begriffe und die diskursiv erzeugten Wirklichkeits- und Wahrheitskonstruktionen mit. Gesellschaftliche Kräfteverhältnisse werden ihrerseits von Wissensformen „legitimiert“ (Berger/Luckmann 1969), delegitimiert bzw. modifiziert.

### 34.5.1 „Funktionsweise“ von Diskursen

Die Diskursforschung zielt dementsprechend in der Regel darauf, die Machtverhältnisse, wie sie sich in den Diskursen ausdrücken, als gesellschaftliche Machtverhältnisse zu rekonstruieren und dadurch den „Naturalismus“ des Wissens zu dekonstruieren. Sozialwissenschaftliche Diskursanalysen gehen damit über Text- oder Inhaltsanalysen (Mayring/Fenzl, Kapitel 38 in diesem Band) hinaus. Sie sind neben ihrem Status als interpretative Analytik konkreter Wissensbestände immer auch Institutionen- bzw. Gesellschaftsanalysen. In dieser Hinsicht schließen sie – insbesondere in ihren deutschsprachigen Ausprägungen – an „klassische“ wissenssoziologische Konzeptualisierungen von materiellen Verhältnissen und Wissensformen an (Scheler 1980).

Für die Analyse bedeutet das, dass – nachdem die Elemente eines Diskurses und die inneren Beziehungen seiner Zeichen und Formen (Abschnitt 34.4.1) identifiziert wurden – im Anschluss erklärt werden muss, mit Hilfe welcher Machtmechanismen der Diskurs stabilisiert wird. Eine Machtanalyse beruht damit nicht allein auf einer Beschreibung der

Diskursformation, sondern fragt nach ihrer Funktion – ganz ähnlich wie auch Mannheim (1982) von einer „Funktionalisierung“ von Wissensformen spricht.

Für eine solche Beschreibung der Funktionsweise von Diskursen ist es notwendig, im Rahmen von Diskursanalysen die rein sprachliche Analyse zu überschreiten: Erst wenn eine Verbindung der sprachlichen Formen mit nichtsprachlichen Objektivierungen des Wissens nachgewiesen werden kann, kann das methodische Potential einer Diskursanalyse als ausgeschöpft gelten. Zu letzteren zählen Sozialstrukturen, Organisationen bzw. Institutionen, Materialitäten (Dinge, Architekturen, Medien), Rechtsverhältnisse und Sprecherpositionen, Handlungsvollzüge individueller wie kollektiver Akteure. Machtanalyse und Wissensanalyse werden dabei im Forschungsprozess immer wieder versuchsweise und zunehmend systematisch aufeinander bezogen, bis ihre wechselseitige Stabilisierung und Transformation in der Darstellung nachvollziehbar gemacht werden kann.

### 34.5.2 Ebenen der Machtanalyse

Der im Zuge der Popularisierung der Diskursanalyse häufig vorgebrachte Slogan von der „Macht der Diskurse“ hat der Diskursanalyse den Vorwurf eingebracht, sie sehe als modische Methodologie die Macht – beinahe paranoisch – überall, weshalb sie zur analytischen Unschärfe verführe. Diese Einschätzung beruht allerdings auf einem mangelnden Verständnis ihrer Vorgehensweisen und Analyseebenen. Analytisch müssen in der sozialwissenschaftlichen Diskursanalyse *zwei Ebenen* unterschieden werden: erstens die Ebene der innerdiskursive Ordnung und zweitens die Ebene der funktionalen Beziehungen eines Diskurses (mit seiner Ordnung von Sprecherpositionen) zu seinem Außen, d.h. zu anderen Diskursen und zur allgemeineren historisch – gesellschaftlichen Situation.

Diese Unterscheidung bietet gegenüber der älteren, von Karl Mannheim vertretenen „funktionalen Analyse“ (Mannheim 1982) den Vorteil, kein abstraktes Verständnis eines historischen Prozesses als Referenzrahmen für das Auftreten eines Wissens zu unterstellen, sondern zwischen den *lokalen Formationsregeln* eines Diskurses und seinen *inter- und außerdiskursiven Anschlussmöglichkeiten* unterscheiden zu können.

Nur eine vorläufige Trennung beider Analyseebenen ermöglicht es, sie in einem letzten Schritt wieder aufeinander zu beziehen und die Frage zu stellen, in welches Verhältnis ein Diskurs als Wissensform zu anderen Wissensformen gesetzt ist, wie außerdiskursive Wirklichkeiten gestützt, destabilisiert, transformiert oder zerstört werden, und welche Akteure mit ihren Interessen, die teils inner-, teils außerdiskursiven Ursprung haben, sich mit eigenen Absichten in dieses Geschehen einschalten.

Auf der ersten Ebene der *innerdiskursiven Ordnung* steht zur Frage, wie eine diskursive Praxis in sich selbst auf Dauer gestellt wird, um zu einer „Wissensordnung“ (Schütz/Luckmann 1984) zu werden. Zu ihrer Beantwortung wird es nötig, die *Formen und Techniken der „Veridiktion“* (Foucault 2009), also der Generierung legitimer Aussagen innerhalb einer spezifischen diskursiven Praxis, zu bestimmen. Diese Veridiktion entsteht dabei allerdings meist nicht isoliert, sondern im Hinblick auf eine öffentliche Geltung von Aussagen.

„Welches Wissen (...) kann mittels welcher Prozeduren in welchen gesellschaftlichen Handlungsfeldern und mit welchen praktischen Konsequenzen für wen Gültigkeit erlangen?“ (Hirselund/Schneider 2011: 408). Im Detail kann gefragt werden, welche Sprecherpositionen und Subjektpositionen vorzufinden sind, welche Verknappungen und Anregungen vorliegen und welche Begriffe und Gegenstände dominant, marginalisiert oder tabuisiert sind.

Auf der zweiten Ebene, der *funktionalen Beziehungen eines Diskurses zu einem Außen*, ist es wichtig, die *Grenzen des Diskurses und ihre Funktion für andere Akteure und Diskurse* zu bestimmen. Wie wird ein Außen des Diskurses konstruiert? Welche Techniken kommen dabei zum Einsatz? In welche strukturellen, institutionellen Rahmenbedingungen ist diese Wissensproduktion eingebettet und welche nicht-diskursiven Praktiken gehen damit einher? Stehen die materialen Einrichtungen und institutionellen Praktiken im Vordergrund der Diskursanalyse, wird auch von „Dispositivanalysen“ gesprochen (Keller 2005; Schneider/Bührmann 2012); wenn die Aneignungen von Sozial- und Selbsttechniken durch Subjekte erklärt werden soll, von Subjektivierungsanalysen (vgl. Pfahl/Traue 2012, Pfahl/Schürmann/Traue 2014).

## Literatur

- Angermüller, Johannes (2007): Nach dem Strukturalismus. Theoriediskurs und intellektuelles Feld in Frankreich. Bielefeld: Transcript
- Berger, Peter/Luckmann, Thomas (1969): Die gesellschaftliche Konstruktion der Wirklichkeit. Frankfurt a.M.: Fischer
- Bergmann, Jörg (1981): Ethnomethodologische Konversationsanalyse. In: Schröder, Peter/Steger, Hugo (Hg.): Dialogforschung. Düsseldorf: Schwann, 9–51
- Bröckling, Ulrich/Krasmann, Susanne/Lemke, Thomas (2000): Gouvernementalität der Gegenwart. Frankfurt a.M.: Suhrkamp
- Bröckling, Ulrich (2007): Das Unternehmerische Selbst. Frankfurt a.M.: Suhrkamp
- Bublitz, Hannelore/Hanke, Christine/Seier, Andrea (2000): Der Gesellschaftskörper. Frankfurt a.M.: Campus
- Diaz-Bone, Rainer (2003): Entwicklungen im Feld der foucaultschen Diskursanalyse. In: FQS 4 (3): Art.1. <http://www.qualitative-research.net/index.php/fqs/article/view/683> (Zugriff am 2.2.2014)
- Douglas, Mary (1991): Wie Institutionen denken. Frankfurt a.M.: Suhrkamp
- Engel, Antke (2009): Bilder von Sexualität und Ökonomie. Bielefeld: transcript
- Fairclough, Norman/Wodack, Ruth (1997): Critical Discourse Analysis. In: Van Dijk, Teun A. (Hg.): Discourse Studies: A Multidisciplinary Introduction. Vol. 2. London: Sage, 258–284
- Foucault, Michel (1981): Archäologie des Wissens. Frankfurt a.M.: Suhrkamp
- Foucault, Michel (1989): Der Wille zum Wissen. Sexualität und Wahrheit I. Frankfurt a.M.: Suhrkamp
- Foucault, Michel (1991): Die Ordnung des Diskurses. Frankfurt a.M.: Fischer
- Foucault, Michel (2009): Die Regierung des Selbst und der anderen. Frankfurt a.M.: Suhrkamp
- Freitag, Walburga (2005): Contergan. Eine genealogische Studie des Zusammenhangs wissenschaftlicher Diskurse und biographischer Erfahrungen. Münster: Waxmann
- Hark, Sabine (2011): Feministische Theorie – Diskurs – Dekonstruktion. In: Keller et al. (Hg.): 381–400
- Hirsland, Andreas/Schneider, Werner (2011): Wahrheit, Ideologie und Diskurse. In: Keller et al. (Hg.): 401–432
- Jäger, Siegfried (2004): Kritische Diskursanalyse. Münster: Unrast
- Keller, Reiner (2004) Diskursforschung. Wiesbaden: VS Verlag
- Keller, Reiner (2005): Wissenssoziologische Diskursanalyse. Wiesbaden: VS Verlag
- Keller, Reiner/Hirsland, Andreas/Schneider, Werner/Viehöver, Willy (2011): Handbuch sozialwissenschaftliche Diskursanalyse. Bd. 1: Theorien und Methoden. Wiesbaden: VS Verlag
- Keller, Reiner/Knoblauch, Hubert/Reichertz, Jo (2012): Kommunikativer Konstruktivismus. Wiesbaden: VS Verlag

- Küçük, Bülent (2008): Die Türkei und das andere Europa. Phantasmen der Identität im Beitrittsdiskurs. Bielefeld: Transcript
- Laclau, Ernesto/Mouffe, Chantal (2001): Hegemony and Socialist Strategy – Towards a Radical Democratic Politics. London/New York: Verso
- Maasen, Sabine/Mayerhauser, Torsten/Renggli, Cornelia (2006): Bilder als Diskurse. Weilerswist: Velbrück
- Mannheim, Karl (1982) [1922]: Ideologische und soziologische Interpretationen der geistigen Gebilde. In: Meja, Volker/Stehr, Nico (Hg.): Der Streit um die Wissenssoziologie. Frankfurt a.M.: Suhrkamp. 213-232
- Pfahl, Lisa (2011): Techniken der Behinderung. Der deutsche Lernbehinderungsdiskurs, die Sonderschule und ihre Auswirkungen auf Bildungsbiografien. Bielefeld: Transcript
- Pfahl, Lisa/Schürmann, Lena/ Traue, Boris (2014): Das Fleisch der Diskurse. Zur Verbindung von Biographie- und Diskursforschung in der wissenssoziologischen Subjektivierungsanalyse. In: Susann Fegter et al. (Hg.): Diskursanalytische Zugänge zu Bildungs- und Erziehungsverhältnissen. Wiesbaden: VS. (im Druck)
- Pfahl, Lisa/Traue, Boris (2012): Die Erfahrung des Diskurses. In: Keller, Reiner/Truschkat, Inga (Hg.): Methodologie und Praxis der wissenssoziologischen Diskursanalyse. Wiesbaden: VS Verlag. 425-450
- Reisigl, Martin (2011): Grundzüge der Wiener Kritischen Diskursanalyse. In: Keller et al. (Hg.): 459-498
- Reisigl, Martin/Angermüller, Johannes/Nonhoff, Martin/Wrana, Daniel (2014): DiskursNetz. Frankfurt a.M.: Suhrkamp
- Saar, Martin (2007): Genealogie als Kritik. Frankfurt a.M.: Campus
- Scheler, Max (1980): Die Wissensformen und die Gesellschaft. Bern: Francke
- Schneider, Werner/Bührmann, Andrea (2012): Vom Diskurs zum Dispositiv. Bielefeld: Transcript
- Schürmann, Lena (2013): Schmutz als Beruf. Prekarisierung, Klasse, Geschlecht. Münster: Westfälisches Dampfboot
- Schütz, Alfred/Luckmann, Thomas (1984): Die Strukturen der Lebenswelt. Frankfurt a.M.: Suhrkamp
- Traue, Boris (2010): Das Subjekt der Beratung. Zur Soziologie einer Psychotechnik. Bielefeld: Transcript
- Traue, Boris (2013): Visuelle Diskursanalyse. In: Zeitschrift für Diskursforschung 1 (2): 117-136
- Viehöver, Willy (2011): Diskurse als Narrationen. In: Keller et al. (Hg.): 193-224

Gabriele Rosenthal

---

## 35.1 Biographie und biographische Thematisierungen

Der Begriff „*Biographie*“ (Griechisch: „Leben“ & „schreiben“) bezieht sich nicht nur auf Geschriebenes, sondern ebenso auf in Gesprächen mitgeteilte biographische Selbst- oder Fremdbeschreibungen. Biographische Beschreibungen werden sowohl in informellen Zusammenhängen, als auch in unterschiedlichen formal organisierten oder durch Institutionen regulierten Kontexten mündlich geäußert oder schriftlich verfasst.

Dies erfolgt zum Beispiel in Anhörungen und Aktenvermerken bei Behörden, in Gerichtsverfahren, in Bewerbungsgesprächen, im ärztlichen Setting, in religiösen Handlungszusammenhängen, in der massenmedialen Kommunikation, im Kontext Sozialer Arbeit oder im Rahmen von Konfliktmanagement und Friedensförderung sowie in vielen weiteren sozialen Feldern. Je nachdem, ob jemand gesteuert durch bestimmte Vorgaben und Fragen seine Biographie als „traurige Geschichte“ zur Erreichung bestimmter Ziele präsentiert oder in einer Freundschaftsbeziehung über in der Vergangenheit Erlittenes spricht, fällt die biographische Selbstpräsentation sehr unterschiedlich aus. Diese unterschiedlichen Kontexte haben ihrerseits unterschiedliche Effekte: sowohl für biographische Thematisierungen der Zukunft, als auch für die Konstruktion der Vergangenheit.

Die *Settings*, in denen biographische Thematisierungen eingefordert oder sozial erwünscht oder gerade nicht erwünscht sind, unterscheiden sich ebenso wie die *sozialen Regeln*, welche Lebensbereiche und -phasen wie thematisiert werden dürfen oder sollten und welche nicht, von Kultur zu Kultur, von Milieu zu Milieu. So haben zum Beispiel MigrantInnen, die in unterschiedlichen Kulturen und politischen Verbänden (bspw. Staaten) sozialisiert sind, zum Teil sehr divergente soziale Regeln für biographische Thematisierungen erlernt und internalisiert. Gesellschaftliche, institutionelle und familiale Regeln bzw. die Regeln unterschiedlicher Diskurse geben vor, *was* und *was nicht* sowie *wie*, *wann* und *in welchen Kontexten* etwas thematisiert werden darf.

Die in unterschiedlichen Situationen hergestellten biographischen Selbst- wie auch Fremdthematisierungen führen dazu, dass Menschen im Laufe ihres Lebens *biographische Konstruktionen* ausbilden, die ihnen dazu verhelfen, ihr Leben zu deuten, d.h. ihrer Vergangenheit, Gegenwart und anvisierten Zukunft eine Bedeutung zu geben und damit eine Orientierung in ihren Handlungs- wie Lebensentwürfen zu gewinnen. Ähnliche Funktionen haben diese Konstruktionen in den Prozessen des Fremdverstehens (Knoblauch, Kapitel 36 in diesem Band), in denen versucht wird, den lebensgeschichtlichen Prozess des So-Gewordenseins bei Anderen nachzuvollziehen.

Diese vom einzelnen Biographen in sozialen Interaktionen immer wieder neu zu leistende Konstruktionsarbeit erfolgt mit Hilfe von *gesellschaftlich* (also durch ein Kollektiv oder vielmehr durch mehrere Kollektive) *vorgegebenen, teilweise institutionalisierten und teilweise im Lauf der Sozialisation internalisierten Mustern*. Weder diese Muster noch die mit ihrer Hilfe vollzogene Konstruktionsarbeit sind dabei als ein (nur) individuelles Geschehen zu begreifen. Diese Konstruktionen können sich zwar im Laufe des Lebens verfestigen, jedoch werden sie auch immer wieder modifiziert und vor allem erfordern sie in veränderten und krisenhaften Lebenssituationen entsprechend der veränderten Gegenwart und Zukunft eine Reinterpretation der Vergangenheit. Die Biographie des jeweiligen Individuums entsteht somit jeweils immer wieder neu in den Gegenwarten des Erzählens oder Schreibens.

---

## 35.2 Ziele und Methodologie soziologischer Biographieforschung

Eine von dem US-amerikanischen Soziologen William Isaac Thomas und dem polnischen Soziologen Znaniecki (1918-1920/1958) durchgeführte Untersuchung zu polnischen MigrantInnen in den USA markiert den Beginn der soziologischen Biographieforschung. Neben Dokumenten zum Migrationsprozess analysierten sie hauptsächlich aus Polen oder in die Heimat gesandte Briefe (Ernst, Kapitel 63 in diesem Band) und eine auf ihre Bestellung geschriebene Biographie eines Migranten. Thomas und Znaniecki vertraten die Forderung, dass die sozialen Probleme von Einwanderern nicht nur aus der Perspektive der „objektiven“ Rahmenbedingungen, sondern auch auf der Basis der „subjektiven“ Sichtweisen der MigrantInnen sowie ihrer lebensgeschichtlichen Erfahrungen – vor und nach der Einwanderung – zu untersuchen seien. Für die gegenwärtige Biographieforschung zu den unterschiedlichsten Gruppierungen von Menschen ist diese Forderung immer noch forschungsleitend.

Mittlerweile hat sich der biographietheoretische Ansatz national wie auch international insbesondere in der Soziologie (Fuchs-Heinritz 2005; Rosenthal 1995; Völter et. al. 2005) und in den Erziehungswissenschaften (Krüger/Marotzki 1999) als eine Teildisziplin mit einer genuin eigenen Theoriegrundlage und Methodologie etabliert, die in der Bundesrepublik vor allem im Ansatz des Sozialkonstruktivismus verankert ist, der durch Berger und Luckmann formuliert wurde. Damit verbindet sich die methodologische und forschungspraktische Forderung, die „subjektiven“ Erfahrungen und Perspektiven der

Alltagshandelnden als Ausgangspunkt der Analyse zu nehmen und deren lebensgeschichtliche Genese zu rekonstruieren.

Zentrales Anliegen der soziologischen Biographieforschung ist es dabei, der *gegenseitigen Konstitution von Individuen und Gesellschaften* gerecht zu werden. Lebensgeschichtliche und kollektivgeschichtliche Prozesse werden in ihren „Wechselwirkungen“ und unhintergehbaren Verflechtungen empirisch untersucht. Biographie wird also nicht als etwas rein Individuelles oder bloß Subjektives, sondern als ein soziales Konstrukt verstanden, das auf kollektive Regeln, Diskurse (Traue et al., Kapitel 34 in diesem Band) und gesellschaftliche Rahmenbedingungen verweist und sowohl in seiner Entwicklung als auch im deutenden Rückblick immer beides zugleich ist: ein individuelles und ein kollektives Produkt. Mit einem biographietheoretischen Ansatz sind neben dem von den Biographien und Handlungsgeschichten von Individuen ausgehenden Versuch, diesen „Wechselwirkungen“ gerecht zu werden, noch zwei weitere Prämissen verbunden.

Zum einen ist dies die Forderung, die Bedeutungen von Erfahrungen *nicht isoliert, sondern im Gesamtzusammenhang der Lebensgeschichte* zu interpretieren, und zum anderen der Anspruch einer *Prozessanalyse, die den historischen Verlauf der Entstehung, Aufrechthaltung und Veränderung von sozialen Phänomenen im Kontext der Erforschung von Lebensverläufen rekonstruiert*. Das methodische Postulat einer konsequenten Einbettung der einzelnen Biographie in ihren geschichtlichen und gesellschaftlichen Kontext wurde in den letzten 20 Jahren vor allem durch die Wende zu Mehrgenerationenanalysen in der Biographieforschung zunehmend umgesetzt (Bertaux/Bertaux-Wiame 1991; Hildenbrand 1999; Rosenthal 2005). Ein auf mehrere Generationen ausgedehnter Zeithorizont der Untersuchung ermöglicht es, die *Genese sozialer einschließlich kollektiver Phänomene* wie der kulturellen Praxis des Erinnerns über einen längeren Zeitraum hin zu erforschen. Angestrebt wird eine Rekonstruktion der individuellen Erfahrungsgeschichte in ihrer „Wechselbeziehung“ mit dem langfristigen Wandel der gesellschaftlich-kulturellen Rahmenbedingungen und vice versa dieses Wandels im Zusammenhang mit den Veränderungen biographischer Verläufe und (Selbst-)Thematisierungen.

Der Einsatz der Methoden biographischer Forschung hat sich in sehr unterschiedliche Untersuchungsfelder ausdifferenziert. Zum einen kann es um die *Analyse des gelebten Lebens bzw. spezifischer Lebensbereiche oder -phasen von bestimmten Personengruppen oder gesellschaftlichen Gruppierungen in bestimmten historischen Zeiträumen* gehen (wie z.B. von MigrantInnen in der Untersuchung von Thomas & Znaniecki). Zum anderen geht es um die *Rekonstruktion bestimmter sozialer Settings aus der Perspektive der Handelnden in spezifischen historischen Epochen und soziokulturellen Kontexten* (z.B. eine Milieustudie über einen sozialen Brennpunkt in einer Großstadt). Ein weiteres Ziel ist die *Analyse biographischer Selbst- und Fremdthematisierungen in sozialen Interaktionen* (z.B. biographische Thematisierungen von KlientInnen auf Ämtern). Für die gegenwärtige soziologische Biographieforschung sind weitere wichtige Anliegen die *Analyse der biographischen Konstruktionen und der biographischen Selbstpräsentation in der Gegenwart* (z.B.: Was sind die Regeln biographischer Selbstthematisierungen von Überlebenden kollektiver Gewalt aus Bosnien oder von ehemaligen Psychiatriepatienten?) und damit verbunden die

---

*Rekonstruktion der Genese und Transformationen dieser Konstruktionen*, wobei auch der Frage nachgegangen wird, inwiefern biographische Thematisierungen in der Vergangenheit innerhalb bestimmter Settings (wie z.B. im Kontext eines Asylverfahrens oder eines Gesprächs in der Psychiatrie) einen nachhaltigen Einfluss auf die Konstruktionen in anderen Situationen und sozialen Kontexten haben.

---

### **35.3 Methoden der Biographieforschung**

Aus den bisher diskutierten Anliegen der Biographieforschung folgen bestimmte Erfordernisse an die Untersuchungsmethoden. Im Unterschied zu anderen qualitativen Methoden sollen nicht nur die *Deutungen der befragten Personen in der Gegenwart* der Befragung offengelegt werden, sondern es wird gezielt versucht, einen Einblick in die *Genese dieser Deutungen* und in die *sequenzielle Gestalt der erlebten Lebensgeschichte* zu erhalten sowie die *Rekonstruktion von Handlungsabläufen in der Vergangenheit und des damaligen Erlebens* zu ermöglichen. Das biographisch-narrative Interview (Küsters, Kapitel 40 in diesem Band) wird diesen Erfordernissen in besonderem Maße gerecht und das von der Autorin vorgestellte Auswertungsverfahren biographischer Fallrekonstruktionen ist vor allem dem Anliegen geschuldet, in analytisch getrennten Auswertungsschritten sowohl die Gegenwartsperspektive des Handelnden als auch dessen Perspektiven in der Vergangenheit vor dem Hintergrund der gesamtgesellschaftlichen und historischen Kontexte zu rekonstruieren.

---

### **35.4 Das biographisch-narrative Interview**

Das narrative Interview (Küsters, Kapitel 40 in diesem Band) basiert auf der Annahme, dass selbsterlebte Ereignisse sich am überzeugendsten in Erzählungen darstellen lassen (Schütze 1977). Lässt sich der Befragte auf eine Stegreiferzählung und einen Prozess des Erzählens von Geschichte zu Geschichte und damit auch auf einen Erinnerungsfluss ein, ermöglicht dies mehr *Nähe* zur damals *erlebten* – nicht einer vermeintlich wirklichen – Vergangenheit als bei anderen Formen der Selbstpräsentation (wie z.B. dem kondensierten Berichten, dem Argumentieren über Bestandteile der Vergangenheit oder Belegerzählungen), die stärker durch die Gegenwartsperspektive und die jetzige Interaktion mit den Zuhörern geprägt werden.

Damit ist jedoch nicht gemeint, dass die Erzählung sich jemals mit dem Erleben in der Vergangenheit decken könnte, sondern nur, dass die Erzählung eine größere Annäherung an den Handlungsablauf und das Erleben in der damaligen Situation ermöglicht als andere Formen der sprachlichen Darstellung. Vor allem ist es möglich, dass in dem durch das Erzählen geforderten oder induzierten Erinnerungsvorgang *Eindrücke, Gefühle, sinnliche und leibliche Empfindungen oder bisher zurückgedrängte Komponenten der erinnerten Situationen* vorstellig werden, an die man sich lange nicht mehr erinnert oder über

die man noch nie gesprochen hat. Dazu können auch solche gehören, die mit der Gegenwartsperspektive nicht kompatibel sind oder nicht dem jetzigen Präsentationsinteresse des Biographen und den Regeln der gegenwärtig wirksamen kollektiven Diskurse (Traue et al., Kapitel 34 in diesem Band) entsprechen.

Dieses Phänomen der durch den Erzähl- und Erinnerungsvorgang ausgelösten zunehmenden Nähe des Autobiographen (Ernst, Kapitel 63 in diesem Band) zur erlebten Vergangenheit nutzen wir in der Biographieforschung mit der zuerst von Schütze entwickelten Technik des *biographisch-narrativen Interviews* (Küsters, Kapitel 40 in diesem Band), die mittlerweile in sehr unterschiedlichen thematischen, aber auch unterschiedlichen geografischen Kontexten erprobt und weiterentwickelt wurde (Rosenthal 2011: 151ff.). Gezielt wird bei einer narrativen Gesprächsführung versucht, die Befragten zu längeren biographischen Erzählungen oder, allgemeiner formuliert, zu autonom gestalteten Präsentationen ihrer Lebensgeschichte aufzufordern und dabei zu unterstützen. Schütze (1983), der dieses Instrument im Kontext einer Untersuchung von Machtstrukturen in Gemeinden in den 1970er Jahren entwickelt hatte, führte es in späteren Jahren in die Biographieforschung ein und schlug vor, unabhängig von dem thematischen Schwerpunkt der empirischen Untersuchung zur Erzählung der gesamten Lebensgeschichte aufzufordern. So wird es möglich, einzelne Lebensbereiche oder -phasen im Gesamtzusammenhang des Lebens betrachten zu können.

Die ForscherInnen orientieren sich zunächst an den Relevanzen der Befragten und deren alltagsweltlichen Konstruktionen. Aus diesem Anspruch ergibt sich ein Gesprächsverlauf, bei dem erst in der dritten Phase des Interviews die Relevanzen der WissenschaftlerInnen für die Gesprächsführung dominant werden:

- In der ersten Phase des Interviews wird den Befragten die Möglichkeit gegeben, ohne Zwischenfragen ihre Familien- und Lebensgeschichte oder bestimmte Phasen und Bereiche ihres Lebens zu erzählen bzw. darzustellen. Diese so genannte *Eingangserzählung* wird nicht durch Detaillierungsfragen unterbrochen, sondern sie wird nur durch parasprachliche Bekundungen wie „mhm“, durch Blickkontakt und andere leibliche Aufmerksamkeitsbekundungen oder (bei Stockungen in der Erzählung) durch motivierende Aufforderungen zum Weitererzählen wie „Und wie ging es dann weiter?“ unterstützt.
- Erst in der zweiten Phase des Gesprächs, nach erkennbarer Beendigung der Haupterzählung durch die Befragten, werden Fragen gestellt. Zunächst werden nur *erzählinterne Nachfragen* gestellt, d.h. Fragen zu bereits Erwähntem, die auf den während der Haupterzählung gemachten protokollförmigen Notizen beruhen. In diesem ersten Nachfrageteil des Interviews werden keine Themenbereiche angesteuert, die nicht schon von den Erzählenden selbst ins Gespräch eingeführt wurden.
- Erst in der dritten Phase des Interviews werden *erzählexterne Fragen* zu bisher nicht erwähnten Themen gestellt (die die ForscherInnen interessieren).

Dank dieser Abfolge von erzählinternen und erzählexternen Nachfragen und der zeitlichen Trennung zwischen den entsprechenden Interviewphasen können wir bei der

Auswertung der Frage nachgehen, welche Themen an welcher Stelle im Gespräch von den Befragten selbst eingeführt wurden, welche Themen vermieden wurden und was die Bedeutung von Thematisierung wie Nichtthematisierung sein kann. Unter anderem um die Bedeutung von Auslassungen erschließen zu können, bedarf es jedoch einer Phase mit erzählexternen Nachfragen.

---

### 35.5 Biographische Fallrekonstruktionen

Die für diese Auswertung am häufigsten verwendete Datengrundlage sind biographisch-narrative Interviews (Küsters, Kapitel 40 in diesem Band), häufig auch in Verbindung mit anderen biographischen Texten. Ebenso wird mit niedergeschriebenen Autobiographien, biographischen Thematisierungen in Alltags- oder Organisationskontexten, (familien-)biographischen Dokumenten (Fotoalben, Tagebücher, Briefe etc.) (Ernst, Kapitel 63 in diesem Band) und personenbezogenen Akten (Lebensläufe in Gerichtsverfahren, Personalakten in Parteien, Anamneseakten etc.) (Salheister, Kapitel 62 in diesem Band) sowie mit der Kombination dieser Materialien gearbeitet.

Die von Schütze (1983) vorgeschlagene textanalytische Auswertung biographischer Interviews kann – wie das von der Autorin vorgeschlagene Verfahren – in modifizierter Form auch auf andere Texte angewendet werden. Bei der *Textanalyse nach Schütze* sind die Unterscheidung nach Textsorten (Erzählung, Argumentation und Beschreibung) und die Analyse ihrer Funktion bedeutsam. Dieses Verfahren wurde von einigen AutorInnen abgewandelt und mit anderen interpretativen Verfahren kombiniert, insbesondere mit der von Oevermann entwickelten „Objektiven Hermeneutik“ (Hildenbrand 1995; Wohlrab-Sahr 1998; Rosenthal 1995).

Gemeinsam ist diesen Verfahren ihr rekonstruktives und sequenzielles Vorgehen. Mit „*rekonstruktiv*“ ist gemeint, dass einzelne Segmente eines Textes nicht unter Kategorien subsumiert werden, sondern die Bedeutung dieser Teile im konkreten Gesamtzusammenhang eines bestimmten Textes erschlossen wird. Dies heißt auch, die Rolle (bzw. „Funktion“) und Position dieser Teile im Prozess der interaktiven Konstitution des Textes in der Gegenwart des Sprechens herauszuarbeiten. Unter einem „*sequenziellen Vorgehen*“ (Kurt/Herbrik, Kapitel 33 in diesem Band) wird ein Vorgehen verstanden, bei dem der Text bzw. kleinere Texteinheiten entsprechend ihrer sequenziellen Gestalt, also in der Ordnung der Abfolge ihres Entstehens, interpretiert werden.

Das von der Autorin vorgestellte *Verfahren biographischer Fallrekonstruktionen* (Rosenthal 1995; 2011) orientiert sich an der Objektiven Hermeneutik in einer Verknüpfung mit dem textanalytischen Verfahren nach Schütze und der phänomenologisch-gestalttheoretischen Feldanalyse nach Gurwitsch (1974). Eine sequenzielle Auswertung erfolgt hierbei hinsichtlich der zeitlichen Struktur sowohl der *erlebten* als auch der *erzählten bzw. präsentierten Lebensgeschichte*, die auch in erster Linie in Argumentationen oder Beschreibungen dargestellt sein kann.

Einerseits wird versucht, die Chronologie der biographischen Erfahrungen im Lebensverlauf und deren Bedeutungen für den Biographen zu rekonstruieren. Andererseits wird die zeitliche Struktur der Lebenserzählung analysiert, d.h. der Frage nachgegangen, in welcher Reihenfolge, in welcher Ausführlichkeit und in welcher Textsorte die BiographInnen ihre Erfahrungen im Kontext der Textproduktion (in einem Interview oder auch in einem anderen Rahmen, wie z.B. einem Familiengespräch oder einer niedergeschriebenen Biographie) präsentieren. Bei diesem Analyseschritt, der sich auf die aktuelle Präsentation der Lebensgeschichte konzentriert, wird vor allem der Interaktionsverlauf zwischen Befragten und Zuhörern bzw. zwischen Autobiographen und imaginierten Lesern rekonstruiert.

Entscheidend bei dieser Auswertungsmethode ist es also, zwischen erzählter und erlebter Lebensgeschichte zu differenzieren und den beständigen Wandel von Bedeutungen im Lebensverlauf zu berücksichtigen. Um dies zu ermöglichen, ist es erforderlich, zunächst in getrennten Analyseschritten den beiden verflochtenen, aber dennoch verschiedenen Phänomenen der erzählten und der erlebten Lebensgeschichte nachzugehen. Das heißt: Ziel der Rekonstruktion ist sowohl die biographische Bedeutung des in der Vergangenheit Erlebten, als auch die Bedeutung der Selbstpräsentation in der Gegenwart. Wird bei der Rekonstruktion der Fallgeschichte nach der biographischen Bedeutung einer Erfahrung zur damaligen Zeit gefragt, so stellt sich bei der Rekonstruktion der Lebenserzählung die Frage nach der Funktion der jetztzeitigen Darstellung des Erlebens für die interviewte Person in ihrem gegenwärtigen sozialen Kontext. Die Autorin geht davon aus, dass man damit methodisch kontrolliert den *Wechselwirkungen zwischen Vergangenem, Gegenwärtigem und Zukünftigem* gerecht werden kann (Rosenthal 1995).

Bei der *Auswertung der erzählten Lebensgeschichte* bzw. der in der Gegenwart des Gesprächs präsentierten Biographie wird deshalb gezielt danach gefragt, inwiefern sich die in der Gegenwart des Sprechens erinnerte Vergangenheit durch die Gegenwartsperspektive, die institutionelle und interaktive Rahmung des Gesprächs sowie den aktuellen Zukunftshorizont des Biographen konstituiert, aber auch, wie die erlebte Vergangenheit die Gegenwarts- und Zukunftsperspektive mitbestimmt.

Beim *kontrastiven Vergleich der erzählten mit der erlebten Lebensgeschichte* geht es zum Abschluss der Fallrekonstruktion um mögliche Erklärungen für die Differenz zwischen diesen beiden Ebenen, d.h. zwischen der Vergangenheits- und der Gegenwartsperspektive. So stellt sich zum Beispiel die Frage, welche Bedeutung in der Gegenwart des Erzählers es für eine Migrantin aus der ehemaligen Sowjetunion hat, wenn sie die in der Sowjetunion erlebten Diskriminierungen als ethnisch Deutsche betont und versucht, dazu passende Erlebnisse erzählerisch auszubauen, ihr ehemaliges parteipolitisches Engagement in der Sowjetunion jedoch nur am Rande bzw. zwischen den Zeilen andeutet. Mit der Rekonstruktion der Gegenwartsperspektive, die die für die Einreise nach Deutschland im Kontext des Aufnahmeverfahrens notwendigen Reinterpretationen der Vergangenheit, die stärkere Betonung der deutschen Abstammung und der Belege von Diskriminierungen deutlich macht, wird es möglich, die in der Darstellung ausgebauten Situationen von Diskriminierung nicht als Abbild des Erlebens in der Vergangenheit zu verstehen. Und ebenso kann die Gegenwartsperspektive durch die Rekonstruktionen der Erlebnisse in der Vergangenheit

auf ihre Bedeutungen bzw. ihre biographische Funktion hin befragt und vielleicht erklärt werden (Rosenthal 2010).

Wie erfolgt nun diese *Auswertung im Einzelnen*, bevor (zum Abschluss der Analyse) die Annahmen zu der in der Gegenwart präsentierten und die zur erlebten Lebensgeschichte miteinander kontrastiert werden?

### 35.5.1 Sequenzielle Analyse der biographischen Daten

Ganz entscheidend für eine analytische Distanzierung der InterpretInnen von der präsentierten Lebensgeschichte und für die Konzentration auf die Perspektiven in der Vergangenheit ist der erste Auswertungsschritt der sequenziellen Analyse der biographischen Daten (Rosenthal 2011; Kurt/Herbrik, Kapitel 33 in diesem Band), der von Ulrich Oermann (et. al. 1980) vorgestellt wurde und in seiner Logik, sich zunächst auf die Strukturdaten eines Falles zu beschränken, dem Vorgehen der Objektiven Hermeneutik entspricht.

Zunächst werden die nur in geringem Maße an die Interpretationen der Biographen gebundenen Daten (z.B. zur Familienkonstellation, zum Ausbildungsverlauf oder zu Krankheiten) in der *zeitlichen Abfolge der Ereignisse im Lebenslauf* analysiert. Diese Daten werden aus dem transkribierten Interview wie aus allen anderen zur Verfügung stehenden Quellen (Archivmaterial, Interviews mit Familienangehörigen oder anderen Bezugspersonen, Arztberichten, behördlichen Akten, z.B. Gerichtsakten) entnommen.

Historische bzw. sozio-politische Daten, die für den vorliegenden Fall relevant sein könnten, werden ebenso in diese Liste aufgenommen. Damit wird gleich zu Beginn der Analyse einer Biographie eine *gezielte Einbettung der lebensgeschichtlichen Erfahrungen in die jeweiligen gesellschaftlichen und politischen Kontexte* vorgenommen und so eine individualisierende Herangehensweise vermieden (Radenbach/Rosenthal 2012).

Mit diesem Analyseschritt soll vor allem auch einer vorschnellen Übernahme der gegenwärtigen Selbstdeutungen und -darstellungen der Interviewten entgegengewirkt werden. Angestrebt ist hier vielmehr das Auffinden bzw. die Konstruktion von Hypothesen über die möglichen Bedeutungen der verschiedenen biographischen Stationen für die Biographen zum Zeitpunkt des Erlebens, und das heißt nicht zuletzt über die sich mit jedem Datum sowohl *eröffnenden als auch verschließenden Handlungsoptionen*. Bei jedem Datum wird danach gefragt, welche Handlungsmöglichkeiten die Biographin oder der Biograph in einer bestimmten Situation hatte, welche er oder sie auswählte und welcher mögliche weitere biographische Verlauf daraus hätte folgen können.

### 35.5.2 Text- und thematische Feldanalyse

Der zweite Auswertungsschritt – die Text- und thematische Feldanalyse – erfolgt ebenfalls in der Logik einer sequenziellen Analyse im Sinne der Objektiven Hermeneutik (Hinrichsen et al. 2013; Rosenthal 2011). Die Hypothesenbildung orientiert sich an Schützes

Vorschlägen zur Sequenzierung des Textes nach Textsorten und an der thematischen Feldanalyse nach Gurwitsch. Ziel ist es, die *Regeln für die Gestaltung der in der Gegenwart des Interviews präsentierten biographischen Selbstpräsentation* herauszufinden. Die Analyse konzentriert sich auf die Frage, weshalb sich ein Biograph oder eine Biographin – ob nun bewusst intendiert oder latent gesteuert – so und nicht anders darstellt.

Zur Vorbereitung der Analyse wird das transkribierte Interview (oder andere biographische Texte) gemäß der zeitlichen Abfolge in der Form eines stichwortartigen Überblicks in Analyseeinheiten gegliedert. Kriterien für die *Sequenzierung*, d.h. für die Definition, wann eine Sequenz beginnt und wann sie endet, sind: Redewechsel, Änderungen der Textsorte und inhaltliche Modifikationen.

Die Hypothesenbildung bei der sequenziellen Analyse der einzelnen Einheiten orientiert sich an den Fragen, wann und wie ein Inhalt präsentiert wird, inwiefern dies durch die aktuelle Interaktionssituation bedingt ist und was das Thema der einzelnen Sequenzen bzw. das thematische Feld sein kann, das die Sequenzen verbindet. Die Frage, ob die einzelnen Sequenzen untereinander in einem Beziehungszusammenhang stehen, orientiert sich an Gurwitschs (1974: 4) Konzeption des „*thematischen Feldes*“, das er definiert als „die Gesamtheit der mit dem Thema kopräsenten Gegebenheiten, die als sachlich mit dem Thema zusammenhängend erfahren werden und den Hintergrund oder Horizont bilden, von dem sich das Thema als Zentrum abhebt“. Bei jeder Sequenz geht es in der Analyse um das *Auffinden der inhärenten Verweisungen auf mögliche thematische Felder und um den hypothetischen Entwurf der jeweils anschlussfähigen weiteren Sequenzen*. Im Fortgang der Analyse zeigt sich dann, welche thematischen Felder von den Biographen (breiter) ausgestaltet werden, welche sich potenziell anbietenden Bestände dieser Felder nicht entwickelt oder nur andeutend thematisiert werden. Ebenso wird dabei deutlich, welche Felder vermieden werden.

### 35.5.3 Rekonstruktion der erlebten Lebensgeschichte

Im dritten Schritt der Auswertung, der Rekonstruktion der erlebten Lebensgeschichte, wird in Verbindung mit den bei der Analyse der biographischen Daten formulierten Hypothesen wieder nach der *biographischen Bedeutung eines Erlebnisses zum Zeitpunkt des Erlebens* gefragt. Die Hypothesen zu den einzelnen biographischen Daten werden jetzt mit den dazu vorliegenden Aussagen der Biographen kontrastiert.

Die bereits abgeschlossene thematische Feldanalyse verhilft hierbei zu einem *quellenkritischen Blick* (Salheiser, Kapitel 62 in diesem Band), der vermeidet, die durch die Gegenwart neu konstituierte Perspektive auf die Vergangenheit naiv als Abbildung des Erlebens in der Vergangenheit zu verstehen. Wissen wir z.B., dass die vermutlich hinter dem Rücken der Biographin wirksam werdende Gestaltung einer biographischen Selbstpräsentation im thematischen Feld „Ich wurde als Deutsche in der Sowjetunion diskriminiert“ vor allem Ausdruck einer sozial eingeforderten Selbstdarstellung ist, sind wir offen für andere

Lesarten auf der Ebene der erlebten Lebensgeschichte und können damit z.B. die Auslassung der Thematisierung eines erfolgreichen Bildungsaufstiegs in der Sowjetunion in der Haupterzählung vor diesem Hintergrund interpretieren.

*Sequenzielle Feinanalysen* einzelner Textsequenzen in Anlehnung an die Objektive Hermeneutik (Oevermann 1983) dienen bei diesem Analyseschritt des Weiteren zur Überprüfung und Erweiterung von aus den bisherigen Auswertungsschritten gewonnenen Hypothesen.

## Literatur

### Klassische Grundlagentexte

- Fuchs-Heinritz, Werner (2005): Biographische Forschung. Wiesbaden: VS-Verlag
- Krüger, Heinz-Hermann/Marotzki, Winfried (Hg.) (1999): Handbuch erziehungswissenschaftliche Biographieforschung. Opladen: Leske + Budrich
- Rosenthal, Gabriele (1995): Erlebte und erzählte Lebensgeschichte. Frankfurt a.M.: Campus
- Schütze, Fritz (1983): Biographieforschung und narratives Interview. In: Neue Praxis 3: 283-293
- Völter, Bettina/Dausien, Bettina/Lutz, Helma/Rosenthal, Gabriele (Hg.) (2005): Biographieforschung im Diskurs. Wiesbaden: VS Verlag

### Zur Kombination von Diskurs- und Biographieanalyse

- Pohn-Weidinger, Maria (2013): Heroisierte Opfer? Wiesbaden: VS Verlag
- Wundrak, Rixta (2010): Die chinesische Community in Bukarest. Wiesbaden: VS Verlag

### Zur Kombination von Biographieforschung und Ethnographie

- Köttig, Michaela (2004): Lebensgeschichten rechtsextrem orientierter Mädchen und junger Frauen. Gießen: Psychosozial Verlag

### Zur Kombination von Biographieforschung und Videoanalyse

- Witte, Nicole (2010): Ärztliches Handeln im Praxisalltag. Frankfurt a.M.: Campus

### Zur Kombination biographischer Interviews und Gruppendiskussionen

- Miethe, Ingrid (1999): Frauen in der DDR-Opposition. Opladen: Leske + Budrich

### Zur Kombination biographischer Interviews mit Familiengesprächen

- Hildenbrand, Bruno (1999): Fallrekonstruktive Familienforschung. Opladen: Leske + Budrich
- Rosenthal, Gabriele (Hg.) (1997): Der Holocaust im Leben von drei Generationen. Gießen: Psychosozial Verlag

### Zur Auswertung von geschriebenen Autobiographien

- Müller-Botsch, Christine (2009): „Den richtigen Mann an die richtige Stelle“. Frankfurt a.M.: Campus
- Völter, Bettina (2003): Judentum und Kommunismus. Opladen: Leske + Budrich

## Weitere Literatur

- Bertaux, Daniel/Bertaux-Wiame, Isabelle (1991): „Was du ererbt von deinen Vätern ...“.  
Transmissionen und soziale Mobilität über fünf Generationen. In: BIOS 4 (1): 13-40
- Gurwitsch, Aron (1957/1974): Das Bewußtseinsfeld. Berlin/New York: De Gruyter
- Hildenbrand, Bruno (1995): Fallrekonstruktive Forschung. In: Flick, Uwe/von Kardorff, Ernst/Keupp, Heiner/von Rosenstiel, Lutz/Wolff, Stephan (Hg.): Handbuch qualitative Sozialforschung. Weinheim: Beltz. 256-259
- Hinrichsen, Hendrik/Rosenthal, Gabriele/Worm, Arne (2013): Biographische Fallrekonstruktionen. Zur Rekonstruktion der Verflechtung ‚individueller‘ Erfahrung, biographischer Verläufe, Selbstpräsentationen und ‚kollektiver‘ Diskurse. PalästinenserInnen als RepräsentantInnen ihrer Wir-Bilder. In: Sozialer Sinn (2), 157-183
- Oevermann, Ulrich (1983): Die Bedeutung von Adornos methodologischem Selbstverständnis für die Begründung einer materialen soziologischen Strukturanalyse. In: Friedeburg, Ludwig von/Habermas, Jürgen (Hg.): Adorno-Konferenz 1983. Frankfurt a.M.: Suhrkamp. 234-289
- Oevermann, Ulrich/Allert, Tilmann/Konau, Elisabeth (1980): Zur Logik der Interpretation von Interviewtexten. In: Heinze, Thomas/Klusemann, Hans-W./Soeffner, Hans-Georg (Hg.): Interpretationen einer Bildungsgeschichte. Bensheim: päd.-extra-Buchverlag. 15-69
- Radenbach, Niklas / Rosenthal, Gabriele (2012): Das Vergangene ist auch Gegenwart, das Gesellschaftliche ist auch individuell. Zur Notwendigkeit der Analyse biographischer und historischer Rahmendaten. In: Sozialer Sinn, 3-37
- Rosenthal, Gabriele (2005): Die Biographie im Kontext der Familien- und Gesellschaftsgeschichte. In: Völter, Bettina et. al. (Hg.): Biographieforschung im Diskurs. Wiesbaden: VS Verlag. 46-64
- Rosenthal, Gabriele (2010): Die erlebte und erzählte Lebensgeschichte. In: Griese, Birgit (Hg.): Subjekt – Identität – Person? Wiesbaden: VS Verlag. 197-218
- Rosenthal, Gabriele (2011): Interpretative Sozialforschung. Weinheim/München: Juventa (Kap. 5.4.: Narratives Interview und narrative Gesprächsführung; Kap. 6.: Biographieforschung und Fallrekonstruktionen)
- Schütze, Fritz (1977): Die Technik des narrativen Interviews in Interaktionsfeldstudien. In: Arbeitsberichte und Forschungsmaterialien der Universität Bielefeld 1. Bielefeld: Fakultät für Soziologie
- Thomas, William Isaac/Znaniecki, Florian (1958): The Polish Peasant in Europe and America. 2 Bände. New York (Neuausgabe der 2. Auflage von 1928, zuerst 1918-1920)
- Wohlrab-Sahr, Monika (1998). Konversion zum Islam. Frankfurt a.M.: Campus

Hubert Knoblauch

---

## 36.1 Was ist Ethnografie?

Ethnographie bezeichnet wörtlich die „Beschreibung“ des „Volkes“. Genauer handelt es sich um eine sozialwissenschaftliche Forschungsstrategie, bei der mehr oder weniger unbekannte ethnische Gruppen, Gemeinschaften oder andere soziale Einheiten und deren Handlungsweisen, Wissensformen und materiale Kulturen untersucht werden. Als ausformuliertes Konzept geht die Ethnographie auf Bronislaw Malinowski (1923) zurück und bezeichnet die Art der ethnologischen Feldforschung durch eigene systematische teilnehmende Beobachtung. Der Begriff der teilnehmenden Beobachtung wurde erstmals 1924 von Lindemann verwendet, der damit zunächst die soziologische Forschung im Rahmen der Chicagoer Schule charakterisierte (Lindner 1990). Auch wenn die teilnehmende Beobachtung (Thierbach/Petschick und Dangschat, Kapitel 66 und 76 in diesem Band) als Königsweg der Ethnographie erscheint, umfasst diese eine Vielzahl verschiedenster qualitativer und gelegentlich sogar quantitativer Methoden sowohl der Datenerhebung, als auch der Auswertung. Dazu zählen neben verschiedenen anderen Formen der Beobachtung vor allem das Interview (Helfferich und Küsters, Kapitel 39 und 40 in diesem Band) sowie Dokumente (Salheiser und Ernst, Kapitel 62 und 63 in diesem Band), audiovisuelle Aufzeichnungen (Tuma/Schnettler, Akremi, Kapitel 62 und 63 in diesem Band), Fotografien (Bohnsack, Kapitel 67 in diesem Band) und materiale Objekte (Schubert, Kapitel 70 in diesem Band). Zu den verschiedenen Methoden der Datenerhebung und den entsprechenden Datensorten kommen noch verschiedene Arten der Dateninterpretation und -analyse sowie der Darstellung hinzu.

### 36.1.1 Varianten der Ethnografie

Anfänglich als langfristige Beobachtung einer „fremden“ Ethnie konzipiert, hat sich die Ethnographie im Laufe der Zeit auch der eigenen Gesellschaft zugewandt. Zugleich hat sie sich in zahllose Varianten aufgefächert. Neben der *funktionalen Ethnographie* Malinowskis hat besonders die „*dichte Beschreibung*“ eine weite Verbreitung gefunden (Geertz 1983). Neben einer *Spezialisierung auf verschiedene Gegenstände* (Ethnographie der Kommunikation, des Diskurses, der Events, der Situation, der Stadt, der Szenen, der Organisationen, des Internets) zeichnet sich auch die Auflösung des engen Zusammenhangs der Ethnographie mit einem besonderen Ort ab, wie etwa in der „*globalen Ethnographie*“ (Burawoy et al. 2000), die den Zusammenhang lokaler mit globalen Prozessen betrachtet, und in der „*multi-sited ethnography*“ (Marcus 1995), die die Verknüpfung verschiedener Situationen und Schauplätze zum Thema hat. Im Unterschied zur konventionellen Ethnographie, die sehr langfristig an festen Orten und zumeist auch mit gleichbleibenden Akteuren betrieben wird, zeigt sich eine zunehmende Tendenz zur Fokussierung auf bestimmte Interaktionssituationen (Knoblauch 2001), auf räumliche Bewegungen von Akteuren („Walking along“, Ethnographie der Migration) und sogar auf die Bewegung und Verschiebung von Gegenständen (etwa in der „Akteur-Netzwerk-Theorie“).

### 36.1.2 Ethnographie und das Problem der Intersubjektivität

In diesem Beitrag fokussiere ich auf die grundsätzliche Methodologie der Ethnographie. Dabei möchte ich mich im ersten Teil darauf konzentrieren, wie die Ethnographie die grundlegende Frage der Intersubjektivität behandelt, wie wir Andere verstehen. Diese Frage ist grundlegend, denn sie betrifft nicht nur die Ethnographie, sondern die Sozialwissenschaften insgesamt. Diese stehen im Unterschied zu den Naturwissenschaften immer vor dem Problem, dass ihr Gegenstand selbst eigene Deutungen von seinen eigenen Handlungen hat – und dass diese Deutungen wiederum entscheidend sind für das, was erklärt werden soll:

„Genau genommen gibt es nirgendwo so etwas wie reine und einfache Tatsachen. Alle Tatsachen sind schon immer aus einem universellen Zusammenhang durch unsere Bewusstseinsabläufe ausgewählte Tatsachen. Somit sind sie immer interpretierte Tatsachen: entweder sind sie in künstlicher Abstraktion aus ihrem Zusammenhang gelöst oder aber sie werden nur in ihrem partikulären Zusammenhang gesehen. Daher tragen in beiden Fällen die Tatsachen ihren interpretativen inneren und äußeren Horizont mit sich“ (Schütz 1971: 7).

Wie im zweiten Teil gezeigt werden soll, kann das Verstehen von Anderen auf zweierlei Weise gelöst werden, und entsprechend finden sich in der Ethnographie eine Ethnographie des Fremden und eine Ethnographie des Anderen. Beide können, drittens, als komplemenäre Strategien zur Verfolgung des zentralen soziologischen Ziels der Ethnographie aufgefasst werden: Soziale Phänomene in ihrem eigenen Recht, also „*sui generis*“, zu verstehen.

## **36.2 Ethnographie als Zugang zur sozialen Wirklichkeit der Anderen**

Die meisten Methoden der Sozialwissenschaften pflegen einen sehr indirekten Zugang zu ihrem Gegenstand. Im Regelfall werden soziale Prozesse, Handlungen und Personen durch Aussagen über diese Ereignisse *rekonstruiert*, wobei sich die qualitativ-interpretativen Methoden (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) durch die Interpretation dieser Aussagen, die quantitativ-standardisierten Methoden (Stein, Kapitel 7 in diesem Band) durch die Kodierung und darauf aufbauende Quantifizierung auszeichnen. Im Unterschied dazu zählt die Ethnographie zu den Methoden, in denen sich die Forschenden einen unmittelbaren Zugang zum Gegenstand suchen. „Unmittelbar“ bedeutet hier nicht, dass der Gegenstand nicht etwa kommunikativ vermittelt sein könnte – was etwa in der „*Webnografie*“ (Strübing 2004) zwangsläufig der Fall ist –, sondern dass sich Forscher und Erforschte auf dem „*Feld*“ der Erforschten begegnen. Ethnographie untersucht also die soziale Wirklichkeit im Vollzug. Darin ähnelt sie dem Experiment (Eifler, Kapitel 11 in diesem Band). Im Unterschied aber zum Experiment, bei dem der Gegenstand wie die Bedingungen der Untersuchung von Seiten der Wissenschaft festgelegt werden, werden „die Anderen“ bei der Ethnographie in ihrem eigenen Recht untersucht. Es sind die Untersuchten selbst, die das Thema, die Situation und die Form der Untersuchung herstellen und damit vorgeben. Die Erhebung von Daten durch den genannten Methodenmix (Kelle, Kapitel 8 in diesem Band) dient dann dem Verstehen (Kurt/Herbrik, Kapitel 33 in diesem Band) dieses Feldes (Thierbach/Petschick, Kapitel 66 in diesem Band). Methodologisch gesprochen geht es der Ethnographie grundsätzlich darum, die Anderen (und ihre – je nach theoretischer Präferenz: Handlungen, Praktiken, Diskurse, insgesamt: Kultur) zu verstehen. Weil sich die Ethnographie von allen sozialwissenschaftlichen Methoden am vollständigsten auf die Anderen einlässt, kann man sie deswegen auch als Paradefall der interpretativen Sozialforschung verstehen: Das Verstehen der Anderen als Einzelne oder als Vertreter von Kulturen bleibt während des gesamten Forschungsprozesses eine Aufgabe, so dass selbst die Fragestellungen der Forschung von diesem Verstehen abhängig sind. Daraus erklärt sich die Nähe der Ethnographie zur Grounded Theory (Strübing, Kapitel 32 in diesem Band).

---

## **36.3 Befremdung und Binnenperspektive – zwei Strategien der Ethnographie**

Das Verstehen Anderer ist ein grundlegendes Thema der Sozialtheorie, das unter dem Titel der „Intersubjektivität“ behandelt wird. Dabei lassen sich zwei idealtypische Modelle unterscheiden (Knoblauch/Schnettler 2004).

### 36.3.1 Fremdheit

Die klassische Form der Ethnographie zielt auf das Andere als ein Fremdes, zunächst auch an einem anderen Ort, das in einem Kontrast zum Eigenen steht. Dieses Modell der *Fremdheit* („Alienetät“) betrachtet das Andere als eine Art „Black Box“. Aus diesem Modell leitet sich auch das „*Othering*“ ab, die Betrachtung des Anderen (und der anderen Kultur) als wesentlich unterschiedlich von der eigenen, der *Ethnozentrismus* (der Andere wird mit den eigenen Kategorien erfasst) und der *Exotismus* (Andere werden als different stilisiert).

### 36.3.2 Andersheit

Das Konzept der *Andersheit* („Alterität“) geht dagegen von einer Differenz aus, unterstellt aber Gemeinsamkeiten. Das Andere wird nicht einseitig nach Maßgabe der Vorstellungswelt der Forschenden und ihrer Kultur erfasst, sondern beidseitig und sozusagen „symmetrisch“. Theoretisch gelten die Strukturen der Lebenswelt (Schütz/Luckmann 1979) als eine gemeinsame Grundlage, die sich empirisch durch die zwischen ihnen ablaufende Kommunikation bewährt. In der Folge von Schütz und der interpretativen Methodologie wird das Andere also wesentlich durch Ähnlichkeit bestimmt, so dass die teilnehmende Beobachtung (Thierbach/Petschick, Kapitel 66 in diesem Band) als Erwerb des Wissens und der Praktiken des Feldes verstanden werden kann, die zu einer Angleichung an die Anderen führt (das als „*going native*“ bezeichnet wird).

### 36.3.3 Befremdung

Auch wenn die Begriffe für diese Unterscheidung nicht immer einheitlich verwendet werden (Rothenburg 2006; Schütz 2009), schlägt sich die Differenz zwischen beiden Vorstellungen in den Methodologien der deutschsprachigen Ethnographie nieder.

Das *Modell der Alienetät* findet sich mustergültig in dem, was Amann und Hirschauer (1997; siehe auch Agar 1996) als *ethnographische Strategie der Befremdung* vorschlagen. Zwar räumen sie durchaus ein, dass es graduelle Anpassungen des Forschenden an das untersuchte Feld gibt, doch besteht ihrer Ansicht nach die zentrale Strategie des ethnographischen Forschens in der Erzeugung von Differenz zum untersuchten Feld. Diese Befremdung wird nicht nur als punktuell verstanden, wie dies in vielen Methodenbüchern vorgesehen ist. Vielmehr gelte es, die Befremdung kontinuierlich zu erneuern.

„Die verständnisvolle Vertrautheit ist kein Telos, sondern ein immer neu zu überwindender Durchgangspunkt“ (Amann/Hirschauer 1997: 29).

Das bedeutet, dass man sich nie in einem Feld einrichtet, dazu immer Distanz wahrt. Das Feld wird nicht nur fortwährend befremdet, es bleibt in einem sehr konkreten Sinne fremd.

### 36.3.4 Binnenperspektive

Die Strategie der Befremdung steht in einem häufig übersehenen Gegensatz zur Forschungsstrategie der *lebensweltlichen Ethnographie* (Honer 1993; Hitzler/Eberle 2000). Anstatt eine Distanz zum Feld aufrecht zu erhalten, zielt die lebensweltliche Ethnographie darauf, Teil des Feldes werden, d.h. die Wirklichkeiten möglichst so zu erfassen, „wie sie von Mitgliedern typischerweise erfahren, erlitten und erhandelt werden“ (Honer 2011: 31). Wer vorhat, Ethnographie zu treiben, will die Perspektive der Handelnden einnehmen, und das bedeutet, dass man selber die Handlungskompetenz erwirbt.

Das entscheidende Merkmal ist die Einnahme dessen, was Malinowski (1923) in seiner ethnologischen Forschung als den „*native's point of view*“ bezeichnete. Darunter wird der Versuch des Forschenden bezeichnet, die Perspektive derer einzunehmen, die untersucht werden, um sie verstehen zu können. Wie bei der *Autoethnographie* (Ellis et al. 2011) wird das eigene Forschungshandeln als Teil der Lebenswelt betrachtet, die untersucht werden soll (Hammersley/Atkinson 1983); doch soll dabei auf die mögliche Differenz zur Perspektive der Untersuchten, etwa hinsichtlich des Geschlechts, der Klasse oder der Ethnie geachtet werden.

---

## 36.4 Ethnografie und Globalisierung

Auch wenn die beiden Strategien der Ethnographie auf kontrastierenden Modellen des Anderen beruhen – als soziologische Erforschung des mehr oder weniger Bekannten und als ethnologische Erforschung des Fremden –, so können sie als zwei Antworten auf dasselbe Problem verstanden werden. Denn angesichts von Prozessen wie Globalisierung, Transnationalisierung, der wachsenden Weltkultur wird Ethnographie zusehends zu einer *Erforschung der eigenen (Welt-)Gesellschaft*. Entsprechend müssen sich die ethnographischen Strategien zunehmend der Tatsache anpassen, dass die zu beschreibende Welt uns in der einen oder anderen Weite immer schon bekannt ist. Die beiden Strategien tragen genau diesem Umstand Rechnung, indem sie einmal das Bekannte sozusagen künstlich befremden und zum anderen das, was different im Gemeinsamen ist, fokussieren. Zudem können beide Strategien auch als *komplementäre und alternierende epistemologische Strategien* verstanden werden: Während die Strategie der *Binnenperspektive* von der Möglichkeit des Verstehens ausgeht und rekonstruiert, dient die *Befremdung* als Korrektiv. Die Verschränkung beider Strategien schützt zum einen vor einer Reduktion auf einen relativistischen Subjektivismus der Forschenden selbst; zum anderen verhindert sie eine naive Vorstellung der Objektivität (Fine 1993). Die Beidseitigkeit ist auch der Kern des ethnographischen Vorgehens: Während die Ethnographie – wie jedes Forschungshandeln – den Fokus ihrer Beobachtung durch ihre eigene Fragestellung bestimmt, geht es ihr als Ethnographie immer auch darum, das Feld so zu bestimmen, wie es von den Akteuren in ihren Handlungen und mit ihrem Wissen geschaffen, gestaltet und aufrechterhalten wird.

## Literatur

### Grundlagenittel zur Ethnographie

- Atkinson, Paul/Coffey, Amanda/Delamont, Sara (2007): Handbook of Ethnography: London: Sage
- Breidenstein, Georg/Hirschauer, Stefan/Kalthoff, Herbert/Nieswand, Boris (2013): Ethnografie. Die Praxis der Feldforschung. Konstanz: UVK
- Girtler, Roland (1984): Methoden der qualitativen Sozialforschung. Anleitung zur Feldarbeit. Wien/Köln/Graz: Böhlau
- Knoblauch, Hubert (2003): Qualitative Religionsforschung. Religionsethnographie in der eigenen Gesellschaft. Paderborn/München/Wien/Zürich: Ferdinand Schöningh
- Lofland, John (2004): Analyzing Social Settings: A Guide to Qualitative Observation and Analysis. Belmont: Wadsworth

### Weiterführende Fragen und spezifische Methoden

- Amann, Klaus/Hirschauer, Stefan (Hg.) (1997): Die Befremdung der eigenen Kultur. Frankfurt a.M.: Suhrkamp
- Burawoy, Michael/Blum, Joseph A./George, Sheba/Gille, Zsuzsa/Gowan, Teresa/Haney, Lynne/Klawitter, Maren/Lopez, Steven H./O Riain, Sean/Thayer, Millie (Hg.) (2000): Global Ethnography. Berkeley: University of California Press
- Denzin, Norman (1997): Interpretive Ethnography. Thousand Oaks: Sage
- Ellis, Carolyn/Adams, Tony E./Bochner, Arthur P. (2011): Autoethnography. In: FQS 12 (1): Art. 10
- Hollan, Douglas (2001): Developments in Person-Centered Ethnography. In: More, Carmella C./Mathews, Holly F. (Hg.): The Psychology of Cultural Experience. Cambridge: CUP. 48-67
- Honer, Anne (1993): Lebensweltliche Ethnographie. Wiesbaden: DUV
- Knoblauch, Hubert (2001): Fokussierte Ethnographie. In: sozialersinn 2 (1): 123-141
- Marcus, George E.: Ethnography In/Of the World System. The Emergence of Multi-Sited Ethnography. In: Annual Review of Anthropology 24: 95-117
- Rammert, Werner/Schubert, Cornelius (2006): Technografie. Frankfurt a.M./New York: Campus
- Smith, Dorothy E. (2005): Institutional Ethnography. Lanham: Altamira
- Strübing, Jörg (2004): Webnografie? Zu den methodischen Voraussetzungen einer ethnografischen Erforschung des Internet. In: Rammert, Werner/Schubert, Cornelius (2006): Technografie. Frankfurt a.M./New York: Campus. 249-274

## Zeitschriften

Ethnography

Journal or Contemporary Ethnography

## Weitere Literatur

- Agar, Michael (1996): *The Professional Stranger*. New York: Academic Press
- Amann, Klaus/Hirschauer, Stefan (1997): Die Befremdung der eigenen Kultur. In: ibides (Hg.): *Die Befremdung der eigenen Kultur*. Frankfurt a.M.: Suhrkamp. 7-52
- Burawoy, Michael/Blum, Joseph A./George, Sheba/Gille, Zsuzsa/Gowan, Teresa/Haney, Lynne/Klawitter, Maren/Lopez, Steven H./O Riain, Sean/Thayer, Millie (Hg.) (2000) : *Global Ethnography*. Berkeley: University of California Press
- Ellis, Carolyn/Adams, Tony E./Bochner, Arthur P. (2011): Autoethnography. In: FQS 12 (1). Art. 10
- Emerson, Robert (2001): The Development of Ethnographic Field Research. In: ibidem (Hg.): *Contemporary Field Research*. Long Grove: Waveland Press. 1 -53
- Fine, Gary Allen (1993): Ten Lies of Ethnography. In: *Journal of Contemporary Ethnography* 22 (3): 267-294
- Geertz, Clifford: Dichte Beschreibung. Frankfurt a.M.: Suhrkamp 1983
- Hammersley, Martyn/Atkinson, Max (1983): *Ethnography*. London/New York: Tavistock
- Hitzler, Ronald/Eberle, Thomas (2000): Phänomenologische Lebensweltanalyse. In: Flick, Uwe/von Kardoff, Ernst/Steinke Ines (Hg.): *Qualitative Forschung*. Reinbek: Rowohlt. 109-118
- Hollan, Douglas (2001): Developments in Person-Centered Ethnography. In: More, Carmella C./Mathews, Holly F. (Hg.): *The Psychology of Cultural Experience*. Cambridge: CUP. 48-67
- Honer, Anne (2011): Das Perspektivenproblem in der Sozialforschung. In: ibidem (Hg.): Kleine Leiblichkeiten. Wiesbaden: VS Verlag. 27-41
- Keller, Reiner (2005): Wissenssoziologische Diskursanalyse. Wiesbaden: VS Verlag
- Knoblauch, Hubert (2001): Fokussierte Ethnographie. In: *sozialersinn* 2 (1): 123-141
- Knoblauch, Hubert/Schnettler, Bernt (2004): „Postsozialität“, Alterität und Alienetät. In: Schetsche, Michael (Hg.): *Der maximal Fremde*. Würzburg: Ergon. 23-42
- Lindner, Rolf (1990): Die Entdeckung der Stadtkultur. Frankfurt a.M.: Suhrkamp
- Malinowski, Bronislaw (1923): The Problem of Meaning in Primitive Languages. In: Ogden, Charles K./Richards, Ivor A. (Hg.): *The Meaning of Meaning*. London: Kegan Paul. 451-510
- Marcus, George E.: Ethnography In/Of the World System. In: *Annual Review of Anthropology* 24: 95-117

- Rothenburg, Richard (2006): Social Constructivism and the Enigma of Strangeness. In: Rothenburg, Richard/Schnepel, Burkhard/Shimada, Shingo (Hg.): *The Making and Unmaking of Differences*. Bielefeld: Transcript. 27-42
- Schütz, Alfred (1971): Wissenschaftliche Interpretationen und Alltagsverständnis menschlichen Handelns In: ibidem (Hg.): *Gesammelte Aufsätze I*. Den Haag: Nijhus. 1-52
- Schütz, Alfred (2009): Das Problem der transzendentalen Intersubjektivität bei Husserl. In: Sebald, Gerd/Grathoff, Richard/Thomas, Michael (Hg.): *Philosophisch-phänomenologische Schriften I*. Konstanz: UVK. 227-257
- Schütz, Alfred/Luckmann, Thomas (1979; 1984): *Strukturen der Lebenswelt*. 2 Bände. Frankfurt a.M.: Suhrkamp

Linda Hering und Robert J. Schmidt

---

## 37.1 Was sind Einzelfallanalysen?

Unter einem *Fall* kann Vielfältiges verstanden werden. Neben klassischen Gegenständen der Sozialforschung, wie Einzelpersonen, Personengruppen, Organisationen/Netzwerken, ganzen Gesellschaften bzw. Kulturen oder anderen Formen sozialer Zusammenhänge, können Einzelfallanalysen auch soziale Prozesse, Episoden, einzelne Situationen oder Ähnliches betreffen (Baur/Lamnek 2005). Zudem gibt es eine Vielzahl von Anwendungsbereichen und sozialwissenschaftlichen Disziplinen, die mit Fallstudien arbeiten, z.B. im Bereich der Evaluations-, Stadt-, Biographie- oder auch Organisationsforschung.

Bei der *Einzelfallanalyse* (*Case Study*) handelt es sich demnach um eine *umfassende Forschungsstrategie, bei der eine abgrenzbare Einheit – ein Fall – in ihren Binnenstrukturen und Umweltverhältnissen umfassend verstanden werden soll* (Yin 2009: 18). Sie kann sowohl standardisierte (Stein, Kapitel 7 in diesem Band) wie nichtstandardisierte Erhebungs- und Auswertungsverfahren umfassen (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band), steht in ihren Erkenntnisinteressen jedoch eher der qualitativen Tradition der Sozialforschung nahe. Sie startet mit einem motivierten Interesse an einer zunächst heuristisch abgegrenzten Untersuchungseinheit und will diese im Verlauf des Forschungsprozesses umfassend verstehen und ihre konkrete Ausprägung erklären. Die Qualifikation der Einzelfallanalyse als Forschungsstrategie beruht darauf, dass sie durch kein spezielles Erhebungs- oder Auswertungsverfahren, sondern vielmehr durch ein umfassendes Forschungsdesign mit dem Ziel eines Tiefenverständnisses des Falls gekennzeichnet ist. Weiterhin wird der Fall in seiner natürlichen Umgebung untersucht – die Abgrenzung zwischen Fall und Umgebung ist fließend (Yin 2009). Anders als im Experiment oder in quasi-experimentellen Designs (Eifler, Kapitel 11 in diesem Band) sind für Fallstudien die umgebenden Rahmenbedingungen nicht als Störvariablen auszuschließen, sondern liegen vielmehr explizit im Erkenntnisinteresse des Vorhabens. Praktisches Leitmotiv für die Durchführung einer Fallstudie sollte es daher sein, in Erhebung, Auswertung und

Interpretation stets den vielfältigen Facetten des Falls zu folgen, ihn in seiner Komplexität möglichst umfassend darzustellen und dabei ein Forschungsdesign anzustreben, das die hierfür nützlichen Verfahrensstrategien vereint.

---

## 37.2 Welche Fragen können durch Fallanalysen geklärt werden?

Insgesamt können für Fallstudien prinzipiell drei Idealtypen wissenschaftlicher Forschungsinteressen analytisch unterschieden werden, die einander in konkreten Analysen häufig ergänzen:

1. *Der per se interessante Fall.* Es gilt das neue, unbeschriebene Phänomen zunächst einmal in allen Facetten möglichst erschöpfend zu beschreiben und so erst einer eventuell beginnenden sozialwissenschaftlichen Debatte zugänglich zu machen. Bei diesem Typus ist eine Begründung des Neuen, Interessanten am Fall notwendig. Die Argumentation bewegt sich jedoch stets auf der Seite des Unbekannten, Problematischen der Situation, die dann auch ungerichtetes Forschen motivieren kann. Die zahlreichen Facetten eines Falls werden so miteinander in Beziehung gesetzt, dass sie für sich interessant sind. In einer seiner bekanntesten Studien beschäftigt sich z.B. Latour (1998) mit dem Verkehrsprojekt „Aramis“, welches in den 1980er-Jahren in Frankreich durchgeführt wurde. Das innovative führerlose, schienengebundene Personentransportsystem versuchte Fahrgäste individuell ans Ziel zu bringen, es scheiterte jedoch aufgrund von sich widersprechenden Interessen. Detailliert präsentiert der Autor den Fall unter der Verwendung von Interviewpassagen sowie Forschungsnotizen.
2. *Erkenntnisinteressen in Bezug auf bestehende Theorien verschiedener Reichweite.* Hierbei ist der Gegenstand für einen theoretischen Diskurs von Interesse, entweder weil er diesem einen neuen Aspekt hinzufügt, bestehende Vorstellungen in Frage stellt oder ein besonders typisches Beispiel für diesen ist. Der Fall wird demnach zu bestehenden Theorien unterschiedlicher Reichweite – also in Bezug zu Konzepten über ein lokales Handlungsfeld bis hin zu gesellschaftsweit tragenden Begrifflichkeiten – in Beziehung gesetzt. So kann etwa eine Untersuchung der Stadt Dortmund Einiges zur bestehenden Theorie postindustrieller Städte beitragen oder aber für die empirische Fundierung eines Theorieansatzes wie dem der „Eigenlogik der Städte“, z.B. mittels Ethnografie (Knoblauch, Kapitel 36 in diesem Band) oder Diskursanalysen (Traue et al., Kapitel 34 in diesem Band) herangezogen werden (Anwendungsbeispiele liefern: Barbehön 2013; Baur et al. 2014).
3. *Erkenntnisinteressen in Bezug auf ein umfassenderes Forschungsdesign oder bereits bestehende Einzelfallstudien.* Hier steht der Wechsel von Fallvergleich und Fallkontrastierung im Fokus der Analyse (Kelle/Kluge 2010; Yin 2009; Vaughan 1992: 199), also die Relation des Einzelfalls zu einem oder mehreren konkreten anderen Fällen. Dabei kann es sich sowohl um die Einbettung in ein größeres Forschungsvorhaben, als auch um den Bezug auf eine spezifische, in der Literatur beschriebene Fallstudie handeln. Für die Arbeit in einem Projektverbund ist eine gute Fallauswahl z.B. in Hinblick auf

das gemeinsame Erkenntnisinteresse äußerst bedeutsam, muss aber immer auch den einzelnen Fragestellungen der Teilprojekte genüge tun. Im Zuge des DFG-Verbundprojektes „Eigenlogik der Städte“ wurden aus diesem Grund vier Städte systematisch ausgewählt, die es den einzelnen Teilprojekten ermöglichen, über verschiedene Forschungsfelder (Friseurbranche, Kriminalliteratur, Problemdiskurse und Stadtmarketing) hinweg empirische Belege für die Beschreibung der Eigenlogik der einzelnen Städte zu erbringen und dabei gleichzeitig eigene Forschungsinteressen zu verfolgen (Barbehön 2013; Baur et al. 2014; Marent/Richter 2013).

---

### 37.3 Welche Fälle sollten ausgewählt werden?

Die Einzelfallanalyse ist auch in der Auswahlstrategie des Gegenstandsbereichs ein spezifischer Typus, der von Untersuchungen mit kleinen und großen Fallzahlen abgegrenzt werden kann: Es handelt sich in der Regel um einen bzw. in einem vergleichenden Design meist um zwei Fälle. Stets erfolgt die Auswahl der Fälle im Fallstudiendesign nicht zufällig, sondern vielmehr bewusst (Akremi, Kapitel 17 in diesem Band) in Hinblick auf das spezifische Erkenntnisinteresse. Sie entspringt einer mehr oder weniger stark strukturierten Vorstellung über den Gegenstand des Forschungsvorhabens zu Beginn der Untersuchung, die sich durchaus auch als problematisch erweisen kann. Die Gegenstände können dabei vorab unterschiedlich stark eingegrenzt werden, allerdings muss ein Interesse in jedem Fall begründet sein (Vaughan 1992).

Diese Vorstellung vom Fall hängt wiederum stark vom oben skizzierten Erkenntnisinteresse ab. In diesem Sinne können dem jeweiligen Erkenntnisinteresse dann auch weitere Typen von Fällen zugeordnet werden, die klassischerweise in Fallstudien untersucht werden:

#### 37.3.1 Der per se interessante Fall

1. Der *neue und unbeschriebene Fall*, der dem bisherigen begrifflichen Instrumentarium der Sozialwissenschaften bislang nicht zugänglich war, wie z.B. Latours (1998) oben genannte Studie zum Verkehrsprojekt „Aramis“.
2. Der *normativ interessante Fall*, z.B. im Sinne von Auftrags- und Evaluationsforschung (Döring und Ziegler, Kapitel 9 und 10 in diesem Band), oder auch ethisch-moralisch begründeten Gegenstandsbereichen. Fallstudien zum Thema „soziale Ungleichheit“ sind in der Soziologie sehr prominent und werden zudem aus gegebenem Anlass immer wieder von Bundes- bzw. Landesregierungen in Auftrag gegeben (z.B. Solga et al. 2009).

### **37.3.2 Auswahl in Bezug auf bestehende Theorien**

1. *Hypothesentestende, kritische Fälle.* In „Die Arbeitslosen von Marienthal“, einem Klassiker der empirischen Sozialforschung, untersuchen Jahoda et al. (1975) die Folgen von Arbeitslosigkeit, welche ein ganzes Dorf betraf. Dabei stellen die Forscher fest, dass Langzeitarbeitslosigkeit nicht wie erwartet zur Revolte führt, sondern sich vielmehr in passiver Resignation Ausdruck verleiht.
2. *Fälle, die neuartige Hypothesen über den Gegenstandsbereich nahelegen.* Berking und Schwenk (2011) wählten die beiden Hafenstädte Bremerhaven und Rostock für einen Fallvergleich zu deren lokalen Besonderheiten aus. Es zeigte sich, dass in beiden Städten Selbst- und Fremdbilder stark variieren und sich dies ebenfalls im Umgang mit auftretenden Problemen widerspiegelt. Dieses neue perspektivenverschränkende Vorgehen auf die Fälle eröffnete den Autoren, dass Bremerhaven eher von einer Krisenstimmung gezeichnet ist, während Rostock sich in der Analyse als gleichermaßen traditions- wie erfolgreiche Stadt zu präsentieren vermag.
3. *Fälle, die im Hinblick auf einen spezifischen theoretischen Diskurs einen neuen Aspekt thematisieren.* Häußermann und Kapphan (2002) gehen etwa der Frage nach, inwiefern sich Berlin nach dem Fall der Mauer von einer ehemals politisch geteilten in eine neu-erndings sozial gespaltene Stadt entwickelt. Hierfür beleuchten sie die Stadt z.B. unter Aspekten der postmodernen Stadt wie Arbeitsmarkt und Wohnungsversorgung, setzen sich weiterhin mit den historisch gewachsenen sozialräumlichen Strukturen wie Arbeiterviertel und Villenkolonien auseinander und gelangen schließlich über aktuelle Zuwanderungs- und Segregationsproblematiken zu stadtpolitischen Anforderungen.

### **37.3.3 Auswahl in Bezug auf ein Forschungsdesign bzw. andere, konkrete Fälle**

1. Man kann nach einem *maximal ähnlichen Fall* in Relation zu einem anderen im Forschungsdesign oder in der Literatur beschriebenen Fall (*Most Similar Cases Designs*, Akremi, Kapitel 17 in diesem Band) suchen. Eine solche Fallauswahl wurde in dem DFG-Verbundprojekt „Eigenlogik der Städte“ herangezogen. Die vier ausgewählten Städte sollten aufgrund ihrer Ähnlichkeiten vergleichend untersucht werden, weshalb bei der Fallauswahl bestimmte Faktoren wie z.B. Größe, Dichte, Heterogenität, aktuelle Problemlagen sowie historische Traditionen konstant gehalten wurden (Baur et al. 2014).
2. Man kann nach einem *maximal unterschiedlichen Fall* in Relation zu einem anderen im Forschungsdesign oder in der Literatur beschriebenen Fall (*Most Different Cases Designs*, Akremi, Kapitel 17 in diesem Band) suchen. Die Fallauswahl für die im vorherigen Punkt beschriebene Forschung des DFG-Projektverbundes „Eigenlogik der Städte“ wurde z.B. nicht ausschließlich aufgrund der Ähnlichkeiten getroffen, auch Differenzen wurden systematisch herangezogen. So ähneln sich die Städte in dem Fakt,

- dass sie sich vergleichbaren Problemen zu stellen hatten, doch unterscheiden sie sich grundlegend darin, wie erfolgreich sie sich diesen Aufgaben stellten (Baur et al. 2013).
3. Man kann nach einem *Fall* suchen, *der im Hinblick auf eine oder mehrere eingrenzbare Dimensionen verschiedenartig ist, in anderen jedoch dem Normalfall in der Population gleicht*. Somit kann der Einfluss dieser Dimension empirisch zugänglich gemacht werden. Das südhessische Darmstadt ähnelt hinsichtlich seiner Größe und formalen Strukturen vielen anderen Städten in Deutschland. Die umfassende Studie „Typisch Darmstadt“ (Löw et al. 2010), die den lokalen Besonderheiten der Stadt nachgeht, belegt jedoch eindrucksvoll, wie vielschichtig dieser Fall hinsichtlich Kreativität, Integration, Verkehr, Sicherheit usw. ist. Statistisch ähnelt er vielen anderen, betrachtet man allerdings die spezifischen Konstellationen von Wissensbeständen und Ausdrucksformen im Detail, ist er doch sehr individuell.
  4. Zudem können *Extremfälle* betrachtet werden, welche die Variation innerhalb einer bestimmten Population abbilden. Detroit wird z.B. gern als Extremfall für den wirtschaftlichen Niedergang postindustrieller Städte herangezogen, weshalb sich viele Studien mit dieser amerikanischen Stadt beschäftigen. Gallagher (2010) geht einen Schritt weiter und gibt nicht nur eine umfassende Beschreibung der vorherrschenden Problematiken Detroits, sondern liefert mit Hilfe seiner Fallstudie auch Vorschläge, welche Potenziale (z.B. Nutzbarmachung urbaner Landschaften) der Stadt noch inne liegen und inwiefern anderorts erfolgreiche Projekte hier etabliert werden könnten.

---

## 37.4 Forschungsdesign von Fallstudien

Allgemein sind bei Fallstudien zwei idealtypische Forschungsdesigns zu unterscheiden, die wiederum das Vorgehen bei der Auswahl, Erhebung und Auswertung der Daten beeinflussen. Einerseits ein *lineares Vorgehen* (Thomas 2011), bei dem die Abgrenzung des Falles mit einem relativ feststehenden Forschungsdesign (Stein, Kapitel 7 in diesem Band) mit konkretem Sampling-Plan einhergeht, der meist einem Ablauf aus der Festlegung der Forschungsfrage und des Designs, der Datenerhebung und -auswertung sowie darauf folgender theoretischer Verallgemeinerung folgt. Andererseits kann auch ein *iteratives Vorgehen* gewählt werden, bei dem im Zuge der Theoriegenese der Fall erst im Verlauf des Forschungsprozesses immer weiter eingrenzt wird (Przyborski/Wohlrab-Sahr und Strübing, Kapitel 6 und 32 in diesem Band).

So lässt sich eine Stadt als Fall in vielfältiger Weise *abgrenzen* (Abbott 1992; Baur 2014), z.B. räumlich über Verwaltungsgrenzen oder hinsichtlich spezifischer Merkmale ihre Wirtschaftsstrukturen betreffend. Insbesondere bei historischen Fallstudien müssen sich die Forscher mit sich verändernden Grenzen auseinandersetzen. Es ist also häufig lohnenswert die Eingrenzung und Abgrenzung des Falls nicht vorab festzustellen, sondern selbst zu einem ersten Ergebnis der Forschung zu machen. Hierbei stellt sich das Forschungsdesign allgemein als ständiger Wechsel aus den verschiedenen Elementen des Forschungsprozesses dar, in dem vor allem im Sinne des *Theoretical Sampling* (Strübing, Kapitel 32 in diesem Band) gearbeitet wird.

Weiterhin eröffnet das Fallstudiendesign dem Forscher die Möglichkeit eines umfassenden *Methoden-Mix* (Kelle, Kapitel 8 in diesem Band), wodurch er verschiedenste Facetten seines Falles durch unterschiedliche Daten- und Auswertungsformen beleuchten kann. Eine grundlegende Auseinandersetzung mit den Besonderheiten der jeweils verwendeten Erhebungs- und Auswertungsmethoden ist dabei unabdingbar. Bei diesem Vorgehen ist es besonders wichtig, sein Erkenntnisinteresse – das Tiefenverständnis des Falls – nicht aus dem Blick zu verlieren. Insbesondere die flexible Handhabung der einzelnen Teilschritte des Forschungsprozesses, welche es ermöglicht, sich dem Gegenstand unvoreingenommen und offen zu nähern, begünstigt zugleich ein Abschweifen und einen Verlust des analytischen Fokus (Baur/Lamnek 2005: 248). Wie schon angesprochen, kann sich die Abgrenzung des einzelnen Falles im Laufe des Forschungsprozesses noch wandeln, was wiederum Auswirkungen auf bereits aufgestellte Hypothesen hat, die dementsprechend verändert oder modifiziert werden müssen und schließlich in einem angepassten Forschungsdesign münden (Walton 1992: 129).

Zudem besteht ein grundlegender Unterschied im Design zwischen *Fallstudien, die für sich stehen* und solchen, die in *fallvergleichendem Sinne* angelegt sind. Bei letzteren kommt der internen Koordination zwischen den beiden Fällen eine wesentlich größere Rolle zu. Im Vergleich der Fälle kann nach Ähnlichkeitsstrukturen gesucht und so ein Typus konstruiert (Kelle/Kluge 2010) oder durch kontrollierte Varianz eines Aspekts nach dem Einfluss dieses gesucht werden.

---

## 37.5 Datenauswahl und Datenerhebung

Die Datenauswahl hängt bei (Einzel-)Fallstudien primär von der Abgrenzung und dem Erkenntnisinteresse ab. Prinzipiell können sowohl *standardisierte* als auch *nicht-standardisierte Daten* verwendet werden. Relevant ist jeweils der Gewinn für das Tiefenverständnis des Falles bzw. das Erkenntnisinteresse, wodurch u.a. Archivdaten (Salheiser, Kapitel 62 in diesem Band), Dokumente (Ernst und Klein, Kapitel 63 und 64 in diesem Band), (nicht-)teilnehmende Beobachtungen (Thierbach/Petschick, Kapitel 66 in diesem Band), qualitative Interviews (Helfferich, Kapitel 39 in diesem Band), Statistiken (Hartmann/Lengerer, Kapitel 71 in diesem Band), Umfragen (Reinecke, Kapitel 44 in diesem Band), aber auch Bild- und Tonmaterialien (Bohnsack; Schnettler/Tuma und Akremi, Kapitel 67, 68 und 69 in diesem Band) sowie Objekte aller Art (z.B. Autos oder Gebäude) (Schubert und Dangschat, Kapitel 70 und 76 in diesem Band) mögliche Informationsquellen darstellen. Möchte ein Forscher z.B. etwas über eine Stadt als Ganzes aussagen, bieten sich vielerlei Datenquellen an: Für die Außendarstellung kann er sich Materialien des Stadtmarketings anschauen, einen Blick in den Alltag über Ethnografien (Knoblauch, Kapitel 36 in diesem Band) erheben, politische Diskurse (Traue et al., Kapitel 34 in diesem Band) in Zeitungen oder über Protokolle von Ratssitzungen verfolgen und sogar literarische bzw. filmische Werke wie Krimis heranziehen, um ein tieferes Verständnis für die fiktionale Darstellung seines Falles zu erlangen.

Der Datenerhebungsprozess selbst beginnt meist bereits vor dem eigentlichen Start der Fallanalyse, da schon vorab Hintergrundinformationen gesammelt werden, die später wieder Verwendung finden können (Stake 1995: 49). Im Vergleich zu anderen Paradigmen des Forschungsdesigns, die in der Regel bestimmte Arten von Datentypen nahe legen, stellt sich der Erhebungsprozess in Fallstudien oft komplexer dar, weshalb die Einhaltung eines vorher zusammengestellten Forschungsdesigns in der Praxis meist unabdingbar ist (Stake 1995: 51; Baur/Lamnek 2005: 244f.). Vor allem, wenn mehrere Forscher an einem Fall (= einem Projekt) arbeiten, ist es sinnvoll, einzelne Schritte – von der Datenauswahl über die Erhebung bis zur Niederschrift der Ergebnisse – aufzulisten und Zuständigkeiten zuzuteilen. Allerdings sind Änderungen des Fallstudiendesigns und damit auch der Datenerhebung während des laufenden Forschungsprozesses möglich und teilweise sogar notwendig, wenn sich z.B. die Fallabgrenzung oder die Datenzugänge ändern. So können sich Stadtgrenzen im historischen Verlauf durch Eingemeindungen verändern, und Unternehmen können Strukturveränderungen durch Fusionen oder Auslagerungen von Tätigkeiten erfahren. Dies kann es für den Forscher notwendig machen, seinen Fall neu abzugrenzen (ihn auszuweiten bzw. einzugrenzen) und ggf. Feldzugänge neu zu eröffnen. Ebenfalls sinnvoll ist es, eine Fallstudiendatenbank anzulegen, in der verschiedene Dokumente wie z.B. Feldnotizen, Protokolle, Tabellen oder auch Interviewtranskripte gesammelt aufgelistet werden (Yin 2009: 118). Dies erlaubt es den beteiligten Forschern, nicht nur zu jedem Zeitpunkt des Forschungsprozesses einen Überblick über alle erhobenen und noch zu erhebenden Daten zu erlangen, sondern insbesondere auch während der Analysephasen, auf schon erarbeitete Ergebnisse oder Erklärungsversuche zurückgreifen zu können (Yin 2009: 119ff.).

---

## 37.6 Auswertung des Falles

### 37.6.1 Analytische Techniken

Anknüpfend an das Vorgehen bei der Datenauswahl, ist es beim Analyseverfahren ebenso möglich bzw. sogar nötig, qualitative Ansätze wie die wissenssoziologische Hermeneutik (Kurt/Herbrik, Kapitel 33 in diesem Band) und/oder quantitative Methoden wie z.B. Zeitreihen- oder Sequenzanalysen (Mika/Stegmann, Kapitel 85 in diesem Band) anzuwenden. Für die Arbeit am Material schlägt Yin (2009: 136ff.) fünf analytische Techniken vor, welche dem Forscher bei der Ergründung seines Falles helfen können: *Musterabgleich*, *Erklärungsbildung*, *Zeitreihenanalyse* (Metz/Thome, Kapitel 84 in diesem Band), *Aufstellen von Logikmodellen* und *Fallvergleich-Synthese*. Die jeweils geeigneten Analysetools wie z.B. die *qualitative Inhaltsanalyse* (Mayring/Fenzl, Kapitel 38 in diesem Band) oder standardisierte *Mustererkennungsverfahren*, wie z.B. die Sequenzmusteranalyse, (Mika/Stegmann, Kapitel 85 in diesem Band), sind je nach vorhandenen Datentypen sowie dem Erkenntnisinteresse auszuwählen und können dementsprechend von Fall zu Fall verschieden sein.

### **37.6.2 Dokumentation**

Die Dokumentation (Kuckartz/Rädiker; Flick, Kapitel 27 und 29 in diesem Band) sollte am Ende allerdings derart dargelegt werden, dass deutlich wird, wer wann was getan hat und welche Folgen dies nach sich zog (Baur/Lamnek 2005: 244). Plausibilität, die durchaus in einer gut erzählten Geschichte und mit Hilfe von Illustrationen zum Ausdruck gebracht werden kann, ist demnach Ziel der Darstellung der Analyseergebnisse (Hammersley et al. 2000).

### **37.6.3 Fallinterpretation**

Die Fallinterpretation erfolgt schließlich in Bezug auf die verwendete bzw. zu generierende Theorie (Reichertz, Kapitel 2 in diesem Band) und in Abhängigkeit vom Erkenntnisinteresse. Allerdings sollte der Forscher hier besonders darauf achten, die Beweiskette, also die Verbindung der Forschungsfrage mit dem Material und der Ergebnisdarlegung, klar aufzuzeigen, da sonst die Nachvollziehbarkeit der Erklärungen für andere Personen nicht gegeben ist. Dem vorbeugend sollte eine Forschertriangulation durchgeführt werden (detaillierter dazu siehe typische Probleme).

---

## **37.7 Möglichkeiten zur Verallgemeinerung**

Insbesondere hinsichtlich der Diskussionen um die Verallgemeiner- bzw. Übertragbarkeit stehen Einzelfallanalysen methodologisch meist der interpretativen Tradition in der empirischen Sozialforschung nahe. Verallgemeinerbarkeit wird dabei nicht durch statistisches Schließen von einer Zufallsstichprobe auf eine umfassendere Grundgesamtheit vollzogen, sondern geschieht vielmehr mittels einer Argumentation für abstraktere Kategorien, Begriffe oder Zusammenhänge, welche die Analyse des Einzelfalls nahelegt. Diese sind auf andere Kontexte und Fälle übertragbar. In jeder Analyse sollte dennoch eine Grundgesamtheit für die Übertragung definiert werden, auch wenn diese nur den Einzelfall selbst umfasst. Erst dies ermöglicht es dem Forscher, seinen Fall klar einzuordnen, gleichzeitig abgrenzen zu können und somit Klarheit für Interpretationen oder eventuelle Verallgemeinerungen zu haben. Des Weiteren ist Verallgemeinerbarkeit stets ein Problem des Gegenstandsbereichs. Bei Einzelfällen sind die Aussagen und beschriebenen Strukturen zunächst prinzipiell auf den konkreten Fall beschränkt. In Anlehnung an die eingangs beschriebenen Forschungsinteressen gibt es allerdings drei verschiedenartige Argumentationslinien die Verallgemeinerbarkeit betreffend. In vielen Fällen können und sollten sich diese Überlagern und wechselseitig stützen:

1. *Theoriegeleitete Verallgemeinerung*: Die Erkenntnisse zeigen sich mit Ergebnissen vorhergehender Forschungen oder Theorien einer bestimmten Reichweite kompatibel, so dass plausibel für eine Übertragung auf einen eingegrenzten Gegenstandsbereich argumentiert werden kann oder bestehende Verallgemeinerungen als unzulässig angesehen werden können.

2. *Empirische Typenbildung durch Wechsel von Fallvergleich und Fallkontrastierung* (Kelle/Kluge 2010). Diese Designs bilden eine Art Hybrid, da sie die kausalanalytische Logik standardisierter Verfahren (Kühnel/Dingelstedt, Kapitel 80 in diesem Band) mit der dichten Interpretation einzelner Fälle verbinden. Die Ergebnisse sind durch die kontrollierte Verbindung der Fälle in Bezug auf eingegrenzte Variablen verallgemeinerbar, die variieren oder konstant gehalten werden. In dieser Argumentation ergeben sich Ähnlichkeiten zu quasi-experimentellen Designs (Eifler, Kapitel 11 in diesem Band).
3. *Verallgemeinerung innerhalb des Falls durch die interne Bildung empirischer Zusammenhänge*. Hierbei werden durch die Integration der zahlreichen Facetten des Falls „dichte Theorien“ im Sinne der Grounded Theory (Strübing, Kapitel 32 in diesem Band) gebildet. Man muss sich dabei stets der Rückführbarkeit der Konzepte auf die Daten selbst gewiss sein, die die Plausibilität des Konzepts ausmachen (Eisenhardt 1989). Hierbei gibt es unterschiedliche Herangehensweisen, um die Variationen innerhalb eines Falls zu Konzepten zu verdichten, z.B. die systematische Variation zweier Variablen (Peters 1998), der Wechsel von offenem, axialem und selektivem Kodieren im Sinne der Grounded Theory (Strübing, Kapitel 32 in diesem Band).

---

### 37.8 Zu treffende Entscheidungen im Forschungsverlauf

Die erste Entscheidung bezieht sich auf die *heuristische Abgrenzung eines Falls nach nachvollziehbaren Kriterien* (Akremi, Kapitel 17 in diesem Band). Der Forscher muss eine grobe Vorstellung über den Fall haben, welche je nach Erkenntnisinteresse einen mehr oder weniger vorläufigen Charakter im Hinblick auf seine Eingrenzung hat. Dies hat vielerlei Auswirkungen auf den Forschungsprozess, z.B. welche Fragen beantwortet werden sollen, wie das Forschungsdesign gestrickt wird und welches Datenmaterial Verwendung findet. Das Forschungsinteresse bildet dabei den Ausgangspunkt, kann sich jedoch im Lauf des Forschungsprozesses selbst verändern, was demnach auch eine Anpassung des Falles selbst nach sich zieht (Walton 1992; Baur/Lamnek 2005). Der nächste Schritt ist die *Aufstellung des Forschungsdesigns*. Dieses wird zu einem großen Teil dadurch bestimmt, ob man einen Einzelfall untersucht oder einen Fallvergleich vornimmt. Der Fallvergleich ist allerdings dann vorzuziehen, wenn eine empirisch begründete Verallgemeinerung der Ergebnisse angestrebt wird (Baur/Lamnek 2005: 244).

Neben diesen grundlegenden Entscheidungen sollte weiterhin eine *Forschungsfrage formuliert* werden, die sich entweder aus dem übergreifenden Forschungsdesign ableitet, an relevanten Fragestellungen des Forschungsfeldes orientiert ist oder das neu zu definierende Phänomen in irgendeiner Art und Weise zu fundieren verhilft. Im Design des Forschungsprozesses geht es dann darum, dass alle hilfreichen Datenquellen Verwendung finden können und zudem insbesondere der methodische Mix von verschiedenen Datentypen wie auch der Erhebungs- und Auswertungsverfahren ausgestaltet wird (Yin 2009: 101).

### 37.9 Typische Probleme

Neben einer Vielzahl allgemeiner Probleme der empirischen Sozialforschung ergeben sich in Einzelfallstudien auch einige spezifische:

1. Zu Beginn einer Fallanalyse ist eine *Abgrenzung des Falls* (Akremi, Kapitel 17 in diesem Band) unabdingbar. Diese kann unterschiedlich problematisch sein. Manche Fälle scheinen klar abgrenzbar und dinglich „vorzuliegen“ (z.B. eine Personen oder ein Gegenstand), bei anderen orientiert sich der Forscher an der Semantik, die das Feld vom Fall hat. Für den Forscher könnte daraus allerdings die Gefahr entstehen, sich in einer vorgegebenen Definition (wie sie z.B. das Forschungsfeld vom Fall hat) zu verlieren. Darüber hinaus können Fälle selbst innerhalb des Feldes z.T. äußerst verschieden abgegrenzt und definiert werden. Städte und ihre je spezifischen Stadtgrenzen werden z.B. von den Bewohnern sehr verschieden interpretiert. Ein zu starres Festhalten an einer vordefinierten Abgrenzung des Falls könnte im Forschungsverlauf wichtige Aspekte von der Analyse ausschließen (Vaughan 1992). Der Forscher neigt gegebenenfalls dazu, sich nicht von den Begriffen im Feld, von bestimmten Orten, Personen etc. zu lösen und folgt somit dem Fall mitunter nicht in adäquater Weise. Dies könnte im Extremfall eine für die Einzelfallstudie sinnvolle Verschiebung des Analysefokus verhindern, da man sich z.B. allein auf vorab getätigte Hypothesen fokussieren und dadurch eventuell neue und vorher noch unbedachte Aspekte von der Analyse ausschließen würde. Es gehört daher zu den wichtigsten Aufgaben des Forschers, die Abgrenzungsproblematik in ihren verschiedenen Facetten während des gesamten Prozesses zu reflektieren.
2. Das *Abschweifen vom ursprünglichen Forschungsinteresse* kann, muss aber kein Problem für die (Einzel-)Fallstudie sein. Da ein Wandel – auch mehrfach – des Falles während des Forschungsprozesses nicht untypisch für dieses Forschungsparadigma ist, sollte man sich immer wieder bewusst machen, ob Erkenntnisinteresse und Fallabgrenzung in einem angemessenen Verhältnis zueinander stehen (Walton 1992; Baur/Lamnek 2005).
3. Weiterhin kann die Entscheidung über den *Zeitpunkt des Abbruchs der Analyse*, also wann der Fall detailliert genug beschrieben ist, schwer fallen. Hier kann es hilfreich sein, eine Art der (Forscher-)Triangulation durchzuführen, bei der einerseits dem Forschungsprozess interne Personen, also verschiedene Forscher ihre Ergebnisse abstimmen („investigator triangulation“). Andererseits werden dem Forschungsprozess externe Personen, die als Insider des Feldes fungieren (z.B. die Untersuchten selbst) nach ihrer Einschätzung befragt („member validation“) (Baur/Lamnek 2005: 249). Mit diesen Personen kann der Forscher bzw. die Forschergruppe z.B. folgende Fragen klären: Ist die Forschungsfrage bereits hinreichend beantwortet oder könnten weitere Daten den Erkenntnisgewinn erhöhen? Ist die Ergebnistiefe vergleichbar mit bereits vorhandenen Studienergebnissen? Welche typischen Besonderheiten bzw. Auffälligkeiten zeigen sich im (jeweiligen) Fall und sollte man diesen weiter nachgehen? Sind die angestrebten Verallgemeinerungen und Interpretationen zulässig?

4. In der Fallinterpretation besteht sowohl das Problem der Überinterpretation, also der Übertragung der Ergebnisse auf unzulässige Abstraktionsniveaus, wie der *Unterinterpretation* bzw. Präsentation des rohen Datenmaterials, also der fehlende Bezug zu bestehenden Erkenntnissen über das Feld oder das fehlende Konzeptionalisieren der Daten.

Fallstudien nehmen in der Beantwortung von Forschungsfragen in thematischen Forschungsfeldern, wie im Design einzelner Studien ganz unterschiedliche Positionen ein. Aus diesen Positionen ergeben sich verschiedene Anforderungen an die und Probleme in der praktischen Durchführung, weshalb Sie stets den Ausgangspunkt für methodisches Design und Reflexion bilden.

## Literatur

- Abbott (1992): What Do Cases Do? In: Ragin/Becker (Hg): 53-82
- Aderhold, Jens/Harm, Katrin (Hg.) (2013): Die subjektive Seite der Stadt. Wiesbaden: VS Verlag
- Barbehön, Marlon (2013): Kommunale politische Eliten und die Eigenlogik der Städte. In: Aderhold/Harm (Hg): 273-294
- Baur, Nina (2014): Comparing Societies and Cultures. In: HSR 39 (2): 257-291
- Baur, Nina/Lamnek, Siegfried (2005): Einzelfallanalyse. In: Mikos, Lothar/Wegener, Claudia (Hg): 241-253
- Baur, Nina/Löw, Martina/Hering, Linda/Raschke, Anna Laura/Stoll, Florian (2014): Die Rationalität lokaler Wirtschaftspraktiken im Friseurwesen. In: Bögenhold, Dieter (Hg): 299-327
- Berking, Helmuth/Schwenk, Jochen (2011): Hafenstädte. Frankfurt a.M.: Campus Verlag
- Bögenhold, Dieter (Hg.) (2014): Soziologie des Wirtschaftlichen. Wiesbaden: Springer Fachmedien
- Brähler, Elmar/Adler Corinne (Hg.) (1996): Quantitative Einzelfallanalysen und qualitative Verfahren. Gießen: Psychosozial Verlag
- Eisenhardt, Kathleen (1989): Building Theories from Case Study Research. In: The Academy of Management Review 14 (4): 532-550
- Fricke, Werner (Hg.) (1998): Innovationen in Technik, Wissenschaft und Gesellschaft. Bonn: Friedrich-Ebert-Stiftung
- Gallagher, John (2010): Reimagining Detroit. Detroit: Wayne State University Press
- Gromm, Roger/Hammersley, Martyn/Foster, Peter (Hg.) (2000): Case Study Method. London et al.: Sage
- Häußermann, Hartmut/Kapphan, Andreas (2002): Berlin: von der geteilten zur gespaltenen Stadt? Opladen: Leske + Budrich Verlag
- Hammersley, Martyn/Foster, Peter/Gomm, Roger (2000): Case study and theory. In: Gromm, Roger/Hammersley, Martyn/Foster, Peter (Hg.) (2000): 234-258
- Jahoda, Marie/Lazarsfeld, Paul Felix/Zeisel, Hans (1975 [1933]): Die Arbeitslosen von Marienthal. Frankfurt a.M.: Suhrkamp Verlag
- Kelle, Udo/Kluge, Susanne (2010): Vom Einzelfall zum Typus. Wiesbaden: VS Verlag
- Latour, Bruno (1998): Aramis – oder die Liebe zur Technik. in: Fricke (1998): 147-164
- Löw, Martina/Noller, Peter/Süß, Sabine (Hg.) (2010). Typisch Darmstadt. Frankfurt a.M./ New York: Campus
- Marent, Johannes/Richter, Ralph (2013): Jede Stadt eine eigene Welt. In: Aderhold/Harm (2013): 59-80
- Mikos, Lothar/Wegener, Claudia (Hg.) (2005): Qualitative Medienforschung. Konstanz: UVK
- Peters, B. Guy (1998): Comparative Politics. Hounds Mills: Macmillan

- Ragin, Charles/Becker, Howard S. (Hg.) (1992): What is a case? New York: Cambridge University
- Stake, Robert E. (1995): The art of case study research. Thousand Oaks: Sage
- Solga, Heike/Powell, Justin/Berger, Peter A. (Hg.) (2009): Soziale Ungleichheit. New York: Campus Verlag
- Thomas, Gary (2011): A Typology for the Case Study in Social Science Following a Review of Definition, Discourse, and Structure. In: Qualitative Inquiry 17 (6): 511-521
- Vaughan, Diane (1992): Theory elaboration. In: Ragin/Becker (Hg): 173-203
- Walton, John (1992): Making the theoretical case. In: Ragin/Becker (Hg): 112-137
- Yin, Robert K. (2009): Case study research. Thousand Oaks: Sage

Philipp Mayring und Thomas Fenzl

---

## 38.1 Was ist qualitative Inhaltsanalyse?

Qualitative Inhaltsanalyse stellt eine *Auswertungsmethode dar, die Texte bearbeitet*, welche im Rahmen sozialwissenschaftlicher Forschungsprojekte in der Datenerhebung anfallen, z.B. Transkripte von offenen Interviews (Helfferich, Kapitel 39 in diesem Band) oder Fokusgruppen (Vogl, Kapitel 41 in diesem Band), offene Fragen aus standardisierten Befragungen (Züll/Menold, Kapitel 52 in diesem Band), Beobachtungsprotokolle aus Feldstudien (Thierbach/Petschick, Kapitel 66 in diesem Band), Dokumente (Ernst, Kapitel 63 in diesem Band), Akten (Salheiser, Kapitel 62 in diesem Band), Zeitungsartikel (Klein, Kapitel 64 in diesem Band) und Internetmaterialien.

Im Vergleich mit anderen textanalytischen Methoden wie z.B. der Grounded Theory (Strübing, Kapitel 32 in diesem Band), wissenssoziologischen Hermeneutik (Kurt/Herbrik, Kapitel 33 in diesem Band), Biographieforschung (Rosenthal, Kapitel 35 in diesem Band) oder der Ethnografie (Knoblauch, Kapitel 36 in diesem Band) wird sie, so die bibliographische Analyse von Titscher et al. (2000), am häufigsten angewandt. Mit der qualitativen Inhaltsanalyse steht ein Verfahren qualitativ orientierter Textanalyse zur Verfügung, das mit dem technischen Know-how der quantitativen Inhaltsanalyse (Quantitative Content Analysis) große Materialmengen bewältigen kann, dabei aber im ersten Schritt qualitativ-interpretativ bleibt und so auch latente Sinngehalte erfassen kann. Das Vorgehen ist dabei streng regelgeleitet und damit stark intersubjektiv überprüfbar, wobei die inhaltsanalytischen Regeln auf psychologischer und linguistischer Theorie alltäglichen Textverständnisses basieren.

### 38.1.1 Begriffsbestimmung

Für eine genauere Begriffsbestimmung der qualitativen Inhaltsanalyse sollte man allerdings präzisieren, was sich auf beide Wortbestandteile bezieht:

- *Inhaltsanalyse* suggeriert, dass es bei dieser Art von Textanalyse nur um „Inhalte“ ginge, oder wie der erste Lehrbuchautor der „Content Analysis“, Bernhard Berelson (1952) formulierte, um manifesten Inhalt. In der Tat ist die Inhaltsanalyse in der ersten Hälfte des 20. Jahrhunderts entwickelt worden, um in den aufkommenden Massenmedien (Zeitung, Radio) (Klein, Kapitel 64 in diesem Band) quantitative Analysen inhaltlicher Aspekte (z.B. Themen, Rubriken) vornehmen zu können. Schon bald wurde auch innerhalb dieser kommunikationswissenschaftlichen Methodenverwendungen ein erweitertes Verständnis entwickelt, das auch latente Sinngehalte und subjektive Bedeutungen zum Ziel der Analyse machte. Neben dieser Erweiterung in Richtung tieferliegender Sinngehalte des Textes ist auch eine Erweiterung in Richtung formaler Textcharakteristika zu verzeichnen (z.B. Stilanalyse, Wortflussanalysen).
- *Qualitative Inhaltsanalyse* suggeriert, dass es ausschließlich um qualitativ-interpretatives Umgehen mit Text ginge. Im Regelfall werden jedoch auch in der qualitativen Inhaltsanalyse Kategorienhäufigkeiten eruiert und statistisch analysiert (Züll/Menold und Kuckartz/Rädicke, Kapitel 52 und 27 in diesem Band). Das Vorgehen besteht prinzipiell aus zwei Schritten. In einem ersten Schritt werden induktiv am Material entwickelte oder vorab theoriegeleitet-deduktiv postulierte (Reichert, Kapitel 2 in diesem Band) Kategorien einzelnen Textpassagen zugeordnet. Dieser Prozess wird zwar von genauen inhaltsanalytischen Regeln begleitet, er bleibt aber ein qualitativ-interpretativer Akt. In einem zweiten Schritt wird dann analysiert, ob bestimmte Kategorien mehrfach Textstellen zugeordnet werden können.

Aus diesem Grund wäre die Bezeichnung „*qualitativ orientierte kategoriengeleitete Textanalyse*“ wohl der bessere Begriff.

### 38.1.2 Abgrenzung von anderen Verfahren

Das Merkmal der *Kategoriengeleitetheit* ist in der Tat das zentrale Unterscheidungskriterium gegenüber anderen Textanalyseansätzen. *Kategorien* stellen Analyseaspekte als Kurzformulierungen dar, sind in der Formulierung mehr oder weniger eng am Ausgangsmaterial orientiert und können hierarchisch geordnet sein (*Ober- und Unterkategorien*). So werden im unten angeführten Beispiel Nennungen von gesundheitsfördernden Maßnahmen im Betrieb wie Umfragen und Gesundheitschecks zur Oberkategorie „Evaluation“ zusammengefasst. Das *Kategoriensystem* (als die Zusammenstellung aller Kategorien) ist das eigentliche Instrumentarium der Analyse. Mit ihm wird das Material bearbeitet und nur die Textstellen berücksichtigt, die sich auf die Kategorien beziehen.

In ihrer Formulierung können Kategorien den „Codes“ in der *Grounded Theory* (Strübing, Kapitel 32 in diesem Band) durchaus ähnlich erscheinen. Während dort die Codes zunächst in einem explorativen Akt („offenes Kodieren“) theoriegenerierend aus dem Material heraus entwickelt werden, vollzieht sich die Zuordnung Textstelle – Kategorie in der qualitativen Inhaltsanalyse jedoch *streng regelgeleitet*.

Anderer qualitative Textanalyseansätze (z.B. Objektive Hermeneutik, Sozialwissenschaftlich-hermeneutische Paraphrase, Psychoanalytische Textinterpretation) sind hier viel offener als die qualitative Inhaltsanalyse und verlangen keine Auswertungskategorien oder Codes. Der Grundvorgang besteht auch dort in der Kennzeichnung von Textstellen, die Auswertung geht aber eher in eine interpretative Erweiterung dieser Textstellen. Durch intensives Deuten der Textstellen, im Sinne der *Hermeneutik* (Kurt/Herbrik, Kapitel 33 in diesem Band) als mehrfaches Herantragen des eigenen Vorverständnisses an den Text und schrittweiser Interpretation der dahinterliegenden Bedeutungen wird das Material intensiv bearbeitet. *Demgegenüber definiert die qualitative Inhaltsanalyse sowohl die Auswertungsaspekte als auch die Auswertungsregeln so exakt, dass ein systematisches, intersubjektiv überprüfbare Durcharbeiten möglich wird.* Das ist dann auch der Grund, weshalb sie, der Tradition der „Content Analysis“ folgend, an *großen Textmengen* angewendet werden kann, während hermeneutisch-interpretative Analysen in der Regel mit kleineren Textmengen arbeiten.

Dadurch und durch die Breite der unterschiedlichen Analysetechniken unterscheidet sie sich auch von eher intuitiven Verfahren im Umkreis der „Content Analysis“. So wird in der „Codebook Analysis“ (Neuendorf 2002) ein Inventar verwendet, das Kategoriennamen und Kurzdefinitionen enthält, allerdings ohne die Genauigkeit des in der qualitativen Inhaltsanalyse für ähnliche Zwecke verwendeten sog. Kodierleitfaden (siehe unten). Auch die „Thematic Text Analysis“ (Braun/Clarke 2006) ist hier zu nennen, die im Material inhaltliche Gegenstandsbereiche erfassen und auszählen will und bei der Suche nach den zentralen Themen entweder mit theoretischen Vorgaben arbeitet oder sich an Worthäufigkeitslisten und Wortkombinationshäufigkeiten orientiert.

---

## 38.2 Grundprinzipien kategoriengeleiteter qualitativ orientierter Textanalyse

Die qualitativ-inhaltsanalytischen Techniken wurden in den 1980ern in einem Forschungsprojekt entwickelt, in dem offene Interviews (Helfferich, Kapitel 39 in diesem Band) mit arbeitslosen Lehrerinnen und Lehrern im Längsschnitt über ein Jahr hinweg durchgeführt wurden (mehr als 500 Interviews), was zu Transkripten von über 10.000 Seiten geführt hat (Ulich et al. 1985). Gesucht war ein Verfahren, das mit solchen Datenmengen umgehen kann (was die quantitative Inhaltsanalyse auch leistet), aber trotzdem interpretative Züge trägt und nicht nur in einem Auszählen von manifesten Textbestandteilen besteht. Die Grundprinzipien der Methodik, die auf diesem Hintergrund entwickelt wurden, sind folgende (Mayring 2010):

- *Das Material wird in ein Kommunikationsmodell eingeordnet*, in dem Textproduzenten, sozio-kultureller Hintergrund, Textproduktionssituation, Textwirkungen und Zielgruppen des Textes eruiert werden und festgelegt wird, auf welche Teile dieses Kommunikationsmodells durch die Textanalyse Aussagen getroffen werden sollen. Es geht also nicht um eine Textanalyse per se, sondern um Schlussfolgerungen über den Text hinaus. Im oben angeführten Projekt beispielsweise sollte von Interviewtranskripten auf den psychischen Zustand der Interviewten geschlossen werden.
- *Die qualitative Inhaltsanalyse geht streng regelgeleitet vor*, statt „freie“ Interpretationen zu kreieren. Sie folgt einem für die jeweilige Analyse zusammengestellten, an den Grundformen der Inhaltsanalyse orientierten *Ablaufmodell* (s.u.). Für die einzelnen Analyseschritte wurden weitere inhaltsanalytische Regeln formuliert.
- Die *Systematik* der Inhaltsanalyse zeigt sich auch darin, dass vorab *Analyseeinheiten* definiert (Akremi, Kapitel 17 in diesem Band) werden (siehe auch unten das Ablaufmodell): die *Kodiereinheit* legt den minimalen Textbestandteil fest, der ausgewertet werden darf (semantische Einheit, Wort, Satz usw.), die *Kontexteinheit* bestimmt, welche Informationen für die einzelne Kodierung herangezogen werden dürfen (Satz, Absatz, Interviewantwort, ganzes Interview, Zusatzkontextmaterial) und die *Auswertungseinheit* definiert die Materialportion, der ein Kategoriensystem gegenübergestellt wird (ganzes Material, Materialteile, Mehrfachkodierungen usw.).
- *Der Grundvorgang besteht nun in der regelgeleiteten Zuordnung von (entweder induktiv am Material entwickelten oder deduktiv vorab aufgestellten) Kategorien zu konkreten Textstellen*, den inhaltsanalytischen Zuordnungsregeln folgend. Diese Regeln werden in einer Pilotphase an das Material angepasst und im Sinne eines zirkulären Modells verfeinert, wonach aufgrund der ersten Auswertungserfahrungen die Struktur der Analyse noch verändert werden kann. Wenn es hier zu Anpassungen gekommen ist, müssen die Regeln jedoch im endgültigen Materialdurchgang konstant gehalten werden (*lineares Modell*).
- *Wie auch in der quantitativen Inhaltsanalyse sollten vorab die inhaltsanalytischen Einheiten festgelegt werden* (vgl. auch Akremi, Kapitel 17 in diesem Band). Die Kodiereinheit bestimmt dabei – wie bereits oben erwähnt – was der minimale Textbestandteil sein kann, der einer Kategorie zugeordnet werden kann, und bestimmt damit auch die Sensibilität der Analyse. Die *Kontexteinheit* legt das Material fest, auf das für die jeweilige Kodierung zurückgegriffen werden darf. Die *Auswertungseinheit* definiert die Textportionen, die dem Kategoriensystem jeweils gegenübergestellt werden (z.B. Satz, Absatz, Antwort auf Interviewfrage, ganzes Interviewtranskript, einzelnes Dokument).
- *Dadurch kann gewährleistet werden, dass inhaltsanalytische Gütekriterien sinnvoll überprüfbar werden*. Zentrale Gütekriterien (Krebs/Menold und Flick, Kapitel 30 und 29 in diesem Band) sind dabei die Intrakoderübereinstimmung und die Interkoderübereinstimmung. Die *Intrakoderübereinstimmung* wird dadurch bestimmt, dass nach Abschluss der Analyse noch einmal von Beginn an ausgewertet wird, ohne die vorher zugeordneten Kategorisierungen anzusehen. Sie ist ein Indikator für die Stabilität des Vorgehens und damit ein Reliabilitätsmaß. Für die Bestimmung der

*Interkoderübereinstimmung* wird ein zweiter Kodierer herangezogen. Übereinstimmungen sind ein Kriterium für die Objektivität (Krebs/Menold, Kapitel 30 in diesem Band) des Verfahrens. Für die qualitative Inhaltsanalyse reicht dabei oft schon eine ausschnittsweise Überprüfung. Ziel kann auch nicht völlige Übereinstimmung sein, da die interpretativen Bestandteile einen gewissen Spielraum bedingen, besonders bei induktiver Kategorienbildung (s.u.). Dafür besteht die Möglichkeit, die gefundenen Nicht-Übereinstimmungen zu diskutieren, über die angemessene Kodierung zu entscheiden und im Datensatz zu bereinigen.

---

## 38.3 Techniken qualitativer Inhaltsanalyse

Es sind eine Reihe von konkreten qualitativ-inhaltsanalytischen Techniken differenziert worden, die sich an den Grundvorgängen der Zusammenfassung, Explikation und Strukturierung orientieren (Mayring 2010). Insgesamt sind verschiedene Techniken entwickelt worden:

### 38.3.1 Zusammenfassende Inhaltsanalyse

Bei der *zusammenfassenden Inhaltsanalyse* wird das Material zunächst in die inhaltstragenden Paraphrasen umformuliert und dann schrittweise, den Makrooperatoren der Zusammenfassung (siehe die Ausführungen im nächsten Abschnitt) folgend, reduziert. Wenn der erste Schritt der Paraphrasierung übersprungen und ein Selektionskriterium eingeführt wird, über welche Textbestandteile die Zusammenfassung laufen soll, so wird von *induktiver Kategorienbildung* gesprochen. Im übernächsten Abschnitt werden wir dies am Beispiel verdeutlichen. Zentral sind hier die inhaltsanalytischen Regeln der *Kategoriendefinition* (Über welche Aspekte sollen Kategorien formuliert werden?) und des *Abstraktionsniveaus* (Wie allgemein sollen die Kategorien formuliert werden?). Nach einem ersten Durchgang der Kategorienbildung können die Kategorien schrittweise zu Hauptkategorien nach den Regeln der Zusammenfassung generalisiert werden.

### 38.3.2 Explikation

Bei der Explikation werden einzelne unklare Textstellen zum Gegenstand gemacht. Von der Zielrichtung her entspricht dies dem hermeneutischen Vorgehen (Kurt/Herbrik, Kapitel 33 in diesem Band), läuft jedoch regelgeleiteter ab. Bei der Explikation wird zwischen einer *engen* und einer *weiten Kontextanalyse* unterschieden, je nachdem, welches explikative Material zugelassen wird. In einem ersten Durchgang wird das für die Explikation der Textstelle heranzuziehende relevante Material identifiziert, in einem zweiten Durchgang nach den Regeln der Zusammenfassung komprimiert.

### 38.3.3 Strukturierende Inhaltsanalyse

Bei den strukturierenden qualitativen Inhaltsanalysen handelt es sich um *deduktive Kategorienanwendungen*, bei denen das Kategoriensystem vorab theoriegeleitet entwickelt und dann an den Text herangetragen wird. Hier unterscheiden wir zwischen *einfachen Kategorienlisten* (nominales Skalenniveau) und *ordinal geordneten Kategoriensystemen* (z.B. viel – mittel – wenig). Mit beiden können dann komplexere quantitative Analysen durchgeführt werden. Zentrales Hilfsmittel stellt hier der *Kodierleitfaden* dar, der für jede Kategorie eine Definition, typische Textpassagen als Ankerbeispiele und Kodierregeln zur Abgrenzung zwischen den Kategorien enthält (zur Begründung siehe die Ausführungen im nächsten Abschnitt). Der Kodierleitfaden wird zunächst theoriegeleitet entwickelt und in einer Pilotphase am Material weiter ausgebaut und ergänzt. Die Regeln werden in Tabellenform zusammengestellt. Ein Beispiel gibt Abb. 38.1 wieder. Anhand des Kodierleitfadens wird dann bestimmt, wann welche Kategorie einer Textstelle zuzuordnen ist.

### 38.3.4 Ablaufmodell

Jede dieser Techniken ist mit eigenen Ablaufmodellen beschrieben worden. Ein allgemeines Ablaufmodell, das die induktive und deduktive Vorgehensweise vereinigt, haben wir in Abb. 38.2 dargestellt. Dieses wird im Rahmen einer Beispielstudie an späterer Stelle im Text erläutert. Es kommt nun darauf an, in Abhängigkeit der jeweiligen konkreten Fragestellung die adäquate Technik auszuwählen. Auch Kombinationen sind dabei möglich (für Beispiele siehe Mayring/Gläser-Zikuda 2008).

---

## 38.4 Theoretischer Hintergrund der inhaltsanalytischen Regeln: Alltagsoperationen der Textverarbeitung

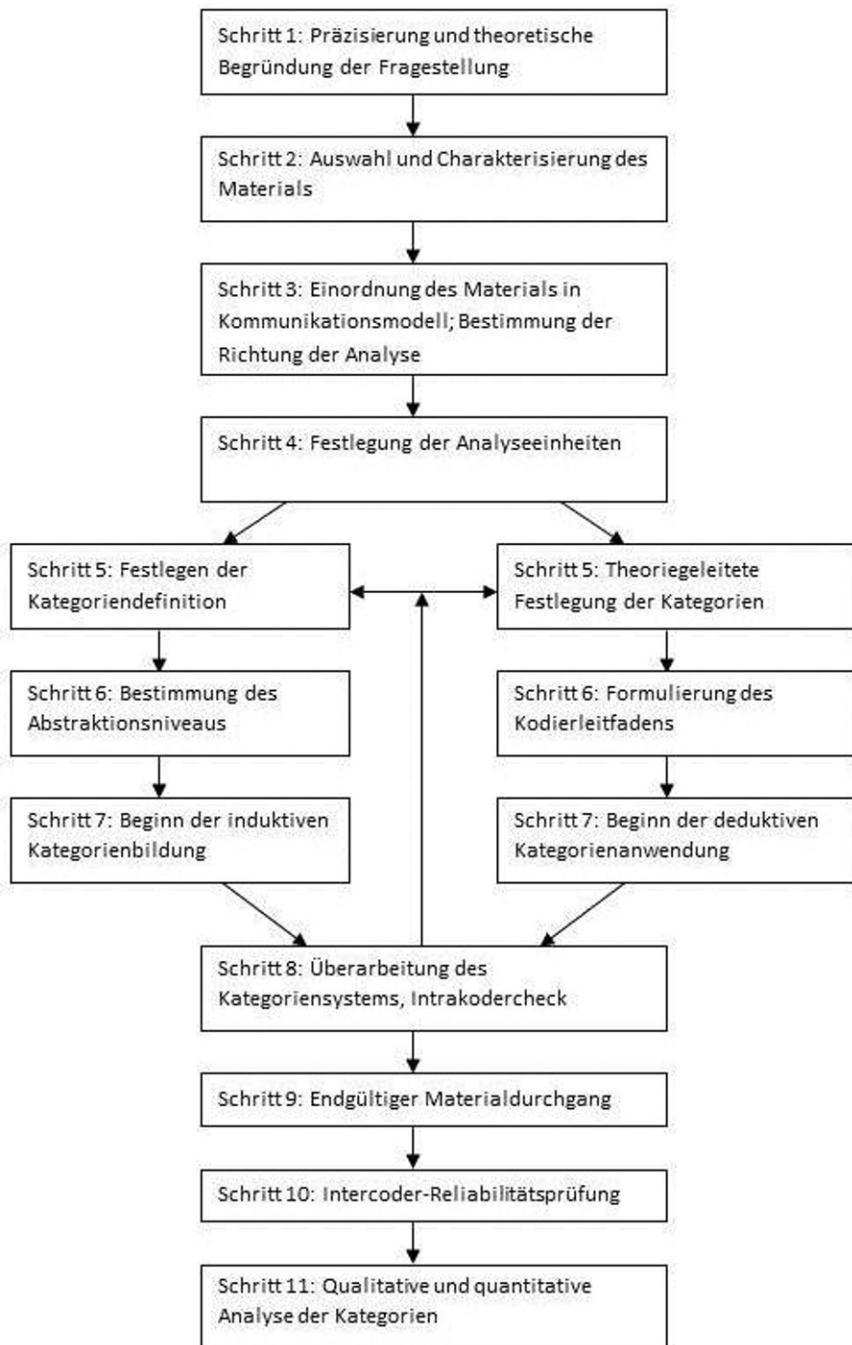
Gerade in der Soziologie wurde immer wieder behauptet, dass die qualitative Inhaltsanalyse kein elaboriertes oder anspruchsvolles Verfahren sei, da ihre Verfahrensweisen zu wenig theoretisch begründet seien. So zählt Reichertz (2007: 198) sie nicht zu den qualitativen Methoden, denn diese „entwickeln vorab nun eine mehr oder weniger explizite Grundlagentheorie“. Dabei übersieht Reichertz, dass die grundlegenden Vorgänge der qualitativen Inhaltsanalyse an empirisch-psychologischen Studien zur Alltagsverarbeitung von Text anknüpfen. So wird für zusammenfassende Analysen auf die Psychologie der Textverarbeitung zurückgegriffen, in der untersucht wird, wie Menschen im Alltag systematisch Texte zusammenfassen. Es wurden Studierende mit der Aufgabe betraut, Texte zusammenzufassen und dabei laut zu denken, welche kognitiven Operationen sie anleiten. Dabei sind folgende reduktive Prozesse differenziert worden: Auslassen, Generalisation, Konstruktion, Integration, Selektion und Bündelung (van Dijk 1980). Auf diese Makrooperatoren reduktiver Textverarbeitungsprozesse wird in den inhaltsanalytischen Regeln

Kategorie	Definition	Ankerbeispiele	Kodierregeln
<b>K1: hohes Selbst- vertrauen</b>	Hohe subjektive Gewissheit, mit der Anforderung gut fertig geworden zu sein, d.h. - Klarheit über die Art der Anforderung und deren Bewältigung, - Positives, hoffnungsvolles Gefühl beim Umgang mit der Anforderung, - Überzeugung, die Bewältigung der Anforderung selbst in der Hand gehabt zu haben.	„Sicher hat's mal ein Problemchen gegeben, aber das wurde dann halt ausgeräumt, entweder von mir die Einsicht, oder vom Schüler, je nachdem, wer den Fehler gemacht hat. Fehler macht ja ein jeder.“ (17,23) „Ja klar, Probleme gab's natürlich, aber zum Schluss hatten wir ein sehr gutes Verhältnis, hatten wir uns zusammengerauft.“ (27,33)	Alle drei Aspekte der Definition müssen in Richtung „hoch“ weisen, es soll kein Aspekt auf nur mittleres Selbstvertrauen schließen lassen. Sonst Kodierung „mittleres S.“
<b>K2: mittleres Selbst- vertrauen</b>	Nur teilweise oder schwankende Gewissheit, mit der Anforderung gut fertig geworden zu sein.	„Ich hab mich da einigermaßen durchlaviert, aber es war oft eine Gratwanderung.“ (3,55) „Mit der Zeit ist es etwas besser geworden, aber ob das an mir oder an den Umständen lag. Weiß ich nicht.“ (77, 20)	Wenn nicht alle drei Definitionsaspekte auf „hoch“ oder „niedrig“ schließen lassen.
<b>K3: niedriges Selbst- vertrauen</b>	Überzeugung, mit der Anforderung schlecht fertig geworden zu sein, d.h. - wenig Klarheit über die Art der Anforderung, - negatives, pessimistisches Gefühl beim Umgang mit der Anforderung, - Überzeugung, den Umgang mit der Anforderung nicht selbst in der Hand gehabt zu haben.	„das hat mein Selbstvertrauen getroffen; da hab ich gemeint, ich bin eine Null – oder ein Minus.“ (5, 34)	Alle drei Aspekte deuten auf niedriges Selbstvertrauen, auch keine Schwankungen erkennbar.

**Abb. 38.1** Beispiel für einen Kodierleitfaden (Ulich et al. 1985)

zurückgegriffen (Mayring 2010). Durch die Inhaltsanalyse soll also an Alltagsprozesse des Textverständens und Textverarbeitens angeknüpft werden.

Ähnliches gilt für die Prozesse deduktiver Kategorienanwendung (Reichert, Kapitel 2 in diesem Band) in der qualitativen Inhaltsanalyse. Im Zentrum der Prozesse steht der Kodierleitfaden, der - wie oben erläutert - jede einzelne Kategorie in drei Schritten festlegt, nämlich durch die Kategoriendefinition, die Ankerbeispiele und die Kodierregeln. Auch hier kann man auf Alltagsoperationen der Kategorisierung zurückgreifen. Im Rahmen der Allgemeinen Psychologie wurde untersucht, wie, auch im menschlichen Entwicklungsprozess, Kategorien gebildet und mit Sinn gefüllt werden. Verschiedene Ansätze sind dabei unterschiedbar (Murphy 2002): Die klassische Sicht der *Kategorienrepräsentation*



**Abb. 38.2** Ablaufmodell induktiver Kategorienbildung und deduktiver Kategorienanwendung (Mayring/Brunner 2006)

postuliert, dass Kategorien durch definierende Merkmale festgelegt sind. Die *Prototypentheorie* (Murphy 2002) geht von typischen Merkmalen aus, die Kategorien definieren. Die *Theorie der Entscheidungsgrenzen* (Murphy 2002) schließlich geht von fließenden Kategorienübergängen aus und legt durch Regeln fest, was gerade noch zu welcher Kategorie gehört. Diese drei Ansätze entsprechen nun genau den drei Spalten des *inhaltsanalytischen Kodierleitfadens: Definition, Ankerbeispiele und Kodierregeln* zur Abgrenzung. Neuere Entwicklungen in der Kategorisierungsforschung zeigen nun, gestützt auch durch neurowissenschaftliche Studien (Pritzel/Markowitsch, Kapitel 78 in diesem Band), dass mehrere Kategorisierungsprozesse gleichzeitig laufen, wenn wir beobachtete Exemplare Kategorien zuordnen (*Theorie multipler Systeme*), und genau das macht der Kodierleitfaden. Qualitative Inhaltsanalyse kann sich in ihren Regeln somit auf psychologische und linguistische Konzepte alltäglicher Textverarbeitung berufen.

---

### 38.5 Die qualitative Inhaltsanalyse als Mixed-Methods-Ansatz

Weiterhin kann gesagt werden, dass die qualitative Inhaltsanalyse sich zu den in den letzten Jahren in der Methodendiskussion immer bedeutsamer gewordenen Mixed-Methods-Ansätzen (Mayring et al. 2007; Kelle, Kapitel 8 in diesem Band) zuordnen lässt. Diese im Wesentlichen auf Denzins Ausarbeitung der Triangulation (Flick, Kapitel 29 in diesem Band) zurückgehende Methodologie hat zu den verschiedensten Modellen der Kombination und Integration qualitativer und quantitativer Analyseschritte geführt. Sie erlaubt induktive wie deduktive Analysevorgänge, ein idiographisch auf einzelne Textstellen Eingehen (Explikation) wie auch ein generalisierend-quantitatives Vorgehen, deskriptive wie auch messende Analysen (z.B. Skalierung, Latechva/Davidov und Blasius, Kapitel 55 und 83 in diesem Band).

---

### 38.6 Eine Beispielstudie

Zur weiteren Vertiefung und Unterstützung von Forschungsprojekten, die mit qualitativer Inhaltsanalyse arbeiten wollen, haben wir Materialien im Internet zusammengestellt ([www.qualitative-content-analysis.aau.at](http://www.qualitative-content-analysis.aau.at)) sowie eine spezielle Software entwickelt ([www.qcamap.org](http://www.qcamap.org)), die am Computer in interaktiver Form durch die einzelnen Schritte der qualitativen Inhaltsanalyse führt. Diese Webapplikation steht frei zur Verfügung (Open Access).

Außerdem soll abschließend anhand eines Praxisbeispiels, das sich mit betrieblicher Gesundheitsförderung (BGF) beschäftigt (Fenzl/Mayring 2009: 60-87), die konkrete, praktische Vorgehensweise bei der inhaltsanalytischen Technik der induktiven Kategoriengenbildung verdeutlicht werden.

Im Rahmen dieses Projekts wurden Auszeichnungen an Unternehmen verliehen, welche als Vorzeigbeispiele für gelungene Gesundheitsförderung am Arbeitsplatz dienen können. Um ein tieferes Verständnis für die in der Praxis relevanten und entscheidenden

Kriterien für eine effiziente (wirtschaftliche) und effektive (wirksame) BGF zu erlangen, wurden vorselektierten Unternehmen, Organisationen und Institutionen quer durch Europa, welche zuvor als Best-Practice-Modelle nach den bisherigen Qualitätsstandards eingestuft wurden, halbstandardisierte Fragebögen mit offenem Antwortformat (Reuband und Züll/Menold, Kapitel 47 und 52 in diesem Band) zugesandt. Die jeweils für BGF im Unternehmen bzw. in der Organisation verantwortlichen Abteilungen bzw. Personen (z.B. Personalabteilung, Betriebspyschologen, Betriebsärzte) wurden darin gebeten, neben einigen Auskünften über die Größe des Betriebs bzw. den Mitarbeiterstand insbesondere Informationen zur Integration der BGF in die Unternehmenskultur und -politik zur Verfügung zu stellen und konkrete Angaben zu den Maßnahmen zur (1) Vorbeugung von Rauchen sowie zur Förderung (2) gesunder Ernährung, (3) körperlicher Aktivität (Bewegung) sowie (4) mentaler Gesundheit (Stress) zu machen. In diesen vier Gesundheitsbereichen sollten einerseits die Organisation der Maßnahmen sowie die geschaffenen Strukturen und andererseits die Strategie und Umsetzung der Maßnahmen beschrieben werden. In Ergänzung dazu wurde auch nach der Evaluation und den Ergebnissen der bisherigen Maßnahmen und Interventionen gefragt. Insgesamt standen 65 halbstandardisierte offene Fragebögen in deutscher bzw. englischer Sprache aus neun Ländern für die durchzuführende Datenanalyse zur Verfügung, welche die Frage nach den Gemeinsamkeiten der Konzepte und Strategien zur BGF, die sich in der Praxis bewähren, beantworten sollte. Der konkrete Auftrag für die Analyse lautete damit, die entscheidenden Faktoren und wesentlichen Aspekte sowie die zentralen Maßnahmen für gelungene BGF in den oben genannten Bereichen herauszuarbeiten.

Basierend auf dieser Fragestellung und dem vorliegenden Datenmaterial ist die Interpretationstechnik der *Zusammenfassung*, konkret die systematische Kategorisierung des Materials mit *induktiver Kategorienbildung*, naheliegend und zweckmäßig (Mayring 2010: 59-77). Die zu analysierenden fünf Bereiche der Gesundheitsförderung (Integration in Unternehmenskultur, Rauchen, Ernährung, Bewegung/körperliche Aktivität, Stress/mentale Gesundheit) können dabei als „*deduktive*“ *Oberkategorien* verstanden werden, die aufgrund der Struktur des Fragebogens vorgegeben waren. Diese Oberkategorien wurden im Zuge der Auswertung mit induktiv entwickelten Kategorien aufgefüllt.

Wenn wir uns das Ablaufdiagramm der *induktiven Kategorienbildung* in Erinnerung rufen (Abb. 38.2), sieht die weitere Vorgehensweise vor, dass bevor die Arbeit am Material beginnen kann, die Kategoriendefinition sowie die Analyseeinheiten und das Abstraktionsniveau festzulegen sind. Im konkreten Beispiel wurden nach einer ersten Sichtung des Materials und in Anlehnung an die Fragestellung zwei Hauptkategorien als Kategoriendefinition formuliert. Kategorie A sollte die Beschreibung zentraler Elemente der förderlichen Maßnahmen bei der Implementierung der Gesundheitsförderung umfassen. In Kategorie B sollten die entscheidenden Faktoren und wesentlichen Aspekte für eine erfolgreiche Umsetzung von BGF erfasst werden. Das Material war also in allen fünf deduktiven Oberkategorien (Integration von BGF in Unternehmenskultur, Rauchen, Ernährung, Bewegung, mentale Gesundheit) hinsichtlich dieser beider Hauptkategorien, welche als

Kategoriendefinitionen (Hauptkategorie A und B) für die Kategorienbildung dienten, zu analysieren.

Was das Abstraktionsniveau (Akremi, Kapitel 17 in diesem Band) betrifft, so sollten alle konkreten Äußerungen im Material zu spezifischen Maßnahmen und deren Aspekten für die BGF in induktiven Kategorien kodiert werden. Um die Präzision der Inhaltsanalyse zu erhöhen, war in weiterer Folge die Analyseeinheit festzusetzen:

- Als *Kodiereinheit* (der kleinste auszuwertende Materialbestandteil) wurden dabei mehrere Wörter mit Sinnzusammenhang – bedeutungstragende Phrasen – in Richtung einer der beiden Hauptkategorien festgesetzt.
- Die *Kontexteinheit* (der größte in eine Kategorie fallende Textbestandteil), wurde mit dem gesamten Antworttext zu einem der fünf Fragebogenthemen (Integration in Unternehmenskultur, Rauchen, Ernährung, Bewegung/körperliche Aktivität, Stress/mentale Gesundheit) definiert, beispielsweise also die gesamte Antwort auf die Frage zu den Maßnahmen zur Förderung gesunder Ernährung.
- Die *Auswertungseinheit* (welche Texte nacheinander ausgewertet werden) stellten schließlich alle Fragebögen dar.

Dabei waren Mehrfachzuordnungen von Materialbestandteilen, welche unterschiedliche Aspekte aufzeigten, zu unterschiedlichen Kategorien zulässig, Mehrfachnennungen von Kategorien in einem Fragebogen wurden jedoch nicht gezählt.

Unter Berücksichtigung der Festlegungen zu den Selektionskriterien und zum Abstraktionsniveau des Kategoriensystems sowie zur Analyseeinheit konnte nun mit der schrittweisen Bildung der induktiven Kategorien aus dem Material begonnen werden. Konkret wurden dabei im ersten Durchlauf für alle fünf Oberkategorien induktive Kategorien aus dem Text entwickelt und die Häufigkeiten der Nennungen der Kategorien notiert.

Bei der Überarbeitung der Kategorien im zweiten Durchlauf nach der Analyse von etwa einem Drittel aller Fragebögen wurden Kategorien, welche sich auf einen ähnlichen Gegenstand bezogen, zusammengefasst und reduziert. Außerdem wurde aufgrund der zahlreichen Quernennungen von Ausprägungen über alle Oberkategorien, insbesondere die vier Gesundheitsbereiche, hinweg eine weitere Oberkategorie für „allgemeine Punkte“ entwickelt. Die darin enthaltenen induktiv formulierten Kategorien zu den förderlichen Maßnahmen und entscheidenden Faktoren, wie etwa die „Bereitstellung einer Auswahl vielfältiger Angebote“ oder die „laufende Fortbildungen zum Thema im eigenen Haus“, besaßen Gültigkeit für alle vier Bereiche der betrieblichen Gesundheitsförderung.

Wir wollen nun für den Gesundheitsbereich und die Oberkategorie „Ernährung“ einige konkrete Beispiele für die induktive Kategorienbildung geben. Die Textstelle „*In der Kantine werden spezielle Diäten, z.B. fettarme Kost, Brain Food etc. angeboten*“ beschreibt eine Maßnahme zur Förderung der gesunden Ernährung und ist deshalb entsprechend der Kategoriendefinition in die Kategorie A einzuordnen. Sie wurde im ersten Durchlauf mit „FA6: Angebot verschiedener Diäten in der Kantine“ kodiert. Im zweiten Durchlauf wurde die Fundstelle dann aufgrund der Zusammenfassung und Reduktion mit „FA6: vielfältiges Angebot gesunder Ernährung im Betrieb“ kodiert. Demgegenüber beschreibt

der Materialbestandteil „... um den Mitarbeiter/innen die Bedeutung von gesunder Ernährung bewusst zu machen“ einen entscheidenden Faktor für den Erfolg zur Umsetzung der Gesundheitsförderung im Bereich Ernährung und erfüllt damit die Kategoriendefinition der Hauptkategorie B. Die Textstelle wurde daher im ersten Durchlauf mit „FB1: Bedeutung von gesunder Ernährung bewusst machen“ kodiert und im zweiten Durchlauf auf „FB1: Bewusstsein zum Thema gesunde Ernährung schaffen“ reduziert.

In der dritten Durchlaufphase wurden dann alle gefundenen Kategorien innerhalb der sechs Oberkategorien (Integration in Unternehmenskultur, generelle Punkte, Rauchen, Ernährung, Bewegung, Stress/mentale Gesundheit) nochmals zu folgenden sechs, auf einem höheren Abstraktionsniveau liegenden „Themengebieten“ bzw. Überkategorien zugeordnet:

1. *Ressourcen*: Kategorien betreffend Materialien, zeitliche/finanzielle Ressourcen etc.
2. *Aus- und Weiterbildung*: Kategorien, welche Trainingsmaßnahmen, Wissensvermittlung und -aufbereitung sowie den Informationsfluss thematisierten.
3. *Struktur*: Kategorien zum Umgangs mit dem Thema, der verfügbaren Infrastruktur (z.B. koordinierende Arbeitsgruppen), der Einbindung von Mitarbeiter/innen etc.
4. *Strategie und Umsetzung*: Kategorien zu konkreten Konzepten, Methoden, Maßnahmen und Angeboten der betrieblichen Gesundheitsförderung
5. *Evaluation*: Kategorien betreffend Umfragen, Gesundheitschecks, Datensammlung etc. bzw. zur Verwertung der Informationen
6. *Ziele*: Kategorien zu Zielvorstellungen und Zweck der Maßnahmen etc.

Im endgültigen Materialdurchlauf erfolgte dann anhand des induktiv entwickelten Kategoriensystems mit den sechs deduktiven Oberkategorien die Kodierung aller 65 Fragebögen. Zu diesem Zweck wurde das Datenmaterial im Text markiert und die Codes der Kategorien wurden am Textrand notiert. Gleichzeitig wurden die Häufigkeiten der Nennungen der einzelnen Kategorien notiert, um diese anschließend quantitativ auszuwerten. Abbildung 38.3 zeigt als Beispiel die induktiven Kategorien für die Oberkategorie „Integration von BGF in die Kultur und Politik der Organisation/des Unternehmens“ inklusive der Zuordnung zu den sechs Überkategorien.

Als Ergebnis liegt somit einerseits das vollständige Kategoriensystem für die sechs analysierten Bereiche betrieblicher Gesundheitsförderung (Integration in Unternehmenskultur, generelle Punkte, Rauchen, Ernährung, Bewegung, Stress/mentale Gesundheit) vor, und andererseits die zugehörigen Häufigkeiten der Einzelkategorien. Anhand dieser beider Quellen erfolgt schließlich die Interpretation des Datenmaterials sowie der Rückbezug zur Fragestellung, also die Klärung der entscheidenden Faktoren und wesentlichen Aspekte sowie der zentralen Maßnahmen effizienter und effektiver betrieblicher Gesundheitsförderung, welche sich in der Praxis bewähren. Für die angegebene Oberkategorie (Abb. 38.3) bedeutet dies, dass für die erfolgreiche Integration betrieblicher Gesundheitsförderung in die Unternehmenskultur

1. eine wirksame und zweckmäßige Evaluation die Wünsche, den Bedarf und die Nutzung der Angebote und Maßnahmen erfassen sollte,

STRATEGIE/UMSETZUNG		
Code	Kategoriename	Häuf.
IA4	Einbeziehung in die Unternehmenskultur	32
IA3	BGF als Verantwortungsbereich des Managements	30
IA1	eigenständiges, namentlich gekennzeichnetes Gesundheitsprogramm	22

STRUKTUREN		
Code	Kategoriename	Häuf.
IA5	Einführung von Arbeitsgruppen zur BGF	55
IA7	Partizipation der MitarbeiterInnen bei der Planung/Entwicklung von BGF	48
IA6	Einbeziehung von Experten/Fachpersonal	45

AUS-/WEITERBILDUNG		
Code	Kategoriename	Häuf.
IA8	Ergebnisse/Wirkung der BGF-Maßnahmen den MitarbeiterInnen darlegen	32

EVALUATION		
Code	Kategoriename	Häuf.
IA13	Erfassung von MitarbeiterInnen-Beschwerden, Anregungen, Ansichten etc.	55
IA12	Bewertung der BGF-Pläne/Aktivitäten durch die MitarbeiterInnen	48
IA10	Bedarfsanalyse zur Planung der BGF-Angebote	38
IA11	Analyse der Nutzung der BGF-Maßnahmen	38
IA14	regelmäßige MitarbeiterInnen-Befragung zum Gesundheitsverhalten	33
IA9	Datenbank über Krankenstände (Dauer, Ursachen)	14

**Abb. 38.3** Kategorien zu den beschreibenden Elementen (Hauptkategorie A) zentraler Maßnahmen zur Oberkategorie „Integration von BGF in die Kultur und Politik der Organisation/des Unternehmens“ bestehend aus Code, Name und Häufigkeit der Kategorien sowie der Zuordnung in die sechs zuvor genannten Themengebiete (Fenzl/Mayring 2009: 68-69).

2. Arbeitsgruppen mit Vertreter/innen aus der Belegschaft und Experten/innen-Kreisen, die Einbindung der Mitarbeiter in den Prozess und die Rückmeldung der Wirkung der Maßnahmen an die Beteiligten eine entscheidende Rolle spielen.

## Literatur

- Berelson, Bernhard (1952): Content Analysis in Communication Research. Glencoe (Ill.): Free Press
- Braun, Virginia/Clarke, Victoria (2006): Using Thematic Analysis in Psychology. In: Qualitative Research in Psychology 3: 77-101
- Fenzl, Thomas/Mayring, Philipp (2009): Results of the MOGP Content Analysis. In: Leitgöb, Heinz/Sebinger Sarah (Hg.): moveeurope. Linz: Gutenberg. 60-87
- Mayring, Philipp (2010): Qualitative Inhaltsanalyse. Weinheim/Basel: Beltz Verlag
- Mayring, Philipp/Gläser-Zikuda, Michaela (Hg.) (2008): Die Praxis der Qualitativen Inhaltsanalyse. Weinheim: Beltz
- Mayring, Philipp/Huber, Günter L./Gürtler, Leo/Kiegelmann, Mechthild (Hg.) (2007): Mixed Methodology in Psychological Research. Rotterdam: Sense Publishers
- Mayring, Philipp/Brunner, Eva (2006): Qualitative Textanalyse – Qualitative Inhaltsanalyse. In: Flaker, Vito/Schmid, Tom (Hg.) (2006): Von der Idee zur Forschungsarbeit. Wien: Böhlau. 453-462
- Murphy, Gregory (2002): The Big Book of Concepts. Cambridge: MIT Press
- Neuendorf, Kimberly A. (2002). The Content Analysis Guidebook. Thousand Oaks: Sage
- Reichert, Jo (2007): Qualitative Sozialforschung. In: EWE 18: 195-208
- Titscher, Stephan/Meyer, Michael/Wodak, Ruth/Vetter, Eva (2000): Methods of Text and Discourse Analysis. London: Sage
- Ulich, Dieter/Haußer, Karl/Mayring, Philipp/Strehmel, Petra/Kandler, Maya/Degenhardt, Blanca (1985): Psychologie der Krisenbewältigung. Weinheim: Beltz
- Van Dijk, Toin Adrianus (1980): Macrostructures. Hillsdale/New York: Erlbaum

---

## **Teil 4**

### **Offene Befragung**

Cornelia Helfferich

---

## 39.1 Definitionen

Qualitative, leitfadengestützte Interviews sind eine verbreitete, ausdifferenzierte und methodologisch vergleichsweise gut ausgearbeitete Methode, qualitative Daten zu erzeugen. *Leitfadeninterviews* gestalten die Führung im Interview über einen vorbereiteten Leitfaden, *Expertneninterviews* sind definiert über die spezielle Auswahl und den Status der Befragten.

Ganz allgemein liegen qualitative Daten in der Form von Texten vor. *Texte* sind definiert als strukturierte, nicht zufällige Anordnung von Symbolen, was z.B. auch Bilder (Bohnsack, Kapitel 67 in diesem Band) mit einschließt (zur Beschaffenheit von Texten als Daten siehe Flick 1996: 43ff.). Eine übliche Form, die für qualitative Auswertungen genutzt wird, sind *schriftliche Texte*, in denen sprachliche Symbole in einer sinnhaft strukturierten Weise angeordnet sind, und die – lässt man die Auswertung bereits vorhandener Texte außer Acht – in einer Interviewsituation generiert wurden.

Auf die *Gestaltung der Interviewsituation* als Gestaltung der Datenerhebung ist große Sorgfalt zu verwenden – vergleichbar mit der Sorgfalt, mit der in der standardisierten Forschung ein Fragebogen (Reinecke, Kapitel 44 in diesem Band) entwickelt wird –, weil davon die Güte (Flick, Kapitel 29 in diesem Band) und die Brauchbarkeit der erhobenen Daten abhängen.

Die erste zentrale Frage bei der Erzeugung von Texten im Interview ist die, *wie und mit welcher Begründung das Sprechen (die Textproduktion) der interviewten Person beeinflusst und gesteuert wird*, indem in den Interviewablauf interveniert und er geformt und vorstrukturiert wird. Das kann z.B. dadurch geschehen, dass die Aufmerksamkeit auf einen im Zusammenhang mit der Forschungsfrage interessierenden Bereich gelenkt wird oder dass Begrifflichkeiten, Relevanzen und Vorverständnisse von den Interviewenden als Vorgaben eingeführt werden.

Die zweite, mit der ersten Frage zusammenhängende zentrale Frage betrifft die *Gestaltung der an dem Interview beteiligten Rollen*. Die Interviewsituation konstituiert in der Regel ein asymmetrisches und komplementäres Rollenverhältnis „Interviewende-Interviewte“, das einen künstlichen Charakter hat, weil die Kommunikation nicht den in der Alltagskommunikation geltenden Regeln (z.B. der Erwartung der Wechselseitigkeit der Beiträge) entspricht. Die Interviewenden sprechen die Interviewten in einer spezifischen Rolle an. Die Interviewten ihrerseits deuten das Setting, die eigene Rolle und ihr Verhältnis zu der interviewenden Person bzw. zur Wissenschaft, für die die Interviewenden stehen können, ordnen dies alles ein und gestalten die Situation mit ihren eigenen Äußerungen aus.

Die Methodologie der Leitfadeninterviews ebenso wie die Methodologie der Experteninterviews betreffen beide die Ausgestaltung der Interviewsituation in der Strukturierung des Interviewablaufs und der Positionierung der beteiligten Rollen.

*Leitfadeninterviews* sind definiert als Interviews, die mit einem Leitfaden den Interviewablauf gestalten.

Der *Leitfaden* ist eine vorab vereinbarte und systematisch angewandte Vorgabe zur Gestaltung des Interviewablaufs. Er kann sehr unterschiedlich angelegt sein, enthält aber immer als optionale Elemente (Erzähl-)Aufforderungen, explizit vorformulierte Fragen, Stichworte für frei formulierbare Fragen und/oder Vereinbarungen für die Handhabung von dialogischer Interaktion für bestimmte Phasen des Interview. Der Leitfaden beruht auf der bewussten methodologischen Entscheidung, eine maximale Offenheit (die alle Möglichkeiten der Äußerungen zulässt) aus Gründen des Forschungsinteresses oder der Forschungspragmatik einzuschränken. Die Erstellung eines Leitfadens folgt dem Prinzip „So offen wie möglich, so strukturierend wie nötig“. Für die meisten Fragestellungen und Forschungsinteressen ist es notwendig, bei aller grundsätzlichen Offenheit den Interviewablauf in einem gewissen Maß zu steuern.

Auf einen Leitfaden verzichten *narrative, monologische Interviews* (Küsters, Kapitel 40 in diesem Band) in einer ersten Phase der Stegreiferzählung, in der allein die interviewte Person spricht. Begründet ist dies darin, dass biografische Erfahrungsausschnitte (Rosenthal, Kapitel 35 in diesem Band) oder fiktive Welten erhoben werden sollen, deren Gesamtfigur nur durch eine ununterbrochene Erzählung erfasst werden kann. Das erfordert aber eine biografische Reflexions- und Erzählkompetenz, die z.B. bei Kindern und teilweise auch bei Jugendlichen nicht vorausgesetzt werden kann. Auch bei sozialen Gruppen, in denen selbstbezügliches Sprechen über die individuellen Besonderheiten nicht üblich ist, sind stärker stützende Erhebungsverfahren wie Fokusinterviews (s.u.) sinnvoll. Narrative Interviews können im Nachfrageteil einen Leitfaden einsetzen. Keinen Leitfaden nutzen zudem *ethnographische Interviews*, die als offenes, spontanes Gespräch ohne weitere Vorgaben angelegt sind.

Während Leitfadeninterviews über die Methode – über das Erhebungsinstrument „Leitfaden“ – definiert sind, sind *Expertенinterviews* über die spezielle Zielgruppe der Interviewten und über das besondere Forschungsinteresse an Expertenwissen als besondere Art von Wissen bestimmt.

Experten können als Ratgeber und Wissensvermittler fungieren, die Fakten- und Erfahrungswissen weitergeben und so wenig aufwändig einen guten Zugang zu Wissensbereichen eröffnen (Typus „*systematisierendes Experteninterview*“). Das Erkenntnisinteresse kann sich aber auch auf die Beschaffenheit des Expertenwissens selbst und z.B. die zugrundeliegenden Konzepte, Deutungen und Konstruktionen von Welt(-ausschnitten) richten (Typus „*theoriegenerierendes Experteninterview*“) (Bogner/Menz 2005: 37). Wer als Experte oder Expertin behandelt werden soll und wem auf diese Weise ein spezifischer Status als Interviewpartner oder -partnerin zugewiesen wird, ist dabei im Einzelnen zu entscheiden.

---

## 39.2 Strukturierung und Rollenzuweisung als Aspekte der Interviewgestaltung

Um die Entscheidung zu treffen, in welchem Ausmaß und auf welche Weise die Interviewsituation vorstrukturiert werden soll und den Beteiligten Rollen zugewiesen werden sollen, sind vier grundlegende Überlegungen zu beachten.

### 39.2.1 Das Interview ist eine Kommunikationssituation, in der interaktiv der Text erzeugt wird

Prinzipiell wird davon ausgegangen, dass es einen authentischen Text nicht gibt. Er müsste losgelöst von allen sozialen Generierungsbedingungen Gültigkeit beanspruchen, und diese idealistische Möglichkeit wird von der Wissenssoziologie verneint. Qualitative Forschung braucht auch keine authentischen Texte, denn sie will – sich wandelnde! – subjektive Wahrheit und soziale Sinnstrukturen rekonstruieren. Man kann von der Konvention ausgehen, dass Personen, die zum Interview bereit sind, nicht offen lügen. Sie erzählen ihre subjektive Wahrheit, die für den spezifischen Erzähltagenblick gültig ist, wobei sie z.B. etwas ausschmücken oder weggelassen, etwas abmildern oder zuspitzen.

Allgemein sind sprachliche Handlungen wie Antworten oder Erzählen immer inter-subjektiv angelegt. Zum sozialen Umfeld, auf das sie ausgerichtet sind, gehören reale oder imaginäre Personen, auf die hin erzählt wird. Den Rahmen bildet die von den Beteiligten geteilte Situationsdefinition: „Dies ist ein Interview“. Mindestvoraussetzungen sind außerdem eine gemeinsame Sprache und das Beherrschen von Basisregeln der Kommunikation. Auf dieser Basis beziehen sich Interviewende und Interviewte wechselseitig aufeinander und stellen zusammen das her, was nachher als Text der Auswertung zu Grunde gelegt wird.

Das Interview ist in diesem Sinn eine Interaktions- und Kommunikationssituation, in der unter den konkreten Bedingungen des Settings, der Interaktionsdynamik und des persönlichen Verständigungsprozesses zwischen den am Interview Beteiligten eine spezifische, kontextgebundene Version einer symbolischen Welt erzeugt wird (Helfferich 2011: 64).

Mit dem Aufgeben der Authentizitätserwartung handelt man sich die Pflicht ein, stets zu reflektieren, unter welchen Bedingungen die spezifische Version entstanden ist, und den Text immer als „Text-im-spezifischen-Entstehungskontext“ und als Abbild der Interviewten-Interviewenden-Interaktion auszuwerten.

### **39.2.2 Qualitative Forschung verpflichtet zu größtmöglicher Offenheit**

Das Aufgeben des Authentizitätsanspruchs entbindet nicht von der Aufgabe, das Interviewsetting und den Verständigungsprozess im Interview so zu gestalten, dass das Sinn- system und die – möglicherweise situative – subjektive Wahrheit der Interviewten entfaltet werden kann. Der größte Fehler qualitativer Interviewdurchführung liegt darin, zu viel vorzugeben und abzufragen sowie in einer Haltung, bestätigt bekommen zu wollen, was man schon weiß. Prüfstein für Offenheit ist daher die Frage, ob die Interviewten in einer für die Interviewenden unerwarteten und mit ihrem Deutungssystem inkompatiblen Weise – aber mit ihrer eigenen Logik – antworten oder erzählen können. Offenheit heißt für die Interviewten, dass sie den Raum haben, das zu sagen, was sie sagen möchten, dass sie z.B. das ansprechen können, was ihnen selbst wichtig ist, dass sie ihre eigenen Begriffe verwenden können. Für die Interviewenden heißt Offenheit, dass sie nicht ihren eigenen Verstehenshorizont als Maß alles Verstehbaren nehmen und sich auf ganz anders geartete Normalitäten, die aus sich heraus ihren eigenen Sinn haben, einlassen können. Hier ist es hilfreich, sich die Wurzeln der qualitativen Sozialforschung in dem ethnographischen Fremdverstehen in Erinnerung zu rufen.

### **39.2.3 Die Notwendigkeit, Offenheit einzuschränken ist im Forschungsinteresse begründet**

Offenheit ist eine Anforderung und eine Ideallinie, die Orientierung gibt, die aber nie vollständig einzulösen ist. Der erste Grund wurde schon genannt: *Das Interview stellt eine soziale Kommunikationssituation dar und schon das allein schränkt die Möglichkeit dessen ein, was alles wie gesagt werden kann.* So wenig, wie man nicht nicht kommunizieren kann, kann man nicht nicht die Äußerungsmöglichkeiten von Interviewten beeinflussen. Die Interviewenden sind immer Mit-Erzeugende des Textes und können ihren Einfluss nicht ins Nichts auflösen. Interaktionssituationen geben immer ein Minimum an nicht hintergehbaren Einschränkungen der offenen Äußerungsmöglichkeiten vor.

Zum zweiten ist der *Verzicht auf jede Strukturierung im Dienste des Forschungsinteresses meist weder vertretbar noch sinnvoll.* So sind Interviewte aus ethischen Gründen immer über den Zweck der Forschung zu informieren – das kann aber bereits die Offenheit der Äußerungsmöglichkeiten einschränken. Die Ankündigung des Forschungsvorhabens, das einen bestimmten Ausschnitt der Welt als forschungsrelevant setzt, und ebenso Frageformulierungen und (Erzähl-)Aufforderungen vermitteln Setzungen und Vorannahmen. Die

Interviewten entnehmen diesen Aspekten Hinweise, was für die Forschung „von Interesse“ sei und welches Vorverständnis gesetzt ist. Auch wenn die Interviewten die Freiheit haben, solche Vorgaben abzulehnen und Fragen zurückweisen, so lassen sie sich dann, wenn sie von ihrer Seite her einen Beitrag zu einer Verständigung leisten wollen, auf diese Relevanzen und das Vorverständnis ein, akzeptieren den gegebenen Äußerungshorizont als gemeinsamen Verständigungsrahmen und versuchen, ihn so gut wie möglich auszufüllen.

Und bei aller Offenheit werden in der Regel Äußerungen zu bestimmten Thematisierungsbereichen, die für die Forschungsfrage relevant sind, benötigt. Eine Offenheit, die den Interviewten nicht die entsprechenden Stichworte liefert und die es ermöglicht, dass sie über alles Mögliche sprechen, nicht aber über die Themen der Forschung, erzeugt dann Daten, die für den konkreten Zweck letztlich nicht brauchbar sind. Die Überlegungen zur Interviewgestaltung „So offen wie möglich“ sind um den Nachsatz „und so strukturiert wie nötig“ zu ergänzen.

Die Beschränkungen von maximaler Offenheit ergeben sich aus dem Forschungsinteresse. Prinzipiell verlangt ein Interesse an der Rekonstruktion von subjektiven Konzepten, Deutungen, Relevanzen, Sinnstrukturen etc. mehr Offenheit und mehr Raum für die Interviewten, ihre Erzählung in ihrer Eigenheit zu entfalten. Ein stärker auf inhaltliche Informationen zu umrissenen Forschungsthemen gerichtetes Interesse bedarf dagegen stärker der Vorgabe von Themen, um entsprechende Thematisierungsanreize einzuführen, und der Leitfaden kann stärker strukturiert sein.

Hierzu zwei Beispiele:

1. *Forschungsinteresse: Rekonstruktion subjektiver Konzepte, Deutungen und Konstruktionen von Bildung und der Selbstdeutung des eigenen Lebens in der Dimension „Was hat Bildung mit meinem Leben gemacht“.* Zumindest in einem Teil des Interviews sollte es möglich sein, den eigenen subjektiven Sinn, der in vielfältiger Weise dem Konzept „Bildung“ verliehen werden kann, zur Entfaltung zu bringen, z.B. als allgemeine Sozialisation oder nichtformale Bildung. Es muss auch möglich sein, ein Verständnis des eigenen Lebens zu präsentieren, in dem Bildung – wie auch immer verstanden – keine Relevanz hat. Fragen wären z.B. „Unter Bildung kann man sehr viel verstehen. Können Sie anhand ihrer eigenen Erfahrungen erzählen, was Bildung für Sie bedeutet?“
2. *Forschungsinteresse: Subjektiv wahrgenommene Bedingungen für Erfolg und Scheitern bei formalen Bildungsabschlüssen in Bildungseinrichtungen, speziell Bedeutung der Lehrenden, der Unterstützung durch Eltern und des Erlebens von Prüfungssituationen.* Hier sollte das Verständnis von Bildung vorgegeben und entsprechende Stichworte sollten eingeführt oder Fragen gestellt werden. Das Forschungsinteresse begründet hier stärkere Einschränkungen der Offenheit mit Vorgaben wie „Wir untersuchen Bildungserfolg in der Schule. Können Sie erzählen, was zu Ihrem Schulerfolg oder -misserfolg beigetragen hat?“ und auch eine stärkere Steuerung des Interviewverlaufs mit spezifischen Fragen.

Je stärker das Forschungsinteresse auf konkrete, offen erhobene Informationen ausgerichtet ist, desto mehr Strukturierung verträgt das Interview und desto mehr Vorgaben sind gerechtfertigt.

### 39.2.4 Die Rollenausgestaltung im Interview

Nicht nur die Strukturierung z.B. durch intervenierende Fragen oder begriffliche Vorgaben, sondern auch die Zuweisung von Rollen an die Beteiligten sind Teil der Gestaltung der Interviewsituation. In der Regel besteht eine *Asymmetrie* in dem Sinn, dass die eine Person aufgrund eines wissenschaftlichen Interesses das Interview führt, die andere trägt als Auskunftsperson zur Erfüllung dieses Interesses bei. Die *Beziehungen* sind durch weitere *Dimensionen* bestimmt:

- *Dimension der Machtrelation und der Sicherheit:* Beide Beteiligte haben in ihren Rollen ein Machtpotenzial: Die Interviewenden, weil sie die wesentliche Kontrolle über die Interviewsituation haben. Sie allein kennen den Fortgang des Interviews und sie haben es als Möglichkeit in der Hand, die Interviewten bloßzustellen oder nicht bloßzustellen. Die Interviewten haben die Macht, etwas zu äußern oder eben nicht zu äußern. Angesichts dieser Machtbalance kann gerade Offenheit und ein geringer Strukturierungsgrad in der Interviewgestaltung Interviewte belasten und verunsichern, wenn sie lieber wissen möchten, was von ihnen konkret erwartet wird. Offenheit, die verunsichert, kann zu defensiven und kargen Antworten führen. Die Unsicherheit ist am größten und die Rollenaushandlung am wichtigsten vor allem in der Eröffnungsphase des Interviews. Der Erzählfluss und positive Erfahrungen mit der situativen Resonanz, die die Interviewenden signalisieren, führen in der Regel zu einer Entspannung und Öffnung im weiteren Interviewverlauf.
- *Dimension Fremdheit/gemeinsamer Erfahrungshintergrund:* Unterschiede zwischen Interviewten und Interviewenden z.B. bezogen auf das Alter, das Geschlecht, den sozialen Status oder ethnische Zugehörigkeit bedeuten, dass ein gemeinsamer Erfahrungshintergrund nicht vorausgesetzt wird. Das kann die Interviewten dazu motivieren, mehr zu erklären und weiter auszuholen, um etwas verständlich zu machen. Es kann aber auch dazu führen, Dinge, für die aufgrund der Unterschiede kein Verständnis erwartet wird, nicht anzusprechen. Ähnlichkeiten im Erfahrungshintergrund dagegen können sich dagegen vertrauensbildend auf die Erwartung auswirken, verstanden zu werden, und die Befragten für Äußerungen öffnen. Bestimmte Aspekte werden aber möglicherweise nicht thematisiert, weil sie als gemeinsam geteiltes Wissen nicht explikationsbedürftig erscheinen.
- *Berücksichtigung unterschiedlich ausgeprägter Diskurskulturen:* Qualitative Interviews verlangen generell eine spezifische sprachliche Vermittlungskompetenz, die nicht in gleichem Maß in allen Sozialgruppen üblich ist. Daher sind die Fragen auf die in der Befragtengruppe üblichen Formen der Reflexions-, Kommunikations- und Verbalisierungskulturen und -kompetenzen zuzuschneiden. In der Ausgestaltung der

Interviewrollen darf den Interviewten nicht eine Erzählrolle zugemutet werden, die diese nicht ausfüllen können. Gegebenenfalls sind Alternativen zu entwickeln wie mit solchen Zielgruppen oder Interviewsituationen umzugehen ist. Besondere Gruppen sind hier z.B. Kinder oder Experten und Expertinnen (Mey 2003).

---

### 39.3 Das Leitfadeninterview

In einem Forschungsprojekt, in dem Leitfadeninterviews Methode der Wahl sind, wird in der Regel ein einziger Leitfaden entwickelt, der allen Interviews zugrunde gelegt wird. Damit sind die einzelnen Interviews gut vergleichbar, weil die Erhebungssituation sich ähnelt und z.B. ähnliche oder gleiche Fragen gestellt werden. Dieser Leitfaden wird schriftlich festgehalten und in der Regel dem Ergebnisbericht als Anhang angefügt.

#### 39.3.1 Formen des Leifadens

Ein Leitfaden kann aus einer Liste vorgegebener und in fester Reihenfolge zu stellender offener Fragen („*Frage-Antwort-Schema*“) bestehen oder aus mehreren „Erzählaufrüfferungen“ („*Erzählaufrüfung-Erzählung-Schema*“) oder beides kombinieren. Fragen können ausformuliert sein, wobei vorgegeben werden kann, wie flexibel oder wörtlich die Formulierung zu übernehmen ist, oder die Interviewenden können Aspekte im Interview flexibel anzusprechen – die Stichworte des Leitfadens dienen lediglich als Erinnerungsstütze.

Ein Leitfaden muss aber keineswegs immer nur verbale Elemente enthalten. Auch der Einsatz von „Stimuli“ kann in einem Leitfaden festgehalten werden. Vereinbarungen, wie eine Erzählung spontan zu unterstützen sei, können aufgenommen werden. Der Leitfaden kann zudem Stimuli unterschiedlicher Art oder aktivierende Aufforderungen einbinden wie z.B. die Aufforderung, ein Bild zu kommentieren, Stellung zu etwas zu beziehen oder ein Problem aufzuarbeiten, jeweils mit entsprechenden Nachfragen. Ein *Stimulus* kann ein Bild (Bohnsack, Kapitel 67 in diesem Band), eine Geschichte (Ernst, Kapitel 63 in diesem Band), eine Fallvignette (Dülmer, Kapitel 53 in diesem Band), ein Film (Akremi, Kapitel 69 in diesem Band) etc. sein, der im Sinne des Forschungsinteresses bei den Befragten eine Reaktion auslöst (s.u. Fokusinterviews). So definiert, deckt das Leitfadeninterview eine Reihe wichtiger Interviewtraditionen ab.

Es ergibt sich damit ein Spektrum an Ausgestaltungen eines Leitfadens. Es reicht von einem Leitfaden, der sehr zurückhaltend mit Vorgaben und Frageformulierungen ausgestaltet ist, und aus wenigen Erzählaufrüfferungen besteht, ergänzt um Hinweise, wie eine Erzählung gestützt und aufrechterhalten werden kann – geeignet für ein hermeneutisch-rekonstruktives Interesse an Sinnstrukturen (Kurt/Herbrik, Kapitel 33 in diesem Band) –, bis zu einem stark strukturierten Leitfaden mit einer Liste von Fragen zu konkreten inhaltlichen Aspekten – geeignet für eine Orientierung an konkreten Informationen zu Inhalten

(Mayring/Fenzl, Kapitel 38 in diesem Band). Im ersten Fall wird der Akzent bei „so offen wie möglich“ gesetzt, im zweiten Fall auf „so strukturiert wie notwendig“.

Es gibt Gründe für und gegen eine (starke) Strukturierung der Interviewsituation. *Gegen eine starke Strukturierung spricht*, dass die Antworten resp. Erzählungen dann stärker als Echo auf die Fragen generiert sind. Wird z.B. in dem obigen Beispiel ein bestimmter Bildungsbegriff vorgegeben, kann nicht mehr erhoben werden, ob die Befragten möglicherweise ein ganz anderes Verständnis von Bildung haben. Die Vielfalt möglicher Äußerungen wird eingeschränkt. Das Hinlenken auf bestimmte Aspekte bedeutet eben auch, dass subjektive Relevanzen der Befragten, die in eine andere Richtung gehen oder die den Setzungen der Forschung zuwiderlaufen, eine geringere Chance der Äußerung haben. Nur in Verständigungskrisen wird offenbar, dass das Vorgegebene nicht dem Vorstellungshorizont der Interviewten entspricht und sie Vorannahmen nicht teilen oder dass die Interviewten zentrale Begrifflichkeiten nur aus Höflichkeit adaptieren, obwohl sie nicht ihrem Verstehenshorizont entsprechen.

Auf der anderen Seite gibt es *Argumente für eine (starke) Strukturierung*. Wenn in dem Leitfaden Relevanzen vorgegeben werden, dann ist sichergestellt, dass das, was für die Forschung an Äußerungen interessant und wichtig ist, angesprochen wird. Die Befragten äußern sich auch dann, wenn der Fokus für sie selbst nicht relevant ist und sie bei einer ganz offenen Erzählauflaufforderung das Thema nicht von sich aus angesprochen hätten. Wenn alle zudem ähnliche Fragen gestellt bekommen, sorgt dies für eine Vergleichbarkeit der Interviews miteinander. Dies kann für die Auswertung, die vor der Herausforderung steht, verallgemeinernde Ergebnisse aus der unreduzierten Vielfalt der individuellen Deutungen zu gewinnen, eine wesentliche Erleichterung sein.

### 39.3.2 Die Erstellung eines Leitfadens

Im Folgenden wird ein Verfahren vorgeschlagen, wie ein Leitfaden nach dem Prinzip „So offen wie möglich, so strukturierend wie nötig“ erarbeitet werden kann. Der formale Aufbau des Leitfadens folgt einem dreistufigen Prinzip:

1. In einem ersten Schritt wird den Befragten die *Möglichkeit* gegeben, *sich so frei wie möglich zu äußern*. Die Erzählauflaufforderung wird dabei so formuliert, dass die Chance besteht, dass möglichst viele für die Forschung interessante und relevante inhaltliche Aspekte spontan angesprochen werden.
2. In einem zweiten Schritt werden diejenigen *Aspekte nachgefragt – und zwar nur diese! –, zu denen keine Texte oder für das Forschungsinteresse nicht in ausreichendem Maß Texte erzeugt wurden*. Diese beiden Schritte können wiederholt eingesetzt werden.
3. In einem dritten Schritt werden in einem abschließenden Teil des Interviews *strukturierte und in der Formulierung vorgegebene Fragen* gestellt.

Die Kombination von offener Erzählaufforderung und in der Offenheit näher zu bestimmenden immanenten Nachfragen hilft, drei wesentliche *Anforderungen an einen Leitfaden* zu erfüllen:

- *Offenheit als Priorität*: Interventionen als Einschränkungen der Äußerungsmöglichkeiten müssen so offen wie möglich gehandhabt werden und müssen mit dem Forschungsinteresse begründet sein.
- *Übersichtlichkeit*: Der Leitfaden muss übersichtlich sein, im (selten realisierbaren) Idealfall so übersichtlich, dass man ihn im Kopf behalten kann. Zu viele Fragen beschränken die für das Generieren von Texten notwendige Erzählzeit.
- *Anschmiegen an den Erzählfluss*: Die Komposition eines Leitfadens sollte dem Erinnerungs- oder Argumentationsfluss folgen und nicht zu abrupten Sprüngen, Themenwechseln oder zum Switchen zwischen der Erinnerungsperspektive und einer Reflexions- oder Bilanzierungsperspektive zwingen. Einstellungs- und Bewertungsfragen gehören in einen gesonderten Abschnitt. Priorität vor dem Einhalten einer Reihenfolge der Fragen soll immer die spontane Äußerung haben. Wenn ein Nachfragen von den Interviewten direkt („Wollen Sie dazu noch etwas hören?“) oder indirekt (Andeutungen, offenkundige Unvollständigkeiten, nonverbal: Zögern) „angeboten“ wird, sollte dies nur aus sehr gewichtigen Gründen übergangen werden.

Das konkrete Vorgehen bei der Erstellung eines Leitfadens (siehe Abb. 1) hat neben der Erstellung des Produktes den zweiten Zweck, das eigene Forschungsinteresse zu realisieren und zu reflektieren und sich von einer Fixierung auf Detailfragen zu lösen, die eine Haltung der Offenheit blockieren. Die *vier Schritte* – leicht mit der *Formel SPSS* zu merken – beinhalten (ausführlicher: Helfferich 2011: 178ff.):

- *das Sammeln von Fragen („S“)*: Es werden möglichst viele Teilespekte des Forschungsinteresses als mögliche Fragen zusammengetragen – ohne Schere im Kopf und ohne Berücksichtigung der methodisch guten Formulierung und Eignung,
- *das Prüfen der Fragen („P“)*: Ein kritischer Blick auf die Liste ist zu verbinden mit einem Rückbezug auf das Forschungsinteresse: Öffnen die Fragen bzw. Erzählaufforderungen die Äußerungsmöglichkeiten so, dass Neues und Fremdes zur Sprache kommen kann, und ermöglichen sie die Generierung von Texten, die bezogen auf ein informationsbezogenes oder rekonstruktives Interesse relevant sind? Welche Fragen drücken die Erwartung aus, ein Vorwissen bestätigt zu bekommen – können widersprechende Äußerungen entfaltet werden? Nach dieser Prüfung werden Fragen gestrichen oder z.B. Fragen mit Ein-Wort-Antworten, die als notwendig erachtet werden, in einen gesonderten Fragebogen ausgelagert, um das Interview zu entlasten,
- *das Sortieren („S“)*: Die verbleibenden Fragen werden nach zeitlicher Abfolge, inhaltlicher Zusammengehörigkeit und Fragerichtung (z.B. Erinnerungsfragen versus Bewertungsfragen) sortiert und gebündelt,
- *das Subsumieren („S“)*: Für jedes Bündel von Fragen wird ein möglichst erzählgenerierend wirkender und möglichst wenig Präsuppositionen enthaltender Impuls gesucht, unter den die Einzelaspekte „subsumiert“ (= untergeordnet) werden können. Es soll

**Tab. 39.1** Schema eines Leitfadens

Leitfrage/Stimulus/ Erzählaufforderung	Inhaltliche Aspekte Stichworte – nur erfragen, wenn nicht von allein thematisiert	(Nach-)Fragen mit obligatorischer Formulierung
Erzählaufforderung		
Erzählaufforderung		
Erzählaufforderung		
Bilanzierungsfragen		
Einstellungsfragen		
Abschlussfrage: Fehlt etwas?		
Nach dem Interview: Ergänzender Fragebogen für Faktenfragen		

damit die Chance gegeben werden, dass die Aspekte spontan angesprochen werden (1. Spalte in Tab. 39.1). Die „untergeordneten“ Erzählaspekte werden als Stichworte für Nachfragen notiert (2. Spalte in Tab. 39.1). Sie kommen dann zum Einsatz, wenn die Aspekte nicht ausreichend spontan zur Sprache kamen. Bei einigen Aspekten kann entschieden werden, sie obligatorisch in einer vorformulierten Weise zu stellen (3. Spalte oder Sammlung stärker strukturierter und Erzählaufforderungen nicht zuordenbarer Fragen im Abschlussteil des Interviews).

Bei einem ausgewogenen Verhältnis von Erzählaufforderungen und Nachfragen wird der Interviewtypus „*teilnarrativ*“ genannt. Die Form kann aber auch stärker strukturierend und weniger mit Erzählaufforderungen gestaltet sein und nimmt dann eher die für Experteninterviews typische Gestalt an.

### 39.3.3 Modifikationen des Leitfadens: Fokusinterviews

Eine Eingrenzung der Offenheit für alle möglichen Äußerungen der Befragten kann darin bestehen, den Fokus des Forschungsinteresses einzugrenzen. In dem Typus des *Fokusinterviews* als besondere Form des Leitfadeninterviews geschieht dies durch die zentrale Einführung eines spezifischen Stimulus. Definiert ist das Fokusinterview als „ein Interviewverfahren, vor dessen Beginn eine von allen Befragten erlebte Stimulussituation (Film, Radiosendung, gelesener Text, erlebtes Ereignis, Experiment) steht. Das Interview ist darauf fokussiert, auszuleuchten, wie diese Situation subjektiv empfunden wurde und was davon wie wahrgenommen wurde.“ (Przyborski/Wohlrab-Sahr 2008: 147)

Fokusinterviews wurden ursprünglich in den 1940er Jahren als Gruppeninterviews (Vogl, Kapitel 41 in diesem Band) in der Forschung zur Rezeption medialer Darstellungen entwickelt; die Fokussierung als strukturierendes Moment kann aber ebenso in Einzelinterviews realisiert werden. Je nachdem, wie weit die Definition die Stimulussituation fasst und je nachdem, ob die Stimulussituation auch während des Interviews anhalten kann, können zwei wichtige Interviewtraditionen hier eingeordnet werden (Übersicht und Quellen: Helfferich 2011: 36f., Flick 1996, Kapitel 8), für die sich die Frage der Erstellung und Handhabung eines Leitfadens in ähnlicher Weise stellt:

- Bei einem *problemzentrierten Interview* soll im Interview zusammen mit der interviewten Person ein Problem ausgearbeitet werden – das Problem ist der Fokus. Dies fällt unter die in der Definition von Przyborski und Wohlrab-Sahr (2008) gefasste Form des Fokus als „erlebtes Ereignis“.
- Bei einem *Dilemma-Interview* wird ein Dilemma vorgestellt und eine Aufforderung an die Interviewten formuliert, mögliche Lösungen zu diskutieren. Hier schließen stärker projektive Verfahren an, bei denen z.B. als Stimulus eine Fallvignette (Dülmer, Kapitel 53 in diesem Band) eingesetzt wird. Das Interesse gilt den möglichen Lösungswegen oder Bewertungen und Handlungsstrategien und den dabei verwendeten Argumentationsebenen (also dem imaginierten und nicht einem faktischen subjektiven Erleben).

Für diese Interviewformen gilt wie für Fokusinterviews allgemein, dass die Handhabung eines Leitfadens durch das fokussierte Erhebungsinteresse und die entsprechende Festlegung der Aufgabe der Interviewenden bestimmt ist: Sie sollen das subjektive Erleben der Stimulussituation erheben und dabei die Interviewten immer wieder auf den Fokus zurückführen. Für den Einstieg sind sehr offene und unstrukturierte Fragen vorgesehen, um die Äußerungen des Erlebten nicht einzuschränken. Es schließen weitere unstrukturierte oder halbstrukturierte Fragen an, die möglichst wenig direktiv sein sollen (Kriterium der Nicht-Beeinflussung) und deren Richtung durch die Kriterien der Spezifität (Fragen, die heranführen an konkrete Erinnerungen einer zurückliegenden Situation, z.B. auch stützend durch Fotos), der Tiefe (Fragen, die Gefühle, Wahrnehmungen und Bedeutungen von affektiven Aspekten der Stimulussituation ansprechen) und der Erfassung eines breiten Spektrums der Reaktionen der Befragten angegeben ist.

Neue Themen können mit vorsichtigen Überleitungen eingeführt werden. Das entwickelte Leitfadenmodell kann als formaler Rahmen genommen werden; das Besondere ist aber die sorgsame Vermeidung stark strukturierender und die Antwortmöglichkeiten einschränkender Fragen auch in der weiteren Interviewabfolge. Für problemzentrierte Interviews sind die weiteren Interviewpassagen offen-dialogisch gestaltet mit spontanen Nachfragen. Hopf (2000: 355) fasst die Vorteile der Interviewführung bei Fokusinterviews zusammen als „Möglichkeit, eine sehr zurückhaltende, nicht-direktive Gesprächsführung mit dem Interesse an sehr spezifischen Informationen und der Möglichkeit zur gegenstandsbezogenen Explikation von Bedeutungen zu verbinden.“ Es gelten insgesamt die gleichen Überlegungen wie oben bei der Abwägung der Argumente für oder gegen eine starke Strukturierung der Interviewführung.

### **39.3.4 Ergänzende Vereinbarungen zum Umgang mit Interviewsituationen und Bedeutung von Interviewschulungen**

Jenseits des konkreten Leitfadens sind im Vorfeld einer Interviewstudie Vereinbarungen mit allen Interviewenden zu treffen, wie mit schwierigen Situationen umzugehen ist. So kann verabredet werden, wie auf *Wünsche nach konkreteren, geschlossenen Fragen* reagiert werden soll („Was genau wollen Sie denn wissen?“), wenn die Befragten durch die Offenheit überfordert sind, weil sie die in der Alltagskommunikation übliche und Sicherheit gebende Ausrichtung an den Erwartungen des Gegenübers vermissen. Auch ist das *Durchsprechen von Interventionen* sinnvoll, wie eingespielte „karge“, kurze Antworten wieder geöffnet werden und umgekehrt, ob und wenn ja mit welchen Interventionen Interviewte in einem ausufernden Wortschwall gebremst werden können. Verabredet werden sollte auch, wie auf *persönliche Rückfragen* der Interviewten an die Interviewenden reagiert werden soll und ob die Interviewten auf *Widersprüche in ihren Antworten* aufmerksam gemacht werden sollen (Helfferich 2011: 142). In jedem Fall ist die Vertrautheit mit dem Leitfaden notwendig und Schulungen im Vorfeld der qualitativen Datenerhebung, in denen eine Haltung der Offenheit eingeübt werden kann, werden empfohlen.

---

## **39.4 Experteninterviews**

### **39.4.1 Allgemeine Bestimmung**

Bei Experteninterviews werden die Befragten aufgrund des ihnen zugeschriebenen Status als Experten und Expertinnen interviewt. Das Forschungsinteresse kann eher informationsbezogen sein; dann werden die Aussagen als faktische Befunde genommen. Es kann sich aber auch auf die Rekonstruktion von Expertenwissen richten, wenn z.B. die Konstitution eines Sonderwissens eingebettet in einem spezifischen sozialen Kontext von spezialisierten Professionen untersucht wird (zur Vereinbarkeit beider Interessen siehe: Przyborski/Wohlraab-Sahr 2008: 132).

Generell wird davon ausgegangen, dass das Expertenwissen in einem gewissen Sinn von der Person gelöst werden kann – wer die gleiche Ausbildung absolviert und langjährig mit einer bestimmten Materie Erfahrungen gesammelt hat, verallgemeinert sein Wissen und stellt es nicht als persönliche und höchst individuelle Besonderheit dar. Sowohl die *historische Wandelbarkeit von Expertenwissen* als auch die *Differenzen in den Meinungen innerhalb der Gruppe von Experten und Expertinnen* zeigen, dass die beanspruchte Verallgemeinerbarkeit des Expertenwissens nicht mit einer Objektivität der Meinung gleichgesetzt werden kann und auch in Experteninterviews subjektive Deutungen gefunden werden.

Es gibt eine längere Diskussion darüber, wer als Experte im Sinne von Experteninterviews bezeichnet werden kann und welche Rolle der sozial zugeschriebene *Status „Experte/Expertin“* oder die faktische berufliche Position spielen soll. Einem weiten Begriff, bei dem

alle Menschen Experten ihres Lebens sind, steht ein enger Begriff gegenüber, der den Expertenstatus einer Funktionselite vorbehält. Przyborski und Wohlraab-Sahr (2008: 133) schlagen als einen brauchbaren Mittelweg vor, dass als Experten solche Personen gelten, „die über ein spezifisches Rollenwissen verfügen, solches zugeschrieben bekommen und eine darauf basierende besondere Kompetenz für sich selbst in Anspruch nehmen.“

Das *Rollenwissen* kann unterschiedliche Formen nehmen: Insiderwissen über institutionelle Abläufe, (professionsbasiertes und Gültigkeit beanspruchendes) Deutungswissen oder Wissen über Hintergründe und Kontexte von sonst schwer zugänglichen, kleinräumigeren Erfahrungsbereichen (Meuser/Nagel 2005). Expertenwissen kann in kanonischer Form über Ausbildungen und eine wissenschaftliche Gemeinschaft vermittelt sein und/oder auf vertieftem Erfahrungswissen beruhen. Besondere Aufmerksamkeit hat in den letzten Jahren das „implizite Wissen“ („*Tacit Knowledge*“) von Experten und Expertinnen im Rahmen der interpretativen Zugänge zu Organisationskulturen erfahren.

Das Rollenwissen kann außer mit Berufsrollen auch mit spezialisiertem, außerberuflichem Engagement verbunden sein. Diese Definition eröffnet die Möglichkeit – begründet mit dem spezifischen Forschungsinteresse –, Privatpersonen, die sich in spezifischen Segmenten in besonderer Weise engagiert und dort Erfahrungen gesammelt haben, als Experten zu interviewen. Auch in diesem Fall ist das Interview auf die Abfrage des Sonderwissens hin ausgelegt (und auch darauf beschränkt) und der Privatperson wird im Unterschied zu anderen qualitativen Interviewformen ein mit besonderer Deutungsmacht versehener Status zugewiesen.

Über die Adressierung der Interviewten als Experten haben Experteninterviews eine *spezifische Rahmung*. Es wird Anerkennung vermittelt in dem Sinn, dass Erfahrungen, Erprobtheit in Routinen und möglicherweise eine profunde Ausbildung unterstellt werden und dass das Wissen wichtig ist, weil es in besonderem Maß verallgemeinerbar und gültig ist. Das Label „Experteneinterview“ signalisiert zudem, dass es nur um diese Expertise als Sonderwissen gehen wird und sehr persönliche Bereiche nicht angesprochen werden. Dies lässt sich bei einer entsprechenden Formulierung der Forschungsfrage auch strategisch nutzen: Betroffene aus diskriminierten Gruppen verfügen ebenfalls über ein Sonderwissen. Ihnen als Experten und Expertinnen einen positiven Status zu verleihen und statt ihrer Benachteiligung dieses Wissen als ihre Ressource in den Mittelpunkt zu stellen, kann die Interviewbereitschaft deutlich stärken.

### 39.4.2 Experteninterviews als Leitfadeninterviews

Auch bei Experteninterviews ist zunächst das Forschungsinteresse, also die Form des Expertenwissens, das zu erheben ist, zu klären. In der Regel ist es stärker informationsbezogen auf die Erhebung von praxis- und erfahrungsbezogenem, technischem Wissen ausgerichtet. Üblicherweise sind Experteninterviews Leitfadeninterviews und allgemein wird eine stärkere Strukturierung mit Sachfragen oder Stimuli im Sinne von Fokusinterviews vorgeschlagen. Auf das formale Schema eines Leitfadens bezogen (s. Abschnitt 3) bedeutet

dies, dass Erzählaufrüfferungen zurückgefahren bzw. Fragen spezifisch enger fokussiert werden. Der Akzent liegt stärker auf einer strukturierten Abfolge von konkret und prägnant beantwortbaren Fragen.

Abläufe, Zusammenhänge, Wissenselemente und Routinen eignen sich für Abfragen. Sind verallgemeinernde Aussagen aufgrund der Heterogenität der Praxis schwierig, kann für umrissene und zeitlich wie aufgabenbezogen eingeschränkte Bereiche um die Darstellung von Fällen oder Beispielen unter spezifischen Maßgaben gebeten werden. Eine weitere Möglichkeit, z.B. implizites Wissen zu erheben, ist die Vorlage einer Fallvignette oder einer Problemkonstellation mit der Bitte um eine Kommentierung aus Expertensicht.

Die stärkere Strukturierung und Fokussierung sowie die sorgsame Prüfung der Beantwortbarkeit der Fragen (Pretest im Vorfeld) unterstreicht den professionellen Charakter des Kontaktes und greift Aspekte der Wissenskulturen und Arbeitszusammenhänge der Befragten auf. So ist z.B. die Expertenzeitz (mit dem Status der Expertise zunehmend) kostbar, die spezialisierten Wissensbereiche sind thematisch eng begrenzte Ausschnitte und die Kompetenz wird daran gemessen, wie effektiv Wissen in einem eigenen Fachjargon reproduziert werden kann. Aufforderungen zu unspezifischen Narrationen entsprechen diesen Regeln nicht, was aber nicht heißt, dass Experten sich auf Sachfragen hin nicht das Recht nehmen, ausführliche Narrationen zu generieren. Nicht zuletzt ermöglicht der Leitfaden den Interviewenden, sich als strukturiert und kompetent, eben auch als professionell, zu inszenieren.

In besonderer Weise bedarf ein Experteninterview der Vorarbeit: Ein Tabu ist es, Informationen abzufragen, die aus anderen Quellen leicht zugänglich sind, da dies als Zeichen von Bequemlichkeit und mangelndem Respekt interpretiert wird (es kann aber um eine Bestätigung der Informationen gebeten werden). Professionelle, die sich zu sensiblen Kontexten äußern sollen, bitten häufig vorab um die Zusendung des Leitfadens.

### **39.4.3 Die Interaktionssituation in Experteninterviews**

Die Gestaltung der Interviewsituation als Kommunikations- und Interaktionssituation bedarf bei Experteninterviews besonderer Aufmerksamkeit, da mit der Zuschreibung des Expertenstatus eine Machtrelation gesetzt ist: Die Interviewten verfügen über ein überlegenes Wissen, das einen anerkannten gesellschaftlichen Status begründet, und die Interviewenden möchten an dem Wissen teilhaben. Die Rolle der Interviewenden ist in einer komplementären Relation zu dem Expertenstatus zu definieren, z.B. als Autorität im gleichen oder einem anderen Feld, als naiver Laie, als potenzieller Kritiker oder Komplize – mit entsprechenden Folgen für die Generierung von Aussagen (Bogner/Menz 2005: 50ff., Gläser/Laudel 2009: 172ff.).

Diese Rollenkonfigurationen folgen entweder dem Prinzip der gleichen Augenhöhe oder einem hierarchischen Gefälle zwischen Experten und Laien. Die Selbstpräsentation der Interviewenden – aber auch die Ausgestaltung und Ausrichtung des Leitfadens – veranlassen die Interviewten, den Interviewenden Kompetenz oder Inkompetenz

zuzuschreiben, wobei auch die Dimensionen Geschlecht und Alter eine wesentliche Rolle spielen (Jüngeren und Frauen wird weniger Expertise zugewiesen: Abels/Behrens 2005). Przyborsik und Wohlrab-Sahr sehen die beste Selbstpräsentation in einer „gelungenen Mischung von Kompetenz und Wissensbedarf“ (2008: 135).

## 39.5 Gütekriterien für die Interviewdurchführung

Die Interviewdurchführung unterliegt auch Gütekriterien (Flick, Kapitel 29 in diesem Band) wie Datenerhebungen in der standardisierten Forschung. Beide Forschungstraditionen stimmen darin überein, dass Beliebigkeit und Zufälligkeit der Datenerzeugung nicht dem wissenschaftlichen Standard entsprechen. Doch kann für qualitative Forschung das Gütekriterium nach „Objektivität“ als Unabhängigkeit der Messergebnisse von den Untersuchenden (Krebs/Menold, Kapitel 30 in diesem Band) sein, denn in der Interviewsituation ist *Subjektivität* grundsätzlich involviert. Während standardisierte Forschung jede Form von Subjektivität als unwissenschaftlich ablehnt, unterscheidet qualitative Forschung zwischen methodisch kontrollierter und reflektierter Subjektivität, die Gütekriterium qualitativer Wissenschaftlichkeit ist, und nicht reflektierter Subjektivität, die den wissenschaftlichen Ansprüchen nicht genügt.

Während die standardisierte Forschung die Datenerhebung entkontextualisieren möchte – jeder Kontext soll die gleichen Ergebnisse erzeugen –, ist in der qualitativen Forschung die *Kontextgebundenheit der erzeugten Texte* nicht hintergehbar. Reliabilität (Krebs/Menold, Kapitel 30 in diesem Band) ist so kein sinnvolles Gütekriterium für qualitative Interviewführung. Das entsprechende Gütekriterium im Qualitativen besteht vielmehr darin, die Kontextgebundenheit der Textgenerierung einzubeziehen, also den Kontext der Interviewsituation nicht nur sorgfältig zu gestalten, sondern auch in der Auswertung die spezifische kommunikative Entstehung des Textes aufzugreifen. Kriterium ist dann nicht die identische Replikation der Daten in einer neuen Erhebung, sondern eine Einschätzung, wie welche Erhebungskontexte welche Versionen erzeugen.

Der Sorge um Validität, dass das erhoben, abgebildet und „verstanden“ (statt im Standardisierten „gemessen“) wird, was erfasst werden soll, lässt sich gerade nur durch die Prinzipien der *Offenheit* und der Fremdheit (Knoblauch, Kapitel 36 in diesem Band) in der Erhebungssituation begegnen. Je offener die Erhebung, desto eher wird die Entfaltung des subjektiven Sinns ermöglicht und umso angemessener und valider wird in diesem Sinn erhoben.

Diese Gütekriterien für die Gestaltung der qualitativen Erhebungen sind den Besonderheiten qualitativer Forschung angemessen: Weil es eine Bereinigung des Interviews um soziale und kommunikative Effekte nicht geben kann, müssen diese Effekte systematisch bei der Gestaltung des Interviews und ebenso bei der Interpretation in Rechnung gestellt und einbezogen werden. Die *methodische Kontrolle des Fremdverständens* in der Interviewsituation beinhaltet hohe Anforderungen an die Reflexion der Subjektivität von den beiden an der Erzeugung der Daten Beteiligten.

## Literatur

- Abels, Gabriele/Behrens, Maria (2005): ExpertInnen-Interviews in der Politikwissenschaft.  
In: Bogner, Alexander/Littig, Beate/Menz, Wolfgang (Hg.): Das Experteninterview.  
Wiesbaden: VS Verlag. 173-190
- Bogner, Alexander/Menz, Wolfgang (2005): Das theoriegenerierende Experteninterview.  
In: Bogner, Alexander/Littig, Beate/Menz, Wolfgang (Hg.): Das Experteninterview.  
Wiesbaden: VS Verlag. 33-70
- Flick, Uwe (1996): Qualitative Sozialforschung. Reinbek b. Hamburg: Rowohlt
- Gläser, Joachim/Laudel, Grit (2009): Experteninterview und qualitative Inhaltsanalyse.  
Wiesbaden: VS Verlag
- Helfferich, Cornelia (2011): Die Qualität qualitativer Daten. Wiesbaden: VS Verlag
- Hopf, Christel (2000): Qualitative Interviews – ein Überblick. In: Flick, Uwe/v. Kardorff,  
Ernst/Steinke, Ines (Hg.): Qualitative Sozialforschung. Reinbek: Rowohlt. 349-359
- Mey, Günter (2003): Zugänge zur kindlichen Perspektive – Methoden der Kindheits-  
forschung. URL: <https://www.familienhandbuch.de/kindheitsforschung/allgemeines-kindheitsforschung/zugange-zur-kindlichen-perspektive-methoden-der-kindheitsforschung#besonderheiten> (Zugriff 19.08.2013)
- Meuser, Michael/Nagel, Ulrike (2005): ExpertInneninterviews – vielfach erprobt,  
wenig bedacht. In: Bogner, Alexander/Littig, Beate/Menz, Wolfgang (Hg.): Das  
Experteninterview. Wiesbaden: VS Verlag. 7-29
- Przyborski, Aglaja/Wohlrab-Sahr, Monika (2008): Qualitative Sozialforschung. München:  
Oldenbourg

Ivonne Küsters

Das narrative Interview ist eine Interviewform, in der ein Befragter auf eine Eingangsfrage bzw. Erzählaufforderung ohne Unterbrechungen, ohne Vorgaben und in großer Ausführlichkeit antworten kann. Es dient der *Erhebung von Handlungsprozessen*, also Ereignisverkettungen (Mika/Stegmann, Pötter/Prein, Kapitel 86 und 87 in diesem Band), an denen der Befragte selbst beteiligt war.

Dabei ist es möglich, die *Anfangs-, Entwicklungs- und Endpunkte der erhobenen Prozesse* mit der Frage nicht bereits vorzugeben (Reinecke, Kapitel 44 in diesem Band), sondern von den Befragten selbst setzen zu lassen, also die Prozesse so zu erfassen, wie sie von den Befragten strukturiert werden. So wird eine große *Offenheit gegenüber den Relevanzen und Deutungen der Informanten* erreicht und entsprechend die Einflussnahme des Interviewers (Glantz/Michael, Kapitel 21 in diesem Band) auf die Darstellung minimiert.

Der Zeitbezug narrativer Interviews ist grundsätzlich retrospektiv. Die Frageformulierung ist jedoch so gestaltet, dass der Befragte sich an einen Zeitpunkt zurückerinnert und von dort aus seine Erlebnisse vergegenwärtigt und zum Prozessende hin erzählt.

Der Schwerpunkt des Einsatzes narrativer Interviews liegt bislang in der *Biographieforschung* (Rosenthal, Kapitel 35 in diesem Band), zur Erhebung von Lebensgeschichten; jedoch sind sie für sämtliche, auch kleinformatige Handlungsprozesse mit Beteiligung des Befragten anwendbar. Wiederkehrende Handlungsroutinen können mit dieser Interviewform nicht erhoben werden, weil sie nicht als einzelner Handlungsverlauf, sondern als Synthese vieler ähnlicher Verläufe erinnert werden.

---

## 40.1 Erzähltheorie

Die Methode stützt sich auf eine Reihe von Annahmen über die Erinnerung und Darstellung eigener Erlebnisse. Diese Grundlage der Methode, die vor allem durch Fritz Schütze (1976a, b; 1977; 1983; 1984) gelegt wurde, wird gemeinhin als „Erzähltheorie“ bezeichnet,

denn sie bezieht sich auf die Reproduktion von eigenen Erlebnissen in Form von Erzählungen, genauer noch: *Stegreiferzählungen*, also unvorbereiteten, ad-hoc formulierten Erzählungen für anwesende Zuhörer.

Einen entscheidenden Impuls für die Entwicklung dieser Erzähltheorie gab eine soziolinguistische Untersuchung von mündlichen Erzählungen eigener Erlebnisse, erstellt an einem Korpus von 600 in diversen Forschungskontexten erhobenen Interviews (Labov/Waletzky 1973). Darin wird Erzählen als eine „verbale Technik der Erfahrungsrekapitulation aufgefasst, im Besonderen als die Technik der Konstruktion narrativer Einheiten, die der temporalen Abfolge der relevanten Erfahrung entsprechen“ (Labov/Waletzky 1973: 79); darin liegt die „referentielle“ Funktion des Erzählers (1973: 79, 95-111). Zugleich wird in der Analyse von Labov und Waletzky die soziale Kontextualisierung des Erzählers hervorgehoben: Erzählungen haben den beiden Autoren zufolge eine „evaluative“ Funktion, indem sie auf einen Stimulus aus dem sozialen Kontext reagieren, in dessen Rahmen erzählt wird (1973: 79, 114-122, 125). Diese Untersuchung lenkte den Blick auf das Potential von mündlichen Erzählungen als Daten für die Soziologie und lieferte die Geltungsbegründung für die Methode des narrativen Interviews: In direkten Interaktionen ist es möglich, durch Initiierung spontaner Erzählers Darstellungen zu erhalten, die analog zu den erinnerten Erlebnissen des Erzählers aufgebaut sind.

Die von Schütze und anderen daran anschließend entwickelte Theorie über die Aufschichtung von Erlebnissen im Gedächtnis nimmt an, dass persönliche Erlebnisse sich in der Situation ihres Erlebens in der Form so genannter „kognitiver Figuren“ (Kallmeyer/Schütze 1977: 176ff.; Schütze 1984:78; zur Kritik an der Erzähltheorie: Küsters 2009: 32-38) im Gedächtnis niederschlagen, die beim spontanen Erzählen reproduziert werden: Der *Erzählträger* ist notwendig eine beteiligte Person des damaligen Geschehens, zumeist der Handlungsträger, mindestens ein Beobachter. Die *Erzählkette* repräsentiert die einzelnen Abschnitte des Geschehensablaufs in ihrem chronologischen und inhaltlichen Zusammenhang vom Beginn bis zum (vorläufigen) Abschluss. *Situationen* sind einzelne, besonders herausgehobene Prozessabschnitte, oft Höhe- oder Wendepunkte des Geschehens, die aus dem Gesamtlauf herausstechen und diesen maßgeblich beeinflusst haben. Die *thematische Gesamtgestalt* ist sozusagen der Titel, unter dem das gesamte Erlebnis verbucht wurde, dessen Bewertung durch den Erzähler: So kann die Erzählung beispielsweise als eine Leidens-, Scheiterns-, Rechtfertigungs- oder Erfolgsgeschichte angelegt sein. Erzählungen eigener Erlebnisse bestehen aus diesen vier Elementen, die jeweils auf Elemente der erlebten und erinnerten Handlungsprozesse verweisen.

In Form von Erzählungen werden diese kognitiven Figuren, in denen sich die erinnerten Erlebnisse niedergeschlagen haben, konsistenter und mit weniger redigierenden Eingriffen durch den Erzähler reproduziert als in Argumentationen oder Beschreibungen. Letztere beide Darstellungsformen ermöglichen es nämlich, die Chronologie der Ereignisse zu verlassen, Handlungsabschnitte auszulassen und die eigene Beteiligung am geschilderten Geschehen zu verschleiern, ohne dass dies die Darstellung inkonsistent werden lässt oder wenigstens indirekt an der Darstellung sichtbar wird.

Weil aber Erzählungen am erinnerten Geschehensverlauf entlang gestaltet werden, wie Labov und Waletzky (1973) gezeigt haben, sind in ihnen solche redigierenden Eingriffe sichtbarer, insbesondere wenn die erzählende Person „live“ und unvorbereitet vor Zuhörern spricht. Dann wird das Redigieren des Dargestellten am Erzähltext zumindest indirekt, zum Beispiel als Zögern, Schweigen, Wechsel der Textsorte, erkennbar. Dies liegt am besonders starken Wirksamwerden der so genannten „*Zugzwänge der Sachverhaltsdarstellung*“ (Kallmeyer/Schütze 1977: 162, 187ff.) in Erzählsituationen:

- Der *Detaillierungszwang* bewirkt die Reproduktion des erinnerten Prozesses entlang der Erzählkette, also in seinem chronologischen und inhaltlichen Zusammenhang, und in dem Detaillierungsgrad, der für die Nachvollziehbarkeit des Erzählten notwendig ist.
- Der *Gestaltschließungszwang* bewirkt die Vollständigkeit bei der Reproduktion der Erzählkette und einzelner Situationen, also deren erzählerisches Zu-Ende-Führen, so dass die Erzählung auch über innere Widerstände hinweg und trotz Scham oder Hemmungen fortgeführt wird.
- Der *Relevanzfestlegungs- und Kondensierungszwang* bewirkt, dass der Erzähler aus der erinnerten Fülle von kleinsten Handlungssplittern thematisch und orientiert am zu erzählenden Prozessgeschehen auswählt und so eine konsistente Geschichte formt; so wird in einer Lebensgeschichtserzählung nicht von jedem Urlaub erzählt werden, wohl aber von dem, in dem entscheidende Begegnungen stattgefunden haben.

Das intensive Wirken der drei Zugzwänge in Stegreiferzählungen beruht darauf, dass es sich um eine direkte Interaktion handelt, in der einem Zuhörer ein vergangenes Prozessgeschehen nachvollziehbar und einsichtig gemacht werden muss (Kallmeyer/Schütze 1977: 168; genauer zur *Motivierung* der Erzählbereitschaft durch den Interviewer: Küsters 2012: 263ff.).

---

## 40.2 Erzählstimulus und Haupterzählung im narrativen Interview

Bei der Erhebung von narrativen Interviews wird eine Doppelstrategie verfolgt: Einerseits imitiert das narrative Interview eine alltäglich, „natürlich“ vorkommende Interaktion: das spontane Erzählen eigener Erfahrungen in Face-to-Face-Interaktionen. Andererseits hebt das narrative Interview die Interaktion dem Kontext der Alltagserfahrung, indem es die normalerweise wechselnden Interaktionsrollen von „Sprecher“ und „Zuhörer“ polarisiert: Die Redeverteilung wird fast völlig asymmetrisch – Einer erzählt, der Andere hört nahezu vollständig schweigend zu. Die dahinter stehende Idee ist zum einen, die kognitiven Figuren möglichst vollständig und mit möglichst geringer Intervention durch die Interviewerin in der Erzählung reproduzieren zu lassen. Zum anderen ist das Zuhören, die Anwesenheit und Aufforderung des Zuhörers an den Erzähler notwendig, denn es sichert das vollständige Wirken der Zugzwänge des Erzählens.

Das narrative Interview, von dem eine vollständige Tonaufzeichnung angefertigt wird, beginnt mit einer *Eingangsfrage*, auch *Erzählstimulus* genannt, die sich auf ein vom Befragten miterlebtes Prozessgeschehen bezieht und eine Beantwortung in Form einer Erzählung ansteuern muss (Schütze 1983: 285). Im Hauptanwendungsbereich Biographieforschung (Rosenthal, Kapitel 35 in diesem Band) lautet dieser Stimulus häufig einfach „Bitte erzählen Sie mir Ihre Lebensgeschichte.“ Darauf reagiert der Befragte, zumeist nach einer kurzen Aushandlung (ausführlicher dazu: Küsters 2009: 56f.), im Erfolgsfall mit einer vollständigen Erzählung des erfragten Prozessgeschehens; dies ist die so genannte *Eingangserzählung* oder *Haupterzählung*, während der er vom Interviewer nicht unterbrochen wird. Der schweigt erzählunterstützend, denn bereits kleine Nachfragen oder Bitten um Präzisierungen von Daten oder Orten können den Erzähler vom Bezug auf die zu reproduzierenden kognitiven Figuren abbringen. Eine Redeübernahme durch den Interviewer erfolgt erst, nachdem vom Befragten die Erzählung des Prozesses ausdrücklich abgeschlossen, d.h. eine so genannte *Koda* (beispielsweise „Das war's.“) gesetzt wurde.

---

### **40.3 Immanente und exmanente Nachfragen im narrativen Interview**

Danach beginnt die Nachfragephase des narrativen Interviews mit den so genannten *immanenten Nachfragen*, mit denen das in der Haupterzählung angedeutete, aber nicht ausformulierte Erzählpotential ausgeschöpft werden soll (Schütze 1983: 285). Dazu hat der Interviewer während der Eingangserzählung auf thematische Brüche, Lücken, angefangene und nicht auserzählte Erzählungsfragmente geachtet. Diese werden nun in der Reihenfolge ihres Erscheinens in der Erzählung durch – ebenfalls wieder erzählgenerierende – Fragen angesteuert. Zum Beispiel rekurriert der Interviewer auf eine zuvor nur angedeutete Situation und fragt: „Wie ging es dann weiter?“

Wenn das Erzählpotential durch die immanenten Nachfragen ausgeschöpft ist, folgt mit den *exmanenten Nachfragen* die abschließende Phase des Interviews. Die hier eingesetzte Fragetechnik zeigt gewisse Ähnlichkeiten zum Leitfadeninterview (Helfferich, Kapitel 39 in diesem Band). Die exmanenten Nachfragen zielen statt auf Erzählungen auf Beschreibungen und Argumentationen, fragen nach dem „Warum“, versuchen den Erzähler zur Verbalisierung seiner Selbstdeutung bezogen auf den dargestellten Prozess zu bewegen (Schütze 1983: 285).

In einem nachgelagerten Auswertungsschritt, der *Wissensanalyse*, kann diese Selbstdeutung des Erzählers mit dem rekonstruierten Prozessverlauf kontrastiert werden (Schütze 1983: 286). Daneben können die exmanenten Nachfragen auch zur explorativen Überprüfung (und kommunikativen Validierung) von ersten Forschungsergebnissen genutzt werden. Zuletzt kann die Erhebung von *sozio-demographischen Daten* (Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band) erfolgen, soweit das Interview selbst diese Informationen noch nicht geliefert hat.

## 40.4 Auswertung

Die Auswertung von narrativen Interviews geschieht anhand einer exakten, Pausen, abgebrochene Worte, parasprachliche Äußerungen miterfassenden Verschriftlichung der Tonaufzeichnung des Interviews. Zunächst wird eine Unterscheidung zwischen Erzählungen und anderen Textsorten wie Argumentationen und Beschreibungen vorgenommen. Danach erfolgt eine weitere Untergliederung der *Narrationspassagen* in einzelne *Segmente*, die jeweils einen Prozessabschnitt repräsentieren. Hierbei orientiert man sich an formalen Hinweisen, wie dem Rahmenschaltelelement „und dann“ und anderen temporalen Konjunktionen, aber auch an inhaltlichen Brüchen, Schweigen, Tempus- oder Textsortenwechseln sowie unterschiedlichen Detaillierungsgraden einzelner Passagen. Die Interpretation geht Wort für Wort, Satz für Satz, Segment für Segment am Text entlang und fokussiert sich auf die Rekonstruktion und Analyse des erinnerten und erzählten Ablaufs des Prozessgeschehens. Die Interpretationshaltung dabei lässt sich als eine „symptomatische Datenanalyse“ (Schütze 1983: 286) beschreiben, die neben den erzählten Inhalten immer auch nach de-thematisierten Inhalten und nicht-erzählten Prozesselementen sucht. Denn gemäß der Grundlagentheorie des narrativen Interviews repräsentiert der Erzähltext die Abfolge des erinnerten Erlebens, so dass auch thematische Aussparungen, Lücken und Brüche im Text ein wichtiges Datum für die Auswertung sind. Ziel der Auswertung von narrativen Interviews ist die Analyse des Verlaufs des Prozessgeschehens in seiner sozialen Rahmung sowie der Erfahrungsaufschichtung und Handlungsorientierungen des Erzählers, und zwar für jeden erhobenen Fall. Danach erfolgt ein Vergleich der dabei identifizierten Verlaufsstrukturen und ihrer jeweiligen Bedingtheit am gesamten Sample (Akremi, Kapitel 17 in diesem Band). Dieser Fallvergleich führt schließlich zur Formulierung von „theoretischen Kategorien“ über den untersuchten sozialen Prozess (Schütze 1983: 286ff.).

## Literatur

- Arbeitsgruppe Bielefelder Soziologen (Hg.) (1976): Kommunikative Sozialforschung. München: Fink
- Aumüller, Matthias (Hg.) (2012): Narrativität als Begriff. Analysen und Anwendungsbeispiele zwischen philologischer und anthropologischer Orientierung. Berlin: de Gruyter
- Ihwe, Jens (Hg.) (1973): Literaturwissenschaft und Linguistik. Bd. 2. Frankfurt: Athenäum
- Kallmeyer, Werner/Schütze, Fritz (1977): Zur Konstitution von Kommunikationsschemata der Sachverhaltsdarstellung. In: Wegner (Hg.): 159-274
- Kohli, Martin/Robert, Günther (Hg.) (1984): Biographie und soziale Wirklichkeit. Stuttgart: Metzler
- Küsters, Ivonne (2009): Narrative Interviews. Grundlagen und Anwendungen. 2. Auflage. Wiesbaden: VS Verlag
- Küsters, Ivonne (2012): Narrationen als Repräsentationen sozialer Prozesse. Erzählungen als empirische Daten in der Soziologie. In: Aumüller (Hg.): 259-279
- Labov, William/Waletzky, Joshua (1973, englisches Original 1967): Erzählanalyse. Mündliche Versionen persönlicher Erfahrung. In: Ihwe (Hg.): 78-126
- Schütze, Fritz (1976a): Zur soziologischen und linguistischen Analyse von Erzählungen. In: Internationales Jahrbuch für Wissens- und Religionsoziologie. Band X. Opladen: Westdeutscher Verlag: 7-41
- Schütze, Fritz (1976b): Zur Hervorlockung und Analyse von Erzählungen thematisch relevanter Geschichten im Rahmen soziologischer Feldforschung. In: Arbeitsgruppe Bielefelder Soziologen (Hg.): 433-495
- Schütze, Fritz (1977): Die Technik des narrativen Interviews in Interaktionsfeldstudien – dargestellt an einem Projekt zur Erforschung von kommunalen Machtstrukturen. Bielefeld. Manuskript
- Schütze, Fritz (1983): Biographieforschung und narratives Interview. In: Neue Praxis. Kritische Zeitschrift für Sozialarbeit und Sozialpädagogik 13: 283-293
- Schütze, Fritz (1984): Kognitive Figuren des autobiographischen Stegreiferzählens. In: Kohli/Robert (Hg.): 78-117
- Wegner, Dirk (Hg.) (1977): Gesprächsanalysen. Hamburg: Buske

Susanne Vogl

Gruppendiskussionen (auch: Fokusgruppen, „focus groups“) sind geplante Diskussionen, um Einstellungen zu einem bestimmten, durch das Forschungsinteresse definierten Bereich in einer offenen, freundlichen Atmosphäre zu erheben. Dazu werden in einer Gruppe Kommunikationsprozesse initiiert, die einem alltäglichen Gespräch ähneln. Dabei geht es nicht (nur) um einen Austausch von Argumenten, sondern es wird auch erzählt, erinnert oder gegenseitig ergänzt. Entscheidend ist die Interaktion der Gruppenmitglieder, wodurch die Methode über die gleichzeitige Befragung (Helfferich, Kapitel 39 in diesem Band) mehrerer Personen hinausgeht.

Gruppendiskussionen sind meist qualitativ ausgerichtet und streben Natürlichkeit der Erhebungssituation, Kommunikativität und Offenheit an. Insbesondere durch die Interaktion der TeilnehmerInnen untereinander und der damit einhergehenden Beeinflussung sind Gruppendiskussionen realistischer und kommen einer natürlichen Gesprächssituation näher als beispielsweise Einzelinterviews. Bedingt durch den diskursiven Austausch können sich im Verlauf der Diskussion Argumente verändern.

Die Gruppendiskussion dient entweder als Medium, um Einzelmeinungen in validierter Form oder um Gruppenmeinungen losgelöst vom Individuum zu erheben. Außerdem werden entweder bereits vorhandene Meinungen, oder eben der Entstehungs- und Aushandlungsprozess von Meinungen untersucht (Flick 1995). Individuelle Einstellungen oder Narrationen subjektiver Erfahrungen werden dagegen mit Gruppendiskussionen nicht erhoben.

---

## 41.1 Konzeptionen und Erkenntnisziele

Die Ursprünge der Gruppendiskussion liegen im angloamerikanischen Raum, wo nach Kleingruppenexperimenten (Eifler, Kapitel 11 in diesem Band) die Erhebung von Informationen im Mittelpunkt stand. In Deutschland geht die Konzeption der Gruppendiskussion

im Wesentlichen auf Pollock (1955) und Mangold (1960) zurück. Pollock verfolgte das Modell des *Individuums in öffentlicher Auseinandersetzung*. Er untersuchte die nicht-öffentliche Meinung der einzelnen TeilnehmerInnen einer Diskussionsgruppe, die den TeilnehmerInnen selbst häufig erst in der Auseinandersetzung mit anderen bewusst wird. Trotz der Kritik an isolierten Interviewten in der Umfrageforschung, standen weiterhin Einzelmeinungen im Zentrum. Bei Mangold (1960) vollzog sich dann ein Wechsel von der Einzel- zur *Gruppenmeinung*. Er ging davon aus, dass durch die Selbstläufigkeit des Diskurses Gruppenmeinungen als Produkt kollektiver Interaktion gewissermaßen arbeits- teilig vorgetragen werden.

Eine Weiterentwicklung erfuhr das Gruppendiskussionsverfahren durch Bohnsack et al. (2010). Klassische Konzeptionen der Gruppendiskussion wurden im Hinblick auf ihr wissenssoziologisches Erkenntnisinteresse an kollektiven Orientierungsmustern erweitert.

---

## 41.2 Vor- und Nachteile

Im Vergleich zu anderen Befragungsarten besteht der größte Vorteil der Gruppendiskussion darin, kollektive Orientierungen sozusagen im Entstehungsprozess herausarbeiten zu können. Erst im Gespräch sieht man sich gezwungen, die eigene Meinung zu benennen und zu behaupten, wodurch tieferliegende Einstellungen und ein größerer Bereich von Reaktionsweisen zum Vorschein kommen. Im Prozess der Auseinandersetzung mit Anderen mag sich die Meinung zwar ändern, dafür zeichnet sie sich aber deutlicher und besser durchdacht ab. Die gegenseitige Beeinflussung der TeilnehmerInnen und die zwischen dem/der ModeratorIn und den Mitgliedern der Gruppe, die bei standardisierten Verfahren als Störvariable gesehen wird, ist bei Gruppendiskussionen konstitutiver Bestandteil des Verfahrens. Da diese Beeinflussung alltäglich ist, erscheinen Gruppendiskussionen realistischer und alltagsrelevanter als Einzelinterviews.

Zwar kann die Gruppensituation einerseits Meinungsausserungen anregen, andererseits aber auch verhindern. Die „öffentliche“ Atmosphäre und die eventuell damit verbundene soziale Erwünschtheit können „private“ Meinungsausserungen verhindern. Der größte Vorteil von Gruppendiskussionen, die Gruppendynamik, kann also auch Nachteile verursachen. Die Abhängigkeit der Ergebnisse von der Dynamik der jeweiligen Gruppe kann außerdem nicht nur die Vergleichbarkeit der Ergebnisse mit anderen Gruppen erschweren, sondern unter Umständen auch die externe Validität beeinträchtigen. Anpassungsmechanismen behindern individuelle Meinungsausserungen und MeinungsführerInnen können die Diskussion monopolisieren, wodurch die Vielfalt möglicher Ausserungen eingeschränkt ist (Lamnek 2005). Wenn Gruppen den Ansichten eines/er Meinungsführers/-in zu leicht folgen, kann dies zu extremen Gruppenmeinungen führen. Eine relativ hohe Schweigerquote kann belastend sein, außerdem kann Schweigen Folge einer abweichen den Meinung sein, die aus Angst vor negativen Sanktionen für sich behalten wird (siehe auch das Konzept der „Schweigespirale“ nach Noelle-Neumann). Nachteilig wirken auch

---

besonders hohe Ansprüche an den/die DiskussionsleiterIn, weil viele Entscheidungen zur Steuerung aus der Situation heraus getroffen werden müssen (Flick 1995).

Der ökonomische Vorteil, mehrere Personen gleichzeitig an einem Ort zu „befragen“, wird durch zeitaufwändigere Organisation, Transkription und Auswertung tendenziell relativiert.

---

### 41.3 Planung und Durchführung

Zwar kann im Rahmen explorativer Studien das Design für die Anwendung von Gruppendiskussionen maximal flexibel gehalten werden, Offenheit heißt jedoch nicht Beliebigkeit. Die Qualität von Gruppendiskussionen hängt stark von guter Planung ab. Es muss entschieden werden,

1. wer die beteiligten Personengruppen (TeilnehmerIn, ModeratorIn, TechnikerIn) sind und wie sie rekrutiert werden sollen,
2. wie die Gruppendiskussion hinsichtlich technischer und lokaler Bedingungen, Moderation und Teilnehmermotivation durchgeführt werden soll – Grad der Strukturierung, Grad der Moderatorenbeteiligung, Gruppengröße und Anzahl an Gruppen –,
3. welches die Erkenntnisabsichten sind,
4. wie transkribiert und analysiert wird und
5. wie die Ergebnisse dargestellt und präsentiert werden sollen.

Die Planung und konkrete Umsetzung muss immer auf Anforderungen an die Daten, Vergleichsmöglichkeiten, ethische Aspekte und Auswirkungen auf DiskussionsteilnehmerInnen abgestimmt werden.

---

### 41.4 Leitfaden

Gruppendiskussionen werden durch einen Leitfaden (Helfferich, Kapitel 39 in diesem Band) strukturiert, dem die klare Definition eines Forschungskonzeptes zugrunde liegt. Der Leitfaden ist in der Regel allgemein gehalten, stellt eine Liste mit offenen Fragen dar und wird flexibel gehandhabt. So kann der/die ModeratorIn auf weitere, nicht antizipierte Aspekte eingehen. Der Leitfaden soll Rahmen und kein Korsett sein (Lamnek 2005). In Abhängigkeit vom Vorwissen sowie der gewünschten Vergleichbarkeit mit anderen Gruppen (und der Erfahrung des/der Moderators/-in) ist der Leitfaden mehr oder weniger stark strukturiert. Meist werden nur grobe Rahmenthemen und eine geringe Anzahl an Einzelaspekten vorgegeben, um einzelne Punkte detailliert diskutieren zu können. Für eine ein- bis zweistündige Diskussion werden vier bis sechs distinkte Themen bzw. Fragen vorgegeben. Die Diskussion sollte sich dabei – wie bei einem Trichter – zunehmend thematisch verengen. Außerdem sollten bei der Ausgestaltung des Hauptteils der Gruppendiskussion verschiedene Phasen der Gruppendynamik berücksichtigt werden. Nach anfänglichen

Orientierungsphasen stellt sich häufig eine zunehmende Tendenz zur Konformität ein. Um diese zu durchbrechen oder die Gruppe zu einer kontroverseren Diskussion anzuregen, können vorab provokative Aussagen überlegt werden (Kühn/Koschel 2011).

---

## 41.5 Gruppengröße und Gruppenzusammensetzung

Sowohl die Gruppengröße als auch die Anzahl der durchzuführenden Gruppendiskussionen hängt von der Anzahl der gewünschten Vergleiche, vom Forschungsgegenstand, der Art der erhobenen Daten und der angestrebten Analyse ab. Als ideal gelten sechs bis zehn (max. zwölf) TeilnehmerInnen. Bei kleineren Gruppen kann es schwierig sein, eine Diskussion aufrechtzuerhalten, größere können schwerer kontrolliert werden. Um das Nichterscheinen mancher TeilnehmerInnen auszugleichen, sollten ca. 20% mehr rekrutiert werden.

TeilnehmerInnen werden in der Regel gezielt bewusst (Akremi, Kapitel 17 in diesem Band) oder nach den Prinzipien des Theoretical Samplings (Strübing, Kapitel 32 in diesem Band) ausgewählt. Dabei kann es sich um Realgruppen oder künstliche, homogene oder heterogene Gruppen handeln. *Realgruppen* sind Gruppen, die auch außerhalb der Diskussionssituation bestehen. Das Gegenteil sind *künstliche Gruppen*, die nur zum Zwecke der Diskussion zusammengestellt werden und deren Mitglieder sich nicht kennen. Einerseits sind Gruppendiskussionen mit Realgruppen realistischer und haben daher eine hohe externe Validität (Eifler, Kapitel 11 in diesem Band). Andererseits können Rollen- und Statusgefüge die Offenheit der Diskussion einschränken, weil Aussagen nicht immer sanktionsfrei bleiben (Lamnek 2005).

Letztlich muss die Gruppenzusammensetzung gewährleisten, dass die TeilnehmerInnen genügend Gemeinsamkeiten aufweisen, um über ein Thema zu sprechen. Ziel ist ein homogener Erfahrungshintergrund, aber keine homogenen Einstellungen (Morgan 2009). Wenn die TeilnehmerInnen sich – überspitzt gesagt – in allem einig sind, führt dies zu einer unproduktiven und oberflächlichen Diskussion.

Bei der Festlegung der Anzahl der Gruppendiskussionen wird in der Regel eine theoretische Sättigung angestrebt, d.h. bei der Durchführung jeder weiteren Gruppendiskussion werden praktisch keine neuen Erkenntnisse gewonnen. Als Richtwert haben sich laut Morgan (2009) drei bis fünf Gruppen bewährt.

---

## 41.6 ModeratorIn

Der/die ModeratorIn ist die zentrale Figur in der Gruppendiskussion. Er/sie ist maßgeblich für das Gelingen und die Verlässlichkeit der Ergebnisse verantwortlich. Er/sie muss zwar kein Experte auf dem Themengebiet sein, sollte aber dennoch eine gute Sach- und Gegenstandskompetenz haben, um die Diskussion verstehen und lenken zu können. Er/

sie verbindet zielgerichtetes Suchen und die Bereitschaft, sich treiben und überraschen zu lassen, und baut eine vertrauens- und verständnisvolle Atmosphäre auf.

Zu den Techniken der Moderation zählen (Kühn/Koschel 2011):

1. Rückkoppelndes Widerspiegeln: Paraphrasieren und aktives Zuhören,
2. Entschleunigung, um Details sichtbar zu machen,
3. projektive Techniken, um Emotionen aufzudecken,
4. Visualisierungen zur Strukturierung und Präzisierung und
5. (mit Vorsicht) Konfrontation und Provokation zur Vertiefung.

Typische Fehler bei der Moderation von Gruppendiskussionen sind nach Kühn und Koschel (2011):

1. zu strikte Orientierung am Leitfaden,
2. Verwicklung in Einzelgespräche,
3. Fokussierung auf die Gesprächsleitung und zu geringe Orientierung an der Aussagekraft von Beiträgen,
4. Bemühen um Unsichtbarkeit (man sieht sich als ModeratorIn als möglichen Störfaktor und will die Gruppe möglichst wenig beeinflussen),
5. zu geringe Berücksichtigung des Prozesscharakters (Phasen einer Diskussion und Gruppendynamik),
6. falsches Zeitmanagement,
7. zu schneller Themenwechsel bei Pausen oder schleppender Diskussion sowie
8. suggestives Auftreten.

Ein Praxisbeispiel illustriert einen Anwendungsbereich: In der Politikforschung wurden anhand von Gruppendiskussionen in einem österreichischen Bundesland verschiedene mögliche Positionierungen von Kandidaten in der Nationalratswahl bei unterschiedlichen Wählergruppen evaluiert (Breitenfelder et al. 2004). Ziel war eine zielgruppenspezifische strategische Positionierung des Spitzenkandidaten zu erarbeiten.

In einem zielgruppenspezifischen Ansatz wurden homogene Gruppen zusammengesetzt (z.B. „Arbeiter über 40 Jahre in der Region X“), mit denen je sehr strukturierte Gruppendiskussionen durchgeführt wurden, um ausführlich Images der Kandidaten, politische Themen und Inserate zu testen. Beispielsweise wurden Imagebilder anhand kurzer Geschichten über Spitzenkandidaten mit verschiedenen Fakten aus dessen Leben oder mit spezifischen Persönlichkeitsaspekten und Kompetenzen erprobt. Die Gruppen diskutierten anschließend, welches Konzept mehr Eindruck hinterließ und lieferten somit wichtige Hinweise auf eine zielgruppengenaue Positionierung eines Kandidaten. Durch die starke Strukturierung von Gruppendiskussionen wurden Vergleiche erleichtert und eine Bearbeitung vieler Themen in relativ kurzer Zeit ermöglicht, allerdings auf Kosten von Tiefe und Detailliertheit der Diskussion.

## Literaturverzeichnis

- Bohsack, Ralf/Przyborski, Aglaja/Schäffer, Burkhard (Hg.) (2010): Das Gruppendiskussionsverfahren in der Forschungspraxis. Opladen: Budrich
- Breitenfelder, Ursula/Hofinger, Christoph/Kaupa, Isabella/Picker, Ruth (2004): Fokusgruppen im politischen Forschungs- und Beratungsprozess. In: FQS 5 (2): Artikel 25.  
URL: <http://www.qualitative-research.net/index.php/fqs/article/download/591/1284>
- Flick, Uwe (1995): Qualitative Forschung. Reinbeck: Rowohlt
- Kühn, Thomas/Koschel, Kay-Volker (2011): Gruppendiskussionen. Wiesbaden: VS Verlag
- Lamnek, Siegfried (2005): Gruppendiskussion. Weinheim: Beltz
- Loos, Peter/Schäffer, Burkhard (2001): Das Gruppendiskussionsverfahren. Opladen: Leske + Budrich
- Mangold, Werner (1960): Gegenstand und Methode des Gruppendiskussionsverfahrens.  
Frankfurt a.M.: Europ. Verl.-Anst.
- Morgan, David L. (2009): Focus Groups as Qualitative Eesearch. Thousand Oaks: Sage
- Pollock, Friedrich (1955): Gruppenexperiment. Frankfurt a. M.: Europ. Verl.-Anst.

Michael Häder und Sabine Häder

---

## 42.1 Einleitung

Im Jahr 1998 wurden verschiedene Experten danach gefragt, wie sie sich die Zukunft der Festnetztelefonie im Jahr 2005 vorstellen. Den Hintergrund für diese Studie bildete die Vermutung, dass sich immer mehr Personen dazu entschließen würden, ihren Festnetztelefonanschluss abzumelden und nur noch über den Mobilfunk zu telefonieren. Dies brächte aber weitreichende Folgen für die empirische Sozialforschung, die sich häufig telefonischer Befragungen bedient, mit sich. Neben zahlreichen anderen Einschätzungen wurde auch nach dem für 2005 vermuteten Anteil an Haushalten mit Festnetzanschlüssen in Deutschland gefragt. Den Teilnehmern an der Delphi-Befragung wurde der Mittelwert aller Schätzungen dann rückgemeldet. Vor dem Hintergrund dieser Information wurden sie dazu aufgefordert, die entsprechende Schätzung zu wiederholen. Im Ergebnis zeigte sich, dass zum gefragten Zeitpunkt nur noch etwa 91 Prozent der Haushalte über das Festnetz telefonieren würden. Es stellte sich schließlich heraus, dass dieser Wert relativ exakt dem tatsächlichen Anteil entsprach und die Forschungen zu Mobilfunkbefragungen rechtzeitig auf den Weg gebracht werden konnten. Wie kam ein solches Ergebnis zustande und wie funktionieren Delphi-Befragungen generell?

Bei der Delphi-Methode handelt es sich um einen relativ stark strukturierten Gruppenkommunikationsprozess (Vogl, Kapitel 41 in diesem Band), in dessen Verlauf Sachverhalte, über die naturgemäß unsicheres und unvollständiges Wissen existiert, von Experten (Helfferich, Kapitel 39 in diesem Band) beurteilt werden. Die beiden wesentlichen *Aspekte des Delphi-Ansatzes* sind damit die Gruppenkommunikation und die Strukturierung unsicheren Wissens.

Der Grundgedanke, der der Delphi-Methode zugrunde liegt, geht davon aus, dass Experten über die Fähigkeit verfügen, unvollständiges bzw. unsicheres Wissen aufgrund probabilistischer mentaler Modelle zu rekonstruieren (Häder 2014: 37ff., Gigerenzer et al. 1991). Wenn für diese Rekonstruktion zusätzlich die in den Rückmeldungen enthaltenen

Informationen genutzt werden, so wird weiter angenommen, kommt es zu einer Qualifizierung dieser Modelle und damit in den Folgebefragungen zu höherwertigen Urteilen.

Delphi-Befragungen können auf sehr unterschiedliche Art und Weise durchgeführt werden. Die folgenden *Designelemente* sind jedoch für die klassische Variante einer Delphi-Befragung charakteristisch:

1. Es erfolgt die Operationalisierung der allgemeinen Frage- bzw. Problemstellung mit dem Ziel, *konkrete Kriterien* abzuleiten, die den Experten im Rahmen einer quantifizierenden Befragung für eine Beurteilung vorgelegt werden können. Dieser erste Schritt kann sowohl von dem die Delphi-Befragung veranstaltenden Forscherteam (dieses wird oft auch als Monitoring-Team bezeichnet) selbst, als auch mithilfe einer offenen, qualitativen Befragung von (externen) Experten (Helfferich, Kapitel 39 in diesem Band) bewältigt werden. Für Prognosezwecke könnte beispielsweise ohne allzu konkrete weitere Vorgaben danach gefragt werden, welche Ereignisse oder Entdeckungen in einem bestimmten Zeitraum auf einem bestimmten Gebiet zu erwarten sind. So lautete zum Beispiel 1998 die entsprechende Frage in der oben bereits erwähnten Studie zur Zukunft des Mobilfunkens: „Zur Zeit verfügen in Deutschland ca. 20% der Menschen über ein Mobilfunktelefon (Handy). In ca. 95% der Haushalte befinden sich Festnetzanschlüsse. Wie werden sich Ihrer Meinung nach diese Anteile in den nächsten fünf Jahren entwickeln? Schätzen Sie bitte den Anteil der Menschen, die dann über ein Mobiltelefon verfügen werden, in Prozent: ...“ (Häder 2000: 30).
2. Dem schließt sich die *Ausarbeitung eines standardisierten Fragebogens* (Reinecke, Kapitel 44 in diesem Band) an. Dieser dient dazu, Experten anonym nach ihren Meinungen zu den interessierenden Sachverhalten zu befragen. Dabei kann es sich beispielsweise auch um die Schätzung des konkreten Jahres handeln, bis zu dem sich eine bestimmte Entdeckung oder Erfahrung durchgesetzt haben wird.
3. Im Folgenden geschieht eine Aufbereitung der Befragungsergebnisse durch das die Befragung veranstaltende Forscherteam sowie eine *anonymisierte Rückmeldung der Ergebnisse an die beteiligten Befragten*. Dabei erhält dieses Feedback lediglich eine Information zur Gruppenmeinung, nicht jedoch zur vorausgegangenen individuellen Schätzung des jeweiligen Teilnehmers.
4. Schließlich kommt es zu einer *ein- oder mehrmaligen Wiederholung der Befragung* (Welten) auf der Grundlage der von den Experten über diese Rückinformation gewonnenen (neuen) Erkenntnisse bis zum Erreichen eines vorher festgelegten Abbruchkriteriums. Auf diese Weise werden die Teilnehmer zur nochmaligen Auseinandersetzung mit dem Gegenstand der Befragung angeregt. Diese Aktion führt dann zu einer Verbesserung der ursprünglichen Urteile. So lag nach der ersten Welle der Median der Schätzungen zum Anteil der Haushalte, die 2005 exklusiv über Mobilfunk telefonieren würden, bei 95 Prozent und nach der zweiten Welle bei 92 Prozent. Dies war ein Wert, der die kommende Wirklichkeit besser abbildete.

Dieses klassische Design wird in der Praxis durch zahlreiche Varianten ergänzt. So existieren unterschiedliche Ansichten über die *erforderliche Zahl, die optimale Struktur und die Art und Weise der Auswahl der Experten(-gruppe)*. Weiterhin gibt es unterschiedliche Auslegungen über die *erforderliche Anzahl der Wellen*. Ebenso wird das den Teilnehmern übermittelte *Feedback* sehr verschieden gestaltet, teilweise werden ergänzend Selbsteinschätzungen der Experten über deren Kompetenz erfragt. Zudem sind von den Teilnehmern im Rahmen der Befragung relativ unterschiedliche *Aufgaben* zu lösen. Schließlich sind die Ansichten über das *Abbruchkriterium* der Studie nicht einheitlich usw.

Für einen Überblick zu den bisherigen Anwendungen des Delphi-Ansatzes bietet sich ein Konzept an, das von vier Typen ausgeht, die im Folgenden dargestellt werden.

### **42.1.1 Typ 1: Delphi-Befragungen zur Ideenaggregation**

Ein erstes Ziel, mit dem Delphi-Befragungen veranstaltet werden können, ist die Aggregation von Ideen. Die Besonderheit von Delphi-Projekten, die diesem Zweck dienen, besteht darin, dass sie – anders als beim klassischen Design – einen *ausschließlich qualitativen Ansatz* darstellen. Hier wird auf die quantifizierenden Wellen verzichtet. Stattdessen werden bereits die Ergebnisse der qualitativen Einschätzungen den Experten auf geeignete Weise rückgemeldet und danach die Fragestellungen erneut – qualitativ, d.h. weitgehend unstrukturiert – erhoben.

Ein typisches Ideenaggregations-Delphi beschreibt Hasse (1999: 213ff.). Mithilfe einer Befragung sollten Argumente im Rahmen der Diskussion um die landschaftsästhetische Beurteilung von Windkraftanlagen gesammelt werden.

### **42.1.2 Typ 2: Delphi-Befragungen für eine möglichst exakte Vorhersage eines unsicheren Sachverhalts bzw. für dessen genaue(re) Bestimmung**

Hier soll erhöhte *Klarheit über eine bestimmte, diffuse Angelegenheit* geschaffen werden. Mit dem Ziel, einen unbekannten Sachverhalt möglichst exakt vorherzusagen, sind gegen Ende der 1940er-Jahre die ersten wissenschaftlichen Delphi-Befragungen angetreten. Damals ging es im Rahmen von Arbeiten der RAND-Corporation zur Landesverteidigung um die Ermittlung möglicher strategischer Angriffsziele sowjetischer Raketen in den USA. Ein typisches Vorhersage-Delphi beschreibt auch Janssen (1978). Das Ziel dieser Befragung waren Preisprognosen in den Obstbaugebieten am Bodensee und an der Niederelbe. Wöchentlich trafen sich die Obstpreisnotierungskommissionen nach dem Muster des Delphi-Designs und gaben Prognosen zur Preisentwicklung in der folgenden Woche ab.

### **42.1.3 Typ 3: Delphi-Befragungen zur Ermittlung und Qualifikation der Ansichten einer Expertengruppe über einen diffusen Sachverhalt**

Bei diesem Typ sollen die Meinungen einer konkret bestimmbarer Expertengruppe erhoben und dabei qualifiziert werden. Die Resultate solcher Studien dienen dann beispielsweise dazu, um gezielte Schlussfolgerungen für erforderliche Interventionen abzuleiten, um auf ein auf diese Weise ermitteltes Problem zu reagieren, oder um eine Sensibilisierung gegenüber befürchteten Fehlentwicklungen zu erreichen. Als ein typisches Beispiel kann wiederum die Delphi-Befragung zur Zukunft des Mobilfunkes herangezogen werden. Eine Konsequenz aus den Befunden der genannten Studie war die Aufsetzung von methodischen Forschungen (Stichproben, Antwortqualität usw.) zu den Möglichkeiten von telefonischen Befragungen auch über den Mobilfunk.

### **42.1.4 Typ 4: Delphi-Befragungen zur Konsensbildung unter den Teilnehmern**

Bei Delphi-Befragungen werden infolge des Feedbacks Gruppenprozesse ausgelöst. Diese dienen dazu, das Ergebnis der Befragung zu qualifizieren. Es gibt auch Ansätze, bei denen Delphi-Studien explizit das Ziel verfolgen, ein möglichst hohes Maß an Konsens unter den Teilnehmern zu schaffen. Beispielsweise im Rahmen der Vorbereitung einer demokratischen Entscheidungsfindung können sich solche Befragungen als sinnvoll erweisen. Ein typisches Konsens-Delphi beschreiben Mettler und Baumgartner (1997: VIIff.). Den Autoren geht es in ihrem Projekt um konkrete Empfehlungen für eine konsensorientierte Politik im Bereich Mikroelektronik und Arbeitsmarkt im Bundesland Nordrhein-Westfalen.

---

## **42.2 Abgrenzung des Delphi-Ansatzes gegenüber qualitativen Befragungsansätzen**

Delphi-Befragungen weisen einige Besonderheiten auf, die eine gewisse Abgrenzung gegenüber anderen Ansätzen erfordern. Im Unterschied zu qualitativen Verfahren (Akremi, Kapitel 17 in diesem Band) bedienen sich Delphi-Befragungen teilweise *großer* Expertengruppen. So nahmen im Jahr 1996 beispielsweise 4.868 Experten an der ersten und 4.196 Experten an der zweiten Welle des sechsten Technologie-Delphi in Japan teil (NISTEP 1997). Die Befragungsunterlagen sind – ebenfalls im Unterschied zu den typischen qualitativen Ansätzen (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) – *hochstandardisiert*. Schließlich wird versucht, eine Gesamtheit zu definieren, aus der die Teilnehmer an der Befragung systematisch rekrutiert werden. So könnten beispielsweise alle Autoren eines bestimmten Zeitraumes in einer (oder in mehreren) Fachzeitschrift(en) die Grundgesamtheit bilden. Hier wäre der Einsatz statischer Auswahlverfahren (geschichtete Zufallsauswahl) (Häder/Häder, Kapitel 18 in diesem Band) prinzipiell denkbar.

Delphi-Befragungen weisen auch gegenüber *Leitfadeninterviews* (Helfferich, Kapitel 39 in diesem Band) bestimmte Besonderheiten auf. Eine Besonderheit ist die Wiederholung der Befragung vor dem Hintergrund der den Befragten mitgeteilten Ergebnisse der ersten Welle. Eine solche erneute Erhebung ermöglicht den Mitgliedern des Panels (Schupp, Kapitel 73 in diesem Band) ein nochmaliges Auseinandersetzen mit der Fragestellung und führt – dies zeigen zahlreiche empirische Experimente (für einen Überblick siehe Häder 2009: 191ff.) – zu höherwertigen Urteilen. Zweitens zeichnen sich Delphi-Befragungen – wie bereits festgestellt – durch einen hohen Grad an Standardisierung aus. Daraus resultiert die Erwartung, dass die zur Diskussion stehende Problematik von den Teilnehmern stringenter bearbeitet wird als bei Leitfadeninterviews. Die Standardisierung hat drittens Konsequenzen auch für die Auswertung der Erhebung.

Gegenüber *Gruppendiskussionen* (Vogl, Kapitel 41 in diesem Band) besitzen Delphi-Befragungen ebenfalls Besonderheiten. Dabei handelt es sich insbesondere um die Anonymität der Teilnehmer untereinander. Diese ermöglicht es den Teilnehmern, Meinungsänderungen vorzunehmen, ohne dabei von den anderen Mitgliedern des Expertenpanels identifiziert werden zu können. Weiterhin trägt die Anonymität zur Vermeidung von Meinungsführerschaften bei, wie sie bei offenen Gruppendiskussionen unvermeidlich sind.

---

### 42.3 Auswertungsstrategien

Die Ergebnisse von Delphi-Befragungen werden vor allem mithilfe der deskriptiven Statistik bearbeitet. Wichtig sind zunächst geeignete Maße, um die ermittelten Antwortverteilungen darzustellen. In Abhängigkeit von der Datenqualität betrifft dies die entsprechenden Mittelwerte und Streuungsmaße. Dazu bieten sich weiterhin grafische Abbildungen an. Außerdem haben sich folgende Strategien bewährt:

Bei der Darstellung der Schätzungen von Subpopulationen ist es gebräuchlich, die Ergebnisse jener Expertengruppe, die in ihren Urteilen besonders sicher war bzw. jener Teilnehmer, die als Insider (vermutlich) besonders kompetent Auskunft geben können, gesondert auszuweisen. Dies kann beispielsweise relativ einfach dadurch geschehen, dass die jeweiligen Mittelwerte bzw. Streuungsmaße separat angegeben werden.

Eine Gewichtung (Kiesl, Kapitel 24 in diesem Band) der inhaltlichen Ergebnisse der Befragung anhand der Kompetenzfrage schlagen Loveridge et al. (1995: 11) vor. Einer solchen Idee liegt die Vorstellung zugrunde, dass kompetentere Experten höherwertige Schätzungen abgeben als andere. Demzufolge wäre es dann sinnvoll, diese Auskünfte stärker zu gewichten als die anderer Teilnehmer.

Faktorenanalysen (Blasius, Kapitel 83 in diesem Band) und Clusteranalysen (Blasius und Baur, Kapitel 79 in diesem Band) können genutzt werden, um die Struktur der Urteile bzw. die der Experten zu ermitteln. Entsprechende Beispiele finden sich bei Kinast (2000: 146) sowie bei Grupp u.a. (2000). Zudem kann der Scree-Test bei Delphi-Studien eingesetzt werden, um von einer ursprünglich relativ großen Anzahl von Items nach jeder Welle jene auszusondern, deren Bewertung sich aufgrund der Expertenurteile als trivial erwiesen hat.

## Literatur

- Gigerenzer, Gerd/Hoffrage, Ulrich/Kleinbölting, Heinz (1991): Probabilistic Mental Models. In: *Psychological Review* 98 (4): 506-528
- Grupp, Hariolf/Blind, Knut/Cuhls, Kerstin (2000): Analyse von Meinungsdisparitäten in der Technikbewertung mit der Delphi-Methode. In: Häder, Michael/Häder, Sabine (Hg.): *Die Delphi-Technik in den Sozialwissenschaften*. Opladen: Westdeutscher Verlag. 43-66
- Häder, Michael (2000): Mobilfunk verdrängt Festnetz. ZUMA-Arbeitsbericht 00/05. URL: [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/zuma\\_arbeitsberichte/00\\_05.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/zuma_arbeitsberichte/00_05.pdf)
- Häder, Michael (2014): *Delphi-Befragungen*. Wiesbaden: VS Verlag
- Hasse, Jürgen (1999): Bildstörung. Oldenburg: Bibliotheks- und Informationssystem der Universität Oldenburg
- Janssen, Hans (1978): Application of the Delphi Method to Short-Range Price Predictions on the Fruit Market. In: *Acta Horticulturae* 77 (1): 223-230
- Kinast, Kurt (2000): Interne und Externe Wirkungszusammenhänge in Geschäftsbeziehungen. In: Häder, Michael/Häder, Sabine (Hg.): *Die Delphi-Technik in den Sozialwissenschaften*. Opladen: Westdeutscher Verlag. 133-160
- Loveridge, Denis/Georghiou, Luke/Nedeva, Maria (1995): United Kingdom Technology Foresight Programme Delphi Survey. University of Manchester
- Mettler, Peter H./Baumgartner, Thomas (1997): Partizipation als Entscheidungshilfe. Opladen: Westdeutscher Verlag
- NISTEP (1997): The Sixth Technology Forecast Survey. NISTEP Report No. 52. Tokyo: National Institute of Science and Technology Policy (NISTEP) Science and Technology Agency

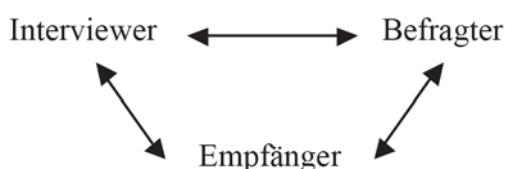
Jürgen Friedrichs und Ulrich Schwinges

Das journalistische Interview ist wie das wissenschaftliche durch eine asymmetrische Kommunikation gekennzeichnet: Der Interviewer fragt, der/die Befragte antwortet. Der wesentliche Unterschied zwischen diesen beiden Interviewformen besteht darin, dass das journalistische Interview für Dritte – die Empfänger – geführt, manchmal geradezu inszeniert wird. Leser/innen, Hörer/innen oder Zuschauer/innen sollen über einen Sachverhalt oder die Meinung, Einschätzung, Bewertung der oder des Befragten informiert werden. Hingegen wird das wissenschaftliche Interview für die fragenden Forscher geführt.

Diese Bedingung verändert die Rollen von Interviewer und Befragtem. Dem Interviewer kommt damit die Aufgabe zu, Fragen für die in der jeweiligen Sendung oder den Beitrag in einer Zeitung/Zeitschrift zu einem gegebenen Problem für die Empfänger zu stellen. Diese Konstellation lässt sich als Beziehungsdreieck darstellen.

Jeder der drei Beteiligten hat unterschiedliche Kenntnisse, Erwartungen, Einstellungen und (u.a. politische) Interessen. Wie sich daraus folgern lässt, müssen die Interessen von Journalisten, Befragten und Empfängern (ebenso wie unter den Empfängern) nicht übereinstimmen. So mag der Befragte eine Botschaft senden wollen, dem Interviewer könnte es vor allem wichtig sein, diesen Gesprächspartner „gehabt zu haben“, und die Empfänger wären gerne über das angekündigte Thema informiert worden – was aber aufgrund der unterschiedlichen Interessen von Interviewer und Befragtem nicht oder nur unzureichend geschieht – so etwa am 25.9.2011, als die Bundeskanzlerin Angela Merkel in Günter Jauchs ARD-Sendung zu Gast war und Jauch keine Nachfragen zu ihrer Politik stellte.

**Abb. 43.1** Beziehungsdreieck



Wenn der Interviewer sein Interview empfängerorientiert anlegen will, sollte er zumindest wissen, wie alt und vorgebildet oder informiert etwa die Leser/innen seines Artikels sind, um einschätzen zu können, welche Kenntnisse die Rezipienten von dem Thema des Interviews haben dürften. Sind diese Vorkenntnisse gering, wird man entweder eine längere Anmoderation, bei Zeitungen und Zeitschriften „Anlauf“ benötigen. Bei komplizierteren Themen, wie z.B. einem Interview zu den Ursachen der Finanzkrise, wird man in Hörfunk und TV einen „Einspieler“, also einen erläuternden Beitrag, vor das Interview setzen, in dem u. a. erklärt wird, was „toxische Papiere“ sind. Nach unseren Erfahrungen kennen aber nur wenige Radio- und Fernseh-Journalisten die sozio-demographische Struktur der Empfänger ihrer jeweiligen Sendungen – und überschätzen vermutlich das Vorwissen der Empfänger/innen.

Wie das leitfadengestützte Interview hat das journalistische Interview keinen Fragebogen. An dessen Stelle tritt das Interview-Konzept. Dieses Konzept wird in mehreren Schritten erarbeitet. Zu einem gegebenen Problem, z.B. der Einführung der Ganztagschule, werden dessen Teilprobleme oder Aspekte bestimmt/erarbeitet, hier z.B.: Lehrkräfte am Nachmittag? Nehmen alle Schüler/innen am Unterricht nachmittags teil? Wo ist die Küche? Wer übernimmt die Betreuung mittags? Aus diesen Aspekten werden nur diejenigen (oder sogar nur diejenige/derjenige) ausgesucht, die für den aktuellen Stand des Problems und die jeweilige Sendung und somit für die Empfänger wichtig sind. Der Interviewer muss nun ein Informationsziel festlegen, d.h. bestimmen, worüber die Empfänger am Ende des Interviews informiert sein sollen. Diesem Ziel werden dann einzelne Fragen in einer überlegten Reihenfolge zugeordnet, z.B. von allgemeinen zu speziellen Fragen. Gemeinhin wird das Interview umso informativer sein, je weniger Aspekte es enthält, je stärker es sich also thematisch beschränkt. Nur dann nämlich hat die Interviewerin/der Interviewer bei einem Standard-Interview von 3-4 Minuten Zeit genug, um bei unklaren oder ausweichenden Antworten Nachfragen stellen zu können, z.B. „Heißt das, dass...“ oder „Sind Sie nun dagegen oder dafür?“ Der Aufbau des Interviews ist bei den Hörfunk- und TV-Interviews wichtiger als bei den Print-Interviews, weil die elektronischen Interviews ein linearer Prozess der Rezeption sind – man kann nicht noch einmal zurück gehen – und man ihnen deshalb besonders folgen müssen.

Das weist auf ein weiteres wichtiges Unterscheidungsmerkmal hin – journalistische Interviews, vor allem in ihrer live-Variante, sind in ein sehr enges Zeitkorsett gepresst – eine Bedingung, die die meisten wissenschaftlichen Interviews nicht erfüllen müssen.

Wie bei dem wissenschaftlichen Interview hängt auch hier das Konzept davon ab, wie lang das Interview sein soll; im Fernsehen dauert es in der Regel 1½ bis 4 Minuten, im Hörfunk in morgendlichen Informationssendungen bis zu 8 Minuten.

## 43.1 Typen von Interviews

Es gibt zahlreiche Vorschläge, wie sich journalistische Interviews typisieren lassen. Wir nehmen nur zwei Differenzierungen auf: Informationsinterview vs. kontroverses Interview und Sachinterview vs. Interview zur Person.

### 43.1.1 Informationsinterview vs. kontroverses Interview

Journalistische Interviews lassen sich auf einem Kontinuum anordnen (und führen): am einen Ende steht das Informationsinterview. Es ist eine Beschreibung eines Sachverhalts, gekennzeichnet durch Fragen „Wie?“, „Warum?“, „Was?“ und „Wo?“. Am anderen Ende steht das kontroverse Interview. Es richtet sich auf Begründungen und Rechtfertigungen. Kennzeichnend sind hier die Fragen „Warum?“, „Weshalb?“ oder „Gab es noch andere Gründe?“. Zwischen beiden Formen sind unterschiedliche Interviews möglich. So kann ein weitgehend als Informationsinterview angelegtes Interview durch die Frage „Warum haben Sie so entschieden?“ mehr Spannung erhalten und rückt in die Nähe eines kontroversen Interviews.

Im Unterschied zum wissenschaftlichen Interview wird vor allem im kontroversen journalistischen Interview häufiger der Fragetypr „Information plus Frage“ verwendet. Ein (fiktives) Beispiel: Ein Interview mit Walter Ulbricht im August 1961 „Sie haben im Dezember 1960 gesagt, es werde in Berlin keine Mauer geben. Nun bauen Sie doch eine. Warum?“

### 43.1.2 Interviews zur Sache vs. Interview zur Person

Sowohl Informationsinterviews als auch kontroverse Interviews sind in aller Regel Interviews zur Sache. Eine weitere Typologie richtet sich daher auf die Trennung von Sachinterviews und Interviews zur Person. Interviews zur Person sollen einen oder mehrere Aspekte einer Person erhellen. Es sind häufig prominente Personen, deren Leben dargestellt werden soll. Die Frage, die hierfür den Ausgangspunkt bildet, lautet „What makes him/her tick?“. Diese Interviews variieren in der Länge, zum Teil unterbrochen durch eingespielte Dokumente und Filmsequenzen aus dem Leben der Person. Berühmt sind bis heute die bis zu 45 Minuten langen TV-Interviews zur Person z.B. von Günter Gaus, Margret Dünser, Georg Stefan Troller, Gero von Böhm und auch die von Franz Xaver Gernstl. Es gibt aber auch kürzere Interviews zur Person. Beispiele hierfür sind Interviews mit Bands oder mit einem Filmregisseur, der sich entschlossen hat, in die USA zu gehen oder einem Schauspieler, der nach den Schwierigkeiten seiner Rolle gefragt wird, oder einer Mutter, die ihr Kind vernachlässigt hat. Gelegentlich geht solchen Interviews ein Bericht voraus, das Interview dient dann dazu, einen Aspekt des Beitrags, eben eine Person, zu vertiefen.

### 43.1.3 Weitere Formen

Eine Mischform stellen Interviews zum Beruf und zur Person dar. In ihnen wird ein Beruf anhand einer Person dargestellt, z.B. die Inhaberin einer Agentur für Zeitarbeit. Dazu werden sowohl Sachfragen als Fragen zur Person gestellt (z.B. „Warum haben Sie dann doch Jura studiert?“). Auch die Interviews in Talkshows sind eine Mischform, weil sowohl Sachfragen („Wann wussten Sie von dieser Entscheidung?“) als auch Fragen zur Person („Was haben Sie in dem Moment empfunden?“) gestellt werden.

Alle drei Formen, „das reine“ Interview zur Person, das Interview zur Sache und zur Person als auch das Interview in der Talkshow haben gemeinsam über Erlebnisse und Entscheidungen von Personen zu informieren. Vermutlich besteht ein Reiz dieser Formen und Fragen darin, den Empfängern zu ermöglichen – oder sogar zu suggerieren –, sich mit den Problemen von Personen zu identifizieren, sie zumindest nachvollziehen zu können, und nun gespannt darauf zu sein, wie die befragte Person ein Dilemma gelöst hat – oder auch nur den Voyeurismus der Empfänger zu bedienen – eine Funktion, die ein sozialwissenschaftliches Interview niemals haben sollte.

## Literatur

- Bommert, Hanko/Kleyböcker, Ralf/Voß-Frick, Andrea (2002): TV-Interviews im Urteil der Zuschauer. Münster: LIT Verlag
- Friedrichs, Jürgen/Schwinges, Ulrich (2009): Das journalistische Interview. 3. überarb. Auflage. Wiesbaden: VS Verlag
- Haller, Michael (2001): Das Interview. Ein Handbuch für Journalisten. 3., überarb. Auflage. München: UVK

---

## **Teil 5**

### **Standardisierte Befragung**

Jost Reinecke

---

## 44.1 Die grundlegende Charakteristik der standardisierten Befragung

Die Befragung kann als ein klassisches Instrument der Datenerhebung für die empirisch orientierten Disziplinen der Sozial- und Wirtschaftswissenschaften bezeichnet werden. Insbesondere durch den Einsatz von Befragungen in der Markt- und Meinungsforschung (Ziegler, Kapitel 10 in diesem Band) hat ihre Bedeutung stark zugenommen. Auch wenn sich die Befragungsformen in den letzten 40 Jahren weiterentwickelt und diversifiziert haben, ist die von Scheuch (1973: 70) aufgestellte Definition der Befragung als Interview auch heute noch zutreffend: „Unter Interview als Forschungsinstrument sei hier verstanden ein planmäßiges Vorgehen mit wissenschaftlicher Zielsetzung, bei dem die Versuchsperson durch eine Reihe gezielter Fragen oder mitgeteilter Stimuli zu verbalen Informationen veranlasst werden soll.“

Die Befragung ist demnach das Ergebnis einer Reaktion seitens der befragten Person basierend auf einem mehr oder weniger komplexen Kommunikationsprozess. Ein wichtiges Unterscheidungskriterium für die Formen der Befragung ist neben den Kommunikationsarten der Strukturierungs- und Standardisierungsgrad. Während innerhalb der qualitativ orientierten Sozialforschung Befragungstechniken mit geringer Standardisierung verbreitet sind (z.B. das Leitfadeninterview, Helfferich, Kapitel 39 in diesem Band), wird innerhalb der quantitativen Sozialforschung eine möglichst hohe Standardisierung angestrebt. Mit der Standardisierung werden der Grad der Festlegung des Fragetextes (Porst, Kapitel 50 in diesem Band), die dazu gehörenden Antwortkategorien (Züll/Menold, Franzen, Kapitel 51 und 52 in diesem Band) und die Reihenfolge im Fragebogen (Klöckner/Friedrichs, Kapitel 49 in diesem Band) charakterisiert. Da sich standardisierte Befragungen in der Regel auf umfangreiche Populationen (Häder/Häder, Kapitel 18 in diesem Band) beziehen, wird der Herstellung gleicher Bedingungen im Befragungsprozess ein hoher Stellenwert eingeräumt. Durch die Standardisierung ist für die Datenauswertung

sichergestellt, dass unterschiedliche Antworten auf eine Frage auch tatsächlich auf unterschiedliche Angaben der befragten Personen zurückzuführen sind und nicht auf unterschiedliche Bedingungen während der Befragungssituation. Die standardisierte Befragung ist eine der gebräuchlichsten Erhebungsmethoden, die zugleich methodologisch am besten erforscht ist.

---

## 44.2 Die standardisierte Befragung im Forschungsprozess

Der Forschungsprozess kann mehr oder weniger differenziert in fünf Phasen unterteilt werden (Diekmann 2007: 192):

- Formulierung und Präzisierung des Forschungsproblems (Stein, Kapitel 7 in diesem Band),
- Planung und Vorbereitung der Erhebung (Stein, Kapitel 7 in diesem Band),
- Datenerhebung,
- Datenaufbereitung (Lück/Landrock, Kapitel 28 in diesem Band)
- Datenauswertung und
- Berichterstattung (Friedrichs, Kapitel 16 in diesem Band).

Mit der Vorbereitung der Erhebung wird die Entscheidung getroffen, welche Art von Fragebogen eingesetzt werden soll, welche Messinstrumente (Burzan, Kapitel 81 in diesem Band) der Fragebogen enthält, wie das Untersuchungsdesign (einmalige oder mehrfache Erhebung, Stein, Kapitel 7 in diesem Band) gestaltet wird, auf welche Population sich die Stichprobe bezieht (Häder/Häder, Kapitel 18 in diesem Band) und letztendlich welche Befragungsform (mündliche, telefonische, internetgestützte oder schriftliche Befragung) eingesetzt werden soll. Besonderes Augenmerk muss auf die Konstruktion von standardisierten Fragebögen gelegt werden. Dies betrifft sowohl die inhaltliche Komplexität des Fragebogens als auch die technische Umsetzung. Ein selbst zu administrierender Fragebogen (Reuband, Wagner/ Hering, Kapitel 47 und 48 in diesem Band) hat höhere Anforderungen an Einfachheit und Verständlichkeit als ein Fragebogen, durch den ein Interviewer (Glantz/Michael, Kapitel 21 in diesem Band) die befragende Person führt. In jedem Fall ist die Vereinbarkeit der Fragebogeninhalte mit forschungsethischen Grundsätzen (Friedrichs, Kapitel 3 in diesem Band) und mit datenschutzrechtlichen Erfordernissen (Mühlchen, Kapitel 4 in diesem Band) zu gewährleisten.

Die theoretische Arbeit muss für eine standardisierte Befragung – im Unterschied zu offenen und teilstandardisierten Befragungen – vor der Konstruktion der Fragebögen erfolgen. Dies bedeutet, dass alle für die inhaltlichen Fragestellungen benötigten Konstrukte in operationalisierter Form als getestete und bewährte Fragen und Skalen (Latcheva/Davidov, Kapitel 55 in diesem Band) vorliegen müssen, damit diese im Fragebogen berücksichtigt werden können. Bevor ein standardisierter Fragebogen in die Feldphase geht, sind oft mehrere Pretests (Weichbold, Kapitel 19 in diesem Band) notwendig, um mögliche Probleme bei der Durchführung der Befragung erfassen und beheben zu können.

Nach Scholl (2003: 26) streben standardisierte Befragungen in erster Linie den Vergleich zwischen den Untersuchungsobjekten an, wobei Stichprobenziehung, der Fragebogen und die Interaktion zwischen Interviewer und Befragtem dazu dienen soll, Vereinheitlichung und Objektivierbarkeit der Untersuchungsergebnisse (Krebs/Menold, Kapitel 30 in diesem Band) zu sichern. Zu den *Voraussetzungen für eine erfolgreiche Durchführung von Befragungen* sind folgende Bedingungen zu nennen (Diekmann 2007: 440f.):

1. Die Existenz einer „gemeinsamen Sprache“, d.h. Interviewer (Glantz/Michael, Kapitel 21 in diesem Band) und die befragte Person interpretieren die Bedeutung von Fragen und Antworten in gleicher Weise
2. Die Kooperation des Befragten (Hlawatsch/Krickl, Engel/Schmidt, Kapitel 20 und 23 in diesem Band), sein Interesse am Thema und die damit verbundene Unterstützung einer wissenschaftlichen Befragung
3. Anerkennung, dass die eigene Meinung vom Interviewer als wichtig erachtet wird und
4. Die Existenz einer „Norm der Aufrichtigkeit“, d.h. die verinnerlichte Tendenz des Befragten, dem Interviewer eine aufrichtige Antwort zu geben.

Insbesondere die letzte Bedingung hat die methodologische Forschung dazu gebracht, den Befragungsprozess theoretisch zu betrachten und Bedingungen zu untersuchen, bei denen die Aufrichtigkeit der befragten Person anzuzweifeln ist und die Antworten als systematische Antwortverzerrungen einzuordnen sind (Esser 1975, Schuman/Presser 1981, Reinecke 1991). Zu den Einflussfaktoren auf derartige Verzerrungen zählen die sichtbaren und nichtsichtbaren Merkmale des Interviewers (z.B. Geschlecht, Alter, Interviewerfahrung), Einflüsse in der Interviewsituation (z.B. Anwesenheit dritter Personen), bestimmte Frageformulierungen und Fragemerkmale (z.B. sensitive persönliche Fragen) sowie die Merkmale des Befragten selber (z.B. Meinungslosigkeit, Tendenz zu sozial erwünschtem Antwortverhalten, inhaltsabhängige Befürwortungstendenzen). Verschiedene Techniken können eingesetzt werden, um das Ausmaß systematischer Antwortverzerrungen in der Befragungssituation gering zu halten. Hierzu zählen sowohl die Trainings der Interviewer (Glantz/Michael, Kapitel 21 in diesem Band) zu möglichst neutralem, einheitlichem Verhalten gegenüber den Befragten als auch die Wahl der richtigen Befragungsform. Auf Grund der Sensitivität bestimmter Befragungsthemen (Hlawatsch/Krickl, Kapitel 20 in diesem Band) sind keine validen Informationen durch anwesende Interviewer zu erwarten, so dass eine Erhebungssituation ohne direkte Kommunikation mit dem Interviewer gewählt werden sollte.

Zur Überprüfung von systematischen Verzerrungen auf Grund der genannten Einflussfaktoren bietet es sich an, die Frageformulierungen mit unterschiedlichen Erhebungsmethoden zu untersuchen und zu überprüfen, ob der Zusammenhang zwischen der zu messenden Dimension (z. B. Fremdenfeindlichkeit) und den operationalisierten Items (z. B. Skala zur Einstellung gegenüber Ausländern) durch die Erhebungsmethode beeinflusst wird. Ein hoher Zusammenhang zwischen Dimension und Messung weist auf eine hohe Konstruktvalidität hin. Wenn mindestens zwei Dimensionen und zwei Erhebungsverfahren eingesetzt werden, kann das nach Campbell/Fiske (1959) benannte

*Multitrait-Multimethod*-Design zur Überprüfung der Konstruktvalidität verwendet werden. Die Autoren differenzieren für die Prüfung nach *Konvergenz* und *Diskriminanz* der Messungen. Konvergenz bedeutet, dass die Messresultate bezogen auf die jeweiligen Dimensionen unabhängig von den eingesetzten Erhebungsmethoden sind. Diskriminanz bedeutet, dass die Messungen inhaltlich verschiedener Dimensionen sich nicht überschneiden und der Zusammenhang zwischen diesen Dimensionen unabhängig von den Erhebungsmethoden ist. Die Prüfung nach Konvergenz und Diskriminanz erfolgte bei Campbell/Fiske (1959) durch Vergleich der Korrelationskoeffizienten in der *Multitrait-Multimethod*-Matrix. Mit konfirmatorischen Faktorenmodellen (Latcheva/Davidov, Kühnel/Dingelstedt, Kapitel 55 und 80 in diesem Band) lässt sich auf Basis dieser Matrix zusätzlich testen, ob konvergente und diskriminante Messungen vorliegen und damit von Konstruktvalidität unabhängig von der Erhebungsmethode gesprochen werden kann. Scherpenzeel/Saris (1997) untersuchten 12 unterschiedliche Studien, in denen *Multitrait-Multimethod*-Experimente enthalten sind. Neben verschiedenen inhaltlichen Dimensionen (z.B. Lebenszufriedenheit) und deren Messungen wurden in den Studien telefonische, schriftliche sowie computergestützte Erhebungstechniken eingesetzt. Jedes dieser Experimente enthält zwischen drei und vier Skalen, die jeweils mit drei oder vier Erhebungsmethoden erfasst wurden. Anhand von 57 *Multitrait-Multimethod*-Matrizen und den darauf basierenden konfirmatorischen Faktorenmodellen können die Autoren sehr detailliert zeigen, bei welcher Erhebungsmethode die jeweiligen Messungen zu zuverlässigen und gültigen Ergebnissen führen (Scherpenzeel/Saris 1997).

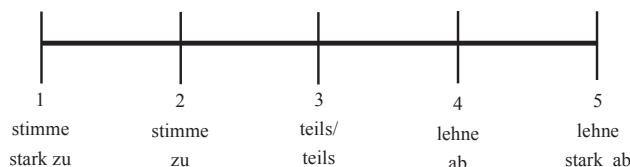
---

### 44.3 Fragevarianten

Standardisierte Befragungen weisen einen hohen Grad an Strukturierung bezogen auf das eingesetzte Messinstrument (Fragebogen) auf. Der überwiegende Teil der Fragen wird mit vorgegebenen Antwortkategorien (Franzen, Kapitel 51 in diesem Band) in einer festgelegten Reihenfolge (Klöckner/Friedrichs, Kapitel 49 in diesem Band) gestellt. Diese sogenannten *geschlossenen Fragen* (Porst, Kapitel 50 in diesem Band) setzen voraus, dass das Spektrum der Antwortalternativen bekannt ist. Davon zu unterscheiden sind *Hybridfragen*, die ebenfalls feste Antwortvorgaben enthalten, zusätzlich aber die Möglichkeit bieten, nicht vorgesehene Antworten zu geben (beispielsweise unter der Rubrik „Sonsstiges“). *Offene Fragen* (Züll/Menold, Kapitel 52 in diesem Band) enthalten dagegen keine Antwortvorgaben, wobei dieses Format in standardisierten Befragungen eher vermieden wird. Sowohl Hybridfragen als auch offene Fragen setzen voraus, dass das Spektrum der Antwortalternativen nicht vollständig bekannt ist (Faulbaum et al. 2009: 20). Eine nach Informationswunsch differenzierte Einteilung der Fragen unterscheidet nach fünf Typen (Diekmann 2007: 471):

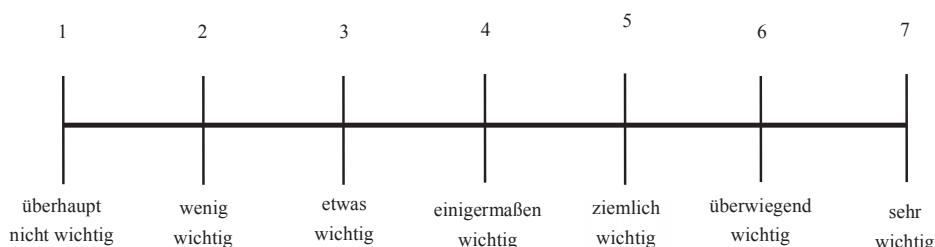
1. Einstellungen
2. Fakten und Wissen
3. Ereignisse, Verhaltensintentionen, Verhalten

### 5-stufige Antwortskala



**Abb. 44.1** 5-stufige Antwortskala

### 7-stufige Antwortskala



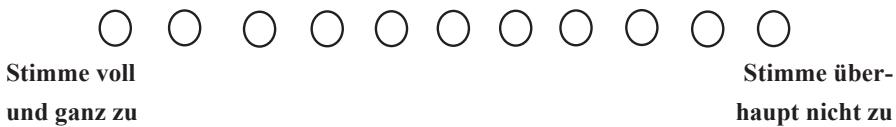
**Abb. 44.2** 7-stufige Antwortskala

4. sozialstatistische Merkmale und
5. Netzwerkfragen

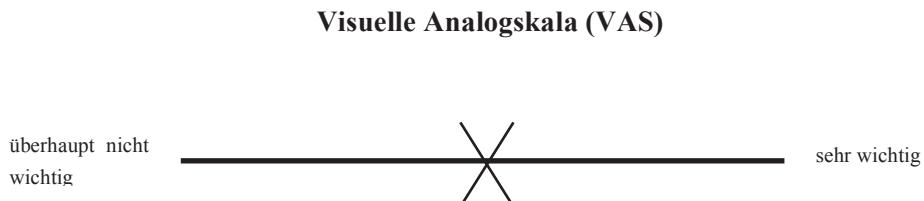
#### 44.3.1 Einstellungen

Einstellungen werden häufig durch eine Reihe von Aussagen (auch *Items* genannt) erfasst. Die befragte Person wird aufgefordert, die vorgelegten Items auf einer Ratingskala (Franzen, Kapitel 51 in diesem Band) nach unterschiedlichen Stufen zu bewerten. Fünf- oder siebenstufige Skalierungen (z.B. mit den Endpunkten „stimme stark zu“ und „lehne stark ab“) werden in größeren Studien wie die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS, [www.gesis.org/allbus](http://www.gesis.org/allbus)) vielfach eingesetzt (Abb. 44.1 und 44.2). Mit diesem auf Likert (1932) zurückgehenden Skalierungsverfahren der summierten Ratings (Blasius, Kapitel 83 in diesem Band) wird versucht, bestimmte Einstellungsdimensionen reliabel und valide (Krebs/Menold, Kapitel 30 in diesem Band) zu messen.

Bei elektronischen Fragebögen (Wagner/Hering, Kapitel 48 in diesem Band) besteht entweder die Möglichkeit, die Skalenstufen durch sogenannte Radioknöpfe (Abb. 44.3) abzubilden oder visuelle Analogskalen einzusetzen (Abb. 44.4), die kontinuierliche Abstufungen durch Markierungen auf einer Linie oder durch einen Schieberegler erlauben (Faulbaum et al. 2009: 27).



**Abb. 44.3** Darstellung der Abstufungen durch Radioknöpfe



**Abb. 44.4** Visuelle Analogskala

Für die Auswertung können die Items mittels multivariater statistischer Verfahren (z.B. Faktorenanalysen) auf ihre Messgenauigkeit geprüft werden (Blasius/Baur, Kapitel 79 in diesem Band). Eine Reihe von Einstellungsskalen ist mit ihren Reliabilitäts- und Validitätsmaßen im sogenannten ZIS-Skalenhandbuch enthalten, welches über die Webseiten der Gesellschaft sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS, [www.gesis.org](http://www.gesis.org)) kostenlos erhältlich ist (ZIS 2012).

#### 44.3.2 Fakten und Wissen

Fakten und Wissen können über zwei oder mehrere Antwortkategorien erfasst werden. Wenn beispielsweise nach dem Familienstand gefragt wird, sind mehrere Antwortkategorien vorgesehen, von denen eine vom Befragten ausgewählt wird (z. B. verheiratet). Wenn beispielsweise nach Geschichtskenntnissen gefragt wird, dann können mehrere Lösungen als Antwortkategorien formuliert werden, von denen nur eine richtig ist.

#### 44.3.3 Ereignisse, Verhaltensintentionen, Verhalten

Ereignisse und Verhalten werden über Fragen gemessen, die sich auf die Vergangenheit beziehen und in der Regel auf einen bestimmten Zeitraum eingegrenzt werden. Beispielsweise unterscheidet man in epidemiologischen und kriminologischen Forschungskontexten nach der *Prävalenz* und der *Inzidenz* des Verhaltens. Die *Prävalenz* erfasst die Ausführung des Verhaltens über dichotome Antwortkategorien (ja/nein), während über die sich anschließende Frage nach der *Inzidenz* die Häufigkeit des gezeigten Verhaltens in einem

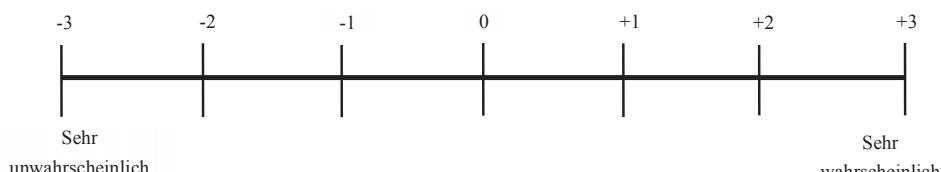
vorgegebenen Zeitraum ermittelt wird. In kriminologischen Studien wird beispielsweise die Frage gestellt, ob in den letzten 12 Monaten ein Ladendiebstahl begangen wurde. Die Häufigkeit der Kategorie „ja“ ist die Prävalenzrate des Ladendiebstahls. Für alle Personen, die diese Frage bejaht haben, kann anschließend die Frage gestellt werden, wie oft sie einen Ladendiebstahl in den letzten 12 Monaten begangen haben. Aus der Antwortverteilung kann die Inzidenzrate für den Ladendiebstahl ermittelt werden. Demgegenüber sind Verhaltensabsichten in die Zukunft gerichtet, wobei diese auch als Wahrscheinlichkeitsabfragen mit einer bipolaren Skalierung versehen werden können, die hier von „sehr unwahrscheinlich“ bis „sehr wahrscheinlich“ reicht (Abbildung 44.5).

Ist die Korrespondenz zwischen Einstellungen, Verhaltensintentionen und Verhalten zentraler Untersuchungsgegenstand, dann ist der mehrfache Einsatz der Messinstrumente in Wiederholungsbefragungen (Paneldesign; Schupp, Kapitel 73 in diesem Band) notwendig. Die Operationalisierungen und Skalen (Latcheva/Davidov, Kapitel 55 in diesem Band) der Theorie des geplanten Verhaltens (Ajzen 1991) können hier als geeignete Instrumentarien genannt werden. Empirische Beispiele zeigen Bamberg et al. (2000) sowie Wittenberg/Reinecke (2003).

#### 44.3.4 Sozialstatistische Merkmale

Die Erhebung sozialstatistischer Merkmale (Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band) erfordert eine sorgfältige Differenzierung der Fragen insbesondere nach Geschlecht, Alter, Bildung, Familienstand, Beruf, Einkommen und Haushalt. Einzelne Fragen können sich nicht nur auf die befragte Person, sondern auch auf (Ehe)Partner und Eltern beziehen. Ein sorgfältiger Aufbau der einzelnen Fragen und eine entsprechende Filterführung (Klöckner/Friedrichs, Kapitel 49 in diesem Band) sind hier notwendig. Mit Filterführung ist eine Verzweigung im Fragebogen gemeint, die mit entsprechender Kennzeichnung für den Interviewer oder den Befragten dazu führt, dass die nachfolgenden Fragen übersprungen und nicht beantwortet werden. Es ist empfehlenswert, die sorgfältig getesteten und in anderen Studien bewährten Fragen aus der ZUMA-Standarddemographie, die auch im ALLBUS ([www.gesis.org/allbus](http://www.gesis.org/allbus)) eingesetzt wird, für den eigenen Fragebogen zu verwenden.

**Bipolare Skala**



**Abb. 44.5** Bipolare Skala

#### **44.3.5 Netzwerkfragen**

Die standardisierte Befragung kann als das wichtigste Erhebungsverfahren für die Gewinnung von Informationen der Befragten zu ihren Beziehungsnetzen angesehen werden. Sogenannte egozentrierte Netzwerkinformationen haben in der sozialwissenschaftlichen Umfrageforschung über die kontinuierliche Implementation der Instrumente im Fragebogen des ALLBUS eine herausragende Bedeutung. Das egozentrierte Netzwerk besteht beispielsweise im ALLBUS aus drei, nicht im gleichen Haushalt wohnenden Personen des Verwandten- und Bekanntenkreises, mit denen der Befragte nach eigener Einschätzung am häufigsten privat zusammen ist. Die abgefragten Informationen können sich sowohl auf sozio-ökonomische Merkmale als auch auf subjektiv wahrgenommene Einstellungsmerkmale des Netzwerks beziehen (Baur, Kapitel 74 in diesem Band).

---

#### **44.4 Der Gesamtfragebogen**

Unter Beachtung von Qualitätsstandards (siehe weiter unten) kann mit der Zusammenstellung der einzelnen Fragebogenteile ein erfolgreich einzusetzender Gesamtfragebogen entstehen. Der Fragebogen darf den zu befragenden Personenkreis einerseits nicht überfordern und muss bei schriftlichen Befragungen oder online-gestützten Erhebungen selbsterklärend sein. Andererseits müssen bei mündlichen oder telefonischen Befragungen für den Interviewer ausreichend verständliche Anweisungen im Fragebogen enthalten sein, die den Befragungsprozess unterstützen und den Anforderungen an die Befragten gerecht werden sollen.

---

#### **44.5 Feldorganisation**

Standardisierte Befragungsinstrumente werden in repräsentativen Umfragen eingesetzt. In den meisten Fällen wird die Feldorganisation von privaten Markt- und Meinungsforschungsinstituten, die über entsprechende technische Einrichtungen (z.B. Telefonlabors) und geschulte Interviewer verfügen, übernommen. Die Entwickler des standardisierten Fragebogens bestehen meistens aus einem Forscherteam, welches das Umfrageinstitut zur Durchführung der Befragung beauftragt. In Abhängigkeit vom Aufwand gehört zu dem Auftrag auch die technische Umsetzung des Fragebogens (z.B. bei Online-Befragungen) oder auch die Dateneingabe. Einschließlich der Schulung der Interviewer werden alle wesentlichen Arbeitsschritte der Feldorganisation zwischen Forscherteam und Umfrageinstitut vertraglich festgehalten.

## 44.6 Der Befragungsmodus

Der Befragungsmodus kann nach drei verschiedenen Dimensionen differenziert werden: nach der Form der Administration, nach der Kommunikationsform und nach der Befragungstechnologie (Faulbaum et al. 2009: 30).

### 44.6.1 Administrationsform

Eine standardisierte Befragung kann entweder mit oder ohne Interviewer administriert werden. Bei Einsatz von Interviewern und Computerunterstützung wird nach *Computer Assisted Personal Interviewing* (CAPI) (Stocké, Kapitel 45 in diesem Band) und *Computer Assisted Telephone Interviewing* (CATI) (Hüfken, Kapitel 46 in diesem Band) unterschieden. Wird ohne Interviewer, aber mit Computerunterstützung gearbeitet, dann ist nach folgenden Verfahren zu differenzieren: *Audio Computer-Assisted Self-Interviewing* (ACASI), *Telephone Audio Computer-Assisted Self-Interviewing* (T-ACASI), *Computer Assisted Self-Administered Interviewing* (CASI) und *Computerized Self-Administered Questionnaire* (CSAQ) (Wagner/Hering, Kapitel 48 in diesem Band). Wird ein Fragebogen mit Interviewer ohne Computerunterstützung eingesetzt, dann ist die Bezeichnung *Paper and Pencil* (PAPI oder P&P) geläufig. Die gleiche Bezeichnung existiert für den bei der schriftlichen Befragung versandten Fragebogen in Papierform (Reuband, Kapitel 47 in diesem Band).

### 44.6.2 Kommunikationsform

In der Markt- und Meinungsforschung (Ziegler, Kapitel 10 in diesem Band) hat sich in den letzten Jahren ein deutlicher Trend durchgesetzt, mehr telefonische (Hüfken, Kapitel 46 in diesem Band) und mehr internetgestützte Befragungen (Wagner/Hering, Kapitel 48 in diesem Band) durchzuführen. Dafür sind die Verwendung persönlich-mündlicher Interviews (Stocké, Kapitel 45 in diesem Band) und der Einsatz schriftlicher Befragungen (Reuband, Kapitel 47 in diesem Band) deutlich zurückgegangen. Nach den Übersichten der Arbeitsgemeinschaft deutscher Marktforschungsinstitute (ADM, [www.adm-ev.de](http://www.adm-ev.de)) lag der Anteil der Studien mit persönlich-mündlichen Interviews im Jahre 1997 noch bei 44%, im Jahre 2012 aber nur noch bei 21%. Dagegen stieg der Anteil der Studien mit telefonischen oder internetgestützten Befragungen im gleichen Zeitraum von 40% auf 73%. Durch den Einsatz von Telefon und Internet lässt sich nicht nur eine günstigere Kostengestaltung erreichen, auch die schnelle Verfügbarkeit der Ergebnisse ist hauptsächlich für diese Entwicklung verantwortlich.

#### 44.6.3 Befragungstechnologie

Die Einrichtung von Telefonlaboren und die Technik *computerunterstützter Telefoninterviews (CATI)* ermöglichen eine effiziente Abwicklung von Umfragen bei gleichzeitiger Steigerung der Erhebungsqualität. Die Antworten der Befragten werden direkt in den Computer eingegeben und können anschließend zentral verarbeitet werden. Die Abwicklung der Feldphase kann vollständig durch ein CATI-System übernommen werden, welches die Kontakt- und Terminverwaltung sowie die Auswahl der zu befragenden Personen automatisch steuert. Die Stichprobenziehung (Häder/Häder, Kapitel 18 in diesem Band) erfolgt über den Ansatz von Gabler/Häder (1997), wobei sich dieser auf die zufällige Auswahl von Festnetztelefonnummern beschränkt. Vor der Liberalisierung des Telefonmarktes war diese Stichprobenziehung für repräsentative Untersuchungen ausreichend. Auf Grund der gestiegenen (und immer noch steigenden Anzahl) exklusiver Nutzer des Mobilfunks können nur noch kombinierte Stichprobenansätze aus Festnetz- und Mobilfunknummern (*Dual-Frame*) eine repräsentative Telefonbefragung mit standardisierten Fragebögen gewährleisten (Hunsicker/Schroth 2007; Häder/Häder 2009).

Die umfangreichste Entwicklung in den letzten 15 Jahren kann für die *internetgestützte Befragung* (Wagner/Hering, Kapitel 48 in diesem Band) mit dem Einsatz standardisierter Fragebögen verzeichnet werden. Erst mit dem nahezu flächendeckenden Angebot der Versorgung von Haushalten mit Zugang zum Internet ist ein breiter Einsatz von netzgestützten Befragungen sinnvoll. Der Online-Fragebogen ist eine Variante des schriftlichen Fragebogens, wobei dessen Bearbeitung ausschließlich über ein technisches Medium (z.B. ein Notebook) erfolgt (CSAQ). Je nach Auswahl der Untersuchungsobjekte werden netzgestützte Befragungen auf Basis von Zufallsstichproben oder selbstselektierten Stichproben durchgeführt (Diekmann 2007: 524f.).

Bei netzgestützten Befragungen basierend auf Zufallsstichproben besteht die Möglichkeit, repräsentative Informationen und Aussagen treffen zu können. Hierzu gehören *listenbasierte Stichproben aus Populationen mit hoher Erreichbarkeit* (z.B. Studierende einer Hochschule), sogenannte *Mixed-Mode-Befragungen* oder vorrekrutierte Panels aus der Bevölkerung.

Bei den *Mixed-Mode-Befragungen* erfolgt die Stichprobenziehung außerhalb des Internets, während die Befragungen selbst online durchgeführt werden. Die ausgewählten Personen können vielfach auch die Erhebungsform wählen, so dass die Informationen aus dem Fragebogen auf wahlweise internetgestützten, schriftlichen, telefonischen *oder* mündlichen Befragungen basieren können.

Bei *vorrekrutierten Panels* werden die Befragungsteilnehmer durch eine Zufallsstichprobe aus der Bevölkerung ausgewählt und mit Geräten für die netzgestützte Befragung ausgestattet (Engel et al. 2012). Eine regelmäßige Teilnahme an unterschiedlichen Studien wird erwartet.

Bei netzgestützten Befragungen basierend auf *selbstselektierten Stichproben* besteht kein Repräsentativitätsanspruch. Hierzu gehören internetbasierte Befragungen zu Unterhaltungszwecken, Befragungen mit uneingeschränkt selbstrekrutierten Teilnehmern und

auch vorrekrutierte Panels. Bei letzteren werden die Teilnehmer ohne Zufallsauswahl auch mit der Erwartung gewonnen, sich regelmäßig an internetgestützten Befragungen zu beteiligen.

Sehr einfach konstruierte Fragebögen sind zu Beginn des Internetzeitalters vielfach in die elektronische Post (Email) eingebaut oder als Anhang versendet worden. Die befragten Personen sollten den Fragebogen ausfüllen und diesen elektronisch zurücksenden. Mittlerweile werden elektronische Fragebögen mit entsprechend technischem Aufwand konstruiert und können über eine Webadresse aufgerufen werden. Die befragten Personen bekommen per E-Mail die Webadresse und Hintergrundinformationen zu der jeweiligen Studie mitgeteilt. Da es bei den zu befragenden Personen viele Unterschiede bezogen auf deren technische Möglichkeiten (Browsersversionen, Programmversionen, Übertragungsgeschwindigkeiten) gibt, ist entsprechender Aufwand notwendig, damit Präsentationsunterschiede des Fragebogens ausgeschlossen werden können. Ansonsten lassen sich vermeintlich inhaltliche Antwortdifferenzen auch als Präsentationsunterschiede erklären. Eine einfach verständliche Nutzerführung während der Beantwortung des Fragebogens ist von besonderer Bedeutung, da von sehr heterogenen Erfahrungen und Fähigkeiten der Befragten auszugehen ist. Der Aufwand steigt mit dem Alter der Personen und mit der Heterogenität der Stichprobe. Intensive Vortests (Weichbold, Kapitel 19 in diesem Band) unter Berücksichtigung unterschiedlicher technischer Rahmenbedingungen sind daher unbedingt notwendig (Dillman 2000).

Gegenüber der Markt- und Meinungsforschung ist bei der wissenschaftlich orientierten Sozialforschung die Entwicklung zu mehr telefonischen oder internetgestützten Befragungen zwar auch zu erkennen, der persönlich anwesende Interviewer wird bei umfangreicher und komplexer gestalteten Fragebögen aber unbedingt benötigt. Große und längerfristig laufende Studien wie die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS, [www.gesis.org/allbus](http://www.gesis.org/allbus)), das Sozio-ökonomische Panel (SOEP, [www.diw.de/soep](http://www.diw.de/soep)) oder der European Social Survey (ESS, [www.europeansocialsurvey.org](http://www.europeansocialsurvey.org)) basieren auf standardisierten *persönlich-mündlichen Befragungen (Face-to-Face-Befragungen)* unter Einsatz umfangreicher Interviewerstäbe. In der Regel handelt es sich um stark strukturierte Interviewsituationen mit gleichen Fragen für alle Befragten mit einer durch entsprechende Interviewertrainings (Glantz/Michael, Kapitel 21 in diesem Band) geprägten neutralen Befragungssituation. Die Papierform des Fragebogens wird heute weitgehend durch den Einsatz von tragbaren Computern (CAPI) oder PDAs (*Personal Digital Assistants*) ersetzt.

Die Verwendung standardisierter Fragebögen für *schriftliche Befragungen* (Reuband, Kapitel 47 in diesem Band), die als repräsentative Befragungen meist postalisch durchgeführt werden, ist gegenüber der mündlichen, telefonischen und mittlerweile auch der internetgestützten Befragung deutlich weniger verbreitet. Sie weist aber einige, nicht zu unterschätzende, Vorteile auf. Die zu befragenden Personen können die Fragen besser durchdenken und haben insgesamt mehr Zeit den Fragebogen auszufüllen. Merkmale und Eigenschaften von Interviewern können das Befragungsergebnis nicht beeinflussen. Wenn es um sensible Themen (z.B. delinquentes Verhalten, Sexualverhalten) in der Befragung geht und damit auch Interviewereffekte zu erwarten sind, sollte ein Fragebogen zum

Selbstausfüllen konzipiert werden. Da bei der schriftlichen Befragung keine Person für Nachfragen oder andere Hilfestellungen zur Verfügung steht, muss der Fragebogen nicht nur einfach gestaltet sondern auch selbsterklärend sein. Komplexe Filterfragen und Verzweigungen im Fragebogen sollten daher vermieden werden (Klöckner/Friedrichs, Kapitel 49 in diesem Band).

Technische Neuerungen sind auch bei der schriftlichen Befragung zu verzeichnen: Wenn Interviewer einen Computer zu den Befragten bringen, diese in die Bedienung des Gerätes einführen und anschließend die Befragten den Fragebogen am Computer selbst ausfüllen lassen, dann werden diese Erhebungen als CASI-Sitzungen bezeichnet. Varianten von CASI sind ACASI und T-ACASI. Bei ACASI werden die Fragen am Bildschirm akustisch unterstützt, bei T-ACASI werden die Fragen in akustischer Form über das Telefon gestellt, die anschließend vom Befragten am Computer beantwortet werden (Faulbaum et al. 2009: 34).

Werden Jugendliche und junge Erwachsene (Heinen/König, Kapitel 58 in diesem Band) in größerem Umfang befragt oder, wie bei der Internationalen Grundschulleseuntersuchung (IGLU, [www.ifs-dortmund.de](http://www.ifs-dortmund.de)), Leistungstests unterzogen, dann sind *Klassenzimmerbefragungen* üblich. Diese Befragungsart kann als eine Variante der schriftlichen Befragung eingestuft werden, da anwesende Personen, die für die örtliche Durchführung der Erhebung verantwortlich sind, in der Regel den zu befragenden Personen Hilfestellung geben können, aber keine vollständige mündliche Präsentation der einzelnen Fragebogen-teile vornehmen. Für die Komplexität des Fragebogens gelten die gleichen Regeln wie bei der postalischen Befragung.

---

## 44.7 Die Qualität von standardisierten Befragungen

### 44.7.1 Klassische Gütekriterien

Die Qualität von Fragebögen hängt entscheidend davon ab, ob die befragte Person in der Lage ist, die ihr oder ihm gestellten Fragen zu beantworten. Dies hängt zunächst davon ab, ob der Fragebogen die klassischen Gütekriterien erfüllt: Durch die Verwendung standardisierter Fragen wird ein möglichst hoher Grad an Objektivität als Voraussetzung für die Reliabilität und Validität der Messungen bzw. der eingesetzten Skalen angestrebt (Krebs/Menold, Kapitel 30 in diesem Band).

Unter *Objektivität* wird die Unabhängigkeit der Fragebogenergebnisse von den Interviewern verstanden. Eine hohe Objektivität liegt beispielsweise dann vor, wenn zwei Interviewer mit gleichem Fragebogen bei dem gleichen Befragten zu fast übereinstimmenden Ergebnissen kommen würden (Diekmann 2007: 249).

Mit der *Reliabilität (Zuverlässigkeit)* eines Messinstrumentes wird die Reproduzierbarkeit der Messergebnisse erfasst. Mit einem Reliabilitätskoeffizienten kann die Zuverlässigkeit auch quantitativ ermittelt werden. Hierzu existieren verschiedene Verfahren, wobei die Methode der Testhalbierung am häufigsten verwendet wird. Besteht eine Skala

beispielsweise aus 20 Items, dann können zunächst Teilsummenscores von je 10 Items berechnet werden. Aus dem durchschnittlichen Zusammenhang zwischen den unterschiedlichen Skalenhälften kann ein Reliabilitätskoeffizient (z.B. Cronbachs-Alpha) für die Skala berechnet werden. Empirische Werte größer 0.8 werden als akzeptabel bezeichnet (Schnell et al. 2011: 144f.).

Unter *Validität (Gültigkeit)* wird das Ausmaß der Genauigkeit, mit dem das Messinstrument eine Einstellung oder eine Verhaltensweise misst, verstanden. Drei verschiedene Arten der Validitätsbestimmung werden unterschieden:

- *Inhaltliche Validität* liegt dann vor, wenn eine Auswahl von Fragen oder Items tatsächlich die zu messende Eigenschaft in hohem Grad repräsentiert.
- Die *Kriteriumsvalidität* gibt den Grad des Zusammenhangs zwischen der Skala und anderen, nicht in der Skala erfassten Einstellungen oder Verhaltensmerkmalen an. Je stärker dieser Zusammenhang ist, desto besser ist die Kriteriumsvalidität der Skala.
- Mit der *Konstruktvalidität* wird die Korrespondenz zwischen den einzelnen Items der Skala und der zu messenden Dimension (z.B. Einstellungen gegenüber Migranten) geprüft. Diese Korrespondenz kann beispielsweise mit faktorenanalytischen Modellen quantifiziert werden und erlaubt damit eine statistische Prüfung der Konstruktvalidität.

Zwischen Objektivität, Reliabilität und Validität besteht ein hierarchisches Verhältnis: Objektivität ist eine notwendige, aber nicht hinreichende Bedingung für Reliabilität, während Reliabilität eine notwendige, aber nicht hinreichende Bedingung für Validität ist. Objektive und reliable Messinstrumente in den Fragebögen sind Minimalvoraussetzungen zur Konstruktion valider Messinstrumente. Durch einzelne Voruntersuchungen (Pretests von Fragebögen) können inhalts valide, kriteriums valide oder konstrukt valide Messungen erreicht werden.

#### 44.7.2 Totaler Umfragefehler (Total Survey Error)

Auch wenn die klassischen Gütekriterien erfüllt werden, können bei Befragten mit heterogenen sozio-ökonomischen Merkmalen entsprechend unterschiedliche kognitive und sprachliche Fähigkeiten ausgebildet sein, die bei gleichen Fragebogeninstrumenten zu unterschiedlich hohen *Ausfällen (Nonresponse)* (Engel/Schmidt, Kapitel 23 in diesem Band) und zu unterschiedlichen *Antwortqualitäten* (Krebs/Menold, Kapitel 30 in diesem Band) führen. Ein mögliches Missverhältnis zwischen den Anforderungen durch die gestellten Fragen und den Fähigkeiten, diese zu beantworten, lässt die Wahrscheinlichkeit ansteigen, dass Befragte zu bestimmten Antwortstrategien neigen. Hierzu gehören Tendenzen, bestimmte Antwortkategorien (Extremkategorien, Mittelkategorien) zu bevorzugen, die zuerst oder zuletzt genannte Antwortalternative zu wählen (*Primacy* oder *Recency Effect*), die inhaltsunabhängige Zustimmungstendenz (Franzen, Kapitel 51 in diesem Band) und das sozial erwünschte Antwortverhalten (Hlawatsch/Krickl, Kapitel 20 in diesem Band). Die Qualität einer standardisierten Befragung richtet sich danach, inwieweit die idealen

Erhebungsbedingungen realisiert werden können. Die Differenz zwischen den idealen und den tatsächlichen Bedingungen einer Befragung kann über den *totalen Umfragefehler* (*Total Survey Error, TSE*) (Faulbaum, Kapitel 31 in diesem Band) quantifiziert werden. Dieser lässt sich unterscheiden nach dem *Stichprobenfehler* (*Sampling Error*) (Häder/ Häder, Kapitel 18 in diesem Band) und den Fehlern, die nicht auf die Stichprobeneziehung zurückzuführen sind (*Non-Sampling Errors*). Hierzu zählen neben Interviewereffekten und Befragteneffekten auch Fälschungen (Blasius, Kapitel 22 in diesem Band) sowie Fehlerquellen, die bei der Konstruktion von standardisierten Fragebögen auftreten können.

#### **44.7.3 Nonresponse**

Eine bedeutende und die Qualität von standardisierten Befragungen beeinflussende Fehlerquelle ist die Nichtantwort (*Nonresponse*) (Engel/Schmidt, Kapitel 23 in diesem Band). In einmaligen oder mehrfach durchgeführten Querschnittsuntersuchungen (Mochmann, Kapitel 14 in diesem Band) können durch Nichtbeantwortung einzelner Fragen oder Items sogenannte *Item Nonresponses* auftreten. Je nach Fragebogenkonstruktion können auch im Fragebogen vorgesehene Kategorien wie „Weiß nicht“ oder „Keine Meinung“ dazu zählen. Quantitativ lassen sich die Verzerrungswirkungen von *Item Nonresponse* durch Vergleiche der Antwortenden und der Nichtantwortenden bestimmen (Groves/ Cooper 1998). Verfahren zur mehrfachen Ersetzung fehlender Werte (*Multiple Imputation*) können dann bei der Datenanalyse eingesetzt werden (Cielebak/Rässler, Kapitel 26 in diesem Band), wenn sich die *Nonresponses* zufällig im Datensatz verteilen und komplexere multivariate Analysetechniken (Blasius, Baur, Kapitel 79 in diesem Band) möglichst vollständige Datenstrukturen erfordern (Little/Rubin 2002; van Buuren 2012). Wenn im Rahmen von Paneluntersuchungen (Schupp, Kapitel 73 in diesem Band) die gleichen Fragen oder Items den Befragten über einen längeren Zeitraum wiederholt vorgelegt werden, dann können neben der Nichtbeantwortung einzelner Fragebogenteile auch temporäre Ausfälle der Befragten zu einer spezifischen Erhebung bzw. Panelwelle auftreten. Diese komplette Nichtbeantwortung von Fragebögen wird als *Unit Nonresponse* bezeichnet.

#### **44.7.4 Voruntersuchung (Pretest)**

Vor dem eigentlichen Einsatz des standardisierten Fragebogens lassen sich durch Pretestverfahren (Weichbold, Kapitel 19 in diesem Band) Mängel von Fragen und Fragebögen reduzieren. Ein wesentliches Pretestverfahren sind *kognitive Interviews* (Willis 2005), die in Laborsituationen die kognitiven Prozesse bei der Beantwortung von Fragen untersuchen sollen. Kognitive Interviews können evaluieren, wie Personen Fragen oder Items verstehen und interpretieren, welche Informationen und Ereignisse Personen aus dem Gedächtnis abrufen, welche Entscheidungen dem Antwortverhalten zugrunde liegen und wie Personen ihre Antwort den zur Verfügung stehenden Antwortkategorien zuordnen (Faulbaum

et al. 2009: 97). Zu den wichtigsten Verfahren kognitiver Interviews gehören die Nachfrage (*Probing*), das Paraphrasieren (*Paraphrasing*), die Bewertung der Antwortverlässlichkeit (*Confidence Rating*), das Sortieren (*Sorting*) und das laute Denken (*Think-aloud*).

Bei den Nachfragetechniken werden Begrifflichkeiten, Fragen und Antworten vom Interviewer durch Zusatzfragen hinterfragt, um Informationen über das Verständnis der Frage zu erhalten. Mit dem Paraphrasieren wiederholen die Befragten nach der Beantwortung den Fragetext mit eigenen Worten. Über die eigene Verbalisierung können Rückschlüsse auf das Frageverständnis gezogen werden. Zur Bewertung der Antwortverlässlichkeit werden die Befragten aufgefordert, hierzu eine Einschätzung abzugeben. Bei den Sortiertechniken sollen vorgegebene Begrifflichkeiten auf Gemeinsamkeiten geprüft werden. Hierzu werden den Befragten Aussagen (Items) vorgegeben, die hinsichtlich interessanter Begriffe gruppiert werden sollen. Bei der Technik des lauten Denkens werden die Befragten aufgefordert, ihre Gedankengänge bezogen auf die Antworten zu den gestellten Fragen zu formulieren, um Informationen über den Prozess zur Antwort zu erhalten (Faulbaum et al. 2009: 98).

Kognitive Interviews werden in einem frühen Entwicklungsstadium des Fragebogens eingesetzt. Die in der Hauptuntersuchung angestrebte Befragungsform ist hierbei von untergeordneter Bedeutung.

Ein weiteres Pretestverfahren stellt der sogenannte *Feldpretest* dar, der als eine reduzierte Testerhebung unter den realistischen Bedingungen der geplanten Haupterhebung charakterisiert werden kann. Der Befragungsmodus sollte dem in der Haupterhebung entsprechen. Mit dem Feldpretest wird angestrebt, Probleme des Frageverständnisses und Probleme bezogen auf den technischen Ablauf der Befragung zu ermitteln (Faulbaum et al. 2009: 99). Während kognitive Interviews sich in den ersten Planungsphasen der standardisierten Befragung anbieten, können Feldpretests den Einsatz und die Handhabung der finalen Version des Fragebogens durch Experten und potentielle zu befragende Personen überprüfen.

Mit ursprünglich in den USA entwickelten *Fragebewertungssystemen* (*FBS*) können Qualitätseinschränkungen des Fragebogens hinsichtlich des Fragetextes und der Antwortformate identifiziert werden. Das System enthält eine sogenannte Checkliste, die eine systematische Prüfung auf das Verstehen des Fragetextes, auf die damit verbundenen Anforderungen an die zu befragende Person sowie auf mögliche Verzerrungseffekte erlauben. Explizit werden im FBS auch Anweisungen an den Interviewer oder – bei schriftlichen Befragungen – Anweisungen an den Befragten überprüft. Die ausführliche Fassung eines FBS wird in Faulbaum et al. (2009: 111f.) vorgestellt.

## Literatur

- Ajzen, Icek (1991): The Theory of Planned Behavior. In: Organisational Behavior and Human Decision Processes 50: 179-211
- Bamberg, Sebastian/Gumbl, Harald/Schmidt, Peter (2000): Rational choice und theoriegeleitete Evaluationsforschung am Beispiel der Verhaltenswirksamkeit verkehrspolitischer Maßnahmen. Opladen: Leske und Budrich
- Campbell, Donald T./Fiske, Donald W. (1959): Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. In: Psychological Bulletin 56: 81-105
- Diekmann, Andreas (2007): Empirische Sozialforschung. Reinbek bei Hamburg: Rowohlt
- Dillman, Don A. (2000): Mail and Telephone Surveys. The Tailored Design Method. New York: Wiley
- Engel, Uwe/Bartsch, Simone/Schnabel, Christiane/Vehre, Helen (2012): Wissenschaftliche Umfragen. Methoden und Fehlerquellen. Frankfurt: Campus
- Esser, Hartmut (1975): Soziale Regelmäßigkeiten des Befragtenverhaltens. Meisenheim: Hain
- Faulbaum, Frank/Prüfer, Peter/Rexroth, Magrit (2009): Was ist eine gute Frage? Die systematische Evaluation der Fragequalität. Wiesbaden: VS Verlag
- Gabler, Siegfried/Häder, Sabine (1997): Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland. In: ZUMA-Nachrichten 41: 7-18
- Groves, Robert M./Cooper, Mick P. (1998): Nonresponse in Household Interview Surveys. New York: Wiley
- Häder, Michael/Häder, Sabine (2009): Telefonbefragungen über das Mobilfunknetz: Konzept, Design und Umsetzung einer Strategie zur Datenerhebung. Wiesbaden: VS Verlag
- Hunsicker, Stefan/Schroth, Yvonne (2007): Die Kombination von Mobilfunk- und Festnetztichproben: Eine praktische Anwendung des Dual-Frame-Ansatzes. In: Methoden – Daten – Analysen 1: 161-182
- König, Rene (Hrsg.) (1973): Handbuch der empirischen Sozialforschung, Band 2: Grundlegende Methoden und Techniken. München: dtv
- Likert, Rensis (1932): A Technique for the Measurement of Attitudes. In: Archives of Psychology 140: 1-55
- Little, Roderick/Rubin, Donald (2002): Statistical Analysis with Missing Data. New York: Wiley
- Mansel, Jürgen/Raithel, Jürgen (Hg.) (2003): Kriminalität und Gewalt im Jugendalter. Weinheim: Juventa
- Reinecke, Jost (1991): Interviewer- und Befragtenverhalten. Theoretische Ansätze und methodische Konzepte. Wiesbaden: Westdeutscher Verlag
- Scherpenzeel, Anette/Saris, Willem E. (1997): The Validity and Reliability of Survey Questions. A Meta-Analysis on MTMM Studies. In: Sociological Methods and Research 25: 341-383

- Scheuch, Erwin K. (1973): Das Interview in der Sozialforschung. In: König (Hg.): 66-190
- Schnell, Rainer/Hill Paul B./Esser Elke (2011): Methoden der empirischen Sozialforschung.  
9. Auflage. München: Oldenbourg
- Scholl, Armin (2003): Die Befragung. Konstanz: UVK
- Schuman, Howard/Presser, Stanley (1981): Questions and Answers in Attitude Surveys:  
Experiments on Question Form, Wording, and Context. New York: Sage
- van Buuren, Stef (2012): Flexible Imputation of Missing Data. Boca Raton: CRC Press
- Willis, Gordon B. (2005): Cognitive Interviewing: A Tool for Improving Questionnaire  
Design. London: Sage
- Wittenberg, Jochen/Reinecke, Jost (2003): Diebstahlskriminalität von Jugendlichen. In:  
Mansel/Raithel (Hg.): 207-225
- Zusammenstellung sozialwissenschaftlicher Items und Skalen, ZIS Version 15.00 2012.  
(ISSN 1867-5972, ISBN 978-3-86819-018-2)13

Volker Stocké

---

## 45.1 Einleitung

In der Bundesrepublik wurden 2012 21% aller standardisierten Befragungen (ADM 2013, <https://www.adm-ev.de/zahlen/#c245>, Zugriff vom 20. Oktober 2013) in den Sozialwissenschaften und der Marktforschung mithilfe eines persönlich-mündlichen Befragungsmodus durchgeführt. Obwohl die Verwendung dieser Befragungsart in den vergangenen Jahren rückläufig ist, stellt sie somit immer noch eine der wichtigsten Arten der Erfassung von Befragungsdaten dar (Reinecke, Kapitel 44 in diesem Band). Der vorliegende Beitrag beschreibt die Besonderheiten dieses Datenerhebungsmodus und diskutiert die hieraus erwachsenden Vor- und Nachteile hinsichtlich der resultierenden Datenqualität, d.h. er befasst sich konkret mit den Befragungsmodi PAPI („Paper and Pencil Interviewing“, auch: interviewer-administrierte Befragung) und CAPI („Computer-Assisted Personal Interviewing“, auch: computergestützte Interviewer-Administration), die nachfolgend auch als „face-to-face“ (FTF)-Befragungen bezeichnet werden (für eine ausführliche Darstellung der FTF-Methode: Biemer/Lyberg 2003: 149ff., Groves et al. 2004: 134ff., Lyberg/Kasprzyk 1991, Schnell 2012: 187ff., Tourangeau et al. 2000: 289ff.). Qualitative Befragungen und spezielle Befragungsformen wie etwa Delphi-Befragungen (Häder/Häder, Kapitel 42 in diesem Band) werden in diversen anderen Beiträgen dieses Bandes behandelt.

Bei den einer FTF-Befragung zugrunde liegenden Stichproben handelt es sich entweder um Personen- oder Haushaltsstichproben (Häder/Häder, Kapitel 18 in diesem Band). Den Zielpersonen bzw. Zielhaushalten wird vor dem Erstkontakt durch den Interviewer häufig ein Ankündigungsschreiben mit dem Ziel der Verbesserung der Kooperationsbereitschaft durch die Zielpersonen zugestellt (vgl. auch Abschnitt 3). Dieses Anschreiben stellt die Inhalte der Befragung, deren Legitimität sowie die Identität des Sponsors und ggf. die des die Umfrage durchführenden Instituts vor. Der Erstkontakt erfolgt normalerweise im Rahmen eines Besuchs durch die Interviewer, kann aber bei Verfügbarkeit von Telefonnummern auch telefonisch erfolgen. Ob dann der Vorteil einer Aufwandreduktion

und Erhöhung der Kontaktrate die Reduktion der Kooperationsneigung am Telefon aufwiegt, bedarf der Abwägung im Einzelfall. Erfolgt die Stichprobenziehung auf der Haushaltsebene, während die Grundgesamtheit auf der Personenebenen definiert ist, so muss der Interviewer nach Einwilligung der Kontakterson in Mehrpersonenhaushalten zuerst eine Zielperson für die Befragung auswählen und deren Teilnahmebereitschaft sicherstellen (Hlawatsch/Krickl, Glantz/Michael, Kapitel 20 und 21 in diesem Band). In einer prototypischen FTF-Befragung lesen die Interviewer die Fragen vom Papierfragebogen oder Computerbildschirm ab und zeichnen die Antworten der Befragten auf. Auch wenn Interviewer die Anwesenheit Dritter im Interview vermeiden sollen, stellen sie selbst Interaktionspartner und Adressaten des Antwortverhaltens dar (vgl. Abschnitt 4.2).

---

## **45.2 Fehlerquellen in persönlich-mündlichen Befragungen**

Die in Befragungsdaten insgesamt vorliegende Fehlervarianz lässt sich nach dem „Total Survey Error“-Ansatz (Faulbaum, Kapitel 31 in diesem Band) grob in stichprobenbezogene („Sampling Error“) und durch den Messprozess („Non-Sampling Error“) erzeugte Varianzkomponenten zerlegen. Hinsichtlich beider Fehlerquellen weisen FTF-Befragungen Besonderheiten gegenüber anderen Befragungsformen auf. Bezuglich des stichprobenbezogenen Messfehlers spielen insbesondere die Realisierbarkeit einer Zufallsstichprobe (Häder/Häder, Kapitel 18 in diesem Band), das Ausmaß deren Klumpung sowie die Erreichbarkeit und Kooperationsbereitschaft der Stichprobenmitglieder eine besondere Rolle (Engel/Schmidt, Kapitel 23 in diesem Band). Merkmale von FTF-Befragungen wirken sich auch auf die Qualität von Angaben im Interview aus. In dieser Hinsicht wirken sich unterschiedliche Aspekte der FTF-Befragung in besonderem Maße auf den Grad der Elaboriertheit des Antwortgenerierungsprozesses („Survey Satisficing“, Krosnick 1991) und die Neigung zu Antwortverzerrungen durch soziale Erwünschtheit aus (Hlawatsch/Krickl, Kapitel 20 in diesem Band).

---

## **45.3 Stichproben und stichprobenbezogene Fehler in persönlich-mündlichen Befragungen**

Stichprobenbezogene Fehler können auf eine mangelnde Abdeckung der Grundgesamtheit durch den Stichprobenplan („Coverage Error“), auf den mangelnden Einbezug relevanter Befragter in die Stichprobe („Sampling Error“) und auf die mangelnde Fähigkeit zur Interviewrealisierung mit den Stichprobenmitgliedern („Nonresponse Error“) zurückgeführt werden. Die unvollständige Beantwortung von Fragen im Interview („Item-Nonresponse“) kann als Spezialfall des „Nonresponse Errors“ aufgefasst werden (Engel/Schmidt, Kapitel 23 in diesem Band). Die Stichprobenkonstruktion und der stichprobenbezogene Fehler unterscheiden sich zwischen den Erhebungsmodi. Stichproben für FTF-Befragungen können auf einer Zufallsauswahl aus Einwohnermeldeamtsregistern beruhen, werden aber

vornehmlich im Rahmen von Zufallsbegehungens („Random Walk“) generiert (Häder/Häder, Kapitel 18 in diesem Band).

Wegen der ansonsten sehr hohen Kosten, werden in FTF-Umfragen überwiegend regional geklumpte Stichproben verwendet, wobei sich die Mitglieder innerhalb der Kluppen überdurchschnittlich ähnlich sind (hohe Intra-Klassenkorrelation). Der statistische Informationsgehalt von Befragungsdaten (effektive Stichprobengröße) ist in geklumpten Stichproben geringer als in einfachen Zufallsstichproben, muss in Auswertungen durch Design-Gewichte angepasst werden (Kiesl, Kapitel 24 in diesem Band), so dass generell eine umfangreichere Nettostichprobe erforderlich ist. Aus diesem Grund ist in FTF-Befragungen eine höhere Fallzahl der Netto-Stichproben erforderlich, als dies bei einfachen Zufallsstichproben der Fall ist.

Sowohl durch Zufallsbegehungens generierte als auch Registerstichproben haben den Vorteil, dass Adressinformationen der Zielhaushalte oder sogar der Zielpersonen der Befragung verfügbar sind (zur datenschutzrechtlichen Problematik, Schnell 2012: 194ff., Mühlichen, Kapitel 4 in diesem Band). Dies erlaubt vor dem eigentlichen Erstkontakt eine postalische Zustellung von Informationsmaterial und Teilnahmeanreizen. Beide Maßnahmen, insbesondere aber vorab überreichte Teilnahmeanreize unabhängig von einer Befragungsteilnahme, haben sich als für die Kooperationsbereitschaft potentieller Befragter förderlich erwiesen. Für die Entscheidung zu einer Befragungsteilnahme spielen die Interviewer (Glantz/Michael, Kapitel 21 in diesem Band) eine zentrale Rolle. Einerseits erwerben sie in der Interviewerschulung geeignete Strategien, um Befragte zu einer Interviewteilnahme zu bewegen. Andererseits sind bestimmte Teilnahmemotive von Befragten, beispielsweise das Streben nach Abwechslung oder nach einer angenehmen Interaktion, explizit auf die Person des Interviewers gerichtet. Hinzu kommt, dass FTF-Umfragen im Vergleich zu alternativen Befragungsarten ein höheres Maß an Legitimität und Bedeutsamkeit zugeschrieben wird: Interviewer können in der Kontakt situation auf ein zuvor zugestelltes Anschreiben verweisen und sind typischerweise mit einem „offiziellen“ Ausweis und einem seriösen Namensschild ausgestattet. Außerdem signalisiert der Aufwand eines persönlichen Besuchs durch die Person des Interviewers, dass es sich bei der Interviewanfrage um ein wichtiges Anliegen handelt. Im Durchschnitt führen FTF-Befragungen aus den angeführten Gründen zu höheren Kooperationsraten, wodurch der potentiell vorliegende „Non Response Error“ (Engel/Schmidt, Kapitel 23 in diesem Band) reduziert wird.

---

#### **45.4 Kognitive und motivationale Determinanten der Datenqualität in FTF-Befragungen**

Eine Beurteilung der im Rahmen von FTF-Interviews resultierenden Qualität von Umfraggedaten ist auf der Grundlage des *Grundmodells der Fragebeantwortung* möglich (Tourangeau et al. 2000: 7ff.). Nach diesem konzeptionellen Modell erfordert die Beantwortung von Fragen in Interviews die Durchführung von vier Schritten durch die Befragten. In einem

ersten Schritt müssen Befragte den *intendierten Frageinhalt verstehen*, zweitens *relevante Informationen aus dem Gedächtnis abrufen*, auf deren Grundlage sie sich drittens *ein auf den Fragegegenstand bezogenes Urteil bilden* und dann im vierten Schritt dieses *Urteil*, nach einer etwaigen Editierung, vermittelt durch die Wahl einer Antwortoption *kommunizieren*. Die Qualität von Befragungsdaten nimmt zu, je vollständiger und kognitiv elaborierter die ersten drei Schritte durchgeführt werden, und je weniger die Befragten im vierten Schritt ihre Antwort vor dem Hintergrund strategischer Überlegungen editieren. Unterschiedliche Aspekte von FTF-Befragungen wirken sich auf die Art und Weise aus, mit der Befragte die vier zur Antwortgenerierung notwendigen Schritte durchführen.

#### 45.4.1 Survey Satisficing

Im Rahmen der *Theorie des Survey Satisficing* wird angenommen, dass Befragte wegen einer zu hohen Komplexität der Befragungsinhalte („task complexity“), wegen unzureichender kognitiver Fähigkeiten („cognitive ability“) und mangelnder Motivation zur Fragebeantwortung („motivation“) einen oder mehrere der ersten drei Schritte des Grundmodells der Fragebeantwortung wenig intensiv durchlaufen („weak satisficing“) oder ganz auslassen („strong satisficing“) (Krosnick 1991). Aspekte der FTF-Befragung wirken sich vermittelt über diese drei Determinanten des Satisficing auf die Elaboriertheit des Antwortgenerierungsprozesses und dadurch auf die Qualität von Befragungsdaten aus. Im Vergleich zu telefonischen Befragungen (Hüfken, Kapitel 46 in diesem Band) wird in FTF-Befragungen die Datenqualität aus den folgenden Gründen weniger durch Satisficing beeinträchtigt.

##### 45.4.1.1 Komplexität von Befragungsinstrumenten

Die kognitiven Anforderungen bei der Beantwortung der Fragen und der Handhabung des Fragebogens generell wirken sich auf die durch die Befragten erlebte Komplexität von Befragungsinstrumenten aus.

In FTF-Befragungen können erstens die Befragten bei der Beantwortung der verbal gestellten Fragen durch *visuelle Hilfsmittel* unterstützt werden. Hierbei handelt es sich beispielsweise um die Präsentation langer Antwortlisten oder differenzierter Antwortskalen mithilfe von Antwortkärtchen. Diese Hilfsmittel reduzieren die Anforderungen an das Arbeitsgedächtnis der Befragten, die damit verbundene Komplexität der Fragen und die Neigung zum Satisficing.

Zweitens stellt die Handhabung von Fragebögen mit *komplexer Filterführung* hohe Ansprüche an die Befragten, lenkt diese von Befragungsinhalten ab und birgt das Risiko einer fälschlichen Auslassung von Fragen. In FTF-Befragungen sind es geschulte und mit dem Fragebogen vertraute Interviewer, welche die Filteranweisungen anwenden müssen (Tobias/Glantz, Kapitel 21 in diesem Band). Bei allen Formen der computergestützten Befragungsdurchführung werden auch die Interviewer durch die in der Befragungssoftware programmierte Filterführung von dieser Aufgabe entlastet.

Die Interviewer spielen drittens eine bedeutende Rolle bei der Aufklärung von *Verständnisproblemen* auf Seiten der Befragten und unterstützen diese bei der Generierung einer auf die Antwortvorgaben passenden Antwort („*Probing*“). Die durch die aktive Rolle der Interviewer in FTF-Befragungen erreichte Reduktion fehlender Werte („*Item Nonresponse*“) geht mit einer stärkeren Beeinflussung der Antwortinhalte und die damit auf die Interviewer zurückführbare Fehlervarianz einher (Engel/Schmidt, Kapitel 23 in diesem Band).

#### 45.4.1.2 Den Befragten verfügbare kognitive Kapazitäten

Merkmale von FTF-Befragungen wirken sich auch auf die effektiv den Befragten verfügbaren kognitiven Kapazitäten aus. Bedeutsam ist in dieser Hinsicht erstens der bei der Beantwortung der Fragen *subjektiv gefühlte Zeitdruck*, der abgesehen von anderen Faktoren von der Kommunikationssituation und dem Medium der Kommunikation abhängt. Bei Interviewer-Administration werden längere Gesprächspausen als unangenehm erlebt und durch relativ schnelle und daher weniger überlegte Antworten vermieden. Dies ist in besonderem Maße bei Verwendung des Kommunikationsmediums Telefon der Fall (Hüfken, Kapitel 46 in diesem Band). In selbst-administrierten Befragungen herrscht dagegen wegen des Fehlens eines direkten Adressaten der Antworten ein deutlich geringerer Zeitdruck. Es kann somit bei FTF-Befragungen, relativ zu selbst-administrierten Befragungen und telefonbasierter Administration, von einem mittleren Maß an zeitdruckbedingtem Satisficing ausgegangen werden.

Zweitens, abgesehen vom sehr selten eingesetzten „*Audio Computer-Assisted Self-Interviewing*“ (ACASI), bei dem die Fragen von einem Computer vorgelesen werden, erfordert die Selbst-Administration von Fragebögen ein Minimum an Leseverständnis bei den Befragten. In der erwachsenen Wohnbevölkerung Deutschlands sind rund 7.5 Millionen Personen von funktionalem *Analphabetismus* betroffen (Schütte/Schmies, Kapitel 61 in diesem Band; zum Problem von Beschränkungen der Literalität bei Migranten, vgl. Abschnitt 6). Das Leseverständnis dieser Gruppe ist soweit eingeschränkt, dass sie bei selbst-administrierten Interviews faktisch als befragungsunfähig gelten müssen. Sofern es keine Sprachbarrieren gibt (El Menouar, Kapitel 60 in diesem Band) erlaubt die FTF-Befragung auch mit Mitgliedern dieser Teilstichprobe die Durchführung von Interviews und vermeidet eine entsprechende Selektivität der Nettostichprobe.

Die Verwendung von FTF-Befragungen bewirkt bei den Befragten eine relativ hohe Motivation zu einer möglichst angemessenen Beantwortung der Fragen im Interview (Lyberg/Kasprzyk 1991). Die Interviewer spielen eine wesentliche Rolle bei der Motivation der Befragten zu einer vollständigen und möglichst kognitiv elaborierten Abarbeitung der ersten drei Schritte des Antwortgenerierungsprozesses (zu deren Bedeutung beim vierten Schritt der Editierung, vgl. Abschnitt 4.2).

Zusammenfassend ist die in FTF-Befragungen realisierte Datenqualität generell weniger durch Satisficing beeinträchtigt als es bei telefonischen Befragungen der Fall ist. Im Vergleich zu selbst-administrierten Befragungen (Reuband, Wagner/Hering, Kapitel 47 und 48 in diesem Band) sollte das FTF-Interview dann Vorteile bei der Vermeidung von

Satisficing bieten, wenn das Befragungsinstrument oder der Befragungsgegenstand zunehmend komplexer ist, bzw. die kognitiven Fähigkeiten der Befragten oder deren Motivation zur Beantwortung der Fragen im Durchschnitt geringer ausfallen.

#### **45.4.2 Soziale Erwünschtheit und Interviewereffekte**

Befragte in Umfragen orientieren sich bei der Beantwortung von Fragen häufig nicht am subjektiv geglaubten „wahren Wert“ sondern an den Erwartungen, die sie in der sozialen Umwelt wahrnehmen (Hlwatasch/Krickl, Kapitel 20 in diesem Band). Unter dem Phänomen der Antwortverzerrungen durch soziale Erwünschtheit lassen sich auch bestimmte Arten der Interviewereffekte (Glantz/Michael, Kapitel 21 in diesem Band) subsumieren. Neben persönlichkeitstheoretischen Ansätzen (Crowne/Marlowe 1960), die situationale Faktoren weitgehend ausklammern, geht die Theorie rationalen Befragtenverhaltens (Esser 1991) von der Bedeutung unterschiedlicher im Rahmen von FTF-Befragungen vorliegender Faktoren aus. Demnach resultieren Antwortverzerrungen durch soziale Erwünschtheit aus den Selbstdarstellungsstrategien von Befragten mit dem Ziel der Erlangung sozialer Anerkennung durch andere und der Vermeidung formeller Sanktionen (Esser 1991). Beide Zielsetzungen wirken sich nur bei fehlender Privatheit und mangelnder Anonymität auf das Antwortverhalten aus. In der Standardvariante des FTF-Interviews werden Fragen durch die Interviewer vorgelesen und die Antworten durch diese aufgezeichnet. Die Interviewer sind somit potentiell Adressaten der Selbstdarstellungsstrategien der Befragten. Entsprechend ist empirisch gut bestätigt, dass Angaben in FTF-Befragungen, im Vergleich zu selbst-administriert erhobenen Daten, stärker durch soziale Erwünschtheit verzerrt sind (z.B. in Befragungen über Einstellungen zu Minoritäten, über Alkohol- und Drogenkonsum sowie über die Häufigkeit des Gottesdienstbesuchs und der Teilnahme an politischen Wahlen). Es lassen sich nur unsystematische Unterschiede in der Stärke der Antwortverzerrungen durch soziale Erwünschtheit zwischen FTF- und telefongestützten Befragungen sowie nach der Computerisierung der Datenerhebung feststellen.

Die Privatheit und Anonymität von Angaben lassen sich im Rahmen von FTF-Befragungen durch unterschiedliche Maßnahmen erhöhen. So können sensible Teile des Fragebogens an die befragte Person übergeben werden, die diese ohne Wahrnehmung der Antwortinhalte durch den Interviewer (gesicherte Privatheit der Antwortsituation) selbst-administriert ausgefüllt. Eine zusätzliche Maßnahme kann in der sogenannten „Secret Ballot Technique“ gesehen werden. Hierbei wird der selbst-administrierte Fragebogenteil durch den Befragten in eine Art Wahlurne geworfen, wodurch das Gefühl der Anonymität erhöht wird (keine nachträgliche Identifikation von Antwortinhalten).

Die Nennung welcher Antwort in einer konkreten Befragungssituation zu einer besonders günstigen Selbstdarstellung führt, ergibt sich aus der wahrgenommenen relativen Erwünschtheit der Antwortoptionen (Stocké 2004). Abgesehen von gesellschaftlichen Normen bilden die mit dem Frageinhalt in Bezug stehenden Merkmale und Verhaltensweisen der Interviewer die Grundlage dieser Erwünschtheitswahrnehmungen (als

Überblick zu Interviewereffekten, Biemer/Lyberg 2003: 149ff., Michael/Glantz, Kapitel 21 in diesem Band). Während diese Merkmale und Verhaltensweisen bei Selbst-Administration naturgemäß keine Rolle spielen, sind diese in FTF-Befragungen im Vergleich zu telefonischen Befragungen entweder überhaupt nur oder zumindest klarer erkennbar. So können beispielsweise US-amerikanische Befragte auch in Telefoninterviews mit überzufälliger Wahrscheinlichkeit die Hautfarbe der Interviewer an stimmlichen Merkmalen identifizieren. In FTF-Befragungen ist aber, wegen der dann sicheren Identifikation der Hautfarbe, mit einer Verstärkung der bereits in den Telefoninterviews beobachteten Effekte der ethnischen Zugehörigkeit der Interviewer zu rechnen. Ähnliche Unterschiede in der Stärke von Effekten der Interviewermerkmale treffen auch auf die beobachteten Effekte des Alters der Interviewers auf den selbst-berichteten Alkoholkonsum und auf diejenigen des Interviewergeschlechts auf die Angaben über die Geschlechterrollenorientierung zu.

In FTF-Interviews anwesende Dritte verstärken bei fehlender Kenntnis der „wahren Werte“ Verzerrungen durch soziale Erwünschtheit, während sie bei deren Kenntnis eine Abschwächung der Verzerrungen bewirken (Lyberg/Kasprzyk 1991). Auch bei der Befragung mithilfe anderer Modi können weitere Personen anwesend sein. Allerdings können diese entweder nur die Antworten, nicht aber die Fragen (Telefoninterviews) oder nur in Ausnahmefällen beide für die Sanktionierung des Antwortverhaltens notwendigen Informationen wahrnehmen (Selbst-Administration). Die nach Kenntnis der „wahren Werte“ entgegengesetzten Effekte anwesender Dritter auf Antwortverzerrungen durch soziale Erwünschtheit fallen in FTF-Befragungen stärker als in anderen Befragungsmodi aus.

Zusammenfassend muss in FTF-Interviews durchschnittlich mit stärkeren Antwortverzerrungen durch soziale Erwünschtheit generell und durch deren Spezialfall der Interviewereffekte gerechnet werden. Dies trifft auch für Einflüsse anwesender Dritter zu, wobei hier in FTF-Interviews sowohl mit stärkeren Antwortverzerrungen wie auch mit einer intensiveren Wahrheitskontrolle gerechnet werden muss.

---

## 45.5 Organisatorische Aspekte bei der Durchführung von FTF-Interviews

Aus der Logik der Durchführung von FTF-Interviews ergibt sich eine Reihe von Besonderheiten hinsichtlich der Organisation der Datenerhebungsarbeiten. Im Vergleich zu den in der Regel zentralisiert durchgeföhrten Telefoninterviews erfordert die *Kontrolle eines möglichen Fehlverhaltens der Interviewer* in FTF-Befragungen besondere Vorkehrungen. Dabei soll erstens sichergestellt werden, dass *Stichprobenpläne eingehalten* werden. Demnach dürfen ausnahmslos die im Rahmen einer Registerstichprobe oder einer Zufallsbegehung ausgewählten Zielpersonen bzw. Haushalte befragt werden. Da Interviewer in FTF-Befragungen typischerweise nach der Zahl realisierter Interviews und nicht nach dem zeitlichen Aufwand entlohnt werden, bestehen Anreize zur Befragung leicht zugänglicher Ersatzpersonen und zur teilweisen oder vollständigen Fälschung der Interviews (Blasius, Kapitel 22 in diesem Band). Besteht bei Verwendung einer Haushaltsstichprobe

die Notwendigkeit der Auswahl einer zu befragenden Zielperson, so sehen sich Interviewer dem Anreiz ausgesetzt, eine Befragung statt mit der durch einen Zufallsprozess (z.B. „Schwedenschlüssel“ oder „letzter Geburtstagsmethode“) bestimmten Person, mit einer leicht zugänglichen und befragungsbereiten Person durchzuführen. Zweitens erfolgt bei Anwendung einer zeit- und kostensparenden Standardversion der Zufallsbegehung die Stichprobengenerierung und möglichst sofortige Befragung durch den gleichen Interviewer. Auch hier können Interviewer mit weniger Aufwand Interviews realisieren, wenn sie Haushalte statt nach der Begehungsanweisung nach deren Verfügbarkeit und Kooperationsbereitschaft auswählen. Anreize zu dieser Art von Fehlverhalten werden durch Verwendung des aufwendigeren „Adress-Randoms“ beseitigt, indem die Listung und Befragung der Haushalte durch unterschiedliche Personen erfolgt. Drittens müssen die generellen Regeln standardisierter Befragungen und die für die konkrete Umfrage spezifischen Intervieweranweisungen eingehalten werden.

Eine umfassende und mit praktischen Elementen angereicherte Interviewerschulung ist eine notwendige, nicht aber hinreichende Voraussetzung für die Vermeidung von Fehlverhalten der Interviewer (Glatz/Michael, Kapitel 21 in diesem Band). Da in FTF-Befragungen eine direkte Kontrolle der Arbeit der Interviewer (im Gegensatz zu zentralisiert durchgeführten Telefoninterviews) während der Durchführung praktisch nicht möglich ist, müssen stichprobenweise Nachkontrollen zur Auswahl und Interviewdurchführung direkt bei den Befragten erfolgen. Diese Kontrollen tragen dann zur Vermeidung und nicht nur zur Entdeckung von Interviewerfehlverhalten bei, wenn sie den Interviewern bereits vor Beginn der Feldarbeit angekündigt werden.

Durch die physische Präsenz der Interviewer können in FTF-Befragungen zusätzlich zu Befragungsdaten noch weitere Daten gesammelt werden. Dabei handelt es sich um unterschiedliche Arten von Testdaten, physiologische und medizinische oder sogar genetische Informationen. Außerdem lassen sich die Lebensbedingungen der Befragten durch Einschätzungen der Interviewer über die Nachbarschaft und die Wohnbedingungen der Befragten abbilden. Bedingt durch den hohen Aufwand bei der Feldorganisation handelt es sich bei FTF-Befragungen insgesamt um einen kostspieligen und durch eine lange Feldzeit gekennzeichneten Datenerhebungsmodus.

---

## **45.6 FTF-Befragungen in kulturell heterogenen und interkulturell vergleichenden Studien**

Die Befragung sprachlich und kulturell heterogener Stichproben im nationalen Kontext und in der interkulturell vergleichenden Sozialforschung stellt eine Herausforderung der empirischen Sozialforschung dar (Braun, Kapitel 56 in diesem Band). So treten bei der Befragung von Zuwanderern der ersten Generation häufig Verständnisprobleme wegen mangelnder Sprachbeherrschung sowohl bei der Rekrutierung als auch der Interviewdurchführung auf. Entsprechend liegen die Ausschöpfungsquoten in dieser Subpopulation oft unter der Quote von Personen ohne Migrationshintergrund. Der Einsatz

bilingual-koethnischer Interviewer, die Verwendung von in die Herkunftssprache der Befragten übersetzten Anschreiben und Fragebögen hat sich als positiv für die Teilnahmebereitschaft türkisch-stämmiger Befragter in Deutschland erwiesen. Für die Feldorganisation von FTF-Befragungen muss insbesondere die Bereitstellung von auf die Befragten passenden bilingualen Interviewern als große Herausforderung betrachtet werden. Dies setzt die Verfügbarkeit einer hinreichenden Zahl entsprechend sprachlich kompetenter Personen und das Wissen über die Notwendigkeit bestimmter sprachlicher Fähigkeiten möglichst schon vor dem Erstkontakt mit dem Zielhaushalt voraus. Dies trifft auch bei multilingualen Nationen, wie etwa der Schweiz, Kanada oder Simbabwe zu. In letzterem Land müssen Umfrageforscher mit 20 unterschiedlichen Sprachen und einer Vielzahl an Dialekten umgehen.

Eine zunehmende Anzahl an Umfragen, wie beispielsweise das „European Social Survey“ (ESS), das „International Social Survey Program“ (ISSP), das „World Value Survey“ (WVS) oder das Afrobarometer zielen auf die vergleichende Beschreibung und Erklärung sozialer Phänomene in einer teilweise sehr großen Anzahl von Ländern und kulturell heterogenen Räumen (vgl. hierzu die Beiträge in Harkness et al. 2003, Braun, Kapitel 56 in diesem Band). Hierbei können leicht Unterschiede zwischen Nationen und kulturellen Räumen diagnostiziert werden, die in Wirklichkeit ein Methodenartefakt sind. So spielen selbstverständlich die Verwendung inhaltlich vollständig vergleichbarer Fragebogenübersetzungen, aber auch die Vermeidung von Unterschieden durch die Verwendung variierender Befragungsmodi eine zentrale Rolle. Da meist in zumindest einigen der an den Umfrageprojekten teilnehmenden Nationen Telefonanschlüsse nicht flächendeckend verfügbar sind oder das Literalitätsniveau der Bevölkerung unzureichend ist, können weder telefonische noch selbst-administrierte Befragungen in allen teilnehmenden Ländern durchgeführt werden. Wird wie beim ESS die Verwendung des gleichen Befragungsmodus in allen Länderkontexten angestrebt, muss mehrheitlich die FTF-Befragung als Methode der Wahl gelten. Die hohen Kosten von FTF-Befragungen und die in vielen Nationen bestehenden finanziellen Restriktionen führen allerdings oft zur Verwendung multipler Befragungsmodi („Mixed Mode Surveys“). So wird beispielsweise das ISSP in den derzeit 48 teilnehmenden Ländern teilweise im FTF-Modus und teilweise selbst-administriert durchgeführt. In den am Afrobarometer teilnehmenden Nationen muss die Wahl zwischen den Administrationsmodi auch von der Verfügbarkeit von Transportmitteln, der Sicherheit der Interviewer hinsichtlich Kriminalitätsrisiken und saisonalen Wetterbedingungen abhängig gemacht werden.

Abgesehen von finanziellen Restriktionen und Problemen rein sprachlicher Art muss in FTF-Interviews auch beachtet werden, dass Menschen aus unterschiedlichen Kulturreisen über alternative normative Vorstellungen über eine angemessene sprachliche Interaktion verfügen können. So unterscheiden sich beispielsweise die deutsche und andere Kulturen darin, wie direkt Gesprächspartner das eigentliche Ziel der Kommunikation angehen sollten, ohne als unhöflich zu erscheinen. Ähnliches trifft auch auf normative Regeln darüber zu, welche Kombination von Geschlecht, Alter und sozialem Status von Interviewer und befragter Person als angemessen angesehen wird. Eine Missachtung von

Unterschieden in diesen Kommunikations- und Interaktionsnormen kann leicht zu einer Reduktion der Kooperationsbereitschaft von Befragten führen (vgl. hierzu die Beiträge in Harkness et al. 2010).

Es liegen Hinweise darauf vor, dass sich die mangelnde Privatheit und Anonymität von FTF-Befragungen je nach kulturellem Hintergrund der Befragten unterschiedlich in Antwortverzerrungen durch soziale Erwünschtheit manifestieren (vgl. die Beiträge in Abschnitt IV von Harkness et al. 2010). Demnach lassen sich individualistische sowie fernöstlich-kollektivistische (z.B. China, Japan) und an Ehre orientierte kollektivistische Kulturen (z.B. Türkei, Italien, Spanien) unterscheiden. Diese Kulturen unterscheiden sich darin, wie zentral Menschen eine positive Selbstdarstellung ansehen. In der fernöstlichen Kultur wird Bescheidenheit, Mäßigung und Harmonie als normativ geboten angesehen, so dass Selbstdarstellungsstrategien im Interview als im Vergleich zu individualistischen Kulturen weniger zur Produktion von sozialer Anerkennung geeignet angesehen werden. Im Gegensatz hierzu werden in der zweiten Gruppe der kollektivistischen Kulturen die Ehre und das Ansehen der Zielperson sowie die von nahe stehenden Personen als zentral erachtet. Demnach sind hier im Vergleich zu individualistischen Kulturen stärkere Tendenzen zur Selbstdarstellung und damit stärker ausgeprägte Verzerrungen durch soziale Erwünschtheit zu erwarten. Empirische Untersuchungen belegen scheinbar entsprechende interkulturelle Unterschiede in der Stärke von Antwortverzerrungen durch soziale Erwünschtheit.

## Literatur

- Biemer, Paul P./Lyberg, Lars E.(2003): Introduction to survey quality. New York: Wiley
- Biemer, Paul P./Groves, Robert M./Lyberg, Lars E./Mathiowetz, Nancy A./Sudman, Seymour (Hg.) (1991): Measurement errors in surveys. New York: Wiley
- Crowne, Douglas P./Marlowe, David: A new scale of social desirability independent of psychopathology. In: Journal of Consulting Psychology 24: 349-354
- Esser, Hartmut (1991): Die Erklärung systematischer Fehler in Interviews. Befragtenverhalten als „rational choice“. In: Wittenberg (Hg.): 59-78
- Groves, Robert M./Fowler, Floyd J. Jr./Couper, Mick P./Lepkowski, James M./Singer, Eleanor/Tourangeau, Roger (2004): Survey methodology. Hoboken New Jersey.: Wiley
- Harkness, Janet A./van de Vijver, Fons J. R./Mohler, Peter P. (Hrsg.) (2003): Cross-cultural survey methods. New York: Wiley
- Harkness, Janet A./Braun, Michael/Edwards, Brad/Johnson, Timothy P. (2010): Survey methods in multicultural, multinational, and multiregional contexts. New York: Wiley
- Krosnick, Jon A. (1991): Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied Cognitive Psychology 5: 213-236
- Lyberg, Lars/Kasprzyk, Daniel (1991): Data collection methods and measurement error: An Overview. In: Biemer et al. (Hg.): 237-257
- Schnell, Rainer (2012): Survey-Interviews. Methoden standardisierter Befragungen. Wiesbaden: VS Verlag
- Stocké, Volker: Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. In: Zeitschrift für Soziologie 33: 303-320
- Tourangeau, Roger/Rips, Lance J./Rasinski, Kenneth (2000): The psychology of survey response. Cambridge: Cambridge University Press
- Wittenberg, Reinhard (Hg.) (1991): Person - Situation - Institution - Kultur. Günter Büschges zum 65. Geburtstag. Berlin: Duncker & Humblot

Volker Hüfken

## 46.1 Hintergrund und Stellenwert der Telefonumfrage

Telefonumfragen werden bereits seit mehr als 90 Jahren angewendet. Damals wie heute kommt den Medien in diesem Zusammenhang eine besondere Bedeutung zu. Der Wettbewerb, insbesondere um eine zeitnahe Berichterstattung gesellschaftlicher oder politischer Ereignisse, begünstigt dabei den Einsatz der Telefonumfrage. So wurden in Chicago bereits in den 1920er-Jahren die ersten Wahlumfragen mittels Telefon durchgeführt (Chicago Tribune 1923, zitiert nach Frankovic 2012). Ende der 1920er-Jahre, mit der Verbreitung der Radionutzung, gewinnt auch das Bedürfnis nach gezielteren Angaben über die Hörgewohnheiten in der Bevölkerung an Bedeutung. Nicht nur in den amerikanischen, auch in den europäischen Rundfunkanstalten wie z.B. in England, Deutschland und Österreich gab es ein entsprechend großes Interesse. Dies geht mit einem enormen Anstieg postalischer (Reuband, Kapitel 47 in diesem Band), persönlich-mündlicher (Stocké, Kapitel 45 in diesem Band) und telefonischer Umfragen einher, wobei der Einsatz von Telefonumfragen lediglich für die Vereinigten Staaten berichtet wird (Bessler 1980).

Der vielfältige Einsatz dieser Erhebungsmethoden mündete bereits in den 1930er-Jahren, insbesondere in den Vereinigten Staaten, in eine intensive Methodendiskussion. Folgende Aspekte wurden für die Telefonumfragen hervorgehoben: Telefonumfragen sind kostengünstiger gegenüber dem persönlich mündlichen Interview, die Untersuchungsergebnisse sind schneller verfügbar, über 95% der Haushalte, die über ein Telefon verfügen, haben auch ein Radio, und Telefonumfragen ermöglichen eine bessere Kontrolle über die Interviewer bzw. Standardisierung der Erhebungsbedingung. Als Nachteil wurde ein wesentlicher Punkt benannt: Einschränkungen in der Generalisierbarkeit der Untersuchungsergebnisse. So wurde vom Einsatz der Telefonumfrage in kleineren Städten oder ländlichen Gebieten auf Grund der schlechteren Telefonversorgung abgeraten. Auch wurde problematisiert, dass die soziale Zusammensetzung der Stichprobe hinsichtlich höherer Einkommensgruppen verzerrt sei (Karol 1937). Gleichwohl, eine der

bekanntesten Umfragen des vergangenen Jahrhunderts, welche die Zeitschrift „Literary Digest“ im Rahmen der Präsidentschaftswahlen im Jahr 1936 durchführte, basierte im Wesentlichen auf Adressen aus Telefonbüchern und anderen ergänzenden Verzeichnissen. Das Resultat der fehlerhaften Wahlprognose aufgrund eben dieser Stichprobenverzerrung ging als Literary Digest Desaster in die Lehrbücher ein. Spätestens seit diesem Zeitpunkt wurden Telefonumfragen, welche auf Adressen bzw. Telefonnummern aus Telefonbüchern basieren, mit dem Stigma „quick and dirty“ etikettiert. Erst mit einer nahezu lückenlosen flächendeckenden Versorgung (von über 90%) der Haushalte mit Telefon, welche in den Vereinigten Staaten in den 1980er-Jahren und in West-Deutschland in den 1990er-Jahren erreicht wurde, war ein zentrales Kriterium zur Durchführung nationaler Umfragen in der Allgemeinbevölkerung erreicht. Bis dahin dominierten die persönlich-mündlichen Umfragen.

Drei weitere Entwicklungen trugen zur Popularität der Telefonumfragen wesentlich bei. Erstens: Auf Grund systematischer Vergleichsuntersuchungen konnte gezeigt werden, dass die Ergebnisse von Telefonumfragen denen der persönlich-mündlichen Umfragen gleichwertig waren (Groves/Kahn 1979). Zweitens: Die Entwicklung und der Einsatz der Computer-Assistierten-Telefoninterviews (CATI) erbrachte eine Vielzahl von Vorteilen gegenüber der herkömmlichen „Papier & Bleistift-Methode“ (Fink 1983). Drittens: Der Einsatz und die Fortentwicklung von Auswahlverfahren, die auch solche Telefonnummern für die Stichprobenziehung berücksichtigen, die in Listen bzw. Telefonbüchern nicht erfasst sind. Diese in den USA entwickelten Zufallsauswahlverfahren, welche als Random Digit Dialing (RDD) oder Randomized Last Digit Dialing (RLD) bekannt sind, schließen auch die Anschlüsse bzw. Telefonnummern in die Stichprobe ein, die nicht öffentlich zugänglich sind (Lavrakas 1987). Beide Auswahlverfahren basieren auf Informationen über Telefonnummern: Beim RLD-Verfahren werden auf der Basis bestehender Listen eingetragene Nummern verwendet, wobei an Stelle der letzten Ziffer(n) – normalerweise wird die letzte Ziffer modifiziert – eine neu generierte Zufallszahl angefügt wird. Beim RDD-Verfahren werden nicht die Informationen aus Listen verwendet, sondern es werden Zufallszahlen generiert, die innerhalb der insgesamt technisch möglichen Nummernbereiche (Angaben dazu erhält man bei der Bundesnetzagentur) liegen. Ein zentraler Nachteil ist, dass außer einer Telefonnummer keine weiteren Angaben über den Haushalt vorliegen. Im Gegensatz dazu können beim Telefonbuch bzw. bei listenbasierten Auswahlverfahren auf der Basis ergänzender Angaben, z.B. der Adresse, Haushalte bzw. Personen vor dem telefonischen Erstkontakt angeschrieben und über die anstehende Studie und das Telefoninterview informiert werden.

Betrachtet man die Entwicklung der Telefonumfragen für Deutschland, vollzog sich diese – ähnlich wie in den USA – zunächst über die kommerziellen Markt- und Meinungsforschungsinstitute, bevor eine intensivere Auseinandersetzung und Etablierung in der akademischen Sozialforschung einsetzte. Technologische Entwicklungen und gesellschaftliche Veränderungen bildeten Anfang der 90er-Jahre für die Umfrageforschung in Deutschland neue Herausforderungen. Mit der Einführung der Mobilfunknetze D1 und D2 im Sommer 1992 und der Liberalisierung des Telekommunikationsmarktes, welche

1996 gesetzlich geregelt wurde, liefen zwei Entwicklungen parallel. Besonders deutlich wurde dies in Ost-Deutschland. Dadurch, dass die Versorgungslage in Ost-Deutschland sich kurz nach der Wende mit ca. 20% auf einem niedrigen Niveau befand, gab es in der Bevölkerung ein entsprechendes Nachholbedürfnis. Sodass bereits 1998, mit einer Quote von ca. 90%, nahezu eine Vollversorgung in den neuen Bundesländern erreicht wurde.

Wie seit den 90er-Jahren üblich, bildeten die Angaben in den Telefonbüchern die Grundlage für die Stichprobenziehung. Jedoch warf dieses Vorgehen, nicht nur auf Grund der zunehmenden Nachfrage von Mobiltelefonen, sondern auch aufgrund der in den 90er-Jahren einsetzenden rechtlichen Entwicklung (Wilms/Jochum, in: Beck'scher TKG-Kommentar, 4. Aufl. 2013, § 104 Rn. 9ff. mit zahlreichen Nachweisen und Quellen) Probleme auf. Während zumindest bis in das Jahr 1991 noch nahezu jeder Haushalt, der über einen Telefonanschluss verfügte, in das Telefonbuch eingetragen war (Grundsatz automatischer Eintragung als Zwangseintragung, es sei denn: besonders begründete Ausnahmen, Geheimnummern), erhielten die Anschlussinhaber zum 1.7.1991 erstmals das Recht, der Eintragung in das öffentliche Kundenverzeichnis (Telefonverzeichnis) gegenüber der Deutschen Bundespost Telekom ganz (bezüglich aller Angaben: Name und Anschrift) oder teilweise (z.B. nur bezogen auf die Straße und die Hausnummer) zu widersprechen. Die Kunden mussten über dieses Widerspruchsrecht informiert werden.

Auch nach der fortschreitenden Liberalisierung des Telekommunikationsmarktes und der Gründung der Deutschen Telekom AG als Nachfolgeunternehmen der Deutschen Bundespost TELEKOM hat die Reform der maßgeblichen Rechtsvorschriften im Jahr 1996 die Widerspruchslösung (sog. Opt-Out-Regelung) beibehalten und einen Anspruch auf kostenfreie Entsprechung des Widerspruchs verankert. Zusätzlich wurde für die Kunden ein Wahlrecht eingeführt, nach dem sie entscheiden konnten, ob sie im gedruckten oder im elektronischen Verzeichnis oder in keinem der beiden Verzeichnisse geführt werden. Sodann setzte zunächst eine differenzierte Regelung der Eintragung in das Telefonbuch ein, die letztlich mit der derzeit geltenden Einwilligungslösung abgeschlossen wurde: Während für Neukunden bereits auf der Grundlage des Telekommunikationsgesetzes 1996 die Widerspruchsregelung nicht mehr galt, wurde sie für die Altkunden erst im Jahr 2000 aufgegeben. Zu diesem Zeitpunkt hat der Verordnungs- und Gesetzgeber davon Abstand genommen, die Eintragung nur auf Widerspruch zu löschen und ist stattdessen dazu übergegangen, die Eintragung nur auf Antrag hin vorzunehmen (sog. Opt-in-Regelung). Seitdem können sich Personen mit ihrem Namen, ihrer Anschrift und zusätzlichen Angaben wie Beruf, Branche und Art des Anschlusses in öffentliche gedruckte oder elektronische Verzeichnisse eintragen lassen (§ 104 TKG). Für das Ausmaß der Abbildung der Telefonanschlüsse durch das Telefonbuch kann aus dieser Rechtsentwicklung vor allem die folgende Problematik geschlussfolgert werden: Während die Zwangseintragung bis in das Jahr 1991 sicherstellt, dass nahezu sämtliche Telefonanschlussinhaber durch das Telefonbuch ermittelt werden können, haben die Widerspruchslösung (opt-out-Regelung) und ihr nachfolgend die Antragslösung (opt-in-Regelung) die Sensibilität der Bevölkerung für die Preisgabe von personenbezogenen Daten (Mühlchen, Kapitel 4 in diesem Band) durch das Telefonbuch erhöht und die Wahrscheinlichkeit des Telefonbucheintrags und damit

der Vollabdeckung der Telefonanschlüsse verringert. Dies hat zur Folge, dass Personen bzw. Haushalte aus den neuen Bundesländern, mobilere, jüngere und Personen, die in Einpersonenhaushalten leben, in Telefonbüchern unterrepräsentiert sind.

Entgegen der bisherigen Praxis, Angaben aus den Telefonbüchern als Auswahlgrundlage verwenden zu können, erforderte der systematische Rückgang der Telefonbucheinträge Lösungen. In dieser Folge entstand eine intensive Diskussion über alternative Zufallsauswahlverfahren. Des Weiteren ist seit einigen Jahren ein Rückgang der Festnetznutzer zu Gunsten der mobile-only-Haushalte zu beobachten. Da die mobile-only-Haushalte sich von den Haushalten, die über ein Festnetz zu erreichen sind, nicht nur auf der Ebene soziodemographischer Merkmale (Jüngere, Mehrpersonenhaushalte, Personen mit niedrigerem Einkommen und Arbeitslose) unterscheiden, sondern auch auf der Ebene von Einstellungen und Verhaltensweisen, berichtet wird dies für die Themen Gesundheit, Mediennutzung und Politik, hat dies zur Folge, dass dem um die Mobilfunknummern erweiterten Auswahlrahmen eine besondere Bedeutung zukommt (vgl. für die USA u.a. AAPOR 2008, für Deutschland Häder/Häder 2009, ADM 2012 und ein Überblick zu Erfahrungen in einigen europäischen Ländern in Häder et al. 2012). Auf Grund nicht vorhandener Listen von Mobilfunknummern werden bei der Konzeption des Auswahlverfahrens die bereits vergebenen Nummernbereiche, welche die Bundesnetzagentur an die Netzbetreiber vergeben hat, berücksichtigt. Da der Auswahlrahmen der Mobilfunknummern mit denen der Festnetznummern kombiniert wird und somit die Auswahlwahrscheinlichkeiten aller Personen erlaubt, spricht man von „Dual Frame“. Bei der Konstruktion solcher Auswahlgrundlagen kommen unterschiedliche Schritte und Berechnungen zur Anwendung (Häder/Häder 2009, ADM 2012).

Trotz einer Vielzahl von Veränderungen, wie oben geschildert, hat die Telefonumfrage – nicht zuletzt auch wegen innovativer Lösungen – nicht an Bedeutung verloren. Sieht man sich die Umfrageaktivitäten der deutschen Markt- und Sozialforschungsinstitute genauer an, so nehmen die Telefoninterviews im Vergleich zu allen anderen Erhebungsverfahren seit 2000 eine Spitzenstellung ein. Ihr Anteil, gemessen an den Interviews, beträgt seit dem Jahr 2000 41% und liegt mit 38% im Jahr 2012 weiterhin vor den anderen Erhebungsverfahren. Nach wie vor dürfte der Erfolg der Telefonumfrage durch eine Vielzahl von Vorteilen getragen sein.

---

## 46.2 Vor- und Nachteile der telefonischen Befragung

### 46.2.1 Vorteile

Verschiedene Aspekte tragen zum Stellenwert und den Vorteilen der telefonischen Befragung bei. Je nach Relevanz werden zeitliche, ökonomische, praktische und methodische Gründe genannt:

1. *Aktualität und Verfügbarkeit der Untersuchungsergebnisse:* Im Vergleich zu persönlich-mündlichen oder postalischen Umfragen, deren Feldphase meist mehrere Wochen

umfasst, ist die Telefonumfrage meist innerhalb weniger Tage abgeschlossen. Daher eignet sich die Telefonumfrage insbesondere auch dazu, auf aktuelle gesellschaftliche Ereignisse relativ kurzfristig zu reagieren, z.B. in der Wahlforschung und der Medienbranche.

2. *Geringe Kosten:* Insbesondere im Vergleich zu persönlich-mündlichen Befragungen fällt der Unterschied relativ deutlich aus (Groves 1989). Wenngleich mit der Zunahme der mobile-only-Haushalte und deren Berücksichtigung, wie beim dual-frame-Ansatz, die Kosten insgesamt höher liegen, sollten diese dennoch deutlich niedriger als bei persönlich-mündlichen Umfragen ausfallen.
3. *Regionale Streuung:* Klumpungen durch räumlich bedingte Ähnlichkeiten der Untersuchungseinheiten (Häder/Häder, Kapitel 18 in diesem Band), wie in persönlich-mündlichen Umfragen gegeben, werden vermieden. Interviewer können per Zufall verschiedenen Ortsnetzen zugewiesen werden, eine Methode, die sich bei CATI relativ problemlos umsetzen lässt.
4. *Erreichbarkeit der Zielperson bzw. Zielhaushalte:* Auf Grund der geringen Kosten sind viele Kontaktversuche möglich.
5. *Kontrolle Feldablauf:* Der Ablauf der Datenerhebung wird über die gesamte Feldphase zentral vom Forscher koordiniert und überprüft. Dies umfasst die Stichprobenziehung, die Unterstützung bei der Kontaktaufnahme, die Auswahl der Zielperson (z.B. „last-birthday-Methode“, also der Person, die als letztes Geburtstag hatte), die Durchführung des Interviews (durch die Möglichkeit des Mithörens eines Supervisors), die Handhabung und den Ablauf des Fragebogens, sowie die Eingabe der Daten.
6. *Spezialpopulationen:* Ein mehrstufiges Screening zur Ermittlung von bestimmten Bevölkerungssegmenten, z.B. Altenpopulationen oder Haushalte mit Internetanschluss, ist relativ gut anwendbar.
7. *Experimentelle Designs:* Auswahllisten, Fragerotationen oder die Steuerung von Fragefolgen sowie das Filtern von Fragen können bei computergestützten Verfahren automatisiert erfolgen.
8. *Vorkontakt:* Kurznachrichten bzw. SMS (Short Message Service), die der Zielperson vor dem telefonischen Erstkontakt zugesendet werden, erhöhen die Kooperationswahrscheinlichkeit. In dieser Kurzmitteilung wird die Zielperson über Ziel und Zweck der Studie sowie den baldigen Kontakt durch den Interviewer informiert.

### 46.2.2 Nachteile

Folgende Nachteile stehen den oben genannten Vorteilen gegenüber. Jedoch werden in der Literatur einige Maßnahmen berichtet, mit denen man diese zumindest reduzieren kann.

1. *Stichprobe:* Personen ohne Telefon können nicht berücksichtigt werden. Dies ist ohnehin kaum erheblich, da der Anteil mit unter 1% in Deutschland gegenwärtig sehr gering ausfällt. Auch in den meisten anderen europäischen Ländern ist mit einer Versorgungsquote von ca. 95% eine Vollversorgung nahezu erreicht.

2. *Auswahlrahmen*: Die Auswahlverfahren, welche die beiden unterschiedlichen Nummernstrukturen – Festnetz und Mobilfunk – für die Bildung von Stichproben berücksichtigen, sind bereits für mehrere Länder entwickelt und etabliert. Problematischer gestaltet sich die Berücksichtigung der Haushalte, die ausschließlich via Internet bzw. Voice-over-Internet Protocol (VoIP) telefonieren. Wie diese in einen Auswahlrahmen einbezogen werden können, ist bisher nicht gelöst (BITKOM 2012). Vor dem Hintergrund der stetig zunehmenden Verbreitung zeichnet sich dringender Handlungsbedarf ab.
3. *Umfang bzw. Interviewdauer*: Im Vergleich zur persönlich-mündlichen oder postalischen Umfrage ist die Interviewdauer bzw. der Umfang der Fragen eher begrenzt.
4. *Vorlagen*: Bild- oder Textvorlagen, die meist als differenzierte Beurteilungs- und Einstellungsskalen verwendet werden, sind im Gegensatz zu persönlich-mündlichen oder selbstadministrierten Erhebungen nicht einsetzbar. Eine Einschränkung, die vermutlich mit der Verbreitung des Smartphone abnehmen wird.
5. *Geographische Stratifizierung*: Während Rufnummern im Festnetz mit einer Ortsnetzkennzahl – der Teilnehmer ist im Gebiet der Rufnummer ansässig – gekennzeichnet sind, gibt es diese Verknüpfung für IP-basierte oder Mobilfunknetze nicht. Eine geographische Stratifizierung ist daher nur mit Festnetznummern möglich. Zu beachten ist zudem, dass die Ortsnetze kaum mit den Verwaltungsgrenzen der politischen Gemeinden übereinstimmen. Hinzu kommt, dass die eindeutige Zuordnung – Ortsnetzkennzahl (Vorwahl) und Gemeinde – abhängig von der Gemeindegröße ist. Je kleiner die Gemeinde desto geringer die Chance einer eindeutigen Zuordnung.
6. *Datenqualität*: Im Vergleich zu persönlich-mündlichen Befragungen gelten Telefonumfragen hinsichtlich sozialer Erwünschtheit und der Vermeidung, wohl überlegte Angaben zu machen (*satisficing*), eher als verzerrt.
7. *Homogenisierung durch Interviewer*: Die Antworten der Befragten innerhalb eines Interviewers ähneln sich untereinander stärker, als es durch Zufall zu erwarten wäre. Daher nehmen die Interviewerbedingten Ähnlichkeiten der Untersuchungseinheiten durch eine deutlich höhere Anzahl von Interviews pro Interviewer zu. Um derartige Intervieweffekte gering zu halten, wird empfohlen, die Anzahl der Interviews pro Interviewer auf max. 15 Interviews zu begrenzen.
8. *Anrufbeantworter*: Der Anteil der nicht erreichten Haushalte bzw. Zielpersonen fällt durch den Anrufbeantworter insgesamt niedriger aus. Jedoch kann der Anrufbeantworter auch als Filter für eingehende Anrufe genutzt werden und somit Ausdruck mangelnder Kooperationsbereitschaft sein. Bei der Frage danach, ob man auf dem Anrufbeantworter eine Nachricht hinterlassen sollte, um die Kontaktwahrscheinlichkeit zu verbessern oder die Verweigerungswahrscheinlichkeit zu verringern, ist die Literaturlage eher uneinheitlich. In Deutschland scheint die Nachricht auf dem Anrufbeantworter die Interviewwahrscheinlichkeit zu erhöhen (Engel et al. 2012). Bei Mobilfunkanschlüssen zeigt sich ein weiteres Problem, das der klaren Zuordnung. So ist bisher unklar, ob eine erreichte Mailbox noch durch den Inhaber genutzt wird oder durch einen anderen Inhaber, oder es sich um eine Nummer handelt, die noch nicht

vergeben aber vom Mobilfunkanbieter bereits freigeschaltet wurde. Vor dem Hintergrund einer verlässlichen Berechnung der Ausschöpfungsquote ist diese Information jedoch unerlässlich.

## 46.3 Praxis der Telefonumfrage: Maßnahmen zur Reduzierung der Ausfälle

Wie eingangs erwähnt, galt die Telefonumfrage lange als „quick and dirty“-Erhebung. Nicht zuletzt basierte dies auch darauf, dass die persönlich-mündliche Befragung als Königsweg der Sozialforschung galt. Dies resultierte u.a. auch daher, dass in Telefonumfragen die Ausschöpfungsquote eher niedriger war als in persönlich-mündlichen Umfragen. Unabhängig von einem allgemeinen Trend, wonach die Bereitschaft zur Teilnahme an Umfragen in den letzten Jahren insgesamt eher rückläufig ist (Engel/Schmidt, Kapitel 23 in diesem Band), ist die schlechtere Erreichbarkeit für den Rückgang mitverantwortlich. Unabhängig von den unterschiedlichen Bemühungen und Strategien zur Maximierung der Teilnahmequote, wie beispielsweise durch eine erhöhte Anzahl der Kontakte, personalisierte Anschreiben, wiederholte Kontaktierung der weniger kooperationsbereiten Personen, Incentives gelten das Aufkommen der technischen Ausstattungen der Telefonanschlüsse (Anrufbeantworter bzw. Mailbox), der technischen Ausstattung der Telefonapparate (Rufnummernanzeige auf dem Display) sowie gesellschaftliche Veränderungen (z.B. der Anstieg des Telefonmarketing und die allgemeine Zunahme an Umfragen), als weitere Faktoren, welche den Rückgang der Teilnahme an Telefonumfragen mitbestimmen (Curtin et al. 2005, Engel et al. 2012).

### 46.3.1 Feldzeiten und Anzahl der Kontakte

Die Feldzeiten belaufen sich in bundesweiten Telefonumfragen meistens über mehrere Tage, wobei darauf geachtet wird, dass die Haushalte bzw. Personen an verschiedenen Tagen und zu unterschiedlichen Uhrzeiten kontaktiert werden. Am ehesten, gemessen an der Anzahl der Kontaktversuche, können Personen an Werktagen in den frühen Abendstunden – in der Zeit von 16:00 bis 20:30 Uhr – erreicht werden. Für das Wochenende werden meist der gesamte Nachmittag und Abend zur Kontaktaufnahme empfohlen.

Nach 10 Kontakten liegen für ca. 90% aller Festnetznummern die abschließend relevanten Informationen – zur Berechnung der Ausschöpfungsquote nach AAPOR-Standard – vor, als optimale Anzahl werden 12 Kontaktversuche genannt. Gehen die Bemühungen weit über das Maß von 12 Kontakten hinaus, steht der Aufwand in keinem angemessenen Verhältnis zum Ergebnis. So werden lediglich geringe positive Effekte in der soziodemographischen Zusammensetzung der Stichprobe und nur marginale Verbesserungen auf der Ebene von Einstellungen und Verhaltensweisen berichtet, die durch eine Erhöhung der Kontaktierung bewirkt wurden. Anders als beim Festnetz sind die Personen, die über

Mobilfunk angerufen werden, leichter zu erreichen. Ein Unterschied, der auch unabhängig vom Geschlecht, Alter und Schulabschluss berichtet wird. Ergänzend sei noch zu erwähnen, dass die zur Berechnung der Ausschöpfungsquote relevanten Informationen sich bei den Festnetznummern relativ klar zuordnen lassen, während u.a. die automatischen Anlagen im Mobilfunk, z.B.: „Der gewünschte Gesprächspartner ist zurzeit nicht erreichbar“, dafür nicht ausreichend sind.

### **46.3.2 Erstkontakt**

Für die (erfolgreiche) Durchführung des Interviews kommt dem Erstkontakt eine besondere Bedeutung zu. Mit dem Erstkontakt wird eine Informationsgrundlage geschaffen, welche für die Zielperson bzw. den Haushalt hilfreich für die Entscheidung zur Teilnahme an der Umfrage ist. Dies hat zur Folge, dass mit dem Erstkontakt auch darüber entschieden wird, ob man unmittelbar oder später (z.B. durch Terminvereinbarung) teilnimmt. Beim Mobilfunk kommt hinzu, dass telefonische Kontaktversuche durch automatisches Hinterlegen der Anschlussnummer als Anrufeingänge festgehalten werden, eine Funktion, die im Festnetz nur eingeschränkt vorhanden ist. Es wird vermutet, dass durch die mitgeteilte bzw. hinterlegte Telefonnummer Unsicherheit abgebaut und Vertraulichkeit erzeugt werden kann. Zudem könnte die Zielperson bzw. der Haushalt den Anrufer durch Rückruf ermitteln. Insgesamt scheint ein positiver Effekt durch die Mitteilung der Rufnummer auf die Kooperationsbereitschaft und auf die Erreichbarkeit auszugehen.

### **46.3.3 Anschreiben**

Wird die Zielperson bzw. der Haushalt vorab mit einer schriftlichen Mitteilung über die anstehende Umfrage informiert (Warm-Kontakt), unterscheidet sich diese Vorgehensweise von der Kalt-Kontaktierung dadurch, dass die Informationsvermittlung ausschließlich durch den Interviewer erfolgt. Bezogen auf den Transfer von Informationen, wie z.B. Erläuterungen über Ziel und Zweck des Anrufs bzw. der Telefonumfrage, ist der Kalt-Kontakt sowohl für die Kontaktperson als auch für den Interviewer eine größere Herausforderung. Für den Interviewer ist die Art der Kontaktierung insofern entscheidend, als dass sich die kommunikative Kompetenz nicht nur auf der sprachlichen sondern auch auf der stimmlichen Ebene beobachten lässt. Der Kalt-Kontakt erfordert eine höhere Flexibilität des Interviewers bezogen auf mögliche Rückfragen und Argumentationen seitens der Zielperson. Dies wirkt sich wiederum auf prosodische Attribute des Interviewers aus.

Welche Wirkung schriftliche Vorabinformationen bzw. deren unterschiedliche Inhalte auf die jeweiligen Beteiligten einer Telefonumfrage ausüben, ist – bis auf die höhere Teilnehmerate – immer noch relativ wenig erforscht. Der positive Effekt auf die Kooperation kann, muss jedoch nicht mit einer verbesserten Zusammensetzung der realisierten Stichprobe einhergehen. Es wird vermutet, dass diejenigen Personen überproportional in

die realisierte Stichprobe einbezogen werden, die im Telefonbuch verzeichnet sind. Wollte man schriftliche Vorabinformationen auch in Mobilfunkumfragen anwenden, so wäre dies – auf Grund ungeeigneter Adressregister – nicht möglich. Stattdessen bietet sich eine verkürzte Vorabmitteilung via short message service (SMS) an. Erste Untersuchungsergebnisse zum Einsatz von SMS in Mobilfunkbefragungen liegen vor und zeigen einen vergleichbar positiven Einfluss.

---

#### **46.4 Kontaktprotokoll**

Informationen aus den Kontaktprotokollen gehören zu den Paradata (Felderer/Birg/Kreuter, Kapitel 25 in diesem Band), die Hinweise darauf liefern, ob sich die Interviewwahrscheinlichkeit dadurch erhöht, dass bei der Terminvereinbarung bereits mit der Zielperson oder einer anderen Person aus dem Haushalt gesprochen wurde. Eine weitere, weitaus häufigere Verwendung der Kontaktdaten findet bei der Erstellung der Ausfallstatistik statt. Hierbei interessiert insbesondere die detaillierte Zuordnung zu den jeweiligen Kategorien (verweigert, nicht erreicht etc.) wie sie beispielsweise in der amerikanischen Standardklassifikation ausführlich dokumentiert ist. Auf der Grundlage der Ausfallstatistik können dann vertiefende bzw. vergleichende Ausfallanalysen durchgeführt werden, um ein systematischeres Bild über die Ausfallproblematik und deren Entwicklung zu geben.

Eine weitere Möglichkeit Kontaktdaten aus einer prozessualen Perspektive zu betrachten, ist die Interviewerproduktivität bzw. die Analyse der Interviewerperformance. Dabei werden unterschiedliche Informationen (z.B. Anzahl der Interviews und/oder Anzahl der Verweigerungen pro Stunde, Anzahl der vereinbarten Termine, Anzahl realisierter Interviews auf der Basis vorheriger Terminvereinbarungen etc.) auf der Ebene der Interviewer berücksichtigt. Eine Übersicht über „Interviewer-Performance-Indikatoren“ und deren Einsatzmöglichkeiten stellen West/Groves (2013) vor. Weichen einige Werte des Interviewers relativ deutlich vom Durchschnitt ab, so bieten diese u.a. die Grundlage für eine weitere Schulung des Interviewers.

---

#### **46.5 Interviewer**

Da der Einsatz zufallsgenerierter Telefonnummern – RDD oder RLD-Stichproben – in der Umfrageforschung üblich ist, erfolgt der Erstkontakt bei Festnetzteilnehmern nicht schriftlich, sondern mündlich durch den Interviewer. In der jeweiligen Situation entscheidet daher der Interviewer darüber, welche Vorgehensweise und (Überzeugungs-) Strategien er – aus der Schulung und seinen eigenen Erfahrungen – anwendet, wenn es während bzw. nach der Einleitung oder bei einem wiederholten Kontakt nicht zu einem Interview zu kommen scheint. Unterschiedliche Situationen und verschiedene Anforderungen erfordern seitens des Interviewers eine besondere kommunikative Kompetenz. Während es bei der Durchführung eines Interviews u.a. auf die Aussprache und Präsentation von Fragen

sowie auf die Beachtung von Wortfolgen oder Fragewiederholungen und eine neutrale aber angemessene Konversation ankommt, ist die Gesprächseröffnung eher durch ein flexibles und konziliantes, den Teilnehmer gewinnendes Auftreten gekennzeichnet.

Im Vergleich zum persönlich mündlichen Interview (hier sind auch Blickkontakte, Mimik, äußere Erscheinung und Gestik relevant), können beim telefonischen Interview nur paralinguistische Merkmale Bestandteil der Interkommunikation zur wechselseitigen Personenwahrnehmung und zum Aufbau der Rollenbeziehungen in der Interaktion sein. In diesem Zusammenhang werden folgende Eigenschaften genannt: Stimmqualität, Artikulation, Modulation sowie sprachfreie Reaktionen, z.B. „mmm-hmm“. Ob und wie sich die stimmliche Realisierung eines Erhebungsinstrumentes auf Qualität und Güte der Angaben des Befragten auswirkt, wurde bisher nicht systematisch untersucht, obwohl sich die Sozialforscher über die grundlegende Bedeutung der Phänomene als solche durchaus einig sind, wenn sie auf den Einfluss von Sprechpausen, den Klang einer Stimme, das Sprechtempo oder andere psychologische Aspekte einer Sprechhandlung hinweisen. Klarer scheint jedoch die Befundlage zum Einfluss stimmlicher Merkmale auf die Interviewwahrscheinlichkeit. Danach gelten folgende soziolinguistische und prosodische bzw. lautliche Eigenschaften als relevant: eine eher langsamere Sprechgeschwindigkeit, eine persönliche Ansprache, eine frei formulierte Gesprächseröffnung, eine eher durchschnittliche Stimmhöhe, eine als passend empfundene Betonung von Worten und Passagen, sowie eine als angenehm empfundene Stimme. Darüber hinaus sind die Erfahrung, Einstellung und Motivation für die Interviewwahrscheinlichkeit von Bedeutung. In aktuellen Arbeiten zeichnet sich zudem ab, dass, in Anlehnung an Befunde aus der persönlich mündlichen Befragung, in Telefonumfragen allenfalls vermittelnde Effekte von den Persönlichkeitseigenschaften auf die Interviewwahrscheinlichkeit ausgehen.

## Literatur

- AAPOR Cell Phone Task Force (2008): Guidelines and Considerations for Survey Researchers when Planning and Conducting RDD and Other Telephone Surveys in the U.S. with Respondents Reached via Cell Phone Numbers. Retrieved from: [http://www.aapor.org/uploads/Final\\_AAPOR\\_Cell\\_Phone\\_TF\\_report\\_041208.pdf](http://www.aapor.org/uploads/Final_AAPOR_Cell_Phone_TF_report_041208.pdf)
- ADM (2012): ADM-Forschungsprojekt „Dual-Frame-Ansätze“ 2011/2012, abrufbar unter: [https://www.adm-ev.de/index.php?eID=tx\\_nawsecuredl&u=0&file=fileadmin/user\\_upload/PDFS/ADM\\_Dual\\_Frame\\_Projekt\\_-\\_Forschungsbericht.pdf&t=1376153983&hash=e29fee026734203fbc0db3eb2a0281f9369fb692](https://www.adm-ev.de/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/user_upload/PDFS/ADM_Dual_Frame_Projekt_-_Forschungsbericht.pdf&t=1376153983&hash=e29fee026734203fbc0db3eb2a0281f9369fb692)
- Bessler, Hansjörg (1980): Hörer- und Zuschauerforschung. München: Deutscher Taschenbuch Verlag
- BITKOM (2012): Internet-Telefonie gewinnt weiter an Boden. abrufbar unter: [http://www.bitkom.org/files/documents/BITKOM\\_Presseinfo\\_Internet-Telefonie\\_08\\_01\\_2012.pdf](http://www.bitkom.org/files/documents/BITKOM_Presseinfo_Internet-Telefonie_08_01_2012.pdf)
- Curtin, Richard/Presser, Stanley/Singer, Eleanor (2005): Changes in Telephone Survey Nonresponse over the Past Quarter Century. In: Public Opinion Quarterly 69: 87-98
- Engel, Uwe/Bartsch, Simone/Schnabel, Christiane/Vehre, Helen (2012): Wissenschaftliche Umfragen. Methoden und Fehlerquellen. Frankfurt: Campus
- Fink, James C. (1983): CATI's First Decade: The Chilton Experience. In: Sociological Methods & Research 12: 153-168
- Frankovic, Kathleen A. (2012): Straw Polls in the U.S.: Measuring Public Opinion 100 Years Ago. In: Haad et al. (Hg): 66-84
- Geppert, Martin/Schütz, Raimund (2013): Beck'scher TKG-Kommentar: Telekommunikationsgesetz. München: C. H. Beck.
- Groves, Robert M./Kahn, Robert L. (1979): Survey by Telephone. New York: Wiley
- Groves, Robert M. (1989): Survey Errors and Survey Costs: An Introduction to Survey Errors. New York: Wiley
- Haad, Hannes/Jerabek, Hynek/Petersen, Thomas (Hg.) (2012): The Early Days of Survey Research and Their Importance Today. Wien: Braumüller
- Häder, Michael/Häder, Sabine (Hg.) (2009): Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung. Wiesbaden: VS Verlag
- Häder Michael/Häder, Sabine/Kühne, Mike (Hg.) (2012): Telephone Surveys in Europe. Research and Practice. Heidelberg: Springer
- Karol, John J. (1937): Measuring Radio Audiences. In: Public Opinion Quarterly 1: 92-96
- Lavrakas, Paul J. (1987): Telephone Survey Methods. Sampling, Selection and Supervision. Beverly Hills: Sage
- West, Brady T./Groves, Robert M. (2013): A Propensity-Adjusted Interviewer Performance Indicator. In: Public Opinion Quarterly 77: 352-374
- Wilms, Heinrich/Jochum, Georg (2013): §104 Rn 9ff. In: Geppert/Schütz et al. (Hg.)

Karl-Heinz Reuband

---

## 47.1 Traditionen schriftlich-postalischer Befragungen

Die schriftlich-postalische Befragung ist eine Variante der standardisierten Befragung (Reinecke, Kapitel 44 in diesem Band), bei der sowohl bei der Kontaktaufnahme als auch bei der Rückgabe eine postalische Zustellung gewählt wird. In manchen Fällen schriftlicher Befragung beschränkt sich der postalische Teil auf die Rücksendung, und der Erstkontakt erfolgt in anderer Weise – z.B. über eine Verteilung des Fragebogens statt Zusendung. In diesen Fällen ist es hilfreich, sich ebenfalls an den grundlegenden Prinzipien postalischer Befragungen zu orientieren.

Postalische Befragungen können auf eine lange Tradition zurückblicken – eine Tradition, die freilich durch Diskontinuitäten, Phasen der Marginalisierung und Abwertung gekennzeichnet ist. Die ersten groß angelegten Erhebungen, welche die Allgemeinbevölkerung zum Thema hatten, bedienten sich – anders als dies heute üblich ist – postalischer Befragungen. Dies geschah in den USA bereits Ende des 19. Jahrhunderts. Federführend waren die Massenmedien, welche die Veröffentlichung von Meinungsumfragen als Mittel der eigenen Verkaufsförderung nutzten. Am häufigsten waren „Probeabstimmungen“ zur Prognose von Wahlergebnissen („straw polls“). Gelegentlich auch wurden aktuelle politische Kontroversen aufgegriffen – z.B. ob man die Prohibition von Alkohol beibehalten solle oder nicht.

Die Umfragen wurden in Eigenregie durch die Zeitungen bzw. Zeitschriften selbst durchgeführt. Dabei beschränkte man sich in der Regel auf ein bis zwei Fragen, gedruckt auf einer Postkarte. Die Zahl angeschriebener Personen war erheblich, sie ging bis in mehrere Millionen. Und teilweise nahmen auch Millionen Menschen an den Befragungen teil (Frankovic 2012). Zu den bekanntesten Umfragen dieser Art gehören die der Zeitschrift „Literary Digest“. Lange Zeit war sie mit ihren Prognosen recht erfolgreich gewesen, bis es im Jahr 1936 in geradezu spektakulärer Weise zum Desaster kam und eine massive Fehlprognose abgegeben wurde. Dabei hatte man diesmal sogar 10 Mio. Postkarten verschickt

und 2,3 Mio. zurückerhalten. George Gallup, der – mitfinanziert von der „New York Herald Tribune“ – zeitgleich mit einer systematischen Quotenstichprobe (Häder/Häder, Kapitel 18 in diesem Band) und mit einem Interviewerstab (Glantz/Michael, Kapitel 21 in diesem Band) „face-to-face“ (Stocké, Kapitel 45 in diesem Band) „lediglich“ mehrere Tausend Personen befragt hatte, hatte demgegenüber den Wahlsieger korrekt vorhergesagt.

Welche Gründe auch für die Fehlprognose des „Literary Digest“ verantwortlich gewesen sein mögen – eine geringe Ausschöpfungsquote (Engel/Schmidt, Kapitel 23 in diesem Band), eine verzerrte Stichprobe durch Auswahl von Personen mit registriertem Telefon- oder Autobesitz oder andere Gründe: mit der massiven Fehleinschätzung des Wahlausgangs wurde nicht nur das Ende der Zeitschrift eingeleitet, sondern ebenso die Praxis der „straw polls“. In den Sozialwissenschaften führte das Desaster nicht nur zu einem verschärften Bewusstsein um die Bedeutsamkeit der Stichprobenziehung (Häder/Häder, Kapitel 18 in diesem Band). Auch die postalische Befragung wurde als solche (vermeintlich) diskreditiert. Andere Verfahren, gestützt auf Quotenstichproben (und später auf Random-Route - Stichproben) und der Einsatz eines Interviewerstabes schienen besser geeignet, die Gesamtbevölkerung in repräsentativer Weise abzubilden. Zudem waren die Ausschöpfungsquoten vielversprechender: Sie lagen weit über denen, die man zuvor bei postalisch durchgeföhrten Umfragen erreicht hatte.

Trotz dieser Abkehr verschwanden postalische Erhebungen nicht völlig aus dem Blickfeld der Sozialforschung. Insbesondere in der Markt- und Mediennutzerforschung fanden sie weiterhin Anwendung und führten in den 1940- und 1950er-Jahren (vereinzelt schon in den 1920er- und 1930er-Jahren) zu methodischen Erkenntnissen, die bis heute Geltung haben. Dazu zählten u.a. die Bedeutsamkeit des Themas für die Teilnahmebereitschaft und die zentrale Rolle von Erinnerungsaktionen für die Ausschöpfungsquote und die Zusammensetzung der Befragten. Des Weiteren gehört dazu die Erkenntnis, dass bei schriftlich-postalischen Befragungen eine größere Bereitschaft zur offenen Beantwortung sensibler Fragen (Hlawatsch/Krickl, Kapitel 20 in diesem Band) besteht als bei face-to-face Befragungen (Lazarsfeld 1935).

Die Erfahrungen mit postalischen Befragungen, insbesondere in den USA, führten in den 1950er- und den 1960er-Jahren in den USA zu größeren Übersichtsarbeiten und Empfehlungen, die sich in leicht abgewandelter Form bis heute in entsprechenden Veröffentlichungen wiederfinden. Die Erkenntnisse fanden frühzeitig ebenfalls Eingang in deutsche Publikationen, wie in dem von René König herausgegebenen Sammelband „Das Interview“ (erstmals 1952, mit mehreren Auflagen bis in die 1970er-Jahre). Eine aktuelle, umfassend und kompetent geschriebene Bestandsaufnahme wurde des Weiteren in den 1970er-Jahren von Klaus Wieken im Rahmen der mehrbändigen Reihe „Techniken der Sozialforschung“ vorgelegt. Damit blieben postalische Befragungen zumindest als Option weiterhin im Methodenkanon der deutschen Sozialforschung sichtbar, wenn auch mit zwiespältiger Einschätzung ihrer Potentiale und Probleme.

Die Bilanz, die in den damaligen Veröffentlichungen über postalische Befragungen gezogen wurde, war alles in allem positiv: Wenn man die geeigneten Strategien einsetzt – so das Fazit – sind in Populationen, bei denen zwischen dem Thema der Untersuchung

und den eigenen Interessen ein enger Bezug existiert, sowohl hohe Ausschöpfungsquoten als auch eine hohe Datenqualität möglich. Erhebliche Zweifel äußerte man jedoch gegenüber einem Einsatz in der Allgemeinbevölkerung.

---

## 47.2 Neue Entwicklungen und alte Vorbehalte

Es blieb der 1978 in den USA erschienenen Monographie von Don Dillman vorbehalten, mit seiner Übersicht über Studien und Strategien zur Erhöhung der Ausschöpfungsquote und hoher Datenqualität die pessimistische Beurteilung nachhaltig zu verändern. Deutlich wurde, dass man auch in der Allgemeinbevölkerung hohe Ausschöpfungsquoten erreichen kann und dass die verbreitete Einschätzung nicht notwendigerweise zutrifft, wonach die Allgemeinbevölkerung mit postalischen Befragungen überfordert ist. Dass das, was Dillman für die USA als Empfehlungen in seiner „*Tailored Design Method*“ entwickelte (später weiter elaboriert in Dillman 2000), ebenfalls auf andere Länder erfolgreich angewandt werden kann und hohe Ausschöpfungsquoten möglich sind, wurde in den 1980er- und 1990er-Jahren dann auch für Deutschland gezeigt (Hippler 1985, Reuband/Blasius 1996, Reuband 2001).

Mit der Erkenntnis, dass postalische Befragungen akzeptable Ausschöpfungsquoten in der Bevölkerung erreichen können, hat sich die Beurteilung des Verfahrens in den letzten Jahren allmählich in der Umfrageforschung verändert. Besonders in den USA zeichnet sich eine vermehrte Aufgeschlossenheit ab, auch wenn postalische Befragungen in der Allgemeinbevölkerung nach wie vor eine Seltenheit darstellen. Ein wesentlicher Grund für die Neubewertung dort liegt einerseits in sinkenden Ausschöpfungsquoten und gestiegenen Kosten bei face-to-face- und Telefonumfragen und andererseits in der erst seit kurzem bestehenden Verfügbarkeit von Adressen auf Bevölkerungsebene. Seitdem Adressen aus der Postzustellung über den „*postal delivery sequence file*“ der U.S. Post verfügbar sind, hat das „*address based sampling*“ in der Umfrageforschung der USA an Bedeutung gewonnen.

Hinzu kommt zu Gunsten postalischer Befragungen, dass selbstadministrierte schriftliche Befragungen in der Online-Forschung (Wagner/Hering, Kapitel 48 in diesem Band) an Bedeutung gewonnen haben und man die Qualität schriftlicher Befragungen inzwischen weniger kritisch beurteilt als noch vor einigen Jahren. Das Bildungsniveau der Bevölkerung ist zudem gestiegen, die Zahl der Personen, die Schwierigkeiten beim Ausfüllen eines Fragebogens haben könnten, ist geringer geworden. Manche Autoren glauben infolgedessen an eine zukünftige Renaissance postalischer Erhebungen, auch in allgemeinen Bevölkerungsbefragungen.

Im Vergleich zu den USA ist in der deutschen Sozialwissenschaft die Beurteilung postalischer Befragungen reservierter. Dabei hätte man eher eine Aufgeschlossenheit erwarten können, da – anders als in den USA – in Deutschland eine brauchbare Adressenbasis für die Stichprobenziehung über die Einwohnermeldeämter seit jeher existiert. Davon wird seitens der städtischen statistischen Ämter im Rahmen ihrer postalischen

„Bürgerbefragungen“ seit mehreren Jahren auch Gebrauch gemacht (DifU 2005), nicht aber von der „academic community“.

Der Grund für die Zurückhaltung der „academic community“ ist methodischer Art. Am häufigsten wird – auch in Lehrbüchern – vorgebracht, postalische Befragungen würden oft geringe Ausschöpfungsquoten erreichen. Die Aussage ist richtig, aber trifft nicht den Kern des Problems. Auch andere Verfahren erreichen mitunter geringe Ausschöpfungsquoten, besonders wenn sie nicht aufwändig genug durchgeführt werden. Die Frage ist daher korrekterweise nicht ob, sondern unter welchen *Bedingungen*, zufriedenstellende Antwortquoten bei postalischen Befragungen möglich sind. Der Grund für geringe Quoten bei postalischen Befragungen liegt häufig darin, dass ein geringerer Aufwand betrieben wird als bei anderen Befragungsverfahren und aus Kostengründen auf Erinnerungsaktionen verzichtet wird. Wird zudem ein Thema gewählt, das in der Bevölkerung auf geringes Interesse stößt, so sind niedrige Ausschöpfungsquoten (und eine höchst selektive Zusammensetzung der Befragten) eine fast zwangsläufige Folge.

---

### 47.3 Ausschöpfung von Stichproben

Postalische Befragungen in der Allgemeinbevölkerung müssen nicht zwangsläufig geringe Ausschöpfungsquoten erzielen. So haben postalische Befragungen in der deutschen Allgemeinbevölkerung, die im Grundsatz den Strategien von Dillman zur Maximierung von Ausschöpfungsquoten gefolgt sind, in den 1980er- und 1990-Jahren selbst bei umfangreichen Fragebögen zum Teil Ausschöpfungsquoten bis über 70% erreicht (Hippler 1985, Reuband/Blasius 1996, Reuband 2001).

Inzwischen sind die Ausschöpfungsquoten bei postalischen Befragungen in der Allgemeinbevölkerung – ähnlich wie bei anderen Befragungsverfahren – gesunken (DifU 2005, Harm/ Jaek 2012). Doch sind unter optimalen Bedingungen (Befragung der Bevölkerung durch die Universität am Ort oder das statistische Amt der Stadt) Werte von 50% bis 60% durchaus noch möglich (auch bei einem 11-seitigen Fragebogen, dessen Beantwortung rund eine halbe Stunde dauert).

Selbst unter weniger günstigen Umständen (etwa bei Umfragen, die an anderen Orten als dem der durchführendes Forschungsinstitution stattfinden und sich kein entsprechender „Heimat-Bonus“ auswirken kann), können Werte von über 40% erzielt werden (Reuband 2011). Die entsprechenden Werte zeitgleicher (telefonischer und persönlich-mündlicher) Befragungen des ALLBUS und SOEP liegen darunter. Damit hat sich das Bild von der Überlegenheit mündlicher Befragungen gegenüber postalischen Befragungen in Fragen der Ausschöpfung ins Gegenteil gekehrt.

Nicht nur in der Allgemeinbevölkerung haben postalische Befragungen zufriedenstellende bis sehr gute Ausschöpfungsquoten erzielen können, sondern auch in Gruppen, die als schwer befragbar gelten, z.B. bei Alten (Motel-Klingenbiel/Klaus/Simonson, Kapitel 59 in diesem Band) und Personen in ungünstigen soziökonomischen Verhältnissen (Schütte/Schmies, Kapitel 61 in diesem Band). So wurde in einer postalischen Befragung

von Personen im Alter ab 71 Jahren mit Fragen primär zur Zeit des Dritten Reiches im Jahr 1999 eine Ausschöpfungsquote von rund 50% erreicht (Reuband 2006), ähnlich hoch fiel die Quote in einer Erhebung in sozial deprivierten Stadtteilen von NRW aus. Des Weiteren haben sich postalische Befragungen auch bei Panelstudien, die einen Zeitraum von mehreren Jahren umfassen und mit einem besonderen Aufwand verbunden sind, bewährt (Reuband 1998).

Postalische Befragungen, welche die nicht-deutsche Bevölkerung (El-Menouar, Kapitel 60 in diesem Band) einbeziehen und sich eines deutschen Fragebogens bedienen, erbringen demgegenüber niedrigere Ausschöpfungsquoten. So erzielte z.B. eine Frankfurter Bürgerbefragung des Amtes für Statistik im Jahr 2012 unter den Ausländern, die seit mindestens einem Jahr in Frankfurt leben, nur eine Ausschöpfungsquote von 29%; unter den Personen mit deutscher Staatsangehörigkeit waren es 55% (Frankfurt Statistik Aktuell 2013/15). Würde man einen Fragebogen einsetzen, der in der Sprache der Migranten bzw. Ausländer verfasst ist (Brick et al. 2012), wäre die Ausschöpfungsquote sicherlich höher. Ob sie jedoch die gleichen Werte wie unter Personen mit deutscher Herkunft erreichen würde, ist fraglich – nicht zuletzt, da man – je nach Geschlecht, ethnischer und regionaler Herkunft – auch bei der eigenen Muttersprache gewisse Probleme der Schriftlichkeit erwarten kann (El-Menouar, Kapitel 60 in diesem Band).

In neuerer Zeit haben postalische Befragungen vermehrt in Form von Mixed-Mode-Designs an Bedeutung gewonnen. Die eine Variante besteht darin, den Zielpersonen mehrere Verfahren zur Auswahl zu stellen. Eine andere besteht im sequentiellen Wechsel der Befragungsformen. Wo Optionen zwischen Online- und postalischer Befragungen bestanden, hat die ausschließliche Verwendung der postalischen Befragungsoption die höhere Ausschöpfungsquote erbracht (Messer/Dillman 2011). Auch hat sich gezeigt, dass in der Allgemeinbevölkerung die Bearbeitung des Fragebogens in Papierform der Bearbeitung im Rahmen einer Online Befragung vorgezogen wird.

---

## 47.4 Potentiale und Probleme

### 47.4.1 Verfügbarkeit einer adressbasierten Stichprobe

Voraussetzung für die Durchführung postalischer Befragungen ist das Vorhandensein von Adressen. Für Forschung, die im öffentlichen Interesse ist (wie universitäre Forschung), ist in Deutschland ein Zugang über die Einwohnermeldeämter möglich. Im Fall rein kommerzieller Forschung entfällt dies. Aus diesem Grund hat es unter den Markt- und Sozialforschungsinstituten auch keine Neigung gegeben, sich in nennenswertem Maße postalischer Befragungen zu bedienen oder ein Interesse daran zu entwickeln. Für universitäre Einrichtungen hätte es solche Tendenzen geben können, aber sie hielten sich in Grenzen.

Einwohneradressen liegen in der Bundesrepublik – anders als etwa in Skandinavien – lediglich auf der Gemeindeebene vor. Bundesweite Erhebungen sind somit nur über den Umweg über die Meldeämter der einzelnen Gemeinden möglich. Dies erfordert eine

aufwändige, mehrstufige Stichprobenziehung und ist ein Grund, warum in Deutschland nahezu alle postalischen Befragungen in der Allgemeinbevölkerung auf Stadt- oder Gemeindeebene beschränkt sind. Dies ist jedoch keine zwingende Notwendigkeit und kein prinzipieller Einwand gegenüber dem Einsatz postalischer Befragungen an sich: ein mehrstufiges Auswahlverfahren, das zu einer personenbezogenen Adressenstichprobe führt, ist inzwischen auch bei anderen bundesweiten Erhebungen, wie dem ALLBUS, üblich.

#### **47.4.2 Kosten- und Qualitätsvorteile**

Für den Einsatz einer postalischen Befragung sprechen bei Verfügbarkeit einer adressbasierten Stichprobe sowohl praktische und ökonomische Gründe als auch solche der Datenqualität.

1. *Geringe Kosten:* Postalische Befragungen sind billiger als face-to-face- und telefonische Erhebungen, und sie erfordern einen geringen personellen und organisatorischen Aufwand. Es bedarf keines aufwändigen und kostspieligen Interviewereinsatzes *und keines aufwändigen technischen und organisatorischen Apparats*.
2. *Erreichbarkeit der Zielperson:* Es werden alle Zielpersonen erreicht, für die eine gültige Adresse vorliegt. Bei face-to-face- und telefonischen Befragungen sind hingegen mehrere Anläufe über einen längeren Zeitraum notwendig. In dem Maße, wie die Zahl der Singlehaushalte gestiegen ist, die Zahl der Haushaltsmitglieder gesunken ist und die Freizeit zunehmend außer Haus verbracht wird, ist es zusehends schwieriger geworden, Personen persönlich zu erreichen.
3. *Reduktion sozialer Erwünschtheitseffekte bei der Beantwortung:* Durch die Abwesenheit eines Interviewers ist eine größere Anonymität als bei face-to-face- und telefonischen Befragungen gegeben. Soziale Erwünschtheitseffekte (Hlawatsch/Krickl, Kapitel 20 in diesem Band) treten seltener auf (Tourangeau et al. 2000, Reuband 2002). Gleichwohl ist auch bei postalischen Befragungen von Beschönigungstendenzen auszugehen - primär aufgrund des Selbstbildes, das der Befragte von sich hat.
4. *Fehlender Beantwortungsdruck:* Der Befragte hat eine Kontrolle über die Interviewsituation, er ist keinem Druck ausgesetzt und kann sich Zeit bei der Beantwortung der Fragen lassen – was bei Erinnerungsfragen der Qualität der Angaben zugute kommen kann. Auch ist denkbar, dass die Kontrolle über die Befragungssituation bei manchen Personen (besonders mit niedriger Bildung) die Angst zerstreut, sich im Interview Blößen zu geben. Dies könnte die Bereitschaft zur Teilnahme (Engel/Schmidt, Kapitel 23 in diesem Band) fördern.

### 47.4.3 Probleme, die aus der Abwesenheit des Interviewers erwachsen

Die Abwesenheit eines Interviewers hat den Vorteil erhöhter Anonymität und einer offeneren Beantwortung der Fragen. Die Kehrseite dessen ist, dass Motivierung, Unterstützung und Kontrolle durch den Interviewer entfallen. Man kann jedoch versuchen, diese Nachteile durch entsprechende Maßnahmen zu kompensieren.

1. *Fehlende Motivierung durch einen Interviewer:* Die Motivierung muss im Wesentlichen durch das Anschreiben geleistet werden, in Kombination mit dem Themenaufbau des Fragebogens. Mehr als bei face-to-face- und telefonischen Umfragen ist daher eine besonders *sorgfältige Gestaltung des Anschreibens und des Fragebogen erforderlich*.
2. *Verständnisschwierigkeiten.* Stärker als bei face-to-face- und telefonischen Befragungen bedarf es sorgfältiger Prüfungen und Pre-Tests des Fragebogens im Hinblick auf Fragekonstruktion, Fragebogenlayout und Filterführung (Weichbold, Kapitel 19 in diesem Band). Dass manche Fragen falsch verstanden oder falsch beantwortet werden (z.B. Ankreuzen zwischen den Kategorien, unzulässige Mehrfachnennungen), ist dennoch nicht zu vermeiden. Daher ist nach Eingang der Fragebögen ein größerer Prüfungsaufwand, ggf. mit Neucodierungen (Lück/Landrock, Kapitel 28 in diesem Band), als bei anderen Erhebungsformen notwendig.
3. *Bildungs-Bias aufgrund eingeschränkter kognitiver Kompetenz?* Personen mit niedrigem Bildungsniveau wird in der Literatur unterstellt, dass sie größere Schwierigkeiten haben, einen Fragebogen zu beantworten und daher seltener teilnehmen als bei face-to-face- und telefonischen Befragungen. Personen mit schlechter Bildung sind in postalischen Umfragen in der Tat unterrepräsentiert. Aber sie scheinen den bisherigen Studien zufolge, die sich auf Personen mit deutscher Staatsangehörigkeit beziehen, nicht stärker unterrepräsentiert zu sein als in face-to-face- und telefonischen Befragungen (Reuband/Blasius 1996, Reuband 2002). Wo in nennenswertem Maße Personen mit ausländischer Herkunft einbezogen sind (die zudem mit geringer Bildungsqualifikation ausgestattet sind, wie in machen U.S. Studien), dürfte sich die Situation anders darstellen.
4. *Fehlende Kontrolle der Interviewsituation:* Dass eine angeschriebene Person die Beantwortung des Fragebogens einem anderen Mitglied des Haushalts überlässt, ist theoretisch zwar denkbar, hat sich aber aufgrund externer Überprüfungen anhand von Registerdaten als ein zu vernachlässigendes Problem erwiesen (Reuband 1999b: 91). Ebenso unwahrscheinlich ist es, dass andere Personen in größerem Umfang beim Ausfüllen des Fragebogens helfen (Kunz 2010: 137). Selbst bei hochbetagten Alten (Motel-Klingenbiel/Klaus/Simonson, Kapitel 59 in diesem Band), bei denen üblicherweise besondere Probleme erwartet werden, kommt dies relativ selten vor (Reuband 2006: 115).

## 47.5 Praktische Durchführung postalischer Befragungen: Was ist zu beachten?

Um hohe Ausschöpfungsquoten zu erreichen, sind mehrere Bedingungen zu beachten, die von der Thematik des Fragebogens bis zur Zahl der Erinnerungen reichen und in der jeweiligen Kombination ihre nachhaltige Wirksamkeit entfalten.

### 47.5.1 Thema der Untersuchung und Fragebogenaufbau

Der Bezug zwischen Thema der Umfrage und Interessenlage auf Seiten der Zielpersonen stellt eine der wichtigsten Einflussgrößen für die Bereitschaft zur Teilnahme an Befragungen dar (Groves et al. 2004 Engel/Schmidt, Kapitel 23 in diesem Band). Je nach Thema und Themenvielfalt fühlen sich viele oder wenige Menschen angesprochen. Bei face-to-face- und telefonischen Befragungen ist es für den Interviewer möglich, die Themenstellung lediglich global und vage zu benennen (hat das Interview erst mal begonnen, wird es selten abgebrochen). Bei postalischen Befragungen jedoch liegt der Fragebogen dem Befragten vor.

Um eine themenspezifische überproportionale Selektion von Interessierten zu reduzieren oder zu vermeiden, ist es ratsam, *Themen von allgemeinem Interesse sowohl im Anschreiben als auch im Fragebogen in den Vordergrund zu rücken*. Die eigentlich den Forscher interessierenden Themen können dann in einer gewissen Mischung an späterer Stelle positioniert werden. Zwar ist davon auszugehen, dass der Befragte den Fragebogen vor der Beantwortung in gewissem Umfang durchblättert, aber die ersten Seiten werden vermutlich den größeren Eindruck hinterlassen (womöglich zunächst auch als einzige betrachtet) und daher in ihrer Themensetzung für die Motivierung umso wichtiger sein.

### 47.5.2 Anschreiben

Da bei postalischen Befragungen der Interviewer entfällt, kommt dem Anschreiben ein *strategischer Stellenwert* zu. Es muss die Informationen liefern, die üblicherweise vom Interviewer vermittelt werden. Und es müssen all jene möglichen Einwände antizipiert und berücksichtigt werden, die gegen die Befragung vorgebracht werden können. Zugleich soll das Anschreiben kurz gefasst und nicht länger als eine Seite sein. In Anlehnung an Dillman müssen die folgenden Angaben enthalten sein:

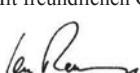
- *Offizieller Briefkopf*
- *Datumsangabe*
- *Personalisierte Anrede*
- *Thema und Ziel der Studie*: Die Angaben dazu sollten eher allgemein als zu spezifisch gehalten sein, um einen breiten Kreis der Zielpersonen mit unterschiedlichen Interessen anzusprechen und zu motivieren.

- *Bedeutsamkeit der Teilnahme:* Die Tatsache, dass die Teilnahme für das Gelingen der Untersuchung bedeutsam ist, sollte betont werden und die Tatsache, dass nur eine begrenzte Zahl von Personen stellvertretend für die Gesamtheit befragt wird.
- *Angaben zum Auswahlverfahren:* Es sollte angegeben werden, wie die Person in der Stichprobe ausgewählt wurde. Dabei geht es nicht um Details, sondern um allgemeine Informationen (wie z.B. Auswahl nach einem wissenschaftlich-statistischen Verfahren aus dem Einwohnermeldeamtsregister).
- *Vertraulichkeitszusicherung, inkl. Erklärung der Fragebogennummer (sofern dort vorhanden):* Vertraulichkeit und Einhaltung der Anonymität ist zuzusichern. Der Hinweis sollte nicht nur aus Platzgründen knapp gehalten werden (es steht ja für den Gesamttext nur eine Seite zur Verfügung), sondern auch weil eine zu starke Hervorhebung des Anonymitätsaspekts Misstrauen wecken kann. Ist der Fragebogen mit einer Nummer versehen (ID), z.B. um nur noch diejenigen Personen in die Erinnerungsaktionen einzubeziehen, die bisher nicht geantwortet haben – ist dies zu erwähnen und knapp zu begründen. Andernfalls könnte die Anonymität der Datenerhebung und – Auswertung bezweifelt werden.
- *Erläuterung der Rücksendung, inkl. Rücksendedatum:* In Bezug auf das Rücksendedatum ist es ratsam, einen Zeitraum zu nennen, der relativ knapp gefasst ist (ca. 3-5 Tage), um ein Verschieben der Beantwortung (und ein Vergessen der Beantwortung) zu vermeiden.
- *Angebot der Kontaktaufnahme (Telefon)*
- *Dank und Unterschrift:* Laut Dillman sollte die Unterschrift des Forschers mit Kugelschreiber geleistet werden. Dabei sollte die Farbe blau und nicht schwarz verwendet werden und bei der Unterzeichnung Druck ausgeübt werden. Dadurch soll dem Empfänger des Schreibens der Eindruck vermittelt werden, dass eine persönliche, eigenständige Unterschrift geleistet wurde. Neuere deutsche Untersuchungen haben jedoch gezeigt, dass es keinen Unterschied für die Ausschöpfungsquote hat, ob eine gescannte Unterschrift oder eine Originalunterschrift gewählt wurde (Reuband 2001).

Ein Beispiel für ein Anschreiben ist im Abb. 47.1 wiedergegeben.

### 47.5.3 Umfang des Fragebogens

Die Bereitschaft zur Teilnahme an der Befragung ist von der Länge des Fragebogens abhängig. *Kürzere Fragebögen (von 4-6 Seiten)* haben höhere Ausschöpfungsquoten als längere Fragebögen. Laut Dillman sollte der Umfang 11 Seiten Text (plus 1 Seite Deckblatt) nicht überschreiten. Liegt die Seitenzahl höher, gäbe es einen stärkeren Abfall der Ausschöpfungsquote. Über die *Zahl der Fragen* ist damit freilich nichts ausgesagt: Je nach *Layout* Gestaltung ist es möglich, unterschiedlich viele Fragen auf einer Seite unterzubringen, ohne die Lesbarkeit zu beeinträchtigen. In welchem Umfang die Dichte der Fragebogen-gestaltung und die Zahl der Seiten über die Teilnahme an der Befragung entscheiden, ist – auch im Kontext unterschiedlicher Themen – bislang nicht eindeutig geklärt.

<p><b>Sozialwissenschaftliches Institut</b>  <b>Lehrstuhl Soziologie II</b>            Prof. Dr. Karl-Heinz Reuband</p>		<p>D-40225 Düsseldorf, den 16.09.2009            Universitätsstr. 1            Tel.: 0211-81-13123            Tel.: 0211-81-15301            Fax: 0211-81-12263</p>
<p><b>WENN UNZUSTELLBAR ZURÜCK</b></p> <p>Universität Düsseldorf · Sozialwissenschaftliches Institut · 40225 Düsseldorf</p>		
<p>Herr            Hans Mustermann            Musterstr. 1            40225 Düsseldorf</p>		
<p>Sehr geehrter Herr Mustermann</p>		
<p>wir führen zurzeit in Düsseldorf eine größere Befragung durch. In dieser geht es um aktuelle Fragen. Wir möchten gern erfahren, wie die Menschen heutzutage in Düsseldorf leben, welche Erfahrungen sie machen und was sie über verschiedene Themen denken, z.B. wie zufrieden sie mit der Stadt sind, ob sie das Internet nutzen, was sie von Rauchverboten halten, wie sehr sie sich durch Kriminalität bedroht fühlen usw.</p>		
<p>Für die Befragung wurden aus dem Einwohnerverzeichnis der Stadt Düsseldorf nach einem wissenschaftlich-statistischen Zufallsverfahren 2000 Adressen ausgewählt und vom Einwohnermeldeamt zur Verfügung gestellt. Diese Bereitstellung geschieht jeweils nur dann, wenn das Projekt - wie unseres - von öffentlichem Interesse ist, wissenschaftlichen Prinzipien folgt und man von einer Einhaltung des Datenschutzes ausgehen kann. Unter den bereitgestellten Adressen befand sich auch Ihre. Ihre Mitarbeit ist freiwillig. Nehmen Sie nicht teil, entstehen Ihnen keine Nachteile.</p>		
<p><b><i>Für das Gelingen der Untersuchung ist Ihre Mitarbeit jedoch unbedingt erforderlich. Um ein richtiges Gesamtbild zu erhalten, ist es äußerst wichtig, dass alle von uns angeschriebenen Bürger den Fragebogen ausfüllen.</i></b></p>		
<p>Ihre Angaben werden selbstverständlich streng vertraulich behandelt. Die Angaben werden in Zahlen umgesetzt und ohne Ihren Namen und Ihre Adresse ausgewertet und ausschließlich für Gruppen zusammengefasst dargestellt. Wer die Angaben gemacht hat, ist somit nicht mehr zu erkennen. Die auf dem Fragebogen angegebene Nummer dient lediglich zu unserer Information: damit wir wissen, wer uns bereits geantwortet hat und niemand unnötigerweise ein weiteres Mal angeschrieben wird.</p>		
<p>Die Beantwortung des Fragebogens nimmt nicht allzu viel Zeit in Anspruch. Wir bitten Sie, den Fragebogen in den nächsten 3-4 Tagen auszufüllen und in dem beiliegenden, bereits frankierten Rückumschlag an uns zurückzusenden.</p>		
<p>Für Rückfragen stehen wir Ihnen gern über die oben angegebenen Telefonnummern zur Verfügung. Für Ihre Teilnahme möchten wir im Voraus recht herzlich danken.</p>		
<p>Mit freundlichen Grüßen</p>		
		
<p>Prof. Dr. Karl-Heinz Reuband</p>		
<p>Anlage: Fragebogen, Rücksendeumschlag für kostenfreien Versand</p>		

**Abb. 47.1** Erstes Anschreiben bei einer postalischen Befragung.

Wählt man beim Fragebogendruck dünnes Papier (etwa um Porto zu sparen), ist zu bedenken, dass dies leicht zum Überblättern der Seiten führen kann. Daher sollte u.E. eher normales 80g- Papier gewählt werden. Ob man den Fragebogen als Booklet in DIN A5 Format druckt (wie einst von Dillman empfohlen) oder in DIN A4, ist für die Ausschöpfungsquote vermutlich irrelevant. Nicht irrelevant ist es jedoch für die Fragebogengestaltung: Wählt man ein kleines Format, muss man der Lesbarkeit wegen (gerade bei Älteren) eine etwa vergleichbare Schriftgröße wählen wie beim DIN A4 Format (11er- oder 12er-Schrift). Dies hieße entweder gegenüber dem DIN A4 Format weniger Fragen zu stellen oder den Seitenumfang zu erhöhen.

#### 47.5.4 Art und Zeitpunkt des Versands

Der Eindruck einer Massensendung soll vermieden werden. Daher wurde in der Literatur lange Zeit empfohlen, eine Briefmarke – am besten eine Sonderbriefmarke – zu verwenden und keine Frankiermaschine zu benutzen. Experimente haben jedoch ergeben, dass es keinen Unterschied zwischen Frankiermaschine und Briefmarke im Hinblick auf die Ausschöpfungsquote gibt (Reuband 2001) – womöglich solange, wie deutlich ist, dass es sich bei dem Absender um eine renommierte Institution (wie eine Universität) handelt. Aus diesem Grund ist es in Deutschland möglich, die kostengünstigste Variante einer Massensendung in Form eines „Infobriefs“ zu wählen, ohne dass diese die Teilnahmewahrscheinlichkeit beeinträchtigt.

Werden die Fragebögen per Infobrief versandt, müssen die Briefe in aufsteigender Reihenfolge nach Postleitzahl geordnet in den Postversand gegeben werden. Dabei ist eine bestimmte Mindestzahl an Briefen pro Postleitzirkusziffer erforderlich, anderenfalls fallen die normalen Portokosten an. Für die Rücksendung des Fragebogens wird dem Anschreiben ein Rückantwortumschlag beigelegt. Er muss nicht mit einer Briefmarke versehen sein. Es reicht erfahrungsgemäß und spart Kosten, ihn mit dem Ausdruck „Gebühr bezahlt Empfänger“ zu versehen.

In der Literatur wird empfohlen, den Versand ungefähr in der Mitte der Woche stattfinden zu lassen, um die Chance einer Beantwortung zu erhöhen. Dabei geht man davon aus, dass die Zielpersonen am Wochenende mehr Zeit für die Beantwortung haben als in der Woche und eher geneigt sind, den Fragebogen zu beantworten. Zeiten, in denen die Zielpersonen vermehrt im Urlaub sind – z.B. in den Schulferien –, sind möglichst zu vermeiden. Gleiches gilt für die Weihnachtszeit (auch wenn der Effekt womöglich nicht so stark ist, wie oft vermutet, Reuband 2001: 313).

Nicht alle der gezogenen *Adressen* sind zum Zeitpunkt der Versandaktion gültig. Zum Teil ist die fehlende *Aktualität* der Melderegister ein Grund, zum Teil die zwischenzeitlich verstrichene Zeit zwischen Stichprobenziehung und Einsatz der Adressen. Um die Zahl der neutralen Ausfälle (wie verzogen, verstorben) für die Berechnung der Ausschöpfungsquote zu bestimmen, sollte der Briefkopf mit dem Hinweis versehen sein „Wenn unzustellbar zurück“. Dieser Aufdruck ist zwar keine Garantie dafür, dass alle nicht zustellbaren Briefe von der Post zurückgeschickt werden (Reuband 1999b: 81), aber für den größeren

Teil dürfte dies in der Regel wohl der Fall sein. Wird auf den Aufdruck verzichtet und werden keine Rückmeldungen erfasst, werden die Zahl neutraler Ausfälle und damit auch die Ausschöpfungsquote als Ganzes unterschätzt.

#### 47.5.5 Erinnerungsaktionen

Erinnerungsaktionen sind als der wichtigste Faktor anzusehen, der zur Erhöhung der Ausschöpfungsquote beiträgt. Erforderlich sind in allgemeinen Bevölkerungsumfragen mindestens zwei Erinnerungsaktionen. Weitere sind je nach Forschungsziel und Erhebungsverlauf zu erwägen. Die dritte Erinnerungsaktion erbringt gewöhnlich nur noch einen minimalen Zuwachs an Befragten und ist kostenaufwändig. Andererseits kann es unter Umständen notwendig sein, die Zahl der Befragten weiter zu erhöhen, die soziale Zusammensetzung der realisierten Stichprobe zu verbessern und/oder nähere Analysen auf der Basis der durch die Mahnaktionen rekrutierten Befragten zu unternehmen. Dabei werden oftmals diejenigen, die erst nach der letzten Mahnung antworten, zur Schätzung des Sozialprofils und der Orientierungen der „harten“ Verweigerer herangezogen. Ein Beispiel für die erste „Erinnerungsaktion“ ist in Abb. 47.2 wiedergegeben.

Auf Erinnerungsaktionen zu verzichten, bedeutet, das Potential postalischer Befragungen und deren Datenqualität nicht auszuschöpfen: nicht nur erhält man eine besonders niedrige Ausschöpfungsquote, man läuft auch Gefahr eine höchst selektive, sozial verzerrte und thematisch unausgewogene Zusammensetzung von Befragten zu erhalten. Durch Erinnerungsaktionen werden gewöhnlich überproportional diejenigen einbezogen, die in Umfragen unterrepräsentiert sind: politische Desinteressierte sowie Personen, die an den Themen der Befragung wenig Interesse haben. Des Weiteren werden häufig die schlechter Gebildeten verstärkt erfasst (Reuband 2001, Petermann 2005, Kunz 2010).

Die Datenqualität, gemessen u.a. an der Zahl ausgelassener Fragen, verschlechtert sich zwar leicht durch die Erinnerungsaktionen. Doch hält sich diese Tendenz für die Gesamtzahl der Befragten in Grenzen. Sie wirkt sich allenfalls unter den Befragten in der dritten Erinnerungsaktion in etwas stärkerem Maße aus. Ein Grund für die Verschlechterung in dieser Subgruppe liegt in deren geringer Teilnahmemotivation. Darüber hinaus spiegelt sich darin wohl auch ein real vorhandener höherer Grad an Meinungslosigkeit und Desinteresse wider (was eine bessere Repräsentanz auch der weniger Informierten durch die erweiterte Umfragebasis bedeutet).

Wo seitens der Forschung auf Erinnerungsaktionen verzichtet wird, geschieht dies oft mit der Begründung, man sei aus Kostengründen dazu gezwungen. Dies ist meist ein Trugschluss: Bei Vorhandensein einer Kennziffer (auf dem Fragebogen oder der beigelegten, getrennt zuzusendenden Karte) kann man sich auf die Information stützen, ob bereits geantwortet wurde, und die erneute Versandaktion selektiv und ökonomisch günstig gestalten. Die Kosten sind unter diesen Bedingungen in vielen Fällen geringer, als wenn einer massenhaften Versendung von Fragebögen der Vorzug gegeben wird, aber nur wenige Fragebögen ausgefüllt zurückkommen.

<p><b>Sozialwissenschaftliches Institut</b> <b>Lehrstuhl Soziologie II</b> Prof. Dr. Karl-Heinz Reuband</p> <p> WENN UNZUSTELLBAR ZURÜCK Universität Düsseldorf · Sozialwissenschaftliches Institut · 40225 Düsseldorf</p>	<p><b>D-40225 Düsseldorf</b> Universitätsstr. 1 Tel.: 0211-81-13123 Tel.: 0211-81-15301 Fax: 0211-81-12263</p>
<p>Herr Hans Mustermann Musterstr. 1</p>	
<p>40225 Düsseldorf</p>	
<p>30. September 2009</p>	
<p>Sehr geehrter Herr Mustermann,</p>	
<p>wir haben Ihnen vor rund einer Woche einen Fragebogen zugesandt, der sich damit befasst, wie die Menschen in Düsseldorf leben und was sie über verschiedene Themen denken.</p>	
<p>Falls der Fragebogen bereits zurückgeschickt wurde, möchten wir dafür recht herzlich danken.</p>	
<p>Sollten Sie den Fragebogen noch nicht zurückgeschickt haben, bitten wir hiermit noch einmal ganz herzlich darum: Weil der Fragebogen nur an einen kleinen, aber repräsentativen Querschnitt der Düsseldorfer Haushalte geschickt wurde, ist es äußerst wichtig, dass sich jeder der Angeschriebenen auch an der Studie beteiligt. Nur so kann eine Repräsentativität der Ergebnisse sichergestellt werden.</p>	
<p>Sollte der Fragebogen nicht eingegangen bzw. abhanden gekommen sein oder sollten Sie Rückfragen haben, rufen Sie uns bitte unter einer der oben angegebenen Telefonnummern an.</p>	
<p>Mit freundlichen Grüßen</p>	
<p> Prof. Dr. Karl-Heinz Reuband</p>	
<p>PS: Sollte die von uns angeschriebene Person verzogen sein, für längere Zeit verreist sein oder aus anderen Gründen an der Befragung nicht teilnehmen können, so wären wir für eine kurze Mitteilung sehr dankbar. Vermerken Sie in diesem Fall bitte den Grund auf diesem Schreiben und senden Sie es in dem beigefügten Rückumschlag zurück.</p>	

**Abb. 47.2** Anschreiben bei erster Erinnerungsaktion

In Anlehnung an die von Dillman entwickelte Methode ist

- ca. *eine Woche nach Erstversand* eine Postkarte oder ein Brief zu versenden, wo für die Teilnahme gedacht wird und diejenigen, die noch nicht geantwortet haben, an die Rücksendung des Fragebogens erinnert werden.
- *Zwei Wochen später* erfolgt eine zweite Erinnerungsaktion mit *erneutem Fragebogen*.
- Und *zwei bis drei Wochen nach der zweiten Erinnerungsaktion* erfolgt eine dritte Erinnerung mit *Fragebogen*.

Weitere Erinnerungsaktionen sind denkbar (auch per Eilboten, Einschreiben, Telefon), haben sich aber als wenig ertragreich erwiesen. Erforderlich sind mindestens zwei Erinnerungen.

Dillman und andere Autoren empfehlen, noch vor Erstversand des Fragebogens ein *Ankündigungsschreiben* zu verschicken. Doch hat sich gezeigt, dass dies die Ausschöpfungsquote zwar erhöht, der Effekt aber nur wirksam ist, wenn auf Erinnerungsaktionen ganz oder teilweise verzichtet wird. Mit zunehmender Zahl an Erinnerungsaktionen verblasst der Effekt des Ankündigungsschreibens auf die Ausschöpfungsquote.

Üblich ist es, bei den Erinnerungsaktionen alle Personen, die bisher noch nicht geantwortet haben, erneut anzuschreiben. Denkbar sind aber auch *selektive Erinnerungen*: wenn z.B. aufgrund entsprechender Informationen aus dem Adressenpool, wie Geschlecht, Alter oder Postleitzahl eine unterdurchschnittliche Teilnehmerate (Engel/Schmidt, Kapitel 23 in diesem Band) ersichtlich ist. Unter diesen Umständen ist es erwägenswert, zielgruppenspezifische Anschreiben zur Motivierung der Personen einzusetzen, die bisher nicht geantwortet haben. Dadurch werden nicht nur Kosten gespart, es kann ggf. auch die soziale Zusammensetzung der realisierten Stichprobe verbessert werden.

Besonderes Gewicht ist bei den Erinnerungsaktionen auf die *Anschreiben* zu legen. Dabei geht es nicht darum, alle Argumente des ersten Schreibens zu wiederholen, sondern einzelne Aspekte, welche die Wichtigkeit der Teilnahme thematisieren, besonders zu betonen. Gegebenenfalls kann dem Anschreiben ein Beiblatt beigelegt werden, in dem die Bedeutsamkeit der Studie hervorgehoben wird oder ausführlicher auf einzelne Fragen der Datenerhebung und Datenauswertung eingegangen wird. In diesem Zusammenhang ist z.B. an Berichte aus der Presse zu denken, die dem Projekt ein besonderes Gewicht geben könnten (in Reaktion auf Pressemitteilungen, die seitens des Projekts versandt wurden).

Bei lokalen Erhebungen auf Stadt- oder Gemeindeebene ist ein Begleitschreiben des Bürgermeisters (oder anderer lokaler Institutionen) denkbar. Ob ein solches Begleitschreiben nützlich sein kann oder nicht, ist natürlich ebenfalls vom Ansehen des Bürgermeisters oder der jeweiligen Institution in der Bevölkerung abhängig und muss jeweils von Fall zu Fall entschieden werden. Auch ist nicht auszuschließen, dass Personen mit Sympathie für den Bürgermeister (bzw. dessen Partei oder Institution) überproportional in die Erinnerungsaktion einbezogen werden. Unter diesen Bedingungen würde die Ausschöpfungsquote erhöht, aber dies durch eine Verzerrung in der Zusammensetzung der Befragten erkauft.

### 47.5.6 Nummerierung des Fragebogens

Eine Nummerierung auf dem Fragebogen ermöglicht es, ökonomisch vorzugehen und sich bei den Erinnerungsaktionen auf die Personen zu beschränken, die bisher nicht geantwortet haben. Eine Nummerierung kann ebenfalls sinnvoll sein, wenn man Kontextdaten dazu lesen will (Cielebak/Rässler, Poetschke, Graeff, Kapitel 26, 87 und 72 in diesem Band) oder mit einem Teil der Befragten weitere Befragungen durchführen will (Schupp, Kapitel 73 diesem Band).

Das Vorhandensein einer Nummer auf dem Fragebogen (etwa auf der ersten Seite rechts oben) hat im Vergleich zu einem Fragebogen ohne Nummerierung nach den bisherigen Erfahrungen in der Regel keinen Einfluss auf die Ausschöpfungsquote und das Antwortverhalten (Reuband 1999a, 2006). Dass Befragte die Nummer vom Fragebogen entfernen, um auf diese Weise Anonymität zu garantieren, ist selten (1-3% der Befragten). Gleichwohl ist es bei manchen sensiblen Themen (Hlawatsch/Krickl, Kapitel 20 in diesem Band), Populationen (Schütte/Schmies, Kapitel 61 in diesem Band) und Zeitumständen erwägenswert, auf eine Nummerierung zu verzichten. Um in einem solchen Fall eine selektive Erinnerung betreiben zu können, sollte dem Anschreiben eine nummerierte Postkarte beigelegt werden, die getrennt vom dem Fragebogen zurückzuschicken ist und auf der von der Zielperson vermerkt wird, ob der Fragebogen ausgefüllt zurückgesandt wurde oder nicht.

Dass Personen die Postkarte zusenden und mitteilen, sie hätten den Fragebogen zurückgeschickt, obwohl sie es nicht getan haben, ist im Individualfall denkbar. Auf der Ebene der Gesamtheit der Befragten stellt dies jedoch eher als Ausnahme dar: Es werden mehr Fragebögen zurückgesandt als Postkarten, in denen eine Zusendung bejaht wird (Reuband 1998, 2006). Offenbar vergessen einige der Befragten die Rücksendung der Postkarte. Oder sie wollen aus Gründen erhöhter Anonymität sichergehen, dass nicht einmal der Tatbestand ihrer Teilnahme an der Befragung bekannt wird.

### 47.5.7 Teilnahmeanreize

Teilnahmeanreize (oder Incentives) können materieller und nichtmaterieller Art sein. Nichtmaterielle Incentives liegen u.a. in dem Gefühl des Befragten, an einer bedeutsamen Studie teilzunehmen, materielle Anreize sind u.a. Geldbeträge oder materielle Sachleistungen (früher z.T. in Form von Telefonkarten). Am wirksamsten haben sich Geldgeschenke erwiesen. Sie wirken in der Regel auch dann, wenn es sich um kleine Beträge handelt. Erklärt werden kann dies durch das Prinzip der Reziprozität: man erwidert das Geschenk durch Teilnahme an der Befragung. Je höher der Betrag ist, desto mehr steigt die Teilnahmebereitschaft. Dabei wirkt allerdings das Prinzip des abnehmenden Grenznutzens ein: Die Teilnahmebereitschaft steigt nicht proportional zur Höhe des Geldbetrags an, sondern flacht schließlich ab.

Am wirksamsten sind Incentives, wenn sie von Anfang an dem Fragebogen beigelegt werden. In späteren Phasen der Erinnerungsaktionen haben sie keine oder nur schwache Auswirkungen. Lotterien oder Preisausschreiben haben allenfalls marginale Effekte (Church 1993, Nakash et al. 2006). Bedacht werden sollte beim Einsatz der Incentives, dass sie nicht zur verstärkten Beteiligung der Personengruppen führt, die ohnehin gut vertreten sind. Die Incentives sind daher ggf. auf die Gruppen hin abzustimmen, die in der Umfrage unzureichend repräsentiert sind.

## Literatur

- Blasius, Jörg/Hox, Joop/Leeuw, Edith de/Schmidt, Peter (Hg.) (2002): Social science methodology in the new millennium [auf CD-Rom]. Opladen: Leske und Budrich
- Brick, J. Michael/Montaquila, Jill M./Han, Daifeng/Williams, Douglas (2012): Improving Response Rates for Spanish Speakers in Two-phase Mail Surveys, in: Public Opinion Quarterly 76: 721-732
- Church, Allan H. (1993): Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. In: Public Opinion Quarterly 57: 62-79
- DifU. Deutsches Institut für Urbanistik (2005): Die Beteiligung an kommunalen Bürgerumfragen 1970-2004. DifU-Berichte 4. Berlin: DifU
- Dillman, Don A. (2000): Mail and Internet Surveys: The Tailored Design Method. 2nd ed. New York: John Wiley
- Frankovic, Kathleen A. (2012): Straw Polls in the U. S.: Measuring Public Opinion 100 Years Ago. In: Haas, Hannes/Jerabek, Hynek/Petersen, Thomas (Hg.): 66-84
- Haas, Hannes/Jerabek, Hynek/Petersen, Thomas (2012): The Early Days of Survey Research and Their Importance Today. Wien: Wilhelm Braumüller
- Harm, Katrin/Tobias Jaeck, (2012): Bürgerumfrage Halle 2012. Der Hallesche Graureiher 2012-2. Halle: Martin-Luther-Universität
- Hippler, Hans-Jürgen (unter Mitarbeit von Kristina Seidel) (1985): Schriftliche Befragungen bei allgemeinen Bevölkerungsbefragungen. In: ZUMA Nachrichten 16: 39-56
- Kunz, Franziska (2010): Mahnaktionen in postalischen Befragungen. In: Methoden-Daten-Analysen 4: 127-155
- Lazarsfeld, Paul (1940): The Use of Mail Questionnaires to Ascertain the Relative Popularity of Network Stations in Family Listening Surveys. In: Journal of Applied Psychology 24: 803-816
- Messer, Benjamin L./Dillman, Don A. (2011): Surveying the General Public over the Internet using Address-based Sampling and Mail Contact Procedures. In: Public Opinion Quarterly 75: 429-457
- Nakash, Rachel A./Hutton, Jane L./Jorstad-Stein, Ellen C./Gates, Simon/Lamb, Sarah E. (2006): Maximising response to postal questionnaires- A systematic review of randomised trials in health research. In: BMC Medical Research Methodology, 6: 5
- Petermann, Sören (2005): Rücklauf und systematische Verzerrungen bei postalischen Befragungen. Eine Analyse der Bürgerumfrage Halle 2003. In: ZUMA-Nachrichten 57: 56-78
- Reuband, Karl-Heinz (1998): Panelmortalität in postalischen Erhebungen und soziale Zusammensetzung der Befragten. In: Planung und Analyse 25 (3/98): 116-21
- Reuband, Karl-Heinz (1999a): Anonyme und nicht-anonyme postalische Bevölkerungsbefragungen. In: Planung und Analyse 26 (1/99): 56-58
- Reuband, Karl-Heinz (1999b): Postalische Befragungen in den neuen Bundesländern. In: ZA Information 45: 71-99

- Reuband, Karl-Heinz (2001): Möglichkeiten und Probleme des Einsatzes postalischer Befragungen. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 53: 338-364
- Reuband, Karl-Heinz (2002): Mail and telephone surveys. In: Blasius, Jörg/Hox, Joop/ Leeuw, Edith de/Schmidt, Peter (Hg.): P021403
- Reuband, Karl-Heinz (2006): Postalische Befragungen alter Menschen. In: ZA Information 59: 100-127
- Reuband, Karl-Heinz (2011): Rücklaufquoten und Repräsentativität. In: Stadtforschung und Statistik. Zeitschrift des Verbandes deutscher Städtestatistiker (1/11): 44-50
- Reuband, Karl-Heinz/Blasius, Jörg (1996): Face-to-face, telefonische und postalische Befragungen. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 48: 296-318
- Tourangeau, Roger/Rips, Lance J./Rasinski, Kenneth (2000): The Psychology of Survey Response. Cambridge

Pia Wagner und Linda Hering

## 48.1 Ursprung und Verbreitung der Online-Befragung

Online-Befragungen platzieren ihre Fragebögen auf dem Server eines Forschungsinstituts oder eines Providers, wo sie von den Befragungsteilnehmern online ausgefüllt werden. Von anderen Befragungsmodi der standardisierten Befragung (Reinecke, Kapitel 44 in diesem Band) unterscheiden sich Online-Befragungen also durch das von ihnen genutzte Medium: das Internet.

Innerhalb eines Jahrzehnts haben Verfahren der internetgestützten Datenerhebung die empirische Markt- und Sozialforschung in Deutschland nachhaltig verändert: Machten Online-Befragungen nach Angaben der ADM-Mitgliedsinstitute um die Jahrtausendwende gerade einmal 3% aller standardisierten Befragungen aus, wurden im Jahr 2011 mit 36% die meisten Erhebungen online durchgeführt (34% telefonisch, 24% persönlich, 6% schriftlich; ADM 2012: 17-18). Damit reagiert die Markt- und Sozialforschung auf die immer umfassender werdende Versorgung der deutschen Haushalte mit Internetanschlüssen. 76,5% der deutschen Bevölkerung ab 14 Jahren nutzten 2013 das Internet, was etwa 53,7 Millionen Menschen entspricht, die potentiell online erreichbar sind (Initiative-D21 2013: 18). Natürlich folgt daraus im Umkehrschluss, dass 23,5% der Deutschen online *nicht* erreichbar sind, was die Möglichkeiten der internetgestützten Datenerhebung deutlich einschränkt.

Technisch möglich ist die Datenerhebung im Internet (genauer: im World Wide Web) seit der Veröffentlichung von HTML 2.0 im Jahr 1994. Mithilfe dieser auszeichnungsisierten Skriptsprache ließen sich die ersten Fragebogenformulare online stellen (Roessing 2009). Einen Impuls zur Professionalisierung der Online-Forschung im deutschsprachigen Raum lieferte die 1996 gegründete German Internet Research List (gir-l), eine Mailingliste für methodische und praktische Fragen rund um die Datenerhebung mit dem Internet. Aus dem lebendigen Austausch ging die Konferenzserie „General Online Research“ (GOR) hervor, die sich inzwischen zu einer der weltweit wichtigsten Fachkonferenzen zum Thema

Online-Forschung entwickelt hat. Methodisches Wissen und forschungspraktische Informationen bündelt die Deutsche Gesellschaft für Online-Forschung (DGOF, [www.dgof.de](http://www.dgof.de)) mittlerweile auch in einer eigenen Buchreihe „Neue Schriften zur Online-Forschung“. Dem Massenmarkt steht das Instrument der Online-Befragung zur Verfügung, seit Google 2008 für jedermann kostenlos einen Fragebogen-Generator mit serverbasierter Antwortspeicherung (heute eine Funktionalität von „Google Drive“) anbietet (Welker/Matzat 2009: 33-34).

Ein großes Problem des Erhebungsverfahrens begründet sich allerdings aus diesen Entwicklungen: Da nun jedermann ohne Aufwand und Kosten Fragebögen erstellen und einem Massenpublikum zugänglich machen kann, kursieren im Internet eine Menge methodisch unzureichender Online-Befragungen, die von geringem wissenschaftlichen Wert und nach Ansicht einiger Autoren sogar feldschädigend sind (Lee et al. 2008, Couper/Coutts 2006). Denn die Güte (Krebs/Menold, Kapitel 30 in diesem Band) empirischer Erhebungsverfahren steht und fällt mit der Teilnahmebereitschaft ihrer Befragungspersonen (Engel/Schmidt, Kapitel 23 in diesem Band), der das Dauerfeuer von Teilnahme-Anfragen jedoch erheblich schadet. Als Konsequenz daraus stehen Forscher vor der Herausforderung, bei der Konzeption und Durchführung von Online-Befragungen besonders sorgfältig vorzugehen.

---

## 48.2 Stärken und Schwächen gegenüber anderen Befragungsmodi

Ob eine Online-Befragung für die Datenerhebung eines Projektes die geeignete Wahl ist, muss anhand ihrer Vor- und Nachteile gegenüber alternativen Befragungsmodi abgewogen werden. Auf den ersten Blick weisen Online-Befragungen im Vergleich mit telefonischen (Hüfken, Kapitel 46 in diesem Band), persönlich-mündlichen (Stocké, Kapitel 45 in diesem Band) und schriftlich-postalischen Befragungen (Reuband, Kapitel 47 in diesem Band) eine ganze Reihe von Vorteilen auf: Allem voran sind Online-Befragungen *zeitlich und räumlich unabhängig*. Das heißt alle Befragungspersonen können über große Entfernung und Grenzen hinweg simultan kontaktiert werden. Dies kann den unerwünschten Methodeneffekten unterschiedlicher Befragungszeitpunkte entgegenwirken (Blasius/Brandt 2009: 158). Online-Befragungen sind des Weiteren besonders offen für *neue, grafisch anspruchsvolle Instrumente*. Hier kamen etwa zuerst interaktive Fragebogenelemente wie Schieberegler für die stufenlose Eingabe von Antwortwerten oder Drag-and-Drop-Module zum Sortieren von Elementen zum Einsatz. Auch *multimediale Inhalte* wie Bilder, Audioelemente oder Videos sind in Online-Befragungen ohne großen Aufwand integrierbar. Sie können die Teilnehmer motivieren und dabei helfen, Missverständnisse zu vermeiden (Schnell et al. 2005: 377).

Anders als bei telefonischen und persönlich-mündlichen Erhebungen werden bei Online-Befragungen *keine Interviewer* eingesetzt. Mit den Interviewern entfallen auch Interviewer-Effekte (Glantz/Michael, Kapitel 21 in diesem Band) und Effekte sozialer Erwünschtheit (Hlawatsch/Krickl, Kapitel 20 in diesem Band). Gleichzeitig bieten

Online-Befragungen dieselben technischen Möglichkeiten wie computergestützte telefonische und persönliche Erhebungsmethoden. Insbesondere die *automatische Filterführung*, aber auch automatische Checks auf Item-Nonresponse und Plausibilität erleichtern den Teilnehmern das Ausfüllen ihres Fragebogens und erhöhen die Datenqualität (Lozar Manfreda/Vehovar 2008: 264). Im Gegensatz zu schriftlich-postalischen und anderen nicht computergestützten Erhebungsmethoden drohen des Weiteren *keine Fehler durch manuelle Datenerfassung*, denn die online eingegebenen Daten werden direkt auf einem Server gesammelt. Das bedeutet auch, dass die Daten auf dem Server sofort verfügbar sind und – beispielsweise zur Vorbereitung der Datenbereinigung oder für Zwischenberichte – noch vor Befragsende abgerufen werden können.

Als ausschlaggebender Vorteil gegenüber allen anderen Erhebungsmethoden wird oft der *geringe finanzielle Aufwand* von Online-Befragungen angeführt. Je nach Ausgestaltung der Teilnehmerverwaltung (postalisch, per E-Mail oder mixed mode) fallen tatsächlich geringere bis gar keine Kosten für Einladungen, Fragebogenversand und Reminder an. Auch die Kosten für Interviewer und manuelle Dateneingabe entfallen. In dieser Rechnung fehlen allerdings die Kosten für das technische Equipment (Hard- und Software), für Webdesign, Pretests usw. Der Kostenvorteil von Online-Befragungen realisiert sich also erst ab einer gewissen Anzahl ausgefüllter Fragebögen (Lozar Manfreda/Vehovar 2008: 282).

Der offensichtlichste *Nachteil* von Online-Befragungen gegenüber anderen Befragungsmodi ist, dass ihre Reichweite von der *Ausstattung der Zielpopulation mit Computer und Internet* abhängt. Es stellt sich stets die Frage der Repräsentativität: Ist die Zielgruppe überhaupt online erreichbar? Vor besonderen Hürden stehen Online-Befragungen auch wenn es darum geht, die erreichbare Population zur Teilnahme zu motivieren. Denn häufig vermuten die von massenhaft versandten E-Mails (unsolicited commercial e-mails) belästigten Internetnutzer auch hinter ernstgemeinten Teilnahmeanfragen kommerzielle Interessen (*UCE-Problematik*; Hauptmanns 1999: 23). Darauf deutet zumindest eine in der Regel *geringe Ausschöpfungsquote* hin (Maurer/Jandura 2009: 66). Daher sind ein seriöses Auftreten, informative Ankündigungsschreiben sowie motivierende Anschreiben und Reminder für die Durchführung einer Online-Befragung unerlässlich.

Stärker als bei anderen Erhebungsmethoden kommt bei Online-Befragungen außerdem ein Problem inhaltlicher Art zum Tragen: Verschiedene Studien deuten darauf hin, dass die Online-Umgebung die Bindung der Befragungsteilnehmer an soziale Normen verringert, weshalb deren individuelle Motive beim Ausfüllen eines Fragebogens gegenüber sozialen Aspekten in den Vordergrund treten. Während die im Internet wahrgeommene Anonymität also einerseits offenere Antworten erwarten lässt, stellt die *soziale Entkontextualisierung* gleichzeitig die Übertragbarkeit der erhobenen Antworten auf das soziale Alltagshandeln infrage (Taddicken 2009). Unter Berücksichtigung dieser Probleme ist zusammenfassend festzuhalten, dass Online-Befragungen immer dann eine sinnvolle Option darstellen, wenn ihr anvisierter Teilnehmerkreis umfangreich ist und über einen Internetzugang verfügt.

## 48.3 Stichprobenziehung bei Online-Befragungen I: Abdeckungsbezogene Probleme

Um repräsentative Befragungsergebnisse zu erhalten, ist die Ziehung von Zufallsstichproben (Häder/Häder, Kapitel 18 in diesem Band) aus der anvisierten Grundgesamtheit unabdingbar. Die Frage, inwieweit diese sogenannte *Zielgesamtheit* tatsächlich online erreichbar ist, spielt deshalb bei der Konzeption von Online-Befragungen eine wesentliche Rolle. Wen kann man also mit Online-Befragungen erreichen? Den Ergebnissen einiger Studien über Internetnutzer zufolge unterscheiden sie sich v.a. in den folgenden soziodemografischen Merkmalen von der Gesamtbevölkerung (Initiative-D21 2013: 22-23; ähnliche Werte berichten Blasius/Brandt 2009: 160, Lee et al. 2008: 11 sowie Couper/Coutts 2006: 230):

- *Alter*: Onliner sind in der Regel jünger als der Durchschnitt der Gesamtbevölkerung. So sind über 90% der 14- bis 40-Jährigen online, während von den Personen ab 70 nur noch 30% im Internet aktiv sind.
- *Geschlecht*: In Deutschland dominieren noch immer Männer die digitale Welt. 81% von ihnen sind online, von den Frauen hingegen nur 72%.
- *Einkommen*: Personen mit höherem Einkommen zählen deutlich häufiger zur Gruppe der Internetnutzer als andere. Haushalte mit einem Monatsnettoeinkommen ab 3.000 Euro nutzen das Internet zu 93%, von den Haushalten mit weniger als 1.000 Euro netto pro Monat nur 55%.
- *Bildung*: Die Gruppe der Internetnutzer setzt sich eher aus Personen zusammen, die einen höheren Bildungsabschluss besitzen. Schüler führen die Statistik mit 98% Onlinern an, gefolgt von Personen mit abgeschlossenem Studium (91%). Das Schlusslicht bilden die Volks- und Hauptschulabsolventen mit 61%.

Diese „digitale Kluft“ besteht auch weiterhin in fast allen hochtechnisierten Ländern. Darüber hinaus belegen Studien, dass selbst Onliner und Offliner, die einander in den oben genannten Merkmalen gleichen, hinsichtlich weiterer Merkmale wie *Religiosität*, *politischer Motivation* oder *Haushaltsgröße* Unterschiede aufweisen (z.B. Blasius/Brandt 2009).

Noch beträchtlicher fallen die Unterschiede im Internetzugang aus, wenn *räumlich kleinteilige Einheiten* wie Landkreise oder Städte als Untersuchungseinheiten betrachtet werden (vgl. die Bundesländeraufschlüsselung der Initiative-D21 2013: 21). Aus internationaler Perspektive klafft der *Graben zwischen den wohlhabenden und den armen Gesellschaften* der Welt auch im Hinblick auf die Internetnutzung (vgl. Chen/Wellman 2004). Kulturvergleichende Studien (Braun, Kapitel 56 in diesem Band) müssen sich besonders achtsam mit den länderspezifischen Unterschieden bezüglich des Internetzugangs auseinandersetzen – und das nicht nur über Ländergrenzen hinweg, sondern auch innerhalb einer Nation.

Allgemeine Bevölkerungsumfragen sind folglich zum jetzigen Zeitpunkt nicht realisierbar. Denn die erreichbare Population – die sogenannte *Auswahlgesamtheit* – beschränkt sich aufgrund der „digitalen Kluft“ auf die aktiven Internetnutzer, die in wesentlichen

Merkmale systematisch von den Offlinern unterschieden sind. Stichproben aus der Auswahlgesamtheit der Internetnutzer können deshalb keine Repräsentativität für die Zielgesamtheit der Gesamtbevölkerung beanspruchen. Die Differenz zwischen Zielgesamtheit (der anvisierten Population) und Auswahlgesamtheit (der online erreichbaren Population) bezeichnet man als *abdeckungsbezogenes Problem* der Stichprobenziehung.

Anders verhält es sich mit Erhebungen, die auf spezifische Internet-affine Teilstichproben abzielen (z.B. die *Mitarbeiter einer Organisation* oder die *Studierenden einer Hochschule*). Hier stimmen Ziel- und Auswahlgesamtheit weitgehend überein, so dass die abdeckungsbezogenen Probleme in den Hintergrund treten. Die erforderlichen E-Mail-Adresslisten können von den zuständigen Personalabteilungen bzw. Verwaltungen eingeholt oder durch gründliche Recherche selbst erstellt werden. Aus den vollständigen Adresslisten können schließlich Zufallsstichproben gezogen werden. Auch Internet-Foren führen solche Mitgliederlisten, wobei jedoch von Doppelanmeldungen und Karteileichen auszugehen ist, die wiederum *Stichprobenfehler* zur Folge haben können (Engel/Schmidt, Faulbaum, Kapitel 23 und 31 in diesem Band).

---

## 48.4 Varianten von Online-Befragungen

Es gibt zahlreiche Spielarten von Online-Befragungen, die sich entlang des Kriteriums der Zufallsstichprobe unterscheiden lassen (vgl. Lozar Manfreda/Vehovar 2008: 265-267; Couper/Coutts 2006: 229-238). Besonders häufig trifft man im Internet sogenannte *Convenience Samples* an, etwa wenn auf Bannern oder in Pop-Ups zur „Frage des Tages“ eingeladen wird. Solche Befragungen sind uneingeschränkt selbstselektierend, in keiner Weise repräsentativ und dienen allenfalls zu Unterhaltungszwecken. Ebenso wenig können *Befragungen mit selbstrekrutierten Freiwilligen-Panels* Repräsentativität für sich in Anspruch nehmen. Zwar wählen Markt- und Meinungsforschungsinstitute die Zielpersonen aus ihren Freiwilligen-Panels (sogenannten Access-Panels) mitunter nach dem Zufallsprinzip aus. Die Zusammensetzung der Panels jedoch erfolgt durch offene Einladungen und ist daher hoch selektiv. In beiden Fällen sollten Verallgemeinerungen von der Stichprobe auf die Zielpopulation kritisch hinterfragt werden.

Demgegenüber sind die folgenden Varianten der Online-Befragung durchaus in der Lage, Zufallsstichproben aus ihrer jeweiligen Zielpopulation zu ziehen: *Intercept-Befragungen* richten sich gezielt an die Besucher einer Webseite, hauptsächlich um Kundenzufriedenheit zu erfragen oder um die Webseite selbst bewerten zu lassen. Jeder n-te Besucher wird um sein Urteil gebeten, wobei natürlich mit hohen Nonresponse-Quoten zu rechnen ist. So muss auch hier hinterfragt werden, ob die Teilnahmeverweigerung nicht systematisch mit der Haltung gegenüber dem Befragungsgegenstand zusammenhängt. Bei *Befragungen mit listenbasierten Stichproben aus Populationen mit einem hohen Abdeckungsgrad* ist es darüber hinaus möglich, Teilnahmeverweigerer zu identifizieren und Ausfallgründe zu formulieren. Ausgangspunkt ist eine Liste von Zielpersonen mit Internetzugang (wie oben beschrieben etwa die Mitarbeiter einer Organisation oder die Studierenden einer

Hochschule). Diese werden per E-Mail oder Post zur Teilnahme eingeladen und meist mit einem Passwort ausgestattet. Passwort und personalisierte Anrede führen zu einer deutlich höheren Beteiligung als beispielsweise Intercept-Befragungen. Davon profitieren auch *Befragungen mit einem vorrekrutierten Panel von Internetnutzern*. Hier werden die Panelmitglieder mithilfe eines zufallsgesteuerten Auswahlverfahrens bestimmt (z.B. einer RDD-Telefonbefragung), so dass sie eine Zufallsstichprobe aus Internetnutzern bilden, aus der dann zum jeweiligen Befragungszweck wiederum Zufallsstichproben gezogen werden können.

Während die Schlussfolgerungen aus listenbasierten Stichproben und vorrekrutierten Panels keine Geltung jenseits ihrer Zielpopulation (z.B. der befragten Studierenden bzw. der Internetnutzer allgemein) besitzen, streben zwei weitere zufallsbasierte Varianten der Online-Befragung nach der Verallgemeinerbarkeit auf die Gesamtbevölkerung. Dies versuchen zum einen *Mixed-Mode-Befragungen mit Wahl der Teilaahmemethode*. Wenn die zufällig ausgewählten Zielpersonen die Wahl haben, so die Annahme, eine Befragung online oder etwa schriftlich-postalisch auszufüllen, entfallen die systematischen Verzerrungen, die sonst mit dem Kriterium Internetzugang einhergehen. Allerdings verursacht der Einsatz mehrerer Befragungsmodi wiederum vielfache Kosten für die parallele Datenerhebung und -bereinigung, nicht zu sprechen von den Auswirkungen unterschiedlicher Datenqualitäten und von Methodeneffekten. Als die einzige Möglichkeit, eine Zufallsstichprobe aus der Gesamtbevölkerung im Online-Modus zu befragen, verbleibt schließlich die *Ausstattung eines vorrekrutierten Panels mit der erforderlichen Technik*. Wegen seiner enormen Kosten wird dieses Vorgehen aber nur sehr selten gewählt. Derzeit wird es u.a. von der holländischen Organisation CentERdata (Universität Tilburg) und dem US-Unternehmen Knowledge Networks praktiziert. Nichtsdestotrotz stellen gerade für kleinere Projekte, die ein methodisch sauberes Vorgehen bei vertretbarem Aufwand anstreben, Befragungen listenbasierter Stichproben den Königsweg dar.

---

## 48.5 Stichprobenziehung bei Online-Befragungen II: Nonresponse und Rücklaufquote

Eine der am häufigsten beschriebenen Ursachen für Stichprobenfehler, nicht nur bei Convenience Samples und Access Panels, sondern bei allen Arten von Online-Befragungen ist die Problematik der *Selbstrekrutierung*: Bis die Befragten zur eigentlichen Umfrage gelangen, müssen sie bereits mehrere Hürden überwinden (Couper/Coutts 2006: 225). Sie müssen erstens auf der Startseite des Fragebogens ankommen, was zumindest bei Convenience Samples und Intercept-Befragungen von der Intensität der Internetnutzung abhängt und bei den anderen Online-Varianten voraussetzt, dass die Befragten dem Link zur Befragung folgen (können). Zweitens müssen sie sich aktiv zur Teilnahme entscheiden und drittens je nach Art der Internetverbindung dafür Kosten in Kauf nehmen (z.B. Providergebühren, die jedoch mit der Verbreitung von Flatrates zunehmend an Bedeutung verlieren). Scheitert die Teilnahme eines Befragten an diesen Hürden, spricht man von *Unit-Nonresponse*

bzw. einem *Nonrespondent*. Brechen Teilnehmer den Fragebogen aus mangelndem Interesse oder aus anderen Gründen während des Ausfüllens ab, nennt man sie *Drop-Outs*. Einzelne nicht beantwortete Fragen schließlich bezeichnet man als *Item-Nonresponse*.

Die Untersuchung von Nonresponse und Ausfallgründen spielt im Forschungsfeld der Online-Befragung eine genauso große Rolle wie die Steigerung der Teilnahmebereitschaft. Nur wenn die Ausschöpfungsquote einer Studie offengelegt und ihre realisierte Stichprobe auf systematische Verzerrungen geprüft wird, sind nämlich Aussagen zur Geltungskraft ihrer Ergebnisse möglich. Als Determinanten der Teilnahmebereitschaft haben sich inzwischen vor allem das Interesse der Befragten an der Thematik, aber auch zeitliche Aspekte erwiesen – idealerweise sollte eine Umfrage nicht länger als 10 bis 15 Minuten dauern. Incentives werden demgegenüber eine geringere Bedeutung zugeschrieben (Batinic/Bosnjak 2000: 307, Gräf/Heidingsfelder 1999: 123-125).

Bei anonym durchgeführten Befragungen – für alle Teilnehmer gibt es nur einen Universal-Link zur Umfrage – ist die Bestimmung der Ausschöpfungs- und Nonresponse-Quoten nur mit Einschränkungen bzw. gar nicht möglich. Der Forscher steht vor dem Problem, eine Quote zu berechnen, die beispielsweise über das Logfile einer Webseite eruiert werden muss, ohne Mehrfachbesucher doppelt zu zählen. „Auswärtige Verweigungsquoten“ von Personen, die über eine externe Webseite auf die Umfrage aufmerksam werden, jedoch nicht teilnehmen, können gar nicht erfasst werden.

Bei personalisierten Umfragen – jeder Teilnehmer wird persönlich angeschrieben und erhält ggf. ein Zugangspasswort – ist die Berechnung einer Ausschöpfungsquote hingegen möglich. Die gängigen Online-Befragungstools bieten i.d.R. integrierte Rücklaufstatistiken an, die detaillierte Informationen zur Nettostichprobe, zur mittleren Bearbeitungszeit, zu den Seiten mit den meisten Abbrüchen und zur Quotenstatistik ausweisen. Diese detaillierten Daten erlauben es den Forschern, noch genauer als bei schriftlich-postalischen Befragungen dem Zeitpunkt von Befragungsabbrüchen nachzugehen, und tragen so zur Qualitätssteigerung künftiger Studien bei.

---

## 48.6 Zur Gestaltung von Online-Fragebögen

Da die Qualität von Online-Befragungen besonders von der Teilnahmebereitschaft ihrer Befragten abhängt, stehen Online-Forscher vor der Herausforderung, ihren Fragebogen so zu gestalten, dass er sowohl die Ernsthaftigkeit der Forschung vermittelt als auch zur Teilnahme motiviert. Orientierung hierfür bieten einerseits die für alle Befragungs-Modi geltenden *Regeln der Fragebogenkonstruktion*. Andererseits muss das Fragebogendesign auch den *Besonderheiten des Mediums Internet* genügen. Detaillierte Hinweise zur guten Gestaltung von Online-Befragungen geben u.a. Lozar Manfreda/Vehovar (2008) und Dillman (2007).

Je nach Länge des Fragebogens ist zu entscheiden, ob die Fragen auf mehrere Seiten verteilt oder auf einer Seite platziert werden sollen. Bei einer aus wenigen Fragen bestehenden Kurzumfrage kann es für die Befragten durchaus angenehmer sein, nur nach unten

scrollen zu müssen. Dies birgt jedoch die Gefahr, dass sie einzelne Fragen übersehen. Aus diesem Grund ist es i.d.R. von Vorteil, nur eine Frage bzw. Fragenbatterie pro Seite vorzulegen. Um die Befragten dennoch nicht im Ungewissen über die Gesamtlänge des Fragebogens zu lassen, sollte dann eine *Fortschrittsanzeige* eingeblendet werden.

Zu einem wohldesignierten und nutzerfreundlichen Online-Fragebogen gehört standardmäßig die *automatische Filterführung*. Aber auch der Einsatz vordefinierter oder während der Umfrage generierter Platzhalter sowie der Einsatz von Schleifen (Loops) erleichtern den Befragten das Antworten, personalisieren die Umfrage und vermindern dadurch Abbrüche. Da die Daten von den Befragten selbstständig eingetragen werden und kein Interviewer als kontrollierende Instanz zwischengeschaltet ist, können zudem Vollständigkeits- oder Plausibilitätschecks sinnvoll sein. Diese sollten allerdings sparsam eingesetzt werden, da immer wiederkehrende Warnhinweise die Befragten irritieren und dadurch Abbrüche provozieren könnten.

Gerade bei den technisch raffinierteren Features ist jedoch stets zu berücksichtigen, dass die Befragten unterschiedlichste technische Voraussetzungen mitbringen. *Kompatibilitäts- und Darstellungsprobleme* führen in der Regel zum sofortigen Abbruch einer Befragung. Umfassende Pretests auf verschiedene Hard- und Software sollten daher bei jeder Studie miteingeplant werden.

---

## 48.7 Export, Aufbereitung und Dokumentation von Online-Befragungsdaten

Ein Vorteil von Online-Befragungen ist, dass die Angaben der Teilnehmer direkt auf dem Befragungsserver gesammelt werden und dem Forscher dort unmittelbar zur Verfügung stehen. An einer umfassenden *Datenbereinigung* führt jedoch trotzdem kein Weg vorbei, zum Beispiel da Befragte unzutreffende Filterwege wählen oder inkonsistente Angaben machen.

Einige Befragungs-Programme sind in der Lage, Statistiken auf dem Niveau einfacher Häufigkeitsverteilungen auszugeben, die erste Hinweise auf ein auffälliges Antwortverhalten liefern können. Für eine systematische Datenbereinigung wird man jedoch nicht umhin kommen, die Daten in ein statistisches Datenverarbeitungs-Programm zu exportieren. Standardmäßig ist der Datenexport in den gängigen Formaten üblich – wie CSV, ASCII, Excel, SPSS oder STATA. Mittels Exportfilter können diese Ausgaben bei Bedarf auf bestimmte Fragen- oder Befragtengruppen eingegrenzt werden.

Ein weiteres nützliches Feature von Online-Befragungen ist ihre unkomplizierte *Dokumentation*. Variablen- und Wertelisten entstehen quasi „automatisch“ parallel zur Fragebogenkonstruktion. Druckversionen der verwendeten Fragebögen, gesammelte Pretest-Kommentare und vieles andere stehen der eigenen Projektdokumentation jederzeit zur Verfügung.

## 48.8 Software zur praktischen Umsetzung

Wer eine eigene Online-Umfrage durchführen möchte, kann im Internet aus einer Vielzahl angebotener Software-Lösungen auswählen. Unter [www.WebSM.org](http://www.WebSM.org) und [www.gesis.org](http://www.gesis.org) sind Aufzählungen von Software-Anbietern und -Paketen zu finden; einen Überblick über die im Einzelnen angebotenen Dienstleistungen geben Vehovar et al. (2012). Welche Software allerdings die richtige ist, hängt ganz von den Bedürfnissen eines Projekts ab: der Anzahl geplanter Befragungen, dem verfügbaren Budget, den Vorkenntnissen der Projektbeteiligten und den gewünschten Funktionalitäten der Software.

Zu Beginn ist die Entscheidung zu treffen, ob die Dienste eines kommerziellen Software-Anbieters in Anspruch genommen werden oder ob eine freie Software zum Einsatz kommen soll. Freie Software mit offenem Quellcode (*open source*) kann kostenlos aus dem Internet heruntergeladen und im Rahmen der Lizenzbestimmungen individuell angepasst werden. Alle Aspekte der Durchführung und Administration liegen damit in der Verantwortung des Nutzers, nicht zuletzt die Einhaltung der Datenschutzbestimmungen. Daher ist der Einsatz freier Software nur Personen mit fundierten technischen Kenntnissen und einer geeigneten Ausstattung (u.a. Web- und Datenbankserver) zu empfehlen (Roessing 2009: 52–53).

*Kommerzielle Software-Anbieter* übernehmen gegen eine Lizenzgebühr das Hosting einer Befragung sowie in Abhängigkeit vom gebuchten Paket auch die Teilnehmerverwaltung und das Datenmanagement. Auf Einsteiger mit geringem Budget sind kostengünstige Software-Pakete zugeschnitten, die auf eine Umfrage oder mehrere einfache Umfragen limitiert sind. Zum nicht-kommerziellen Einsatz der Software an Universitäten oder in Abschlussarbeiten stehen meist Campus- oder Studenten-Lizenzen zur Verfügung. Für Großprojekte mit mehr als 1000 Teilnehmern pro Umfrage ebenso wie für anspruchsvolle Nutzer, die spezielle Anforderungen an die Software stellen (etwa Kompatibilität mit mobilen Geräten, Mehrsprachigkeit oder Paneldesigns) reichen diese Lizenzen nicht aus. An sie richten sich fortgeschrittene Software-Angebote bis hin zu individuell maßgeschneiderten und entsprechend teuren Versionen.

Unabhängig vom Modus der Administration und vom Grad ihrer Komplexität sollte eine Umfrage-Software mindestens folgende Standards erfüllen (Lozar Manfreda/Vehovar 2008: 281–282):

- *Teilnehmerverwaltung* mit Serienbrief-Option für Ankündigungs-, Einladungs- und Erinnerungsschreiben;
- *nutzerfreundliches Interface* bei der Fragebogenprogrammierung einschließlich diverser Hilfsfunktionen (Handbuch, Onlinehilfe, Tutorials, Fragen-Bibliotheken usw.);
- *flexible Fragebogengestaltung* hinsichtlich Layout (Hintergrund, Schrifttypen, Multimedia-Elemente, Fortschrittsbalken), Fragetypen (offene und geschlossene Fragen, semantische Differenziale, Mehrfachnennungen) und computerunterstützter Fragerführung (automatische Filterführung, Loops, Randomisierung, Plausibilitäts-Checks etc.);
- zuverlässige und sichere Daten-Verwaltung und -Übermittlung.

Perspektivisch wird die Integration anderer Befragungsmodi in die Online-Software eine immer prominentere Rolle einnehmen: Ein einmal programmiert Online-Fragebogen soll künftig automatisch in andere Befragungsformen, etwa einen Papierfragebogen oder Vorlagen für CATI- oder CAPI-Befragungen, konvertiert werden können (Couper/Miller 2008: 835).

---

## 48.9 Datenschutz

Für Online-Befragungen gelten dieselben datenschutzrechtlichen Bestimmungen wie für andere schriftliche Befragungsmodi. Aktuelle Diskussionen zu Datensicherheit und Privatsphäre im Internet geben allerdings Anlass, allein schon aus Gründen der Vertrauensbildung besonders sensibel mit den Daten der Befragten umzugehen. Mehrere professionelle Markt- und Sozialforschungsinstitute haben Richtlinien zur Internetforschung formuliert, die dabei Orientierung geben können (ADM 2001, ESOMAR 2011).

Das Risiko eines nicht autorisierten Zugriffs durch Dritte ist am höchsten, solange Daten auf einem Server gespeichert und während sie elektronisch übermittelt werden. Alle Server müssen daher durch aktuellste Sicherheitstechnologien gegen Angriffe geschützt sein. Grundsätzlich sollten Befragungsdaten nur temporär auf Servern gespeichert und so bald wie möglich auf lokale, am besten nicht mit dem Internet verbundene Speichermedien transferiert werden. Vom Datenversand via E-Mail ist abzuraten und stattdessen der persönlichen Übergabe oder dem postalischen Versand einer Daten-CD der Vorzug zu geben. Vor ihrem Versand empfiehlt es sich auf jeden Fall, die Daten mit einem Verschlüsselungs-Programm (z.B. TrueCrypt) zu schützen.

Dies gilt umso mehr, da Anonymität bei vielen Online-Befragungen weder möglich noch erwünscht ist: Um den Zugang zu einer Befragung zu regulieren und Mehrfachteilnahmen auszuschließen, müssen Forscher die Identität ihrer Teilnehmer kennen. Nur dann lassen sich auch Teilnahmeverweigerer gezielt erinnern oder im Falle von Panels Ergebnisse über die Zeit vergleichen. Die Identifizierung der Teilnehmer erfolgt in der Regel über ein vorab zugewiesenes Passwort oder – beim automatischen Login – über eine individuelle URL. Dass die beteiligten Forscher diese Informationen nicht dazu missbrauchen, um personalisierte Daten zu veröffentlichen, ist Bestandteil jeder Datenschutzerklärung. Um Missbrauch vorzubeugen, werden Adressdaten von allen übrigen Angaben getrennt abgespeichert, so dass sich nur noch über die Identifikationsnummer eine Verbindung herstellen lässt. Sofern sich die Teilnehmer nicht ausdrücklich anders erklärt haben, sind nach Erfüllung des angegebenen Forschungszwecks alle Adressdaten zu löschen.

## 48.10 Ausblick

Abschließend bleibt mit Lee et al. (2008: 18f.) festzuhalten, dass sich das Internet als Technologie noch in seinem Anfangsstadium befindet. Und auch seine methodologische Betrachtung als ein Instrument der empirischen Wissenschaften steht erst am Anfang. Mit Web 2.0 und den rasant wachsenden Social Media werden die Möglichkeiten der Interaktion mit den Befragungsteilnehmern immer komplexer. Mobile Web-Anwendungen erschaffen neue Spielräume und lassen den Unterschied zwischen Online- und Telefonbefragungen immer weiter schmelzen (Couper/Miller 2008: 835). Solange sich das Internet wandelt und weiterentwickelt, werden sich auch die Bedingungen und Möglichkeiten der Forschungspraxis ändern.

## Literatur

- ADM (2001): Standards zur Qualitätssicherung für Online-Befragungen. Online-Ressource (01.07.2013): [http://www.adm-ev.de/fileadmin/user\\_upload/PDFS/Online-standards\\_D.PDF](http://www.adm-ev.de/fileadmin/user_upload/PDFS/Online-standards_D.PDF)
- ADM (2012): Jahresbericht 2011. Online-Ressource (01.07.2013): <https://www.adm-ev.de/index.php?id=jahresberichte>
- Batinic, Bernad (Hg.) (2001): Internet für Psychologen. Göttingen et al.: Hogrefe
- Batinic, Bernad/Bosnjak, Michael (2000): Fragebogenuntersuchungen im Internet. In: Batinic (2001): 287-317
- Blasius, Jörg/Brandt, Maurice (2009): Repräsentativität in Online-Befragungen. In: Weichbold et al. (Hg): 157-177
- Chen, Wenhong/Wellman, Barry (2004): The Global Digital Divide – Within and Between Countries. In: IT & Society 7: 39-45
- Couper, Mick P./Coutts, Elisabeth (2006): Online-Befragung. Probleme und Chancen verschiedener Arten von Online-Erhebungen. In: Diekmann (Hg): 217-243
- Couper, Mick P./Miller, Peter V. (2008): Web Survey Methods. Introduction. In: Public Opinion Quarterly 72: 831-835
- Diekmann, Andreas (Hg.) (2006): Methoden der Sozialforschung. Wiesbaden: VS Verlag
- Dillman, Don A. (2007): Mail and Internet Surveys: The Tailored Design Method. New Jersey: Wiley
- ESOMAR (2011): ESOMAR Guideline for Online Research. Online-Ressource (01.07.2013): [http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR\\_Guideline-for-online-research.pdf](http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR_Guideline-for-online-research.pdf)
- Fielding, Nigel/Lee, Raymond M./Blank, Grant (Hg.) (2008): The Sage Handbook of Online Research Methods, Los Angeles et al.: Sage
- Gräf, Lorenz/Heidingsfelder, Martin (1999): Bessere Datenqualität bei WWW-Umfragen – Erfahrungen aus einem Methodenexperiment mit dem Internet-Rogator. In: Batinic et al. (2001): 113-126
- Hauptmanns, Peter (1999): Grenzen und Chancen von quantitativen Befragungen mit Hilfe des Internet. In: Batinic et al. (2001): 21-38
- Initiative-D21 (2013): D21-Digital-Index. Auf dem Weg in ein digitales Deutschland. Online-Ressource (01.07.2013): <http://www.initiatived21.de/wp-content/uploads/2013/04/digitalindex.pdf>
- Jackob, Nikolaus/Schoen, Harald/Zerback, Thomas (Hg.) (2009): Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung. Wiesbaden: VS Verlag
- Lee, Raymond M./Fielding, Nigel/Blank, Grant (2008): The Internet as a Research Medium: An Editorial Introduction to The Sage Handbook of Online Research Methods. In: Fielding et al. (Hg): 3-20
- Lozar Manfreda, Katja/Vehovar, Vasja (2008): Internet Surveys. In: de Leeuw et al. (Hg): 264-284

- Maurer, Marcus/Jandura, Olaf (2009): Masse statt Klasse? Einige kritische Anmerkungen zu Repräsentativität und Validität von Online-Befragungen. In: Jackob et al. (Hg): 61-73
- Roessing, Thomas (2009): Internet für Online-Forscher: Protokolle, Dienste und Kommunikationsmodi. In: Jackob et al. (Hg): 49-58
- Schnell, Rainer/Hill, Paul B./Esser, Elke (2005): Methoden der empirischen Sozialforschung. München, Wien: R. Oldenbourg Verlag
- Taddicken, Monika (2009): Methodeneffekte von Web-Befragungen: Soziale Erwünschtheit vs. Soziale Entkontextualisierung. In: Weichbold et al. (Hg): 85-104
- Vehovar, Vasja et al. (2012): Survey software features overview. A WebSM Study Online-Ressource (01.07.2013): [www.websm.org/upload/editor/1362408700WebSM\\_WSS\\_report\\_b\\_v04.pdf](http://www.websm.org/upload/editor/1362408700WebSM_WSS_report_b_v04.pdf)
- Weichbold, Martin/Bacher, Johann/Wolf, Christof (Hg.) (2009): Umfrageforschung. Herausforderungen und Grenzen. Wiesbaden: VS Verlag
- Welker, Martin/Matzat, Uwe (2009): Online-Forschung: Gegenstände, Entwicklung, Institutionalisierung und Ausdifferenzierung eines neuen Forschungszweiges. In: Jackob et al. (Hg): 33-47

Jennifer Klöckner und Jürgen Friedrichs

Der Fragebogen ist ein Instrument – ebenso wie eine Waage oder ein Mikroskop. Hätten wir unterschiedliche austarierte Waagen, so würden Gegenstände gleichen Gewichts dennoch zu unterschiedlichen Anzeigen des Gewichts führen. Hätten sie zudem verschiedene Gewichte, so könnten wir zunächst nicht entscheiden, welchen „Anteil“ an dem gemessenen Gewicht auf die unterschiedliche Tarierung der Waagen und welcher auf die unterschiedlichen Gewichte der Gegenstände zurückzuführen ist. Der Fragebogen sollte demnach so konstruiert sein, dass er Interviewereffekte (Glantz/Michael, Kapitel 21 in diesem Band) minimiert und durch die Frageformulierung (Porst, Kapitel 50 in diesem Band) auch unterschiedliche semantische Interpretationen der Fragen durch die Befragten ausschließt. Das gilt für alle Formen der Befragung: face-to-face, schriftlich, telefonisch oder online. Und ebenso, ob der Fragebogen schriftlich oder elektronisch vorliegt.

---

## 49.1 Inhalt des Fragebogens

Dem Fragebogen liegt ein Untersuchungsziel zugrunde. Das Ziel besteht darin, Sachverhalte zu beschreiben und Hypothesen zu testen. Sowohl die angestrebte Beschreibung als auch die Hypothesen enthalten Konzepte, die durch einzelne Variablen operationalisiert sind und durch Fragen erhoben (= gemessen) werden. Der Fragebogen kann mehr oder minder stark standardisiert sein. Die Spanne reicht von einem Fragebogen, der nur geschlossene Fragen enthält, die auch fest angeordnet sind, bis zu einem Fragebogen mit überwiegend offenen Fragen in nur zum Teil festgelegter Reihenfolge (Leitfaden-gestütztes Interview). Der Fragebogen ist dabei sowohl Instrument der Erhebung als auch „Aufzeichnungsmedium“ für die gesammelten Antworten und Daten (Scholl 2003: 177).

Der Fragebogen soll durch Fragen und Skalen diejenigen Variablen der Konzepte messen (Latcheva/Davidov, Kapitel 55 in diesem Band), die im theoretischen Teil, den Hypothesen, angeführt werden; sie sollten in möglichst valide Fragen operationalisiert werden

(Holm 1975: 66, Raithel 2008: 35ff., Scholl 2003: 141, 139, Porst, Kapitel 50 in diesem Band). Die verschiedenen Dimensionen der Forschungsfrage müssen daher vorab spezifiziert und in Fragen der Studie einbezogen werden (dimensionale Auflösung) (Holm 1975: 66, Porst 2009: 15, Scholl 2003: 141).

Ein Beispiel ist die Studie „Nachbarschaftseffekte: Leben in einem benachteiligten Stadtteil“. Hier wurden face-to-face Befragungen durchgeführt. Um die soziale Kontrolle in einem Wohngebiet zu erfassen, wurde in der Studie nicht nur erfragt, ob sich die Leute in der Nachbarschaft kennen und sich gegenseitig helfen, sondern auch, ob es in der Nachbarschaft Vorbilder für die dort lebenden Kinder gibt und ob das Wohngebiet ein guter Ort für das Aufwachsen von Kindern ist (Blasius et al. 2008: 175f.).

---

## 49.2 Anordnung der Fragenkomplexe

Für den Einstieg eignen sich besonders einfache Fakt- oder Einstellungsfragen, die unabhängig vom Wissen der Befragten sind (Diekmann 1997: 414, Holm 1975: 54f., Raithel 2008: 75, Scholl 2003: 143f.). Am besten ist, man beginnt mit offenen Fragen, die auch einen für die Befragten erkennbaren Bezug zu dem Thema der Studie haben, wie es im Anschreiben oder bei der ersten Kontaktaufnahme angekündigt wurde. Offene Fragen motivieren und fordern die Befragten (Züll/Menold, Kapitel 52 in diesem Band). Die ersten Fragen sollten leicht verständlich sein und keine komplexen Inhalte oder komplizierte Formulierungen enthalten. So sollte das Interesse geweckt und die Befragten dazu motiviert werden, den gesamten Fragebogen zu beantworten. Der Eindruck, dass die Befragung langweilig und schwierig zu beantworten sein wird, sollte damit vermieden werden (Porst 2009: 138f.).

Gemeinhin ist der Teil mit den sozio-demographischen Fragen (Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band) unverzichtbar. Man kann diesen Teil an den Anfang des Fragebogens stellen. Das hat jedoch den Nachteil, dass es wenig motiviert. Sinnvoller ist es, diesen Teil an das Ende zu stellen. Zudem enthält auch dieser Teil sensitive Fragen, wie z.B. das Einkommen, die zum Abbruch oder zur Verweigerung führen können.

Die Fragen sollten in Blöcke (Module) zusammengefasst werden, die den Fragebogen thematisch gliedern (Kirchhoff et al. 2003: 20; Porst 2009: 142f.). Fragen, die zum selben Thema gehören, sollten also hintereinander abgefragt werden, weil es den Befragten hilft, sich auf einen Themenblock nach dem anderen zu konzentrieren. Das Hin- und Herspringen zwischen verschiedenen Themen verwirrt. Wenn es der Forschungsfrage entspricht, ist es auch möglich, chronologisch vorzugehen (Kirchhoff et al. 2003: 20). Bezieht sich die Befragung zum Beispiel auf die aktuelle Lebenssituation oder auf Entscheidungen, die im Lebensverlauf getroffen wurden, ist es sinnvoll, zuerst die Vorgeschichte zu erfragen und wie es zu der aktuellen Situation oder Entscheidung gekommen ist.

Die zentralen Fragen und Themenkomplexe sollten in der Mitte des Fragebogens stehen. Falls die befragte Person das Interview abbricht, hat man immerhin die wichtigsten Fragen beantwortet.

Der sorgfältige Umgang mit *intimen* oder *heiklen Themen* ist besonders wichtig, weil diese nicht nur zu Verweigerung oder Abbruch der Befragung, sondern auch zu nicht beabsichtigten Nebeneffekten führen können, da die Befragten „Angst“ vor Sanktionen haben könnten (Reuband/Blasius 1996: 311, Holm 1975: 82, 119, Porst 2009: 124f.). Besonders wichtige oder schwierig zu beantwortende Fragen gehören ebenfalls in den Mittelblock eines Fragebogens. Generell sollte berücksichtigt werden: je heikler eine Frage ist, desto näher sollte sie am Ende eines Fragebogens gestellt werden. Würde diese z.B. als Einstiegsfrage gewählt werden, könnte dies zu einem schnellen Abbruch führen und damit die Verweigerungsrate der gesamten Befragung erhöhen.

Am Ende des Fragebogens sollten – wie zu Beginn – einfach zu beantwortende Faktfragen gestellt werden, da die Befragten bereits ermüdet sein könnten und dadurch die Konzentrationsleistung, aber auch die Motivation fortzufahren, abnimmt. Dazu sind Fragen nach der Standarddemographie besonders gut geeignet, da sich diese auf die eigene Person beziehen und auch nicht mit Erinnerungs- und Interpretationsleistungen verbunden sind (Scholl 2003: 170f.). Durch einen Hinweis darauf, dass die Befragung bereits am Ende ist und nun nur noch „für die Statistik“ Personendaten erfragt werden, wird suggeriert, dass die Demographie gar nicht mehr zum eigentlichen Fragebogen dazugehört (Scholl 2003: 170f.). Um Teil-Verweigerungen bei den Antworten zu vermeiden, empfiehlt es sich, hier noch einmal auf die Anonymität der Befragung hinzuweisen. In face-to-face und telefonischen Befragungen liest der/die Interviewer/in den entsprechenden Hinweis vor, in einer postalischen und in einer Online-Befragung steht er im Überleitungstext zum letzten Block.

Auf der letzten Seite des Fragebogens sollte den Teilnehmern/innen gedankt und Platz für Kommentare oder Anmerkungen geboten werden (Porst 2009: 157, Raithel 2008: 77). Sie dient der Wertschätzung der Befragten und kann zudem auf interessante Aspekte zum Thema der Studie führen. In postalischen Befragungen sollte zumindest eine halbe Seite für Kommentare reserviert sein.

---

### 49.3 Überleitungen

Fragebögen sind in verschiedene Themenkomplexe (Module) eingeteilt. Um von einem Komplex zum nächsten zu gelangen, wird ein Überleitungstext formuliert. Dieser ist wichtig, damit die Befragten sich im Fragebogen orientieren, der Logik des Fragebogens folgen sowie sich auf einen Themenkomplex einstellen können. Die Überleitungstexte sollten die Themenblöcke klar voneinander abgrenzen: „Kommen wir nun zu Fragen, die sich auf die Wahl Ihres Studienfachs beziehen“. Aber auch eine kurze Erläuterung und Beispiele, wie die folgenden Fragen zu beantworten sind, können in Überleitungstexten gegeben werden.

Überleitende Texte sollten flüssig und alltagssprachlich formuliert sein. Supervisiert man telefonische Befragungen, so stellt man fest, dass die Interviewer eben diese Texte variieren. Deshalb sollten bei face-to-face-Befragungen die Texte im Pretest (Weichbold,

Kapitel 19 in diesem Band) ausprobiert und dann mit den Interviewern einheitliche Texte formuliert und festgelegt werden.

Zur Einleitung in den Netzwerkteil des Fragenbogens der Studie „Nachbarschaftseffekte: Leben in einem benachteiligten Stadtteil“ (Blasius et al. 2008: 178) wird dieses Modul durch folgenden Text eingeleitet: „Im Folgenden möchte ich Sie nach Personen fragen, mit denen Sie in den letzten 14 Tagen in einer oder anderen Weise Kontakt hatten. Denken Sie dabei bitte an alle Personen; es können Verwandte und Nicht-Verwandte, Mitglieder Ihres Haushalts und Menschen, die nicht in Ihrem Haushalt leben, sein (...).“

Damit die Befragten gerade bei längeren Befragungen „am Ball“ bleiben, kann bei face-to-face-Interviews von den Interviewern darauf hingewiesen werden, an welcher Stelle im Fragebogen sich das Interview bereits befindet. Äußerungen wie „Nun haben wir das Schwierigste bereits geschafft ...“ oder „Zum Schluss möchte ich Ihnen nur noch folgende Frage stellen ...“ können in Überleitungstexten platziert und von den Interviewern vor-gelesen werden. Solche Anmerkungen erkennen die Leistung der Befragten an und sollen dazu motivieren, den letzten Teil nun auch noch zu Ende zu bringen (Holm 1975: 94).

---

#### **49.4 Positionseffekte von Fragen**

Fragen können sich durch den Kontext und die Reihenfolge, in der sie stehen, gegenseitig beeinflussen (Diekmann 1997: 398, Raithel 2008: 76, Schnell et al. 2008: 342). Es können sogenannte *Positions-, Ausstrahlungs- oder Halo-Effekte* entstehen. Eine Frage kann bestimmte Themen aktivieren und die anschließenden Antworten beeinflussen, über die die Befragten nicht nachgedacht hätten, wäre sie vorher nicht angeschnitten worden. Die Frage danach, wie wichtig freiwilliges Engagement für die Gesellschaft ist, scheint allein noch kein Problem. Folgt anschließend jedoch die Frage, was die Befragten tun würden, um die Gesellschaft selber zu unterstützen, wäre das Thema freiwilliges Engagement bereits aktiviert und würde häufiger angeben werden, als ohne die vorangestellte Frage. Eine Aktivierung bestimmter Themen oder Assoziationen durch eine Frage kann das Antwortverhalten stark beeinflussen.

---

#### **49.5 Offene und geschlossene Fragen**

Sehr viele Fragebögen enthalten nur noch geschlossene Fragen, weil sie schneller zu erheben und einfacher auszuwerten sind. Das Kodieren der offenen Fragen kostet Zeit und verteuert die Befragung. Diesen zwei Nachteilen steht der Vorteil gegenüber, dass offene Fragen ein umfangreicheres und differenzierteres Material liefern (Züll/Menold, Kapitel 52 in diesem Band). So werden im Freiwilligensurvey 1999, 2004 und 2009 die Motive für freiwillige Arbeit mit fünf Items erfasst: „Ich will durch mein Engagement ... a) die Gesellschaft zumindest im Kleinen mitgestalten, b) vor allem mit anderen Menschen zusammenkommen, c) wichtige Qualifikationen erwerben, d) Ansehen und Einfluss in meinem

Lebensumfeld erwerben, e) auch beruflich vorankommen“; diese Items decken aber weder das aus der Literatur bekannte Spektrum der Motive ab, noch sind sie eine gute Zusammenfassung zu Motivbündeln. Eindeutig wäre eine offene Frage informativer gewesen; sie hätte es auch erlaubt, die Befragten besser zu differenzieren.

---

## 49.6 Filter

Filterfragen werden eingesetzt, wenn verschiedene Personengruppen mit einem Fragebogen untersucht werden und je nach Gruppe nur bestimmte Fragen sinnvoll sind. So sollten Personen, die nicht freiwillig tätig sind, auch nicht nach ihrem Zeitaufwand oder ihren Motiven für freiwillige Arbeit gefragt werden. Vor allem in postalischen Befragungen, in denen kein geschulter Interviewer durch den Fragebogen führt, sondern die Befragten diesen selbst ausfüllen müssen, sind Filter eine Fehlerquelle (Porst 2009: 153). Deshalb sollten in postalischen Befragungen möglichst wenige und nur absolut notwendige Filter verwendet werden (Reuband/Blasius 1996). Die visuelle Unterstützung durch Pfeile, Markierungen und Farben hilft den Befragten, sich im Fragebogen zu orientieren, Filterfehler (überflüssige oder fehlende Angaben) zu vermeiden und die richtigen Fragen anzukreuzen.

Vor allem für die Filterführung ist es wichtig, dass die Fragen nummeriert sind, da hinter einer Antwortmöglichkeit der Hinweis platziert wird, mit welcher Frage fortgefahren werden soll (vgl. Abb. 49.1). Filter können global, intern oder punktuell sein. Werden alle Befragten, die eine bestimmte Frage beantwortet haben, „gefiltert“, ist dies ein globaler Filter. Interne Filter werden bei Verweigerung von Fragen verwendet. Auf die nicht beantwortete Frage folgt dieselbe Frage noch einmal mit veränderter Technik, z.B. wird eine zuvor offene Frage dann durch eine Frage mit vorgegebenen Antwortmöglichkeiten ersetzt. Bei den Beispielen in Abb. 49.1 handelt es sich um punktuelle Filter, denn nur denjenigen Befragten, die keine Sozialhilfe beziehen, wird die nachfolgende Frage P gestellt; im zweiten Beispiel werden diejenigen Befragten, die nie einen gemeinsamen Haushalt hatten, auf die Frage 12 geschickt, die anderen auf Frage 11.

Online-Erhebungen und telefonische Befragungen über ein CATI-Labor werden so programmiert, dass je nach angeklickter Antwort bestimmte Fragen oder ganze Fragenblöcke automatisch gefiltert werden, so dass hier auch komplexere Filterführungen verwendet werden können, weil die Befragten nicht selbstständig nach den zu beantwortenden Fragen suchen müssen. Daraus resultiert auch, dass bei komplizierten Frageführungen Online-, telefonische, oder CAPI-Befragungen gewählt werden sollten oder auch face-to-face-Befragungen, bei denen ein gut geschulter Interviewer durch den Fragebogen navigiert.

<b>O. Haben Sie oder ein anderes Haushaltsmitglied im letzten Jahr (2003) eine oder mehrere der folgenden Leistungen erhalten?</b>			
	Ja	Nein	k.A.
Wohngeld	1	2	9
Kindergeld	1	2	9
Unterhaltszahlungen	1	2	9
Arbeitslosengeld	1	2	9
Arbeitslosenhilfe	1	2	9
Unterhaltsgeld vom Arbeitsamt für Umschulung/ Fortbild.	1	2	9
Sozialhilfe	1	2	9
<b>Nein ⇒ Weiter mit P!</b>			
Laufende Hilfe zum Lebensunterhalt	1	2	9
Hilfe in besonderen Lebenslagen	1	2	9
Einmalige Hilfe zum Lebensunterhalt	1	2	9

**10. Haben oder hatten Sie jemals einen gemeinsamen Haushalt mit einem Ehe- oder Lebenspartner bzw. einer Ehe- oder Lebenspartnerin?**

1 Ja

5 Nein → **Weiter mit Frage 12.**



**Abb. 49.1** Beispiele von Filterfragen (Quellen: Face-to-face -ragebogen des Projekts „Nachbarschaftseffekte: Leben in einem benachteiligten Stadtteil“, (2004-2008). Blasius, Friedrichs und Klöckner 2008: 188. Schriftlicher Fragebogen des „Share“-Projekts, gefördert vom Bundesministerium für Bildung und Forschung, von der Europäischen Kommission und dem US National Institute on Ageing. Schröder 2011.)

## 49.7 Layout des Fragebogens

Das Layout des Fragebogens ist vor allem für schriftliche Fragebögen von Bedeutung, da der Fragebogen zum einen ansprechend gestaltet sein sollte, um die Befragten zu motivieren den Bogen auszufüllen, zum anderen, um die Navigation durch die Fragen zu erleichtern (Porst 2009: 165). Aber auch für Interviewer muss der Fragebogen so gestaltet sein, dass diese sich ohne Schwierigkeiten darin zurechtfinden. Dabei sollte darauf geachtet werden, dass die Schrift gut lesbar ist und nicht zu viele Fragen auf einer Seite platziert werden. Der ausgefüllte Fragebogen kann per Hand oder auch automatisch über einen

Scanner eingelesen werden. Dazu ist es erforderlich, den Fragebogen genau auf den Scanner auszurichten. Dazu werden an den vier Kanten jeder Seite Markierungen angebracht, meist kleine schwarze Quadrate. Diese Technik erfordert aber auch, bei Skalen genügend Zwischenraum zwischen den Zeilen zu lassen, damit die Kreuze auf den Kategorien jedes Items nicht zu sehr aneinander kommen. Man benötigt mehr Platz für den Fragebogen, gewinnt aber eine preiswerte Übertragung der Daten in einen File, der dann u.a. mit SPSS oder STATA weiter verarbeitet werden kann.

---

## 49.8 Sonstige Hinweise

### 49.8.1 Listen

Das Abfragen von längeren Listen ist nicht nur demotivierend für die Befragten, sondern auch für die Interviewer. Der Fragebogen ist kein Geheimnis. Die befragte Person sollte den Bogen einsehen können. Deshalb sollte man der befragten Person entweder den Fragebogen hinüber schieben oder ihr eine Liste auf einem gesonderten Blatt überreichen (das sollte allerdings in der Interviewer-Anweisung vorgesehen sein). Dadurch wird die befragte Person einbezogen und sie bleibt aufmerksam. Das ist besonders wichtig, wenn z.B. lange Listen mit Netzwerkpersonen oder Aktivitäten im Tageslauf (Zeitbudgets) erhoben werden.

### 49.8.2 Titelblatt

Ein ansprechend gestaltetes Titelblatt gehört vor allem bei postalischen Befragungen zu einem Fragebogen dazu (vgl. Abbildung 2). Dazu zählen der Titel der Untersuchung, Projektleiter/in und Institut sowie dessen Adresse, Befragungszeitraum bzw. -datum (Kallus 2010: 82f., Porst 2009: 31f., Raithel 2008: 77). Der Titel der Untersuchung sollte dabei allgemein gehalten sein und möglichst keine Erwartungen bei den potentiellen Befragten wecken. Ein Hinweis auf einen Ansprechpartner bei Fragen zur Untersuchung (mit Telefonnummer) erhöht die Vertrauenswürdigkeit und bietet den Befragten die Möglichkeit, Rückfragen an die Studienleiter zu stellen. Ein ansprechendes Logo oder Titelbild kann die Teilnahme noch unterstützen und die Seriosität der Befragung und des Instituts unterstreichen (Porst 2009: 37 ff.). Allerdings sollten verwendete Fotos, Karikaturen oder anderes Bildmaterial nicht provozierend wirken oder die Befragten verärgern (Porst 2009: 40). Die Informationen des Anschreibens dürfen sich auf dem Titelblatt in verkürzter Form wiederholen, für den Fall, dass das Anschreiben und damit wichtige Informationen, wie z.B. die Rückantwortadresse, verloren gehen (Porst 2009: 35).

Bei face-to-face-Befragungen, in denen ein Interviewer die Befragung durchführt, haben Titelblätter nicht dieselbe Relevanz für die Studienteilnehmer wie in schriftlichen Befragungen. Die Befragten betrachten den Fragebogen in den meisten Fällen nur flüchtig,

da der Interviewer die Aufzeichnungen der Antworten übernimmt (Porst 2009: 31f.). Hier können die Titelblätter bereits für die ersten Fragen verwendet werden, wie das linke Beispiel zeigt (s. Abbildung 2). Das Titelblatt der Studie „Nachbarschaftseffekte: Leben in einem benachteiligten Stadtteil“, nimmt die Daten des Interviewers auf, beinhaltet den Titel der Befragung und die durchführende Organisation (in diesem Falle die Universität zu Köln) sowie einen einleitenden Text, den die Interviewer den Befragten vorlesen. Hier wird auf ein bereits erhaltenes Anschreiben Bezug genommen. Weiterhin führt das Titelblatt einige Anmerkungen für den Interviewer auf. Anschließend werden die ersten Fragen aufgelistet.

Das rechte Beispiel ist der schriftliche Zusatzfragebogen des „Share“-Projekts. Hier finden sich die Logos des Projekts sowie das Logo des Befragungsinstituts (in diesem Falle Infas). Der Titel der Studie wird aufgeführt, Daten des befragten Haushalts und der befragten Person, das Interviewdatum und Interviewer-Nummer.

Laufende Nummer des Interviews:		<i>(Bitte nicht eingraben! Wird im Institut vergeben)</i>	
Interviewer-Nr.:		<i>(Interviewer-Nr.)</i>	
<b>Lebensqualität in Köln Vingst/Höhenberg</b> ( <i>© Universitätsbibliothek Köln</i> )			
<b>Fragebogen der Universität zu Köln</b>			
Guten Tag, mein Name ist ... Ich komme vom Forschungsinstitut für Soziologie der Universität zu Köln. Ich würde gern Frau/ Herr ... (Vor- und Zuname) sprechen. Wir führen eine wissenschaftliche Untersuchung durch. Es geht dabei um Ihre Ansichten zu Ihrem Wohnviertel und um Ihre Zufriedenheit im Viertel.			
<b>INT.: FOLGENDES NUR DER ZIELPERSON VORLESEN!</b> Wir haben Ihnen vor kurzen geschrieben und Ihnen mitgeteilt, dass wir Sie befragen möchten. Haben Sie den Brief bekommen?  JA: → Ich möchte das Interview jetzt mit Ihnen durchführen. NEIN: → Hier ist der Brief. Ich möchte das Interview jetzt mit Ihnen durchführen.  Meine ersten Fragen beziehen sich auf Ihre Zufriedenheit mit Ihrer Wohnung und Ihrer Wohnungsguthaltung.			
<b>1. Seit wann wohnen Sie in dieser Wohnung?</b> Hierzu schließen sich Bad, Flur, Räume, etc. an. Jahr: _____ → W.h.: 9999 k.A.: 9999			
<b>4. Haben Sie einen WBS-Schein benötigt, um diese Wohnung zu bekommen?</b> Ja      1 Nein    2 w.h.   3 k.A.   4			
<b>5. Wohnen Sie bzw. Ihre Familie hier zur Miete oder als Eigentümer?</b> Miete    1 Eigentümer    2 k.A.   3			
<b>6. Wie hoch ist Ihre monatliche Miete (inkl. Nebenkosten, Heizung)?</b> Hier im Haus    01 Hier im Viertel    02 Hier in Köln    03 Außenhalb Kölns, u. zwar: k.A.    99 Bin nie umgezogen (INP: Weiter mit Frage 6) 97			
<b>7. Wie haben Sie zuletzt gewohnt?</b> Hier im Haus    01 Hier im Viertel    02 Hier in Köln    03 Außenhalb Kölns, u. zwar: k.A.    99 Bin nie umgezogen (INP: Weiter mit Frage 6) 97			
<small>Hauptstudie April 2004 / 3263</small>			

<b>SHARE</b>		<small>Survey of Health, Ageing and Retirement in Europe</small>	
<b>infas</b>			
Lfd. Nr. des Haushalts	Pers.-Nr.		
1   2   0   4   2	0   0		
Interviewdatum: <input type="text"/> / <input type="text"/> / <input type="text"/>			
Interviewer-Nr.: _____			
Namenskürzel d. Befragten: _____			
<b>„50+ in Europa“</b> <b>Schriftlicher Fragebogen</b>			

**Abb. 49.2** Beispiele für Titelblätter von Fragebögen (Quellen: Face-to-face-Befragung des Projekts „Nachbarschaftseffekte: Leben in einem benachteiligten Stadtteil“, (2004-2008) (Blasius, Friedrichs und Klöckner 2008: 171); Schriftlicher Fragebogen des „Share“-Projekts, gefördert vom Bundesministerium für Bildung und Forschung, der Europäischen Kommission und dem US National Institute on Ageing (Schröder 2011). )

### 49.8.3 Anschreiben

Das Anschreiben sollte die wichtigsten Informationen zur Untersuchung enthalten, eine Beschreibung des Vorhabens und, falls diese nicht bereits aus dem Briefkopf des Anschreibens hervorgehen, die Kontaktdaten der Forscher für Rückfragen; außerdem sollte es die Anonymität der Befragten betonen. Die potentiellen Befragten sollten durch die persönliche Ansprache im Anschreiben motiviert und darüber informiert werden, warum es interessant und sinnvoll für sie ist, teilzunehmen. Sie sollten gut erkennen, welchen Nutzen sie von der Befragung haben. Aber welchen Nutzen können wir einer arbeitslosen Person nennen, der „Universität“ einen Fragebogen zu beantworten? Und wenn die Antwort lautet, die Befragung diene auch dazu, die Lebensbedingungen im Wohngebiet zu verbessern („also auch Ihre“), dann sollten wir das ernst meinen und auch etwas für die Umsetzung der Erkenntnisse tun.

Das Anschreiben sowie die Erläuterungen eines Begleitschreibens sind von großer Bedeutung für die Teilnahme, weil sie den ersten Eindruck vermitteln. Sie sollten vor allem bei postalischen Befragungen darstellen, um welches Thema es sich handelt, wer die Forscher und Institutionen sind und zu welchem Zweck die Untersuchung durchgeführt wird. In face-to-face-Befragungen können die Interviewer diese Erläuterungen vorlesen. Eine Originalunterschrift der Projektleiter sollte ebenfalls enthalten sein (Friedrichs 1990: 238, Kirchhoff et al. 2003: 29). Auch sollte hier auf die Anonymität der Teilnehmer hingewiesen und für Rückfragen Ansprechpartner mit Telefonnummer genannt werden.

Sofern die Namen der zu befragenden Person bekannt sind, ist es nachgerade zwingend, sie in das Anschreiben aufzunehmen, um dieses zu personalisieren. Das war in dem abgedruckten Beispiel nicht möglich. Deshalb musste besonders viel Mühe darauf verwendet werden, ein einladendes Anschreiben zu formulieren. Ein Beispiel für ein Anschreiben ist in Abbildung 3 wiedergegeben.

*Am Ende der Konstruktion eines Fragebogens* sollte man nochmal zurück zu den Hypothesen gehen und prüfen: Wozu brauchen wir diese Frage? Ist sie Teil des Tests einer Hypothese? Ist sie zumindest zwingend, um einen für die Studie zentralen Sachverhalt zu beschreiben, z.B. prekäre Einkommensverhältnisse? Es könnte sein, dass sich bei dieser Prüfung einige Fragen als überflüssig erweisen.

**Universität zu Köln**



Universität zu Köln • Forschungsinstitut für Soziologie • Greifstr. 2 • 50939 Köln

Frau Mia Mustermann  
Musterstraße 1  
50000 Köln

**Wirtschafts- und Sozialwissenschaftliche Fakultät**

**Forschungsinstitut für Soziologie**  
Prof. em. Dr. Jürgen Friedrichs  
Telefon +49 221 470-2409  
Telefax +49 221 470-5025  
Sekretariat +49 221 470-5652  
friedrichs@wiso.uni-koeln.de  
[www.wiso.uni-koeln.de/fifa](http://www.wiso.uni-koeln.de/fifa)  
Greifstr. 2 - 50939 Köln

**Befragung**

— Sehr geehrte Frau Mustermann

wie Sie vielleicht in der Zeitung gelesen haben, führen wir zurzeit eine wissenschaftliche Untersuchung zu den Lebensbedingungen in Höhenberg und Vingst durch. Wir möchten gerne erfahren, wie zufrieden Sie mit Ihrer Wohnumgebung sind, was Sie gut finden, was Sie stört und was man verbessern könnte.

Um diese Fragen zu beantworten, haben wir aus der Einwohnerstatistik eine zufällige Auswahl von Personen getroffen, die wir befragen wollen. Zu diesem ausgewählten Personenkreis gehören auch Sie. Damit die Studie gelingt, ist es wichtig, dass jede der ausgewählten Personen auch an der Befragung teilnimmt. Deshalb wären wir Ihnen sehr dankbar, wenn Sie zu dem Interview bereit wären.

Die von Ihnen gemachten Angaben werden selbstverständlich vertraulich behandelt und bleiben anonym. *Es ist absolut sicher gestellt, dass Ihre Angaben nicht mit Ihrer Person in Verbindung gebracht werden können.* Die Ergebnisse unserer Untersuchung werden veröffentlicht und der Stadt Köln vorgelegt. Wir versprechen uns davon, die Lebensbedingungen mit ihrer Hilfe verbessern zu können. *Pastor Meurer hat uns bei dieser Studie beraten und unterstützt die Befragung.*

Als Dankeschön senden wir Ihnen auf Wunsch gerne eine Kurzfassung der Ergebnisse zu.

In den nächsten Tagen wird Sie einer/r unserer Interviewer/innen aufsuchen. Ein von uns ausgestellter Ausweis kennzeichnet ihn/sie als Mitarbeiter/in der Universität zu Köln.

Sollten Sie Rückfragen haben, so können Sie uns gerne unter einer der unten angegebenen Telefonnummern in Köln anrufen. Für Ihre Teilnahme an der Befragung möchten wir Ihnen im Voraus herzlich danken.

Mit freundlichen Grüßen

— Prof. Dr. Jürgen Friedrichs 

— Prof. Dr. Jörg Blasius 

**Abb. 49.3** Beispiel für ein Anschreiben eines Fragebogens (Quellen: Anschreiben der Befragung des Projekts „Nachbarschaftseffekte: Leben in einem benachteiligten Stadtteil“, (2004-2008) von Friedrichs und Blasius.)

## Literatur

- Blasius, Jörg/Friedrichs, Jürgen/Klöckner, Jennifer (2008): Doppelt benachteiligt? Leben in einem deutsch-türkischen Stadtteil. Wiesbaden: VS Verlag
- Diekmann, Andreas (1997): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Hamburg: Rowohlt
- Friedrichs, Jürgen (1990): Methoden empirischer Sozialforschung. Opladen: Westdeutscher Verlag
- Holm, Kurt (1975): Die Befragung 1. München: Francke
- Kallus, Wolfgang (2010): Erstellung von Fragebogen. Wien: facultas.wav
- Kirchhoff, Sabine/Kuhnt, Sonja/Lipp, Peter/Schlawin, Siegfried (2003): Der Fragebogen. Datenbasis, Konstruktion und Auswertung. Opladen: Leske + Budrich
- Porst, Rolf (2009): Fragebogen. Ein Arbeitsbuch. Lehrbuch. Studienskripten zur Soziologie. Wiesbaden: VS Verlag
- Raithel, Jürgen (2008): Quantitative Forschung. Ein Praxiskurs. Wiesbaden: VS Verlag
- Reuband, Karl-Heinz/Blasius, Jörg (1996): Face-to-face, telefonische und postalische Befragungen): Ausschöpfungsquoten und Antwortmuster in einer Großstadt-Studie. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 48: 296-318
- Schnell, Rainer/Hill, Paul B./Esser, Elke (2008): Methoden der empirischen Sozialforschung. München/Wien: Oldenburg Verlag
- Scholl, Achim (2003): Die Befragung. Konstanz: UVK
- Schowder, Mathis (2011): Retrospective Data Collection in the Survey of Health, Ageing and Retirement in Europe. SHARELIFE Methodology. Mannheim: MEA

## Projekte, deren Fragebögen als Beispiele dienten

- Friedrichs, Jürgen/Blasius, Jörg (2004-2008): Projekt „Nachbarschaftseffekte: Leben in einem benachteiligten Stadtteil“
- Munich Center for the Economics of Aging (MEA) (Projektkoordinator Deutschland) (2004 bis heute): „Share-Projekt“, gefördert vom Bundesministerium für Bildung und Forschung, Europäische Kommission und US National Institute on Aging

Rolf Porst

## 50.1 Informationsverarbeitung bei der Frageformulierung

„...., the reader will be disappointed if he expects to find here a set of definite rules or explicit directions. The art of asking questions is not likely ever to be reduced to easy formulas.“  
(Payne 1951: xi)

Ungeachtet dieser selbst formulierten Einschränkung legt Stanley L. Payne mit seinem 1951 erschienenen Buch „The Art of Asking Questions“ erstmals eine systematische Darstellung und Diskussion von Regeln vor, die bei der Formulierung von Fragebogen-Fragen hilfreich und deshalb zu berücksichtigen wären. Waren bis zu diesem Zeitpunkt ausschließlich einzelne Kapitel in Büchern (Rugg/Cantril 1944) oder Beiträge in Fachzeitschriften (Rugg 1941; Hubbard 1950) erschienen, wird hier zum ersten Mal ein komplettes „Frageformulierungs-Buch“ vorgelegt. Alle späteren Versuche, Regeln für die Formulierung von Fragebogenfragen zusammenzustellen, gehen explizit oder implizit auf Payne (1951) zurück, die meisten davon wieder „nur“ als Einzelbeiträge in Handbüchern, in Büchern zu „Methoden der empirischen Sozialforschung“ oder sonstigen methodischen Abhandlungen (Converse/Presser 1986; Porst 2000; Groves et al. 2004; Diekmann 2007; Häder 2010; Porst 2014).

Auf der anderen Seite impliziert das Buch von Payne schon im Titel, dass es sich bei der Fragebogenkonstruktion und insbesondere bei der Formulierung von Fragebogen-Fragen um eine „Kunst“ oder „Kunstlehre“ handle, eine Sichtweise, die in Umfrageforschung und -praxis relativ lange tradiert wird. In neueren Einführungen in die empirische Sozialforschung (Diekmann 2007; Häder 2010) hat sich dagegen die Erkenntnis verfestigt, dass neben individuellen Fertigkeiten wie z.B. informellem Wissen, persönlicher Erfahrung oder Begabung in zunehmendem Maße die Rezeption und Umsetzung wissenschaftlicher Erkenntnisse, insbesondere aus der kognitionspsychologischen Forschung, wesentlich zur Optimierung von Fragebogen beitragen können.

Seit Ende der Siebziger Jahre des Zwanzigsten Jahrhunderts gibt es systematische Versuche, Fragebogengestaltung als integrierten Bestandteil eines theoretischen Konzepts der Befragung zu verstehen. Seit dieser Zeit arbeiten Sozialpsychologen und Psychologen gemeinsam mit Umfrageforschern an der Erforschung der kognitiven und kommunikativen Prozesse, die der Befragungssituation zugrunde liegen, und die aus ihrer Forschung abgeleiteten konkreten Empfehlungen für die Gestaltung von Fragebogen (Sudman/Bradburn 1982; Schwarz/Sudman 1996; Sudman et al. 1996; Klöckner/Friedrichs, Kapitel 49 in diesem Band) haben das Forschungs- und Arbeitsgebiet „Fragebogenentwicklung“ entscheidend zum Positiven und das heißt hin zu einer eigenständigen wissenschaftlichen Teildisziplin, verändert. Mit den Arbeiten der als „KognUm“ („Kognitionspsychologie und Umfrageforschung“) bezeichneten Kooperation von Psychologen, Sozialpsychologen und Umfrageforschern hat die Entwicklung von Fragebogen eine neue Qualität erfahren.

Ausgangspunkt der Überlegungen zur Qualität von Fragen ist das „model of information processing in a survey situation“ von Strack/Martin (1987: 124ff.; sehr gut dargestellt auch in Sudman et al. 1996: Kapitel 3). Diesem Modell zufolge haben Personen, die an einer Befragung teilnehmen, mehrere Aufgaben zu lösen; sie müssen ...

1. die gestellte Frage verstehen (interpreting the question)
2. relevante Informationen zum Beantworten der Frage aus dem Gedächtnis abrufen und auf der Basis dieser Informationen ein Urteil bilden (generating an opinion)
3. dieses Urteil ggf. in ein Antwortformat einpassen (formatting the response) und
4. ihr „privates“ Urteil vor Weitergabe an den Fragebogen gegebenenfalls „editieren“ (editing the response).

---

## 50.2 Die Frage verstehen

Ob und wie gut Befragungspersonen diese Aufgaben bewältigen können, hängt sehr stark von der Qualität der gestellten Frage und damit nicht zuletzt von der Qualität der Frageformulierung ab. Dies wird besonders deutlich bei Punkt 1 des Modells, dem Verstehen einer Frage.

Verständnis hat – aus Sicht der Befragungspersonen – zwei Dimensionen:

1. Semantisches Verständnis: Was soll eine Frage oder ein Begriff in einer Frage „heissen“?
2. Pragmatisches Verständnis: Was will der Forscher/die Forscherin – oder der/die sie repräsentierende InterviewerIn – eigentlich wissen?

Eine Frageformulierung ist dann gut, wenn sowohl Semantik wie auch Pragmatik für die Befragungspersonen einleuchtend und nachvollziehbar sind. Ist dies nicht der Fall, sehen sich die Befragungspersonen einer Vielzahl von Problemen gegenüber.

---

*Probleme mit dem semantischen Verständnis* treten bei Befragungspersonen u.a. dann auf, wenn ...

1. Begriffe in Fragen unbekannt sind
2. Fragen unklar formuliert sind, auch wenn die darin verwendeten Begriffe „bekannt“ sind
3. Fragen zu schwierig formuliert sind
4. Begriffe in Fragen mehrdeutig sind
5. Begriffe verwendet werden, die legitimer Weise von jeder Befragungsperson individuell verstanden und individuell interpretiert werden können
6. Begriffe verwendet werden, die von unterschiedlichen Befragungsgruppen unterschiedlich verstanden werden.

*Probleme mit dem pragmatischen Verständnis* treten dann auf, wenn eine Frage nicht erkennen lässt, was der Forscher durch sie eigentlich „wissen“ will.

---

### 50.3 Die zehn Gebote der Frageformulierung

Um Probleme dieser Art zu vermeiden oder zumindest zu mildern, sollte man sich bei der Formulierung von Fragebogen-Fragen einschlägiger Regeln bzw. Richtlinien bedienen. Exemplarisch für solche Regeln und Grundsätze betrachten wir im Folgenden die „10 Gebote der Frageformulierung“ von Porst (2000; siehe auch Porst 2014, Kapitel 7), auch diese inspiriert von Payne (1951):

#### **1. Gebot:**

**Du sollst einfache, unzweideutige Begriffe verwenden, die von allen Befragten in gleicher Weise verstanden werden!**

Die Vorstellung, dass eine Frage von allen Befragten in gleicher Weise verstanden werden soll, ist von zentraler Bedeutung für die Durchführung standardisierter Befragungen. Die Chance, diesem Ziel näher zu kommen, wächst in dem Maße, in dem die Fragen einfach und unzweideutig formuliert werden. Was wiederum eine „einfache“ oder „unzweideutige“ Formulierung ist, hängt allerdings sehr von den Personen ab, die man zu befragen gedenkt: bestimmte Formulierungen mögen für VWL-Professoren einfach und verständlich sein, müssen aber deshalb von anderen Personen noch lange nicht verstanden werden. Beispiel:

Was glauben Sie: Wird sich die Konjunktur in Deutschland bis Ende des Jahres 2014 im Vergleich zu heute sehr positiv entwickeln, eher positiv, eher negativ oder sehr negativ, oder wird alles so bleiben, wie es heute ist?

Während die erwähnten VWL-Professoren diese schwierige Frage wohl verstehen und beantworten können, sollte man das bei anderen Bevölkerungsgruppen nicht unbedingt erwarten. Hier sind Probleme mit dem semantischen Verständnis vorprogrammiert: Ist der Begriff „Konjunktur“ überhaupt (und im richtigen Sinne) bekannt? Sind die Begriffe in den Antwortkategorien verständlich? Hier böte sich eher die folgende Formulierung an:

Was glauben Sie: Wie wird die wirtschaftliche Lage in Deutschland am Ende des Jahres 2014 sein: wesentlich besser als heute, etwas besser, gleichbleibend, etwas schlechter oder wesentlich schlechter?

Nun ist der Begriff „Konjunktur“ nicht vollkommen deckungsgleich mit dem Begriff „wirtschaftliche Lage“, und bei einer Befragung von VWL-Professoren sollte man den Begriff „Konjunktur“ ganz bestimmt nicht durch „wirtschaftliche Lage“ ersetzen. Bei anderen Zielgruppen oder bei Allgemeinpopulationen sollte man sich dagegen eher für „wirtschaftliche Lage“ entscheiden, weil man hier nicht sicher sein kann, dass der Begriff „Konjunktur“ verstanden wird oder überhaupt bekannt ist. Es ist aber letztendlich besser, in Umfragen belastbare Informationen über die Erwartungen an die Entwicklung der „wirtschaftlichen Lage“ zu erhalten als nicht-belastbare oder gar keine Informationen über die Erwartungen an die Entwicklung der „Konjunktur“ bis zu einem bestimmten Zeitpunkt.

Man sieht an diesem Beispiel – und das gilt für die meisten Regeln zur Formulierung von Fragebogenfragen – dass man bei jeder Frage, die man konstruiert, den Blick immer auf diejenigen richten muss, die letztendlich die Fragen beantworten oder den Fragebogen ausfüllen sollen. Und man sieht auch, dass es zu einem Konflikt kommen kann zwischen der Forderung, Fragen eindeutig („Konjunktur“) und der Forderung, Fragen einfach („wirtschaftliche Lage“) zu formulieren. In welche Richtung man diesen Konflikt löst, hängt ebenfalls in erster Linie davon ab, wer die Fragen beantworten soll. Wichtig ist, dass die Frageformulierung dem Sprachwissen, der Sprachfähigkeit und dem Sprachgebrauch der Zielpersonen möglichst nahe kommt (was umso schwieriger wird, je heterogener die Zielgruppen sind und/oder je entfernter die Lebenswelt der Zielpersonen von der Lebenswelt der ForscherInnen ist). Damit ist aber keineswegs gemeint, dass man sich als Umfrageforscher „mit Gewalt“ den sprachlichen Gepflogenheiten von „Subkulturen“ anpassen soll. Befragungen von Jugendlichen z.B. sollten zwar auf deren typische Begriffe und Idiome zurückgreifen, wo es sinnvoll und nötig ist, aber auch in diesem Falle sollte die Sprache des Fragebogens insgesamt eine gute Umgangssprache sein; es ist auch nicht ratsam, Fragebogen in Dialekt zu formulieren, selbst wenn man im „tiefsten Niederbayern“ oder in der „hintersten Westpfalz“ befragt – auch hier ist gute hochdeutsche Umgangssprache am Platze.

**2. Gebot:****Du sollst lange und komplexe Fragen vermeiden!**

Lange und komplexe Fragen bergen das Risiko, dass sie schnell unverständlich werden und die Befragungsperson verwirren, dass sie Begriffe enthalten, die redundant oder schlicht überflüssig sind, und/oder – vielleicht sogar unbeabsichtigt – zu viele unterschiedliche Stimuli beinhalten. Wenngleich auch hier die Frage, wie komplex eine Fragebogen-Frage sein darf, in hohem Maße von der Zielgruppe der Befragung abhängt, ist der Unterschied zwischen einer langen und komplexen Frage und einer kurzen und einfachen Frage doch oft schon augenscheinlich:

Wie Sie wissen, sind manche Leute politisch ziemlich aktiv, andere Leute finden dagegen oft keine Zeit oder haben kein Interesse, sich an politischen Dingen aktiv zu beteiligen. Ich lese Ihnen jetzt eine Reihe von Sachen vor, die Leute tun. Bitte sagen Sie mir jedes Mal, wie oft Sie persönlich so etwas tun bzw. wie häufig das bei Ihnen vorkommt. (Liste mit den Antwortkategorien oft - manchmal - selten - niemals). Zuerst: wie oft führen Sie eine politische Diskussion?

Hier werden im einleitenden Satz Informationen gegeben, die zur Beantwortung der eigentlichen Frage absolut nicht erforderlich sind und die den Fragetext unnötig aufzublähen. Alleine das Streichen des einleitenden Satzes reduziert den Umfang der Frage um ein Drittel; ein paar kleine Änderungen, und die Frage könnte – ohne Informationsverlust zu befürchten – wie folgt aussehen:

Wie häufig nehmen Sie aktiv an Diskussionen über politische Themen teil, oft, manchmal, selten oder nie?

Oder, von der Zielrichtung der Frage leicht anders, aber von der Formulierung her noch einfacher.

Wie häufig nehmen Sie aktiv an Gesprächen über politische Themen teil, oft, manchmal, selten oder nie?

**3. Gebot:****Du sollst hypothetische Fragen vermeiden!**

Hypothetische Fragen sind solche, zu deren Beantwortung sich die Befragungsperson in Situationen versetzen muss, in denen sie tatsächlich nicht ist. Ob ihr das gelingt oder nicht hängt zum einen davon ab, in wie fern sich die Person mit der hypothetischen Situation überhaupt schon einmal auseinandergesetzt hat und zum andern davon, wie nahe oder weit die hypothetische von der realen Situation entfernt ist. Die Frage, ob man am Montagmorgen noch zur Arbeit gehen würde, wenn man am Samstag 11 Millionen Euro im Lotto gewonnen hätte, ist sicherlich außerordentlich hypothetisch (wenn man sie wörtlich

nimmt), aber dennoch zu beantworten, weil sich fast jeder schon mal gefragt hat, ob er arbeiten würde, wenn er sein Leben auch ohne Erwerbseinkommen gut fristen könnte. Stellt man dagegen bei einer Befragung von Jugendlichen zwischen 14 und 20 Jahren die Frage, was sie tun würden, wenn sich ihr 16-jähriger Sohn vom Christentum ab- und dem Islam zuwenden würde, wird das eher zu wenig belastbaren Informationen führen, weil die meisten der befragten Jugendlichen sich mit dieser Thematik wohl noch nicht auseinandergesetzt haben (alleine schon die Vorstellung, einen 16-jährigen Sohn zu haben, dürfte die befragten Jugendlichen etwas überfordern).

#### **4. Gebot: Du sollst doppelte Stimuli und Verneinungen vermeiden!**

Die Frage...

Hören Sie gerne Musik von Chopin und Richard Wagner?

ist an sich gar nicht so unsinnig, weil Chopin wie Richard Wagner große Meister der (wenn man das einfach mal so ganz allgemein sagen darf) „klassischen Musik“ sind. Nun ist aber nicht nach klassischer Musik gefragt, sondern nach Musikschaffenden recht unterschiedlicher Ausrichtung.

Deshalb ist die Frage nicht eindeutig zu beantworten, wenn man z.B. zwar gerne Musik von Chopin, aber nicht gerne Musik von Wagner hört. Als denkbare und richtige Antwort wäre „Chopin ja, Wagner nein“ zu erwarten. Da die Befragungsperson aber nur mit „ja“ oder „nein“ antworten darf und dies deshalb auch tut, weiß man am Ende nicht, ob ein „ja“ heißt, dass sie sowohl gerne Chopin als auch Wagner hört oder ob sie mit ihrem „ja“ nur auf einen der beiden Stimuli reagiert hat – und auf welchen?

Wenn man also die beiden Meister erfragen will, bleibt nichts anderes übrig, als zwei Fragen zu stellen, einmal für Chopin, einmal für Wagner.

Problematisch ist auch der Einsatz *doppelter Verneinungen*, insbesondere im Zusammenspiel zwischen Frage- und Skalenformulierung. Wo platziert man bei der Aussage ...

Es ist nicht gut, wenn Kinder ihren Eltern nicht gehorchen

und einer Skala von 1 = „trifft überhaupt nicht zu“ bis 7 = „trifft voll und ganz zu“ seine Antwort, wenn man der Ansicht ist, es sei gut, wenn Kinder ihren Eltern (gelegentlich) nicht gehorchen? Man muss sich für einen Skalenwert entscheiden, der (noch eine Verneinung!) in Richtung „trifft überhaupt nicht zu“ geht. Da wäre eine positive Formulierung wie ...

Es ist gut, wenn Kinder ihren Eltern gelegentlich widersprechen.

bei gleicher Skala mindestens einfacher; jetzt kann man sich leicht für einen Skalenwert entscheiden, der in Richtung „trifft voll und ganz zu“ geht.

**5. Gebot:****Du sollst Unterstellungen und suggestive Fragen vermeiden!**

*Unterstellungen* in Frageformulierungen führen dazu, dass Befragungspersonen Fragen deshalb nicht vernünftig beantworten können, weil sie zwar der eigentlichen Aussage zustimmen könnten, nicht aber dem Satzteil, der die Unterstellung enthält (dies gilt auch für Items, die explizite Prämissen enthalten, denen man zustimmen muss bevor man zur „eigentlichen“ Aussage kommt). Ein Beispiel:

Hat der mangelnde Respekt der Schüler vor ihren Lehrern Ihrer Ansicht nach Einfluss auf die tägliche Unterrichtsgestaltung in den Schulen?

In der Formulierung dieser Frage wird Schülern mangelnder Respekt vor ihren Lehrern attestiert. Dies mag faktisch vielleicht sogar zutreffend sein, aber: Wie soll eine Befragungsperson reagieren, die diese eindeutige Unterstellung gar nicht teilt? Wenn die Befragungsperson nicht der Ansicht ist, dass es Schülerinnen und Schülern am Respekt vor den Lehrern mangelt, wird sie die Frage (eigentlich!) nicht beantworten (können).

*Suggestive Fragen* haben den Nachteil, dass sie die Befragungspersonen in die Ecke drängen, ihren Freiheitsspielraum beim Beantworten beinträchtigen. Auch hier ein Beispiel:

Führende Wissenschaftler sind der Ansicht, dass Autoabgase das Wachstum von Kindern hemmen können. Halten Sie diese Ansicht für richtig, oder halten Sie diese Ansicht für falsch?

Phrasen wie „führende Wissenschaftler“ führen dazu, dass Befragungspersonen sich nicht trauen, den vorgegebenen Autoritäten zu widersprechen, und deshalb konform antworten. Sie können aber auch genauso gut dazu führen, dass Personen auf die so entstandene Freiheitseinengung beim Beantworten der Frage dadurch reagieren, dass sie sich ganz bewusst gegen die Ansicht der Experten stellen, reaktant werden, und zwar unabhängig von ihrer wirklichen Meinung. Also sollte man die Suggestion aus der Frageformulierung nehmen und fragen:

Halten Sie die Aussage, dass Autoabgase das Wachstum von Kindern hemmen können, für richtig oder für falsch?

**6. Gebot:****Du sollst Fragen vermeiden, die auf Informationen abzielen, über die viele Befragte mutmaßlich nicht verfügen!**

Wie beim ersten Gebot ist auch die Umsetzung des sechsten Gebotes sehr stark von der Zielgruppe abhängig, die man befragen möchte. Die Beantwortung der Frage ...

Sind in Ihrer Gemeinde bereits Maßnahmen zur Umsetzung der lokalen Agenda 21 getroffen worden?

dürfte für Lokalpolitiker (hoffentlich!) keine großen Schwierigkeiten mit sich bringen, für den „Durchschnitt“ aber kaum zu beantworten sein, nicht nur alleine wegen des Begriffes „Lokale Agenda 21“, sondern weil man sich mit diesem Problem überhaupt noch nicht beschäftigt hat – viele Befragungspersonen werden mutmaßlich nicht über die Informationen verfügen, die zur Beantwortung der Frage notwendig wären (was ist die lokale Agenda 21? Was passiert in der Gemeinde damit?).

Das Problem stellt sich für Wissensfragen allgemein.

Man muss sich deshalb bei Fragen, die auf Wissen abzielen, immer vor Augen halten, inwiefern die konkret zu befragende Zielgruppe mutmaßlich über Informationen verfügt, die zur Beantwortung der Fragen erforderlich sind. Bestehen hier Bedenken, ist auf entsprechende Fragen besser zu verzichten, es sei denn, der eigentliche Sinn der Frage besteht wirklich darin, Wissen oder Nicht-Wissen festzustellen. Wenn nämlich Wissensfragen dazu führen, dass sich die Befragungsperson ihres Nicht-Wissen bewusst und dieses für sie selbst oder erst recht nach außen hin verdeutlicht wird, kann das schnell (weil sich die psychologischen „Kosten“ erhöhen) die Motivation zum Beantworten der weiteren Fragen beschädigen. Und das wiederum kann zu einer Verschlechterung der Datenqualität führen, in Einzelfällen sicherlich auch zu Interview-Abbrüchen.

## **7. Gebot:**

### **Du sollst Fragen mit eindeutigem zeitlichen Bezug verwenden!**

Ob Fragen zum Verhalten innerhalb bestimmter Zeiträume überhaupt funktionieren oder nicht hängt zum einen von der Bedeutung des Verhaltens ab, das erfragt werden soll, zum andern davon, wie weit der Forscher oder die Forscherin in die Vergangenheit (oder auch in die Zukunft) blicken lassen wollen. Ist das abgefragte Verhalten wichtig, kann man durchaus lange Zeitperioden abfragen; wenig Sinn dagegen machen zeitlich weit reichende Fragen zu banalen oder alltäglichen Ereignissen.

Unabhängig davon muss bei Fragen, die auf Sachverhalte oder Meinungen innerhalb eines bestimmten vergangenen, gegenwärtigen oder zukünftigen Zeitraumes abzielen, dieser Zeitraum genau und eindeutig definiert werden. Die Frage...

Sind Sie in den letzten Wochen irgendwann einmal zum Essen in ein Restaurant gegangen?

ist alleine schon deshalb schlecht, weil der angegebene Zeitraum, an den man sich zu ihrer Beantwortung erinnern soll, vollkommen vage und somit im interpretativen Belieben der Befragungspersonen bleibt.

Eindeutige Formulierungen verwenden ganz konkrete Zeitangaben als „Anker“ wie z. B. „seit dem 1. April“ (sofern der nicht allzu weit zurückliegt), „an Ihrem 16. Geburtstag“ oder „bis zum 31. Januar“:

Sind Sie im letzten Monat, also im Mai 2014 irgendwann einmal zum Essen in ein Restaurant gegangen?

Solche präzisen Anker garantieren natürlich immer noch nicht, dass die Befragungsperson ihre Antworten auch wirklich auf diese Zeitphasen beschränkt, aber sie geben doch immerhin genau vor, auf welche sie sich beschränken sollte.

Ein letztes schließlich: kalendarische Daten sind nur schlecht in den Köpfen der Menschen repräsentiert, und je länger sie zurückliegen und je unwichtiger Ereignisse sind, umso schlechter sind sie repräsentiert. Besser als kalendarische Daten eignet sich allemal der Bezug auf besondere Ereignisse wie „seit Weihnachten“, „seit Sie Ihre derzeitige Arbeitsstelle angetreten haben“ oder „seit Sie aus dem Krankenhaus entlassen worden sind“ - je nach Gegenstand der Befragung oder nach Richtung der Frage.

## 8. Gebot:

### Du sollst Antwortkategorien verwenden, die erschöpfend und disjunkt (überschneidungsfrei) sind!

Antwortkategorien (Franzen, Kapitel 51 in diesem Band) sind *disjunkt*, wenn sich jede Person zweifelsfrei einer einzigen davon (vorausgesetzt natürlich, dass keine Mehrfachnennungen erlaubt sind) zuordnen kann. Bei der Frage ...

Wie viele Vorträge zum Thema „Gesundes Leben“ haben Sie im Jahre 2014 bisher gehört?

... und den Antwortkategorien ...

- keinen
- einen Vortrag
- zwei bis fünf Vorträge
- fünf Vorträge oder mehr

... werden diejenigen Personen, die genau fünf Vorträge gehört haben, beim Beantworten ins Schlingern geraten, weil sie sich sowohl in der dritten wie auch in der vierten Antwortkategorie wiederfinden können.

Disjunkt wären die folgenden Antwortkategorien:

- keinen
- einen Vortrag
- zwei bis vier Vorträge
- fünf Vorträge oder mehr

*Nicht erschöpfend* sind Antwortkategorien dann, wenn die Antwort, die eine bestimmte Person geben möchte, durch die Antwortkategorien nicht abgedeckt ist. Beispiel:

Wie viele Stunden beschäftigen Sie sich in einer normalen Arbeitswoche mit der Entwicklung von Fragebogen?

... und die Antwortkategorien ...

- überhaupt nicht
- bis unter 3 Stunden
- 3 bis unter 5 Stunden
- 5 bis unter 10 Stunden

Wer das Glück hat, sich in einer normalen Arbeitswoche 10 oder mehr Stunden mit der Entwicklung von Fragebogen beschäftigen zu dürfen, könnte hier keine korrekte Angabe machen. Deshalb sollten die Antwortvorgaben korrigiert werden:

- überhaupt nicht
- bis unter 3 Stunden
- 3 bis unter 5 Stunden
- 5 bis unter 10 Stunden
- 10 Stunden oder mehr.

In beiden Fällen könnte man natürlich und sollte besser auch auf die Vorgabe von Antwortkategorien ganz verzichten und die Frage offen stellen, weil die Befragungsperson ohnehin keine Zeitspanne erinnert, sondern eine absolute Häufigkeit, die nach dem Erinnern dann in die vorgegebenen Kategorien formatiert wird; da sie also die absolute Häufigkeit erinnert, kann man Fragen dieser Art auch ohne Antwortvorgaben stellen.

## **9. Gebot:**

**Du sollst sicherstellen, dass der Kontext einer Frage sich nicht (unkontrolliert) auf deren Beantwortung auswirkt!**

Dass Fragen und die dazugehörigen Antwortvorgaben Auswirkungen auf Folgefragen haben können, ist unumstritten und hinreichend belegt. Nur: welche Fragen wirken sich auf Folgefragen wie aus? Antwort: Darüber kann man in der Phase der Fragebogenformulierung meist nur spekulieren. Hier geben erst ein Pretest oder – im schlechtesten Falle – erst die Daten der Befragung selbst Auskunft, sofern man überhaupt Mechanismen (z. B. unterschiedliche Fragebogenversionen mit unterschiedlichen Vorfragen vor der interessierenden Frage) in den Fragebogen eingebaut hat, um Kontexteffekte kontrollieren zu können. Kontexteffekte in Fragebogen lassen sich durch Nachdenken zwar ggf. erahnen, aber nur durch systematische kognitive Pretests (Prüfer/Rexroth 2005; Weichbold, Kapitel 19 in diesem Band) prognostizieren.

**10. Gebot:****Du sollst unklare Begriffe definieren!**

Begriffe sind dann unklar, wenn sie von den Befragungspersonen gar nicht bzw. nicht von allen Befragungspersonen in der gleichen Weise verstanden werden. Das Problem unklarer Begriffe ähnelt ein wenig dem Problem der einfachen Begriffe aus dem 1. Gebot und dem Problem der Verfügbarkeit über benötigte Informationen aus dem 6. Gebot. Im Grunde ist nur die Lösung eine andere: Nicht Vereinfachung der Frage (ist manchmal halt nicht möglich) und nicht Anpassung der Frage an die Zielpersonen (dadurch würde der Begriff verwaschen) heißt die Strategie, sondern Definition des unklaren Begriffes. Betrachten wir die folgende Frage:

Was glauben Sie: In welchem Alter beginnt bei Männern normalerweise die Andropause?

Die Frage wäre gar kein Problem bei einer Stichprobe von Andrologen. Für den Rest der Welt muss der Begriff Andropause allerdings erklärt werden. Dies kann man, je nach Stichprobe, z. B. in Analogie zur Menopause machen:

Unter Andropause versteht man eine Entwicklung im Alternsprozess von Männern, die mit der Menopause bei Frauen vergleichbar ist. Was glauben Sie ...

Diese Frage ist jetzt zwar recht klar geworden, aber dafür auch etwas länger als die erste Variante dieses Beispiels. Der Versuch, eines der Gebote der Frageformulierung (hier das 10. Gebot) einzuhalten, führt in diesem Falle dazu, dass man gegen ein anderes verstößt (hier das zweite Gebot der kurzen und einfachen Frage).

---

**50.4 Zur Handhabung der zehn Gebote**

Allgemein lässt sich zu Regeln der Frageformulierung abschließend feststellen: Wie alle Regeln des täglichen Lebens (sofern sie nicht strafrechtlich oder anderswie massiv sanktioniert werden können) haben sie einen entscheidenden Nachteil: Sie sind nur von beschränktem Nutzen! – Und damit wären wir wieder bei dem eingangs dargestellten Zitat von Payne (1951).

Regeln zur Formulierung von Fragebogen-Fragen sind geeignet als „allgemeine Wegweiser“, die eine grobe Richtung vorgeben, aber ihre Bedeutung relativiert sich und schwindet manchmal sogar ganz, wenn es um die Formulierung konkreter Fragen für einen konkreten Fragebogen geht.

Dann nämlich muss über jede Formulierung neu nachgedacht werden, und Regeln zur Frageformulierung können dabei unterstützend wirken oder sogar hilfreich sein, aber sie ersparen es nicht, ihre Gültigkeit und Wirksamkeit bei jeder neuen Frage jeweils neu zu überdenken und ggf. zu testen.

Man beachte weiterhin: Bei Regeln zur Formulierung von Fragebogenfragen handelt es sich nicht um feste, unumstößliche Vorgaben, die man „blind“ anwenden kann. Die meisten dieser Regeln lassen Interpretationsspielraum zu und stehen gelegentlich sogar in Konkurrenz zueinander, sind also nicht in jedem Falle hundertprozentig einzuhalten.

Man sollte mit solchen Regeln kritisch umgehen, wenn es darum geht, ob eine konkrete Frage „gut“ ist. Regeln zur Formulierung von Fragebogenfragen können genutzt werden als Instrumentarium, das hilft, Fragen „gut“ zu formulieren. Sollte der Fragenentwickler aber trotz all dem Nachdenken, Formulieren und Reformulieren noch immer zweifeln, ob eine Fragebogen-Frage wirklich gut gelungen ist, bleibt immer noch die Möglichkeit, die Qualität dieser Frage in einem kognitiven Pretest zu überprüfen. Weitere Empfehlungen zur Formulieren von Fragen anhand von zahlreichen Beispielen gibt Porst (2014).

## Literatur

- Cantril, Hadley (1944): Gauging Public Opinion. Princeton: University Press
- Converse, Jean M./Presser, Stanley (1986): Handcrafting the Standardized Questionnaire. Beverly Hills: Sage
- Diekmann, Andreas (2007): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. 17. Aufl. Reinbek: Rowohlt
- Groves, Robert M./Fowler, Floyd J./Couper, Mick P./Lepkowski, James M./Singer, Eleanor/Tourangeau, Roger (2004): Survey Methodology. New Jersey: Wiley Hoboken
- Häder, Michael (2010): Empirische Sozialforschung. Eine Einführung. 2. Aufl. Wiesbaden: VS Verlag
- Hippler, Hans-Jürgen/Schwarz, Norbert/Sudman, Seymour (1987): Social Information Processing and Survey Methodology. New York: Springer
- Hubbard, Alfred W. (1950): Phrasing Questions. In: Journal of Marketing 15: 48-56
- Payne, Stanley L. (1951): The Art of Asking Questions. Princeton: University Press
- Porst, Rolf (2000): Praxis der Umfrageforschung. 2., überarbeitete Aufl. Teubner-Studien-skripten zur Soziologie 126. Stuttgart, Leipzig, Wiesbaden: Teubner
- Porst, Rolf (2014): Fragebogen. Ein Arbeitsbuch. 4. Aufl. Wiesbaden: VS Verlag
- Prüfer, Peter/Rexroth, Margrit (2005): Kognitive Interviews. ZUMA How-to-Reihe Nr. 15; [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/howto/How\\_to15PP\\_MR.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/How_to15PP_MR.pdf)
- Rugg, Donald (1941): Experiment in Wording Questions II. In: Public Opinion Quarterly: 91-92
- Rugg, Donald/Cantril, Hadley (1944): The Wording of Questions. In: Cantril, Hadley (Hg.): 23-50
- Schwarz, Norbert/Sudman, Seymour, Hg. (1996): Answering Questions. Methodology for Determining Cognitive and Communicative Processes in Survey Research. San Francisco: Jossey-Bass
- Strack, Fritz/Martin, Leonard L. (1987): Thinking, Judging, and Communicating: A Process Account of Context Effects in Attitude Surveys. In: Hippler/Schwarz/Sudman (Hg.): 123-148
- Sudman, Seymour/Bradburn, Norman M. (1982): Asking Questions. San Francisco: Jossey-Bass
- Sudman, Seymour/Bradburn, Norman M./Schwarz, Norbert (1996): Thinking About Answers. The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey-Bass

Axel Franzen

---

## 51.1 Einleitung

Jeder, der eine standardisierte Befragung plant, muss sich früher oder später überlegen, welche Fragen gestellt werden sollen, wie die Fragen formuliert sein sollen und wie die Antwortkategorien zu gestalten sind. Die Antworten auf diese drei Fragen sind nicht unabhängig, sondern hängen eng miteinander zusammen. In der Regel ist es sinnvoll in der genannten Reihenfolge vorzugehen, also sich erst mit den Inhalten und der Formulierung von Fragen zu beschäftigen und danach die Konstruktion der Antwortskalen vorzunehmen. Ein Forscher ist aber gut beraten, schon bei der Formulierung der Fragen an die spätere Datenauswertung zu denken, und spätestens bei der Festlegung der Antwortkategorien stehen Überlegungen zum Skalenniveau und den statistischen Analysemöglichkeiten im Vordergrund. Deshalb ist es durchaus sinnvoll, sich unabhängig von den Inhalten und Formulierungen der Fragen mit den Konstruktionsmöglichkeiten von Antwortkategorien zu beschäftigen. Die Wahl der Antwortskalen beeinflusst nämlich die Formulierung der Fragen. Aus diesem Grund ist es auch sinnvoll, allgemeine Überlegungen zur Formulierung von Fragen und zur Konstruktion der Antwortskalen in zwei Kapitel zu trennen, wie das in diesem Handbuch erfolgt (Porst, Kapitel 50 in diesem Band).

Die Fragen, die man in ein Erhebungsinstrument aufnimmt, hängen natürlich vom Untersuchungsgegenstand, dem Befragungsmodus und der Zielpopulation ab, an die sich die Befragung wendet. Fragebögen sind deshalb vor allem an die spezifischen Untersuchungszwecke anzupassen und das gilt natürlich auch für die Antwortskalen. Es ist deshalb nicht ganz einfach, allgemeine Regeln zur Konstruktion von Antwortskalen zu formulieren. Vielleicht wird das Thema deshalb in den einschlägigen Lehrbüchern zur empirischen Sozialforschung auch sehr kurz gehalten (Diekmann 2011; Schnell et al. 2011), lediglich die auf die Umfrageforschung spezialisierten Lehrbücher von Faulbaum et al. (2009) und Schnell (2012) beschäftigen sich ausführlicher mit Antwortskalen. In einer älteren Auflage des Lehrbuchs von Schnell et al. findet sich sogar die Aussage „Die Literatur zu

*Response-Errors* umfasst die langweiligsten Texte der empirischen Sozialforschung“ (2008: 420 [8. Auflage]). Vielleicht haben die Autoren sogar recht. Aber nichtsdestotrotz stehen Forscher häufig vor der Frage, wie eine Antwortskala auszusehen hat, um möglichst geringe Antwortverzerrungen auszulösen und dieses Kapitel versucht, hier einige Hilfestellungen zugeben.<sup>1</sup>

Bei der Durchführung einer Befragungsstudie kommen häufig weitere Aspekte hinzu, die man in Betracht ziehen sollte. Viele Untersuchungen werden erst dann so richtig interessant, wenn sie einen Trend beschreiben oder die Beobachtung der Untersuchungseinheiten über einen gewissen Zeitraum erlauben. Ein Trend- oder Paneldesign erfordert natürlich, dass die exakt gleichen Frageformulierungen und die exakt gleichen Antwortskalen wiederholt werden (Mochmann, Schupp, Kapitel 14 und 73 in diesem Band). Ansonsten ist es unmöglich, die realen Veränderungen von den methodisch induzierten Veränderungen zu trennen. Untersuchungen gewinnen in der Regel auch durch die Möglichkeit internationaler Vergleiche. Es kann sehr aufschlussreich sein, Trends einer Organisation oder einer Gesellschaft mit den Veränderungen in anderen Gesellschaften oder Organisationen zu vergleichen. Überlegungen zur Vergleichbarkeit können und sollten daher die Wahl von Antwortskalen mitbestimmen und können dazu führen, dass suboptimale Antwortskalen übernommen werden bzw. den Konstrukteur eines neuen Fragebogens mit schwierigen Abwägungsproblemen konfrontieren.

Sind die Probleme und Ziele hinsichtlich zeitlicher und internationaler Vergleichbarkeit geklärt, dann stellen sich bei der Konstruktion von Antwortskalen mindestens sechs Fragen:

1. Wie viele Antwortkategorien soll die Skala enthalten?
2. Soll eine gerade oder ungerade Anzahl verwendet werden?
3. Ist die Beschriftung aller Kategorien besser als nur die Bezeichnung der Endpunkte?
4. Sollte die Skalenbeschriftung bipolar oder unipolar erfolgen?
5. Sollte eine Skala positiv (zustimmend) oder negativ (ablehnend) beginnen?
6. Sollte die Verbalisierung itemspezifisch oder standardisiert vorgenommen werden?

Die Methodenforschung hat sich intensiv mit diesen Fragen beschäftigt. Dabei geht es darum, das Abschneiden unterschiedlicher Varianten von Skalen in Hinblick auf die zentralen Gütekriterien Objektivität, Reliabilität und Validität empirisch zu testen (Krebs/Menold, Kapitel 30 in diesem Band). Objektivität ist in standardisierten Befragungsstudien in der Regel kein Problem. Es gibt zwar in mündlichen Befragungen Intervieweffekte (Glantz/Michael, Kapitel 21 in diesem Band), aber es gibt keine plausiblen Hinweise, dass solche Probleme durch die Gestaltung von Antwortskalen wesentlich beeinflusst werden. Antwortskalen können sich dagegen in Bezug auf ihre Reliabilität unterscheiden. Gibt man den Befragten die Möglichkeit, ihre Meinungen anhand vielstufiger Skalen differenziert zu äußern, dann sollte eigentlich auch die Gefahr größer sein, dass bei Messwiederholungen

---

<sup>1</sup> Offensichtlich haben die Autoren ihre Meinung geändert. Wenigstens fehlt die zitierte Aussage in der überarbeiteten 9. Auflage.

Abweichungen auftreten und die Test-Retest-Reliabilität im Vergleich zu wenig Antwortmöglichkeiten geringer ausfällt (Eifler, Kapitel 11 in diesem Band). Erstaunlicherweise lässt sich aber empirisch das Gegenteil beobachten: Mehr Antwortkategorien erhöhen in der Regel die Reliabilität.

Weil Antwortskalen den Befragten Hinweise zur Interpretation der Fragen liefern können, dies allerdings auch in Bezug auf die vermeintlich erwünschte Antwort, beeinflussen sie natürlich auch die Validität (Krebs/Menold, Kapitel 30 in diesem Band). Allerdings sind Untersuchungen zur Reliabilität sehr viel einfacher zu bewerkstelligen (z.B. durch einfache Messwiederholungen) als Studien zur Validität, und letztere sind daher eher selten. Darüber hinaus lässt sich aber auch vermuten, dass die Validität eines Messinstruments in der Regel stärker von der Frage selbst und ihrer Formulierung abhängt als von der verwendeten Antwortskala. So gesehen erscheinen Tests zur Reliabilität auch aus inhaltlichen Erwägungen in der Methodenforschung zu Antwortskalen den wichtigsten Stellenwert einzunehmen.

Der zweite Abschnitt des Kapitels versucht, die wichtigsten Forschungsergebnisse zu den oben beschriebenen sechs Fragen zusammenzufassen. Einige Ergebnisse scheinen hinreichend konsistent und zuverlässig zu sein, so dass sich daraus Regeln für die Konstruktion von Antwortskalen formulieren lassen. Diese Regeln werden in der Zusammenfassung beschrieben.

---

## 51.2 Die Gestaltung von Antwortkategorien

### 51.2.1 Offene versus geschlossene Fragen

Antwortskalen liefern den Befragten Hinweise zur Interpretation und der Bedeutung von Fragen. Darüber hinaus können sie auch Hinweise über die vermeintlich sozial erwünschte Antwort liefern und damit die Verzerrung des Antwortverhaltens beeinflussen. Ein bekanntes Beispiel stammt aus einer Studie von Schwarz et al. (1985). In einem randomisierten Labor-Experiment wurden den Befragten zwei Versionen von Antwortskalen bei einer Frage nach ihrem täglichen Fernsehkonsum vorgelegt. In einer Version begann die vorgegebene Antwortskala mit einem kleinen Intervall (0 bis  $\frac{1}{2}$  Stunde, mehr als  $\frac{1}{2}$ , aber weniger als 1 Stunde etc.), in der anderen Version mit einem großen Intervall. In der letzteren Version ergab sich ein bedeutend höherer berichteter Fernsehkonsum als bei der ersten Variante. Weil die Schwarz-Studie schon etwas in die Jahre gekommen ist, kleine Fallzahlen aufweist und in einem Labor stattfand, haben wir die Studie in einer Online-Befragung unter Studierenden der Universität Bern im Jahr 2012 wiederholt. Die Frage lautete „Wie viele Stunden pro Woche sind Sie durchschnittlich sportlich aktiv? (Zählen Sie bitte auch Fortbewegungen mit dem Velo und zu Fuß dazu)“. Die Frage war in einen Online-Fragebogen zur Studienmotivation und zum Studierverhalten integriert, wobei je eine unterschiedliche Version von Antwortmöglichkeiten randomisiert den Befragten vorgelegt wurde. In der ersten Variante begannen die Antwortkategorien mit „bis zu einer

Stunde“, „ca. zwei Stunden“ und so weiter bis zu „mehr als 5 Std.“. In der zweiten Variante startete die Skala mit „bis zu 5 Stunden“, „ca. 6 Stunden“ und so weiter bis zu „mehr als 9 Stunden“. Die Ergebnisse sind in Tab. 51.1 dargestellt.

Während in der ersten Variante 78% der Befragten angaben, bis zu 5 Stunden sportlich aktiv zu sein, sind es in der zweiten Variante 44,5%, ein Unterschied von mehr als 30 Prozentpunkten. Die Studie repliziert damit die Ergebnisse der Schwarz-Studie und zeigt darüber hinaus, dass Antwortskalen auch außerhalb von Labors und selbst bei nicht-heiklen Fragen die Befragungsergebnisse beeinflussen können. Offensichtlich orientieren sich Befragte nicht nur bei sensitiven Fragen an der mittleren Kategorie einer vorgegebenen Antwortskala, weil sie dieses Verhalten als erwünscht oder erwartet interpretieren. Falls möglich, sollte aus diesem Grund auf die Vorgabe von Antwortkategorien verzichtet und stattdessen offenen numerischen Antwortmöglichkeiten der Vorzug gegeben werden. Für die offene numerische Abfrage spricht auch, dass die natürlichen Einheiten angegeben werden und damit die Variable auf dem höchstmöglichen Skalenniveau gemessen wird. Verhaltensfragen wie nach der Anzahl der Theaterbesuche im letzten Halbjahr oder der Anzahl der sozialen Kontakte sollten daher offen gestellt werden. Offene numerische Antwortmöglichkeiten liefern in diesen Fällen genaue Daten auf Absolutskalenniveau.

Die Vermeidung von Effekten der sozialen Erwünschtheit und die Realisierung des höchstmöglichen Skalenniveaus würden auch bei der Frage nach dem Einkommen eine offene Abfrage empfehlenswert erscheinen lassen. Allerdings treten bei der Einkommensfrage in der Regel die höchsten Antwortverweigerungen in einer Befragung auf. Offene numerische Abfragen können den Eindruck erwecken, dass es der Interviewer ganz genau wissen will und sie können damit die Verweigerungsrate bei der Einkommensfrage noch erhöhen (Engel/Schmidt, Kapitel 23 in diesem Band). Die Vermeidung von Item-Nonresponse kann natürlich ein sehr guter Grund sein, bei der Einkommensfrage doch geschlossene Antworten vorzugeben.

**Tab. 51.1** Effekte von Antwortskalen: Wie viele Stunden pro Woche sind Sie durchschnittlich sportlich aktiv?

Version 1	N	in %	Version 2	N	in %
bis zu 1 Std.	167	11,1	bis zu 5 Std.	690	44,5
ca. 2 Std.	287	19,1	ca. 6 Std.	280	18,1
ca. 3 Std.	305	20,3	ca. 7 Std.	213	13,7
ca. 4 Std.	210	14,0	ca. 8 Std.	118	7,6
ca. 5 Std.	201	13,4	ca. 9 Std.	78	5,0
mehr als 5 Std.	330	22,0	mehr als 9 Std.	171	11,0
Summe	1500	100%		1550	100%

### 51.2.2 Zur Anzahl der Antwortkategorien

Bei vielen Fragen sind allerdings offene numerische Antwortkategorien nicht einsetzbar. Dies trifft auf Fragen zu, bei denen es um die Erhebung von Einstellungen, Werten, Meinungen oder Einschätzungen geht. So wäre zum Beispiel die Frage nach der Sorge um den Umweltschutz oder nach der wirtschaftlichen Lage eines Landes mit einer offenen numerischen Antwort nur schwer umsetzbar. Außerdem ist hier im Unterschied zur Messung des Einkommens oder des Fernsehkonsums eine hohe Genauigkeit oft nicht möglich und auch nicht immer notwendig. Aus diesem Grund sind solche Fragen mit geschlossenen Antwortkategorien zu versehen. Hierbei wird zwischen ungeordneten und geordneten Antwortvorgaben unterschieden. Ein Beispiel für Fragen mit ungeordneten Antwortkategorien ist die Frage „Was machen Sie in Ihrer Freizeit?“, die dann von den Antwortvorgaben „Sport“, „lesen“, „Musik“ etc. gefolgt wird. Aus ungeordneten Antwortvorgaben resultieren in der Regel, wie auch in diesem Fall, Nominalskalen. Die Gestaltung von ungeordneten Antwortkategorien lässt sich kaum vom Inhalt der Frage trennen. Bei ungeordneten Antwortkategorien gelten nur drei Regeln, nämlich dass die Antwortkategorien erschöpfend, disjunkt und nicht zu umfangreich sein sollten.

Anders sieht es dagegen bei geordneten Antwortvorgaben aus, bei denen in der Regel das Ausmaß an Zustimmung oder Ablehnung zu einer Aussage, oder die Intensität der Beliebtheit oder des Vertrauens von Personen oder Organisationen gemessen wird. Hierbei stellt sich die Frage, wie viele Abstufungen geschlossene Antwortskalen (auch Ratingskalen genannt) enthalten sollen. In der Praxis (z.B. im ALLBUS 2012) sind fast alle Möglichkeiten anzutreffen, beginnend mit zwei Antwortkategorien über dreistufige, vierstufige, fünfstufige oder bis zu 10er und 100er Skalen. Im World Values Survey (WVS) werden Einstellungen und Werte überwiegend mittels vierstufiger Skalen erhoben. Im International Social Survey Programme (ISSP) finden sich dagegen hauptsächlich fünfstufige Antwortskalen, was die Vergleichbarkeit der Ergebnisse natürlich erschwert (Franzen/Vogl 2013).

In der methodischen Forschung wird vor allem untersucht, welche Skalen in Test-Retest-Untersuchungen die höchste Reliabilität aufweisen. Zunächst erscheint es plausibel, dass eine hohe Anzahl an Antwortkategorien den Befragten mehr Optionen eröffnet und eine feinere Abstufung und damit eine genauere Messung zulässt. Viele Abstufungen erhöhen aber auch die kognitiven Anforderungen an die Befragten. Nach Miller (1956) kann die optimale Anzahl an Antwortkategorien durch die Regel „sieben plus/minus zwei“ beschrieben werden. Dieses Ergebnis wird auch durch neuere Studien bestätigt (Preston/Colman 2000, Kieruj/Moors 2010, Svensson 2000). Abgeraten wird demnach von Skalen mit weniger als fünf Antwortkategorien, die paradoxe Weise über eine geringere Reliabilität verfügen. Skalen mit mehr als 7 oder 9 Abstufungen bringen dagegen keine weiteren Vorteile. Zusätzlich wird davon ausgegangen, dass Befragte mit höherem Bildungshintergrund besser in der Lage sind, mit ausdifferenzierten Skalen umzugehen. Für allgemeine Bevölkerungsbefragungen sollten dagegen eher weniger Kategorien verwendet werden (Weijters et al. 2010). Einen Einfluss auf die Abstufungen hat dabei auch der Befragungsmodus. Schriftliche Befragungen erlauben mehr Antwortkategorien, während

in mündlichen (face-to-face) oder telefonischen Befragungen die Teilnehmer besser mit weniger Kategorien zurechtkommen. Alles in allem sprechen die einschlägigen Ergebnisse dafür, in allgemeinen Bevölkerungsbefragungen fünfstufige oder siebenstufige Skalen zu verwenden (vgl. auch Faulbaum et al. (2009) für eine ähnliche Empfehlung).

### **51.2.3 Gerade oder ungerade Antwortskalen?**

Ausführlich wird in der Literatur auch die Frage diskutiert, ob eine gerade oder ungerade Anzahl von Antwortkategorien die Reliabilität erhöht. Eine gerade Anzahl von Antwortkategorien zwingt die Befragten zu einer eher zustimmenden oder ablehnenden Entscheidung. Ungerade Antwortmöglichkeiten enthalten dagegen eine Mitte. Diese ist allerdings nicht immer eindeutig zu interpretieren. Sie könnte auch von Befragten gewählt werden, die zum Thema keine Meinung haben, was dann zu Fehlmessungen führen würde. Eine gerade Anzahl birgt dagegen die Gefahr, der Akquieszenz Vorschub zu leisten, also der Tendenz der Befragten, eine Frage im Zweifelsfall zu bejahen. Auch dies würde einen Messfehler verursachen. Empirisch sprechen die Ergebnisse eher für ungerade Skalen. Bei einem Vergleich zwischen 5-stufigen und 6-stufigen Antwortkategorien konnte Moors (2008) allerdings keine wesentlichen Unterschiede erkennen. O'Muircheartaigh et al. (2000) berichten dagegen eine bessere Reliabilität für Skalen mit einer mittleren Kategorie.

Um zu vermeiden, dass Befragte die mittlere Kategorie wählen, weil sie damit ihre Meinungslosigkeit zum Ausdruck bringen wollen, wird gelegentlich die explizite Verwendung einer „weiß nicht“ Kategorie empfohlen (Converse/Presser 1986). Dies wird in der einschlägigen Literatur aber sehr kritisch gesehen (vgl. den ausführlichen Literaturüberblick in Krosnick 1999). Eine explizite „weiß nicht“ Kategorie erhöht den Anteil an Befragten, die angeben, keine Meinung zu einem Thema zu haben. Dieser höhere Anteil ist aber nur teilweise auf wirkliche Meinungslosigkeit zurückzuführen. Darüber hinaus wird die „weiß nicht“ Option aber auch gerne genutzt, um den kognitiven Aufwand zu reduzieren, eine Frage zu verstehen und zu beantworten. So wird sie beispielweise von einigen Befragten selbst bei völlig unterschiedlichen Frageinhalten konsistent genutzt. Das explizite Hinweisen auf eine „weiß nicht“ Option könnte Interviewten auch suggerieren, dass sie Experten sein sollten, um Fragen zu beantworten. Insbesondere Befragte mit geringem Bildungshintergrund fühlen sich deshalb entmutigt oder verunsichert. Insgesamt zeigen die Studien, dass explizite „weiß nicht“ Kategorien die Reliabilität nicht erhöhen.

### **51.2.4 Die Beschriftung der Antwortkategorien**

Eine Beschriftung oder Verbalisierung aller Kategorien einer Antwortskala ist nur bei Skalen bis zu 7 oder 9 Kategorien sinnvoll. Bei längeren Skalen ist es schwierig, eine angemessene verbale Abstufung zu konstruieren. In diesem Fall werden in der Regel nur die Endpunkte mit z.B. „stimme sehr stark zu“ oder „stimme überhaupt nicht zu“ bezeichnet.

Auf den ersten Blick erscheinen nummerierte Antwortkategorien präzisere Abstufungen zu ermöglichen als verbale Differenzierungen. Empirische Studien zeigen aber, dass die semantische Kennzeichnung der einzelnen Antwortkategorien zu einer besseren Reliabilität führt (z.B. Krosnick 1999, Tourangeau et al. 2007). Falls eine Beschriftung aufgrund der Abstufungen möglich ist, dann sollte sie auch gewählt werden.

Bei der Verbalisierung der Kategorien von Ratingskalen ist darauf zu achten, dass die Bezeichnungen möglichst gleiche Intervallabstände nahelegen. Das ist bei fünf oder siebenstufigen Skalen leicht durchführbar. Die Antwortkategorien bei Zustimmungsfragen lauten dann z.B. „stimme sehr zu“, „stimme eher zu“, stimme teilweise zu/teilweise nicht“, „stimme eher nicht zu“ und „stimme überhaupt nicht zu“. Bei siebenstufigen Skalen würde man die Skala um extremere Endpunkte z.B. „stimme absolut zu“ und „stimme absolut nicht zu“ erweitern. Falls in schriftlichen Befragungen gleichzeitig numerische Werte präsentiert werden, dann sollten sie gleichsinnig vergeben werden, also in unserem Fall die Ziffern 5 der Ausprägung „stimme sehr zu“ und die Ziffer 1 der Ausprägung „stimme überhaupt nicht zu“ zugeordnet werden. Streng genommen handelt es sich bei Ratingskalen um ordinale Messniveaus. Ob die Skalenabstände darüber hinaus gleich sind und die Skala damit Intervallskalenniveau aufweist, ist eine empirische Frage, die durch geeignete statistische Verfahren (z.B. Item-Response-Theorie) getestet werden kann (Geiser/Eid 2010, Rost 2004). Es empfiehlt sich auch, bei der statistischen Auswertung die Ergebnisse der Analyseverfahren für intervallskalierte Daten (z.B. OLS-Regressionen) mit Ergebnissen für ordinale Daten (z.B. logistische Regressionen) zu vergleichen (Blasius/Baur, Kapitel 79 in diesem Band). Auf der sicheren Seite ist man dann, wenn unterschiedliche Analyseverfahren zu vergleichbaren Ergebnissen kommen.

### 51.2.5 Bipolare oder unipolare Beschriftung?

Eine bipolare Beschriftung von Antwortskalen verwendet das semantische Differential (Gegensatzpaare). Diese sogenannten Polaritätsprofile werden gerne in der Psychologie und der Marktforschung eingesetzt, in der es häufig um die Einschätzung bzw. Zuschreibung von Eigenschaften zu Personen oder Marken geht. Beispiele für Gegensatzpaare sind „stark – schwach“, „richtig – falsch“ oder „gesund – ungesund“. In der soziologisch orientierten Sozialforschung sind Polaritätsprofile eher selten anzutreffen. Bei einer unipolaren Bezeichnung wird dagegen die sprachliche Abstufung eindimensional erreicht. Werden die Kategorien etwa mit „stimme sehr stark zu“ bis „lehne sehr stark ab“ bezeichnet, so handelt es sich um eine bipolare Beschriftung. Unipolar wird die Abstufung mit den Labels „stimme sehr stark zu“ bis „stimme überhaupt nicht zu“ erreicht. Befragte kommen bei einer Befragung besser mit möglichst einfachen Antwortskalen zurecht, die schnell erlernt werden können und die nicht bei jeder Frage neu beachtet werden müssen. Diese Regel, Befragungen möglichst einfach zu halten, spricht daher für die Verwendung unipolarer Skalen. Dies scheinen auch empirische Studien zu bestätigen (Schaeffer/Presser 2003).

Die Beschriftung durch semantische Differenziale kann ebenso wie eine unipolare Bezeichnung mit positiven numerischen Werten kombiniert werden (beispielweise von 1 bis 7) oder aber auch negative Ziffern umfassen (z.B. von -3 bis +3). Empirische Studien haben aber gezeigt, dass Befragte zusätzlich zu der Verbalisierung den verwendeten Ziffern eine Bedeutung zuschreiben. So werden negative Ziffern eher gemieden als positive Zahlen (Schwarz et al. 1991). Gleichmäßige Abstufungen werden daher besser erreicht, wenn unipolare Formulierungen auch mit einem positiven Wertebereich der Ziffern kombiniert werden.

### **51.2.6 Von positiv zu negativ oder umgekehrt?**

Bei beiden Formaten (unipolar oder bipolar) stellt sich die Frage, ob eine Antwortskala zuerst die positiven oder zustimmenden Kategorien verwenden oder aber mit den negativen oder ablehnenden Alternativen beginnen sollte. Hierzu gibt es eine Hypothese aus der kognitiven Psychologie. Optimalerweise sollten Befragte zunächst alle Antwortmöglichkeiten abwarten (lesen oder hören) und dann diejenige wählen, die ihrer Meinung am besten entspricht. Diese Maximierung oder Optimierung des Antwortverhaltens wird aber häufig zugunsten von „satisficing“ aufgegeben. Beim satisficing wählen Befragte unmittelbar diejenige Alternative, die als erstes akzeptabel erscheint. Ein solches Verhalten würde implizieren, dass die Reihenfolge einer Antwortskala (z.B. von positiv zu negativ) von Bedeutung sein kann. Krosnick (1999) führt eine Vielzahl von Studien auf, in denen Reihenfolgeeffekte besonders bei ungeordneten mehrkategorialen Antwortskalen auftraten. Werden unterschiedliche Kategorien als Antwort vorgelegt, dann werden besonders in selbstdurchführten Befragungen die zuerst genannten Antwortmöglichkeiten häufiger gewählt („primacy effect“). In telefonischen und mündlichen Interviews werden dagegen die zuletzt vorgelesenen Antworten besser im Gedächtnis behalten und häufiger gewählt, was als „Recency-Effekt“ bezeichnet wird. Bei Ratingskalen (geordnete Antwortvorgaben) scheinen diese Effekte dagegen weniger stark ausgeprägt zu sein und die Reihenfolge für das Antwortverhalten keine entscheidende Rolle zu spielen (Krebs/Hoffmeyer-Zlotnik 2010).

### **51.2.7 Standardisierte Antwortskalen oder itemspezifische Skalen?**

Im Prinzip spricht die Regel, Befragungen möglichst einfach zu halten, für die Verwendung der gleichen Antwortskalen bei allen Fragen. Die Fragen sollten also so formuliert sein, dass beispielsweise immer eine bestimmte (z.B. fünfstufige) Zustimmungsskala verwendet werden kann. Dennoch zeigen einige Studien, dass itemspezifische Antwortskalen die Reliabilität der Messung erhöhen können (Krebs 2011). Itemspezifische Antwortskalen sind möglichst genau an die Formulierung der Frage angepasst und erleichtern dadurch das Verständnis der Antwortskalen. Die Verwendung immer gleicher Skalen könnte auch zur Ermüdung der Befragten führen. Allerdings sind die Evidenzen für itemspezifische Antwortskalen noch unsicher.

### 51.3 Zusammenfassung

Zusammenfassend lassen sich aus dem Literaturüberblick sieben Regeln zur Gestaltung von Antwortkategorien ableiten:

1. Falls möglich, sollten keine geschlossenen Antwortkategorien vorgegeben werden, sondern die Antworten offen numerisch erhoben werden.
2. Antwortkategorien sollten disjunkt und erschöpfend sein.
3. Antwortkategorien sollten das höchstmögliche Skalenniveau realisieren.
4. Die optimale Anzahl an Antwortkategorien ist sieben plus/minus zwei.
5. Antwortkategorien sollten ungerade sein.
6. Falls möglich, sollte jede Antwortkategorie beschriftet sein.
7. Zur Vermeidung von Antwortmustern sollte die Reihenfolge von positiv zu negativ formulierten Antwortkategorien wechseln.

Aber: keine Regel ohne Ausnahme. Mit Ausnahme von Regel zwei kann es gute Gründe geben, von den formulierten Empfehlungen abzuweichen. Die Einkommensfrage wurde als mögliche Ausnahme schon erwähnt. Auch bei Fragen zur Zufriedenheit mit verschiedenen Lebensbereichen oder der allgemeinen Lebenszufriedenheit haben sich beispielsweise 11er Skalen (von 0 bis 10) etabliert und bewährt. Die zweite Regel ist dagegen eine Mindestanforderung an Antwortskalen, bei der keine Ausnahmen gemacht werden sollten. Schließlich ist darauf zu achten, dass die Gestaltung von Antwortskalen vom Befragungsmodus und natürlich von der Zielpopulation abhängen. Es ist deshalb empfehlenswert, neue Erhebungsinstrumente vor der Feldphase einem gründlichen Pretest zu unterziehen.

## Literatur

- Converse, Jean M./Presser, Stanley (1986): Survey Questions. Beverly Hills: Sage
- Diekmann, Andreas (2011): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. 5. Aufl. Reinbek bei Hamburg: Rowohlt
- Faulbaum, Frank/Prüfer, Peter/Rexroth, Margit (2009): Was ist eine gute Frage? Die systematische Evaluation der Fragequalität. Wiesbaden: VS Verlag
- Franzen, Axel/Vogl, Dominikus (2013): Acquiescence and the Willingness to Pay for Environmental Protection: A Comparison of the ISSP, WVS, and EVS. In: Social Science Quarterly 94: 637-659
- Geiser, Christian/Eid, Michael (2010): Item-Response-Theorie. In: Wolf/Best (Hg.): 311-332
- Kieruj, Natalia D./Moors, Guy (2010): Variations in Response Style Behavior by Response Scale Format in Attitude Research. In: International Journal of Public Opinion Research 22: 320-342
- Krebs, Dagmar (2011): Vortrag am Jahrestreffen der European Social Research Association (ESRA) in Lausanne, Schweiz
- Krebs, Dagmar/Hoffmeyer-Zlotnik, Juergen H.P. (2010): Positive First or Negative First? In: Methodology 6: 118-127
- Krosnick, Jon A. (1999): Survey Research. In: Annual Review of Psychology 50: 537-567
- Miller, George A. (1956): The Magical Number Seven Plus or Minus Two: Some Limits on our Capacity for Processing Information. In: Psychological Review 63: 81-97
- Moors, Guy (2008): Exploring the Effect of Middle Response Category on Response Style in Attitude Measurement. In: Quality and Quantity 42: 779-794
- O'Muircheartaigh, Colm/Krosnick, Jon A./Helic, Armin (2000): Middle Alternatives, Acquiescence, and the Quality of Questionnaire Data. Working paper. University of Chicago
- Preston, Carolyn C./Colman, Andrew M. (2000): Optimal Number of Response Categories in Rating Scales: Reliability, Validity, Discriminating Power, and Respondent Preferences. In: Acta Psychologica 104: 1-15
- Rost, Jürgen (2004): Lehrbuch Testtheorie – Testkonstruktion. Bern: Huber
- Schaeffer, Nora C./Presser, Stanley (2003): The Science of Asking Questions. In: Annual Review of Sociology 29: 65-88
- Schnell, Rainer (2012): Survey-Interviews: Methoden standardisierter Befragungen. Wiesbaden: VS Verlag
- Schnell, Rainer/Hill, Paul B./Esser, Elke (2011): Methoden der empirischen Sozialforschung. 9. Auflage. München: Oldenbourg
- Schwarz, Norbert/Hippler, Hans-J./Deutsch, Brigitte/Strack, Fritz (1985): Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgements. In: Public Opinion Quarterly 49: 388-395

- Schwarz, Norbert/Knäuper, Bärbel/Hippler, Hans J./Noelle–Neumann, Elisabeth/Clark, Leslie (1991): Rating Scales: Numeric Values May Change the Meaning of Scale Labels. In: *Public Opinion Quarterly* 55: 570-582
- Svensson, Elisabeth (2000): Concordance between Ratings Using Different Scales for the Same Variable. In: *Statistics in Medicine* 19: 3483-3496
- Tourangeau, Roger/Couper, Mick P/Conrad, Frederick (2007): Color, Labels and Interpretive Heuristics for Response Scales. In: *Public Opinion Quarterly* 71: 91-112
- Weijters, Bert/Cabooter, Elke/Schillewaert, Niels (2010): The Effect of Rating Scale Format on Response Style: The Number of Response Categories and Response Category Labels. In: *International Journal of Research in Marketing* 27: 236-247
- Wolf, Christof/Best, Henning (Hg.) (2010): *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS Verlag

Cornelia Züll und Natalja Menold

---

## 52.1 Begriffsdefinition

Lange Zeit standen offene Fragen in standardisierten Interviews eher im Hintergrund, da die Analyse als sehr aufwändig im Vergleich zum Zugewinn an Information galt. Mit dem Aufkommen von Online-Umfragen wird die Erhebung von offenen Fragen aber deutlich erleichtert, denn diese können hier kostengünstiger erhoben werden und stehen unmittelbar danach für die Auswertung zur Verfügung.

Als offene Fragen bezeichnet man all jene, bei denen es keine Antwortvorgaben gibt: Der Befragte kann in eigenen Worten antworten, so wie er es gewohnt ist zu sprechen. Kategorien sind nicht vorgegeben. Die Frage wird je nach Interview-Modus entweder vom Interviewer vorgelesen und die Antwort wird protokolliert oder die Frage ist abgedruckt/ erscheint am Bildschirm und der Befragte schreibt selbst seine Antwort in ein vorgegebenes Textfeld. Abb. 52.1 zeigt ein Beispiel zu einer offenen Frage aus dem ALLBUS 2008 (persönlich-mündliches Interview; Quelle: <http://www.gesis.org/allbus>).

Bei der Entwicklung eines Fragebogens stellt sich nun die Frage, ob und wann eine offene Frage sinnvoll ist. Eine allgemeingültige Regel kann hierzu nicht gegeben werden, sondern die Antwort hängt von verschiedenen Rahmenbedingungen ab. Dabei sollte eine Entscheidung für oder gegen eine offene Frage nicht so sehr am Aufwand der Analyse oder an persönlichen Präferenzen für eine Frageform festgemacht, sondern vielmehr von der Forschungsfrage und vom Forschungsstand abhängig gemacht werden.

---

## 52.2 Ziele

Offene Fragen sind in zwei Fällen sinnvoll: zum Informationsgewinn und zur Motivation des Befragten (Porst 2011). Werden Sie zur *Informationsgewinnung* eingesetzt, kann man verschiedene Situationen unterscheiden:

Würden Sie mir bitte sagen, was Sie mit dem Begriff „links“ verbinden?

⇒ Antworten bitte genau notieren!

✉:

---

---

---

**Abb. 52.1** Auszug aus dem Fragebogen des ALLBUS 2008

1. Bei der Fragebogenentwicklung kann man *den Befragungsgegenstand noch nicht eingrenzen oder erwartet neue Aspekte herauszufinden*. Ein Beispiel dazu ist die Frage aus dem ALLBUS 2008 zum Verständnis von links bzw. rechts. Es ist zwar aus anderen Studien bekannt, was Befragte in den 1980er- und 1990er-Jahren damit verbunden, aber es besteht die Annahme, dass die beiden Dimensionen an Bedeutung verlieren und sich das Bild in der Bevölkerung gewandelt hat: Links und rechts wird mit anderen Dimensionen/Themen in Verbindung gebracht als noch vor 20 Jahren. Um dieses neue Spektrum zu erheben, bietet sich eine offene Frage an.
2. *In kognitiven Pretests werden offene Fragen häufig als Nachfrage verwendet, um festzustellen, warum ein bestimmter Skalenwert gewählt wurde oder um zu überprüfen, ob der Sinn einer Frage verstanden wurde* (vgl. Weichbold, Kapitel 19 in diesem Band).
3. *Das Spektrum der möglichen Antworten ist sehr groß und kann nicht mit vorgegebenen Kategorien abgefragt werden*. Ein Beispiel dazu ist die Frage nach der beruflichen Tätigkeit. Es ist zwar bekannt, welche Berufe es gibt, aber eine Liste von mehreren hundert Antwortmöglichkeiten kann einem Befragten natürlich nicht vorgelegt werden.
4. *Man möchte vermeiden, den Befragten in eine bestimmte Richtung zu lenken*. Porst (2011: 67, Kapitel 50 in diesem Band) nennt als Beispiel die Frage nach der besten Rockband aller Zeiten. Die Zahl der Gruppen ist sehr groß und eine Auswahl der bekanntesten Bands als Kategorien gibt zunächst einmal nur eigene Vorstellungen des Fragebogenentwicklers vor. Der Befragte wird in Richtung dieser Vorgaben gelenkt und man risiert, dass Kategorien genannt werden, einfach weil sie vorgegeben waren und nicht, weil sie die wirkliche Präferenz/Meinung des Befragten wiedergeben.
5. Offene Fragen eignen sich zur *Abfrage von Wissen* besser als geschlossene Fragen, weil sie nicht nur die Wahrscheinlichkeit, durch das Raten eine richtige Antwort zu erzielen, minimieren, sondern auch häufig zu reliableren und validierenden Angaben führen (Krebs/Menold, Kapitel 30 in diesem Band). Bei offenen Fragen bekommt man aber auch mehr „weiß nicht“-Antworten oder Verweigerungen, wenn entweder die Befragten die Antwort zwar wissen, aber die Beantwortung Nachdenken, d.h. Aufwand, erfordert oder

---

wenn sich die Befragten bei der Beantwortung der Frage nicht sicher sind (vgl. Krosnick/Presser 2010).

Neben der Informationsgewinnung können offene Fragen zur *Motivation der Befragten* in einer Umfrage eingesetzt werden. Dem Befragten wird die Möglichkeit gegeben, zwischen vielen geschlossenen Fragen und vorgegebenen Antwortkategorien auch einmal frei und in eigenen Worten seine Meinung äußern zu können. Zudem wird häufig auf der letzten Seite eines Fragebogens eine offene Frage der Form gestellt „Gibt es von Ihrer Seite noch etwas, das Sie zu unserer Umfrage sagen möchten?“ oder „Haben Sie noch Anmerkungen zum Thema unserer Befragung?“ (siehe Porst 2011: 157).

---

### 52.3 Die Gestaltung von offenen Fragen

Das Beantworten offener Fragen stellt höhere Ansprüche an die kognitiven Fähigkeiten eines Befragten als das Auswählen einer Kategorie. Die Antwort erfordert die Formulierung in eigenen Worten, was nicht nur eine Bereitschaft hierzu, sondern auch die Fähigkeit zur Verbalisierung notwendig macht. In der Regel sind alle Befragten in der Lage, derartige Fragen zu beantworten, wichtig ist aber das Interesse des Befragten am Thema der Untersuchung (Holland/Christian 2009): Interessierte Befragte sind eher bereit, offene Fragen zu beantworten. Der Befragte muss bei einer offenen Frage sehr klar verstehen, was von ihm erwartet wird. Dies betrifft nicht nur den Rahmen für mögliche Antworten, sondern auch den Umfang der Antworten. Es empfiehlt sich immer, den Befragten darauf hinzuweisen, welche Antwortform man erwartet (nur Stichwörter, eine Angabe, mehrere Angaben/eine Auflistung, eine kurze Beschreibung etc.).

Bei persönlich-mündlichen Interviews kommt dem Interviewer bei offenen Fragen eine wichtige Rolle zu: Der Interviewer kann durch Zusatzinformationen und Nachfragen dem Befragten helfen und ihn zu besseren und vollständigeren Antworten motivieren (Dillman/Christian 2005, Glantz/Michael, Kapitel 21 in diesem Band). Er ist aber auch für die Niederschrift der Antwort verantwortlich, d.h. er muss das von Befragten Geäußerte auf den Fragebogen (möglichst ohne Veränderung) übertragen.

---

### 52.4 Auswertung offener Fragen

Offene Fragen wurden in der Vergangenheit häufig gemieden, denn nicht nur für den Befragten stellen sie eine höhere Bürde da, sondern auch für den Forscher, denn die Daten müssen vor der eigentlichen Analyse zunächst codiert werden. Dies führt häufig dazu, dass offene Angaben zwar erhoben, aber letztendlich in der Analyse nicht verwendet oder nur einige interessante Zitate in den Texten gesucht und berichtet werden. Ein solches Vorgehen kann jedoch mit einer wissenschaftlichen Vorgehensweise nicht vereinbart werden. Im Folgenden sollen einige Analyse-Ansätze zur Codierung und Analyse von offenen Fragen vorgestellt werden.

### 52.4.1 Quantitative Inhaltsanalyse

Die klassische Auswertungsmethode für offene Fragen ist die quantitative Inhaltsanalyse (Mayring/Fenzl, Kapitel 38 in diesem Band). Dabei werden die offenen Antworten nach Vorgabe eines Kategorienschemas durch einen oder mehrere Codierer codiert. Jede Kategorie repräsentiert einen Bedeutungsaspekt des zu codierenden Gegenstandes und wird durch eine Definition und aussagekräftige Beispiele aus dem Datenmaterial, sog. Ankerbeispielen, beschrieben (Früh 2007). Die Entwicklung des Kategorienschemas kann deduktiv oder induktiv z.B. mit Hilfe einer qualitativen Inhaltsanalyse erfolgen (Mayring/Fenzl, Kapitel 38 in diesem Band). Das Kategorienschema sollte dabei folgenden Anforderungen genügen (Früh 2007):

- Die Kategorien sollten mit den Zielen der Untersuchung/der Forschungsfrage korrespondieren.
- Das Kategoriensystem sollte vollständig sein, d.h. es sollte den durch die Forschungsfrage vorgegebenen Bedeutungsaspekt sowohl auf der Begriffsebene als auch auf der Ebene des Datenmaterials vollständig erfassen.
- Die Kategorien sollten voneinander trennscharf, d.h. wechselseitig exklusiv sein. Dazu sind klare Definitionen und ggf. Abgrenzungen zwischen einzelnen Kategorien notwendig.
- Die Kategorien sollten eindeutig sein.

Ein Beispiel für ein Kategorienschema mit Definitionen und Ankerbeispielen findet sich in Züll et al. 2010. Im zweiten Schritt der Inhaltsanalyse folgten auf Basis des Kategorienschemas, das die Codierregeln vorgibt, die Codiererschulung und erste Probecodierungen. Darauf aufbauend kann das Schema ggfs. noch einmal überarbeitet werden.

Daran anschließend folgt die eigentliche Codierung des Datenmaterials. Die Codierung kann auf Papier oder direkt am Computer in einem Statistikprogramm, in einem Excel-Formular oder in spezieller Software wie z.B. MaxQda (<http://www.maxqda.de>) erfolgen. Das Ergebnis der Codierung (ein oder mehrere Codes pro Antwort) kann direkt zu den anderen Variablen der Umfrage hinzugefügt und gemeinsam mit ihnen analysiert werden.

Im Anschluss an die Codierung sollte die Reliabilität der Codierungen überprüft werden. Dazu wird eine Stichprobe des Materials (abhängig vom Umfang des Materials, z.B. 20% aller offenen Angaben) von einem zweiten Codierer codiert, und es wird ein Reliabilitätsmaß berechnet, das zur Aussage über die Qualität der Codierung verwendet werden kann. Dazu stehen verschiedene Maße, z.B. ein einfaches Verhältnismaß, Cohen's Kappa, Scott's Pi oder Krippendorff's Alpha, zur Verfügung (siehe z.B. Früh 2007). Ist die Reliabilität gering, wird das Kategorienschema überarbeitet und das Datenmaterial erneut kodiert.

### 52.4.2 Computerunterstützte Inhaltsanalyse

Will man die Codierung der offenen Angaben weitgehend automatisieren, empfiehlt sich die Anwendung der computerunterstützten Dictionär-basierten Inhaltsanalyse. Sie unterscheidet sich von der oben beschriebenen Inhaltsanalyse dadurch, dass die Codierung auf Basis eines inhaltsanalytischen Dictionärs/Wörterbuchs erfolgt und ein spezielles Computerprogramm zur Codierung der Texte herangezogen wird. Das Dictionär entspricht dem oben beschriebenen Kategorienschema und enthält die Codierregeln in Form von Wortlisten. Anstelle einer verbalen Definition und anstelle von Ankerbeispielen werden Wörter und Phrasen definiert, die eindeutige Indikatoren für eine Kategorie sind. Immer wenn eines dieser Wörter/Phrasen in einer Antwort auftritt, wird der entsprechende Code vergeben. Als Beispiel für eine solche Kategoriedefinition sei die Kategorie „Haushaltssolidität“ aus dem Kategorienschema zur Bedeutung von „Links“ und „Rechts“ genannt. Die Wortliste umfasst hier „Abbau der Staatsverschuldung“, „Haushaltsdisziplin“, „Haushaltssolidität“, „keine Staatsverschuldung“ und „solide Finanzpolitik“. Eine detaillierte Beschreibung zum Vorgehen bei einer computerunterstützten Inhaltsanalyse offener Fragen findet sich bei Züll/Mohler (2001).

Einige Vorkehrungen bei der Fragestellung vereinfachen und verbessern die computerunterstützte Codierung:

- Je enger und klarer die Frage formuliert ist, desto einfacher die Analyse.
- Die Fragen sollten von vornherein darauf beschränkt werden, was wirklich analysiert werden soll.
- Wenn möglich, sollte eine positive und negative Einstellung/Meinung in zwei Fragen getrennt abgefragt werden.

Das Ergebnis der Codierung sind ein oder mehrere Codes pro Antwort, die wie bei der oben beschriebenen Inhaltsanalyse direkt zu den anderen Variablen der Umfrage hinzugefügt werden können. Wie bei einer Inhaltsanalyse sollte auch bei einer computerunterstützten Inhaltsanalyse die Qualität der Codierung überprüft werden. In diesem Fall wird eine Stichprobe der Texte von einem menschlichen Codierer codiert, und es wird wie oben beschrieben die Übereinstimmung zwischen per Hand und maschinell erstellen Codierungen berechnet.

### 52.4.3 Sonstige Analyse-Ansätze

Es gibt Versuche, die Codierung von Texten mit Hilfe von halbautomatischen Verfahren durchzuführen. Dazu gehört das von Giorgetti/Sebastiani (2003) vorgeschlagene Verfahren des „supervised machine learning“, bei dem auf Basis eines zunächst von Hand codierten Textbeispiels das Programm automatisch weitere Texte entsprechend codiert.

Neben der Codierung der Texte gibt es verschiedene Analyse-Verfahren für Antworten auf offene Fragen, die die Texte direkt auswerten. Diese Ansätze verfolgen das Ziel der direkten Analyse von offenen Angaben, ohne jedem einzelnen Befragten (jeder einzelnen Antwort) einen oder mehrere Codes zuzuweisen. Dazu gehören Verfahren wie die Co-Occurrence-Analyse, bei der das gemeinsame Auftreten von Wörtern in einer Antwort als Analysegrundlage dient. Die dadurch erzeugte Ähnlichkeits- oder Distanzmatrix wird dann in einer Clusteranalyse, Korrespondenzanalyse oder in einer Multidimensionalen Skalierung weiter analysiert (Kronberger/Wagner 2002). Ein ähnliches Vorgehen stellt das von Jackson/Trochim (2002) vorgeschlagene Verfahren des „conceptual mapping“ dar.

## Literatur

- Bauer, Martin W./Gaskell, George (Hg.) (2002): Qualitative researching with text, image and sound. A practical Handbook. Thousand Oaks, CA: Sage
- Dillman, Don A./Christian, Leah M. (2005): Survey mode as a source of instability in responses across surveys. In: *Field Methods* 17: 30-52
- Früh, Werner (2007): Inhaltsanalyse. Theorie und Praxis. Konstanz: UVK Medien
- Giorgetti, Daniela/Sebastiani, Fabrizio (2003): Automatic survey coding by multiclass text categorization techniques. In: *Journal of the American Society for Information Science and Technology* 54: 1269-1277
- Holland, Jennifer L./Christian, Leah M. (2009): The influence of topic interest and interactive probing on responses to open-ended questions in web surveys. In: *Social Science Computer Review* 27: 197-212
- Jackson, Kristin M./Trochim, William M. K. (2002): Concept mapping as an alternative approach for the analysis of open-ended survey responses. In: *Organizational Research Methods* 5: 307-336
- Kronberger, Nicole/Wagner, Wolfgang (2002): Keywords in Context. Statistical Analysis of text features. In: Bauer/Gaskell (Hg.): 299-317
- Krosnick, Jon A./Presser, Stanley (2010): Question and questionnaire design. In: Wright/Marsden (Hg.): 263-314
- Marsden, Peter V./Wright, James D. (Hg.) (2010): Handbook of survey research. Bingley, UK: Emerald Group Publishing Limited
- Porst, Rolf (2011): Fragebogen. Ein Arbeitsbuch. Wiesbaden: VS Verlag
- Züll, Cornelia/Scholz, Evi/Schmitt, Hermann (2010): Kategorienschema für die Vercodung von Assoziationen mit den politischen Richtungsbegriffen „links“ und „rechts“. GESIS Technical Report 2010/19. Bonn: GESIS
- Züll, Cornelia/Mohler, Peter P. (2001): Computerunterstützte Inhaltsanalyse. Codierung und Analyse von Antworten auf offene Fragen. GESIS-How-To-Reihe 8. Mannheim: GESIS

Hermann Dülmer

---

## 53.1 Einleitung

Beim faktoriellen Survey (Vignettenanalyse) handelt es sich um ein experimentelles Design (Eifler, Kapitel 11 in diesem Band), bei dem der Forscher variiierende Situations- oder Personenbeschreibungen, die sogenannten Vignetten, zu einem Thema erstellt und unter einem bestimmten Gesichtspunkt beurteilen lässt. Die vom Forscher festgelegten Situations- oder Personenbeschreibungen, wie etwa das Geschlecht, die Hautfarbe, die Deutschkenntnisse, die Konfession und die Erwerbstätigkeit einer fiktiven Person bilden die unabhängigen Variablen (X-Variablen, experimentelle Einwirkungen) für das zu erklärende Urteilsverhalten (Y-Variable, Messung), also etwa dafür, wie sehr ein Befragter der jeweils beschriebenen Vignettenperson vertrauen würde. Beurteilt jeder Befragte mehr Vignetten als es X-Variablen für die Vignetten gibt, dann lässt sich für jeden einzelnen Befragten die Wichtigkeit, die er oder sie den entsprechenden Vignettenmerkmalen bei der Beurteilung der Vignetten bemisst, mit Hilfe der Regressionsanalyse schätzen. Werden die Vignetten von vielen Befragten beurteilt, dann lässt sich über alle Befragten hinweg auch der Einfluss schätzen, den bestimmte Befragtenmerkmale als weitere unabhängige Variablen auf das Urteilsverhalten (Y-Variable) ausüben. Damit lassen sich dann beispielsweise Fragen danach beantworten, ob es Unterschiede zwischen weiblichen und männlichen Befragten hinsichtlich des Vertrauens in die auf den Vignetten beschriebenen Personen gibt.

Ein Vorteil der Beurteilung konkreter Vignettenbeschreibungen besteht darin, dass dies unserem Urteilsverhalten im alltäglichen Leben erheblich näher kommt als die Beantwortung allgemeiner, häufig abstrakter oder aus dem Kontext herausgelöster Fragen. Dies dürfte vor allem der Gültigkeit (Krebs/Menold, Kapitel 30 in diesem Band) der Ergebnisse zugutekommen (Alexander/Becker 1978: 93). Da die Situations- oder Personenbeschreibungen als experimentelle Einwirkungen (Stimuli) gezielt vom Forscher variiert werden, jede Vignette quasi als „Kontrollgruppe“ für die anderen Vignetten dient und die

Befragten den Fragebögen per Zufall zugewiesen werden (bzw. alle dieselben vom For- scher ausgewählten Vignetten beurteilen), so dass die Vignettenmerkmale unabhängig von den Merkmalen der Befragten sind, erlauben faktorielle Surveys eine *kausale* Erklärung (Kühnel/Dingelstedt, Kapitel 80 in diesem Band) des beobachteten Urteilsverhaltens. Die Variation der Vignettenmerkmale ist somit ursächlich für die beobachtete Änderung des Urteilsverhaltens der Befragten verantwortlich (Taylor 2006: 1196).

---

## 53.2 Der Aufbau der Vignetten: Faktoren, Merkmalsausprägungen und Antwortskala

Der faktorielle Survey lässt sich für die Untersuchung ganz unterschiedlicher Fragestellungen verwenden: In der Literatur finden sich etwa faktorielle Surveys zur sozialen Mindest- sicherung (Liebig/Mau 2002), zur Gültigkeit von Normen (Beck/Opp 2001), zu deviantem Verhalten (Eifler 2007) oder zur Einwanderung (Jasso 1988). Zur Veranschaulichung des faktoriellen Designs ist in Abb. 53.1 neben dem Einleitungstext eine Beispielvignette zum Vertrauen in andere Menschen (Bartolomé/Dülmer 2013) wiedergegeben. Die Vignetten können dabei, wie im verwendeten Beispiel, eher in tabellarischer Form oder als Kurzge- schichte präsentiert werden. Bei Kurzgeschichten werden die variierenden Größen häufig durch Unterstreichung hervorgehoben (Jasso 2006: 412).

### 53.2.1 Faktoren und Merkmalsausprägungen

Eine Hauptaufgabe bei der Erstellung eines faktoriellen Surveys besteht darin festzulegen, welche *Merkmale* (beim faktoriellen Survey als *Faktoren* oder *Dimensionen* bezeichnet) einen bedeutsamen Einfluss auf das Urteilsverhalten ausüben. Bei dieser Aufgabe kann auf existierende Theorien, Forschungsergebnisse und konventionelles Alltagswissen zurückgegriffen werden (Jasso 2006: 342). Anschließend ist eine Entscheidung über die einzubeziehenden *Merkmalsausprägungen* (*Level*) notwendig. Während die Anzahl der Merkmalsausprägungen bei vielen qualitativen Variablen wie dem Geschlecht gegeben ist, stellt sich bei quantitativen Variablen wie dem Alter die Frage, wie viele und welche Merkmalsausprägungen für den faktoriellen Survey verwendet werden sollen. Für die Schätzung einer linearen Beziehung reicht es aus statistischer Sicht aus, zwei Merkmalsausprä- gungen auszuwählen, die die Spannweite der Variable möglichst realitätsgerecht abdecken. Wird eine u-förmige Beziehung erwartet, so sind mindestens drei Merkmalsausprägungen erforderlich.

Die Festlegung der Anzahl der Faktoren und deren Merkmalsausprägungen entschei- det über die Größe des Vignettenuniversums, das sich aus der Anzahl aller möglichen Vignetten zusammensetzt. Für das oben angeführte Beispiel zum Vertrauen wurden die folgenden Merkmalsausprägungen in die Untersuchung einbezogen (Tab. 53.1):

Auf den folgenden Seiten werden wir Ihnen 20 ähnliche Situationen vorstellen, in denen eine Person ihr offenes Handgepäck im Zugabteil zurückgelassen hat, um zur Toilette zu gehen. Im Handgepäck befindet sich jeweils eine Brieftasche mit **etwa 100 Euro** und das **Handy**.

Bevor die Person das Zugabteil verließ, befand sich dort eine einzige weitere Person. Diese andere Person ist jeweils **35 Jahre alt**, **ledig** und besitzt **keine deutsche Staatsbürgerschaft**, unterscheidet sich aber in ihren übrigen Merkmalen von Situation zu Situation. Für jede der 20 Situationen möchten wir Sie bitten anzukreuzen, wie wahrscheinlich es Ihrer Meinung nach ist, dass das **Eigentum** der Person während ihres Toilettenganges von der anderen Person **gestohlen wurde**. Benutzen Sie dabei jeweils bitte die folgende Antwortskala:

Es wurde ...

vollkommen sicher nicht gestohlen	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	vollkommen sicher gestohlen
--------------------------------------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------	--------------------------------

- 1 Frau J. befindet sich in einem Zug, der gerade in den nächsten Bahnhof einfährt. Sie kommt von der Toilette als ihr einfällt, dass sie ihr unverschlossenes Handgepäck im Zugabteil zurückgelassen hat. In dem Handgepäck befinden sich die Brieftasche mit **etwa 100 Euro** und das **Handy**.  
Bevor sie das geschlossene Zugabteil verließ, befand sich dort nur eine weitere, **35jährige, ledige Person ohne deutsche Staatsbürgerschaft** mit folgenden weiteren Merkmalen:

Geschlecht:	weiblich
Hautfarbe:	weiß
Deutschkenntnisse:	spricht gebrochenes Deutsch
Konfession:	muslimisch
Erwerbstätigkeit:	Geschäftsführerin eines mittelständischen Unternehmens

**Wie wahrscheinlich ist es Ihrer Meinung nach, dass das Eigentum von Frau J. während ihrer Abwesenheit von der beschriebenen Person gestohlen wurde? Es wurde ...**

vollkommen sicher nicht gestohlen	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	vollkommen sicher gestohlen
--------------------------------------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------	--------------------------------

**Abb. 53.1** Vignettenbeispiel zum Vertrauen in andere Menschen

Das vorliegende Beispiel setzt sich aus fünf Faktoren zusammen. Weitere Informationen über die Wertgegenstände im Handgepäck sowie das Alter, der Familienstand und die Staatsbürgerschaft der Vignettenperson wurden im vorliegenden Beispiel konstant gehalten, da deren Einfluss für die Forschungsfragen von nachrangiger Bedeutung war. Von den fünf Faktoren weisen vier jeweils zwei Merkmalsausprägungen auf. Die Konfession ist der einzige Faktor, für den fünf Merkmalsausprägungen von Interesse waren. Bei vier Faktoren

**Tab. 53.1** Merkmale und Merkmalsausprägungen des faktoriellen Survey zum Vertrauen

<b>Merkmale (Faktoren/Dimensionen):</b>	<b>Merkmalsausprägungen (Level):</b>
Geschlecht:	– männlich – weiblich
Hautfarbe:	– schwarz – weiß
Deutschkenntnisse:	– spricht gebrochenes Deutsch – spricht fließend Deutsch
Konfession:	– keine – evangelisch – katholisch – orthodox – muslimisch
Erwerbstätigkeit:	– arbeitslose/r Verkäufer/in – Geschäftsführer/in eines mittelständischen Unternehmens

mit je zwei Merkmalsausprägungen und einem Faktor mit fünf Merkmalsausprägungen gibt es genau  $2 \cdot 2 \cdot 2 \cdot 2 \cdot 5 = 80$  Merkmalskombinationen. Das vollständige Vignettenuniversum setzt sich somit aus insgesamt 80 Vignetten zusammen.

Bei der Auswahl der Faktoren und der Merkmalsausprägungen sollte darauf geachtet werden, dass mit zunehmender Anzahl von Faktoren die Komplexität der Aufgabenstellung für die Befragten stark ansteigt (Beck/Opp 2001: 287). Folgt man einer Empfehlung von Green und Srinivasan (1978: 108) für die Conjoint-Analyse, einem eng mit der Vignettenanalyse verwandten Verfahren, das vor allem in der Marktforschung eingesetzt wird (Dülmer 2007: 406), dann ist es angeraten, nicht mehr als sechs Faktoren mit jeweils bis zu vier oder fünf Merkmalsausprägungen in den faktoriellen Survey einzuschließen. Werden weniger Faktoren einbezogen, dann können mehr Merkmalsausprägungen berücksichtigt werden und umgekehrt.

### 53.2.2 Die Antwortskala

Das Antwortverhalten der Untersuchungsteilnehmer wird beim faktoriellen Survey zumeist auf einer Rating-Skala (Franzen, Kapitel 51 in diesem Band) erfasst, wobei empirisch vor allem 11er-Skalen vorherrschen (Wallander 2009: 511). Werden jedem Befragten mehr Vignetten zur Beurteilung vorgelegt als es vorgegebene Antwortkategorien gibt, bei einer 11er-Skala also 12 oder mehr, dann müssen zwangsläufig mindestens zwei Vignetten gleich beurteilt werden. Dies würde bei differenzierterem Urteilsvermögen zwangsläufig

mit einem Informationsverlust einhergehen, da die entsprechenden Unterschiede auf der Antwortskala nicht mehr erfasst werden könnten (Lodge 1981: 5). Die Vorgabe von einer zu hohen Anzahl von Antwortkategorien mag Befragte mit weniger differenziertem Urteilsvermögen hingegen überfordern. Als Alternative zur einfachen Rating-Skala kann die Magnitude-Messung verwendet werden, bei der die Befragten ihr Urteil im Vergleich zu einem zuvor festgelegten Referenzwert ausdrücken (dazu weiterführend Lodge 1981: 7, Jasso 2006: 345-346, Liebig/Mau 2002: 119). In manchen Untersuchungen sind die Befragten schließlich aufgefordert, einen absoluten Betrag zu nennen (etwa den gerechten Verdienst einer beschriebenen Vignettenperson) oder die Wahrscheinlichkeit eines Ereignisses anzugeben (etwa die angenommene Wahrscheinlichkeit einer Scheidung, Jasso 2006: 344).

---

### 53.3 Die Auswahl des experimentellen Designs: Quoten- und Zufallsdesigns

Mit der Anzahl der Faktoren und der Merkmalsausprägungen steigt auch die Größe des Vignettenuniversums, so dass es für den einzelnen Befragten sehr schnell unmöglich wird, alle Vignetten zu beurteilen. Beim faktoriellen Survey hat jeder Untersuchungsteilnehmer daher in der Regel nur eine Stichprobe des Vignettenuniversums (Vignettenset) zu bearbeiten. In repräsentativen Umfragen sollten keinem Befragten mehr als 10 bis 20 Vignetten vorgelegt werden (Beck/Opp 2001: 291). Um systematischen Ordnungseffekten (serielle Korrelationen, die dadurch entstehen können, dass zuvor beurteilte Vignetten einen Einfluss auf die Beurteilung der folgenden Vignetten ausüben) bei der Schätzung der Regressionskoeffizienten vorzubeugen, ist es angeraten, die Vignetten jedes Vignettensets in zufälliger Anordnung beurteilen zu lassen (Jasso 2006: 343). Logisch unmögliche Kombinationen, wie sie bei der Einbeziehung von Bildung und Beruf auftreten können (etwa Physiker ohne Hauptschulabschluss), sind auszuschließen.

Zur Erzeugung der Vignettenstichproben existieren zwei grundsätzlich unterschiedliche Ansätze: dies sind zum einen Zufallsdesigns und zum anderen Quotendesigns (Dülmer 2007: 383). Die Grundidee für die Verwendung eines einfachen *Zufallsdesigns* besteht darin, das Vignettenuniversum möglichst genau durch verschiedene Vignettenstichproben gleicher Größe zu repräsentieren. Hierzu wird aus dem Vignettenuniversum für jeden einzelnen Befragten eine eigene Zufallsstichprobe (Häder/Häder, Kapitel 18 in diesem Band) gleichen Umfangs gezogen (Rossi/Anderson 1982: 40-41, Jasso 2006: 343). Die Ziehung jedes einzelnen Vignettensets erfolgt mit oder ohne Ersetzen bereits gezogener Vignetten, wobei das Ziehen ohne Ersetzen den Vorteil hat, dass keinem Befragten eine Vignette wiederholt vorgelegt wird (Dülmer 2007: 384). Vor der Erhebung ist für jede Vignettenstichprobe zu kontrollieren, ob es eine Vignettenvariable gibt, die zufallsbedingt nur eine Merkmalsausprägung aufweist (Konstante), und ob es eine Vignettenvariable gibt, die sich innerhalb eines Vignettensets perfekt durch die Vignettenvariablen anderer Faktoren

vorhersagen lässt (Linearkombination). Ohne die Ersetzung solcher Vignettenstichproben lässt sich für die betreffenden Befragten nicht schätzen, welchen Einfluss die Variable auf das Urteilsverhalten dieser Person besitzt.

Alternativ wird beim *Quotendesign* (Akremi, Häder/Häder, Kapitel 17 und 18 in diesem Band) versucht, die Merkmale des Vignettenuniversums durch die Konstruktion eines oder vergleichsweise weniger Vignettensets abzubilden. Das Vignettenuniversum zeichnet sich durch die beiden Merkmale aus, dass die Faktoren untereinander unkorreliert sind (*orthogonal*) und dass die Merkmalsausprägungen innerhalb eines Faktors gleich häufig auftreten (*balanciert*). Bei den Quotendesigns lassen sich grundsätzlich die *klassischen faktoriellen Designs* (fraktionelle faktorielle Designs und konfundierte faktorielle Designs, Gunst/Mason 1991, Kirk 1995, Steiner/Atzmüller 2006: 132-133) von den *D-effizienten Designs* (D-effiziente Designs und konfundierte D-effiziente Designs, Kuhfeld et al. 1994, Dülmer 2007: 387, 2013) unterscheiden.

Bei den klassischen faktoriellen Designs sind die Faktoren zwar immer orthogonal, die Merkmalsausprägungen innerhalb der Faktoren aber nicht immer balanciert. Da die Faktoren orthogonal sind, lassen sich die Effekte aller Vignettenmerkmale unabhängig voneinander schätzen. Der Preis für die Verwendung eines *fraktionellen faktoriellen Designs* statt des vollständigen Vignettenuniversums besteht darin, dass bestimmte Vignettenvariablen innerhalb des verwendeten Vignettensets perfekt mit Wechselwirkungen (Interaktionen) zwischen verschiedenen Vignettenvariablen konfundiert sind (Aliasing, Gunst/Mason 1991: 40-41, Steiner/Atzmüller 2006: 126). Perfektes Aliasing zwischen der Vignettenvariable  $X_1$  und der Wechselwirkung zwischen den Vignettenvariablen  $X_2$ ,  $X_3$  und  $X_4$  bedeutet beispielsweise, dass die Korrelation zwischen der Vignettenvariable  $X_1$  und dem Interaktionsterm  $X_2 \cdot X_3 \cdot X_4$  innerhalb des verwendeten Vignettensets 1 (oder aber -1) beträgt, so dass sich deren Effekte statistisch nicht mehr trennen lassen. Um eine systematische Verzerrung (Bias) des geschätzten Regressionskoeffizienten der Vignettenvariable ausschließen zu können, ist Aliasing daher nur zulässig, wenn der Einfluss der Wechselwirkung im Vergleich zum Einfluss der Vignettenvariable (Haupteffekt) vernachlässigbar ist (Gunst/Mason 1991: 41-42). Beim *konzertierten faktoriellen Design* werden verschiedene fraktionelle faktorielle Designs für die Erhebung verwendet. Hierdurch wird eine Konfundierung über verschiedene Vignettensets hinweg möglich: Liegt etwa zwischen einer Vignettenvariable und einer Wechselwirkung ein perfektes Aliasing vor und ist die hierdurch verursachte perfekte Korrelation im ersten fraktionellen faktoriellen Design positiv und im zweiten negativ, dann bietet das konfundierte faktorielle Design die Möglichkeit, zumindest die Größe der Wechselwirkung mit Hilfe einer sogenannte Set-Variable über die beiden fraktionellen faktoriellen Designs hinweg abzuschätzen.

Für manche Fragestellungen mag innerhalb einer vertretbaren Setgröße jedoch kein faktorielles Design existieren. Da die Erzeugung *fraktioneller faktorieller Designs* auf Teilbarkeitsregeln beruht, ist das Finden eines faktoriellen Designs innerhalb einer zumutbaren Setgröße immer dann besonders schwierig bzw. manchmal unmöglich, wenn die Anzahl der Merkmalsausprägungen der Faktoren auf unterschiedlichen Primzahlen basieren (also etwa 2, 3, 5, Gunst/Mason 1991: 51-52, eine sehr gute Veranschaulichung des

grundlegenden Verfahren zur Erzeugung solcher Designs findet sich bei Backhaus et al. 2000: 575-576). D-effiziente Designs erlauben zumindest leichte Korrelationen zwischen den Faktoren, wobei dann quasi im Gegenzug angestrebt wird, dass die Vignettenmerkmale innerhalb eines Faktors gleich häufig auftreten (Kuhfeld et al. 1994: 545). Im Rahmen dieser Möglichkeiten lässt sich für jede gewünschte Setgröße ein D-effizientes Design finden.

Die D-Effizienz ist ein Maß für die relative Güte eines ausgewählten Designs im Vergleich zu einem balancierten, orthogonalen Design. Die Berechnung der D-Effizienz bezieht sich immer nur auf die Vignettenvariablen (und gegebenenfalls Wechselwirkungen), deren Effekte später zu schätzen sind. Werden ausschließlich qualitative Variablen verwendet, dann ist die D-Effizienz auf Werte im Bereich von 0 bis 100 beschränkt (Kuhfeld et al. 1994: 547, 549). Der Maximalwert von 100 bedeutet, dass das ausgewählte Design balanciert und orthogonal ist. Da eine größere Anzahl von Vignetten pro Vignettenset nicht unbedingt zu einer höheren oder zumindest gleichhohen D-Effizienz führt (Dülmer 2007: 394), ist ein Vergleich über verschiedene Setgrößen hinweg angeraten. Werden quantitative Variablen einbezogen, dann kann die D-Effizienz auch Werte größer als 100 annehmen. Ein Designvergleich bleibt aber auch in diesem Fall möglich.

Im Unterschied zum einfachen Zufallsdesign führt das Zusammenfügen einfacher D-effizienter Designs, die Untersuchungsteilnehmern zur Beurteilung vorgelegt wurden, weder zu einer Erhöhung der Balance noch zu einer Reduktion der Korrelationen zwischen den Faktoren. Beim *konfundierten D-effizienten Design* werden Unbalanciertheit und Korrelationen über verschiedene einfache D-effiziente Designs hinweg ausgeglichen bzw. beseitigt (Dülmer 2013). Die D-Effizienz des konfundierten D-effizienten Designs ist somit höher als die des einfachen D-effizienten Designs. Werden alle verwendeten D-effizienten Designs gleich häufig bearbeitet, dann ist das zusammengefügte konfundierte D-effiziente Design balanciert und orthogonal.

Die Erzeugung D-effizienter Designs erfolgt über spezialisierte Computerprogramme (SAS), wobei die D-Effizienz automatisch ausgegeben wird. Durch das Permutieren (Vertauschen) der Zuweisung von Merkmalsausprägungen der Vignetten zu den entsprechenden Variablenausprägungen des gewählten D-effizienten Designs lässt sich ein konfundiertes D-effizientes Design erzeugen (Dülmer 2013). Zur Erzeugung relativ einfacher fraktioneller faktorieller Designs werden zumeist Computerprogramme (SPSS, SAS) oder vorgefertigte Pläne aus der Literatur verwendet (etwa Gunst/Mason 1991, Kirk 1995: 549-654). Die Erzeugung komplexer konfundierter faktorieller Designs kann häufig nur über die Veränderung vorhandener, aber nicht zufriedenstellender Designs erfolgen (Steiner/Atzmüller 2006: 137). Diese Arbeit ist jedoch im Allgemeinen sehr anspruchsvoll und zeitaufwendig.

### 53.4 Die Auswertung faktorieller Surveys

Bei den allermeisten faktoriellen Surveys wird der Einfluss von Vignetten- und Befragtenmerkmalen auf die Beurteilung der Vignetten analysiert. Wird mehr als eine Vignette von jedem Befragten beurteilt, ist das Antwortverhalten in den Kontext eines jeden Untersuchungsteilnehmers eingebettet. Für solche hierarchisch strukturierten Daten empfiehlt sich die Auswertung über die Mehrebenenanalyse (Pötschke, Kapitel 87 in diesem Band). Aus Gründen der Übersichtlichkeit ist im Folgenden die Mehrebenengleichung für eine dichotome oder metrische Vignettenvariable X und eine dichotome oder metrische Variable Z für ein Befragtenmerkmal (keine Wechselwirkung zwischen X-Variable und Z-Variable) wiedergegeben. Da der Einfluss der Set-Variablen eines konfundierten faktoriellen Designs nur über die verschiedenen fraktionellen faktorielle Designs hinweg schätzbar ist, sind die Set-Variablen wie Z-Variablen auf der Befragtenebene in die Mehrebenenanalyse einzubeziehen.

$$\text{Vignettenebene (Level 1): } Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$$

$$\text{Befragtenebene (Level 2): } \beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

In der Zweiebenen-Regressionsgleichung steht „i“ für die jeweilige Vignette, „j“ für einen jeweiligen Befragten, „Y“ für die Beurteilung einer jeweiligen Vignette, „ $\beta$ “ für die befragenspezifischen b-Koeffizienten (die ausschließlich der Schätzung des Mehrebenenmodells dienen), „ $\gamma$ “ für den durchschnittlichen b-Koeffizienten über alle Befragten hinweg (Grand Mean), „u“ für die Residuen (Zufallskomponenten) auf der Befragtenebene und „r“ für das Residuum auf der Vignettenebene. Der Einschluss einer Zufallskomponente bedeutet, dass der b-Koeffizient eines jeden Befragten durch einen befragenspezifischen u-Term vom jeweiligen Grand Mean abweicht. Die statistische Mindestanforderung für die simultane Einbeziehung einer eigenen Zufallskomponente für jede Vignettenvariable (einschließlich des Intercepts und eingeschlossener Wechselwirkungen) besteht folglich darin, dass die gewählte Setgröße mehr Vignetten umfasst als b-Koeffizienten auf der Vignettenebene zu schätzen sind. Ob die Varianz eines u-Terms (Varianzkomponente) signifikant wird oder nicht, ist eine empirische Frage (Snijders/Bosker 2012: 43-44). Je weniger Varianzkomponenten signifikant sind, desto homogener ist das Antwortverhalten der Befragten. Erweisen sich alle Varianzkomponenten als nicht signifikant und können damit alle u-Terme aus der Gleichung entfernt werden, wird aus der Mehrebenenregression eine einfache OLS-Regression. In Abhängigkeit von der Heterogenität des Antwortverhaltens beruhen die geschätzten b-Koeffizienten also stärker auf den Schätzern der einzelnen Vignettensests oder auf den Schätzern der kombinierten Vignettenstichprobe.

### 53.5 Abschließende Bemerkungen

Die Frage nach einem möglichst optimalen experimentellen Design ist beim faktoriellen Survey eng mit der *Heterogenität* des Antwortverhaltens der Befragten verknüpft. Ein optimales Design erlaubt eine effiziente (also möglichst präzise und damit reliable sowie möglichst unverzerrte und damit valide) Schätzung aller Regressionskoeffizienten. Der Unterschied zwischen einem einfachen Quotendesign (unbalanciertes fraktionelles faktorielles oder D-effizientes Design) und einem Zufallsdesign besteht darin, dass durch das Zusammenfügen der verschiedenen Vignettensets eines Zufallsdesigns sich die D-Effizienz erhöht: Die Gesamtstichprobe der von allen Befragten bearbeiteten Vignetten erreicht somit eine größere Balance und Orthogonalität (D-Effizienz) als die einzelnen Vignettensets. Beim einfachen Quotendesign hat jedes Vignettenset hingegen die gleiche D-Effizienz wie die Gesamtstichprobe aller bearbeiteten Vignetten. Bei homogenem Antwortverhalten der Befragten werden die b-Koeffizienten mit Hilfe der OLS-Regression auf der Grundlage der Gesamtstichprobe aller bearbeiteten Vignetten geschätzt. Unter dieser Bedingung weist das Zufallsdesign mit zunehmender Anzahl der Befragten sehr schnell eine höhere D-Effizienz als ein unbalanciertes fraktionelles faktorielles Design und ein D-effizientes Design auf. Dies gilt insbesondere für kleine Setgrößen (Dülmer 2007). Bei sehr heterogenem Antwortverhalten verhält es sich umgekehrt: Das durchschnittliche Vignettenset einer Zufallsstichprobe weist eine geringere D-Effizienz auf als das Vignettenset eines D-effizienten Designs.

Bei konfundierten Quotendesigns werden verschiedene Vignettensets mit gleicher D-Effizienz eingesetzt. Wie beim Zufallsdesign erhöht sich mit dem Zusammenfügen der verschiedenen Vignettensets somit auch die D-Effizienz für die Gesamtstichprobe der von allen Befragten bearbeiteten Vignetten. Damit gewinnen konfundierte Quotendesigns eine gewisse Unabhängigkeit von der Heterogenität des Antwortverhaltens der Befragten: Bei sehr homogenen Antworten weisen sie eine höhere D-Effizienz als das D-effiziente und das unbalancierte fraktionelle faktorielle Design auf. Bei sehr heterogenem Antwortverhalten ändert sich hingegen nichts: Das konfundierte Quotendesign besitzt in diesen Fällen eine ebenso hohe D-Effizienz wie das entsprechende unbalancierte fraktionelle faktorielle bzw. D-effiziente Design und eine höhere D-Effizienz als das Zufallsdesign. Die einzigen Designs, die die maximale D-Effizienz von 100 vollkommen unabhängig von der Heterogenität des Antwortverhaltens erreichen, sind das balancierte fraktionelle faktorielle und das balancierte konfundierte faktorielle Design. Insofern handelt es sich hierbei um ideale Designs.

Für komplexe Fragestellungen und/oder für faktorielle Surveys mit unterschiedlicher Anzahl von Merkmalsausprägungen existiert innerhalb einer zumutbaren Setgröße jedoch nur selten ein konfundiertes faktorielles Design (Steiner/Atzmüller 2006: 137, 144). Unter diesen Umständen bietet das konfundierte D-effiziente Design eine attraktive Alternative. Da die Erzeugung konfundierter D-effizienter Designs bei einer hohen Anzahl von Vignettenvariablen und/oder Merkmalsausprägungen jedoch zeitaufwendig ist, mag man unter diesen Umständen weiterhin einfache Zufallsdesigns verwenden. Sind Vignetten

aufgrund logisch unmöglich Merkmalskombinationen auszuschließen, dann lassen sich konfundierte Quotendesigns nicht anwenden. Fraktionelle faktorielle Designs und einfache D-effiziente Designs lassen sich schließlich verwenden, wenn, wie häufig in repräsentativen Umfragen, kleine Vignettensets benötigt werden und das Antwortverhalten sehr heterogen ist (Dülmer 2007: 406). D-effiziente Designs lassen sich auch noch beim Ausschluss von logisch unmöglichen Merkmalskombinationen einsetzen (Kuhfeld et al. 1994: 551). Im Unterschied zu den anderen Designs lässt sich bei einfachen Quotendesigns im Nachhinein allerdings nicht mehr prüfen, ob es nicht vernachlässigbar große Wechselwirkungen (Interaktionen) zwischen Vignettenvariablen gibt, die bei der Konstruktion des verwendeten Vignettensets nicht berücksichtigt wurden (Dülmer 2013).

## Literatur

- Alexander, Cheryl S./Becker, Henry Jay (1978): The Use of Vignettes in Survey Research.  
In: Public Opinion Quarterly 41: 93-104
- Backhaus, Klaus/Erichson, Bernd/Plinke, Wulff/Weiber, Rolf (2000): Multivariate Analyse-methoden. Eine anwendungsorientierte Einführung. 9. Auflage. Berlin: Springer
- Bartolomé, Edurne/Dülmer, Hermann (2013): Decomposing the Determinants of (Dis-) Trust in Germany and Spain. Results from an Experimental Design Using the Factorial Survey. Vortrag auf der ESRA-Konferenz 2013. Ljubljana, 15.-19. Juli 2013
- Beck, Michael/Opp, Karl-Dieter (2001): Der faktorielle Survey und die Messung von Normen. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 53: 283-306
- Dülmer, Hermann (2013): The Factorial Survey: Design Selection and its Impact on Reliability and Internal Validity. Unveröffentlichtes Manuskript
- Dülmer, Hermann (2007): Experimental Plans in Factorial Surveys: Random or Quota Design? In: Sociological Methods & Research 35: 382-409
- Eifler, Stefanie (2007): Evaluating the Validity of Self-Reported Deviant Behavior Using Vignette Analysis. In: Quality & Quantity 41: 303-318
- Green, Paul E./Srinivasan, V. (1978): Conjoint Analysis in Consumer Research: Issues and Outlook. In: Journal of Consumer Research 5: 103-123
- Gunst, Richard/Mason, Robert L. (1991): How to Construct Fractional Factorial Experiments. Volume 14. The Basic References in Quality Control: Statistical Techniques. Milwaukee, WI: ASQC Quality Press
- Jasso, Guillermina (2006): Factorial Survey Methods for Studying Beliefs and Judgments. In: Sociological Methods & Research 34: 334-423
- Jasso, Guillermina (1988): Whom Shall We Welcome? Elite Judgments of the Criteria for the Selection of Immigrants. In: American Sociological Review 53: 919-932
- Kirk, Roger E. (1995): Experimental Design: Procedures for the Behavioral Sciences. 3. Auflage. Pacific Grove: Brooks/Cole Publishing
- Kuhfeld, Warren F./Tobias, Randall D./Garrat, Mark (1994): Efficient Experimental Design with Marketing Research Applications. In: Journal of Marketing Research 21: 545-557
- Liebig, Stefan/Mau, Steffen (2002): Einstellungen zur sozialen Mindestsicherung. Ein Vorschlag zur differenzierten Erfassung normativer Urteile. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 54: 109-134
- Lodge, Milton (1981): Magnitude Scaling. Quantitative Measurement of Opinions. Beverly Hills, CA: Sage
- Rossi, Peter H./Anderson, Andy B. (1982): The Factorial Survey Approach: An Introduction. In: Rossi/Nock (Hg.): 15-67
- Rossi, Peter H./Nock, Steven L. (Hrsg.) (1982): Measuring Social Judgments: The Factorial Survey Approach. Beverly Hills, CA: Sage
- Snijders, Tom A. B./Bosker, Roel J. (2012): Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Analysis. 2. Auflage. Los Angeles: Sage

- Steiner, Peter M./Atzmüller, Christiane (2006): Experimentelle Vignettendesigns in faktoriellen Surveys. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 58: 117-146
- Taylor, Brian J. (2006): Factorial Surveys: Using Vignettes to Study Professional Judgement. In: British Journal of Social Work 36: 1187-1207
- Wallander, Lisa (2009): 25 Years of Factorial Surveys in Sociology: A Review. In: Social Science Research 38: 505-520

Jürgen H.P. Hoffmeyer-Zlotnik und Uwe Warner

---

## 54.1 Soziodemographische Variablen

Soziodemographische Variablen stellen Hintergrundmerkmale der Bevölkerung dar, über die die Population einer Stichprobe oder einer Zielgruppe eines Forschungsprojektes beschrieben werden kann. Die soziodemographischen Merkmale unterteilen sich einerseits in demographische Merkmale wie Kohortenzugehörigkeit, Geschlecht, ethnische Zugehörigkeit und Familien- und Verwandtschaftsordnungen bzw. Haushaltszusammensetzung, und andererseits in soziale Ungleichheit beschreibende sozioökonomische Merkmale wie Bildung, Ausbildung, Erwerbsstatus, berufliche Tätigkeit und Einkommen. Diese Merkmale dienen auch dazu, die zu analysierenden Populationseinheiten einer Umfrage in Gruppen und Untergruppen zu untergliedern.

In der Sozialforschung werden soziodemographische und sozioökonomische Merkmale bei der Stichprobenziehung (Häder/Häder, Kapitel 18 in diesem Band), für Schichtung (Häder/Häder, Kapitel 18 in diesem Band) und Gewichtung (Kiesl, Kapitel 24 in diesem Band) und bei der Datenanalyse (Blasius/Baur, Kapitel 79 in diesem Band) zu einer Untergliederung der zu analysierenden Populationseinheiten einer Umfrage (Reinecke, Kapitel 44 in diesem Band) nach sozialstrukturellen Merkmalen in Gruppen und Untergruppen genutzt. Bei der Interpretation von Einstellungen und Verhalten sind die Hintergrundmerkmale oft die unabhängigen Variablen. In der Marktforschung (Ziegler, Kapitel 10 in diesem Band) werden soziodemographische Daten auch eingesetzt, um Produkte besser auf spezifische Zielgruppen abstimmen zu können, um hierüber Rückschlüsse auf die Gestaltung von Marketingstrategien zu erhalten.

„Geschlecht“, „Alter“ und „Bildung“ sind die meist genutzten demographischen Variablen in Umfragen. Diese drei Variablen stellen die Minimalinformation an soziologischen Hintergrundvariablen dar. Für Sozialstrukturanalysen (Huinkink, Kapitel 82 in diesem Band) werden darüber hinaus in der Regel noch die Variablen „Familienstand“, „Beruf“, „Einkommen“ und „Haushalt“ benötigt.

## 54.2 Die Variablen im Einzelnen

Im Folgenden werden die in den Demographischen Standards (Statistisches Bundesamt 2010) dargestellten und von den Autoren als Schlüsselvariablen betrachteten Merkmale aufgeführt:

### 54.2.1 Geschlecht

Das „Geschlecht“ der Zielperson wird definiert über deren primäre Geschlechtsmerkmale. Die Operationalisierung von „Geschlecht“ findet in einer Unterteilung in „männlich“ und „weiblich“ statt.

### 54.2.2 Alter

„Alter“ wird definiert über die Lebensjahre. Für die Operationalisierung von „Alter“ reicht die Angabe des Geburtsjahres, idealerweise ergänzt durch den Geburtsmonat.

### 54.2.3 Staatsangehörigkeit

Mit der Einbeziehung von Zuwanderern in die Umfragen gewinnt die Variable der derzeitigen „Staatsangehörigkeit“ an Bedeutung. Diese Variable ist eines von fünf Merkmalen für ethnische Zugehörigkeit und Migrationshintergrund (siehe Hoffmeyer-Zlotnik/Warner 2012: 281-299). In den Demographischen Standards wird lediglich erhoben, ob die befragte Person die deutsche Staatsangehörigkeit, bzw. bei Verneinung die Staatsangehörigkeit eines der Mitgliedsstaaten der Europäischen Union hat.

### 54.2.4 Familienstand

„Familienstand“ wird definiert über die rechtlichen Kategorien eines staatlich legalisierten Zusammenlebens, auch bezogen auf gleichgeschlechtliche eingetragene Partnerschaften. Abgefragt wird in den juristischen Kategorien: ledig, verheiratet, verheiratet aber getrennt lebend, geschieden, verwitwet. Nicht-eingetragene Partnerschaften werden für all jene erfasst, die nicht mit einem legal angetrauten oder eingetragenen Partner zusammenleben.

### 54.2.5 Bildung

Die Variable „Bildung“ stellt einen der zentralen Indikatoren für soziale Ungleichheit dar. Bildung umfasst sowohl die Abschlüsse der allgemeinbildenden Schule als auch die

Ausbildungsabschlüsse in betrieblicher und schulischer Ausbildung sowie Abschlüsse, die an Hochschulen und Universitäten erworben werden. In den Umfragen wird nicht das Wissen erfasst, sondern die erworbenen Zertifikate. In der heutigen Bundesrepublik Deutschland ist hierbei darauf zu achten, dass die Befragten unterschiedliche Bildungssysteme besucht haben können, z.B. das der ehemaligen DDR oder das der Bundesrepublik Deutschland, wobei sich letzteres über die Bildungshoheit der heute 16 Bundesländer weiter unterteilt und teils über länderspezifische Begrifflichkeiten schwer abzufragen ist. Erfasst werden sollten alle erworbenen Zertifikate, da in der Bundesrepublik Abschlüsse auch indirekt über Äquivalenzen erreicht werden können. So ist z.B. der Zugang zu einer Hochschule nicht nur über das Abitur, sondern in einzelnen Bundesländern auch über eine berufliche Ausbildung zu erreichen, die mit der Prüfung zu einem Handwerksmeister endet. Für „noch-Schüler“ wird in den Demographischen Standards der angestrebte Schulabschluss erfasst, damit auch diese Subgruppe von Personen zuzuordnen ist.

#### 54.2.6 Erwerbsstatus

Beim „Erwerbsstatus“ geht es in sozialwissenschaftlichen Umfragen nicht um einen Rückschluss auf das Bruttonationaleinkommen, wie in wirtschaftswissenschaftlichen oder statistischen Erhebungen, sondern um eine zentrale Statusvariable. Daher sollen die Befragten sowohl in den Kategorien eines Lebensphasenkonzeptes (nach Bildungsphase, Erwerbsphase und Phase des Ruhestandes untergliedert) als auch in der Erwerbsphase in arbeitsmarktrelevanten Kategorien erfasst werden, soweit letztere dazu dienen, den sozialen Status zu beeinflussen. Bei der Abfrage werden die Erwerbstätigen nach Umfang des zeitlichen Engagements (Vollzeit, Teilzeit, stundenweise), nach temporären Ruhephasen (Mutterschafts-, Erziehungsurlaub, Elternzeit, Altersteilzeit) oder nach spezifischen Sonderformen einer Erwerbstätigkeit (freiwilliges soziales Jahr etc.) sowie nach Arbeitslosigkeit unterschieden. In den Demographischen Standards wird in Nachfragen geklärt, wie viele parallele Beschäftigungsverhältnisse eine befragte Person hat, ob darunter auch eine in Selbstständigkeit ausgeübte Tätigkeit ist und wie viele Stunden normalerweise pro Woche gearbeitet wird. Für die Nicht-Erwerbstätigen wird geklärt, welcher Gruppe diese angehören, wobei hier unterschieden wird in Bildungsphase, Erwerbslosigkeit und der Phase des Altersruhestandes.

#### 54.2.7 Stellung im Beruf

„Stellung im Beruf“ unterscheidet die Erwerbstätigen in grober Klassifikation nach Selbstständigen und abhängig Beschäftigten. Die Selbstständigen, nach Tätigkeitsfeldern unterschieden (Landwirte, akademisch freie Berufe und Selbstständige), werden entsprechend der Betriebsgröße klassifiziert. Die Nicht-Selbstständigen, in Deutschland unterteilt in die Kategorien „Arbeiter“ (für manuelle Tätigkeiten in der Produktion), „Angestellte“ (für

Tätigkeiten in Handel, Dienstleitung, Verwaltung, Forschung und Lehre) und „Beamte“ (für Personen im Staatsdienst mit Pensionsberechtigung), werden nach der Autonomie der Tätigkeit klassifiziert. Autonomie der Tätigkeit bei den Nicht-Selbstständigen einerseits und Verantwortung durch Betriebsgröße bei den Selbstständigen andererseits kann grob in Kategorien für berufliches Prestige übersetzt werden (Hoffmeyer-Zlotnik 2003).

#### **54.2.8 Berufliche Tätigkeit**

Die zentrale Variable für die Messung von sozialer Ungleichheit ist bei Voll- oder Teilzeitbeschäftigen die Beschreibung der „beruflichen Tätigkeit“. Über die berufliche Tätigkeit, klassifiziert nach der International Standard Classification of Occupations (ISCO 2008, siehe : ILO 2014) kann man sowohl berufliches Prestige als auch den sozioökonomischen Status ableiten. Die ISCO ist ein von der International Labour Organisation (ILO) erstelltes Instrument zur Klassifikation von beruflichen Tätigkeiten – nicht von Berufsbezeichnungen. Die ISCO ist das einzige Instrument, über das sich (a) im ökonomischen Sinn die Tätigkeit, (b) im medizinischen Sinn die körperliche Gefahr durch das Tun, (c) im soziologischen Sinn eine Statusvariable wie berufliches Prestige (Standard International Occupational Prestige Scale, SIOPS, siehe: Treiman 1977; Ganzeboom/Treiman 2003) oder sozioökonomischer Status (Standard International Socio-Economic Index of Occupational Status, ISEI, siehe: Ganzeboom/Treiman 1996) oder Klassenzugehörigkeit (EGP, siehe: Erikson et al. 1979) generieren lässt.

#### **54.2.9 Haushaltsgröße**

Die Variable der „Haushaltsgröße“ setzt eine eindeutige und allgemeinverständliche Definition von Haushalt voraus, denn zunächst muss die Frage geklärt werden, wer zum Haushalt dazu zählen soll. In den 28 Mitgliedsstaaten der Europäischen Union gibt es, kulturgebunden, 28 unterschiedliche Definitionen von Haushalt. Zentrale Dimensionen hierbei sind a) ein gemeinsames Budget, b) eine gemeinsame Haushaltsführung, c) eine gemeinsame Wohnung oder d) die gemeinsame Familie. Da die Befragten, auch innerhalb eines Kulturreiches, „Haushalt“ sehr unterschiedlich verstehen, ist es wichtig, die Beschreibung derer, die zum Haushalt zählen, genau zu definieren. Nur so erlangt man eine vergleichbare Stichprobenziehung, vor allem beim Umstieg von einer Haushalts- auf eine Personenstichprobe: Wer soll befragt werden? Für die Sozialstrukturanalyse wird Haushalt nur über die gesamte Anzahl der Personen, die dazu zählen, und über jene Personen, die zum Gesamteinkommen beitragen, beschrieben.

### 54.2.10 Einkommen

Das Einkommen wird in sozialwissenschaftlichen Bevölkerungsumfragen meistens nur über zwei Variablen erfasst: das „Nettoeinkommen der befragten Person“ und das „gesamte Nettohaushaltseinkommen“ der Haushaltseinheit, in der die befragte Person eingebunden ist. In der Regel ergibt sich hier eine grobe Schätzung des Einkommens, das einem Haushalt zur Verfügung steht. Denn das Einkommen jeder einzelnen Person setzt sich aus einer Reihe von Teileinkommen unterschiedlicher Art zusammen – nur jede vierte erwachsene Person in Deutschland hat ein Einkommen, welches sich aus weniger als sechs Teileinkommen zusammensetzt (wie z.B. monatliches Arbeitseinkommen, Gratifikationen, Rückzahlungen von der Krankenkasse oder der Steuer, Einkommen aus Vermögen, staatliche Transfers, private Transfers). Damit hat sich eine Referenzperson für einen Haushalt nicht nur an die eigenen Teileinkommen zum Zeitpunkt der Befragung zu erinnern, sondern muss dieses für alle Mitglieder des Haushaltes gleichzeitig können. Daher werden in sozio-ökonomischen Studien wie dem Soziökonomische Panel (SOEP) alle Haushaltsmitglieder nach jeder möglichen Einkommensvariante befragt. In Deutschland kann nach der Expert Group on Household Income Statistics (Canberra Group) (2001) nach 72 unterschiedlichen Typen von Teileinkommen unterschieden werden.

---

## 54.3 Standardisierung auf nationaler Ebene

Um über Umfragen hinweg vergleichen zu können, ist es notwendig, soziodemographische Variablen zu standardisieren. Hierzu bedarf es nicht nur einer valide messenden Formulierung der relevanten soziodemographischen Variablen, sondern es bedarf auch des politischen Willens und der politischen Macht, solch ein Instrument auf den Markt zu bringen und durchzusetzen. 1976 hat Franz Urban Pappi (1979) am ZUMA (Zentrum für Umfragen, Methoden und Analysen in Mannheim) mit der Entwicklung einer „Standarddemographie“ für die Bundesrepublik Deutschland begonnen, die es ermöglichen sollte, eine Konvention für den sozialstatistischen Teil von Umfragen zu schaffen. Mitte der 1980er Jahre wurden diese Idee und die Vorarbeiten des ZUMA von den drei Dachorganisationen der Marktforschung (ADM, Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.), der akademischen Sozialforschung (ASI, Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V.) und der amtlichen Statistik (Statistisches Bundesamt) aufgegriffen und zu diesem Zweck hat man eine gemeinsame Arbeitsgruppe für die Entwicklung und Pflege der „Demographischen Standards“ für Deutschland eingerichtet. Bisher sind nach zwei Vorabversionen 1988 und 1991 zum Test von Durchführbarkeit und Akzeptanz die „Demographischen Standards“ in fünf Auflagen erschienen (1993, 1995, 1999, 2004, 2010) (Statistisches Bundesamt 2010). Notwendig werden Überarbeitungen in etwa fünfjährigem Abstand, um die Demographischen Standards an neues Recht und aktuelle gesellschaftliche Entwicklungen anzupassen. Dabei achtet die Arbeitsgruppe darauf, dass die Zeitreihen so wenig wie möglich beeinträchtigt werden.

Die Demographischen Standards sind ebenfalls mit ständig aktualisierten Auszählungen des Mikrozensus vergleichbar. Das ermöglicht dem Umfrageforscher

1. Bevölkerungsgruppen zu definieren, die von der Befragung ausgeschlossen sind,
2. nicht oder nur schwer erreichbare soziodemographische Gruppen (Schütte/Schmies, Kapitel 61 in diesem Band) des Surveys zu identifizieren,
3. das Ausmaß der Antwortausfälle (Engel/Schmidt, Kapitel 23 in diesem Band) zu schätzen,
4. die Verzerrung der Stichprobe zu benennen (Engel/Schmidt, Kapitel 23 in diesem Band) und
5. Hochrechnungsfaktoren (Kiesl, Kapitel 24 in diesem Band) zu ermitteln, um die Population der Antwortenden an die Zielpopulation der Grundgesamtheit anzupassen.

---

## **54.4 Die Harmonisierung soziodemographischer Variablen für den internationalen Vergleich**

Auch für einen internationalen Vergleich (Braun, Kapitel 56 in diesem Band), sei es in einem Projektverbund oder über unterschiedliche Projekte, wäre eine „Standardisierung“ notwendig. Aber die Möglichkeiten des Übersetzens eines standardisierten Umfrageinstruments hören dort auf, wo der Umgang mit soziodemographischen Variablen beginnt. Die Kategorien nationaler Bildungssysteme lassen sich ebenso wenig übersetzen wie Besonderheiten des nationalen Arbeitsmarktes. Daher müssen die Messungen harmonisiert werden, um soziodemographische und sozioökonomische Variablen über unterschiedliche Kulturen oder Länder vergleichbar zu machen.

Harmonisierung ist die Entwicklung gemeinsam nutzbarer Messinstrumente, um soziodemographische Indikatoren aus einer Kultur mit denen aus einer anderen Kultur vergleichbar zu machen. Denn hinter soziodemographischen und sozioökonomischen Indikatoren stehen einerseits kulturelle Konzepte und andererseits auch nationale Strukturen, gegeben durch Organisationen, Traditionen und Gesetze.

Bei der Datenerhebung im internationalen Vergleich werden unterschiedliche Techniken der Harmonisierung benutzt (Hoffmeyer-Zlotnik/Wolf 2003: 389ff.):

Die *Input-Harmonisierung* setzt bei international akzeptierten Standards wie Definitionen, Konzepten, Aggregationen und Klassifikationen an und nutzt diese, um die Daten zu erheben. „All survey countries use precisely the same survey procedures in an ideal case. Country-specific particularities are only permissible where they are indispensable“ (Information Society Technologies 1999: 1). Das bedeutet, dass die für den Vergleich benutzten Kategorien schon für die Datenerhebung eingesetzt werden. Input-Harmonisierung ist immer eine ex-ante-Harmonisierung.

Die *Output-Harmonisierung* nutzt, im Gegensatz zur Input-Harmonisierung, die nationalen Kategoriensysteme. Sie ist eine ex-post durchgeführte Harmonisierung. Man startet auch hier mit einer über alle beteiligten Länder oder Kulturen abgestimmten Definition dessen, was der gemeinsame Indikator messen soll. Der Transfer von nationalen zu

international vergleichenden Kategoriensystemen wird allerdings erst nach der Datenerhebung, ausgehend von nationalen Kategoriensystemen, geleistet.

Bei der Harmonisierung von Variablen sollen die Forscher sich in fünf Schritten vom sozialwissenschaftlichen Konzept zum Messinstrument des Surveys vorarbeiten (siehe: Hoffmeyer-Zlotnik/Warner 2012: 31-40; Hoffmeyer-Zlotnik/Wolf 2003: 395ff.):

- Im ersten Schritt müssen die beteiligten Forscher für jede *Hintergrundvariable* klären, wozu sie diese benötigen (bewirkt eine gezielte Auswahl) und was sie damit messen wollen (Definition der Variable). Hierbei muss beachtet werden, dass die gemeinsam gefundene Definition in jedem der beteiligten Länder Vergleichbares bezeichnet.
- Im zweiten Schritt muss geklärt werden, welches *kulturelle oder nationale Konzept* hinter einem Begriff steht. Dieses ist Aufgabe der nationalen Experten. Die politische Umsetzung der erarbeiteten Konzepte geschieht eher national über Institutionen, Organisationen und/oder Traditionen und ist damit länderspezifisch zu sehen.
- Im dritten Schritt werden über alle beteiligten Länder die *Strukturen* analysiert, die sich aus der nationalen Umsetzung der einzelnen Konzepte ergeben. Hierunter versteht man z.B. die Organisation von Bildung, die Steuergesetzgebung, die Organisation des nationalen Wohlfahrtsstaates, etc. Für die Umsetzung in ein Messinstrument müssen zunächst die Gemeinsamkeiten hinter den nationalen Konzepten und Strukturen für die einzelnen Variablen herausgearbeitet werden.
- Im vierten Schritt muss ein *Messinstrument* ausgewählt werden, sofern Instrumente zur Verfügung stehen (*Output-Harmonisierung*), oder es muss ein für die Fragestellung passendes Messinstrument (bestehend aus Frage und Antwortkategorien) entwickelt werden (*Input-Harmonisierung*). Dieses Messinstrument sollte sowohl ein für nationale Umfragen valides als auch für international vergleichende Umfragen nutzbares Instrument sein. Die beteiligten Forscher/innen müssen testen, ob das entwickelte Messinstrument bzw. Kategoriensystem die empirischen Strukturen in den einzelnen beteiligten Ländern realistisch widerspiegelt und ob das entwickelte Messinstrument die Logik der gemeinsam entwickelten Definition der zu messenden Variable trifft. Auch müssen sie darauf achten, dass das entwickelte Messinstrument von den Befragten, die in unterschiedlichen nationalen und kulturellen Kontexten leben, im Sinne der Intention verstanden wird und beantwortet werden kann.
- Im fünften Schritt findet, sofern *Output-Harmonisierung* gewählt wurde, die *Harmonisierung der nationalen Messinstrumente* statt. Im Fall von *Input-Harmonisierung* findet die Harmonisierung vor der Datenerhebung statt und es wird für die Datenerhebung kein nationales Instrument eingesetzt.

Das *Produkt der Harmonisierung*, unabhängig ob Input oder Output, ist ein *gemeinsames Kategoriensystem*, das – entsprechend dem gemeinsamen Konzept – nationale Ausprägungen vergleichend zusammenfasst.

Um die Vergleichbarkeit wichtiger Schlüsselvariablen der europäischen Statistik zu gewährleisten, hat die Europäische Kommission eine Arbeitsgruppe zur Standardisierung soziodemographischer Variablen für Europa eingesetzt (European Communities 2007).

## 54.5 Weiterführende Materialien zu soziodemographischen Variablen im nationalen und internationalen Vergleich

*Statistisches Bundesamt (2010): Demographische Standards, Ausgabe 2010. <https://www.destatis.de/DE/Methoden/DemografischeRegionaleStandards/Standards.html> (Zugriff: 05.06.2013).* Die drei herausgebenden Verbände entwickeln und pflegen seit Mitte der 80er-Jahre gemeinsame Empfehlungen zu Demographischen Standards in Umfragen für die Bundesrepublik Deutschland. Sie vereinheitlichen soziodemographische Merkmale, um sozialstrukturelle Messungen und Indikatoren zwischen einzelnen Umfragen (und mit jährlichen Auszählungen des Mikrozensus) zu vergleichen.

*European Communities (2007): Task Force on Core Social Variables. Final report. [http://epp.eurostat.ec.europa.eu/cache/ITY\\_OFFPUB/KS-RA-07-006/EN/KS-RA-07-006-EN.PDF](http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-07-006/EN/KS-RA-07-006-EN.PDF) (Zugriff: 05.06.2013).* Der Endbericht der Arbeitsgruppe zu den Kernvariablen europaweiter Umfragen schlägt eine begrenzte Anzahl von Variablen aus den Bereichen Demographie, Geographie und Sozialökonomie vor, die eine bessere Identifikation der Bevölkerung in den jeweiligen Surveys ermöglichen und eine bessere vergleichende Beschreibung der Bevölkerungsgruppen über die Länder Europas und zwischen den Datenquellen gewährleisten sollen.

*Hoffmeyer-Zlotnik, Jürgen H. P./Wolf, Christof (Hg.) (2003): Advances in Cross-National Comparison. A European Working Book for Demographic and Social -Economic Variables.* Im Zentrum dieses Arbeitsbuches steht die Harmonisierung von Messungen soziodemographischer Hintergrundvariablen für den Vergleich in den Ländern Europas. Die Autoren stellen zu ausgewählten Variablen den Einfluss der nationalen Kontexte und Strukturen auf die vergleichende Messung dar, illustrieren die historisch gewachsenen, kulturellen und politischen Unterschiede der nationalen Messstandards, stellen Harmonisierungsstrategien dar und enden mit verallgemeinerten Regeln der Harmonisierung sozialdemografischer und sozioökonomischer Variablen für den Ländervergleich.

Zu einzelnen *Schlüsselvariablen* haben Hoffmeyer-Zlotnik/Warner (2012) Formulierungsvorschläge für Abfrage und Antwortkategorien der zentralen demographischen und sozio-ökonomischen Variablen entwickelt und (2013) erläutert, wie in internationalen Projekten mit dem entsprechenden Instrumentarium umzugehen ist.

Des Weiteren haben vor allem Unterorganisationen der Vereinten Nationen und das Europäische Amt für Statistik (EUROSTAT), die sich um die (international vergleichende) Messung von soziodemographischen Variablen bemühen, *Instrumente zur Verfügung* gestellt:

- *UNESCO-UIS: International Standard Classification of Education. ISCED 2011, [http://www.uis.unesco.org/Education/Documents/UNESCO\\_GC\\_36C-19\\_ISCED\\_EN.pdf](http://www.uis.unesco.org/Education/Documents/UNESCO_GC_36C-19_ISCED_EN.pdf) (Zugriff: 05.06.2013)* beschreibt die internationale Standardklassifikation der Bildung, die zum internationalen Vergleich des höchsten erreichten und zertifizierten Bildungsniveaus dient.

- ILO: International Standard Classification of Occupations, ISCO-08 Structure and preliminary correspondence with ISCO-88, <http://www.ilo.org/public/english/bureau/stat/isco/isco08/> (Zugriff: 05.06.2013). Die Internationalen Standardklassifikationen der beruflichen Tätigkeiten (zurzeit sind die Versionen 1988 und 2008 im Gebrauch) gruppieren Informationen zur Ausübung von beruflichen Tätigkeiten auf der Grundlage von Ähnlichkeiten der Fertigkeiten und Kenntnisse, die diese Tätigkeiten im Berufsalltag voraussetzen.

Zum Abschluss werden noch zwei *Server der Europäischen Kommission und des Statistischen Amtes der Europäischen Union* vorgestellt, die sich als sehr hilfreich erweisen, wenn man mit soziodemographischen Variablen arbeiten will:

- EURYDICE (<http://eacea.ec.europa.eu/education/eurydice/am> 05.06.2013): Seit 1980 unterstützt Eurydice die Europäische Kommission und die Mitgliedsstaaten der Europäischen Gemeinschaft in Bereichen der Bildung und des Lebenslangen Lernens. Das Netzwerk stellt Informationen zu den Schulsystemen und Bildungspolitiken der 28 Mitgliedsländer der EU sowie der vier Mitgliedsländer der EFTA (Schweiz, Liechtenstein, Norwegen und Island) zur Verfügung. Im Zentrum der Information stehen detaillierte Beschreibungen zu Struktur und Organisation des jeweiligen Schul- und Bildungssystems.
- RAMON (<http://ec.europa.eu/eurostat/ramon/am> 05.06.2013): RAMON ist der Metadatenserver von Eurostat. Die Hauptaufgabe dieses Klassifikationsservers ist es, Informationen und Fakten über statistische Daten (Metadaten) bereitzustellen: „Konzepte und Begriffe“ ist ein Online-Glossare im Bereich der Europäischen Gemeinschaftsstatistik, „Klassifikationen“ stellt internationale statistische Klassifikationen, Nomenklaturen und Systematiken zusammen, „Standard-Kodelisten“ werden in den Datenbanken von Eurostat verwendet, „Gesetzgebung und Methodik“ stellt EU-Rechtsakte und Methodikhandbücher im Bereich der Statistik zusammen, „Glossare und Thesuren“ gibt Zugang zum ISI-Glossar (International Statistical Institute) und zu Thesuren im Bereich der Statistik, „nationale Methodiken“ beschreiben auf nationaler Ebene benutzte Erhebungsmethodiken um EU-Statistiken, Qualitätsberichte, usw. zu erarbeiten.

Im Bereich „Klassifikationen“ werden Standardklassifikationen detailliert dargestellt. Neben den aktuellen Schemata und Instrumenten werden auch deren Vorgänger beschrieben und dokumentiert.

## Literatur

- Erikson, Robert/Goldthorpe, John H./Portocarero, Lucienne (1979): Intergenerational Class Mobility in Three Western European Societies. In: British Journal of Sociology 30: 415-441
- European Communities (2007): Task Force on Core Social Variables. Final Report. Luxembourg: Office for Official Publications of the European Communities
- Canberra Group (2001): Expert Group on Household Income Statistics: Final Report on Recommendations. Ottawa
- Ganzeboom, Harry B.G./Treiman, Donald J. (1996): Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. In: Social Science Research 25: 201-239
- Ganzeboom, Harry B.G./Treiman, Donald J. (2003): Three Internationally Standardised Measures for Comparative Research on Occupational Status. In: Hoffmeyer-Zlotnik, Jürgen H. P./Wolf, Christof (Hg.): Advances in Cross-National Comparison. A European Working Book for Demographic and Social -Economic Variables. New York: Kluwer Academic/Plenum Publisher. 159-193
- Hoffmeyer-Zlotnik, Jürgen H.P. (2003): „Stellung im Beruf“ als Ersatz für eine Berufsklassifikation zur Ermittlung von sozialem Prestige. In: ZUMA-Nachrichten 53: 114-127
- Hoffmeyer-Zlotnik, Jürgen H.P./Warner, Uwe (2012): Harmonisierung demographischer und sozio-ökonomischer Variablen. Instrumente für die international vergleichende Sozialforschung. Wiesbaden: VS Verlag
- Hoffmeyer-Zlotnik, Jürgen H.P./Warner, Uwe (2013): Soziodemographische Standards für Umfragen in Europa. München/Mering: Rainer Hampp Verlag
- Hoffmeyer-Zlotnik, Jürgen H.P./Wolf, Christof (2003): Comparing Demographic and Socio-Economic Variables Across Nations: Synthesis and Recommendations. In: Hoffmeyer-Zlotnik, Jürgen H. P./Wolf, Christof (Hg.): Advances in Cross-National Comparison. A European Working Book for Demographic and Social -Economic Variables. New York: Kluwer Academic/Plenum Publisher. 389-406
- Information Society Technologies (1999): CHINTEX Synopsis. Contract no IST-1999-11101. URL: [https://www.destatis.de/DE/Methoden/Methodenpapiere/Chintex/ResearchResults/Downloads/Synopsis.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Methoden/Methodenpapiere/Chintex/ResearchResults/Downloads/Synopsis.pdf?__blob=publicationFile), 4.April.2014
- International Labour Organisation (2014): International Standard Classification of Occupations 2008. Structure Definitions and Correspondence Tables. <http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm> 5. Juni. 2013
- Pappi, Franz. U. (Hg.) (1979): Sozialstrukturanalysen mit Umfragedaten. Probleme der standardisierten Erfassung von Hintergrundvariablen in allgemeinen Bevölkerungs-umfragen. Königstein (Ts.): Athenäum

Statistisches Bundesamt (2010): Demographische Standards, Ausgabe 2010. Eine gemeinsame Empfehlung des ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V., der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (ASI) und des Statistischen Bundesamtes. 5. überarbeitete und erweiterte Auflage. Statistik und Wissenschaft Band 17

Treiman, Donald J. (1977): Occupational Prestige in Comparative Perspective. New York: Academic Press

Rossalina Latcheva und Eldad Davidov

---

## 55.1 Indexbildung, Skalierungsverfahren, Skalen und Skalieren: eine kurze begriffliche Einleitung

In der sozialwissenschaftlichen Literatur existiert keine einheitliche Definition, was unter Index oder Indexbildung subsumiert werden kann und ob und wie sich Indexbildung von Skalierungsverfahren abgrenzt. Diese fehlende definitorische Eindeutigkeit entspringt nicht zuletzt der Frage, ob Indizes auch als Instrumente zur Messung sozialwissenschaftlicher Konzepte bezeichnet werden dürfen (Diekmann 2009: 230ff.).

### 55.1.1 Index

Unter dem Begriff *Index* wird zum einen die Zusammenfassung mehrerer Einzelindikatoren zu einer neuen Variable nach festgelegten Vorschriften verstanden, z.B. dem (Auf) Summieren der Anzahl zustimmender Antworten zu einem bestimmten Thema. Damit ist die Indexbildung explizit ein Auswertungs- und kein Erhebungs- oder Messverfahren. *Soziale Schicht* ist ein Beispiel für einen Index auf der Individualebene, er wird meistens aus den Individualmerkmalen Bildung, Einkommen und Berufsposition errechnet (Huinink, Kapitel 82 in diesem Band). Ein Beispiel auf der Aggregatebene (Graeff, Kapitel 72 in diesem Band) wäre der *Human Development Index* (HDI), der die Lebensverhältnisse in einem Land u.a. auf Basis der Lebenserwartung bei Geburt, des Pro-Kopf-Sozialprodukts und der durchschnittlichen Ausbildungsdauer eines Landes (UNDP 2010: 143) erfasst.

Zum anderen wird unter dem Begriff *Index*  $I_k$  auch ein *Messmodell* der allgemeinen Form  $I_k = f(X_1, X_2, \dots, X_k)$  verstanden, in dem die Variablen  $X_1, X_2, \dots, X_k$  als *Indikatoren* (Burzan, Kapitel 81 in diesem Band) der zu messenden Variable  $X$  bezeichnet werden und die Elemente des Index  $I_k$  darstellen (Besozzi/Zehnpfennig 1976: 12). Beispiele für letzteres wären die Messung der sozialen Schicht, der Lebenszufriedenheit oder des Intelligenzquotienten – allesamt *nicht direkt beobachtbare (latente) Merkmale*, die meistens über eine

Reihe von Indikatorvariablen, welche dieselbe latente Eigenschaft reflektieren, gemessen werden (wie etwa mittels Intelligenztests). Die Antworten der einzelnen Befragten auf die unterschiedlichen Testfragen (bzw. ihre Testwerte) werden dann nach einem vorgeschlagenen statistischen Verfahren zu dem Gesamtwert eines Befragten zusammengefasst.

### 55.1.2 Skala

*Skalen* beinhalten meistens eine Vielzahl von Indikatoren (Items), die dasselbe theoretische Konstrukt reflektieren (z.B. analytische Intelligenz). „Eine Person wird dann dadurch *skaliert*, dass man aus ihren Antworten auf mehrere Items einen sie charakterisierenden Skalenwert in dem zu messenden Merkmal berechnet“ (Borg/Staufenbiel 2007: 389; Hervorhebung durch die Autoren).

Der Begriff „*Skala*“ wird im Kontext der Befragungsforschung aber häufig schon dann verwendet, wenn Befragte anhand einer abgestuften *Antwortskala* (Franzen, Kapitel 51 in diesem Band) eine Frage beantworten.

Es wird weiterhin auch das *Skalenniveau* (Blasius/Baur, Kapitel 79 in diesem Band) von einzelnen Variablen diskutiert.

In dieser Mehrdeutigkeit sei betont, dass mit einer *Skala*, die ein Merkmal *skalieren soll*, das gesamte Instrument gemeint ist, d.h. alle Indikatoren (Burzan, Kapitel 81 in diesem Band) sowie die Vorschrift ihrer Verrechnung, welche durch ein Skalierungsverfahren erarbeitet wurden. Beispielsweise werden die folgenden zehn Indikatoren, welche „Zentrale Arbeitsperzeptionen Berufstätiger“ messen, durch Addition der Itemwerte zu *einem* Skalenwert zusammengeführt (GESIS 2010, Fischer/Kohr 1980):

1. (+) Irgendwie, meine ich, ist meine Arbeit doch wichtig.
2. (-) Bei meiner Arbeit fühle ich mich oft irgendwie leer.
3. (+) Ich habe das Gefühl, bei meiner Arbeit etwas Sinnvolles zu tun.
4. (+) Ohne Arbeit hat das Leben wenig Sinn.
5. (-) Meine Arbeit ist eine einzige Tretmühle.
6. (-) Ich sehe kaum, welchen Sinn meine Arbeit hat.
7. (+) Bei meiner Arbeit kann man immer noch etwas dazu lernen.
8. (-) Meine Arbeit bietet mir kaum Abwechslung.
9. (-) Mir fehlt der Überblick über das, was ich bei meiner Arbeit tue.
10. (-) Manchmal ist es mir fast unangenehm, zu sagen, was ich arbeite.

Dabei bedeutet (+), dass die Frage im Sinne der zu erfassenden Dimension positiv formuliert ist und (-), dass die Frage im Sinne der zu erfassenden Dimension negativ formuliert ist. Jede Frage hat fünf Antwortkategorien: „1“ steht für „unzutreffend“ und „5“ für „zutreffend“. In diesem Sinn wird in diesem Beitrag der Begriff Skala verwendet.

### 55.1.3 Skalierungsverfahren und Indexbildung

*Skalierungsverfahren* (Blasius, Kapitel 83 in diesem Band) sind Techniken zur Konstruktion von Messinstrumenten bzw. von Skalen und beinhalten jeweils eigene Vorschriften, wie einzelne Messwerte zu einem Skalenwert verrechnet werden müssen. Aus dem Gesagten folgt, dass Skalierungsverfahren nicht nur als Auswertungs-, sondern auch als Erhebungsverfahren von latenten Variablen charakterisiert werden können. Es sei jedoch angemerkt, dass Erhebungs- und Analyseverfahren untrennbar miteinander verbunden sind: die Konstruktion einer Skala wird primär von den Zielen darauffolgender Analysen geleitet.

Sowohl bei der Berechnung eines Index als auch bei der Berechnung eines Skalenwertes werden mehrere Variablen nach bestimmten Rechenregeln (z.B. Addition, Multiplikation) und auf Basis (mess-)theoretischer Überlegungen – wie etwa Fragen der Validität und Reliabilität (Krebs/Menold, Kapitel 30 in diesem Band) – zu einer neuen Variable kombiniert.

*Skalierungsverfahren* greifen in erster Linie auf Items zur Erfassung subjektiver Eigenschaften resp. Einstellungen zurück, welche *dasselbe* theoretische Konzept bzw. *dieselbe* Dimension reflektieren – als Beispiele können Einstellungen zu Immigranten, Zufriedenheit mit der Demokratie oder nationale Identität genannt werden. Skalierungsmodelle bieten zudem standardisierte Kriterien zur Beurteilung, welche Items zu einer Skala gehören.

Bei der *Indexbildung* hingegen können auch Indikatoren, die unterschiedliche Dimensionen indizieren, zu einer neuen Größe zusammengefasst werden; eine solche ist die soziale Schicht (s. oben).

---

## 55.2 Gründe für die Verwendung multipler Indikatoren für die Konstruktion von Skalen und Indizes

Bevor wir uns mit weiteren Details und Beispielen zu Indexbildung und Skalierungsverfahren befassen und die Diskussion auf Gemeinsamkeiten bzw. Unterschiede zwischen Skalen und Indizes ausdehnen, wollen wir wichtige Gründe für die Verwendung von multiplen Indikatoren anstelle von einzelnen Items umreißen. Drei Gründe seien hier besonders hervorgehoben:

### 55.2.1 Erfassung möglichst vieler relevanter Aspekte eines theoretischen Begriffs

Ein wesentliches Ziel sozialwissenschaftlicher Theoriebildung und Analysen ist die *Reduktion von Komplexität*. Zur Beschreibung komplexer sozialer Phänomene werden meist eine Reihe theoretischer Konzepte bzw. Begriffe verwendet, die ihrerseits vorwiegend mehrere Aspekte der zu untersuchenden sozialen Wirklichkeit erfassen. Werden in einem Begriff

mehrere Aspekte gebündelt, wie etwa beim Begriff der *Lebensqualität* (Huinink, Kapitel 82 in diesem Band), der meistens „Wohlbefinden“, „Glück“ und „Zufriedenheit“ umfasst, braucht es auch eine Vielzahl an isolierten Indikatoren, welche die interessierenden Dimensionen in der Datenerhebung repräsentieren.

Zudem ist es wesentlich einfacher, in der weiteren Forschung mit dem Begriff bzw. Konstrukt „Lebensqualität“ zu arbeiten, als mit dessen Einzelindikatoren.

Diese Überlegung ist insofern für die Indexbildung von Bedeutung, als die Kombination mehrerer Variablen zur Messung eines theoretischen Begriffs *validere Messungen* produziert als die Verwendung einer einzelnen Variable (Besozzi/Zehnpfennig 1976: 12). Jeder Indikator repräsentiert nur partiell die spezifischen Aspekte eines Konstrukts, d.h. die Korrelationen zwischen den gemessenen Variablen und dem dahinter liegenden theoretischen Konstrukt sind nicht „perfekt“ (oder mit anderen Worten kleiner als + 1 oder - 1).

Skalen kommen daher besonders dann zur Anwendung, wenn theoretische Begriffe bzw. gesellschaftliche Phänomene als mehrdimensional definiert werden (wie etwa verbale, analytische und emotionale Dimensionen in der Definition von Intelligenz) und wenn jede Subdimension durch eine Anzahl empirischer Variablen repräsentiert wird. Im Rahmen der Skalenbildung könnte dann für jede Subdimension eine eigene *eindimensionale* Skala berechnet werden oder mehrere Dimensionen könnten in einer neuberechneten (*mehrdimensionalen*) *Indexvariable* kombiniert werden. Beispiele sind die bereits oben erwähnte soziale Schicht auf der Mikro- und der HDI auf der Makroebene.

### **55.2.2 Verringerung sozialer Erwünschtheit**

Im besonderen Fall der Erfassung subjektiver Eigenschaften wird die Verwendung mehrerer Indikatoren um einen weiteren Grund ergänzt. Die interessierende Eigenschaft – wie z.B. Einstellung(en) zur Todesstrafe – kann unter anderem in der Umfragesituation einer allgemein gültigen sozialen Norm unterliegen, was eine Selbsteinschätzung der Befragten bezüglich des interessierenden Merkmals erschwert oder dazu führt, dass die als unerwünscht wahrgenommene Eigenschaft nur ungern angegeben wird (zum Problem der *sozialen Erwünschtheit im Antwortverhalten* siehe: Hlawatsch/Krickl, Kapitel 20 in diesem Band). Soll die Einstellung zur Todesstrafe erfasst werden, so würden wir mit hoher Wahrscheinlichkeit kein realistisches Bild der Einstellungen der Zielpersonen erhalten, wenn sie direkt mithilfe einer Skala von 1 bis 10 (wo 1 eine starke Ablehnung und 10 eine starke Zustimmung misst) abgefragt werden.

Deshalb wird diese Einstellung meistens *indirekt* über verschiedene, mit ihr eng verbundene, weniger abstrakte Inhalte erfasst. Damit laufen Forschende aber auch Gefahr, nur einen Teil des Konstrukts einzufangen und andere Teile zu vernachlässigen. Dies wird

als *Problem der Validität* oder der *inhaltlichen Gültigkeit der Messung* diskutiert (Krebs/Menold, Kapitel 30 in diesem Band).

### 55.2.3 Reduktion von Messungenauigkeiten

Der dritte Grund für die Verwendung multipler Indikatoren ist eine *Reduktion des Messfehlers* im Vergleich zur Nutzung mehrerer Einzelindikatoren. Solche Messfehler, die sowohl zufällig, als auch systematisch sein können und Messungen verschlechtern, wiegen bei der Verwendung mehrerer Indikatoren weniger schwer, da angenommen wird, dass sich die Zufallsfehler der Einzelindikatoren gegenseitig ausgleichen (zum Problem der *Reliabilität* oder *Zuverlässigkeit* der Messung siehe: Krebs/Menold, Kapitel 30 in diesem Band). „Die Verwendung mehrerer Indikatoren erhöht die Chance, Messungenauigkeiten zu verringern und so den gemeinten, ‚wahren‘ Sachverhalt eher abzubilden“ (Kromrey 2009: 184).

Diese Idee entspringt der *klassischen Testtheorie* (KTT), welche das theoretische Fundament für den Großteil der heute bestehenden Skalen bildet (Borg/Staufenbiel 2007: 313). Ausgangspunkt der KTT ist die Vorstellung von einem hypothetischen „wahren“ Wert, der die Beobachtungsperson entlang der interessierenden Eigenschaft charakterisieren würde. Bei dem Versuch, diese Eigenschaft mittels eines Indikators zu messen, würde die Beobachtung, die wir erhalten, durch zufällige Fehler mal nach oben und mal nach unten von diesem „wahren“ Wert abweichen. Wenn davon ausgegangen wird, dass der *Zufallsfehler* (Häder/Häder, Kapitel 18 in diesem Band) mit gleicher Wahrscheinlichkeit positive wie negative Werte annimmt und sich somit bei *unendlich vielen Messungen* zu Null summieren würde, dann könnte der „wahre Wert“ hypothetisch erhalten werden, indem unendlich viele Male gemessen und dann der Durchschnitt berechnet würde. Da sich diese theoretische Idee aber nicht praktisch umsetzen lässt, entstand der Ansatz, *multiple Indikatoren zur Messung der gleichen Eigenschaft* zu verwenden. Es wird angenommen, dass sich auch die Zufallsfehler unterschiedlicher Indikatoren zu Null ausgleichen.

Davon abgesehen müssen auch *systematische Messfehler* (Krebs/Menold, Kapitel 30 in diesem Band) verhindert werden, welche unter dem Begriff der *Validität* behandelt werden. Hierbei handelt es sich um Messfehler, die sich nicht zu Null ausgleichen würden, sondern die Werte in eine bestimmte Richtung verzerrten. Zum Beispiel würden, wenn Einstellungen gegenüber der Todesstrafe erfragt werden und wenn die Zustimmung zur Todesstrafe sozial unerwünscht ist, die Antworten zu allen Fragen zur Todesstrafe der Gefahr ausgesetzt sein, nicht nur die wahre Einstellung zu messen, sondern auch den Einfluss der (erwarteten) sozial erwünschten Antwort.

Mit zwei Indikatoren können zufällige Messfehler kontrolliert werden. Mit drei Indikatoren können auch systematischen Messfehlern Rechnung getragen werden (Brown 2006, Saris/Gallhofer 2007).

### 55.3 Manifeste versus latente Variablen und reflektive versus formative Indikatoren

Im Folgenden gehen wir kurz auf zwei Differenzierungen ein, die bei der Diskussion von Indexbildung und Skalierungsverfahren relevant sind: manifeste versus latente Variablen und reflektive versus formative Indikatoren.

*Latente Variablen* werden häufig auch als *Konstrukte* bezeichnet, weil sie von den Forschenden mittels Explikation *konstruiert* werden, um theoretische Konzepte bzw. Begriffe in mathematischen Modellen zur Überprüfung von Zusammenhangsannahmen zu repräsentieren. „Latent“ bedeutet „nicht direkt beobachtbar“ und damit auch „nicht direkt messbar“. Latente Variablen repräsentieren theoretische Konstrukte und sind damit einer direkten Messung oder Beobachtung nicht zugänglich.

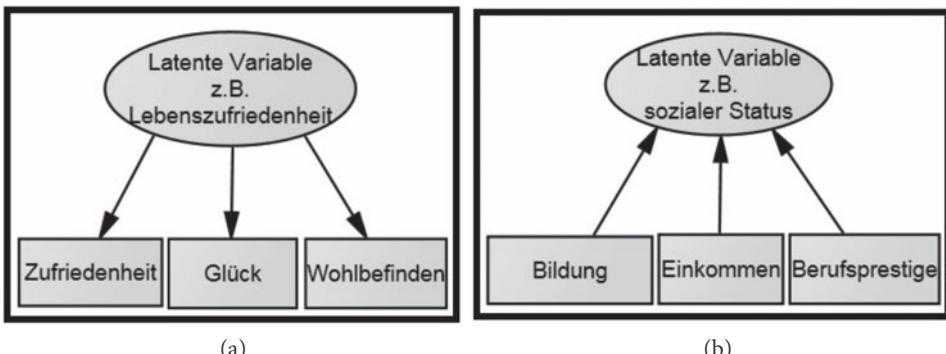
Die Übersetzung theoretischer Konzepte in messbare Größen (*Indikatoren*, die auch *Items* oder *manifeste Variablen* genannt werden) wird als *Operationalisierung* bezeichnet (Bollen 1989: 181, Saris/Gallhofer 2007: 6; Burzan, Kapitel 81 in diesem Band). Bekannte Beispiele für konstruierte latente Variablen können in unterschiedlichen sozialwissenschaftlichen Disziplinen gefunden werden, z.B. „analytische Intelligenz“, „Lebensqualität“, „Anomie“, „Ethnozentrismus“ und „politische Partizipation“.

Abbildung 55.1 ((a) und (b)) zeigt die Verknüpfung zwischen latenten Variablen (Ellipsen) und manifesten Variablen oder Indikatoren (Rechtecke), so wie sie konventionell graphisch dargestellt werden. Die Richtung des unterstellten Kausalzusammenhangs zwischen latenten und manifesten Variablen ist maßgeblich für die Unterscheidung zwischen reflektiven und formativen Indikatoren.

*Reflektive Indikatoren* bzw. *effect indicators* (Bollen 1989: 65) spiegeln die dahinter liegende latente Variable wider. Ihre Zusammenhänge (d.h. statistisch ihre Korrelationen) entstehen dadurch, dass die Indikatoren die *gleiche* latente Variable reflektieren bzw. dass die Indikatoren von der *gleichen* latenten Variable beeinflusst oder verursacht werden (Kühnel/Dingelstedt, Kapitel 80 in diesem Band). Reflektive Indikatoren weisen typischerweise hohe Korrelationen miteinander auf.

Demgegenüber werden *formative Indikatoren* als *causes* bzw. als *induced indicators* bezeichnet (Bollen 1989: 65), weil sie die latente Variable nicht reflektieren, sondern diese bedingen. Somit können formative Indikatoren unterschiedliche Dimensionen erfassen und müssen nicht miteinander korrelieren.

In Abb. 55.1 (b) wird das latente Konstrukt „sozialer Status“ als Kombination bzw. Ergebnis von drei inhaltlich unterschiedlichen Variablen – Bildungsabschluss, Einkommen und Berufsprestige – dargestellt, die jeweils unterschiedliche inhaltliche Dimensionen erfassen und folglich nicht notwendigerweise miteinander korrelieren müssen. Das Beispiel in Abb. 1b kann über einen einfachen Summenindex oder durch eine andere mathematische Verknüpfung (z.B. indem jede Variable ein unterschiedliches Gewicht erhält) berechnet werden.



**Abb. 55.1** (a) Latente Variable mit reflektiven Indikatoren, (b) Latente Variable mit formativen Indikatoren

Dementgegen setzt das Beispiel in Abb. 55.1 (a) voraus, dass die Variablen korrelieren. Diese reflektieren die Antworten zu drei Umfrage-Items über Zufriedenheit, Glück und Wohlbefinden (erhoben durch direkte Fragen, z.B. wie zufrieden, wie glücklich und wie wohl sich die Befragten im Alltag fühlen) und beschreiben damit dasselbe latente Konstrukt „allgemeine Lebenszufriedenheit“. Indikatoren dieser Art werden vorwiegend im Konstruktionsprozess einer Skala entwickelt und somit können auch signifikante (positive) Korrelationen zwischen den drei Variablen erwartet werden. Es wird also angenommen, dass die Antworten auf die Fragen zum Glück, zur Zufriedenheit und zum Wohlbefinden Aufschluss über drei zusammenhängende Nuancen eines latenten Inhalts (der Lebenszufriedenheit) geben. Die drei reflektiven Indikatoren sollten daher die gleiche und möglichst nur diese latente Eigenschaft widerspiegeln bzw. messen, wobei die Eindimensionalität solcher reflektiven Indikatoren mithilfe einer konfirmatorischen Faktorenanalyse (Bollen 1989, Blasius, Kapitel 83 in diesem Band) getestet werden kann.

Die Unterscheidung zwischen reflektiven und formativen Indikatoren ist insofern von Bedeutung, als dass bei den gängigen Methoden zur Überprüfung der Dimensionalität von *Skalen* vor allem reflektive Indikatoren zur Anwendung kommen, bei der *Indexbildung* jedoch sowohl reflektive als auch formative Indikatoren verwendet werden können.

Fassen wir zusammen: Ein *Index* kann auf Basis von mehreren (reflektiven) Indikatoren gebildet werden, die einer *Skala* entstammen bzw. die gleiche Dimension messen, aber er kann auch unabhängig von einer Skala konstruiert werden, d.h. Variablen unterschiedlicher Dimensionen (formative Indikatoren) werden zu einer neuen Größe verrechnet (wie etwa beim „sozialen Status“).

Nach den bisherigen Ausführungen sollte klar geworden sein, dass sowohl für Skalen- als auch Indexbildung Informationen über mehrere Variablen/Indikatoren benötigt werden, die erstens nach bestimmten Regeln ausgewählt und zweitens zu etwas Neuem verrechnet werden. Im Folgenden gehen wir kurz auf einige Punkte ein, die bei der Indexkonstruktion besonders wichtig sind.

## 55.4 Indexbildung

Rohwer und Pötter (2002: 88) weisen darauf hin, dass jede *Indexbildung* eine „Datenreduktion“ darstellt. Die Aussagekraft von Indizes hängt folglich stark von der Methodik ihrer Berechnung bzw. dem Konstruktionsverfahren ab (Rohwer/Pötter 2002: 65f.).

Für die Berechnung von Indizes können unterschiedliche mathematische Funktionen resp. Rechenoperationen zur Anwendung kommen. Je nach Rechenoperation wird meistens zwischen additiven und multiplikativen sowie zwischen gewichteten und ungewichteten Indizes unterschieden. Daher sind sowohl eine grundlegende theoretische Begründung als auch die Transparenz der Indexberechnung (d.h. welche Indikatoren werden wie bei der Berechnung berücksichtigt) von besonderer Bedeutung.

Die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) hat zum Beispiel ein eigenes Handbuch über die Konstruktion von „*composite indicators*“ (ein anderer Begriff für Indizes) herausgegeben, in dem die Methodologie und wichtige (technische) Leitlinien für die Berechnung von Indizes zusammengefasst sind, anhand derer Länder bezüglich ihrer industriellen Wettbewerbsfähigkeit, nachhaltigen Entwicklung, Globalisierung und Innovation klassifiziert werden können (für Details siehe OECD 2008, Saisana/Tarantola 2002).

Andere Autoren legen das Hauptaugenmerk für die Indexbildung auf folgende *Regeln*:

1. Bei reflektiven Indikatoren können unterschiedliche Items nur dann zu einer Skala zusammengefasst werden, wenn sie untereinander positiv korrelieren. Eine negative Korrelation zwischen zwei Items würde sich in diesem Fall nur dann ergeben, wenn eine der beiden Fragen positiv, die andere jedoch negativ formuliert ist. Tabelle 55.1 zeigt, wie eine Korrelationsmatrix mit reflektiven Indikatoren aussehen könnte. Alle Korrelationen zwischen den Indikatoren sind positiv und verhältnismäßig hoch, d.h. höher als 0.30.
2. Bei formativen Indikatoren dürfen die Indikatoren auch deutlich schwächere Korrelationen aufweisen. Allerdings müssen alle Indikatoren (in unserem Beispiel: Zufriedenheit, Glück und Wohlbefinden), die zu einem Index (hier „Lebensqualität“) zusammengefasst werden, in die gleiche Richtung mit einer abhängigen Variable (z.B. „politische Partizipation“) korrelieren – d.h. alle Korrelationen zwischen der abhängigen Variable („politische Partizipation“) und dem jeweiligen Indikator (Zufriedenheit, Glück, Wohlbefinden) des Index („Lebensqualität“) müssen entweder positiv oder negativ sein (Kromrey 2009: 184).

Zu bedenken ist auch, dass bei der Konstruktion von *additiven Indizes* durch das Zusammenzählen von Ausprägungen einer Vielzahl dichotomer Variablen (d.h. Variablen mit nur zwei Ausprägungen) neue ordinal- bzw. intervallskalierte Variablen entstehen, welche die soziale Realität nicht notwendigerweise wiedergeben (Rohwer/Pötter 2002). Deshalb muss die *praktische Tauglichkeit* sowie *analytische Anwendbarkeit* von Indizes stets an der Realität überprüft werden. Nehmen wir als Beispiel das Konzept des freiwilligen Engagements, das meistens mittels einer Reihe dichotomer Items zu Mitgliedschaft und

**Tab. 55.1** Beispiel für eine Korrelationstabelle mit reflektiven Indikatoren

	Zufriedenheit	Glück	Wohlbefinden
Zufriedenheit	1.00		
Glück	0.40	1.00	
Wohlbefinden	0.35	0.55	1.00

Tätigkeiten in „Non-Governmental Organizations“ (NGOs) abgefragt wird. Befragte können jede Frage, die eine Organisation aufzählt, in der man Mitglied sein könnte, mit „Ja“ oder „Nein“ beantworten, was mit „1“ für „Ja“ und „0“ für „Nein“ kodiert wird. Durch ein Zusammenzählen aller mit „1“ kodierten Antworten (d.h. aller Mitgliedschaften) entsteht eine neue Variable (Anzahl von Mitgliedschaften), die mehr Ausprägungen beinhaltet als die ursprünglichen Variablen (z.B. könnten bei zehn Fragen zu Mitgliedschaften durch das Zusammenzählen Werte zwischen einem Minimalwert von „0“ und einem Maximalwert von „10“ erreicht werden). Eine höhere Anzahl an Mitgliedschaften muss jedoch nicht notwendigerweise mit einer höheren Intensität des freiwilligen Engagements einhergehen. Intensität könnte zum Beispiel viel besser mit einer anderen Frage, etwa der Anzahl der Wochenstunden, in denen Befragte freiwillig tätig sind, gemessen werden.

## 55.5 Skalenbildung

Wie bereits erwähnt sind *Skalierungsverfahren* (Blasius, Kapitel 83 in diesem Band) Techniken zur Konstruktion von Messinstrumenten bzw. von Skalen und beinhalten jeweils eigene Vorschriften, wie einzelne Messwerte zu einem Skalenwert verrechnet werden. In diesem Kapitel wurden keine bestimmten Skalierungsverfahren erwähnt. In Methoden-Lehrbüchern werden neben *Likert* meistens die Verfahren nach *Thurstone* und *Guttman* erklärt, obwohl diese in ihrer ursprünglichen Form praktisch kaum mehr eine Rolle spielen (Borg/Staufenbiel 2007: 308). Deren Studium kann aber sehr lehrreich sein, weil sie „Grundideen vorweg [nehmen], die in späteren Skalenkonstruktionsverfahren [...] wieder aufgegriffen wurden“ (Borg/Staufenbiel 2007: 308), welche wesentlich komplexer und voraussetzungsvoller im Verständnis sind. Neben der hier erwähnten *Klassischen Testtheorie* gewinnt die *Probabilistische Testtheorie* immer mehr an Bedeutung. Als umfassende deutschsprachige Lektüre zu Skalierung im Allgemeinen und auch zu sehr vielen spezifischen Verfahren wird das viel zitierte „Lehrbuch – Theorien und Methoden der Skalierung“ von Borg Staufenbiel (2007) empfohlen.

Skalierungsverfahren können als *theoretische Modelle* aufgefasst werden (u.a. Borg/Staufenbiel 2007, Schnell et al. 2008: 183). Am wichtigsten sind dabei die Annahmen darüber, wie die Reaktionen der Befragten auf die ihnen präsentierten Stimuli mit der Ausprägung der latenten, interessierenden Eigenschaft zusammenhängen. Aus diesen Annahmen wird schließlich abgeleitet, welche Art von Indikatoren benötigt wird, wie deren Messqualität geprüft wird und wie sie zu einem Skalenwert verrechnet werden sollten. Um den

Zusammenhang von latentem Konstrukt und Beobachtung modellhaft darzustellen, sei die folgende Gleichung erläutert:  $P = f(T, S)$ , wobei  $P$  die Wahrscheinlichkeit ist, dass ein Item im Sinne der zu messenden Eigenschaft beantwortet – man sagt auch „gelöst“ – wird. Das hieße z.B. bei einem Test der Rechenfähigkeit die korrekte Lösung der Aufgabe.  $T$  bezeichnet die vorgestellte „wahre“ Ausprägung der latenten Variablen bei den Befragten, was – um bei dem Beispiel zu bleiben – die „wahre“ Rechenfähigkeit wäre.  $P$  wird also in Abhängigkeit von  $T$  betrachtet, wobei  $f$  die Art des Zusammenhangs symbolisiert. Die *Schwierigkeit*  $S$  des Items ist eine weitere Größe, welche die Wahrscheinlichkeit der „Lösung“ eines Items beeinflusst. Ein Beispiel zur Verdeutlichung: Ein Schüler mit einer gegebenen Rechenfähigkeit wird die folgenden zwei Aufgaben mit einer unterschiedlichen Wahrscheinlichkeit lösen, weil sie unterschiedlich schwierig sind und somit die Eigenschaft „Rechenfähigkeit“ unterschiedlich stark abfragen: „ $4 + 6 - 20$ “ und „ $14 \times 7 - 164 : 4$ “. An diesem Beispiel wird auch deutlich, weshalb sich die Worte „Lösung“ und „Schwierigkeit“ durchgesetzt haben. Viele Skalierungsverfahren basieren auf der Annahme eines monotonen Zusammenhangs zwischen der Wahrscheinlichkeit der Lösung und dem Niveau ( $T$ ) der latenten Variable. Das würde z.B. für *Lebenszufriedenheit* bedeuten, dass für jedes Item die Wahrscheinlichkeit einer Antwort, die auf Zufriedenheit hinweist, mit steigender tatsächlicher Zufriedenheit bei den Personen zunimmt.

Drei bis vier „gute“ Items können dabei hinreichend sein, um eine theoretische Dimension (z.B. eine Eigenschaft, eine Einstellung etc.) zu messen (zu reflektieren) (s. z.B. Bollen 1989: 179–225). Die *Qualität* („Güte“) der Items einer Skala wird mittels *Itemanalyse* (die den Regeln des jeweiligen Skalierungsverfahrens folgt), der Analyse von *Validität*, *Reliabilität* und bei vergleichenden Studien insbesondere auch der Äquivalenz der Skala (d.h. der Angemessenheit der Skala in unterschiedlichen Erhebungen) überprüft (Krebs/Menold, Kapitel 30 in diesem Band). Hierzu gehört z.B. auch die Prüfung der *Eindimensionalität* und wenn möglich die Überprüfung der theoretischen Annahmen, die jedem Skalierungsverfahren zu Grunde liegen. Im Folgenden werden zwei Aspekte der Qualitätsüberprüfung einer Skala herausgegriffen und im Detail besprochen.

### 55.5.1 Eindimensionalität vs. Multidimensionalität einer Skala

Wenn ein Skalierungsverfahren zum Ziel hat, *eine* Eigenschaft in *einer* Variable zu erfassen, ist ein wesentlicher Teil der Aufgabe, die *Eindimensionalität* der Skala sicherzustellen. Die vielen Indikatoren, die zu einem Skalenwert verrechnet werden, sollten alle die gleiche und möglichst nur diese Eigenschaft messen.

Das Syndrom der „gruppenbezogenen Menschenfeindlichkeit“ stellt beispielsweise ein mehrdimensionales Konzept dar (Zick et al. 2008), das folgende Subdimensionen umfasst: Ausländerfeindlichkeit, Islamophobie, Antisemitismus, Homophobie, Sexismus, Rassismus, Abwertung von Obdachlosen, gesundheitlich beeinträchtigten Personen und Langzeitarbeitslosen, sowie Etabliertenvorrechte. Selbst bei dem Versuch, *eine* dieser Subdimensionen zu messen, muss wiederum die Eindimensionalität sichergestellt

werden, bevor Items, welche zu dem Zweck erhoben wurden, zu einer neuen Variable (z.B. zu einem additiven Index) verrechnet werden. So kann beispielsweise mit *Faktorenanalysen* (Brown 2006, Blasius, Kapitel 83 in diesem Band) herausgefunden werden, ob einem Bündel von Indikatoren eine oder mehrere Dimensionen zu Grunde liegen.

### 55.5.2 Konstruktäquivalenz

Bei der Anwendung von Messinstrumenten (Skalen) in verschiedenen nationalen Kontexten (Braun, Kapitel 56 in diesem Band) oder über die Zeit hinweg (Schupp, Kapitel 73 in diesem Band) ist die Sicherstellung, dass die Messung des relevanten theoretischen Konstrukts äquivalent ist, d.h. über Kontexte und Zeitpunkte invariant ist und damit das Gleiche misst, ein wesentliches Anliegen.

Die Überprüfung der Messinvarianz ist notwendig, bevor zum Beispiel Vergleiche zwischen Ländern oder über die Zeit gemacht werden können. Horn und McArdle (1992: 117) definieren *Messinvarianz* als die Frage „ob Messoperationen unter verschiedenen Bedingungen der Beobachtung und des Studiums von Phänomenen Messungen desselben Attributs erzielen“. In anderen Worten garantiert die Messinvarianz, dass Items in unterschiedlichen Kontexten in ähnlicher Weise verstanden werden.

Fehlende Invarianz bzw. Äquivalenz der Messung führt zum Beispiel dazu, dass Mittelwertunterschiede *nicht* notwendigerweise auf substantielle Unterschiede zurückzuführen sind, sondern ein methodisches Artefakt sein können, das sich aus unterschiedlichen Antwortverzerrungen über Länder und Zeitpunkte bzw. unterschiedlichen Verständnissen der Konzepte ergibt (Steenkamp/Baumgartner 1998). Ebenso wichtig ist, dass fehlende empirische Unterschiede zwischen Ländern dann die Abwesenheit von „echten“ Unterschieden nicht sicherstellen. Steenkamp und Baumgartner (1998) liefern Leitfäden darüber, wie Konstruktäquivalenz bei reflektiven Indikatoren untersucht werden kann.

---

## 55.6 Fazit

Abschließend möchten wir nochmals hervorheben, dass die Aussagekraft von Indizes und Skalen stark von der Methodik ihrer Berechnung bzw. dem Konstruktionsverfahren abhängt. Dabei werden mehrere Variablen nach bestimmten Rechenregeln und auf Basis (mess-)theoretischer Überlegungen zu neuen Variablen kombiniert. Eine theoretische Fundierung der Skalenkonstruktion sowie die Transparenz der Indexberechnung sind aus diesem Grund unerlässlich. Dass Erhebungs- und Analyseverfahren untrennbar miteinander verbunden sind, dass sie theoriegeleitet angewendet werden müssen, haben wir in diesem Kapitel anhand verschiedener Beispiele illustriert.

## Literatur

- Besozzi, Claudio/Zehnpfennig, Helmut (1976): Methodologische Probleme der Index-Bildung. In: Koolwijk, Jürgen von /Wieken-Mayser, Maria (Hg.): Techniken der empirischen Sozialforschung 5: Testen und Messen. München: Oldenbourg. 9-55
- Bollen, A. Klenneth (1989): Structural Equations with Latent Variables. New York: Wiley
- Borg, Ingwer/Staufenbiel, Thomas (2007): Lehrbuch Theorien und Methoden der Skalierung. Bern: Hans Huber
- Brown, Timothy A. (2006): Confirmatory Factor Analysis for Applied Research. New York: The Guilford Press
- Diekmann, Andreas (2009): Empirische Sozialforschung. Reinbek: Rowohlt
- Fischer, Arthur/Kohr, Heinz-Ulrich (1980): Politisches Verhalten und empirische Sozialforschung. München: Juventa
- GESIS (2010): ZIS – Zusammenstellung sozialwissenschaftlicher Items und Skalen- ZIS Version 14.00. URL: <http://www.gesis.org/unser-angebot/studien-planen/zis-ehes/zis/>
- Horn, John L./McArdle, John J. (1992): A Practical and Theoretical Guide to Measurement Invariance in Aging Research. In: Experimental Aging Research 18: 117-144
- Kromrey, Helmut (2009): Empirische Sozialforschung. Stuttgart: Lucius + Lucius
- OECD (2008): Handbook on Constructing Composite Indicators. Methodology and User Guide. OECD Publishing
- Rohwer, Götz/Pötter, Ulrich (2002): Methoden der sozialwissenschaftlichen Datenkonstruktion. Weinheim/München: Juventa
- Saisana, Michaela/Tarantola, Stefano (2002): State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development. Ispra: European Commission
- Saris, Willem E./Gallhofer, Irmtraud N. (2007): Design, Evaluation and Analysis of Questionnaires for Survey Research. New York: Wiley
- Schnell, Rainer/Hill, Paul B./Esser, Elke (2008): Methoden der empirischen Sozialforschung. München: Oldenbourg
- Steenkamp, Jan-Benedict E.M./Baumgartner, Hans (1998): Assessing Measurement Invariance in Cross-National Consumer Research. In: Journal of Consumer Research 25: 78-90
- United Nations Development Programme (2010): The Real Wealth of Nations: Pathways to Human Development. URL: [http://hdr.undp.org/en/media/HDR\\_2010\\_EN\\_Table1\\_reprint.pdf](http://hdr.undp.org/en/media/HDR_2010_EN_Table1_reprint.pdf) am 24.5.2011
- Zick, Andreas/Wolf, Carina/Küpper, Beate/Davidov, Eldad/Schmidt, PeterHeitmeyer, Wilhelm (2008): The Syndrome of Group-Focused Enmity. In: Journal of Social Issues 64 (2): 363-383

Michael Braun

Die interkulturell vergleichende Umfrageforschung hat in den Sozialwissenschaften in den letzten Jahrzehnten stark an Bedeutung gewonnen. Dies betrifft zum einen die Verfügbarkeit von international vergleichbaren Umfragedaten und die auf ihnen basierenden substanzwissenschaftlichen Untersuchungen, zum anderen die Beschäftigung mit methodologischen Fragestellungen. Für interkulturelle Umfragen sind nicht nur die Probleme, die für alle Umfragen allgemein gelten, zu berücksichtigen. Vielmehr bemisst sich die Qualität eines interkulturellen Projekts an der Qualität der einzelnen nationalen Umfragen und insbesondere auch daran, dass diese „vergleichbar“ sind. Dabei spielt es zunächst keine Rolle, ob unterschiedliche ethnische Gruppen in einem einzigen Land oder unterschiedliche Länder miteinander verglichen werden sollen. Bei der Untersuchung verschiedener ethnischer Gruppen (El-Menouar, Kapitel 60 in diesem Band) innerhalb eines Landes sind in der Regel z.B. die Stichprobenziehung (Häder/Häder, Kapitel 18 in diesem Band) und die Übersetzungsproblematik zu berücksichtigen. Bei einem Vergleich unterschiedlicher Länder kommen noch unterschiedliche institutionelle Randbedingungen im weitesten Sinn hinzu (von den Bedingungen der Stichprobenziehung zu den sozioökonomischen und politischen Gegebenheiten). Wegen der größeren Allgemeinheit erfolgt im Folgenden eine Beschränkung auf den Vergleich unterschiedlicher Länder, der Vergleich verschiedener kultureller Gruppen ist aber in der Regel ebenso angesprochen. Zunächst stellt sich die Frage, welche Länder überhaupt in ein Umfrageprojekt oder die Analyse einbezogen werden sollen.

---

## 56.1 Auswahl der untersuchten Länder oder kultureller Gruppen

Die Auswahl der zu vergleichenden Länder oder kulturellen Gruppen spielt bei dem Design eines Forschungsprojektes und der Analyse von Sekundärdaten (Mochmann, Schupp, Kapitel 14 und 73 in diesem Band) eine Rolle. Hierfür bieten sich je nach Forschungsfrage

unterschiedliche Strategien an (z.B. Przeworski/Teune 1970, Scheuch 1968): Unähnliche *Kontextbedingungen*, d.h. die Auswahl möglichst unterschiedlicher Länder, sind ein Vorzug, wenn die Generalisierbarkeit von Zusammenhängen zwischen den untersuchten Variablen aufgezeigt werden soll, wie im „most different systems“-Design (Przeworski/Teune 1970: 34f). Stellt sich dann heraus, dass diese Zusammenhänge in den untersuchten Ländern tatsächlich gleich sind, besteht begründeter Anlass, von deren genereller Geltung auszugehen. Es kann dann nämlich weitgehend ausgeschlossen werden, dass die in einem Land gefundenen Zusammenhänge nur bei einem Zusammentreffen ganz spezifischer Kontextbedingungen bestehen. Die *Systemebene*, d.h. die Merkmale der einzelnen Länder, wird nur dann als Erklärungsfaktor berücksichtigt, wenn die Daten der Annahme der Generalisierbarkeit widersprechen, d.h. wenn in den einzelnen Ländern unterschiedliche Zusammenhänge gefunden werden. Soll demgegenüber nachgewiesen werden, dass eine in einem Land oder einer kulturellen Gruppe gefundene Beziehung nicht überall gilt, erweist sich die Verwendung unterschiedlicher Kontexte (d.h. von Ländern, die sich in einer Vielzahl von Merkmalen unterscheiden) als Nachteil. Wenn die Daten nämlich der Erwartung unterschiedlicher Zusammenhänge scheinbar entsprechen, können Alternativerklärungen eine eindeutige Interpretation unmöglich machen. In diesem Fall ist es vielmehr sinnvoll, möglichst ähnliche Länder auszuwählen (das „most similar systems“-Design von Przeworski/Teune 1970). Werden nämlich schon bei Ländern, die sich hinsichtlich vieler relevanter Merkmale ähnlich sind, Unterschiede in den Zusammenhängen gefunden, so ist davon auszugehen, dass dies erst recht bei sehr verschiedenartigen Ländern der Fall sein wird.

In der Forschungspraxis verlieren diese Überlegungen allerdings zunehmend an Relevanz, zumindest was die Auswahl von Ländern anbetrifft. Viele Projekte der international vergleichenden Umfrageforschung, wie z.B. das International Social Survey Program (ISSP, <http://www.issp.org/>) oder der World Values Survey (WVS, <http://www.worldvaluessurvey.org/>) bemühen sich um eine immer umfassendere Abdeckung aller Länder. Auch die auf einzelne Weltregionen beschränkten Umfrageprojekte wie der European Social Survey (ESS, <http://www.europeansocialsurvey.org/>), die European Values Study (EVS, <http://www.europeanvaluesstudy.eu/>) oder die Eurobarometer-Umfragen ([http://ec.europa.eu/public\\_opinion/index\\_en.htm](http://ec.europa.eu/public_opinion/index_en.htm)) verfolgen in ihrem Bereich eine möglichst komplette Abdeckung der existierenden Länder (zu diesen Datensätzen siehe Mochmann, Kapitel 14 in diesem Band). Zudem existiert mit der Mehrebenenanalyse (Pötschke, Kapitel 87 in diesem Band) ein für interkulturelle Vergleiche besonders geeignetes statistisches Verfahren, mit welchem eine große Zahl von Ländern verglichen werden kann. Mit diesem Verfahren kann z.B. untersucht werden, ob Unterschiede in den Zusammenhängen zwischen den einzelnen Ländern bestehen und wie diese gegebenenfalls mit Merkmalen auf der Länderebene erklärt werden können.

## 56.2 Probleme der Vergleichbarkeit der Projektkomponenten

In der interkulturell vergleichenden Forschung hängen die Schlussfolgerungen von der Qualität und Vergleichbarkeit der einzelnen nationalen Studien ab. Sind einige von ihnen mit Fehlern behaftet, dann entsprechen beobachtete Ähnlichkeiten und Unterschiede möglicherweise lediglich methodischen Artefakten. Dabei kann auch die Verwendung identischer Prozeduren in den einzelnen Ländern keine Vergleichbarkeit garantieren. Teilweise ist die Verwendung unterschiedlicher, aber den jeweiligen Kontexten angepasster Untersuchungsstrategien vorzuziehen. So macht es offensichtlich wenig Sinn, die Durchführung von Telefonumfragen (Hüfken, Kapitel 46 in diesem Band) für alle Länder verbindlich vorzuschreiben. In Ländern mit sehr geringer Telefondichte würde dadurch ein großer Teil der Bevölkerung von vorneherein von der Umfrage ausgeschlossen. Stattdessen sollte man in diesen Ländern auf einen anderen Erhebungsmodus ausweichen, also auf eine persönlich-mündliche (Stocké, Kapitel 45 in diesem Band) oder schriftliche Befragung (Reuband, Kapitel 47 in diesem Band). Auch bei der Stichprobenziehung wird man Unterschiede in der konkreten Durchführung akzeptieren müssen, wenn man die in den einzelnen Ländern jeweils besten Verfahren anwenden möchte. So gilt in Deutschland etwa eine aus den Registern der Einwohnermeldeämter gezogene Stichprobe (Häder/Häder, Kapitel 18 in diesem Band) als optimal für persönlich-mündliche Befragungen. In den Vereinigten Staaten gibt es solche Register allerdings nicht. Dort besteht das Standardverfahren bei wissenschaftlichen Umfragen darin, zuerst zufällig eine Stichprobe von Häuserblöcken zu ziehen, in den ausgewählten Häuserblöcken dann eine weitere Stichprobe von Haushalten und in den ausgewählten Haushalten schließlich die zu befragenden Personen mit einem Zufallsverfahren zu ermitteln. In einigen Entwicklungsländern ist auch ein solches Verfahren nicht durchführbar, z.B. weil keine belastbaren Informationen über die (kleinräumige) Verteilung der Bevölkerung vorliegt, Häuserblocks nicht identifizierbar sind oder ein Teil der Bevölkerung aus Nomaden besteht. In diesen Fällen muss auf weniger stringenten Vorgehensweisen zurückgegriffen werden, die in der Regel keine Zufallsstichproben mehr gewährleisten. Nun wäre es kaum vertretbar, nur um der Gleichheit des Vorgehens Genüge zu tun, solche suboptimalen Verfahren auch in den Ländern einzusetzen, in denen gute Zufallsstichproben gezogen werden können. Schließlich müssen auch bei der Formulierung der Items länderspezifische Besonderheiten beachtet werden. So hat der Staatspräsident in den Vereinigten Staaten eine andere Funktion als in Deutschland. Sollen etwa Einstellungen zu dem jeweiligen Regierungsoberhaupt erfasst werden, müsste daher in der deutschen Fragebogen-Version „President“ durch „Bundeskanzlerin“ ersetzt werden.

Aus diesen Beispielen wird auch deutlich, dass man zwei große Klassen von Fehlern unterscheiden kann (Faulbaum, Kapitel 31 in diesem Band): Erstens repräsentieren die in den unterschiedlichen Ländern gezogenen und realisierten Stichproben die jeweiligen Grundgesamtheiten in vergleichbarer Weise (*coverage and non-response error*) (Engel/Schmidt, Kapitel 23 in diesem Band)? Zweitens funktionieren die Messinstrumente, d.h. die Fragebögen und die einzelnen Fragen nach der Übersetzung, überall gleich (*Messfehler*) (Krebs/Menold, Kapitel 30 in diesem Band)?

Ein „coverage error“ (Häder/Häder, Kapitel 18 in diesem Band) liegt vor, wenn nicht alle Einheiten der zu untersuchenden Population eine von Null verschiedene Wahrscheinlichkeit aufweisen, in die Stichprobe zu gelangen. Der „non-response error“ (Engels/Schmidt, Kapitel 23 in diesem Band) bezieht sich auf die Nicht-Teilnahme von für die (Brutto-) Stichprobe ausgewählten Einheiten der Grundgesamtheit. Er setzt sich zusammen aus dem Anteil der Nicht-Erreichten, der Nicht-Kooperativen und der aus anderen, wie z.B. gesundheitlichen oder sprachlichen Gründen nicht befragungsfähigen Personen. Alle Komponenten können interkulturell stark variieren, z.B. in Abhängigkeit vom Umfrageklima in einer Gesellschaft (d.h. der Wertschätzung von Umfragen und der Bereitschaft, an ihnen teilzunehmen), von der Mobilität der Bevölkerung, der durchschnittlichen Haushaltsgröße, den verfügbaren personellen, finanziellen und organisatorischen Ressourcen des Umfrageinstituts oder der Art des Auftraggebers. Die Auswirkungen des Ausschlusses bestimmter Bevölkerungsteile – sei es durch coverage oder non-response error – hängen unter anderem vom Thema der Umfrage und von der Größe der betroffenen Gruppen ab sowie davon, wie stark sich diese Gruppen von dem in die Umfrage einbezogenen Bevölkerungsteil unterscheiden. Als Messfehler („measurement error“) wird eine Reihe von miteinander interagierenden Fehlerquellen bezeichnet, die auf das Messinstrument (Reinecke, Kapitel 44 in diesem Band), die Interviewer (Michael/Glantz, Kapitel 21 in diesem Band), die Befragten oder den Umfragemodus zurückgeführt werden können.

---

### 56.3 Das Problem der Äquivalenz von Konstrukten und Items

Die Vergleichbarkeit von Konstrukten und Items über die unterschiedlichen Länder hinweg wird in der Regel als „Äquivalenz“ bezeichnet. Dabei lassen sich mehrere Typen von Äquivalenz unterscheiden (van de Vijver/Leung 1997): Konstruktäquivalenz („construct equivalence“), Äquivalenz der Maßeinheit („measurement unit equivalence“) und skalare Äquivalenz („scalar equivalence“).

Interkulturell vergleichenden Studien haben unabhängig von ihren Zielen und der Art der erhobenen Daten eine unverzichtbare Voraussetzung: die *Konstruktäquivalenz* (z.B. Scheuch 1968). Sinnvoll kann in jedem Fall nur dann verglichen werden, wenn in den unterschiedlichen Ländern dieselbe zugrunde liegende Dimension erfasst wird. Ist dies nicht der Fall, haben die Konstrukte in den einzelnen Ländern eine unterschiedliche Bedeutung. Möglicherweise existieren in einigen Ländern für ein theoretisches Konstrukt oder die konkreten, bei der Messung verwendeten Begriffe keine Entsprechungen in der Realität. Dann ist aber überhaupt kein internationaler Vergleich mehr möglich. Ein häufig genanntes Beispiel ist das Konzept der „filial piety“, d.h. der Erwartungen an einen guten Sohn oder eine gute Tochter. In kollektivistischen Gesellschaften (wie z.B. China) schließen diese Erwartungen in stärkerem Maße als in individualistischen Gesellschaften (wie z.B. Deutschland oder die Vereinigten Staaten) die Betreuung der eigenen Eltern im Alter ein. Ein anderes Beispiel wäre ein Vergleich von Einstellungen zu Geschlechterrollen, bei denen in der Regel Vorstellungen und Einstellungen zur Erwerbstätigkeit der Frauen eine

wichtige Rolle spielen. In weitgehend agrarisch geprägten Gesellschaften, in denen Frauen (und auch Männer) selten außerhalb des Haushaltes arbeiten, werden entsprechende Fragen möglicherweise anders (oder aber gar nicht) verstanden als in Gesellschaften, in denen Erwerbstätigkeit räumlich vom Familienleben getrennt ist.

Konstruktäquivalenz ist auch die Voraussetzung für die Äquivalenz der Maßeinheit und der skalaren Äquivalenz. Die Sicherstellung von Konstruktäquivalenz kann einen Einsatz verschiedener Messinstrumente in unterschiedlichen Kulturen erfordern (siehe dazu den nächsten Abschnitt). Dann sind allerdings die beiden anderen Äquivalenzarten möglicherweise nicht mehr gegeben.

Äquivalenz der Maßeinheit bedeutet, dass die Maßeinheit der Skala (Franzen, Kapitel 51 in diesem Band) in unterschiedlichen Kulturen gleich ist. Konkret bedeutet das, dass der Unterschied zwischen zwei Skalenpunkten, z.B. „unzufrieden“ und „ziemlich unzufrieden“ bei einer Frage zur Zufriedenheit mit der Arbeit in den einzelnen Ländern auch gleichen Unterschieden auf der zugrundeliegenden (nicht direkt messbaren) Dimension entspricht. Für den Ursprung der Skala ist aber eine solche Gleichheit zwischen den Ländern dann nicht noch nicht unbedingt gegeben. So könnte in einem Land die Beantwortung der Frage durch einen externen Störfaktor in eine bestimmte Richtung verzerrt werden, zum Beispiel durch eine Tendenz, extreme Antworten generell zu vermeiden. In diesem Land würde dann z.B. die Antwortalternative „völlig unzufrieden“ überhaupt nicht verwendet, auch nicht von denen, die tatsächlich völlig mit ihrer Arbeit zufrieden sind (und somit den gleichen „Nullpunkt“ auf der entsprechenden Dimension aufweisen wie diejenigen, die in einem anderen Land mit „völlig unzufrieden“ antworten. Obwohl dies nichts mit der eigentlich zu messenden Dimension zu tun hat, beeinflusst dies die Ergebnisse. Damit sind auch die gefundenen Mittelwertunterschiede zwischen den einzelnen Ländern irreführend, da sie ja neben der eigentlich zu messenden Dimension auch noch externe Störfaktoren enthalten.

*Skalare Äquivalenz* bedeutet demgegenüber, dass sowohl Maßeinheit als auch Ursprung der Skala gleich sind. Letzteres setzt voraus, um bei dem Beispiel der Arbeitszufriedenheit zu bleiben, dass in allen Ländern diejenigen, die völlig mit ihrer Arbeit unzufrieden sind, bei der entsprechenden Frage auch „völlig unzufrieden“ angeben (oder aber auch nur „sehr unzufrieden“, solange das nur in allen Ländern gleich geschieht). Dann haben Personen in verschiedenen Ländern, die die gleiche Ausprägung auf der zugrundeliegenden Dimension haben, auch den gleichen Nullpunkt bei der für die Messung der Dimension verwendeten Frage. Erst wenn skalare Äquivalenz gegeben ist, können auch die Unterschiede in den Mittelwerten der Fragen und Items über die einzelnen Länder hinweg inhaltlich interpretiert werden.

Wie bereits erwähnt, kann ein Einsatz kulturspezifischer Messinstrumente erforderlich sein, wenn identische Fragen unabhängig von einer möglicherweise fehlerhaften Übersetzung in verschiedenen Ländern entweder überhaupt nicht oder unterschiedlich verstanden würden, und zwar nicht nur aufgrund der Formulierung, sondern auch aufgrund des Realitätsbezugs von Items. Das von Przeworski/Teune (1966) vorgeschlagene „identity-equivalence“-Verfahren ist in solchen Situationen einsetzbar: Zunächst werden die

für alle Länder identischen Indikatoren bestimmt und mit den jeweils kulturspezifischen verbunden. Daraus resultieren unterschiedliche Skalen für die einzelnen Länder. Mithilfe der identischen Items lässt sich dann die funktionale Äquivalenz der kulturspezifischen Items überprüfen. Insgesamt dürften in der Politikwissenschaft stark kontextabhängige Messungen, die sich nicht durch identische Instrumente durchführen lassen, häufiger sein als in anderen sozialwissenschaftlichen Bereichen (Przeworski/Teune 1970). Allerdings ist dies dort auch offensichtlicher, was die Problemlösung erleichtern kann: Bei hoher Kontextabhängigkeit ist von vorneherein klar, dass kein identisches Instrument konstruiert werden kann. Bei mittlerer Kontextabhängigkeit fehlt nicht nur meist die Bereitschaft der Forscher, für die einzelnen Länder manifest unterschiedliche, aber funktional äquivalente Messinstrumente zu konstruieren. Häufig ist auch nicht einfach bestimmbar, wie solche Varianten einer Frage auszusehen haben, mit denen in allen Ländern das Gleiche erfasst werden kann. Die Lösung besteht dann oft darin, sich auf den kleinsten gemeinsamen Nenner zu beschränken. Ein Messinstrument wird also so konstruiert, dass es in allen Ländern eingesetzt werden kann, wobei eine möglicherweise unvollständige Abdeckung der eigentlich intendierten Dimension in Kauf genommen wird. So lassen sich etwa beim Thema Religion nicht alle für christliche Gesellschaften sinnvollen Aspekte abfragen, wenn sie in anderen Religionen keine Entsprechung haben.

---

## 56.4 Herstellung der Äquivalenz der Messinstrumente

Notwendig ist die Herstellung einer Vergleichbarkeit der Stimuli (Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band), die die Befragten letztendlich verarbeiten. Ist diese Vergleichbarkeit nicht gegeben, reflektieren die Daten nicht nur reale Unterschiede zwischen den untersuchten Kulturen, sondern auch Artefakte der Messung. Die Trennung zwischen beiden stellt eine große Herausforderung dar. Es ist natürlich sinnvoll, durch eine sorgfältige Konstruktion der Erhebungsinstrumente ex-ante sämtliche Inäquivalenzen zu vermeiden. Dies ist jedoch in vielen Studien in der Vergangenheit trotz großer Anstrengungen nicht vollständig gelungen. Daher bleibt für den Nutzer von Sekundärdaten die Aufgabe, Äquivalenz ex-post zu überprüfen.

### 56.4.1 Herstellung der Äquivalenz der Messinstrumente ex-ante

Die Vergleichbarkeit der durch bestimmte Messinstrumente erhobenen Daten kann sowohl durch eine unangemessene Übertragung des Fragebogens bzw. einzelner Items in andere Sprachen als auch durch Unterschiede in der sozialen Realität in den unterschiedlichen Ländern beeinträchtigt werden. Die Herstellung von funktionaler Äquivalenz ex-ante wird in der Regel durch eine sorgfältige Konstruktion des Source-Fragebogens unter Einbezug der vorhandenen interkulturellen Kompetenz sowie eine fachgerechte Übersetzung und Adaption von Fragebögen angestrebt. Beispielsweise wird beim ISSP zunächst

ein englischsprachiger Fragebogen von einer sogenannten Drafting Group vorbereitet, die aus den Vertretern von etwa einem halben Dutzend möglichst unterschiedlicher Länder besteht. In einem iterativen Verfahren wird dieser Entwurf mit den Vertretern der übrigen Mitgliedsländer des ISSP diskutiert, in einer kleineren Anzahl von Ländern getestet und schließlich durch das ISSP-Plenum nach ausführlicher Diskussion und Abstimmung verabschiedet. In diesem Prozess werden länderspezifische Besonderheiten ausdrücklich berücksichtigt, sowohl was die Einbeziehung einzelner Items als auch deren Formulierung betrifft. So haben etwa ostasiatische Länder darauf hingewirkt, dass Fragen zur Religion auf einem etwas höheren Abstraktionsniveau formuliert wurden, als es nötig gewesen wäre, wenn alle Länder einen christlichen Hintergrund gehabt hätten. Anschließend wird der Source-Fragebogen in die Sprachen der beteiligten Länder übersetzt. Empfohlen wird hierbei ein Teamansatz, bei dem zwei Übersetzer den Fragebogen zunächst unabhängig voneinander in die Landessprache übertragen und dann gemeinsam mit Experten für das betreffende inhaltliche Thema und Umfrageexperten besprechen. Dabei wird besonders berücksichtigt, ob die Items in der Übersetzung genau so verstanden werden wie im Ausgangsfragebogen. Die daraus hervorgehende landessprachliche Fragebogen-Version wird einem (kognitiven) Pretest unterzogen, wobei insbesondere die Übersetzung überprüft wird, also ob es Probleme mit der Verständlichkeit gibt und ob tatsächlich jeweils die intendierten Dimensionen gemessen werden.

Die Unterschiede zwischen Übersetzung und Adaption sind fließend. Teils wird zwischen beiden Prozessen nicht deutlich unterschieden, obwohl dies sinnvoll wäre. Übersetzung (im engeren Sinne) meint die sprachlichen Aspekte. Da aber auch sprachlich gut übersetzte Fragen in verschiedenen Kulturen unterschiedliche Stimuli darstellen können, ist zusätzlich eine Adaption, die auf (nicht-sprachliche) kulturelle Besonderheiten Rücksicht nimmt, erforderlich. Fehler werden hierbei paradoxe Weise nicht nur dort vorkommen, wo die interkulturellen Unterschiede offensichtlich sind, sondern gerade auch dort, wo sie scheinbar geringer sind und leicht übersehen werden können (z.B. wenn wegen der gemeinsamen Sprache keine Übersetzung erforderlich ist). So konnte etwa gezeigt werden, dass bei Fragen zu den Auswirkungen der Berufstätigkeit einer Mutter auf ihre Kinder ostdeutsche Befragte eher an jüngere Kinder und einen größeren Umfang der Berufstätigkeit der Mutter denken als westdeutsche Befragte. Die Angaben der Befragten sind damit nicht direkt vergleichbar.

Als Hilfsmittel bei der Sicherstellung von Äquivalenz ex-ante bieten sich qualitative Verfahren und quantitative Pretests an. Bei den qualitativen Verfahren dominieren kognitive Interviews, bei denen Probleme bei der Beantwortung der Fragen herausgearbeitet werden sollen (Willis 2005). International vergleichende kognitive Interview-Studien sind allerdings im Vergleich zu *interkulturell* vergleichenden kognitiven Interview-Studien in einem Land wegen des hohen Koordinationsaufwands sehr selten (als eine Ausnahme: Miller et al. 2011). Bei quantitativen Pretests wird der für die Hauptstudie vorgesehene Fragebogen unter – mehr oder weniger – realen Bedingungen getestet. Wenn Pretests in den unterschiedlichen Ländern durchgeführt werden und dabei entsprechend hohe Fallzahlen realisiert werden, können auch statistische Verfahren zur Äquivalenzüberprüfung eingesetzt werden.

In der Realität tritt die Situation, dass in allen beteiligten Ländern große quantitative Pretests oder qualitative Vorstudien durchgeführt werden, wegen des hohen Aufwandes praktisch nicht auf. Selbst für den European Social Survey (ESS) waren qualitative Studien die Ausnahme und ein quantitativer Pretest wurde vor der Verabschiedung des englisch-sprachigen Source-Fragebogens nur in zwei der beteiligten Länder durchgeführt. Zudem werden diese Pretests – wie auch beim ISSP – vor allem unter dem Gesichtspunkt der Auswahl von inhaltlich interessanten und methodisch adäquaten Fragen durchgeführt und nicht im Sinne eines Tests des endgültigen, auch in der Hauptstudie verwendeten Messinstruments. Daher ist es oft erforderlich, zusätzlich eine Überprüfung der Äquivalenz ex-post, d.h. auf der Grundlage der in der Hauptstudie erhobenen Daten, vorzunehmen.

#### **56.4.2 Überprüfung der Äquivalenz der Messinstrumente ex-post**

Zur Überprüfung von Äquivalenz ex-post gibt es eine Vielzahl quantitativer Verfahren. Das wohl mit Abstand am häufigsten angewandte Verfahren ist die konfirmatorische Faktorenanalyse. Braun/Johnson (2011) wenden eine ganzer Reihe komplexer und weniger komplexer Verfahren auf das gleiche inhaltliche Problem an und können dadurch zeigen, dass sie im Wesentlichen zu den gleichen Schlussfolgerungen führen (siehe dort auch Literatur zu den einzelnen Verfahren).

Ein Nachteil aller quantitativen Verfahren ist aber, dass sie zwar das Vorhandensein von Problemen feststellen, deren eigentliche Ursachen aber nicht erklären können. Hier bieten sich dann – wie bereits bei der Sicherstellung von Äquivalenz ex-ante (also im Rahmen von Pretests) Probingtechniken an. Da die Durchführung kognitiver Interviews – zumal im internationalen Vergleich – äußerst aufwendig ist, erscheint hier die Durchführung zusätzlicher webbasierter Studien eine mögliche Vorgehensweise. Dabei können Probingfragen – entsprechend Schumans (1966) Vorschlag für „random probes“ – in einen normalen Web-Fragebogen eingeschaltet werden.

Behr et al. (2012) nutzen beispielsweise zur Analyse eines Items zu zivilem Ungehorsam aus dem ISSP sowohl „comprehension“ („Was verbinden Sie mit dem Begriff ‚ziviler Ungehorsam‘, nennen Sie Beispiele“) als auch „category-selection probes“ („Bitte begründen Sie, warum Sie sich für [Antwortalternative] entschieden haben“). Sie können dabei zeigen, dass die geringe Befürwortung von zivilem Ungehorsam in Ländern wie den Vereinigten Staaten und Kanada im Vergleich zu Ländern wie Deutschland und Spanien im Wesentlichen zwei Gründe hat. Zum einen wird in der ersten Gruppe von Ländern ziviler Ungehorsam stärker mit Gewalt assoziiert als in der zweiten. Dies ist ein Methodenartefakt, d.h. die Befragten beantworten faktisch unterschiedliche Fragen. Zum anderen ist das Vertrauen in die Politiker in der zweiten Gruppe von Ländern (noch) geringer als in der ersten. Dies ist ein inhaltliches Ergebnis. In den Daten sind aber faktisch Methodenartefakte und reale Unterschiede konfundiert, die nicht (oder nur sehr schwer) unterschieden werden können.

Braun et al. (2012) können zeigen, dass Befragte aus unterschiedlichen Ländern bei Fragen zu Migranten zwar weitgehend an vergleichbare, der tatsächlichen Migrationsrealität entsprechende Gruppen denken, dass es dabei aber dennoch zu charakteristischen Verzerrungen kommt. Sie verwenden dazu „specific probes“ („An welche Gruppen haben Sie bei der Beantwortung der Frage gedacht?“). Ob die Einstellungen der Befragten verschiedener Länder zu Migranten möglicherweise deshalb nicht sinnvoll miteinander verglichen werden können, weil trotz vergleichbarer Migrationsrealität an unterschiedliche Gruppen gedacht wird, hätte mit den herkömmlichen statistischen Verfahren alleine nicht untersucht werden können.

---

## 56.5 Weitere Aspekte und weiterführende Literatur

Die zuvor aufgeführten Fehlerquellen betreffen alle in Umfragen erhobenen Variablen, z.B. demographische Informationen, Verhaltensberichte und Einstellungsfragen. Demografische Fragen können z.B. genauso falsch übersetzt werden wie Einstellungsfragen. Auch die Abfrage von objektiven Informationen kann Verständnis- und Interpretationsschwierigkeiten aufwerfen und mit Schwierigkeiten beim Abruf der gewünschten Informationen aus dem Gedächtnis verbunden sein. Möglicherweise unterschiedliche Definitionen in den verschiedenen Ländern betreffen nahezu alle sozio-demographischen Variablen (Hoffmeyer-Zlotnik/Wolf 2003, Scheuch 1968). Bei einigen dieser Variablen, z.B. bei der Bildung, werden die Probleme noch durch die große Unterschiedlichkeit der zugrundeliegenden Systeme verschärft.

Das Äquivalenzproblem betrifft im Übrigen nicht nur die interessierenden Variablen, sondern auch die Auswahl von zu vergleichenden Gruppen in den einzelnen Ländern. Bei internationalen Vergleichen ist stets zu berücksichtigen, ob vielleicht gar keine Unterschiede zwischen der gleichen Gruppe hinsichtlich einer Zielvariablen in verschiedenen Ländern festgestellt worden sind, sondern stattdessen unterschiedliche Gruppen miteinander verglichen wurden. Scheuch (1968) diskutiert beispielsweise, ob Bauern in den USA mit denen in Europa vergleichbar sind, ob man also bei einem interkulturellen Vergleich der Antworten nicht ähnliche, sondern verschiedene Gruppen miteinander vergleicht.

In den letzten Jahren sind eine ganze Reihe von Sammelbänden, Monographien und Einzelartikeln erschienen, die sich mit den methodologischen Problemen der interkulturell vergleichenden Umfrageforschung beschäftigen, z.B. Davidov et al. (2011), Harkness et al. (2010), Harkness et al. (2003), Hoffmeyer-Zlotnik/Wolf (2003), van Deth (2013[1998]) sowie van de Vijver/Leung (1997).

## Literatur

- Behr, Dorothée/Braun, Michael/Kaczmarek, Lars/Bandilla, Wolfgang (2012): Item comparability in cross-national surveys: Results from asking probing questions in cross-national Web surveys about attitudes towards civil disobedience. In: *Quality & Quantity*. 2012. DOI: 10.1007/s11135-012-9754-8
- Braun, Michael/Behr, Dorothée/Kaczmarek, Lars (2012): Assessing cross-national equivalence of measures of xenophobia: Evidence from probing in Web Surveys. In: *International Journal of Public Opinion Research*. DOI:10.1093/ijpor/eds034
- Braun, Michael/Johnson, Timothy P. (2010): An illustrative review of techniques for detecting inequivalences. In: Harkness et al. (Hg.): 375-393
- Davidov, Eldad/Schmidt, Peter/Billiet, Jaak (Hg.) (2011): *Cross-cultural Analysis: Methods and Applications*. New York: Routledge
- Harkness, Janet A./Braun, Michael/Edwards, Brad/Johnson, Timothy P/Lyberg, Lars/Mohler, Peter P./Pennell, Beth-Ellen/Smith, Tom (Hg.) (2010): *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Hoboken, NJ: Wiley
- Harkness, Janet A./van de Vijver, Fons J.R./Mohler, Peter P. (Hg.) (2003): *Cross-cultural Survey Methods*. Hoboken, NJ: Wiley
- Hoffmeyer-Zlotnik, Jürgen H.P./Wolf, Christof (Hg.) (2003): *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*. New York: Kluwer/Plenum
- Miller, Kristen/Fitzgerald, Rory/Padilla, José-Luis/Willson, Stephanie/Widdop, Sally/Caspar, Rachel/Dimov, Martin/Grey, Michelle/Nunes, Cátia/Prüfer, Peter/Schöbi, Nicole/Schoua-Glusberg, Alisú (2011): Design and analysis of cognitive interviews for comparative multinational testing. In: *Field Methods* 23: 379-396
- Przeworski, Adam/Teune, Henry (1966): Equivalence in cross-national research. In: *Public Opinion Quarterly* 30: 551-568
- Przeworski, Adam/Teune, Henry (1970): *The Logic of Comparative Social Inquiry*. New York: Wiley
- Rokkan, Stein (Hg.) (1968): *Comparative Research across Cultures and Nations*. Paris: Mouton
- Scheuch, Erwin K. (1968): The cross-cultural use of sample surveys: Problems of comparability. In: Rokkan (Hg.): 176-209
- Schuman, Howard (1966): The random probe: A technique for evaluating the validity of closed questions. In: *American Sociological Review* 31: 218-222
- Van Deth, Jan W. (Hg.) (2013[1998]): *Comparative Politics: The Problem of Equivalence*. ECPR Classics. Colchester: ECPR Press
- Van de Vijver, Fons J.R./Leung, Kwok (1997): *Methods and Data Analysis for Cross-cultural Research*. Thousand Oaks: Sage
- Willis, Gordon B. (2005): *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage

Ingwer Borg

## 57.1 Merkmale und Varianten von Mitarbeiterbefragungen

Eine Mitarbeiterbefragung (MAB) fragt Mitarbeiter (aus allen oder ausgewählten Ebenen bzw. Bereichen einer Organisation) nach ihren Meinungen und Einstellungen zu arbeitsbezogenen Themen mit der Absicht, mit der Befragung selbst, ihren Messwerten und ihrer Aufarbeitung das Erreichen der Ziele der Organisation zu fördern. Man kann einige gängige MAB-Typen unterscheiden, auch wenn diese in der Praxis meist als Mix realisiert werden:

(1) *MAB als demoskopisches Instrument*: Sie soll zeigen, was „die Mitarbeiter“ über wichtige arbeits- und unternehmensbezogene Themen denken, wo sie der Schuh drückt, welche Einstellungen und Pläne sie haben, wo sie Verbesserungspotentiale sehen. Die Befragungsdaten sollen zudem Unterschiede zwischen Teilen der Organisation (z.B. Männer vs. Frauen, verschiedene Niederlassungen, Führungskräfte und Nicht-Führungskräfte) klären.

(2) *MAB als Benchmarking-Instrument*: Sie soll Kennzahlen generieren, die es erlauben, die Organisation mit anderen Organisationen in Bezug auf Kriterien wie z.B. das Engagement oder die Zufriedenheit der Mitarbeiter zu vergleichen.

(3) *MAB als Klimabefragung mit Rückspiegelung*: Sie soll die Zufriedenheit der Mitarbeiter und das Betriebsklima messen, um diese Ergebnisse dann in zahlreichen „Survey Feedback“-Veranstaltungen mit dem Ziel zu diskutieren, gezielte Verbesserungsmaßnahmen zu lancieren.

(4) *MAB als Instrument zur Messung von Führungsleistung*: Sie soll Kennzahlen erheben, die zur Bewertung (z.T. auch zur Bonusfindung) der Führungskräfte dienen.

(5) *Strategische MAB*: Die MAB soll die Aufmerksamkeit aller Mitarbeiter auf wichtige strategische Themen lenken und die Befragungsergebnisse dann in nachfolgenden Aktionen gezielt zur Verbesserung der Zielerreichung verwenden.

Die verschiedenen MAB-Typen verwenden unterschiedliche Designs. Demoskopische MABs und Benchmarking-MABs kommen z.B. mit echten (oft sogar recht kleinen) Stichproben aus, während die anderen MAB-Typen die Einbindung möglichst aller Mitarbeiter erfordern und demnach von vornherein als Vollbefragungen angelegt sind. Eine Klimabefragung ist, ebenso wie die demoskopische MAB, ein singuläres Projekt, während für die anderen MAB-Typen stets Wiederholungen im ein- oder zweijährigen Zyklus üblich sind. Beim Benchmarking müssen z.B. in ähnlicher Weise Zeitreihen hergestellt werden, weil nur so Trendverläufe und Veränderungen frühzeitig erkennbar sind. Für strategische MABs sind Wiederholungen immer notwendig, weil Lernen Feedback und Verstärkung benötigt.

---

## 57.2 Besonderheiten von Mitarbeiterbefragungen

Zur optimalen Planung eines konkreten MAB-Projekts ist es erforderlich, eine Reihe von Entscheidungen zu treffen, die z.T. deutlich von denen einer Bevölkerungsbefragung abweichen.

### 57.2.1 Schutz des einzelnen Befragten

In Deutschland kommen bislang nur *anonyme* MABs zum Einsatz. Eine anonyme MAB ist immer freiwillig und erhebt nie den Namen der Befragten. Zusätzlich wird meist Folgendes (oft in einer Betriebsvereinbarung) festgelegt: (1) Die Durchführer der MAB (meist ein externes Institut) liefern nur genau einen Auswertungsbericht pro Führungskraft. Zusätzliche Berichte (insbesondere solche ohne bestimmte Teilnehmergruppen, z.B. ohne die Antworten der weiblichen Mitarbeiter) können nicht ohne besondere Genehmigung nachbestellt werden. (2) Auswertungsberichte werden erst ab einer minimalen Gruppengröße (meist: fünf aktive Teilnehmer) erstellt. (3) Um die Möglichkeit der „Rasterfahndung“ zu vermeiden, werden nur wenige (meist nur zwei oder drei) demografische Variablen erhoben. Unverzichtbare demografische Informationen (Organisationseinheit, Führungsebene) werden im Fragebogen automatisch codiert. Demografische Variablen (z.B. Geschlecht), die zur Klärung besonderer Fragen manchmal von Interesse sind, werden dem Befragten zur freiwilligen Bearbeitung angeboten.

In Nordamerika und in weltweiten MABs gibt es zunehmend *vertrauliche* MABs, bei denen jede Person einen „*identifying code*“ bekommt, der es der Personalforschung erlaubt, die MAB-Daten anderen Daten derselben Person zuzuordnen. Das ermöglicht weit bessere Analysen als anonyme MABs. So kann man z.B. die MAB-Daten mit Informationen aus Einstellungsinterviews, Leistungsindikatoren, Krankheitsdaten oder aus alten MABs verknüpfen. Voraussetzung für die Validität solcher Analysen ist jedoch, dass die Befragten davon überzeugt sind, dass ihre Identität tatsächlich vertraulich behandelt wird und dass die Daten nur für allgemeine Forschungszwecke verwendet werden (Saari et al. 2011, Weiner et al. 2011). Dazu gehören u.a. gewisse Regeln, z.B. dass die identifizierenden

Informationen zu den Personen-Codes in einer Extra-Datei abgespeichert werden, zu der der Zugang strikt auf wenige Personen der Personalforschung begrenzt wird.

### **57.2.2 Schutz der einzelnen Führungskraft**

Führungskräfte, die in einer MAB bloßgestellt oder unqualifiziert bewertet werden, haben verminderte Chancen, ihre Führungsleistung zu verbessern. Um dieses Risiko zu reduzieren, werden in einer MAB statistische Quervergleiche ähnlicher Organisationseinheiten nur dem direkt zuständigen Manager, nicht aber den Einheiten selbst, zugänglich gemacht. Der Manager ist aufgefordert, die Ergebnisse mit jeder unterstellten Führungskraft (meist in einem Vier-Augen-Gespräch) zu besprechen und zu re-kontextualisieren, d.h. in den Gesamtzusammenhang (Ziele, Rahmenbedingungen, Kunden usw.) zu stellen. Die Mitarbeiter werden aufgefordert, sich auf „ihre“ Ergebnisse (und auf Aufwärtsvergleiche) zu konzentrieren und keine Quervergleiche anzustellen. An die einzelne Führungskraft wird schließlich die Erwartung gerichtet, die MAB-Ergebnisse mit ihren Mitarbeitern zu diskutieren und darauf sichtbar zu reagieren, ggf. mit (in der Führungslinie abgestimmten) Aktionen, die den (gegebenen) Zielen dienen. Das soll sie vor unnötiger Zusatzarbeit und Aktionismus schützen.

### **57.2.3 Stichprobe**

Fast jede MAB wird heute als *Vollerhebung* durchgeführt (Häder/Häder, Kapitel 18 in diesem Band), weil die Daten später bis auf einzelne Abteilungen heruntergebrochen werden.

### **57.2.4 Typischer MAB-Fragebogen**

Ein MAB-Fragebogen wird so konstruiert, dass er die für die Ziele der MAB wichtigsten Themen hinreichend genau und valide in einem vorgegebenen Zeitfenster von nicht mehr als einer halben Stunde adressiert. Bei guten Formulierungen und Formatierungen kann man in diesem Zeitfenster Antworten für ca. 50-60 Items erheben.

Die ersten MABs dienten der Messung von *Arbeitszufriedenheit*. Ihre Items fokussierten auf die Zufriedenheit mit der Tätigkeit selbst, der Bezahlung, den Kollegen, den Vorgesetzten usw. (Neuberger 1974). Heute steht meist das Engagement der Mitarbeiter (Macey/Schneider 2008; Wiley 2010) im Vordergrund, das mit einer Reihe von Proxy-Variablen (insb. zur Arbeitszufriedenheit und zum Commitment) gemessen wird (Harrison et al. 2006). Da außer Engagement für die Leistungserbringung auch geeignete Arbeitsmittel, Arbeitsabläufe und entsprechendes Know-how nötig sind, werden diese (von Schiemann (2009) zusammenfassend als „*Capabilities*“ bezeichnet) ebenfalls stets in einer MAB adressiert. Dazu kommt noch das „*Alignment*“ der Mitarbeiter in Bezug auf die Ziele und die Strategie des Unternehmens.

Schiemann (2009) fasst die Themenbereiche Alignment, Engagement und Capabilities im Akronym ACE zusammen. Nachhaltige Leistung hängt zudem noch davon ab, ob die Mitarbeiter die Ergebnisse ihrer Anstrengungen sehen („Results“: zeitnah, kausal richtig attribuiert usw.), und wie sie die arbeitsbezogenen Belohnungen wahrnehmen („Rewards“: fair, gerecht, angemessen usw.). Damit wird aus ACE dann RACER, eine Merkformel für die wichtigsten Themenkategorien einer MAB (Borg/Mastrangelo 2008). Vom RACER-Baukasten kann man noch weitergehen und die Zusammenhänge der Variablen in systemischen Theorien wie dem Leistungs-Zufriedenheits-Motor (Borg 2003) artikulieren.

### **57.2.5 Einzel-Item, Skalen und Standard-Skalen**

In einer MAB hat man nur bis zu 30 Minuten Zeit zur Messung von zahlreichen verschiedenen Themen. Man muss sich daher oft mit *einem einzigen Item* pro Thema begnügen. „Single Item“-Messungen sind aber nicht prinzipiell nachteilig (Wanous et al. 1997), vor allem nicht bei Messungen, die über Personen aggregiert werden. Sie können sogar valider sein als Multi-Item-Standardskalen, wenn sie einen komplexen Gegenstand (wie z.B. Arbeitszufriedenheit) zusammenfassend ansprechen (Nagy 2002).

### **57.2.6 Funktionen von MAB-Items**

Primäre Aufgabe von Items in einer MAB ist das *Messen* von Meinungen und Einstellungen. Das kann verschiedenen Zwecken dienen (z.B. Benchmarking, Bonusallokation, Früherkennung von Handlungsbedarfen). Diese Funktion gilt aber nicht für alle Items. Manche MAB-Items sollen vor allem ein *Thema platzieren*. Andere Items sollen bestimmte *Folgeprozesse vorbereiten und beschleunigen*, z.B. das Item: „Bei gewissen Veränderungen in meinem Arbeitsumfeld könnte ich viel produktiver sein“ (mit der üblichen Zustimmung- vs. Ablehnung-Antwortskala). Hier stimmen normalerweise ca. zwei Drittel der Mitarbeiter zu, ein Ergebnis, das das Management dazu legitimiert, das Thema Produktivität anzusprechen und mit den Mitarbeitern zu diskutieren.

### **57.2.7 Besonderheiten bei der Formulierung von MAB-Items**

MAB-Items werden fast ausschließlich als *geschlossene Items* formuliert und zwar i.d.R. als 5-stufige Fragen mit einer Antwortskala von „trifft voll zu“ bis „trifft überhaupt nicht zu“ (Franzen, Kapitel 51 in diesem Band). Der Grund hierfür ist, dass sich so Fragebögen konstruieren lassen, die *schnell* und *reliable* zu beantworten sind, weil immer das gleiche Frage- und Antwortschema verwendet wird. Die Fragebögen sind zudem *kompakt*. Das führt zu größeren Beteiligungsquoten und kostet weniger Arbeitszeit.

## 57.3 Zusätzliche Aspekte von besonderer Bedeutung

### 57.3.1 Regeln

Regeln zur Durchführung und Auswertung müssen für jede MAB vorab ausformuliert und an alle Mitarbeiter kommuniziert werden. Man kann nicht einfach einen Fragebogen verteilen und die Mitarbeiter bitten, diesen auszufüllen. Das würde zu Unruhe und zu vielen Fragen führen, die von den einzelnen Vorgesetzten oder vom MAB-Projektteam nicht mehr beantwortet werden können.

### 57.3.2 Datenanalyse

In jeder MAB ist es wichtig, die Auswertung der Befragung *zügig* zu erstellen, damit der Schwung nicht verloren geht und damit die Ergebnisse aktuell bleiben. In der Praxis wird heute erwartet, dass die *Berichte* für jede Organisationseinheit – und das sind in größeren Organisationen Tausende! – *wenige Tage* nach der Befragung vorliegen. Das ist nur möglich, wenn diese per Computer-Algorithmus erstellt werden. Dieser muss vor Ende der Befragung programmiert sein, auf der Basis von Entscheidungen zu Fragen wie: „Wer bekommt was wann zu sehen? Welche Statistiken werden verwendet? Welche Vergleichswerte kommen zum Einsatz? Was wird in der Einleitung geschrieben?“ Sobald die Daten der Befragung vorliegen, laufen sie durch das Programm. Interpretationen der Ergebnisse sind dabei nicht möglich. Die Berichte enthalten nur vorab formulierte Texte, Statistiken und Vergleichswerte. Des Weiteren werden in MAB-Berichten überwiegend *einfache, verständliche* Statistiken verwendet, wie „Prozent Zustimmung“, oft kombiniert mit einfachen Grafiken wie gestapelten Histogrammen oder Ampeln, und stets kombiniert mit Vergleichswerten (historische Vergleiche, Aufwärtsvergleiche, externe Benchmarks).

### 57.3.3 MAB-Folgeprozesse

Für die Prozesse, die der Befragung nachfolgen, muss frühzeitig ein Plan gemacht und allgemein kommuniziert werden, weil sonst das MAB-Projekt schnell versickert oder in Chaos und Aktionismus endet. Das heute übliche Design für die Folgeprozesse ist eine *Top-Down*-Vorgehensweise. Sie beginnt mit einer Präsentation der Befragungsergebnisse (mit vertiefter Datenanalyse insb. auch von Zusammenhängen) durch einen Experten an die Geschäftsleitung, die sich zu den Befunden positioniert und einige Handlungsfelder definiert. Anschließend verfährt man in ähnlicher Weise auf der nächsten Ebene (Bereichsmanagement) und setzt den Prozess kaskadisch bis zu den einzelnen Teams oder Arbeitsgruppen an der Basis fort.

In jeder MAB wird erwartet, dass diese Ergebnisse der Geschäftsleitung (oft auch dem Betriebsrat, den Werksleitungen oder anderen Interessensgruppen) präsentiert werden. Da die Rezipienten einer MAB-Präsentation keine Wissenschaftler sind, müssen die Statistiken so übersetzt werden, dass sie verstanden und als Fakten akzeptiert werden. Dabei ergeben sich besondere Probleme, wenn die Befunde „negativ“ sind, vor allem für das Selbstbild der Rezipienten. Sie versuchen dann meist, die Zuverlässigkeit und Validität der Aussagen mit zahlreichen „Gegenargumenten“ (z.B. „Alles nur Meinungen!“, „Stimmt objektiv nicht!“) zu testen. Solchen Gegenargumenten ebenso wie dem Eindruck, dass alles sowieso „offensichtlich“ ist, kann man dadurch entgegen wirken, dass die Rezipienten vorab schriftliche Prognosen abgeben, die man später mit tatsächlichen Befunden abgleicht. Man kann zeigen, dass sich dadurch nicht nur die Interessantheit der Befunde erhöht, sondern dass diese später auch besser erinnert werden (Borg/Hillenbrand 2003).

Bei der reinen Meinungsbefragung und bei Benchmark-Umfragen ist die MAB mit der Präsentation der Befunde an das obere Management für den Befrager abgeschlossen. Bei anderen MAB-Ansätzen soll mit den Befunden intensiv weiter gearbeitet werden (Diskussion, qualitative Nacharbeit, Vorschläge zur Verbesserung, etc.). Das geschieht meist in speziellen „Survey Feedback Workshops“ (Nadler 1977), die vorab entsprechend geplant und befähigt werden müssen.

Auf eine MAB „muss“ vom Management *reagiert* werden (Kraut 2006). Reaktionen können Stellungnahmen hierzu sein, Entscheidungen für oder gegen bestimmte Aktionslinien oder die Einleitung bestimmter Aktionen (auf verschiedenen Ebenen).

#### **57.3.4 Weiterführende Literatur**

MABs sind komplexe Projekte, die ein optimales Gesamt-Design ihrer vielen Vorgänge unter den jeweils gegebenen Rahmenbedingungen erfordern. Detaillierte Informationen und viele Praxis-Beispiele hierzu finden sich z.B. in Borg (2003), Borg (2014) oder Borg/Mastrangelo (2008).

## Literatur

- Borg, Ingwer (2003): Führungsinstrument Mitarbeiterbefragung. (3. Aufl.) Göttingen: Hogrefe
- Borg, Ingwer (2014). Mitarbeiterbefragungen in der Praxis. Göttingen: Hogrefe
- Borg, Ingwer/Gabler, Siegfried (2002): Zustimmungsanteile und Mittelwerte von Likert-skalierten Items. In: ZUMA Nachrichten 50: 7-25
- Borg, Ingwer/Hillenbrand, Carola. (2003): Prognosen als Methode zur Reduktion der Offensichtlichkeit von Umfragebefunden: Ein Pilotexperiment. In: ZUMA-Nachrichten 52: 7-19
- Borg, Ingwer/Mastrangelo, Paul M. (2008): Employee surveys in management. Cambridge, MA: Hogrefe-Huber
- Harrison, David A./Newman, Daniel A./Roth, Philip L. (2006): How important are job attitudes? Meta-analytic comparisons of integrative behavioral outcomes and time sequences. In: Academy of Management Journal 49: 305-325
- Kraut, Allen I. (2006): Getting action from organizational surveys. San Francisco, CA: Jossey-Bass
- Macey, William H./Schneider, Benjamin (2008): The meaning of employee engagement. In: Industrial and Organizational Psychology 1: 3-30
- Nadler, David A. (1977): Feedback and organization development: Using data-based methods. Reading, MA: Addison-Wesley
- Nagy, Marc S. (2002): Using a single-item approach to measure facet job satisfaction. In: Journal of Occupational and Organizational Psychology 75: 77-86
- Neuberger, Oswald (1974): Messung der Arbeitszufriedenheit. Stuttgart: Kohlhammer
- Saari, Lise M./Scherbaum, Charles A. (2011): Identified employee surveys: Potential promise, perils, and professional practice guidelines. In: Industrial and Organizational Psychology 4: 435-448
- Schiemann, William A. (2009): Reinventing talent management: How to maximize performance in the new marketplace. Hoboken, N. J.: Wiley
- Wanous, John P./Reichers, Arnon E./Hudy, Micheal J. (1997): Overall job satisfaction: How good are single-item measures? In: Journal of Applied Psychology 82: 247-252
- Weiner, Sara P/Jolton, Jeffrey A./Dorio, Jay M./Klein, Cameron/Herman, Anne E. (2011): Attributed surveys from the perspective of practitioners: Providing more value than risk. In: Industrial and Organizational Psychology 4: 468-472
- Wiley, Jack W. (2010): Strategic employee surveys: Evidence-based guidelines for driving organizational success. San Francisco, CA: Jossey-Bass

Julia Heinen und Susanne König

---

## 58.1 Einleitung

In der Geschichte der empirischen Sozial- und Marktforschung hat sich hinsichtlich der Befragung von Kindern und Jugendlichen ein Einstellungswandel vollzogen: Wurden Heranwachsende zunächst für den relativ komplexen Frage-Antwort-Prozess als unzureichend kompetent eingeschätzt, gelten sie spätestens seit Mitte der 1980er-Jahre als ernstzunehmende Interviewpartner (France 2004: 177). Die Gründe für diese Einstellungsänderung sind vielfältig. Zum einen haben sich Heranwachsende zu einer wichtigen Zielgruppe in der Markt- und Medienforschung entwickelt, was sicherlich zu einer Verbreitung von Befragungen im Kindersegment beigetragen hat (Lipski 2000: 77). Kinder- und Jugendliche sind sowohl aufgrund ihres Einflusses auf das Konsumverhalten ihrer Eltern (Scott 1997: 331) als auch aufgrund eigener finanzieller Mittel durch an sie gerichtete Taschengeldzahlungen, Geldgeschenke u.ä. von ökonomischer Bedeutung (Lipski 2000: 77). Zum anderen wird den Interessen und Rechten von Kindern und Jugendlichen zunehmend eine höhere Relevanz beigemessen (Scott 1997: 331). Gesellschaftspolitisch findet dies insbesondere in der am 20. November 1989 verabschiedeten UN-Kinderrechtskonvention seinen Ausdruck, mit der Kindern eigene Rechte zugesprochen werden. Dennoch bleibt die Frage zu klären, ob relativ junge Menschen für den komplexen Frage-Antwort-Prozess ausreichend kompetent sind.

Die Auffassung, dass die intellektuellen Fähigkeiten von Kindern und Jugendlichen für die Befragung unzureichend sind, gilt bei einem Teil der Wissenschaftler als „overstated, if not simply untrue“ (Scott 1997: 331). Studien, die Personen unter 18 Jahren als Befragte in Untersuchungen mit einbeziehen, setzen jedoch einen ausreichenden kognitiven, sozialen und kommunikativen Entwicklungsstand voraus, der vor allem in Abhängigkeit vom Alter, aber auch vom Geschlecht und vom sozioökonomischen Hintergrund der Kinder und Jugendlichen steht (Scott 1997: 335).

Festzuhalten bleibt, dass in der aktuellen Literatur eine Präferenz dafür besteht, in empirischen Forschungsvorhaben über Heranwachsende diese selbst in die Untersuchung mit einzubeziehen. Dabei ist die Anerkennung von Wahrnehmungsunterschieden ein wesentliches Argument dafür, Kinder und Jugendliche selbst anstelle der Eltern stellvertretend für die Heranwachsenden zu befragen; die von den Erziehungsberechtigen perzipierten Antworten ihres Nachwuchses und die eigene kindliche Perspektive können verschieden sein und zu unterschiedlichen Angaben führen (Heinen 2012). Welche Besonderheiten bei der standardisierten Befragung (Reinecke, Kapitel 44 in diesem Band) von Heranwachsenden zu beachten sind, ist Thema des vorliegenden Artikels (zur qualitativen Befragung von Kindern verweisen wir auf Vogl (2014)). Dabei sind die sich noch in der Entwicklung befindenden Kompetenzen der relativ jungen Befragten sowie die sich daraus ergebende forschungsethische und moralische Verantwortung zu berücksichtigen. Daher steht insbesondere die Frage nach dem Feldzugang zu Minderjährigen im Fokus. So geben z.B. beim Feldzugang zu Heranwachsenden die „Richtlinie für die Befragung von Minderjährigen“ sowie bei Schulbefragungen die landesspezifischen Schulgesetze verbindliche Handlungsanleitungen und Orientierungshilfen. Zudem sind Aspekte der Datenqualität stets dem (auch von Erziehungsberechtigen antizipierten) Wohle des Kindes unterzuordnen. Die Formulierung von Fragen und die Auswahl von Fragetypen und Antwortoptionen sind so einfach wie möglich zu gestalten.

---

## 58.2 Feldzugang

Bei der Befragung von Heranwachsenden stellen in der Regel dritte Instanzen wie u.a. Schule und Familie den Zugang zu Kindern und Jugendlichen her. Es ist aber auch möglich, mit Hilfe von Massenmedien auf die jeweiligen Forschungsvorhaben aufmerksam zu machen. Dabei sind bei der Verallgemeinerung von Ergebnissen dieser Studien stets die besonderen Eigenschaften des rekrutierten Samples zu berücksichtigen. Aber wer gehört eigentlich zur Gruppe der „Kinder- und Jugendlichen“? Und vor allem: Welche rechtlichen Aspekte sind beim Feldzugang zu Minderjährigen zu berücksichtigen?

Im deutschen Jugendschutzgesetz (JuSchG) werden Kinder als Personen definiert, die noch keine 14 Jahre alt sind; Jugendliche sind Personen, die 14, aber noch nicht 18 Jahre alt sind (§ 1 JuSchG). Im Folgenden wird diese Definition berücksichtigt und ggf. auf Charakteristika spezifischer Altersgruppen hingewiesen.

Der Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. (ADM), die Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (ASI), der Bundesverband Deutscher Markt- und Sozialforscher e.V. (BVM) sowie die Deutsche Gesellschaft für Online-Forschung e.V. (D.G.O.F.) haben gemeinsam eine „Richtlinie für die Befragung von Minderjährigen“ herausgegeben. „Diese Richtlinie ist Teil der Standesregeln der deutschen Markt- und Sozialforschung, wie sie sich aus dem Gesetz und den methodischen Standards, aber auch aus der Verkehrssitte ergeben“ (ADM 2006: 2). Sie basiert insbesondere auf einer Abwägung zwischen Persönlichkeitsrechten der Befragten und dem Recht

auf Forschung der Markt- und Sozialforscher (ADM 2006: 2). Sowohl die Persönlichkeits- als auch die Forschungsrechte nehmen Bezug auf das Grundgesetz (GG): Auf der einen Seite ist die Würde des Menschen – und somit auch die der Kinder und Jugendlichen – unantastbar (Artikel 1 Absatz 1 GG) und es gilt das Recht auf freie Persönlichkeitsentfaltung, soweit dadurch nicht die Rechte anderer verletzt werden und gegen die verfassungsmäßige Ordnung oder das Sittengesetz verstößen wird (Artikel 2 Absatz 1 GG). Auf der anderen Seite sind „Kunst und Wissenschaft, Forschung und Lehre [...] frei“ (Artikel 5 Absatz 3 Satz 1 GG). Aber was heißt das konkret für die Ausgestaltung der Befragung mit Minderjährigen?

Bei Befragungen mit Kindern und Jugendlichen ist – ebenso wie bei Interviews mit Volljährigen – auf die Freiwilligkeit der Teilnahme an der Untersuchung hinzuweisen und die Einwilligung des Befragten in die Verarbeitung und anonymisierte Nutzung der erhobenen Daten einzuholen. Die Fähigkeit zur Einwilligung ist jedoch von der sogenannten „Einsichtsfähigkeit“ der Betroffen abhängig (ADM 2006: 1). Als Kriterium für diese Einsichtsfähigkeit im Zusammenhang mit der Befragung von Kindern und Jugendlichen gilt den Herausgebern der o. g. Richtlinie zufolge, „ob der Minderjährige die Konsequenzen der Verwendung seiner Daten übersehen und dementsprechend Stellung nehmen kann“ (ADM 2006: 1) oder nicht.

Während bei Jugendlichen zwischen 14 und 17 Jahren grundsätzlich eine ausreichende Einsichtsfähigkeit unterstellt werden kann und diese somit selbst – anstelle ihrer Eltern – in die Befragung einwilligen können, ist bei Kindern unter elf Jahren i.d.R. eine ungenügende Einsichtsfähigkeit vorhanden und es ist zwingend, das Einverständnis der gesetzlichen Vertreter des Kindes in das Interview einzuholen. Dabei ist eine mündliche Einwilligung ausreichend (ADM 2006: 1).

Bei der Gruppe der 11- bis 13-Jährigen gestaltet sich die Beurteilung der Einsichtsfähigkeit schwieriger und es liegt in der Verantwortung des durchzuführenden Instituts entweder die Einsichtsfähig feststellen zu lassen oder die Einwilligung der Eltern einzuholen (ADM 2006: 1). Da die Herausgeber der Richtlinie keine Operationalisierung der „Einsichtsfähigkeit“ vorgeben, sind die Forscher auch aus forschungsethischer Sicht auf der „sicheren Seite“, wenn sie sich die Einverständniserklärung der Eltern zur Teilnahme ihres Kindes an der Untersuchung einholen.

---

### **58.3 Befragungssituation**

Bei Befragungen der Kinder an ihrem Wohnort ist zu beachten, dass diese nicht ohne Kenntnis eines in der Wohnung anwesenden Erwachsenen stattfinden sollen (ADM 2006: 1). Auch ist grundsätzlich dem Wunsch des gesetzlichen Vertreters, bei einem Interview mit dem Kind anwesend zu sein, entgegenzukommen. Dies kann zwar die Datenqualität mindern, dennoch ist das Anliegen der Eltern als vorrangig zu behandeln (ADM 2006: 1). Der Grund für den negativen Einfluss auf die Datenqualität bei Anwesenheit der Eltern ist, dass das Interview im Kontext familiärer Machtkonstellationen stattfindet.

Kinder richten sich in einer solchen Situation relativ häufig nach den perzipierten Anforderungen der Erziehungsberechtigten, wodurch ein verstärkt sozial erwünschtes Antworten begünstigt wird (Barker/Weller 2003: 211).

Bei Interviews von Kindern und Jugendlichen in den Räumen einer Schule – sogenannte „Schulklassenbefragungen“ – ist grundsätzlich in Deutschland das jeweilige Landesschulrecht zu beachten. So besagt z.B. das Schulgesetz (SchulG) für das Land Nordrhein-Westfalen, ähnlich wie die o.g. Richtlinie, dass Daten von Schülern nur mit deren Einwilligung erhoben werden dürfen. Minderjährige Schüler gelten als einwilligungsfähig, „wenn sie die Bedeutung und Tragweite der Einwilligung und ihre rechtlichen Folgen erfassen können und ihren Willen hiernach zu bestimmen vermögen“ (§ 120 Absatz 2 Satz 2 und 3 SchulG NRW). Zudem sind Befragungen nur zulässig, „wenn dadurch die Bildungs- und Erziehungsarbeit und schutzwürdige Belange einzelner Personen nicht beeinträchtigt werden oder die Anonymität der Betroffenen gewahrt bleibt“ (§ 120 Absatz 4 SchulG NRW). Die Entscheidung trifft die Schulleitung und in Angelegenheiten besonderer oder überörtlicher Bedeutung ist die obere Schulaufsichtsbehörde zu unterrichten.

Hinsichtlich der Datenqualität ist bei Befragungen in Schulen zu berücksichtigen, dass das Antwortverhalten der Schüler insbesondere unter der Kontrolle offizieller Instanzen verfälscht werden kann (Barker/Weller 2003: 211ff.): Die Schule kann als Ort der Befragung bei dem Kind die Assoziation mit einer Testsituation hervorrufen. Das Kind ist dann verstärkt darauf bedacht, die Fragen „richtig“ anstatt ehrlich zu beantworten (Punch 2002: 328).

Einen Sonderfall, um einen Zugang zu Kindern und Jugendlichen zu bekommen, stellen Internetbefragungen dar. Wie kann online eine Einwilligung der Eltern zu der Befragung eingeholt werden? In der Praxis werden die Kinder relativ häufig zu Beginn der Onlineumfrage darum gebeten, zunächst ihre Eltern zu fragen, ob sie an der Befragung teilnehmen dürfen. Füllen sie den Fragebogen aus, wird ein Einverständnis der Erziehungsberechtigten unterstellt. Dieses Vorgehen wird von Forschungsinstituten vergleichsweise oft aus finanziellen Gründen angewendet. Wie viele befragte Kinder dabei tatsächlich bei ihren Eltern die Einwilligung in die Teilnahme an der Befragung einholen, ist bisher unbekannt.

Auch bei der Onlinebefragung mit Heranwachsenden kann die Datenqualität durch die Anwesenheit Dritter beeinflusst werden. Bei einer Studie mit 8- bis 17-jährigen Nutzern spezifischer Kinderwebseiten gaben etwa drei Viertel der Interviewten an, zum Zeitpunkt der Befragung alleine vor dem Computer zu sitzen (Heinen 2012: 148f.).

---

## 58.4 Fragen und Fragetypen

Ganz allgemein kann die Qualität der von Kindern und Jugendlichen erhobenen Daten verbessert werden, wenn die Verwendung von komplizierten Formulierungen, langen Sätzen sowie Fremdwörtern im Fragebogen vermieden wird (Punch 2002: 324). Die Fragen sollten kurz und eindeutig formuliert sein und die Bearbeitung des gesamten Fragebogens sollte relativ wenig Zeit in Anspruch nehmen, da Kinder im Vergleich zu Erwachsenen eine

geringere Konzentrationsspanne aufweisen (Arendt/Rössler 2009: 360; Punch 2002: 324). Konkret wird bei 6- bis 13-Jährigen empfohlen, eine Befragungsdauer von fünf Minuten nicht wesentlich zu überschreiten (Arendt/Rössler 2009: 360).

Des Weiteren können Heranwachsende ausschließlich Fragen beantworten, die ihren eigenen Lebensbereich betreffen. Wird hingegen nach Merkmalen gefragt, die außerhalb ihrer unmittelbaren Erfahrung liegen, sind ungültige Antworten zu erwarten (Scott 1997: 332, 340; Lipski 2000: 82). So können Kinder i.d.R. keine validen Auskünfte über sozio-demografische Kennzeichen ihrer Eltern geben. Zwar werden mit zunehmendem Alter z.B. die Angaben zur Bildung und zum Beruf der Eltern zuverlässiger, bis zum zwölften Lebensjahr ist jedoch erfahrungsgemäß mit fehlerhaften Antworten zu rechnen (Lipski 2000). Dagegen können Fragen zur Anzahl der im Haushalt lebenden Personen oder zum Alter sowie zum Geschlecht des Heranwachsenden bereits früher richtig beantwortet werden, da sie auf den direkten Lebensbereich des Kindes Bezug nehmen (Lipski 2000: 80f.; Kränzl-Nagl/Wilk 2000: 65). Dazu gehören z.B. auch Fragen nach Themeninteressen, Freizeit- oder Musikvorlieben sowie nach Lieblingsfernsehsendungen oder -internetseiten der Kinder (Heinen 2012). Dagegen sollten Fragen zu individuellen Wertorientierungen insbesondere bei Heranwachsenden unter zwölf Jahren vermieden werden (Kuchenbuch 2003: 3). Verinnerlichte Werte können erst im Laufe der Adoleszenz abgefragt werden, weil sich diese in früheren Lebensjahren noch nicht ausreichend gefestigt haben (Kuchenbuch 2003: 3). Im Rahmen von Jugendstudien – wie dem „Jugendsurvey“ vom Deutschen Jugendinstitut (DJI) oder der „Shell Jugendstudie“ – ist es durchaus üblich, Wertvorstellung bei Personen ab dem zwölften Lebensjahr abzufragen. Auch sei angemerkt, dass Kindern das Schätzen von Zeiträumen und räumlichen Distanzen schwer fällt (Kränzl-Nagl/Wilk 2000: 66).

Ganz allgemein ist davon auszugehen, dass die Datenqualität bei der standardisierten Befragung von Kindern und Jugendlichen bei zunehmender Komplexität der Fragen sinkt. So weisen Ranking-Fragen, bei denen verschiedene Antwortkategorien nach einem vorgegebenen Kriterium in eine Rangfolge gebracht werden müssen, ein höheres Komplexitätsniveau auf als Fragen des Paarvergleichs, bei denen sich der Befragte lediglich zwischen zwei Antwortalternativen entscheiden soll. Auch gelten Matrixfragen, bei denen i.d.R. mehrere untereinander aufgeführte Items mit Hilfe einer mehrstufigen Skala zu beurteilen sind, als relativ komplex. Neben Skaleneffekten und Effekten der numerischen Skalenbenennung nehmen Fragen- und Antwort-Reihenfolgeeffekte mit zunehmendem Alter von Kindern und Jugendlichen ab (Fuchs 2005).

König (2012) konnte mittels Onlinebefragungen 8- bis 13-Jähriger zeigen, dass den Kindern die Beantwortung von Ranking-Fragen mit drei Antwortoptionen keine Schwierigkeiten bereitete. Des Weiteren konnte die Datenqualität bei fünfstufig skalierten, positiv formulierten Items, die den Kindern am Bildschirm in Form einer Matrix-Frage gezeigt wurden, als gut beurteilt werden. Dabei ist zu beachten, dass die Befragten kaum zwischen den Skalenpunkten „stimme eher nicht zu“ und „stimme überhaupt nicht zu“ differenzieren. Auch hat sich herausgestellt, dass bei Internetbefragungen mit Kindern die Datenqualität durch den Einsatz von multimedialen Elementen zur Unterstützung des Frageninhalts verbessert werden kann.

## Literatur

- ADM (2006): Richtlinie für die Befragung von Minderjährigen.([https://www.adm-ev.de/index.php?eID=tx\\_nawsecuredl&u=0&file=fileadmin/user\\_upload/PDFS/R05\\_D.pdf&t=1365964220&hash=06c127737c6862b8c40323d9ed2f3176e0ab5f70](https://www.adm-ev.de/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/user_upload/PDFS/R05_D.pdf&t=1365964220&hash=06c127737c6862b8c40323d9ed2f3176e0ab5f70) verfügbar [Stand 13.04.2013])
- Arendt, Kathleen & Patrick Rössler (2009): Kinder online befragen: Herausforderungen und erste Erfahrungen am Beispiel einer Kampagnenevaluation. In: Jackob et al. (Hg.): 355-370
- Barker, John/Weller, Susie (2003): ‘Never Work with Children?’: the Geography of Methodological Issues in Research with Children. In: Qualitative Research 3: 207-227
- France, Alan (2004): Young People. In: Fraser et. al. (Hg.): 175-190
- Fraser, Sandy/Lewis, Vicky/Ding, Sharon/Kellett, Mary/Robinson, Chris (Hg.) (2004): Doing Research with Children and Young People. London: Sage
- Fuchs, Marek (2005): Children and Adolescents as Respondents. Experiments on Question Order, Response Order, Scale Effects and the Effect of Numeric Values Associated with Response Options. In: Journal of Official Statistics 21: 701-725
- Heinen, Julia (2012): Internetkinder. Eine Untersuchung der Lebensstile junger Nutzergruppen. Opladen, Berlin & Toronto: Budrich UniPress
- Heinzel, Friederike (Hrsg.) (2000): Methoden der Kindheitsforschung. Ein Überblick über Forschungszugänge zur kindlichen Perspektive. Kindheiten, Band 18. Weinheim: Juventa-Verlag
- Jackob, Nikolaus/Schoen, Harald/Zerback, Thomas (Hg.) (2009): Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung, Wiesbaden: VS Verlag
- König, Susanne (2012): Online-Forschung mit Kindern. Opladen, Berlin & Farmington Hills, MI: Budrich UniPress
- Kränzl-Nagl, Renate/Wilk, Liselotte (2000): Möglichkeiten und Grenzen standardisierter Befragungen unter besonderer Berücksichtigung der Faktoren soziale und personale Wünschbarkeit, in: Heinzel (Hg.): 59-75
- Kuchenbuch, Katharina (2003): Die Fernsehnutzung von Kindern aus verschiedenen Herkunftsmilieus. In: Media Perspektiven 1: 2-11
- Lipski, Jens (2000): Zur Verlässlichkeit der Angaben von Kindern bei standardisierten Befragungen. In: Heinzel (Hg.): 77-86
- Lyberg, Lars/Biemer, Paul P./Collins, Martin/De Leeuw, Edith/Dippo, Cathryn/Schwarz, Norbert/Trewin, Dennis (Hg.) (1997): Survey Measurement and Process Quality. [International Conference on Survey Measurement and Process Quality, held April 1 – 4, 1995 in Bristol, United Kingdom]. New York: John Wiley
- Punch, Samantha (2002): Research with Children. The Same or Different from Research to Adults? In: Childhood 3: 321-341
- Scott, Jacqueline (1997): Children as Respondents: Methods for Improving Data Quality. In: Lyberg et. al. (Hg.): 331-350
- Vogl, Susanne (2014): Befragung von Kindern – ein Praxisleitfaden. Weinheim und München: Juventa

Andreas Motel-Klingebiel, Daniela Klaus und Julia Simonson

---

## 59.1 Alter, Altern und demografischer Wandel

Der demografische Wandel ist eine der großen aktuellen gesellschaftlichen Herausforderungen. Niedrige Fertilität, zunehmende Migration und die Verlängerung der Lebensspanne verändern die Gesellschaft. Die Lebenserwartung hat im 20. Jahrhundert stetig um zwei bis drei Jahre pro Jahrzehnt zugenommen und steigt weiter. Das lange Leben bringt gesellschaftliche und individuelle Entwicklungschancen mit sich, geht aber zugleich mit bisher unbekannten Herausforderungen einher. Somit wächst der gesellschaftspolitische, wirtschaftliche und wissenschaftliche Bedarf nach Informationen über die Lebenssituationen Älterer, über die Pfade in das Alter und die Dynamiken in dieser Lebensphase. Entsprechendes Wissen und angemessene Altersbilder können durch Befragungen der betroffenen Personen generiert werden. Diese Umfragen stellen allerdings besondere Anforderungen an die wissenschaftliche Methodik.

Überlegungen zu Befragungen von älteren und alten Menschen setzen grundsätzlich deren altersbezogene Abgrenzung voraus. Die Orientierung an der Regelaltersgrenze der gesetzlichen Rentenversicherung liegt nahe, doch befindet sie sich im Wandel und bildet die empirische Realität des Ruhestandsübergangs nicht sinnvoll ab. Zugleich stehen häufig Übergangsprozesse im Zentrum des Analyseinteresses. Entsprechend ist eine Altersabgrenzung allenfalls forschungspragmatisch zu begründen. Die Forschung zeigt zudem, dass keineswegs von einer einheitlichen Lebensphase Alter auszugehen ist, so dass die Herausforderungen für Studien mit älteren und alten Menschen sowohl in zeitlicher wie struktureller Perspektive heterogen sind.

---

## 59.2 Datenquellen für die Sekundäranalyse

Das Thema ‚Alter‘ wird seit einigen Jahrzehnten zunehmend im Rahmen großer Studien erforscht: In Deutschland in der Berliner Altersstudie (BASE) und im Deutschen

Alterssurvey (DEAS) sowie aus europavergleichender Perspektive im Survey of Ageing, Health and Retirement in Europe (SHARE). Während BASE die detailliertesten Daten zur Lebenssituation alter Menschen bereithält (regional und auf über 70-Jährige beschränkt), zielt der DEAS auf Basis für Deutschland bevölkerungsrepräsentativer Daten auf umfassende Aussagen zur Lage der ab 40-Jährigen und ihrem Wandel. Dafür verzichtet der DEAS auf die Erschließungstiefe von BASE. SHARE bietet Möglichkeiten des Gesellschaftsvergleiches, beinhaltet jedoch eine weitere Reduktion der Analysebreite und -tiefe. Studien wie die Interdisziplinäre Längsschnittstudie des Erwachsenenalters (ILSE), die nach Bedingungen gesunden und zufriedenen Älterwerdens fragt, die Heidelberger Hundertjährigen-Studie (HD100), die die Ältesten und ihre Lebenssituation in das Zentrum stellt, oder die Studien zu Möglichkeiten und Grenzen selbständiger Lebensführung im Alter (MUG) betrachten thematische oder regionale Ausschnitte. Die genannten Studien ergänzen sich wechselseitig.

Überdies werden zunehmend allgemeine Surveys (Mochmann, Schupp, Kapitel 14 und 73 in diesem Band) für alter(n)swissenschaftliche Analysen nutzbar gemacht, wie beispielsweise das sozio-oekonomische Panel (SOEP), der Bundes-Gesundheitssurvey des Robert-Koch-Instituts (RKI) und das Familien- und Beziehungspanel (pairfam).

Einen Überblick über europäische Datenquellen liefert beispielsweise das Projekt ‚Data Mapping‘ der Joint Programming Initiative (JPI) ‚More Years, Better Lives – The Potential and Challenges of Demographic Change‘ ([www.jpi-dataproject.eu](http://www.jpi-dataproject.eu)). Eine Darstellung weltweiter Längsschnittstudien zum Altern findet sich bei Kaiser (2013).

Die Datenlage in Deutschland ist im internationalen Vergleich zufriedenstellend. Da vorangegangene Lebensverläufe späte Lebensphasen mitbestimmen und häufig auf die Prüfung von Kausalthypothesen (Kühnel/Dingelstedt, Kapitel 80 in diesem Band) abgehoben wird, werden in Studien zu Altern und Alter meist längsschnittliche Ansätze (Pötter/Prein, Kapitel 86 in diesem Band) bevorzugt. Wachsende theoretische wie empirisch-analytische Ansprüche und die damit zunehmende Komplexität lassen allerdings Befragungen mit älteren und alten Menschen rasch an Grenzen stoßen. In der Konsequenz wird der Altersbereich häufig in der Stichprobendefinition (Häder/Häder, Kapitel 18 in diesem Band) begrenzt, die Beteiligung älterer und alter Befragter vernachlässigt oder die Differenzierung von Altersphasen mit Kategorien wie ‚65+‘ nur unzureichend vorgenommen.

---

### **59.3 Stichprobenziehung**

Mit Blick auf alternde und alte Menschen gelten standardisierte Befragungen (Reinecke, Kapitel 44 in diesem Band) oft als Methode der Wahl, die umfangreiches, möglichst bevölkerungsrepräsentatives Datenmaterial verspricht. Doch stellen bereits Stichprobenziehung und Zugang zu den Befragten (Häder/Häder, Kapitel 18 in diesem Band) erste Hürden dar. So beschränken sich bevölkerungsrepräsentative Stichproben häufig auf *Privathaushalte*, statt den Aufwand zu betreiben, Personen in Institutionen zu berücksichtigen und zu

befragen, ggf. unter Zuhilfenahme von Kurzerhebungen und Proxyinterviews mit Angehörigen und Pflegepersonen. Den im Alternsverlauf zunehmenden Anteil an dauerhaft oder temporär in *Institutionen wie Krankenhäusern, Alters- oder Pflegeheimen* Wohnenden zu vernachlässigen, führt regelmäßig zur Unterschätzung sozialer und gesundheitlicher Problemlagen unter Hochbetagten (Kelfve et al. 2013).

---

## 59.4 Feldzugang

Vor allem in Pflegeeinrichtungen fungieren Leitung und Personal oft als *Gatekeeper*, was bewirken kann, dass ein Kontakt mit Zielpersonen nicht oder nur eingeschränkt zustande kommt. Auch bei hilfebedürftigen Älteren in Privathaushalten erfolgt der Erstkontakt nicht selten über Partner und Familienmitglieder, die oft stellvertretend Teilnahmeentscheidungen treffen. Sieht man von diesen spezifischen Zugangsbedingungen ab, so zeigt sich jedoch, dass die *Nichterreichbarkeit* (Engel/Schmidt, Kapitel 23 in diesem Band) bei älteren und alten Befragten eine geringere Rolle spielt als bei Erwerbstätigen.

---

## 59.5 Befragbarkeit und Befragungsmodus

*Gesundheitliche Einschränkungen* begrenzen nicht nur die Zugangsmöglichkeiten zu älteren und alten Befragungspersonen, sondern setzen auch deren Befragbarkeit herab (Kühn/Porst 1999, Knäuper et al. 2002). Probleme können insbesondere der Rückgang des Hörvermögens, eine verlangsamte Informationsverarbeitung und eingeschränkte Gedächtnisfunktionen verursachen (Wagner/Motel 1996).

Zugleich ist jedoch bekannt, dass dieser Abbau vielfältig kompensiert werden kann, etwa durch den Rückgriff auf Lebenserfahrung oder die Beschränkung auf verbleibende Kompetenzen und vertraute Inhalte. Das jedoch kann bereits aus sich heraus zu Verzerrungen von Befragungsergebnissen in einer Weise führen, die die Durchführung von Erhebungen infrage stellt.

Generell können ältere und alte Menschen im Rahmen persönlicher (Stocké, Kapitel 45 in diesem Band), schriftlicher (Reuband, Kapitel 47 in diesem Band) und telefonischer Interviews (Hüfken, Kapitel 46 in diesem Band) befragt werden. Doch sollte im Interview auf eine deutliche Ansprache und die Vermeidung von Nebengeräuschen geachtet werden sowie im schriftlichen Interview auf die Auswahl von Schrifttyp und -größe, Farben und Papier. Vor allem bei Telefonstudien kann es zu Hör- und Verständnisschwierigkeiten kommen. Im persönlichen Gespräch lassen sich diese Schwierigkeiten durch die visuelle Darstellung beispielsweise langer Antwortskalen mildern (Kühn/Porst 1999: 19). Online-Befragungen (Wagner/Hering, Kapitel 48 in diesem Band) sind aufgrund der noch geringen Internetverbreitung und Computeraffinität bei Älteren bislang wenig praktikabel. Persönliche Befragungen sind in jedem Fall anderen Vorgehensweisen vorzuziehen.

## 59.6 Nichtteilnahme und Stichprobenselektivität

Auch wenn einzelne Befunde für die ältere Bevölkerung höhere Teilnahmequoten ausweisen (Carsjö et al. 1994), ist die Ausschöpfungsquote in dieser Altersgruppe typischerweise gering. Grundsätzlich ist die Teilnahme älterer und alter Befragter von ähnlichen Determinanten (Engel/Schmidt, Kapitel 23 in diesem Band) abhängig wie in jüngeren Bevölkerungsgruppen. Doch kumulieren ungünstige Voraussetzungen der Teilnahme, seien sie nun alterskorreliert wie *schlechte physische Gesundheit* und *geringes kognitives Leistungsniveau*, kohortenspezifisch wie das durchschnittlich *geringere Bildungsniveau* oder sicherungssystembedingt wie die *oftmals schlechtere sozio-ökonomische Situation* der heutigen Älteren (Kelfve et al. 2013). Zudem scheinen alte Menschen ihre Entscheidung zur Teilnahme zu einem höheren Grad als Jüngere vom persönlichen Interesse an der Befragung abhängig zu machen (Kühn/Porst 1999: 17f.). Die genannten Faktoren begründen spezifische Selektivitäten.

Bei der Teilnahme an Wiederholungsbefragungen (Schupp, Kapitel 73 in diesem Band) greifen im Wesentlichen dieselben Mechanismen wie bei der Erstbefragung alter Menschen. Entsprechend besteht das Risiko einer beständigen Zunahme der Selektivität (*Panelmortalität*) über nachfolgende Panelwellen. Vor allem der Gesundheitseffekt gewinnt durch die starke Häufung von gesundheitlichen Problemlagen im Alter an Bedeutung. Ein häufiges Vorgehen zur Reduktion selektiver Panelmortalität besteht darin, einmalig nicht teilnehmenden, aber panelbereiten Studienteilnehmern die Rückkehr in die Befragung zu ermöglichen (Banks et al. 2011). Dieses Verfahren ist effektiv, da sich im Alter die Anlässe für temporäre Befragungsunfähigkeit aufgrund von Krankheit oder Krankenhausaufenthalten häufen (Engstler/Motel-Klingebiel 2010).

Ergänzend erweist sich die Durchführung von Proxyinterviews, also zumeist verkürzte Befragungen mit vertrauten Personen, Verwandten oder Pflegepersonal, als geeignet, um für zwischenzeitlich beeinträchtigte und/oder verhinderte Befragte relevante Informationen zu sammeln. Hierdurch kann der Erkenntnisstand über den temporären oder endgültigen Ausfall erweitert werden. Zudem lässt sich auf diese Weise auch die Stichprobenverzerrung abmildern.

Neben der Panelmortalität nimmt mit zunehmendem Alter auch die *echte Mortalität* als Ausfallgrund bei Längsschnittbefragungen an Bedeutung zu. Um die Ausfallprozesse untersuchen zu können, empfehlen sich Recherchen zum Verbleib der Befragten, die dauerhaft aus der Untersuchung ausscheiden. Ihre Nachverfolgung über Einwohnermeldeämter bzw. die Einholung von Informationen über nahe stehende Personen oder Verwandte liefern Informationen über Ursachen der Nichtwiederbefragung und ggf. neue Adressen oder Sterbedaten. So lässt sich der Ausfall in Analysen einbeziehen. Um weitere Beschreibungen der Lebensumstände der Befragten in der Zeit vor ihrem Ausscheiden zu erfahren, werden häufig so genannte Exit-Interviews eingesetzt, die meist mit Hinterbliebenen geführt werden.

## 59.7 Validität und Reliabilität

Die Datenqualität (Krebs/Menold, Kapitel 30 in diesem Band) wird unter anderem von der Qualität der erfassten Antworten bestimmt. Die Validität von Angaben sinkt mit steigendem Alter. So finden sich Hinweise, dass die Antworten älterer und alter Befragter verstärkt an *sozialer Erwünschtheit* (Hlawatsch/Krickl, Kapitel 20 in diesem Band) und Stereotypen ausgerichtet sind (Kühn/Porst 1999: 9). Auch der Antwortkontext spielt wegen verstärkter *Recency-Effekte* (Franzen, Kapitel 51 in diesem Band) eine größere Rolle als bei Jüngeren (Knäuper 1998). Zudem ziehen Ältere *verbale Angaben* numerischen Werten vor und Studien verweisen auf eine *Ja-Sage-Tendenz* sowie häufiger *fehlende Werte* (Engel/Schmidt, Kapitel 23 in diesem Band) durch Antwortverweigerung oder „weiß nicht“-Nennungen. Für altersspezifisches Antwortverhalten werden allgemein ähnliche Ursachen verantwortlich gemacht wie für die Nichtteilnahme.

Daneben ist eine möglicherweise eingeschränkte *Konstruktvalidität* zu bedenken. Konstrukte bzw. ihre Operationalisierung können in unterschiedlichen Altersgruppen Unterschiedliches messen (Kühn/Porst 1999: 30ff.), so dass eine alterssensible Validierung eingesetzter Instrumente erforderlich ist (Borchelt et al. 1996). Über die Validität der Messungen hinaus, kann auch mangelnde *Reliabilität* die Qualität erhobener Informationen negativ beeinflussen (Krebs/Menold, Kapitel 30 in diesem Band). Wagner und Motel (1996) zeigen, dass bei komplexen Abfragen, die hohe kognitive Ansprüche stellen, die Test-Retest-Reliabilität in höheren Altersgruppen geringer ist.

---

## 59.8 Rechtliche, ethische und pragmatische Probleme

Ethische und rechtliche Fragen ergeben sich bei eingeschränkt einwilligungsfähigen Personen, beispielsweise Demenzkranken. Hier müssen die Voraussetzungen für die Befragbarkeit und Einwilligungsfähigkeit der betreffenden Personen im Einzelfall durch den Interviewenden abgewogen werden. Sie sind nicht allein anhand klinischer Diagnosen oder neuropsychologischer Testergebnisse zu bestimmen (Vollmann et al. 2004: 34), denn selbst bei schwerer Demenz kann eine Einwilligungsfähigkeit gegeben sein (Geiselmann et al. 1996). Bislang sind die Vorgehensweisen allerdings unklar geregelt. Verwiesen werden kann auf die Berliner Altersstudie (BASE), bei der die Einwilligungsfähigkeit an drei Kriterien geknüpft wurde: Entscheidungsfähigkeit, Verständnis und rationale Verarbeitung der relevanten Informationen (Geiselmann et al. 1996: 446). Die Einschätzung ist mehrstufig vorzunehmen und im Laufe der Befragung zu überprüfen.

Darüber hinaus sind Kosten und Nutzen einer Teilnahme zu berücksichtigen (Geiselmann et al. 1996). Schließlich kann die Einwilligungsfähigkeit insgesamt je nach Befragungsdesign, -komplexität oder -thema variieren. Sie ist also immer nur bezogen auf ein Forschungsvorhaben zu bestimmen.

## Literatur

- Banks, James/Muriel, Alastair/Smith, James P. (2011): Attrition and Health in Ageing Studies. In: Longitudinal and Life Course Studies 2: 101-126
- Borchelt, Markus/Gilberg, Reiner/Horgas, Ann L./Geiselmann, Bernhard (1996): Zur Bedeutung von Krankheit und Behinderung im Alter. In: Lindenberger, Ulman/Smith, Jacqui/Mayer, Karl Ulrich/Bates, Paul B. (Hg.): Die Berliner Altersstudie. Berlin: Akademie Verlag. 473-498
- Carsjö, Kerstin/Thorslund, Mats/Wärneryd, Bo (1994): The Validity of Survey Data on Utilization of Health and Social Services Among the Very Old. In: The Journals of Gerontology 49: 156-164
- Geiselmann, Bernhard/Helmchen, Hanfried/Nuthmann, Reinhard (1996): Einwilligungsfähigkeit in der Demenzforschung. In Peters, Uwe H./Schifferdecker, Michael/Krahl, Andreas (Hg.): 150 Jahre Psychiatrie. Köln: Martini. 446-449
- Engstler, Heribert/Motel-Klingebiel, Andreas (2010): Datengrundlagen und Methoden des Deutschen Alterssurveys (DEAS). In: Motel-Klingebiel, Andreas/Wurm, Susanne/Tesch-Römer, Clemens (Hg.): Altern im Wandel. Stuttgart: Kohlhammer. 34-60
- Kaiser, Angelika (2013): A Review of Longitudinal Datasets on Ageing. In: Journal of Population Ageing 6: 5-27
- Kelfve, Susanne/Thorslund, Mats/Lennartsson, Carin (2013): Sampling and Non-Response Bias on Health-Outcomes in Surveys of the Oldest Old. In: European Journal of Ageing 10: 237-245
- Knäuper, Bärbel (Hg.) (1998): Age Differences in Question and Response Order Effects. Philadelphia: Psychology Press
- Knäuper, Bärbel/Schwarz, Norbert/Park, Denise C. (2002): Selbstberichte im Alter. In: Motel-Klingebiel, Andreas/Kelle, Udo (Hg.): Perspektiven der empirischen Alter(n)ssoziologie. Opladen: Leske + Budrich. 75-98
- Kühn, Konstanze/Porst, Rolf (1999): Befragung alter und sehr alter Menschen. Mannheim: ZUMA
- Vollmann, Jochen/Kühl, Klaus-Peter/Tilmann, Amely/Hartung, Heinz-Dieter/Helmchen, Hanfried (2004): Einwilligungsfähigkeit und neuropsychologische Einschränkungen bei dementen Patienten. In: Der Nervenarzt 75: 29-35
- Wagner, Michael/Motel, Andreas (1996): Die Qualität der Einkommensmessung bei alten Menschen. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 48: 493-511

Yasemin El-Menouar

---

## 60.1 Die Unterrepräsentation von Migranten in Bevölkerungsumfragen

Über Befragungen der Allgemeinbevölkerung hinaus sind repräsentative Umfragen von Bevölkerungsminderheiten für viele Fragestellungen zentral. So leben in Deutschland viele Arbeitsmigranten, insbesondere aus der Türkei, Italien und anderen südeuropäischen Ländern, mittlerweile auch in der zweiten und dritten Generation – und durch Familiennachzug oder Bildungsmigration immigrieren auch stetig Neuzuwanderer aus diesen Ländern. Nach dem Fall der Mauer 1989 und mit Aufhebung der EU-Binnengrenzen kamen zunehmend Migranten aus anderen Ländern nach Deutschland, zunächst hauptsächlich aus den ehemaligen GUS-Staaten und zunehmend auch aus den Staaten des Nahen und Mittleren Ostens.

Kulturelle Vielfalt und Pluralismus führen zu Fragen, die sich mit der Integration dieser Gruppen befassen: Gelingt es allen Gruppen gleichermaßen an der Gesellschaft teilzuhaben? Aus welchen Gründen sind bspw. Türkeistämmige stärker benachteiligt als andere Migranten? Wie kann die gesellschaftliche Integration von Migranten verbessert werden? Um Antworten auf diese und ähnliche Fragestellungen zu finden, sind standardisierte Befragungen (Reinecke, Kapitel 44 in diesem Band) unumgänglich. Allgemeine Bevölkerungsbefragungen wie etwa der ALLBUS und Mikrozensus (Mochmann, Hartmann/Lengerer, Kapitel 14 und 71 in diesem Band) sind jedoch aus zwei Gründen dafür eher ungeeignet.

1. *Migranten stellen nur einen geringen Anteil der Gesamtbevölkerung, so dass sie im Rahmen von repräsentativen Bevölkerungsbefragungen nicht in ausreichender Anzahl erfasst werden.* Auf Grundlage von Berechnungen des statistischen Bundesamtes beträgt der Anteil an Personen mit Migrationshintergrund rund 19% (Statistisches Bundesamt 2011a, b). Wird diese Population zusätzlich nach Herkunftsland differenziert, fragmentiert sie sich zum Teil in sehr kleine Gruppen (Tab. 60.1).

**Tab. 60.1** Die zehn größten Bevölkerungsgruppen mit Migrationshintergrund (MH) nach Herkunftsland, in Tausend (Quelle: Migrationsbericht 2011, 158f.; eigene Darstellung)

Herkunftsland	absolut	% Bevölk. mit MH	% Gesamtbevölk.
Türkei	2.956	18,5	3,6
Polen	1.473	9,2	1,8
Russische Föderation	1.227	7,7	1,5
Kasachstan	922	5,8	1,1
Italien	780	4,9	0,9
Rumänien	484	3,0	0,6
Griechenland	392	2,5	0,5
Kroatien	360	2,3	0,4
Serbien	286	1,8	0,3
Ukraine	273	1,7	0,3
Gesamt	9.153	57,4	11,0

Berücksichtigt sind Personen mit derzeitiger oder früherer ausländischer Staatsangehörigkeit bzw. mit mindestens einem Elternteil mit derzeitiger oder früherer ausländischer Staatsangehörigkeit.

So sind die Türkeistämmigen die größte Zuwanderergruppe in Deutschland; ihr Anteil an der Gesamtbevölkerung beträgt 3,6% (Tab. 60.1). In einer Zufallsstichprobe von 1.000 Befragten werden somit lediglich rund 30 bis 40 Personen mit türkischem Hintergrund gezogen, und diese sind zudem auf die drei Generationen verteilt. Für aussagekräftige statistische Analysen ist diese Zahl zu gering.

- Ist geplant, bei einer sozialwissenschaftlichen Befragung (auch) Migranten zu interviewen, so muss das *unterschiedliche Niveau der Deutschkenntnisse* genauso beachtet werden wie mögliche *kulturelle Barrieren*, die Migranten davon abhalten an solchen freiwilligen Umfragen teilzunehmen. Allgemeine Bevölkerungsumfragen richten sich in der Regel an die deutschsprachige Bevölkerung, was auch innerhalb der Gruppe der Migranten zu systematischen Ausfällen (Engel/Schmidt, Kapitel 23 in diesem Band) relevanter Teilpopulationen führen kann (Wasmer et al. 2007: 68; Babka von Gostomski/Pupeter 2008: 170), z.B. bei Frauen mit einem geringen Bildungsabschluss (siehe Abschnitt 60.4). Um solche Ausfälle zu minimieren, können Fragebögen in *verschiedene Sprachversionen* übersetzt und *bilinguale Interviewer* in die Studie eingebunden werden. Aufgrund des geringen Anteils von Menschen mit Migrationshintergrund an der Gesamtbevölkerung wird dieser Aufwand aber im Rahmen von allgemeinen Bevölkerungsbefragungen nur selten betrieben, und wenn, dann wird oft nur die türkische Wohnbevölkerung einbezogen. In einigen Ländern werden weitere Sprachversionen eingesetzt, wenn die Größe einer Teilpopulation mit Migrationshintergrund den Anteil von 5% überschreitet (Méndez/Font 2013: 282). Aus ökonomischer Perspektive sind die Kosten für die Übersetzung des Fragebogens bei kleineren Gruppen meist zu hoch.

In der Regel werden Daten von Migranten nicht im Rahmen von Bevölkerungsbefragungen, sondern durch gezielte *Migrantenbefragungen* erhoben. Diese erfordern je nach Fragestellung und Zielpopulation ein maßgeschneidertes Studiendesign. Meist muss ein Kompromiss zwischen den Zielen der Studie und der Machbarkeit bzw. den Kosten gefunden werden. So werden bei Migrantenbefragungen in der Regel die Gruppen ausgewählt, die in der Gesellschaft relativ stark vertreten sind, z.B. türkische Migranten in Deutschland.

---

## 60.2 Definition der Grundgesamtheit

Zunächst stellt sich die Frage, *wer überhaupt als Migrant gezählt wird*. Je nachdem, welche Definitionskriterien angelegt werden, kann sich der Bestand an Migranten stark unterscheiden. So beinhalten die so genannten Ausländerstatistiken (bspw. das Ausländerzentralregister) lediglich Ausländer im rechtlichen Sinne, also Personen, die ausschließlich eine ausländische *Staatsangehörigkeit* besitzen. Eingebürgerte Migranten mit einer zusätzlichen oder ausschließlichen deutschen Staatsangehörigkeit sind in diesen Statistiken nicht enthalten.

Wird der *Geburtsort* als Kriterium verwendet, so beschränken sich die Daten auf Migranten, die Zuwanderer im eigentlichen Sinne sind (*Foreign-Born-Prinzip*). Statistiken auf der Basis des Geburtsortes umfassen zudem zugewanderte Spätaussiedler, welche die deutsche Staatsbürgerschaft haben, aber eigentlich zu den Migranten zu zählen wären.

Für viele Fragestellungen ist es jedoch relevant, den Migrationshintergrund wesentlich genauer zu erfassen, so dass auch eingebürgerte Nachkommen von Migranten identifizierbar sind (bspw. bei Themen zum Integrationsverlauf über Generationen bzw. zur Bestimmung der gesellschaftlichen Teilhabe). Seit 2005 erlaubt das Mikrozensusgesetz (MZG) die Erhebung wesentlich umfangreicherer Informationen über den Migrationshintergrund (§ 4 Abs. 1 Nr. 2a MZG 2005; § 4 Abs. 2 Nr. 4 MZG 2005). So werden neben der aktuellen auch die *frühere Staatsangehörigkeit* sowie der Geburtsort erfasst. Zusätzlich werden alle vier Jahre der Geburtsort sowie die aktuelle und *frühere Staatsangehörigkeit der Eltern* erhoben.

Nach der aktuellen in Deutschland gültigen Definition sind *Personen mit einem Migrationshintergrund* „alle nach 1949 auf das heutige Gebiet der Bundesrepublik Deutschland Zugewanderten, sowie alle in Deutschland geborenen Ausländer und alle in Deutschland als Deutsche Geborenen mit zumindest einem zugewanderten oder als Ausländer in Deutschland geborenen Elternteil“ (Statistisches Bundesamt 2011a: 6). Die Zahlen im Migrationsbericht 2011 (vgl. auch Tab. 60.1) basieren auf dieser Definition. Im Zensus 2011 hingegen wurden für die Berechnungen lediglich nach 1955 zugewanderte Personen und ihre Nachkommen berücksichtigt, so dass die Zahlen des Zensus vom Mikrozensus abweichen (Statistisches Bundesamt 2013: 26).

Insgesamt kann die Anzahl der Personen mit Migrationshintergrund mittlerweile bis zur dritten Generation relativ gut geschätzt werden. Im Rahmen von Umfragen werden teilweise auch der *Geburtsort der Großeltern* und die *zuhause gesprochenen Sprachen* berücksichtigt, um den Migrationshintergrund möglichst genau bestimmen zu können.

## 60.3 Stichprobenziehung

Einen guten Überblick über Stichprobenziehungsverfahren (Häder/Häder, Kapitel 18 in diesem Band) bei Migranten bieten Salentin (1999) und, stärker praxisorientiert, der Sammelband von Font/Méndez (2013). Zusammenfassend lässt sich festhalten, dass die Möglichkeiten der Stichprobenziehung davon abhängen, welche Auswahlgrundlagen zur Verfügung stehen. Je nach Land sind Informationen über den Migrationsstatus unterschiedlich erfasst. Relativ gute Auswahlbasen stehen bspw. in den Niederlanden, in Schweden und Dänemark zur Verfügung. Dort enthält die amtliche Statistik über die Nationalität und den Geburtsort hinaus auch Informationen über den Geburtsort der Eltern. In England hingegen erfassen amtliche Statistiken den Migrationshintergrund nur sehr unzureichend (Méndez/Font 2013; Erens 2013). In der Regel existieren keine vollständigen Listen von Personen mit Zuwanderungsgeschichte, so dass die Ziehung einer einfachen Zufallsstichprobe nur selten möglich ist. In Deutschland werden im Wesentlichen zwei Methoden für die Ziehung von Zufallsstichproben von Migranten eingesetzt: merkmalsbasierte Ziehungsverfahren auf Basis der Register der Meldeämter und namensbasierte (onomastische) Ziehungsverfahren.

### 60.3.1 Merkmalsbasierte Ziehungsverfahren auf Basis der Register der Meldeämter

In Deutschland werden Merkmale wie die Nationalität sowie der Geburtsort jeder Person statistisch erfasst. Somit können über die *Einwohnermeldeämter* (Häder und Häder, Kapitel 18 in diesem Band) Stichproben in Abhängigkeit der *Nationalität* (Granato 1999) oder des *Geburtsortes* (Salentin 2007) gezogen werden. Die Verwendung des Geburtsortes für die Stichprobenziehung ist bspw. bei Aussiedlerbefragungen sinnvoll, da diese nicht über ihre Nationalität identifiziert werden können. Zu berücksichtigen ist jedoch, dass Einwohnermelderegister lediglich regionale Daten enthalten. Für eine bundesweite Stichprobe müssten entsprechend sehr viele Kommunen einbezogen werden. Zudem können die Kommunen die Herausgabe von Adressdaten aus Gründen des Datenschutzes verweigern, wenn die Anonymität der entsprechenden Personen nicht in ausreichendem Maße gewährleistet wird (Haug et al. 2009: 41). Dies kann bei einer relativ geringen Größe der Zielpopulation in Kombination mit einer starken regionalen Streuung dieser Gruppe auftreten.

Im Unterschied zu den Registern der Meldebehörden enthält das *Ausländerzentralregister* (<http://www.bva.bund.de/DE/Themen/Sicherheit/Auslaenderzentralregister/auslaenderzentralregister-node.html>), das vom Bundesamt für Migration und Flüchtlinge (BAMF) geführt wird, Daten für das gesamte Bundesgebiet. Darin sind *alle Zuwanderer* enthalten, *die am Stichtag nicht die deutsche Nationalität besitzen und länger als drei Monate in Deutschland leben*. Das BAMF nutzt dieses Register für die Ziehung von Stichproben für

eigene Studien. Eine Nutzung des AZR für die Durchführung wissenschaftlicher Studien ist gemäß §24a des Gesetzes über das Ausländerzentralregister aufgrund der Gesetzeslage nur für Forschungszwecke des BAMF möglich.

Stichprobenziehungen sowohl über die Meldeämter als auch auf Basis des AZR sind jedoch mit *Einschränkungen* verbunden. So können z.B. keine Stichproben von Migranten gezogen werden, die in Deutschland geboren sind und bereits die deutsche Staatsbürgerschaft erworben haben (Babka von Gostomski/Pupeter 2008). Somit eignen sich diese Vorgehensweisen nicht für Fragestellungen, die sich auf die zweite und dritte Generation von Migranten beziehen. Auch zur Untersuchung von Themen der Integration sind sie damit nur bedingt geeignet. Da in Deutschland geborene und eingebürgerte Personen mit Migrationshintergrund im Vergleich besser integriert sind als Migranten der ersten Generation, sind die Ergebnisse von Integrationsstudien mit meldeamtbasierter Stichprobenziehung oft verzerrt. Neuerdings erfassen die Meldeämter jedoch ergänzende Informationen, die eine wesentlich genauere Ermittlung des Migrationshintergrundes erlauben.

### 60.3.2 Namensbasierte (onomastische) Ziehungsverfahren

Zunehmend werden auch namensbasierte Ziehungsverfahren für Migrantenstichproben eingesetzt (Schnell et al. 2013; Salentin 2007; Humpert/Schneiderheinze 2000). Dabei werden typische Namen aus einem bestimmten Herkunftsland bzw. eine typische Buchstabenabfolge oder Silben, die in Namen enthalten sind, genutzt, um Register bzw. Verzeichnisse (bspw. das Telefonverzeichnis) nach diesen *Namen(skombinationen)* oder *Silben* abzusuchen. Bspw. kommt die Silbe „öz“ sehr häufig in türkischen Namen vor (Öztürk, Öztoprak etc.). Bei der Verwendung von vollständigen Namen kann es vorkommen, dass aufgrund von verschiedenen Schreibweisen oder Schreibfehlern im Telefonverzeichnis nicht alle Nummern bzw. Namen identifiziert werden können (Granato 1999: 56). Darüber hinaus werden nur typische Namen für das Screening verwendet, so dass Personen mit einem untypischen Namen keine Chance haben, in die Stichprobe zu gelangen. Ein neues Verfahren, das anhand der so genannten n-Gramme des Namens und der Anwendung von statistischen Verfahren Namen einem Herkunftsland zuordnet, ist fehlertoleranter und kann auch untypische Namen mit einbeziehen (Schnell et al. 2013).

Das Zentrum für *Türkeistudien und Integrationsforschung* (ZfTI) (<http://www.zft-online.de/>) besitzt seit 1999 einen angeblich repräsentativen *Telefonnummern-Pool* für Personen mit türkischem Migrationshintergrund, aus dem regelmäßig Stichproben für Telefonumfragen zu verschiedenen Themen gezogen werden (Sauer/Halm 2009: 132ff.). Der Vorteil dieses Verfahrens ist, dass auch Personen der zweiten und dritten Generation, die bereits eingebürgert sind, in die Stichprobe gelangen können. Mit zunehmender interethnischer Heirat wird dies für nachfolgende Generationen jedoch schwieriger werden, weil Personen anhand ihres Namens nicht mehr eindeutig einem bestimmten Herkunftsland zugeordnet

werden können. Zudem besteht das Problem, dass nicht alle Telefonanschlüsse im Telefonverzeichnis gelistet sind. Auch wenn die Telefondichte mit über 90% der Haushalte mit Festnetzanschluss relativ hoch ist, sind lediglich zwischen 50 und 70% aller Anschlüsse im öffentlichen Telefonverzeichnis eingetragen (Schneiderat/Schlinzig 2009: 100). Unter Migranten sind Festnetzanschlüsse zudem weniger üblich als unter Deutschen, und sie sind auch noch seltener im Telefonbuch verzeichnet (Lipps/Kissau 2012).

---

## 60.4 Interviewsprache

Die Deutschkenntnisse von Migranten variieren stark, so dass es zu systematischen Verzerrungen (Engel/Schmidt, Kapitel 23 in diesem Band) kommt, wenn auf *Fragebogenübersetzungen* verzichtet wird (Blohm/Diehl 2001: 228; Salentin 1999: 125; Smith 2013). Personen, die den Fragebogen in der Sprache des Aufnahmelandes bearbeitet haben, unterscheiden sich systematisch von Personen, die eine Version in der Sprache des Herkunftslandes genutzt haben. Unterschiede sind vor allem bei demografischen Merkmalen und Dimensionen der Integration festzustellen (Smith 2013). So hat sich gezeigt, dass ohne eine entsprechende Fragebogenübersetzung überdurchschnittlich häufig Migranten der ersten Generation, Frauen sowie Personen mit einem niedrigen Bildungsniveau aufgrund von Sprachproblemen aus der Befragung herausfallen würden (Becher/El-Menouar 2013). Studien zufolge ist bspw. bei Personen türkischer Herkunft mit Ausfällen zwischen einem Viertel und einem Drittel zu rechnen (bspw. Babka von Gostomski/Pupeter 2008: 170). Entsprechend ist es wichtig, verschiedene Sprachversionen anzubieten und den Befragten die Wahl zu geben, in welcher Sprache sie letztlich am Interview teilnehmen möchten.

Die Schwierigkeit bei Fragebogenübersetzungen besteht darin, die *Indikatoren* und *Konzepte* (Burzan, Kapitel 81 in diesem Band) so zu übersetzen, dass sie nicht nur formal korrekt, sondern auch dem Sprachgebrauch der Zielpopulation angepasst sind (Braun, Kapitel 56 in diesem Band). Es gibt im Wesentlichen zwei Verfahren, die bei mehrsprachigen Studien genutzt werden:

- Bei der *back-translation*-Methode wird der übersetzte Fragebogen wieder in die Ausgangssprache zurückübersetzt und mit dem Originalfragebogen verglichen.
- Bei der *committee-* oder *team*-Methode übersetzen mehrere Übersetzer den Fragebogen unabhängig voneinander und vergleichen anschließend ihre Versionen. Gibt es Diskrepanzen in den Übersetzungen, einigen sich die Übersetzer auf die der jeweiligen Originalfrage am besten entsprechende Version (Schoua-Glusberg 2006). Meist arbeiten die Übersetzer mit den Umfrageexperten zusammen an der Übersetzung des Fragebogens (Harkness 2003). Mit den verschiedenen Sprachversionen entstehen jedoch neue Probleme in Bezug auf die Reliabilität (Krebs/Menold, Kapitel 30 in diesem Band) der Daten, d.h. inwieweit die eingesetzten Indikatoren in den unterschiedlichen Sprachen tatsächlich das Gleiche messen (Harkness et al. 2010).

## 60.5 Interviewerauswahl

Im Rahmen von Migrantenbefragungen sind das *Geschlecht* und die *ethnische Herkunft* des Interviewers wichtige Faktoren, um überhaupt Zugang zur Zielpopulation zu erhalten. Generell ist es wichtig auf kulturelle Besonderheiten bei bestimmten Gruppen zu achten (Erens 2013). Bei Face-to-Face-Befragungen (Stocké, Kapitel 45 in diesem Band) unter Muslimen ist es bspw. unangemessen, wenn eine Frau ohne Anwesenheit eines männlichen Familienmitglieds von einem männlichen Interviewer befragt wird; in diesem Fall wird die Zielperson die Teilnahme höchstwahrscheinlich verweigern. Daher sind weibliche Interviewer hier die bessere Wahl.

Darüber hinaus hat sich bei bestimmten Populationen und Fragestellungen ein *ethnisches Matching* zwischen Interviewten und Interviewern als sinnvoll erwiesen, weil das soziale Vertrauen in Personen der gleichen ethnischen Herkunft häufig größer ist (Erens 2013). Bei bestimmten Themen kann ein ethnisches Matching aber auch unzweckmäßig sein. Bspw. kann ein für Deutschland unproblematisches Thema in anderen Kulturreihen durchaus ein sensibles sein (bspw. Alkoholgenuss bei Muslimen). Je nach Thema und Population sollte entsprechend abgewogen werden, inwieweit ein ethnisches Matching sinnvoll ist (Méndez/Font 2013).

Die Umfrageinstitute (Ziegler, Kapitel 10 in diesem Band) in Deutschland verfügen jedoch nur in Ausnahmefällen über einen entsprechenden Interviewerstab (Glantz/Michael, Kapitel 21 in diesem Band). Muttersprachliche Interviewer werden meist je nach Bedarf durch Zeitungsannoncen oder über Aushänge an Universitäten rekrutiert und nur kurzfristig eingestellt. Erschwerend kommt hinzu, dass meist kein Fachpersonal zur Verfügung steht, das die muttersprachlichen Interviewer adäquat schulen könnte. Zudem ist die Interviewerkontrolle unter diesen Umständen kaum möglich. Generell ist zu empfehlen, bei Migrantenbefragungen in einem Team mit Experten zusammen zu arbeiten, die sich mit der entsprechenden Zielpopulation sowie ihren sprachlichen und kulturellen Besonderheiten auskennen.

## 60.6 Ausgewählte Beispielstudien aus Deutschland

Mittlerweile gibt es in Deutschland zahlreiche Datensätze von Personen mit Migrationshintergrund, die auch für Sekundäranalysen zugänglich sind. Das *sozioökonomische Panel (SOEP)* (Schupp, Kapitel 73 in diesem Band) ist eine Haushaltsbefragung, die seit der ersten Welle im Jahr 1984 auch Migranten befragt. Auch wenn die Fallzahlen begrenzt sind, bietet das SOEP eine einzigartige Grundlage, um Integrationsprozesse im Zeitverlauf zu analysieren. Bisher sind jedoch Migrantengruppen, die in großer Zahl im Rahmen der Gastarbeiteranwerbung nach Deutschland gekommen sind, überrepräsentiert (Haushaltsvorstand ist türkischer, italienischer, spanischer, griechischer oder ehemals jugoslawischer Nationalität). Es ist jedoch geplant, die Migrantenstichprobe des SOEP um neue

Gruppen zu erweitern sowie neue Indikatoren einzubinden, die ausführlichere Analysen zu Integrationsprozessen erlauben (Brücker 2013).

Mit der 2009 durchgeföhrten Studie „*Muslimisches Leben in Deutschland*“ des Bundesamtes für Migration und Forschung (BAMF) ([http://www.bmi.bund.de/SharedDocs/Downloads/DE/Themen/Politik\\_Gesellschaft/DIK/vollversion\\_studie\\_muslim\\_leben\\_deutschland\\_.pdf?\\_\\_blob=publicationFile](http://www.bmi.bund.de/SharedDocs/Downloads/DE/Themen/Politik_Gesellschaft/DIK/vollversion_studie_muslim_leben_deutschland_.pdf?__blob=publicationFile)) gibt es erstmals eine repräsentative Befragung von Muslimen aus verschiedenen Herkunftsländern (N=6.004). Einbezogen wurden insgesamt 49 überwiegend muslimisch geprägte Herkunftsländer. Mithilfe des AZR wurden für jedes Herkunftsland Namenslisten erstellt und das öffentliche Telefonverzeichnis danach abgesucht. Im Rahmen einer telefonischen Befragung wurden der tatsächliche Migrationshintergrund sowie die Konfession aller Personen eines Haushaltes erfasst, so dass auf dieser Basis erstmals eine Schätzung der Größe und Struktur der muslimischen Population in Deutschland möglich geworden ist (Haug et al. 2009: 36ff.). Der Fragebogen wurde in acht Sprachen übersetzt, bilinguale Interviewer kamen zum Einsatz (Haug et al. 2009: S. 46ff.).

Ein sehr aufwendiges Design hat das BAMF auch für ihre im Jahr 2011 durchgeföhrte BAMF-*Einbürgerungsstudie* ([https://www.bamf.de/SharedDocs/Anlagen/DE/Publikationen/Forschungsberichte/fb15-einbuergerungsverhalten.pdf?\\_\\_blob=publicationFile](https://www.bamf.de/SharedDocs/Anlagen/DE/Publikationen/Forschungsberichte/fb15-einbuergerungsverhalten.pdf?__blob=publicationFile)) eingesetzt. Um Eingebürgerte, nicht Eingebürgerte, Optionspflichtige und sich im Einbürgerungsverfahren befindende Zuwanderer zu befragen, wurde mehrstufig vorgegangen. Während die ersten beiden Gruppen mittels repräsentativer Telefonbefragungen (Hüfken, Kapitel 46 in diesem Band) auf Basis einer onomastischen Ziehung und einem anschließenden Telefonscreening befragt wurden, erfolgte die Befragung der letzten beiden Gruppen durch computergestützte persönliche Interviews (CAPI). Die Stichprobenziehung der optionspflichtigen Personen erfolgte über die Staatsangehörigkeitsbehörden, und die der Einbürgerungsverfahren erfolgte über die Einwohnermeldeämter (Weinmann et al. 2012: 51ff.). Beispielstudien aus anderen Ländern mit praktischen Hinweisen und Empfehlungen werden im Sammelband von Font/Méndez (2013) ausführlich dargestellt.

---

## 60.7 Zusammenfassung

Eine Goldene Regel gibt es bei Migrantbefragungen nicht. Je nach Fragestellung und Zielpopulation müssen zum Teil individuelle methodische Wege gefunden werden. Meist basieren Studien auf einem Kompromiss zwischen den Zielen und den praktischen Möglichkeiten bzw. den Kosten. Fragen, die bei der Planung einer Migrantbefragung berücksichtigt werden sollten, sind die Folgenden:

1. *Stichprobenziehung*: Wie ist der interessierende Migrationshintergrund in für Forschungszwecke zugänglichen Listen erfasst? Wie groß ist die Population und wie stark ist sie gestreut?

2. *Übersetzung des Fragebogens:* Wie groß ist der Anteil, der die deutsche Sprache nicht beherrscht? Sind Ausfälle aufgrund von Sprachproblemen zu erwarten? Welche Sprachen sind sinnvoll? Bei einer sprachlich sehr diversen Zielpopulation muss der Fragebogen nicht immer in alle Heimatsprachen übersetzt werden. Bspw. sprechen Personen aus Afrika, wo sehr viele verschiedene Sprachen und Dialekte gesprochen werden, entweder Französisch oder Englisch.
3. *Interviewerauswahl:* Sind kulturelle Besonderheiten zu berücksichtigen (bspw. Sozialnormen)?

*Best-Practice-Beispiele* oder *Hinweise auf Probleme* sind in teilweise sehr ausführlichen Methodenbeschreibungen der bereits durchgeführten Studien zu Migranten enthalten.

## Literatur

- Babka von Gostomski, Christian/Pupeter, Monika (2008): Zufallsbefragung von Ausländern auf Basis des Ausländerzentralregisters. Erfahrungen bei der Repräsentativbefragung „Ausgewählte Migrantengruppen in Deutschland 2006/2007“ (RAM). In: Methoden – Daten – Analysen 2: 149-177
- Becher, Inna/El-Menouar, Yasemin (2013): Lost without translation? Respondents' reasons for using translated questionnaires and their relevance for data quality ESRA The fifth Conference of the European Survey Research Association, Ljubljana, Juli 2013
- Blohm, Michael/Claudia Diehl (2001): Wenn Migranten Migranten befragen. Zum Teilnahmeverhalten von Einwanderern bei Bevölkerungsbefragungen. In: Zeitschrift für Soziologie 30: 223-242
- Brücker, Herbert (2013): Stand und Perspektiven der Migrationsforschung. In: Schimany/ von Loeffelholz (Hg.): 64-106
- Bundesregierung (2013): Migrationsbericht des Bundesamtes für Migration und Flüchtlinge im Auftrag der Bundesregierung. Migrationsbericht 2011. Nürnberg: Bundesamt für Migration und Flüchtlinge
- Erens, Bob (2013): Designing high-quality surveys of ethnic minority groups in the United Kingdom. In: Font /Méndez (Hg.): 45-67
- Font, Joan/Méndez, Mónica (Hg.) (2013): Surveying Ethnic Minorities and Immigrant Populations: Methodological Challenges and Research Strategies. Amsterdam: Amsterdam University Press
- Granato, Nadia (1999): Die Befragung von Arbeitsmigranten: Einwohnermeldeamt-Stichprobe und telefonische Erhebung? In: ZUMA-Nachrichten 45: 44-60
- Häder, Michael/Häder, Sabine/Kühne, Mike (Hg.) (2012): Telephone Surveys in Europe: Research and Practice. Heidelberg: Springer
- Häder, Michael/Häder, Sabine (Hg.) (2009): Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung. Wiesbaden: VS Verlag
- Harkness, Janet A./van de Vijver, Fons J. R./Mohler, Peter Ph. (Hg.) (2003): Cross-Cultural Survey Methods. Hoboken: Wiley
- Harkness, Janet A./Edwards, Brad/Braun, Michael/Johnson, Timothy P./Lyberg, Lars E./Mohler, Peter/Pennell, Beth-Ellen/Smith, Tom (Hg.) (2010): Survey methods in multi-cultural, multinational and multiregional contexts. New York: Wiley
- Harkness, Janet (2003): Quastionnaire Translation. In: Harkness et al. (Hg.): 35-56
- Haug, Sonja/Müssig, Stephanie/Stichs, Anja (2009): Muslimisches Leben in Deutschland. Nürnberg: Bundesamt für Migration und Flüchtlinge
- Humpert, Andreas/Schneiderheinze, Klaus (2000): Stichprobenziehung für telefonische Zuwandererumfragen. Einsatzmöglichkeiten der Namensforschung. In: ZUMA-Nachrichten 47: 36-64

- Lipps, Oliver/Kissau, Kathrin (2012): Nonresponse in an Individual Register Sample Telephone Survey in Lucerne/Switzerland. In: Häder et al. (Hg.): 187-208
- Méndez, Mónica/Font, Joan (2013): Surveying immigrant populations: Methodological strategies, good practices and open questions. In: Font, Joan/Méndez, Mónica (Hg.): 271-290
- Salentin, Kurt (2007): Die Aussiedler-Stichprobenziehung. In: Methoden – Daten – Analysen 1: 25-44
- Salentin, Kurt (1999): Die Stichprobenziehung bei Zuwandererbefragungen. In: ZUMA-Nachrichten, 45: 115-135
- Sauer, Martina/Halm, Dirk (2009): Erfolge und Defizite der Integration türkeistämmiger Einwanderer- Entwicklung der Lebenssituation 1999 bis 2008. Wiesbaden: VS Verlag
- Schimany, Peter/von Loeffelholz, Hans Dietrich (Hg.) (2013): Beiträge zur Migrations- und Integrationsforschung. Aus Anlass des 60-jährigen Bestehens des Bundesamtes für Migration und Flüchtlinge. Nürnberg: Bundesamt für Migration und Flüchtlinge
- Schneiderat, Götz/Schlünzig, Tino (2009): Teilnahmebereitschaft an Mobilfunkbefragungen, Design der Studie und Befunde. In: Häder, Michael/Häder, Sabine (Hg.): 83-164
- Schnell, Rainer/Gramlich, Tobias/Bachteler, Tobias/Reiher, Jörg/Trappmann, Mark/Smid, Menno/Becher, Inna (2013): Ein neues Verfahren für namensbasierte Zufallsstichproben von Migranten. In: Methoden – Daten – Analysen 7: 5-33
- Schoua-Glusberg, Alicia (2006): Translating research instruments: Committee approach focus groups. Evanston, IL: Research Support Services Report
- Smith, Tom (2013): An evaluation of Spanish questions on the 2006 and 2008 US General Social Surveys. In: Font /Méndez (Hg.): 219-240
- Statistisches Bundesamt, 2011a: Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund. Ergebnisse des Mikrozensus 2010. Wiesbaden: Statistisches Bundesamt
- Statistisches Bundesamt, 2011b: Bevölkerung und Erwerbstätigkeit. Ausländische Bevölkerung. Ergebnisse des Ausländerzentralregisters. Wiesbaden: Statistisches Bundesamt
- Statistisches Bundesamt 2013: Zensus 2011: Ausgewählte Ergebnisse. Wiesbaden: Statistisches Bundesamt
- Wasmer, Martina/Scholz, Evi/Blohm, Michael (2007): Konzeption und Durchführung der ‚Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften‘ (ALLBUS) 2006. ZUMA Methodenbericht (09/2007), Mannheim: ZUMA
- Weinmann, Martin/Becher, Inna/Babka von Gostomski, Christian (2012): Einbürgerungsverhalten von Ausländerinnen und Ausländern in Deutschland sowie Erkenntnisse zu Optionspflichtigen. Ergebnisse der BAMF-Einbürgerungsstudie 2011. Nürnberg: Bundesamt für Migration und Flüchtlinge

Miriam Schütte und Tobias Schmies

Häufig möchte man in der empirischen Sozialwissenschaft Personen aus speziellen oder seltenen Bevölkerungsgruppen befragen. Speziell ist eine Subgruppe aufgrund einer gemeinsamen, teilweise auch seltenen Eigenschaft, gemeinsamer Interessen, einer Situation, in der sich alle Mitglieder dieser Gruppe befinden oder aufgrund ihres ähnlichen sozioökonomischen Status. Das können alleinerziehende Väter genauso sein wie Homosexuelle oder Drogenabhängige. Dabei besteht das Problem, dass es häufig nur wenige, unge nau e oder gar keine Informationen über die Mitglieder dieser Gruppe gibt. Die Herausforderung besteht bei einer auf solch eine Population gerichteten Befragung also zunächst darin, diese Population einzugrenzen und deren Mitglieder zu finden. Im Folgenden nicht berücksichtigt sind die Besonderheiten von Mitarbeiterbefragungen (Borg, Kapitel 57 in diesem Band), Betriebsbefragungen (Liebig/Meyermann/Gebel, Kapitel 67 in diesem Band), Befragungen von Minoritäten (El-Menouar, Kapitel 60 in diesem Band), Kindern und Jugendlichen (Heinen/König, Kapitel 58 in diesem Band) sowie Alten (Motel-Klingenbiel/Klaus/Simonson, Kapitel 59 in diesem Band), weil diesen jeweils eigene Kapitel in diesem Handbuch gewidmet sind.

---

## 61.1 Spezielle Gruppen in allgemeinen Bevölkerungsumfragen

Auch wenn allgemeine Bevölkerungsumfragen zum Ziel haben, die gesamte Bevölkerung abzubilden, sind spezielle Teilgruppen dennoch nicht angemessen repräsentiert. Ursache ist zum einen, dass sie, etwa aus forschungspragmatischen Gründen, nicht zur Grundgesamtheit gezählt werden und aus dem Untersuchungsdesign herausfallen. Zum anderen unterscheiden sich Teilgruppen auch hinsichtlich ihres Teilnahmeverhaltens, ihrer Erreichbarkeit und ihrer Fähigkeit, an einer Befragung teilzunehmen (Engel/Schmidt, Kapitel 23 in diesem Band) bzw. stellen für Auswertungen eine zu kleine Stichprobe dar.

### 61.1.1 Nichtzugehörigkeit zur Grundgesamtheit

Bestimmte Bevölkerungsgruppen wie in Deutschland lebende Ausländer, Deutsche, die sich im Ausland aufhalten oder in Anstalten leben, sind regelmäßig sowohl bei nicht-amtlichen wie teilweise auch bei amtlichen *allgemeinen* Bevölkerungsumfragen durch das Untersuchungsdesign nicht abgedeckt und so faktisch aus der Grundgesamtheit ausgeschlossen. Zur Anstaltsbevölkerung bzw. den Sonderbereichen zählt etwa der Zensus Wohnheime oder Gemeinschaftsunterkünfte, wobei letztere noch einmal in sensible und nicht-sensible Sonderbereiche unterteilt wurden. Zu den nicht-sensiblen Sonderbereichen zählen z.B. Studentenwohnheime, Seniorenheime, Kloster und Internate. Beispiele für sensible Sonderbereiche sind Behindertenwohnheime, Notunterkünfte für Obdachlose oder Justizvollzugsanstalten (Statistische Ämter des Bundes und der Länder 2011: 16-17). Auch Angehörige ausländischer Streitkräfte und ausländischer diplomatischer Vertretungen mit ihren Familienangehörigen werden oft nicht erfasst (Boehle/Schimpl-Neimanns 2010).

Eine weitere Ursache für den systematischen Ausschluss bestimmter Bevölkerungsgruppen aus der Grundgesamtheit bei allgemeinen Bevölkerungsumfragen liegt in der Ermittlung der Stichprobe. Da bei der Stichprobenziehung lediglich Anschriften und nicht Personen gezogen werden, fallen Obdachlose und im Ausland lebende Deutsche, die nicht mehr in Deutschland gemeldet sind, schon von vornherein aus der Stichprobe (Schnell 1991). Durch diesen *Undercoverage-Fehler* ist vor allem der Faktor der Allgemeinheit dieser Bevölkerungsbefragungen fraglich (Schnell 1991).

### 61.1.2 Verweigerung

Die Motive zur Verweigerung der Teilnahme an einer Befragung sind zwar individuell verschieden, dennoch ist bei bestimmten Gruppen vermehrt damit zu rechnen. Eine überdurchschnittliche Tendenz der Verweigerung ist dabei vor allem bei älteren Menschen (Motel-Klingenbiel/Klaus/Simonson, Kapitel 59 in diesem Band), Angehörigen bildungsferner Schichten und Eliten zu finden, die sich jedoch in ihren Beweggründen unterscheiden (Scholl 2003: 223-226).

#### 61.1.2.1 Angehörige bildungsferner Schichten

Bei der Teilnahmebereitschaft an allgemeinen Bevölkerungsumfragen lassen sich auch schichtspezifische Unterschiede feststellen. Diese Unterschiede ergeben sich dabei nicht aus der Schichtzugehörigkeit an sich, sondern stellen einen Bildungseffekt dar. Bei Personen mit nur geringem Bildungsstand ist die Verweigerungsrate demnach größer als bei Personen mit höherem Bildungsstand. Ein Grund dafür ist die Angst weniger gut gebildeter Personen vor etwaigen negativen Konsequenzen, die sie bei einer Teilnahme befürchten (Schnell 1997: 202-205). Unterschichtangehörige sehen in einer Befragung häufig eine Art Test und befürchten, den kognitiven Anforderungen einer Befragung nicht gewachsen zu sein, weshalb sie oft aufgrund von Minderwertigkeitsgefühlen verweigern. Um diesem

Problem zu begegnen, sollte bei der Interviewführung eine Vertrauen fördernde Atmosphäre geschaffen werden. Narrative Interviews als Befragungsmethode helfen, den Testcharakter eines Fragebogens zu umgehen (Scholl 2003: 224-226). Bei Befragungen, die in Armutsgebieten stattfinden, hat es sich als hilfreich erwiesen, bei der Erhebung verschiedene Methoden einzusetzen (u.a. Jahoda et al. 1975).

### 61.1.2.2 Eliten und Funktionseliten

Eliten, also sehr Vermögende und Personen, die Spitzenpositionen in Bereichen der Politik, der Verwaltung, des Militärs, der Kirchen, des öffentlichen Lebens etc. innehaben (Hartmann 2007: 18), sind für allgemeine Befragungen oft nur sehr schwer oder gar nicht erreichbar. Für diese Gruppe ist es schwer, Anreize zu setzen, die zur Teilnahme an einer Befragung motivieren können (Schnell 1991).

Bei Topmanagern kann das Interesse an Unterstützung bei eigenen Entscheidungsfindungen durch Forschung oder die Aussicht, durch die Teilnahme einen Innovationsvorsprung zu erhalten, einen entsprechenden Anreiz darstellen. Ein weiteres Motiv kann in der gesellschaftlichen Reputation der forschenden Institution liegen. Ist diese hoch, so kann die Mitwirkung einen Vorteil versprechen (Brandl/Klinger 2006).

Scholl sieht eine weitere Ursache für die Verweigerung der Befragungsteilnahme von Angehörigen der Eliten in der Standardisierung von Befragungen. Eliten möchten nicht mit der allgemeinen Bevölkerung gleichgesetzt werden und empfinden, dass sie ihre Meinung durch standardisierte Antwortvorgaben nicht ausreichend zum Ausdruck bringen können. Eine Möglichkeit, dieses Problem zu umgehen, ist, diese Gruppe beispielsweise per Leitfadeninterviews zu befragen (Scholl 2003: 225).

### 61.1.3 Nicherreichbarkeit

Neben der bewussten Verweigerung, sich an Umfragen zu beteiligen, ist ein weiter Grund für Nonresponse der, dass Teile der Stichprobe für den Interviewer nicht erreichbar sind. So sind sehr mobile Menschen, die häufig auch im Ausland unterwegs sind oder dort dauerhaft leben sowie Menschen ohne festen Wohnsitz schwierig für Befragungen zu kontaktieren (Schnell 1991).

#### 61.1.3.1 Mobile Deutsche

Mobile Deutsche sind beispielsweise Personen, die immer nur kurzfristig im Ausland arbeiten, Menschen, die im Schaustellergewerbe tätig sind oder Personen, die berufsbedingt unter der Woche auf Montage und nur am Wochenende an ihrem eigentlichen Wohnsitz zu erreichen sind (Schnell 1991). Zu der Gruppe der Personen, die häufig aufgrund ihres Berufes mobil sind, zählt auch die oben genannte Gruppe der Eliten (Schnell 1997: 204f.).

Eine weitere mobile Gruppe sind die Sinti und Roma. Zwar sind beide Bevölkerungsgruppen mittlerweile zum großen Teil sesshaft, jedoch gibt es immer noch einen Teil ohne festen Wohnsitz (Grienig 2010). Das Dokumentations- und Kulturzentrum Deutscher

Sinti und Roma schätzt, dass heute rund 70.000 Roma und Sinti mit deutscher Staatsbürgerschaft in Deutschland leben (Dokumentations- und Kulturzentrum Deutscher Sinti und Roma). Dazu kommt eine Dunkelziffer an illegal in Deutschland lebenden Roma und Sinti (Grienig 2010).

#### 61.1.3.2 Menschen mit längeren Auslandsaufenthalten

Zur Gruppe der Menschen, die temporär im Ausland leben, zählen diejenigen, die längerfristig im Ausland arbeiten, aber auch Studierende, die im Ausland Praktika oder Auslandssemester absolvieren (Schnell 1991) sowie eine gewisse Gruppe von Pensionären, die einige Monate im Ausland überwintern. Für ältere Menschen mit türkischem, aber auch anderem Migrationshintergrund, stellen Blohm und Diehl (2001) in ihrer Studie fest, dass sich diese oft für eine längere Dauer in ihrem Heimatland aufhalten, da sie nicht mehr an Schulferienzeiten oder Urlaubstage im Beruf gebunden sind.

#### 61.1.3.3 Obdachlose und Nichtsesshafte

Obdachlose und Nichtsesshafte verfügen über keinen festen Wohnsitz und fallen aus Befragungen meistens generell heraus, wenn sie sich nicht gerade in Notunterkünften speziell für Obdachlose befinden und diese mit in die Befragung aufgenommen werden (Boehle/Schimpl-Neimanns 2010). Ein weiterer Grund für die Unterrepräsentanz von Obdachlosen in Befragungen liegt unter anderem darin, dass die Lebensbedingungen der Betroffenen abschrecken und zu einem Vermeidungsverhalten auf Seiten des Interviewers führen können (Schnell 1991).

Nach der Definition der Bundesarbeitsgemeinschaft Wohnungslosenhilfe (BAG-W) zählen zu den Wohnungslosen obdachlose Personen, aber auch Personen, die in Heimen, Anstalten, Notüberwachungen, Asylen, Frauenhäusern oder übergangsweise bei Freunden und Verwandten untergebracht sind. Insgesamt zählen 2010 rund 248.000 Personen als Wohnungslose, wobei davon rund 9% „auf der Straße leben“ (Jordan 2011). Das entspricht einem Anteil von rund 0,3% Wohnungslosen, bzw. 0,03% Obdachlosen in der Bevölkerung.

Als eine gängige und bewährte Möglichkeit, Mitglieder dieser Gruppen ausfindig zu machen, hat sich das Schneeballverfahren erwiesen. Dabei geht man davon aus, dass sich die Mitglieder untereinander kennen und miteinander vernetzt sind. Entsprechend nutzt man einen Ausgangskontakt zur Weiterverteilung der Fragebögen (Häder 2010: 173-175).

#### 61.1.4 Nichtbefragbarkeit

Die Gründe dafür, warum Menschen für eine Untersuchung nicht befragbar sind, sind vielfältiger Art. Ob und inwieweit eine Person – etwa mit einer speziellen Behinderung – befragt werden kann, liegt unter anderem im subjektiven Ermessen des Interviewers. Die Wahl der Datenerhebungsmethode kann in gewissen Gruppen eine entscheidende Rolle spielen. Neben offensichtlichen Gründen gibt es aber auch, wie beim Analphabetismus,

für den Interviewer verborgene Gründe, die für betroffene Personen die Teilnahme an bestimmten Untersuchungen nahezu unmöglich machen können. Sind Hinderungsgründe im Vorhinein bekannt, so kann unter Umständen rechtzeitig vor der Erhebung, also bei der Planung des Designs darauf reagiert werden.

#### 61.1.4.1 Analphabeten

Von funktionalen Analphabeten wird gesprochen, wenn Menschen zwar lesen und schreiben können, diese Fähigkeiten aber nicht ausreichen, um im gesellschaftlichen Leben problemlos zurechtzukommen. Die Schreib- und Lesefähigkeit beschränkt sich in diesen Fällen auf einen einzelnen Satz, das Verfassen oder Lesen kurzer Fließtexte ist dagegen nicht ohne Weiteres möglich. Zu den funktionalen Analphabeten zählen auch die totalen Analphabeten, die höchstens in der Lage sind, einzelne Wörter zu identifizieren wie den eigenen Namen, nicht aber Sätze. In Deutschland leben schätzungsweise rund 7,5 Millionen funktionale Analphabeten (14% der erwerbsfähigen Bevölkerung) und 4% totale Analphabeten (Grotlüschen/Riekmann 2011). Insbesondere bei postalischen (Stocké, Kapitel 45 in diesem Band) und Online-Befragungen (Wagner/Hering, Kapitel 48 in diesem Band) ergeben sich aus diesen Gruppen starke Ausfälle. Da Analphabetismus nicht zufällig über die gesamte Bevölkerung verteilt ist, sondern beispielsweise Männer, ältere Menschen und Menschen, deren Eltern eine niedrigere Bildung besitzen, stärker betroffen sind, bedeutet ein Wegfall dieser Personengruppe eine starke Verzerrung der Stichprobe (Schnell 1991, Grotlüschen/Riekmann 2011). Die Unterrepräsentanz dieser Bevölkerungsgruppen aufgrund von Analphabetismus kann man minimieren, indem man für die Befragung Techniken verwendet, die für eine Teilnahme keine Lese- und Schreibkompetenz voraussetzen wie telefonische oder Face-to-Face Befragungen, in denen alle Fragen vorgelesen werden.

#### 61.1.4.2 Pflegebedürftige Personen

Pflegebedürftige Personen sind vor allem ältere und sehr alte Menschen, da diese häufiger mit Krankheiten konfrontiert sind, die es ihnen nicht ermöglichen, an Befragungen teilzunehmen. Die Pflegequote entspricht einem Anteil von knapp 20% bei den 80- bis 85-Jährigen und steigt stetig mit zunehmendem Alter (Statistische Ämter des Bundes und der Länder 2010: 25). In Folge dessen steigt auch die Ausfallquote bei Befragungen älterer Menschen. In Deutschland werden zwei Drittel der pflegebedürftigen Menschen zu Hause versorgt, während die restlichen pflegebedürftigen Menschen in Heimen untergebracht sind. Insgesamt sind 2,8% der Menschen in Deutschland, die 15 Jahre und älter sind, pflegebedürftig (Statistisches Bundesamt 2012b: 126). Befinden sich pflegebedürftige Menschen noch zu Hause, liegt es auch hier im Ermessen des Interviewers, inwieweit diese Menschen in die Befragungen aufgenommen werden können. Wenn eine Befragung möglich ist, bietet es sich in der Regel an, Face-to-Face Befragungen durchzuführen. Um den Ausfall bei Menschen, die aufgrund ihres geistigen Gesundheitszustandes nicht mehr in der Lage sind, selbstständig an einer Befragung teilzunehmen, zu vermeiden, kann die Durchführung von Proxy-Interviews sinnvoll sein.

#### 61.1.4.3 Psychisch kranke Personen

Psychisch kranke Personen sind Menschen mit Psychosen, Suchtkranke und Depressive, aber auch Menschen mit psychiatrischen Alterserkrankungen. Nach Schnell (1991) werden psychisch kranke Menschen nicht befragt, da zum einen ein potentieller Interviewer von der pflegenden Person keinen Zugang zur zu befragenden Person bekommt, zum anderen ein Interviewer in der Regel bereits von sich aus eine psychisch kranke Person aufgrund der geistigen Leistung als nicht befragbar einstufen würde (1991).

#### 61.1.4.4 Menschen mit körperlicher oder geistiger Behinderung

Inwieweit eine Befragung bei Menschen mit speziellen Behinderungen durchführbar ist, kann nicht generell beantwortet werden, da dies vom Grad der Behinderung abhängig ist. Gemäß der Schwerbehindertenstatistik von 2009 sind rund 350.000 Menschen in Deutschland blind oder sehbehindert, während rund 270.000 taub oder schwerhörig sind oder unter Sprach- und Sprechstörungen leiden. Da rund 98% der Schwerbehinderten in Deutschland 18 Jahre und älter sind, zählen ca. 0,8% der Bevölkerung zu dieser Gruppe (Statistisches Bundesamt 2012a: 124). Gemäß des Statistischen Jahrbuchs für die Bundesrepublik sind rund 911.000 Menschen geistig behindert oder leiden unter zerebralen Störungen, was einem Anteil von 1,1% der Bevölkerung über 14 Jahre entspricht (2012b: 124).

Die Befragungsmethode kann deutliche Auswirkungen besitzen: Beispielsweise können Menschen mit starker Beeinträchtigung des Sehvermögens postalische oder Online-Befragungen nicht problemlos ausfüllen, dafür aber an telefonischen oder Face-to-Face Befragungen teilnehmen. Für Menschen mit Hörschädigungen stellen dagegen telefonische und Face-to-Face Befragungen eine Hürde dar (Schnell 1991).

---

## 61.2 Gezielte Befragung spezieller Populationen

Bisher wurde die Problematik thematisiert, dass in allgemeinen Bevölkerungsbefragungen gewisse Teile der Bevölkerung aus verschiedenen Gründen nicht in der Brutto- oder Netstichprobe zu finden sind. Nun soll dargelegt werden, welche Problematiken und Besonderheiten zu beachten sind, wenn spezielle Populationen im Fokus des Forschungsinteresses stehen. Feldzugang und Interviewführung sind der jeweiligen Gruppe anzupassen.

### 61.2.1 Befragung spezieller und seltener Bevölkerungsgruppen

#### 61.2.1.1 Szenenbefragung

Für bestimmte Fragestellungen interessieren oft Personengruppen, die bestimmten Szenen zuzuordnen sind. Das können Homosexuelle, Angehörige bestimmter Jugendkulturen oder Sportarten, aber auch Kriminelle und Drogenabhängige sein. Spezifisch an diesen Gruppen ist dabei, dass es häufig keine genauen Verzeichnisse über ihre Mitglieder gibt. Gruppen, von denen kein Verzeichnis existiert, kann man über Zeitschriften, die

thematisch zu der interessierenden Gruppe passen, und über gruppenspezifische Webseiten oder Vereine erreichen und befragen (Dillman 2000: 245-262) oder über das Schneeballverfahren rekrutieren (Häder 2010: 173-175).

Erschwerend kommt hinzu, dass für einige dieser speziellen Gruppen vermehrte Schließtendenzen bestehen. So haben Mitglieder der Mafia aus offensichtlichen Gründen kein Interesse daran, ihre Verbindungen untereinander offenzulegen (Wagner 2009: 130f.). Auch muss – je nach Sensibilität der Gruppe und der interessierenden Fragestellung – mit einer generellen Verweigerung, nicht wahrheitsgemäßen oder sozial erwünschten Antworten gerechnet werden. Häufig gibt es keine genauen Angaben wie über die Anzahl der Opfer häuslicher Gewalt. Diese werden in den offiziellen Statistiken nur erfasst, wenn die Gewalttaten zur Anzeige gebracht werden. Da dies in vielen Fällen nicht geschieht, bleibt ein großer Teil im Dunkeln (Bundesministerium des Inneren 2011: 3). Eine Möglichkeit, Erkenntnisse zu wenigstens einem Teil zu erhalten, stellen Dunkelfeldbefragungen dar. Ein häufig angewandtes Instrument für die Erhebung delinquenter Verhaltens im Dunkelfeld ist die von Short/Ney (1957) eingeführte Methode der selbstberichteten Delinquenz (Self-Reported Delinquency), bei der Befragte Auskunft über ihre eigenen Straftaten geben.

Das Aufkommen des Problems der sozialen Erwünschtheit ist dabei abhängig von der Art der Frage, den dahinterliegenden vermuteten normativen Erwartungen sowie der Persönlichkeit des Befragten. Nach Scholl (2003) werden von der Norm abweichende Verhaltensweisen wie kriminelle Handlungen, Drogenkonsum oder ungewöhnliche sexuelle Vorlieben tendenziell eher von Menschen mit geringerem Selbstbewusstsein und einem Bedürfnis nach Anerkennung abgestritten. Fragen, die bedrohlich oder unangenehm wirken, verursachen beim Befragten soziale Kosten. Das erhöht wiederum die Wahrscheinlichkeit, dass Befragte entweder nicht wahrheitsgemäß antworten oder Antworten geben, die ihnen selbst fälschlicherweise sozial erwünschte Eigenschaften zuschreiben und durch die eigene sozial unerwünschte Eigenschaften abgestritten werden. Des Weiteren steigt ebenso auch die Tendenz, Fragen durch Meinungslosigkeit auszuweichen. Grund für dieses abweichende Antwortverhalten sind die Befürchtung von Konsequenzen oder das Ziel der persönlichen Anerkennung. Die Aufdeckung sozial erwünschter und damit fälschlicher Antworten ist äußerst schwierig, da sich diese je nach Persönlichkeitstyp, Milieu und kulturellem Hintergrund unterscheiden (Esser 1986, Scholl 2003: 207-210). Man kann aber davon ausgehen, dass schriftliche und Online-Befragungen den Effekt von sozialer Erwünschtheit im Vergleich zu telefonischen Befragungen senken, da hier auf einen persönlichen Kontakt durch den Interviewer verzichtet wird (Tourangeau/Smith 1996, Krysan et al. 1994). Ist die Verwendung von persönlichen Interviews unerlässlich, empfiehlt es sich, dass der Interviewer den Befragten ungestört, also ohne Anwesenheit Dritter, interviewt, dabei so neutral wie möglich erscheint und das Vorgehen der Fragestellung angepasst wird (Scholl 2003: 209-212).

### 61.2.1.2 Kontextbefragungen

Oft kann es von Interesse sein, Personen zu befragen, die sich in einem konkreten Kontext aufhalten bzw. aufgehalten haben. Das können beispielsweise Besucher einer bestimmten

Veranstaltung, Bahnfahrer oder Hotelgäste sein. Wie bei Szenenbefragungen stellt sich hier das Problem der Adressengenerierung. Zwar wäre ein telefonisches Screening möglich, das Herausfiltern der passenden Personen ist jedoch sehr aufwendig und kostspielig. Eine alternative Möglichkeit, an Adresslisten zu gelangen, stellen beispielsweise Face-to-Face Befragungen dar, die situativ durchgeführt werden. Problematisch ist hier aber der Einfluss Dritter oder des Interviewers, der eventuell als Teil des Kontextes angesehen wird. Häufig bietet es sich daher an, Menschen zu einem späteren Zeitpunkt zu befragen, um ihnen die Möglichkeit zu geben, Eindrücke mit Abstand zur Situation wiederzugeben. Gibt man Befragten lediglich einen Fragebogen mit und bittet sie, diesen im Nachhinein auszufüllen, ist die Wahrscheinlichkeit einer geringen Rücklaufquote groß. Nach Dillman kann man in solch einer Situation eine hohe Ausschöpfung erreichen, indem man Befragten zum einen den Befragungsinhalt näher erläutert, um die Befragung stärker im Gedächtnis zu verankern, und ihnen zum anderen die Wichtigkeit der Befragung und ihrer Teilnahme daran vermittelt. Des Weiteren sollte den Befragten die Dauer der Befragung mitgeteilt werden sowie ein Anreiz gesetzt werden, beispielsweise durch ein thematisch zum Inhalt der Befragung bzw. des Kontextes passendes Geschenk (Dillman 2000: 245-262).

## **61.2.2 Befragung von Institutionen und Institutionsmitgliedern**

Manchmal kann es von Interesse sein, Mitglieder oder Mitarbeiter bestimmter Institutionen oder Organisationen zu befragen. Ist dies der Fall, muss man bei der Untersuchungsplanung einige Besonderheiten beachten. Im Regelfall können diese Gruppen nicht direkt für eine Befragung kontaktiert werden, da die zuständige Institution vorab informiert werden muss und die Befragung gegebenenfalls einer Genehmigung bedarf.

### **61.2.2.1 Befragungen von Schülern**

Inwieweit Schüler befragt werden dürfen, ist in den jeweiligen Schulgesetzen der Länder geregelt. Für nähere Informationen sind in diesem Fall die einzelnen Schulgesetze einzubeziehen. In der Regel sehen diese jedoch vor, dass die Schulleitung oder Schulaufsichtsbehörde vorab eine Befragung genehmigen muss. Zu beachten ist dabei, dass Befragungen von Kindern im Vergleich zu Befragungen von Volljährigen einigen Besonderheiten unterliegen. Die Befragung von Kindern wird von Heinen/König (Kapitel 58 in diesem Band) detailliert erläutert.

### **61.2.2.2 Befragungen von Studierenden**

Studierende bilden eine Gruppe, die aufgrund des Alters und des Bildungsstandes weniger für methodische Probleme anfällig ist als andere Gruppen. Ein weiterer Vorteil liegt darin, dass es heutzutage Standard ist, dass Studierende bei der Einschreibung eine E-Mail Adresse der Universität erhalten. Infolgedessen sind sie mit einer nahezu hundertprozentigen Abdeckungsrate über diese erreichbar. Entsprechend bietet sich bei dieser Gruppe

eine Online-Befragung beispielsweise über den Universitätsverteiler an, vorausgesetzt, dass diese Vorgehensweise in den jeweiligen Universitätsgesetzen genehmigt ist.

---

### 61.3 Schlussbetrachtung

In allgemeinen Bevölkerungsumfragen besteht in Folge der Stichprobenselektivität immer die Gefahr, verzerrte oder falsche Schlussfolgerungen in Bezug auf einzelne Faktoren zu ziehen. Nehmen beispielsweise Menschen mit schlechtem Gesundheitszustand seltener an Befragungen teil, kann eine Verallgemeinerung der Ergebnisse dazu führen, die generelle Lebenszufriedenheit der Bevölkerung zu überschätzen. Dieses Problem ist insbesondere bei alten oder hochbetagten Menschen der Fall, die häufiger an Krankheiten leiden oder sich aufgrund ihrer Pflegebedürftigkeit in Heimen befinden und dadurch ebenso aus der Stichprobe ausgeschlossen werden.

Da sich der Ausschluss bzw. die Unterrepräsentanz bestimmter Bevölkerungsgruppen aus allgemeinen Bevölkerungsbefragungen aufgrund der vorgestellten Problematiken nicht vollständig vermeiden lassen, werden unter anderem für Analysen nachträglich Gewichtungsmethoden eingesetzt, um Verzerrung entgegenzuwirken. Diese Gewichtungsverfahren sind allerdings umstritten und mit Vorsicht zu verwenden (Kiesl, Kapitel 24 in diesem Band).

Möchte man auf der Basis allgemeiner Bevölkerungsumfragen zu deskriptiven Aussagen über spezielle Bevölkerungsgruppen gelangen, ist also Vorsicht geboten. Für die Untersuchung weiterführender inhaltlicher Fragestellungen, die sich auf spezielle Populationen beziehen, bedarf es aufgrund der Unterrepräsentanz in allgemeinen Bevölkerungsbefragungen und der häufig auf die spezielle Population ausgerichteten Fragestellung gezielter Befragungen, die genau auf die interessierende Populationen ausgerichtet sind und auch nur diese in der Stichprobe beinhalten. Datenerhebungsmethode, Feldzugang und Interviewführung müssen dann an die jeweilige Zielpopulation angepasst werden.

## Literatur

- Blohm, Michael/Diehl, Claudia (2001): Wenn Migranten Migranten befragen. Zum Teilnahmeverhalten von Einwanderern bei Bevölkerungsbefragungen. In: Zeitschrift für Soziologie 30: 223-242
- Boehle, Mara/Schimpl-Neimanns, Bernhard (2010): Mikrozensus Scientific Use File 2008. Dokumentation und Datenaufbereitung. In: GESIS-Technical Report 13
- Brandl, Julia/Klinger, Stephan (2006): Probleme des Feldzugangs zu Eliten. In: Österreichische Zeitschrift für Soziologie 31: 44-65
- Bundesministerium des Inneren (2011): Die Kriminalität in der Bundesrepublik Deutschland. Polizeiliche Kriminalstatistik für das Jahr 2011. Berlin: Bundesministerium des Inneren
- Dillman, Don A. (2000): Mail and Internet Surveys. The Tailored Design Method. New York: Wiley
- Dokumentations- und Kulturzentrum Deutscher Sinti und Roma. Sinti & Roma. <http://www.sintiundroma.de/sinti-roma.html> (Januar 2013)
- Esser, Hartmut (1986): Können Befragte lügen? Zum Konzept des „wahren Wertes“ im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. In: ZUMA-Arbeitsbericht 02
- Grienig, Gregor (2010): Roma in Deutschland. <http://www.berlin-institut.org/online-handbuchdemografie/bevoelkerungsdynamik/regionale-dynamik/roma-in-deutschland.html> (Februar 2013)
- Grotlüschen, Anke/Riekmann, Wibke (2011): leo. – Level-One Studie. Presseheft. Hamburg: Universität Hamburg
- Häder, Michael (2010): Empirische Sozialforschung. Eine Einführung. Wiesbaden: VS Verlag
- Hartmann, Michael (2007): Eliten und Macht in Europa. Ein internationaler Vergleich. Frankfurt/Main, New York: Campus
- Jahoda, Marie/Lazarsfeld, Paul F./Zeisel, Hans (1975): Die Arbeitslosen von Marienthal. Ein soziographischer Versuch über die Wirkungen langandauernder Arbeitslosigkeit, mit einem Anhang zur Geschichte der Soziographie. Frankfurt am Main: Suhrkamp
- Jordan, Rolf (2011): Statistikbericht 2011. Bielefeld: Bundesarbeitsgemeinschaft Wohnungslosenhilfe e.V.
- Kriwy, Peter/Gross, Christiane (Hg.) (2009): Klein aber fein! Quantitative empirische Sozialforschung mit kleinen Fallzahlen. Wiesbaden: VS Verlag
- Krysan, Maria/Schuman, Howard/Scott, Lesli J./Beatty, Paul (1994): Response Rates and Response Content in Mail Versus Face-To-Face Surveys. In: Public Opinion Quarterly 58: 381-399
- Schnell, Rainer (1991): Wer ist das Volk? Zur faktischen Grundgesamtheit bei „allgemeinen Bevölkerungsumfragen“: Undercoverage, Schwererreichbare und Nichtbefragbare. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 43: 106-137

- Schnell, Rainer (1997): Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske + Budrich
- Schnell, Rainer (2012): Survey-Interviews. Methoden standardisierter Befragungen. Wiesbaden: VS Verlag
- Scholl, Armin (2003): Die Befragung. Sozialwissenschaftliche Methode und kommunikationswissenschaftliche Anwendung. Konstanz: UVK Verlagsgesellschaft
- Short, James E./Nye, F. I. (1957): Reported Behavior as a Criterion of Deviant Behavior. In: Social Problems 5 (3): 207-213
- Statistische Ämter des Bundes und der Länder (2010): Auswirkungen auf Kranken-  
ausbehandlungen und Pflegebedürftige im Bund und in den Ländern. Demografischer  
Wandel in Deutschland 2. Wiesbaden: Statistisches Bundesamt
- Statistische Ämter des Bundes und der Länder (2011): Das registergestützte Verfahren  
beim Zensus 2011. Wiesbaden: Statistisches Bundesamt
- Statistisches Bundesamt (2012a): Sozialleistungen schwerbehinderten Menschen. 2009. In:  
Fachserie 13 Reihe 5.1. Wiesbaden: Statistisches Bundesamt
- Statistisches Bundesamt (2012b): Statistisches Jahrbuch für die Bundesrepublik  
Deutschland 2012. Wiesbaden: Statistisches Bundesamt
- Tourangeau, Roger/Smith, Tom W. (1996): Asking Sensitive Questions. The Impact of  
Data Collection Mode, Question Format and Question Context. In: Public Opinion  
Quarterly 60: 275-304
- Wagner, Simone (2009): Datenerhebung bei Spezialpopulationen am Beispiel der  
Teilnehmer lokaler Austauschnetzwerke. In: Kriwy et al. (Hg.): 127-147

---

## **Teil 6**

### **Weitere Datentypen**

Axel Salheiser

---

## 62.1 Was sind Dokumente und was ist Dokumentenanalyse?

### 62.1.1 Dokumente

Unter *Dokumenten* sollen in diesem Kapitel natürliche Daten verstanden werden, die in schriftlicher Form als Texte vorliegen. Dokumente sind insofern *natürliche Daten*, als dass sie nicht zu Forschungszwecken und ohne die Beteiligung oder Intervention der Forschenden entstanden sind. Für die Sozialwissenschaft sind folgende *Dokumententypen* relevant:

- *Offizielle Dokumente bzw. Publikationen*, z.B. Hand- und Jahrbücher, Bücher (Belletristik, Fachliteratur) (Ernst, Kapitel 63 in diesem Band), Manuskripte öffentlicher Reden, Zeitschriften und Zeitungen (Klein, Kapitel 64 in diesem Band), Broschüren, Flugblätter, Karten usw.;
- *Interne Dokumente aus öffentlichen/staatlichen Verwaltungsorganisationen und privaten Unternehmen/Organisationen* (Liebig et al, Kapitel 75 in diesem Band), z.B. Akten, Statistiken, Listen, d.h. Buchhaltungs- und Registerdaten (Hartmann/Lengerer, Kapitel 71 in diesem Band); Dienst- bzw. Geschäftskorrespondenz, Gesprächsprotokolle/-notizen und andere Arbeitsmaterialien;
- *Egodokumente* (Ernst, Kapitel 63 in diesem Band), d.h. persönliche Unterlagen wie z.B. Briefe, private Aufzeichnungen und Notizen, Tagebücher, autobiographische Zeugnisse wie z.B. verschriftlichte Augenzeugenberichte und Gedächtnisprotokolle.

Dokumente können sowohl in gedruckter als auch in handschriftlicher Form auf Papier vorliegen, aber auch auf Mikrofiche, Mikrofilm oder anderen Datenträgern und (elektronischen) Medien gespeichert sein.

Bei Dokumenten handelt es sich um (Träger von) *prozessproduzierten Daten* (Müller 1977: 1): „Unter prozessproduzierten Daten werden all diejenigen Daten verstanden, die als Aufzeichnungen öffentlicher und privater Organisationen im Rahmen ihrer Tätigkeit

und nicht nur zum Zwecke wissenschaftlicher bzw. statistischer Auswertung gesammelt werden bzw. wurden.“

Diese Definition wurde in erster Linie für Daten aus der „sozialen Buchführung“ (der öffentlichen Sozialverwaltung, des Personalwesens in Unternehmen usw.; Bick/Müller 1984: 42) gefunden. Sie lässt sich jedoch auch auf schriftliche Quellen ausweiten, die Individualpersonen zuzurechnen sind, z.B. private Aufzeichnungen, Briefe oder andere Ego-Dokumente (Ernst, Kapitel 63 in diesem Band).

Dokumente sind *Artefakte kommunikativer Praxis* (Schubert, Kapitel 70 in diesem Band) und können deshalb eine Datenbasis für sozialwissenschaftliche Interpretationen darstellen (Hitzler/Honer 1997: 8, Prior 2003). Maßgeblich ist, welche Erkenntnisse über soziales Handeln sich aus einem Dokument erschließen lassen und welche sozialen Prozesse und Strukturen daraus rekonstruiert werden können. Beispielsweise ermöglicht die Auswertung von Akten Aussagen über das Verwaltungshandeln, das in ihnen abgebildet ist (Bick/Müller 1984: 74).

In der Geschichtswissenschaft spricht man von Dokumenten als *schriftliche Quellen* oder *Zeugnisse*, in denen historische Ereignisse geschildert werden oder historische Sachverhalte zum Ausdruck kommen. Typische Beispiele dafür sind historische Briefe, Chroniken und Urkunden.

### **62.1.2 Ziele und Varianten der Dokumentenanalyse**

Da schriftliche Quellen traditionell die wichtigste Datengrundlage für die Geschichtsforschung sind, kommt einer *Quellenkunde* eine zentrale Bedeutung zu, in der die kritische Bewertung der Authentizität, Verlässlichkeit und Aussagekraft der Quellen gewährleistet wird (Beck/Henning 2003, Rusinek et al. 1992). Bei historischen Dokumenten sind neben der Autorenschaft sowie der zeitlichen und geographischen Einordnung auch der Informationsgehalt und die Bedeutung des Dokumentes zu klären, was üblicherweise im Vergleich mit anderen Quellen geschieht, die als weitestgehend gesichert gelten können.

Doch auch die Erforschung von Gegenwartsgesellschaften und neueren und neuesten sozialen Phänomenen kann sich auf die Auswertung entsprechender schriftlicher Quellen, also Dokumente, stützen. Grundsätzlich unterscheidet sich *sozialwissenschaftliche Dokumentenanalyse* von der historischen Quellenarbeit zunächst durch den *zeitlichen Referenzrahmen* und das damit verknüpfte fachspezifische Erkenntnisinteresse. Die Geschichtswissenschaft stellt meist eine narrativ-deskriptive bzw. „philologisch-hermeneutische“ (Weyrauch 1977: 141) Quellenarbeit in den Mittelpunkt, um beispielsweise Ereignis- oder Ideengeschichte zu rekonstruieren. Die Untersuchung historischer Prozesse bzw. von Entwicklungslinien sozialen Wandels (besonders der Neuzeit) hat vor allem innerhalb der historischen Sozialforschung ihre Heimat gefunden, wobei die interdisziplinäre Verschränkung mit der Soziologie auch eine Öffnung der Geschichtswissenschaft für neue Fragestellungen und neue methodische Ansätze bewirkt hat, darunter quantitative bzw. statistische Methoden.

In der modernen Soziologie ist hingegen zwar die Befragung (Helfferich und Reinecke, Kapitel 39 und 44 in diesem Band) als „Königsweg“ (Scheuch 1977) der Datenerhebung etabliert, die Dokumentenanalyse (als Sekundärdatenanalyse) kann aber gerade dort wertvolle Impulse liefern, wo Primärerhebungsdaten nicht vorliegen, dem Erkenntnisinteresse entsprechend ergänzt werden sollten oder ein anderweitiger Feldzugang erschwert ist. Abhängig vom Dokumententyp sind sowohl *explorative* als auch *konfirmatorische Dokumentenanalysen* möglich, was auf die Anknüpfungspunkte zu diversen qualitativen (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) und quantitativen (Stein, Kapitel 7 in diesem Band) Arbeitsweisen und Forschungstraditionen verweist.

*Ziel einer sozialwissenschaftlichen Dokumentenanalyse* kann es sein, die *Interaktionsmuster, institutionellen Handlungskontexte sowie „Wertorientierungen oder Meinungen von Individuen oder Gruppen aufzudecken“* (Best 1977: 162). Anschließend an die klassischen Definitionen der *Inhaltsanalyse* soll dabei von manifesten Textinhalten auf latente soziale Kontexte geschlossen werden (Früh 2011, Mayring/Fenzl, Kapitel 38 in diesem Band).

In erster Linie ist entscheidend, in welcher Struktur und in welchem Umfang die Dokumente vorliegen, so dass beispielsweise im Fall der massenhaften Verfügbarkeit des Dokumententyps B (interne Dokumente aus Verwaltungsorganisationen) eine quantitative Analyse (Hartmann/Lengerer, Kapitel 71 in diesem Band) „prozessproduzierter Daten“ (Scheuch 1977) sinnvoll ist. Hingegen dürften bei einzelnen, unter Umständen sogar nur fragmentarischen Dokumenten (vor allem der Dokumententypen A, C) qualitative, d.h. hermeneutische bzw. interpretative (Kurt/Hebrik, Kapitel 33 in diesem Band) und inhaltsanalytische Verfahren (Mayring/Fenzl, Kapitel 38 in diesem Band) angemessen sein. Nachfolgend werden Grundlagen, Anwendungspotentiale und Probleme der Dokumentenanalyse diskutiert.

Prinzipiell ist zu unterscheiden zwischen der primären, *manifesten* Sinnebene des Textes, auf der soziologisch relevante Sachverhalte durch die Darstellungen und Bewertungen des Autors wiedergegeben und vom Forscher direkt „abgelesen“ und interpretiert werden können (z.B. Ablauf und Bewertung einer Interaktionssituation), und sekundären, *latenten* Sinnebenen, auf die aufgrund des Textes geschlossen werden kann (z.B. Bedeutung der Interaktion). Beispielsweise ermöglichen das Vokabular und der Sprachstil in einem Dokument Aussagen über die Bildung und den Status des Autors, während der geschilderte Sachverhalt eher „unpersönlich“ ist. Strukturierende und zusammenfassende inhaltsanalytische Auswertungsstrategien können hier zwar für viele Fragestellungen zielführend sein, oftmals stellt sich aber die Frage eines tieferen Sinnzusammenhangs zwischen Dokument (Text) und sozialem Umfeld (Kontext). Ebenso wird der Ganzheitlichkeit eines Textes durch Kategorienbildung (Textvercodung) und zusammenfassende bzw. quantifizierende Auswertungen nur selten Rechnung getragen. Ein weiteres Problem konventioneller inhaltsanalytischer Methoden (Mayring/Fenzl, Kapitel 38 in diesem Band) ist, dass eine Bedeutungsäquivalenz der Begriffe im Verständnis von Textproduzent und Analytiker unterstellt wird (Hitzler/Honer 1997: 23). Oftmals lassen sich Sinnmuster in einem Dokument jedoch nur durch eine hermeneutische Textinterpretation, z.B. mittels der Sequenzanalyse, rekonstruieren. Besonders zu beachten sind dabei implizite oder explizite Verweise auf *Außertextliches*, wie metaphorische Anspielungen und charakteristische

sprachliche Floskeln. Auch die Bedeutung intenderter oder unintenter Auslassungen und Verkürzungen kann zentral für ein soziologisches *Textverstehen* (Kurt/Herbrik, Kapitel 33 in diesem Band) sein.

Neben inhaltlichen und sprachlichen Aspekten können außerdem eine Reihe weiterer (formaler) Textgestaltungsmerkmale und Dokumenteneigenschaften sozialwissenschaftlich relevant sein; von der Schriftart und Schriftfarbe über das Trägermaterial, auf dem das Dokument verfasst wurde, bis zum Erhaltungszustand (z.B. Beschädigungen). Die eigentliche *Bedeutung* und der Kontext eines Druckerzeugnisses oder einer maschinengeschriebenen Akte mögen sich erst durch einen eher unscheinbaren handschriftlichen Zusatz oder eine hastige Streichung erschließen. Die Handschrift und Papierqualität eines Briefes wiederum können wichtige Aussagen über den Charakter der Beziehung zwischen Sender und Empfänger sowie allgemeine gesellschaftliche Normen und Konventionen liefern.

---

## 62.2 Methodologische Grundlagen der Dokumentenanalyse

Anders als bei der Erhebung von Primärerhebungsdaten aus sozialwissenschaftlichen Befragungen oder in Experimenten (Eifler, Kapitel 11 in diesem Band) bietet sich mit der Dokumentenanalyse ein *nicht-reaktives Verfahren* (Webb et al. 1975), d.h. die Reaktion der Untersuchungsobjekte wird nicht durch die Datenerhebungstechnik selbst verzerrt. Die außerwissenschaftliche Herkunft der Daten legt allerdings nicht nur dem Historiker, sondern auch dem Sozialwissenschaftler die Pflicht auf, die Validität der Datenbasis zu kontrollieren und zu gewährleisten. Eine Dokumentenanalyse kann niemals nur *datenimmmanent* bleiben, sich also allein auf die Dokumenteninhalte berufen, sondern muss mit einer *Rekonstruktion des Entstehungs- und Nutzungskontextes* des Dokumentes einhergehen: „Texte sind als Teile größerer Kommunikationsprozesse aufzufassen. Die Interpretation des Einzeltextes erfordert die Rekonstruktion der Gesamtstruktur des Kommunikationsprozesses, dessen Teilelement vorliegt“ (Heckmann 1992: 152).

Scheuch (1977: 25) hat auf zwei wichtige Probleme der Nutzung prozessproduzierter Daten für die Forschung verwiesen, nämlich auf

1. die *Selektivität der Daten*, die sich aus dem Entstehungszusammenhang und Verwendungszweck der Daten (z.B. dem Auftrag bzw. der Zuständigkeit einer aktenführenden Verwaltungsbehörde) ergibt, und
2. die *Kategorisierung* bzw. Struktur der *Daten*, die den Erfordernissen der Verwaltungstätigkeit, aber nicht zwangsläufig wissenschaftlich-methodischen Standards entspricht.

Nach dem Verständnis sozialwissenschaftlicher Forschung sind nicht selbst erhobene *Daten* also systematisch verzerrt (Müller 1977: 2, Häder/Häder und Engel/Schmidt, Kapitel 18 und 23 in diesem Band). Um jedoch den Ansprüchen an Reliabilität und Validität (Menold/Krebs, Kapitel 30 in diesem Band) gerecht zu werden, ergibt sich zunächst das Erfordernis einer *Fehlerlehre* (Baur 2009: 21). Diese Fehlerlehre ist analog einer historischen Quellenkritik (Bick/Müller 1984: 42) und muss klären, *zu welchem Zweck* und

wie die Daten produziert, genutzt und überliefert wurden. Nur so können die *Ursachen* und das *Ausmaß* von Verzerrungen und Datenlücken evaluiert werden, um anschließend Anpassungs- und Korrekturmaßnahmen vorzunehmen.

### 62.2.1 Umgang mit der Selektivität der Daten

Hinsichtlich der Selektivität der Daten sind besonders jene *systematischen Abbildungsverzerrungen* zu beachten, die sich aus einer heterogenen Datenerfassungs-, -selektions- und -nutzungspraxis innerhalb der datenerzeugenden (aktenführenden) Verwaltungsorganisation ergeben und als „*shifting validity*“ bezeichnet werden. Insofern entspricht die Fehlerlehre prozessproduzierter Daten einer „*organisationssoziologischen Quellenkritik*“ (Heckmann 1992: 154):

1. Einerseits können *innerhalb einer Organisation ganz unterschiedliche Dokumentations- und Aufbewahrungsvorschriften bzw. Zuständigkeiten der Informationsverarbeitung* existieren, denen zudem im betrieblichen bzw. Behörden-Alltag unterschiedlich folgeleistet wird. Nicht alle Verwaltungsvorgänge sind dokumentationspflichtig oder werden als dokumentationswürdig betrachtet und werden deshalb (in gleichem Umfang) aktenkundig. Aber auch Manipulationen von Buchhaltungsunterlagen bzw. Berichtsfälschungen (vor allem statistischer Angaben) können relevant sein (Webb et al. 1975: 79). Unter Umständen existiert eine administrative Geheimhaltungspraxis, die zu interner Intransparenz und einer unvollständigen Überlieferung von Dokumenten bzw. Daten führt. Beispielsweise kann bei Personalakten die Verarbeitung und Kontrolle den Mitarbeitern eine unterschiedliche Priorität eingeräumt werden. So kommt es schließlich dazu, dass die Vollständigkeit, die Aussagekraft und die Fehleranfälligkeit von personenbezogenen Angaben über Funktionsträger von ihrer vertikalen und horizontalen Position in der Organisationshierarchie abhängen (Salheiser 2009).
2. Andererseits sind *technische Probleme bei der Datenerfassung* (z.B. die Störanfälligkeit von EDV-Anlagen), „*subjektive*“ *Opportunitätsstrukturen und die individuelle Motivation des Personals* für die Datenqualität relevant (z.B. Alltagsroutinen der Eingabe-, Kodier- und Kontrollpraxis, mangelnde Sorgfalt oder Kompetenz usw.), sie sind mitunter eine unsystematische und nur schwer zu rekonstruierende Fehlerquelle.

Um die Fehleranfälligkeit und Phänomene der „*shifting validity*“ beurteilen zu können, sind *Strategien der externen Validierung* unabdingbar: So muss neben einer *Analyse von Vergleichsdaten* (sofern verfügbar) auch *Kontextwissen* über den primären Entstehungs- und Nutzungsprozess der Daten und die Struktur der aktenführenden Verwaltungsorganisation generiert und genutzt werden. Dies erfordert i.d.R. die Analyse weiterer Dokumente (Akten, Organigramme usw.) sowie die externe Validierung anhand von Vergleichsstatistiken. Gegebenenfalls können Experteninterviews (Helfferich, Kapitel 39 in diesem Band) mit solchen Personen hilfreich sein, die selbst an der Generierung der Daten bzw. der Erstellung der Dokumente beteiligt waren.

## 62.2.2 Umgang mit der Struktur (Kategorisierung) der Daten

Bezüglich der Kategorisierung prozessproduzierter Daten ist wichtig, dass in vielen Fällen keine unmittelbare Kompatibilität mit den Standards sozialwissenschaftlicher Messung (Stein, Krebs/Menold und Faulbaum, Kapitel 7, 30 und 31 in diesem Band) gegeben ist. Die in den Dokumenten verwendeten *Begriffe* können beispielsweise unpräzise oder widersprüchlich sein und einem spezifischen Verwaltungsjargon unterliegen. Oftmals folgen *Merkmale* und *Kategorienschemata* (z.B. zur Erfassung personenbezogener Daten) der Eigenlogik der Verwaltungsorganisation. Die Aussagekraft der Daten für wissenschaftliche Auswertungen kann beeinträchtigt sein, wenn bei der Datenerfassung unterschiedliche Merkmale oder Kategorien zusammengefasst wurden (Informationsverlust). Andererseits können bestimmte Merkmale, Merkmalsausprägungen bzw. Unterkategorien existieren, die für Forschungszwecke irrelevant sind (redundante Informationen). Möglicherweise sind zusammengehörende Angaben auch über verschiedene Teildatensätze verstreut, so dass sie aus unterschiedlichen Merkmalen extrahiert, neu geordnet und zusammengeführt werden müssen, um vollständige und konsistente Informationen zu gewinnen.

Die Verwendung von Begriffen und Kategorienschemata kann außerdem nach politisch-ideologischen oder strategischen Vorgaben erfolgt sein (u.a. zur Verschleierung unerwünschter Sachverhalte etc.; Salheiser 2009), was eine direkte Übernahme der Daten bzw. der Datenstruktur für wissenschaftliche Auswertungszwecke und eine neutrale Beurteilung unmöglich macht.

Die Sprache der Dokumente ist per se nicht die Sprache der Wissenschaft, sondern sie unterliegt den Konventionen und der Logik des sozialen Kontextes, aus dem die Dokumente stammen, und bedarf deshalb einer Übersetzung. Daraus ergibt sich die praktische Notwendigkeit einer nachträglichen Operationalisierung (Stein, Kapitel 7 in diesem Band), was der Analyse prozessproduzierter Daten eine Sonderstellung innerhalb sozialwissenschaftlicher Ex-post-facto-Forschungsdesigns (Stein, Kapitel 7 in diesem Band) einräumt. Dabei gilt es, die aufgefundenen Merkmalskategorien der sozialen Buchführung in brauchbare Indikatoren zu überführen und die prozessproduzierten Daten dementsprechend datentechnisch aufzubereiten und umzustrukturieren.

Unter Umständen treten bei historischen prozessproduzierten Daten Konsistenzprobleme auf, wenn Begriffe, Bezeichnungen, Maßeinheiten, statistische Kennziffern usw. während der Datenproduktion über längere Zeiträume einem Bedeutungswandel unterlagen und Kodierschemata aus technischen oder organisatorischen Gründen abgeändert wurden. Hier ist prozessbezogenes Kontextwissen unerlässlich, um die Reliabilität und Validität der Daten zu kontrollieren und sicherzustellen.

Auch Dokumente, deren Inhalte als „qualitative Daten“ zu klassifizieren sind, bedürfen zunächst einer grundlegenden Validierung und Plausibilisierung (Flick, Kapitel 29 in diesem Band). Betrachtet man beispielsweise einen persönlichen Bericht als Zeugnis sozialen Handelns, so sind, wie in der historischen Quellenkunde, der zeitliche, der geographische und der soziale Entstehungskontext des Berichtes sowie die Autorität, Kompetenz und

Glaubwürdigkeit des Verfassers einzuordnen bzw. zu beurteilen (Howell/Prevenier 2004: 82ff.). Wo lassen sich vergleichbare Angaben nachweisen, wo treten Widersprüche zu Tage? Inwieweit korrespondieren die Dokumente mit anderen Quellen, gibt es etwa spiegelbildlich überlieferte Bestände, die sich aus internen Querverweisen und Belegen rekonstruieren lassen (z.B. Verteiler-Abkürzungen auf Verwaltungsakten)?

Briefwechsel, die nur auf der Empfängerseite überliefert sind, gestatten mitunter erst durch die Zusammenführung aus verschiedenen Beständen und durch eine komparative Textanalyse valide Aussagen über den Forschungsgegenstand. Grundsätzlich gilt: Keine Quelle und kein Dokument kann Anspruch auf Objektivität erheben (Menold/Krebs, Flick, Herbrik/Kurt, Kapitel 30, 27 und 33 in diesem Band). Jede Wiedergabe von Sachverhalten, mag sie auch noch so neutral anmuten, unterliegt einem Konstruktionsprozess der Wahrnehmung und der kognitiven Verarbeitung und bildet einen der Intention, Motivation und Emotion des Autors entsprechend gefilterten *Ausschnitt* sozialer Realität ab, sie ist also subjektiv gefärbt und „verzerrt“.

---

## 62.3 Datenzugang

Wie alle anderen Datentypen können auch Dokumente nur einen Ausschnitt (subjektiv wahrgenommener) sozialer Wirklichkeit wiedergeben. Forschungspraktisch ergibt sich zunächst die Problematik des Zugangs zu geeigneten Dokumenten, insofern sie nicht ohne weiteres als Publikationen (Dokumententyp 1) verfügbar sind.

### 62.3.1 Erhebung zusammen mit anderen Daten

Persönliche Dokumente (Dokumententyp 3), die in mikrosoziologischen Forschungskontexten relevant sind, wie z.B. Egodokumente, Briefe, Urkunden, werden in der Regel im Zusammenhang mit dem Feldzugang anderer Datenerhebungsmethoden – der narrativen Interviewführung (Küsters, Kapitel 40 in diesem Band), der teilnehmenden Beobachtung (Thierbach/Petschick, Knoblauch, Kapitel 66 und 36 in diesem Band) usw. – verfügbar, zumindest wenn sie im sozialen Umfeld des Verfassers und seiner Interaktionspartner verblieben sind. Andernfalls stellt sich die Frage nach der Archivwürdigkeit bzw. einer archivischen Überlieferung.

### 62.3.2 Dokumente im Privatbesitz

Ein persönliches Dokument kann aufgrund seines Inhaltes (z.B. des biographischen und familienbiographischen Bezugs) von hohem ideellem Wert für seinen Besitzer sein. Wird die Einsicht, die Reproduktion, die vorübergehende oder gar dauerhafte Überlassung bzw. der Erwerb solcher Dokumente beabsichtigt, sollte der Forscher besonderes Augenmerk

auf die Vermittlung von Vertrauenswürdigkeit und der Seriosität seines Vorhabens legen. Eine schriftliche Übereinkunft über die Nutzung der Dokumente (insbesondere von Angehörigen bzw. von Hinterbliebenen) sollte auf jeden Fall vorliegen, auch wenn die Veröffentlichung der Dokumente bzw. der Forschungsergebnisse in der Regel als völlig unproblematisch eingeschätzt wird. Erfahrungsgemäß kann sich jedoch während des Forschungsprozesses die Kooperationsbereitschaft derer ändern, die persönliche Dokumente von sich oder aus dem Nachlass verstorbener Angehöriger zur Verfügung gestellt haben. Dem ausdrücklichen Wunsch der Nichtveröffentlichung zuwiderzuhandeln, könnte einen ernstzunehmenden Verstoß gegen Prinzipien der Forschungsethik darstellen und u.U. sogar straf- und zivilrechtliche Konsequenzen nach sich ziehen.

### **62.3.3 Dokumente aus Organisationen**

Die Verfügbarkeit von internen Dokumenten öffentlicher und privater Organisationen (Dokumententyp 2) für eine wissenschaftliche Nachnutzung, darunter auch der Zugriff auf prozessproduzierte Massendaten, ist zunächst abhängig von der Kooperationsbereitschaft der Organisationen (Liebig et al, Kapitel 75 in diesem Band) und von übergeordneten datenschutzrechtlichen Belangen (Mühlchen, Kapitel 4 in diesem Band). Insbesondere bei historischen Dokumenten stellt sich aber auch die Frage nach einer archivischen Überlieferung, der Vollständigkeit und dem Erhaltungszustand solcher Daten (Webb et al. 1975: 80). Wie Historiker wissen, kann die Erschließung archivalischer Datenquellen deshalb Detektivarbeit sein. Vor allem der Einsicht in umfangreiche Aktenbestände bzw. von prozessproduzierten Massendaten sollte eine gründliche Planung vorausgehen.

### **62.3.4 Dokumente aus Archiven und Dokumentationsstellen**

Öffentliche und private Archive oder Dokumentationsstellen gehören zu den wichtigsten Institutionen, die Dokumente zur Verfügung stellen. Das geeignete Archiv muss entsprechend seiner sachlichen und räumlichen Zuständigkeit ausgewählt werden. Auf nationaler und internationaler Ebene stehen dafür Archivverzeichnisse und Online-Portale zur Verfügung (Arathymou 2009). Archive sind in der Regel an die Verwaltungen bzw. Einrichtungen angegliedert, deren Akten sie aufbewahren. So ist zwischen Archiven des Bundes, den Landesarchiven, kommunalen Archiven, Archiven von Universitäten und anderen wissenschaftlichen Einrichtungen sowie Archiven von Wirtschaftsunternehmen, Verbänden, Stiftungen und Vereinen zu unterscheiden. Bei historischen Akten muss beachtet werden, dass die Zuständigkeit des Archivs von einer Rechtsnachfolge der jeweiligen Organisation bzw. politischen Einheit (Staat, Bundesland, Kommune) und der Territorialgeschichte abhängen kann.

Der Zugang zu staatlichen Archiven in Deutschland ist in den Archivgesetzen des Bundes und der Länder geregelt. Bei privaten Archiven liegt es hingegen im Ermessensspielraum

der Organisationen, inwiefern sie ihre Archivalien zur wissenschaftlichen Nutzung bereitstellen. Öffentliche und private Archive können sich außerdem erheblich bezüglich der Nutzungsbedingungen unterscheiden. Dies betrifft insbesondere die Veröffentlichung der gewonnenen Befunde bzw. die Reproduktion der benutzten Dokumente. So muss beispielsweise ein wissenschaftliches Interesse an größtmöglicher Transparenz mit den Interessen eines Unternehmens an seinem Image in der Öffentlichkeit abgestimmt werden. Es ist unabdinglich, sich im Vorfeld eines Forschungsvorhabens möglichst umfangreich über die allgemeinen Archivnutzungsvorschriften, die Bedingungen der Akteneinsicht und die verfügbaren *Findmittel* zu informieren. (*Findmittel* sind schriftliche oder elektronische Verzeichnisse des Aktenbestandes.)

Staatliche bzw. öffentliche Archive wie das Bundesarchiv stellen auf ihren Internetseiten hilfreiche Angaben zur *Archivstruktur*, *Bestandsübersichten* und weitreichende elektronische Findmittel zur Verfügung. Bei kleineren öffentlichen Archiven und Unternehmensarchiven erfolgt die Auskunft üblicherweise mittels schriftlicher Kontaktaufnahme mit den Archivmitarbeitern bzw. einer wissenschaftlichen Begleitinstanz der archivführenden Institution.

Die Einsicht und die Nutzung von Archivgut bedürfen in der Regel einer vorherigen *Anmeldung* bzw. *Nutzungserlaubnis*. Dabei sind mitunter Sonderregelungen und Nutzungseinschränkungen für Teilbestände zu beachten, die (a) sensible Daten enthalten und damit besonderen Datenschutzvorschriften (z.B. Sperrfristen) unterliegen oder (b) aus archivischen Gesichtspunkten (Alter, Erhaltungszustand, historische Bedeutung) als besonders schützenswert beurteilt werden.

In vielen Fällen ist ein zeitlicher Abstand zwischen Anmeldung und gewünschtem Zeitpunkt der Dokumenteneinsicht von mehreren Tagen oder sogar Wochen notwendig, da das Archivgut zur Nutzung bereitgestellt (ausgehoben) und ggf. zuvor von den Archiv-Mitarbeitern gesichtet und aufbereitet werden muss, bevor es dem Nutzer vorgelegt werden kann. Im erforderlichen *Nutzungsantrag* sollte das Forschungsvorhaben knapp geschildert werden, wobei insbesondere darzustellen ist, in welcher Form eine Veröffentlichung der Forschungsergebnisse beabsichtigt ist und inwieweit später die Reproduktion des Archivgutes (z.B. die Verwendung von Fotos aus dem Archiv in der geplanten Publikation) gewünscht ist.

Viele Archive machen hier Einschränkungen bzw. Auflagen gemäß ihren Bild- bzw. Urheberrechten und dem Datenschutz, insbesondere wenn es sich um sensible, personenbezogene Unterlagen (z.B. Personalakten von Verwaltungsorganisationen oder private Korrespondenz lebender Personen) handelt. Zu Zwecken wissenschaftlicher Forschung können hier Ausnahmeregelungen getroffen werden, solange den Anforderungen der allgemeinen Datenschutzauflagen entsprochen wird. In der Regel ist eine Anonymisierung bzw. Pseudonymisierung personenbezogener Unterlagen zu gewährleisten. Insofern die Anonymisierung nicht bereits von der aktenausgebenden Stelle vorgenommen wurde, hat der Forscher dafür Sorge zu leisten, dass die Persönlichkeitsrechte der in den Dokumenten erwähnten Personen gewahrt bleiben. Dies kann einerseits durch eine gründliche Unkenntlichmachung von Namen erfolgen, andererseits kann es erforderlich sein, im Text

auf die Schilderung charakteristischer privater Details zu verzichten, die eine Identifikation der entsprechenden Personen durch den Leser ermöglichen würde. Im Zweifelsfall ist vor der Veröffentlichung eine Rücksprache mit den Archivmitarbeitern bzw. eine Beratung mit anderen wissenschaftlichen Experten geboten.

Eine wichtige Ausnahme von der grundsätzlichen Anonymisierungspflicht bei der Veröffentlichung personenbezogener Unterlagen ergibt sich in Bezug auf sogenannte „Personen der Zeitgeschichte“. Darunter werden Personen verstanden, die aufgrund ihrer herausragenden gesellschaftlichen Bedeutung im Fokus öffentlichen Interesses und medialer Aufmerksamkeit stehen, wie z.B. Politiker oder prominente Künstler. Erstens zeigt diese Definition jedoch, dass es interpretationsbedürftige Grenzfälle geben kann, und zweitens sind in jedem Fall der Schutz der Privat- und Intimsphäre der betreffenden Personen zu wahren. Im Sinne der wissenschaftlichen Forschung und der Information der Öffentlichkeit ergibt sich prinzipiell jedoch ein ausreichender Spielraum für die wissenschaftliche Verwertung personenbezogener Dokumente, wenn die darin dargestellten Sachverhalte von öffentlichem Belang sind.

Viele Archive bieten ihren Nutzern die grundsätzliche Möglichkeit, sich im Anschluss an die Akteneinsicht *Fotokopien von Auszügen des Archivgutes* erstellen und diese zuschicken zu lassen. Dies hat den Vorteil, dass eine intensive Dokumentenanalyse nach einer ersten Sichtung des Materials auch unabhängig vom Archivstandort erfolgen kann. In jedem Fall muss in einer Publikation, die sich auf Archivgut stützt, ein sorgfältiger und vollständiger Nachweis der verwendeten Quellen erfolgen (Angabe des Archivs sowie der Signaturen, d.h. der archivinternen Bestandskürzel und Bestellnummern zur Dokumentenidentifikation).

---

## 62.4 Auswahl von Dokumenten

Während man üblicherweise kleine Dokumentenbestände insofern hypothesen- und materialgeleitet sichtet, dass man während der Einsicht die Auswahl weiterer Teilbestände und Einzeldokumente in Abhängigkeit von deren Relevanz für die Forschungsfrage trifft (Akremi, Kapitel 17 in diesem Band), kann bei größeren Beständen, z.B. bei alphabetischen Personenregistern, eine systematische Zufallsauswahl (random sampling) der Dokumente notwendig sein. Dabei können Verfahren der Stichprobenziehung zum Einsatz kommen, die denen der schriftlichen oder mündlichen Befragung (Häder/Häder, Kapitel 18 in diesem Band) ähneln. So könnte beispielsweise eine bestimmte Anzahl von Akten festgelegt werden, die jedem n-ten Aktenbehälter (Karteikasten o.ä.) zu entnehmen ist, bis eine ausreichende Stichprobengröße erreicht worden ist (vgl. Buchholz 2002).

## 62.5 Analysestrategien und Anwendungsmöglichkeiten

### 62.5.1 Methoden-Mix

Bei der sozialwissenschaftlichen Dokumentenanalyse gilt es, je nach der Beschaffenheit der Dokumente und dem jeweiligen Erkenntnisinteresse entsprechend, etablierte Verfahren der qualitativen und quantitativen Sozialforschung und anderen Sozial- und Geisteswissenschaften zu adaptieren. Sowohl die ausschließliche Analyse von Dokumenten als auch die Verschränkung der Dokumentenanalyse mit anderen Forschungsstrategien (Methoden-Mix, Kelle, Kapitel 8 in diesem Band) sind möglich.

### 62.5.2 Qualitative Inhaltsanalyse

So kann eine qualitative Inhaltsanalyse (Mayring/Fenzl, Kapitel 38 in diesem Band) von schriftlichen Zeugnissen aus dem zu untersuchenden sozialen Gegenstandsbereich als unterstützendes, ergänzendes Verfahren im Zusammenspiel mit der qualitativen oder quantitativen Analyse von Befragungsdaten (Reinecke und Helfferich, Kapitel 39 und 44 in diesem Band) oder Beobachtungsdaten (Thierbach/Petschick, Kapitel 66 in diesem Band) angewandt werden, um sich in der Phase der Exploration und Hypothesengenerierung den Forschungsgegenstand näher zu erschließen oder um Fallbeispiele (Hering/Schmidt, Kapitel 37 in diesem Band) zu extrahieren, die die Grundlage für weitere Forschungstätigkeit bilden. Hier wäre auch eine Kombination ganz unterschiedlicher Dokumenten- bzw. Texttypen denkbar. Beispielsweise könnten für stadtsoziologische Untersuchungen (Dangschat, Kapitel 76 in diesem Band) zur räumlichen Konzentration bzw. Segregation sozialer Gruppen im Vorfeld oder als Ergänzung zu einer repräsentativen Befragung auch öffentliche Broschüren der Stadt, Verwaltungsstatistiken zur Soziodemographie der Bewohner (Auszüge aus dem Einwohnermelderegister), Telefon- und Adressbücher, Kartenmaterial des Katasteramtes zur städtischen Siedlungsstruktur und Raumplanung (Lakes, Kapitel 77 in diesem Band) sowie die Regionalpresse und Stadtteilzeitungen gesichtet und quantitativ oder qualitativ ausgewertet werden.

Ein anderes Beispiel für die Dokumentenanalyse von Publikationen bietet eine qualitative Inhaltsanalyse von Populärliteratur mit dem Ziel, den Wandel gesellschaftlicher Diskurse und kultureller Normen (u.a. von Geschlechterverhältnissen und Rollenvorstellungen) zu untersuchen (Ernst 2009).

### 62.5.3 Biographieforschung und sozialwissenschaftliche Hermeneutik

Für mikrosoziologische Einzelfallstudien (Hering/Schmidt, Kapitel 37 in diesem Band), z.B. der Familien- oder Biographieforschung (Rosenthal, Kapitel 35 in diesem Band), ist denkbar, in einem mehrstufigen Prozess (im Sinne eines „theoretical samplings“ im

Rahmen der Grounded Theory; Strübing, Kapitel 32 in diesem Band) (Ego-)Dokumente (Ernst, Kapitel 63 in diesem Band) und Transkripte narrativer Interviews (Küsters, Kapitel 40 in diesem Band) miteinander zu vergleichen und dabei bereits gewonnene Befunde zu überprüfen und weiterzuentwickeln. Dem Forschungsgegenstand und dem korrespondierenden Datenmaterial gemäß ergibt sich ein *interpretativer* und *verstehender* Zugang zu den Dokumenten im Mittelpunkt (Kurt/Herbrik, Kapitel 33 in diesem Band). Bei einer biographischen Fallrekonstruktion können Ego-Dokumente zur Familiengeschichte (Briefe, Tagebücher) helfen, das in narrativen Interviews Geschilderte einzuordnen, oder aus ihnen lassen sich wertvolle Hinweise für weitere Interviews extrahieren. Die Sequenzanalyse (Kurt/Herbrik, Kapitel 33 in diesem Band) eines Briefes oder Tagebuchauszugs etwa erfolgt dabei ähnlich der eines Interviews.

#### **62.5.4 Quantitative Analysestrategien**

Quantitative Analysestrategien für Dokumente, die in der methodischen Tradition indikatorengestützter empirischer Sozialforschung (Burzan und Huinink, Kapitel 81 und 82 in diesem Band), stehen, bieten sich im Fall „von in großen Mengen vorliegenden Quellengruppen wie zum Beispiel Personalakten, Steuerlisten und Kirchenbücher“ (Best/Mann 1977: 1) an.

Bei maschinengeschriebenen Akten (Hartmann/Lengerer, Kapitel 71 in diesem Band) sowie Büchern, Zeitungen (Klein, Kapitel 63 in diesem Band) und anderen gedruckten Dokumenten kann beispielsweise eine *computerunterstützte Inhaltsanalyse* (CUI, Rössler/Geise 2013) stattfinden, insofern sich das Datenmaterial digitalisieren, d.h. scannen und mit automatischen Texterkennungsprogrammen auslesen lässt (Kuckartz/Rädiker, Kapitel 27 in diesem Band).

Enthalten Dokumente eine Vielzahl von (numerischen) Informationen, die homogenen Erfassungsmustern folgen, also Daten in Tabellen, Listen oder Formularen, eignen sich diese besonders für quantitative Analysen. So erlauben etwa die Geburts-, Tauf- und Sterbelisten der Kirchengemeinden mit ihren personenbezogenen Angaben zu Lebensdaten, sozialer und geographischer Herkunft, dem Beruf, dem gesellschaftlichen Status usw. die Rekonstruktion lokaler Sozialstruktrentwicklung eine *Verlaufsdatenanalyse* (Pötter/Prein, Kapitel 86 in diesem Band), sind doch die Veränderungen der Mortalitäts- und Geburtenraten und des Heiratsalters aussagekräftige Indikatoren langfristigen sozialen bzw. kulturellen Wandels (Best/Mann 1977).

Historische Personalakten (die u.a. auch Ego-Dokumente wie hand- oder maschinen-schriftliche Lebensläufe enthalten) können insbesondere bei erheblichem Arbeitsaufwand der Aktenrecherche, Kategorienbildung, Vercodung (Züll/Menold, Kapitel 52 in diesem Band), (manuellen) Datenerfassung und -aufbereitung zu statistischen Datensätzen aggregiert werden, die dann beispielsweise die multivariate Analyse von Kollektivbiographien in der empirischen Elitenforschung ermöglichen (Best/Hornbostel 1998). Die *Aggregation*

*von biographischen Individualdaten* aus Dokumenten (Graeff, Kapitel 72 in diesem Band) bedarf einer intensiven Informationsselektion und -reduktion. Die Datenerhebung kann besonders dann sehr aufwändig sein, wenn Daten zu Einzelpersonen aus verschiedenen Quellen fusioniert, ergänzt und abgeglichen werden müssen (ebd.) und wenn es sich teilweise um handschriftliche Dokumente handelt (Best 1977).

## Literatur

- Arathymou, Spyridoula (2009): Finding and Accessing the Right Archive and Archival Data. In: HSR 34 (3): 71-77
- Baur, Nina (2009): Measurement and Selection Bias in Longitudinal Data. In: HSR 34 (3): 9-50
- Beck, Friedrich/Henning, Eckart (Hg.) (2003): Die archivalischen Quellen. Köln et al.: Böhlau
- Best, Heinrich (1977): Die quantitative Analyse inhaltlicher und kontextueller Merkmale historischer Dokumente. In: Best, Heinrich/Mann, Reinhard (Hg.): Quantitative Methoden in der historisch-sozialwissenschaftlichen Forschung. Stuttgart: Klett-Cotta. 162-205
- Best, Heinrich/Hornbostel, Stefan (1998): Prozeß-produzierte Daten als empirisches Material für eine Soziologie des realen Sozialismus. In: GESIS (Hg.): Materialien zur Erforschung der DDR-Gesellschaft. Opladen: Leske + Budrich. 201-221
- Bick, Wolfgang/Paul J. Müller (1984): Sozialwissenschaftliche Datenkunde für prozeß-produzierte Daten. In: Müller, Paul J./Mann, Reinhard (Hg.): Sozialforschung und Verwaltungsdaten. Stuttgart: Klett-Cotta. 123-159
- Buchholz, Matthias (2002): Stichprobenverfahren bei massenhaft gleichförmigen Einzelfallakten. In: HSR 27 (2/3): 100-223
- Ernst, Stefanie (2009): Using Qualitative Content Analysis of Popular Literature for Uncovering Long-Term Social Processes. In: HSR 34 (1): 252-269
- Früh, Werner (2011): Inhaltsanalyse: Theorie und Praxis. 7., überarb. Auflage. Konstanz et al.: UVK-Verlags-Gesellschaft
- Heckmann, Friedrich (1992): Interpretationsregeln zur Auswertung qualitativer Interviews und sozialwissenschaftlich relevanter „Texte“. In: Hoffmeyer-Zlotnik, Jürgen H. P. (Hg.): Analyse verbaler Daten. Opladen: Westdeutscher Verlag. 142-167
- Hitzler, Ronald/Honer, Anne (Hg.) (1997): Sozialwissenschaftliche Hermeneutik. Opladen: Leske + Budrich UTB
- Howell, Martha/Prevenier, Walter (2004): Werkstatt des Historikers. Köln/Weimar/Wien: Böhlau UTB
- Müller, Paul J. (Hg.) (1977): Die Analyse prozeß-produzierter Daten. Stuttgart: Klett-Cotta
- Prior, Lindsay (2003): Using Documents in Social Research. London: Sage
- Rössler, Patrick/Geise, Stephanie (2013): Standardisierte Inhaltsanalyse. In: Möhring, Wiebke/Schlütz, Daniela (Hg.): Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft. Wiesbaden: VS Verlag. 269-287
- Rusinek, Bernd-A./Ackermann, Volker/Engelbrecht, Jörg (Hg.) (1992): Einführung in die Interpretation historischer Quellen: Schwerpunkt: Neuzeit. Paderborn et al.: Schöningh
- Salheiser, Axel (2009): Handling Ideological Bias and Shifting Validity of Longitudinal Data. In: HSR 34 (1): 197-210

- Scheuch, Erwin K. (1977): Die wechselnde Datenbasis der Soziologie – Zur Interaktion zwischen Theorie und Empirie. In: Müller, Paul J. (Hg.): Die Analyse prozeß-produzierter Daten. Stuttgart: Klett-Cotta. 5-41
- Weyrauch, Erdmann (1977): Datenverarbeitung als Quellenkritik? In: Müller, Paul J. (Hg.): Die Analyse prozeß-produzierter Daten. Stuttgart: Klett-Cotta. 141-197
- Webb, Eugene J./Campbell, Donald T./Schwartz, Richard D./Sechrest, Lee (1975): Nichtreaktive Messverfahren. Weinheim und Basel: Beltz

Stefanie Ernst

---

## 63.1 Was sind literarische Quellen und persönliche Dokumente?

Zumeist verwendet die Soziologie Interviews (Helfferich und Küsters, Kapitel 39 und 40 in diesem Band), Beobachtungen (Thierbach/Petschick, Kapitel 66 in diesem Band) oder Gruppendiskussionen (Vogl, Kapitel 41 in diesem Band), um zu untersuchen, wie Menschen bestimmte Lebensbedingungen oder Herausforderungen bewältigen. Die gesellschaftliche Wirklichkeit bietet aber noch eine weitere spezifische Quelle unmittelbarer Art an: literarische Quellen auf der einen und persönliche Dokumente auf der anderen Seite. Beide Materialsorten bieten einmalige Einblicke in die habituelle, emotionale und psychische Dimension von Lebenspraxis, die sich in kurz-, mittel- und langfristiger Perspektive ausdrückt. Gesellschaftliche Auseinandersetzungen, soziale Bewegungen und individuelle Formen der Alltagsbewältigung können in diesen spezifischen Dokumenten gut analysiert werden.

Diese Materialien besitzen damit eine *kollektive Komponente* (etwa Gruppengefühle) und eine *subjektive Komponente*. Literarische Quellen und persönliche Dokumente werden vom Sozialforscher nicht erst produziert bzw. erhoben. Zudem unterliegen sie unterschiedlichen künstlerischen Ansprüchen, Gattungen und Stilen sowie Zielsetzungen. Damit können nicht nur Konstruktionen sozialer Lebensrealität eingefangen werden, die retrospektiv angelegt sind und mehr als einen reinen statistischen Wert besitzen. Sie zeigen vielmehr, was bestimmte gesellschaftliche Ereignisse für den Einzelnen oder Gruppen bedeuten bzw. bedeutet haben.

Man denke etwa an die Stilisierung des Nationalgefühls im Zuge der Nationalstaatsbildung im 19. Jahrhundert und in den beiden Weltkriegen. Briefe und Tagebuchaufzeichnungen der Soldaten im Feld, der Daheimgebliebenen oder Kriegsgefangenen weisen dieses Gefühl ebenso auf wie kriegskritische oder kriegsverherrlichende Romane dieser Zeit.

Auch das zeitgenössische Massenphänomen Fußball am Ende des 20. Jahrhunderts, das Emotionen und sportlichen Wettkampf par excellence ausdrückt, lässt einen spezifischen

Habitus, spezifische Emotionen und wechselnde gesellschaftliche Trägergruppen erkennen. Fußball ist inzwischen vom wenig regulierten Arbeitersport und Freizeitvergnügen des 19. Jahrhunderts zum millionenschweren, mit entwickelten Spiel- und Fairnessregeln versehenen Geschäft von Profikickern geworden, das über alle Alters- und Geschlechtsgruppen sowie Milieus reicht. Der Habitus unbedingter Loyalität mit dem Verein und das Entstehen einzigartiger Gruppengefühle können dabei aber ganz besonders gut in literarischem Material verdeutlicht und aufgespürt werden.

Obgleich in der Wissenschaft oft Überschneidungen von Literatur und persönlichen Dokumenten und Grenzgänger anzutreffen sind (etwa in der Tagebuch- und Briefliteratur), sollen im Folgenden die unterschiedlichen Textgattungen hinsichtlich des fiktionalen Gehalts, der Zielsetzung und Verfasserschaft sowie des Authentizitätsanspruch voneinander abgegrenzt werden.

### 63.1.1 Literatur

Literatur als *Unterhaltungsmedium, Ort der Reflexion und Kritik* bewegt sich zwischen den gesellschaftlichen Bewertungshorizonten von Populär- und Hochkultur und ist auch wissenschaftsimmanenter Kriterien unterworfen. Literatur soziologisch zu verwenden, kann soziologische Erkenntnisse liefern, die auf andere Weise kaum greifbar sind.

Will man Literatur als Datenquelle soziologisch nutzen, ist zunächst zu bedenken, dass es sich im Allgemeinen mehr noch als bei persönlichen Dokumenten um *fiktive Erzählungen* handelt, die auch spezifischen künstlerischen Ansprüchen unterliegen. Die sozialwissenschaftliche Auswertung von Literatur als einem bestimmten zentralen Orientierungs-, Kommunikations- und Unterhaltungsmedium der Gesellschaft ist dabei in die *Prämissen* eingebettet, dass sich in den Aussagen von Literatur auch Machtverhältnisse zwischen sozialen Gruppen mit ihren besonderen Regeln, Normen und Werten ausdrücken. Dabei geht es über die *symbolische Ebene* hinaus um die *Funktion*, um den *spezifischen sozialen Eigensinn von Vorstellungen im Prozess gesellschaftlicher Entwicklungsschübe*.

Kuzmics und Motzic (2003) zeigen zum Beispiel anhand der Analyse von Romanen, wie Patriotismus und Kriegsbegeisterung im 19. oder Fankulturen im 21. Jahrhundert funktionieren und ein Wir-Gefühl erzeugen, das loyalitäts- und identitätsstiftend ist. Die Autoren interpretieren dabei den Roman „Fever Pitch“ von Nick Hornby auf der einen und Leo Tolstois „Anna Karenina“ auf der anderen Seite. So wird die aktuelle Konstruktion des Fußballkults und der Gruppenzugehörigkeit bei Hornby eindringlich reflektiert, während ländliche Arbeits- und Lebenswelten, Körpererfahrungen und Schamgefühle im Zeitalter der Leibeigenschaft im Russland des ausgehenden 19. Jahrhunderts zum Ausdruck kommen. Tolstoi verdeutlicht den epochalen Wandel des russischen Zarenreichs, von dem der Adel, das Bürgertum und der Bauernstand zwischen Leibeigenschaft und Freiheitsbewegungen gleichermaßen, aber doch unterschiedlich betroffen sind. Gesellschaftliche Norm- und Wertvorstellungen sowie Tabus lassen sich hier unmittelbar rekapitulieren und sind mehr als reine Illustrationen gesellschaftlichen Wandels. Lebensweltliche

Wirklichkeiten einer Generation, eines Milieus oder gar einer ganzen Epoche drücken sich also auch in Literatur aus. Damit ist sie für Längsschnittanalysen ebenso geeignet wie für Gegenwartanalysen.

Literarische Quellen sind weiterhin aufgrund ihrer *Stilistik, Gattung und Zielsetzung* zu differenzieren. So besteht ein Unterschied, ob es sich um Lyrik, Belletristik, Prosa oder konventionelle, populäre und künstlerisch anspruchsvolle, gesellschaftskritische Literatur handelt. So formuliert etwa Ratgeberliteratur explizit das Ziel, „erbauliche“ unterweisende und hilfreiche Hinweise zur erfolgreichen Lebensführung zu geben, während manch ein Roman wie etwa „Im Westen nichts Neues“ von Erich Maria Remarque sicherlich andere Intentionen verfolgt. Auch Theodor Fontanes „Effi Briest“, Gustave Flauberts „Madame Bovary“ oder Emil Zolas „Nana“ erteilen weniger gefällige, erbauliche Ratschläge, sondern problematisieren vielmehr die Enge bürgerlich-moralischer Normvorstellungen und Machtverhältnisse für die weitestgehend rechtlosen Frauen in ihrer Zeit. Ratgeberliteratur als spezifische Literaturform gewährt mit ihrem direkten, normativen Duktus dagegen besonders in Zeiten des gesellschaftlichen Umbruchs und der sozialen und räumlichen Mobilität Einblicke in die vielfältigen Antwortmöglichkeiten und Orientierungen über erwünschte Verhaltensideale.

### **63.1.2 Persönliche Dokumente bzw. Ego-Dokumente (Briefe, Tagebücher, Autobiografien)**

Persönliche Dokumente sind *demgegenüber schriftlich fixierte (Lebens-)Geschichten und Erfahrungen von gesellschaftlicher Wirklichkeit* u.a. in Form des Tagebuchs, der Lebensbeichte und des Briefes sowie der Autobiografie. Eine besondere Textgattung ist dabei zunächst die *Autobiografie*, die neben *Briefen* und *Tagebüchern* unter dem Begriff der „*Ego-Dokumente*“ (Schulze 1996) oder „*personal documents*“ firmiert.

Ego-Dokumente stehen als prozessproduzierte Daten in einen interdisziplinären Forschungs- und Methodenzusammenhang, der die Geschichts-, Literatur- und Sozialwissenschaften gleichermaßen interessiert und unterschiedliche Etappen markiert. Neben der Mentalitäts- und Sozialgeschichte hat etwa auch die Sozialgeschichte Interesse an persönlichen Dokumenten und Aufzeichnungen gefunden. Besonders seit der westeuropäischen Frühneuzeit und ersten Individualisierungsprozessen in der Renaissance liefern diese besonderen Quellen wichtige Beiträge zur „*Erforschung der menschlichen Seele und der Affekte*“ (Schulze 1996: 19). Eine zeitgenössische Lesart könnte vermuten lassen, dass Tagebücher, Autobiografien und Briefe einen bestimmten Ausschnitt persönlich erlebter Geschichte gewissermaßen „authentisch“ darstellen. Dies macht zugleich auch ihre hohe Popularität aus. Menschen reflektieren in subjektiver Perspektive unmittelbar eigenes Erlebtes oder Beobachtetes in ihrer Zeit.

Dabei sind Ego-Dokumente aber in ihren Ursprüngen nicht maßgeblich aus Gründen der Selbstreflexion oder -stilisierung und des Räsonierens über ein vermeintlich erfolgreiches oder beschwerliches Leben abgefasst worden. Vielmehr treten bereits im Zeitalter

der Inquisition, im Ringen um säkulare oder religiöse Welt- und Menschenbilder besonders Tagebücher und Autobiografien als *Rechtfertigungs-, Unterweisungs- und Erziehungsdokumente* auf, die etwa in Gerichtsverfahren verwendet werden, um die Läuterung des vermeintlichen „Delinquenten“ nachzuweisen, der sein „lasterhaftes“ Leben selbstkritisch reflektiert und rekapituliert.

Damit verdichten Autobiografien, Tagebücher und Memoiren als *Alltagsquellen* zwar Geschichte; sie verwandeln sie aber durchaus auch in erfahrene Geschichte von literarischem Wert und damit zu einer Quelle literaturwissenschaftlicher Disziplinen. Viele Anspielungen, Prozesse und Werthaltungen sind dabei als Produkt einer individuellen, aber auch kollektiven Meinung zu entschlüsseln. Der Autor selbst wie sein Zielpublikum stehen mithin in gesellschaftlichen Bezügen, die interpretiert und re-kontextualisiert werden wollen.

Besonders mit der Weiterentwicklung von Ego-Dokumenten im 19. Jahrhundert ist eine spezifische „*Gattungspoetik*“ (Scheutz 2006: 12) zu bemerken, die besonders nach der Authentizität und Verfasserschaft der Texte fragen lässt. Nicht nur werden mit der Erfindung des modernen Menschen *Einblicke in persönliche Beziehungsgeflechte und die Lebensführung* gegeben, sondern auch *Konflikte, Emotionen und Prozesse der Subjektwerdung und Selbstbehauptung* erkennbar. Darüber hinaus zeigen sich *gesellschaftliche Normalitätsvorstellungen* in autobiografischem Material, die sich mit dem gesellschaftlichen Wandel entsprechend verändern. Besonders gesellschaftliche Außenseiter mit außergewöhnlichen Sozialerfahrungen werden daher für die biografische Forschung (Rosenthal, Kapitel 35 in diesem Band) zum expliziten Thema und stehen häufig im Mittelpunkt der Texte.

In den 1920er Jahren sozialwissenschaftlich erschlossen und aus einer hermeneutischen Tradition (Kurt/Herbrik, Kapitel 33 in diesem Band) kommend, wird die Einzigartigkeit der *Autobiografie* schließlich seit den 1960er-Jahren zusehends hinterfragt. Sie ist mithin zum Feld politischer und sozialer Kämpfe geworden. Inzwischen werde der Blick jedoch mehr auf die „Professionalisierung mündlich erzählter Lebensgeschichten“ in Form von Interviews geworfen, was sich zuungunsten persönlicher Dokumente wie Tagebücher und Briefe und der „öffentlichtkeitsadressierten Kommunikationsform“ der Autobiografie auswirke (Heinze 2010: 215).

Ganz besonders *Tagebücher*, die ihrem Zweck nach nicht nur zeitgenössisches Erleben versammeln, sondern explizit geheime Dokumente symbolisieren, lassen persönliche Wünsche, Sehnsüchte und Ängste sowie innere Dialoge und Hoffnungen erkennen. Daher wird ihnen ein besonders hoher Authentizitätsgrad zugeschrieben. Im Unterschied zu oft politisch ambitionierten und beauftragten Memoiren herausragender Persönlichkeiten des öffentlichen Lebens sind Tagebücher zudem über alle gesellschaftlichen Schichten hinweg ein Medium der Selbstauskunft und des Selbstausdrucks (Hämmerle 2006: 28ff., Müller 2006: 81).

Speziell die Gattung der *Mädchen Tagebücher* und der *Arbeiterliteratur* problematisiert diese quer zur ‚Herrschaftsgeschichte‘ liegende Perspektive auf die gesellschaftliche Wirklichkeit mit ihrem eindimensionalen Konzept bürgerlich-männlicher, moderner Subjektvorstellungen. Man gewinnt weiterhin Einblicke in die Selbst- bzw. Ich-Entwürfe

der Autorinnen und Autoren jenseits klassischer, homogener und autonomer Subjektvorstellungen. Gerade die Brüchigkeit und Vieldeutigkeit dieser Selbstzeugnisse liefert neue Erkenntnisse.

Müller (2006) spricht vor diesem vielschichtigen Hintergrund lebensgeschichtlichen Schreibens daher von „popularer Autobiographik“, während Heinze (2010: 204) dagegen besondere „Deutungseliten“ im autobiografischen Schreiben erkennt. Gemeint sind mit „popular“ *Briefe, Tagebuchaufzeichnungen und autobiografische Texte bildungsferner Schichten mit nicht-professioneller Autorenschaft*. Sie bewegen sich damit gemeinsam mit der Arbeiter- oder auch populären Ratgeberliteratur (Ernst 2009) gerade *jenseits* bürgerlicher oder adeliger Autobiographien und klassischer literaturwissenschaftlicher Quellen. Mehr als über den Aspekt hinaus, Nähe, Intimität und Information zu vereinen, liefern sie Einblicke in die „konkreten Lebens- und Beziehungssituationen“ (Müller 2006: 81) der Verfasser und damit in die sozialwissenschaftlich relevanten Vergesellschaftungsprozesse mit ihren biographischen Einschreibungen.

---

### **63.2 Besonderheit und Nutzen literarischer Quellen und persönlicher Dokumente**

Mit Müller (2006), Fuchs-Heinritz (2009) und besonders Heinze (2010) sollen im Folgenden die Besonderheit und Disparität von literarischen und Ego-Dokumenten speziell für die Sozialwissenschaften weiter differenziert werden. Persönliche Dokumente und literarische Quellen sind zunächst Texte über Sachverhalte oder Vorgänge, die weder aktenförmig (Salheiser, Kapitel 62 in diesem Band) noch anderweitig formal abgefasst oder standarisierbar sind. Dieses besondere Datenmaterial kann vielmehr als ein Ausgangsmaterial gesehen werden, das nicht erst durch den Forscher produziert oder erhoben werden muss, sondern sich in der untersuchten gesellschaftlichen Wirklichkeit bereits vorfinden lässt.

Zu unterscheiden sind als Material *biografischer* Forschung (Fuchs-Heinritz 2009: 32) dabei *Memoiren* als neutral anmutende Zeitgeschichtsdarstellungen, ohne das seelische Eigenleben des Verfassers zu dokumentieren, während *Tagebücher* als regelmäßige, unmittelbare Selbstbeobachtung und Gewissenserforschung gelten und sich Briefe als besondere Gattungsform sowie als Kommunikationsmittel der Neuzeit darstellen. Ego-Dokumente wecken daher neben der Sozialgeschichte und Literaturwissenschaft besonders das Interesse einer an *subjektiven Deutungsweisen und Bewältigungsformen* interessierten qualitativen Sozialforschung. Sie können, wie auch literarische Zeugnisse, vielmehr zum „ausagekräftigen ‚Zeugnis‘ des zivilisatorischen Wandels“ (Wild 1996: 195) werden. Diese besondere Literatur- und Textgattung ist zugleich als ein „auf Erkenntnis ausgehender und autonomer Diskurs – ‚Reflexion‘ dieser Lebenspraxis“ (Wild 1996: 196).

Mit der zusehenden Fokussierung auf *mündlich erzählte Lebensgeschichten* (Küsters, Kapitel 40 in diesem Band) können allerdings Leerstellen über die Entstehung und Einbettung autobiografischer Texte entstehen. Dabei ist diese spezifische Kommunikationsform weder nur rein illustrativ oder exemplarisch für persönliche Lebensläufe zu sehen,

sondern in sich sozialkommunikativ besonders reizvoll, weil die Verfasserschaft vielfältig und widersprüchlich ausgestaltet ist: als quasi mehrfacher „Rollenträger“ vom reflektierenden Subjekt zum produzierendem Autor und Protagonisten muss im Erzählen eine Identität geformt werden.

Im Gegensatz zum unterschiedlichen Entstehungskontext narrativer Face-to-Face Interviews zeichnet die besondere Intersubjektivität und Diskursivität diese Textgattung aus. Autobiografien und persönliche Dokumente liefern Informationen über persönliche Lebenswege und Schicksale in zeitlicher, rekonstruktiver Perspektive und verweisen auf Generationsbeziehungen, Milieus und Mentalitäten, was eine besondere Erzählstruktur und Diktion beinhaltet. Der hohe Erfolg und die Popularität von Autobiografien machen sie zudem als Vermittlungsform von Wirklichkeit nicht nur soziologisch relevant, sondern methodisch unterschiedlich interpretierbar. Mehr als durch narrative oder rein literaturwissenschaftliche Methodenzügänge gegeben, kann eine sozialkommunikative und biografiewissenschaftliche Sicht auf Autobiografien, Erkenntnisse über gesellschaftliche Wissens- und Deutungsformen liefern. Teils mit elitärem, borniertem oder belehrendem Zug ausgestattet, reklamieren diese persönlichen Dokumente eine Deutungshoheit über den jetzigen gesellschaftlichen Kontext.

Wenn sich bereits die Analyse autobiografischen Textmaterials als besondere Herausforderung für eine *prozessorientierte Sozialforschung* schlechthin darstellt, so scheint auch die Auswertung *literaturhistorischer* Dokumente für die Soziologie nicht sofort ersichtlich. Das auszuwertende Textmaterial basiert zum einen nicht auf der Wechselseitigkeit zwischen Frager und Befragtem. Zum anderen verweist diese besondere Textgattung eher auf Analysetechniken der Literaturwissenschaft.

---

### 63.3 Stärken und Schwächen

Diese unterschiedliche Lesart autobiografischen und literarischen Materials deutet bereits auf seine besonderen Stärken und Schwächen für sozialwissenschaftliche Verwendungszusammenhänge hin. Häufig werden Ego-Dokumente vor allem hinsichtlich ihrer Authentizität befragt. Besonders angesichts der oben ausgeführten, unterschiedlich motivierten Verfasserschaft, etwa in Form der Lebensbeichte oder als beauftragte Memoiren sind hier mehr oder minder sozial erwünschte, stilisierte oder aber auch erzwungene Selbstauskünfte vorzufinden. Der methodologische Nutzen dieser Texte liegt aber auch für eine mittel- oder langfristig angelegte soziologische Analyse darin, bestehende Diskurskonstruktionen (Traue et al., Kapitel 34 in diesem Band) bestimmter Themen in ihrer Genese aufzuspüren zu können. Sie bereichert unser Wissen etwa zum Thema Geschlechterbeziehungen (Ernst 2009) neben Experteninterviews, teilnehmenden Beobachtungen und statistischen Erhebungen um die Dimension ihres historischen Gewordenseins. Zudem zeigt gerade der Blick auf explizit normative Texte das vorreflexive, unbewusste und zeitgenössische (Geschlechter-)Wissen, bei dem sich jeder als Experte des Alltags sieht. Was in offiziellen Verlautbarungen und geplanten Interviews kaum über Standards gewünschten

Verhaltens zu ermitteln ist, können diese Texte liefern, ohne dass man in konkrete Interaktionen mit den Befragten involviert ist.

Liest man literarische und Ego-Dokumente lediglich als „Herrschaftsgeschichte“ mit „bornierter Deutungshoheit“ oder gerade konträr als „Geschichte von unten“ mit vermeintlich hohem Echtheitsgrad, dann wird man zu unterschiedlichen Interpretationsweisen und Schlussfolgerungen über die *dargestellte* Lebensgeschichte kommen. Daher sind diese Texte nicht als wahrhafte Augen- oder Zeitzeugenberichte misszuverstehen. So gilt auch für geschriebene Lebensgeschichte, das, was Rosenthal schon mit Bezug auf „erzählte Geschichte“ festhält: verharrt man bei der traditionellen Quellenkunde in methodischen Prüfverfahren von „wahr“ vs. „unwahr“ im „Sinne von tatsächlich Stattgefundenen und Erfundenem“ (1995: 14), vernachlässigt man den Blick auf Bedeutungsräume, Verweisungen und Kontextdaten. Und ganz besonders vor dem Hintergrund, dass im Gegensatz zu beobachtbaren Face-to-Face- Interaktionen bei retrospektiv verfassten Texten keine Möglichkeit der intersubjektiven Überprüfbarkeit besteht, zeigen sich die Schwächen dieses Materials.

Werden Briefe, Tagebuchaufzeichnungen und Autobiografien jedoch lediglich als mehr oder minder authentische Narrationen gedeutet, entzieht man das Material einer Lesart, die besonders auf den sozialphänomenologischen Blick abhebt und seine sozial-kommunikative Komponente hervorhebt. Heinze schlägt daher vor, von der prinzipiellen Konstruiertheit autobiografischer Dokumente schlechthin auszugehen:

„Autobiografien arbeiten maßgeblich (...) an der kommunikativen Konstruktion von Wirklichkeit mit, so dass ein implizites Gattungswissen und Gattungserkennen beim Leser vorausgesetzt werden muss.“ (Heinze 2010: 202)

Anders verhält es sich mit literarischen Texten. Der Rezipient weiß im Vorhinein, dass es sich bei diesem Textmaterial um fiktive, gleichwohl aber gesellschaftlich zu verortende Texte handelt. Die kommunikative Situation basiert zum einen, wie autobiografische Zeugnisse, nicht auf der Wechselseitigkeit zwischen Frager und Befragtem. Zum anderen verweist diese Textgattung eher auf die Literaturwissenschaft, mit ihren spezifischen Analysetechniken wie etwa die Diskursanalyse oder die Soziolinguistik. Der methodologische Nutzen der soziologisch-historischen Inhaltsanalyse von Literatur und Ego-Dokumenten besteht jedoch darin, bestehende vorherrschende Konstruktionen zeitgenössischer Texte in ihrer langfristigen Genese und gesellschaftlichen Verflechtung aufzuspüren.

---

### 63.4 Methodologische Behandlungsweisen

Die Interpretation von persönlichen Dokumenten und literarischen Quellen ist an den Forschungsverlauf gekoppelt. Die Ebenen zwischen theoretischer Sättigung und Kontrastierung wechseln mitunter einander ab.

„Auch wenn autobiografisches Schreiben in Form von Erinnerungen, Tagebüchern, Briefen o.ä. zunächst als individuelle Vergegenwärtigung eines Lebens bzw. Lebensabschnitten zu einem bestimmten Zeitpunkt verstanden wird, so ist dem Prozess des

autobiografischen Schreibens ein mitgedachtes Gegenüber in Form eines angesprochenen, potentiellen Lesers oder im weitesten Sinne der interessierten Öffentlichkeit eingeschrieben“ (Heinze 2010: 205f.).

Latent oder explizit, wie die erforderliche Analyse von Einleitung, Vor- oder Nachwort der Autobiografien und belletristischen Literatur zeigt, wird der fiktive Leser zum Gesprächspartner, so dass durchaus eine spezifische Handlungspraxis gesehen werden kann. Alle lebensgeschichtlichen Selbstberichte fallen derzeit unter den Begriff Autobiografie, gleichgültig ob sie über narrative Interviews oder eigene Verfasserschaft erzeugt sind.

Obgleich die fiktiv beschriebenen Welten der Literatur dagegen nicht als empirisch reliabel und theoriebildend gelten, kann mit Literatur mehr geleistet werden, als Lebensgeschichten „schön“ zu illustrieren.

So hat etwa Ratgeberliteratur als populäre, nicht-künstlerische Literatur einen besonderen Gattungs- und Formwandel durchlaufen, bei dem sich zwar die Formen wie auch Inhalte im historischen Verlauf geändert haben. Das anhaltende Bedürfnis nach Orientierung, der lukrative Markt wie auch das Erfolgsversprechen zeichnen die *Sittenlehrer* des 18., die *Anstands- und Manierenbücher* des 19. wie auch die *moderne Ratgeberliteratur* des 20. und 21. Jahrhunderts aber nach wie vor aus. Die auktoriale Erzählerin, der erfahrene Freund und der fiktive Dialog, geshmückt mit vermeintlich wahren Begebenheiten, laden zur Identifikation des Lesers ein. Suggeriert wird, am gesellschaftlichen Leben teilzuhaben und sich mit einem zusehends als individuell zugerechneten Problem gut aufgehoben zu fühlen. Diese stilistische und gattungsgeschichtliche Analyse nebst der Erschließung der innerliterarischen Logik von Genres, Stilen und Moden erfordert Kenntnisse in literatur- und sozialwissenschaftlichen Methoden und Modellen. Weiterhin ist wichtig, für welche und in welcher literarischen Szene zeitgenössische Texte produziert wurden, und wer sie unter welchen Bedingungen rezipiert. So war es etwa im 18. Jahrhundert jungen bürgerlichen Frauen, die weder Zugang zu höheren Schulen und Universitäten noch zu Literaturvereinen und -cafés hatten, untersagt, höhere oder gar die neue modische Literatur der „jungen Wilden“ wie Goethe, Schiller oder Kleist zu lesen. Junge Damen wurden daher auf bildlichen Darstellungen äußerst züchtig und allenfalls mit der Bibel in der Hand abgebildet. Man befürchtete, dass die um sich greifende „weibliche Leselust“ zu moralischen Verwerfungen der Frauenwelt führte und verspottete das frivole „gelehrte Frauenzimmer“ der höfischen Salons (Ernst 2009).

Ein weiteres Beispiel für gesellschaftlichen Wandel und einen Epochenumbruch liefert das in Goethes und Kleists Werken ausgedrückte neuartige Naturverständnis: Die zeitgenössische, statische Naturbetrachtung wurde durch ihre neuartige Stilistik und Poetik gesprengt, indem sie dynamische Bewegungsverben nutzte und sich auf naturwissenschaftliche Erkenntnisse ihrer Zeit stützte (Wild 1996).

Eine soziologische Auswertung setzt sich damit auch zum Ziel, *Verdrängtes, Latentes und Neuartiges in literarischen und Ego-Dokumenten zu erschließen und in den Kontext des historischen und gesellschaftlichen Wandlungsprozesses zu setzen*. Ausgewertet werden können literarische Quellen und Ego-Dokumente z.B. mit der Inhaltsanalyse (Mayring/Fenzl,

Kapitel 38 in diesem Band), der dokumentarischen Methode oder der sozialwissenschaftlichen Hermeneutik (Kurt/Herbrik, Kapitel 33 in diesem Band).

---

### 63.5 Datenauswahl und -auswertung

Zunächst gilt es grundsätzlich, gegenstandsbezogene, spezifische Kriterien dafür zu entwickeln, welche Texte für eine soziologische Analyse ergiebig sind ab und wann die Materialsammlung abgeschlossen werden kann, wann also eine empirische Sättigung angezeigt ist. Es sollten daher Aussagen untersucht werden, die in Bruchstücken oder versteckt Informationen liefern, die die frei oder zwanghaft wahrgenommene Lebenspraxis des Einzelnen in seinen sozialen Bezügen, seiner Schicht, Region oder seinem Milieu dokumentieren. Gefühle, Wert- und Normvorstellungen wie auch Wissensbestände, Erwartungen und Erfahrungen sollten sich darin spiegeln.

Folgendes knappes Beispiel aus Theodor Fontanes „*Effi Briest*“ von 1896 soll hier zur Veranschaulichung dienen, um die Enge bürgerlicher Ehe- und Sexualnormen des 19. Jahrhunderts zu analysieren. Fontane (1819-1898) war zunächst als Apotheker tätig, wurde, engagiert in den Unruhen der 1848er-Revolution, Journalist für ein konservativ-preußisches Blatt und schließlich Schriftsteller. Er gilt als Vertreter des poetischen Realismus und verbindet häufig aus der Perspektive des auktorialen Erzählers Einzelschicksale mit Gesellschaftskritik. Damit begründete er neben Thomas Mann den deutschen Gesellschaftsroman, der mit scheinbarer Leichtigkeit, Ironie, Bilderdichte und Metaphorik arbeitet. Der im Märkischen Adel und der städtisch-bürgerlichen Oberschicht angesiedelte Roman erhielt im Erscheinungsjahr fünf Auflagen mit jeweils bis zu 3.000 Exemplaren, bis 2001 erfolgten sieben Neuauflagen.

„*Effi Briest*“ zeigt, wie im Vorfeld der Individualisierungsschübe des 20. Jahrhunderts persönliches Schicksal, gesellschaftliche Ehrvorstellungen und Geschlechtergrenzen geformt sind. Die bürgerliche Ehe legt unmittelbar und prekär eine bestimmte Form der intimen Geschlechterbeziehung dauerhaft und strukturell fest. Die Ehe galt dabei nahezu bis in die sechziger Jahre des 20. Jahrhunderts für die Frau als die einzige legitime und sozial anerkannte Form der intimen Geschlechterbeziehung. Die gute Quellenlage erlaubt es, dass man auch die zentrale Frage der Eheschließung (freiwillig oder arrangiert) historisch nachzeichnen kann. So räsoniert die 17-jährige Braut Effi Briest: Er „ist ein Mann, ein schöner Mann, ein Mann, mit dem ich Staat machen kann und aus dem was wird in der Welt“ (Fontane 1976: 42).

Die Protagonistin wird – aus der heutigen Perspektive wenig überraschend – in dieser mehr arrangierten als aus Liebe eingegangenen Ehe mit einem 21 Jahre älteren Baron zusehends unglücklich und verliebt sich in einen Offizier. Sie beginnen eine kurze Affäre, die lange unerkannt bleibt und mit dem Umzug nach Berlin endet. Jahre später entdeckt der Ehemann alte Liebesbriefe des Offiziers, den er nach so langer Zeit zum Duell fordert. Die Ehe wird geschieden, und Effi ist sich der Tragweite ihres Ehebruchs schuldhaft bewusst. Ihre Affäre wird aber, im Gegensatz zu aktuellen pikanten Liebesromanen, nicht detailliert

geschildert, sondern nur leicht angedeutet. Besonders daran, dass explizit sexuelle Szenen unerwähnt bleiben und nur noch zu erahnen sind, wird die Verdeckung von Körperlichem und Sexuellem sowie die typische Stilistik des Gesellschaftsromans deutlich. Der Druck gesellschaftlicher Normen für den einzelnen nimmt hier noch besonders tragische Züge an. Er ist so groß, dass Effies Eltern drohen:

„Du wirst einsam leben, und wenn Du das nicht willst, wahrscheinlich aus Deiner Sphäre herabsteigen müssen. Die Welt, in der Du gelebt hast, wird Dir verschlossen sein“ (Fontane 1976: 305).

Während Effi schwer erkrankt, mit 30 Jahren an der Schwinducht stirbt und der Geliebte im Duell sterben muss, führt heute ein außerehelicher „Seitensprung“ oder eine Affäre allenfalls zu einer Scheidung. Die massenhaft bei Frauen in dieser Zeit diagnostizierte „Kränklichkeit“ und Hysterie verdeutlicht dabei, dass sich den Frauen oft allein hier eine anerkannte Fluchtmöglichkeit vor den strengen gesellschaftlichen Rollenanforderungen des ehelichen Gehorsams und der Loyalität bietet. Die bürgerliche, soziale Existenz erfordert hier demnach eine eigentümliche Persönlichkeitsstruktur. Sexualität wird hier nicht nur mit der Institution der Ehe verkoppelt, sondern mehr noch der individuellen Selbstkontrolle des Einzelnen übertragen. Die Frau wird in radikalierter Form zur Adressatin des Triebverzichts und der Disziplinierung, während die soziale Wirklichkeit für den Mann Freiräume ermöglicht. Er gilt in der Ehe letztlich als sexuell überlegen und kann die um Triebverzicht und *geschlechtliche* Ehre bemühte Frau verführen (Ernst 1996).

Für den methodischen Umgang mit Ego-Dokumenten und literarischen Materialien sind Empfehlungen der literaturwissenschaftlichen Leseforschung, der Rezeptionsgeschichte und der sozialhistorischen Quellenkunde hilfreich. So sollten bei der Analyse des Materials folgende Fragen beantwortet werden:

- Welche Identität, welches Geschlecht haben der Verfasser, die Handelnden und das Zielpublikum?
- Sind Verschleierungsversuche, Ausklammerungen, Sublimierungen im Text festzustellen, und wie gestaltet sich der zeitliche Horizont der Textabfassung und der dargestellten Sachverhalte?
- Wie lässt sich die geographische, soziale und lebensphasenspezifische Lage des Textes einordnen?
- Ist der Text als ein Bekenntnis, als Legitimationsschrift oder Statement zu begreifen?
- Welche sprachlich-stilistischen Elemente wie Metaphern, Paraphrasen, Codes, Personalpronomina, Passiv-, Aktivformen, Ironisierungen etc. werden wie zum Einsatz gebracht?
- Welche Zusammenhänge stehen hinter eingebrachten Zitaten?
- Wie ist das Lese-, Wahrnehmungs- und Deutungsverhalten der Zeit ausgebildet?
- Welchen Neuigkeitswert, welche Detailtreue, Milieubezüge, Motivationen, Interessen und Absichten sind im Material erkennbar?

Zu finden sind persönliche Dokumente und literarisches Material vornehmlich in Archiven vielfältiger Art: Privat- und Staats- sowie Stiftungsarchive sind hier ebenso zu nennen wie Länder- und Stadtarchive, Antiquariate sowie thematische und fachwissenschaftliche Archive und Bibliotheken. Oft sind spezielle Handschriftenabteilungen vorhanden, die das historisch wertvolle Material gut geschützt aufbewahren. Erste Anlaufstellen sind zum Beispiel:

- *Archiv für alltägliches Erzählen*, Institut für Volkskunde, Hamburg,
- *Archiv Deutsches Gedächtnis*, FU Hagen,
- *Archiv Kindheit und Jugend*, Universität Siegen,
- *Deutsches Literaturarchiv*, Marbach,
- *Dokumentation lebensgeschichtlicher Aufzeichnungen*, Wien,
- *Helene Lange-Archiv*, Berlin,
- *Internationales Institut für Sozialgeschichte*, Amsterdam,
- *Interdisziplinäres Frauenforschungszentrum*, Bielefeld,
- *Kempowski Archiv für unpublizierte Biographien*, Nartum,
- *Nachlässe in den deutschen Archiven* (mit Ergänzungen aus anderen Beständen). (Verzeichnis der schriftlichen Nachlässe in den deutschen Archiven und Bibliotheken, Band 1, Teil I und II), bearbeitet von Wolfgang A. Mommsen.
- *Schriften des Bundesarchivs 17/I und 17/II*,
- *Schweizerisches Sozialarchiv*,
- *Stiftung Archiv der deutschen Frauenbewegung*, Kassel.

## Literatur

- Berger, Bennett M. (2000): Autobiographien. In: Hoerning, Erika M. (Hg.): Biographische Sozialisation. Stuttgart: Lucius/Lucius. 21-32
- Ernst, Stefanie (1996): Machtbeziehungen zwischen den Geschlechtern. Opladen: WDV
- Ernst, Stefanie (2009): Using Qualitative Content Analysis of Popular Literature for Uncovering Long-Term Social Processes. In: HSR 34 (1): 252-269
- Fuchs-Heinritz, Werner (2009): Biografische Forschung. Wiesbaden: VS Verlag
- Hämmerle, Christa (2006): Ein Ort für Geheimnisse? Jugendtagebücher im 19. und 20. Jahrhundert. In: Eigner, Peter/Hämmerle, Christa/Müller, Günter (Hg.): Briefe, Tagebücher, Autobiografien. Innsbruck/Wien/Bozen: Studienverlag. 28-45
- Heinze, Carsten (2010): Zum Stand und den Perspektiven der Autobiographie in der Soziologie. In: BIOS 23 (2): 201-231
- Kuzmics, Helmut/Mozetic, Gerald (2003): Literatur und sozialer Wandel. Konstanz: UVK
- Müller, Günter (2006): „Vielleicht interessiert sich mal jemand ...“ Lebensgeschichtliches Schreiben als Medium familiärer und gesellschaftlicher Überlieferung. In: Eigner, Peter/Hämmerle, Christa/Müller, Günter (Hg.): Briefe, Tagebücher, Autobiografien. Innsbruck/Wien/Bozen: Studienverlag. 76-94
- Rosenthal, Gabriele (1995): Erlebte und erzählte Lebensgeschichte. Frankfurt M.: Campus
- Scheutz, Martin/Tersch, Harald (2006): Selbstzeugnisse der Frühen Neuzeit. In: Eigner, Peter/Hämmerle, Christa/Müller, Günter (Hg.): Briefe, Tagebücher, Autobiografien. Innsbruck/Wien/Bozen: Studienverlag. 10-27
- Schulze, Winfried (1996): Ego-Dokumente. In: ibidem (Hg.): Ego-Dokumente. Berlin: Akademie Verlag. 11-32
- Wild, Reiner (1996): Literarischer Wandel und Zivilisationsprozess. In: Rehberg, Karl-Siegbert (Hg.): Norbert Elias und die Menschenwissenschaften. Frankfurt a.M.: Suhrkamp. 194-212

Harald Klein

Während die Umfrageforschung in den Sozialwissenschaften in jedem Methodenbuch umfangreich abgehandelt wird, gilt dies nicht für die Verfahren, die Texte analysieren. In diesem Beitrag geht es darum, Artikel aus Printmedien – Zeitungen und Zeitschriften – als Datenmaterial für die sozialwissenschaftliche Forschung zu erschließen.

---

## 64.1 Beschaffung

Die Beschaffung von Zeitungs- oder Zeitschriftenartikeln ist für aktuelle Ausgaben leicht, weil sie im Handel erhältlich sind. Alle anderen Ausgaben von Zeitungen oder Zeitschriften können entweder beim Verlag erworben werden oder sind in Archiven zugänglich. Hierbei kann es sich um Verlagsarchive oder auch um frei zugängliche Archive handeln. Das Zeitungsarchiv der Westfälischen-Wilhelms Universität in Münster ist eines der größten Pressearchive im deutschsprachigen Raum, das für jedermann zugänglich ist. Aber auch an anderen Universitäten gibt es Pressearchive. Einen aktuellen Überblick gibt die Website <http://www.textanalysis.info>.

Oft sind Zeitungen und Zeitschriften nicht mehr nur in gedruckter Form zugänglich, weil sie zum einen sehr viel Platz benötigen und zum anderen gerade bei Zeitungen die Qualität des Druckbildes immer mehr abnimmt. Der Grund dafür ist das qualitativ minderwertige Papier, auf dem Zeitungen gedruckt werden – das muss nicht lange haltbar sein. Daher werden Zeitungen und Zeitschriften verfilmt, entweder auf Mikrofilm oder Mikrofiche. Mit entsprechenden Geräten kann man sich dann die Ausgaben anschauen und auch die gewünschten Seiten ausdrucken oder abspeichern.

## 64.2 Textaufbereitung

Bei textanalytischen Verfahren wird das Material codiert (Mayring/Fenzl, Kapitel 38 in diesem Band), das kann man konventionell mit Papierausdrucken machen, aber auch mit entsprechender Software, entsprechende Übersichten sind im Internet zu finden (Kuckartz/Rädiker, Kapitel 27 in diesem Band). Dazu muss das Material digitalisiert werden, dafür gibt es einige Möglichkeiten wie Abschreiben, Scannen, Diktieren. Werden Texte abgeschrieben, so ist dies mit hohem Zeit- und Kostenaufwand verbunden, so dass man Alternativen wie das Scannen oder Diktieren in Betracht ziehen kann. Diktieren ist über längere Zeit gesehen anstrengend und für große Textmengen nicht geeignet, dafür aber sehr schnell. Die Anstrengung resultiert daraus, dass für eine gute Spracherkennung eine entsprechend gute Aussprache notwendig ist, denn diese bestimmt die Qualität der Umsetzung von Sprache in Text und ist nie ganz fehlerfrei. Der gesamte diktierte Text muss auf Fehler überprüft werden, so dass der Geschwindigkeitsvorteil der Spracherkennung gegenüber dem Abschreiben schrumpft. Gute Schreibkräfte schaffen mehr als 250 Zeichen pro Minute, sind meist schneller und machen weniger Tippfehler. Im Vergleich zur Spracherkennung muss auch nicht der gesamte Text Korrektur gelesen werden, weil Schreibkräfte wissen, ob und an welcher Stelle im Text sie Tippfehler gemacht haben.

Beim Scannen können insbesondere bei Zeitungen erhebliche Probleme auftreten, da die Druckqualität so schlecht ist, dass man mit Scannern und entsprechender Texterkennungssoftware erheblich langsamer ist als wenn der Text abgeschrieben würde (Klein 2013: 18-20). Bei Zeitschriften ist das anders, da das Papier erheblich besser und so der Einsatz von Scannern sinnvoll ist. Von großer Bedeutung ist die Auswahl von Medien und Untersuchungszeitraum, gute Beispiele finden sich bei Früh (2001) und Merten (1995).

---

## 64.3 Datenauswahl

Bevor man anfängt die Artikel zu sammeln, sollte man sich darüber klar sein, ob man alle Artikel – also eine Vollerhebung – oder eine Stichprobe analysieren möchte. Wie bei (quantitativen) empirischen Projekten üblich sind dabei die Hypothesen von zentraler Bedeutung, denn diese bestimmen den Umfang einer Datenerhebung. Diese wird weiterhin durch die zur Verfügung stehenden Ressourcen finanzieller und/oder personeller Art determiniert. Bevor man eine Stichprobe ziehen kann, muss man die Grundgesamtheit definieren. Geht es in einer Hypothese beispielsweise darum, wie ein Thema in deutschen Zeitungen dargestellt wird, so muss man einige Entscheidungen treffen. Allerdings hat nicht jede Zeitung eine eigene Redaktion, das gilt insbesondere für Zeitungen der Funke Mediengruppe Essen, dessen größte Zeitung WAZ in vielen regionalen Ausgaben erscheint, die keine eigenen Redaktionen haben und den Mantelteil der WAZ (also alles außer dem Lokalteil) übernehmen. Daher können nur Zeitungen mit eigener Redaktion – publizistische Einheit – als Untersuchungsgegenstand dienen. „Als publizistische Einheiten versteht

man redaktionell selbstständige Tageszeitungen mit Vollredaktionen. Dazu gehören alle Blätter, die den gesamten redaktionellen Teil (und damit sämtliche Ressorts) wie auch den Anzeigenteil selbstständig erarbeiten und verantworten.“ (Pürer/Raabe 1996: 111)

Die publizistischen Einheiten in Deutschland verringerten sich im Verlauf der letzten Jahrzehnte und erreichten 2009 eine Anzahl von 134 (Schütz 2009). Aus dieser Grundgesamtheit ist dann eine Stichprobe zu ziehen. Dazu muss ein Untersuchungszeitraum definiert werden, denn die Anzahl der publizistischen Einheiten ist davon abhängig. An dieser Stelle sei auf Stamm (2013) verwiesen, der jährlich erscheint und in dem alle publizistischen Einheiten aufgeführt sind.

Viele Abschlussarbeiten in den Medienwissenschaften oder der Publizistik untersuchen die Darstellung eines Themas in der Presse und beschränken sich dabei auf die Tagespresse. Somit werden in größerem Abstand erscheinende Zeitungen und Zeitschriften ausgeschlossen. Da eine Vollerhebung bei 134 Zeitungen nicht machbar ist, muss man aus dieser Grundgesamtheit eine Stichprobe ziehen. Während man bei Bevölkerungsumfragen die Personen entweder rein zufällig oder nach Quoten auswählt, ist dieses Vorgehen bei Zeitungen unangemessen. Diese unterscheiden sich insbesondere in ihrer Auflage bzw. Reichweite, so dass Stichprobenverfahren aus der Umfrageforschung nicht sinnvoll erscheinen (Früh 2001: 138-141).

Je nach Erkenntnisinteresse wird meist eine geschichtete Stichprobe gewählt, die auch die Auflagenhöhe berücksichtigt. Stichprobekriterien sind das Verbreitungsgebiet und die Vertriebsform, so dass aus nationalen überregionalen Abonnementszeitungen (Die Welt, FAZ, SZ und FR), Boulevardzeitungen (Bild, Abendpost Nachtausgabe, Express und Morgenpost Hamburg) und regionalen Abonnementzeitungen (z.B. Westdeutsche Allgemeine Zeitung (WAZ), Neue Osnabrücker Zeitung, Nürnberger Nachrichten) ausgewählt wird. Bei den ersten beiden Gruppen sind Vollerhebungen möglich, während man aus den regionalen Abonnementzeitungen eine Stichprobe ziehen sollte (dazu Früh 2001: 137-141 und Merten 1995: 283-292).

Die nächste Entscheidung betrifft die Analyseeinheit, den statistischen Fall. Dieser ergibt sich aus den Hypothesen und ist meist der Artikel. Selten ist auch eine ganze Tagesausgabe ein statistischer Fall, man sollte aber dennoch auf der Basis von Artikeln analysieren, denn aggregieren kann man dann immer noch.

Sind Zeitungen und Untersuchungszeitraum bestimmt, müssen die Artikel ausgewählt werden. Bei vielen Analysen werden alle zum Untersuchungsthema erschienenen Artikel ausgewählt. Muss aus diesen eine Stichprobe gezogen werden, dann bietet sich die *künstliche Woche* an, um Tageseffekte zu vermeiden. Damit wird vermieden, dass eine Verzerrung durch Inhalte der Tageszeitungen stattfindet, denn in Montagsausgaben ist der Sport überproportional vertreten, während bei den Samstagsausgaben der Unterhaltungsanteil umfangreicher als an anderen Wochentagen ist. Tageszeitungen erscheinen an jedem Wochentag ausgenommen am Sonntag. Eine künstliche Woche wird erreicht, in dem in der ersten Woche des Untersuchungszeitraums die Ausgabe vom Montag, in der zweiten Woche die Ausgabe vom Dienstag usw. genommen wird.

Für die Analyse des Untersuchungsmaterials bieten sich verschiedene Verfahren an:

- hypothesenprüfende: die sozialwissenschaftliche Inhaltsanalyse (content analysis)
- hypothesensuchende: die qualitative Datenanalyse (QDA – qualitative data analysis)

Bei hypothesenprüfenden Verfahren werden Kategorien gebildet und diese ausgezählt. Man erhält für jede Analyseeinheit die Häufigkeiten der Kategorien innerhalb dieser Analyseeinheit. Diese Daten können dann statistisch ausgewertet werden. Wichtig ist dabei, auch externe Merkmale des Mediums wie Name der Zeitung, Erscheinungsdatum, Position des Artikels (Seite), Artikelart (z.B. Aufmacher, Kommentar, Glosse, Reportage) zu erheben, damit später Hypothesengeleitet ausgewertet werden kann. Es wird ein Kategoriensystem und darauf basierend ein Codebuch entwickelt, so dass Codierer anhand von Beispielen und ihrer Sprachkompetenz den Text codieren. Beim Codievorgang wird anhand des Kategoriensystems der Text codiert, wobei man sich ein Kategoriensystem als Fragebogen an den Text vorstellen kann: Die Antworten sind standardisiert und jede Antwort hat einen numerischen Code, der dann in ein Codesheet eingetragen wird. Je Fall gibt es ein Codesheet. Die Daten werden mittels einer Software eingegeben und können dann statistisch analysiert werden. Mit einem Pretest wird das Kategoriensystem auf seine Brauchbarkeit getestet, bevor die Codierung des Textes beginnt. Die Reliabilität einer Codierung kann man dadurch gewährleisten, dass mehrere Codierer den gleichen Text codieren und die Abweichungen analysiert werden – man misst so die Intercoderreliabilität. Durch den Codievorgang selbst treten Gewöhnungs- und Lerneffekte bei den Codierern auf, diese Intracodierreliabilität kann dadurch gemessen werden, dass man den gleichen Text idealerweise zu Anfang und zum Ende des Projektes von denselben Codierern codieren lässt. Der Codievorgang selbst nimmt viel Zeit in Anspruch und wird meist nur einmal gemacht. Änderungen des Kategoriensystems sind nur während des Pretests möglich, nicht aber später.

Statt einer konventionellen Inhaltsanalyse kann man auch eine computerunterstützte durchführen, die auf rein syntaktischer Ebene arbeitet. Es werden Suchbegriffe – das können ganze Wörter, Wortteile oder Kombinationen davon sein – gezählt. Auf der Basis von Wörterlisten wird dann ein Kategoriensystem entwickelt, das geschieht meist mit entsprechender Software. Dabei kann das Kategoriensystem jederzeit geändert werden. Die Bedeutung der Wörter wird in Form von Kategorien berücksichtigt: ähnliche Suchbegriffe werden in einer Kategorie mit einem numerischen Code zusammengefasst und entsprechend ausgezählt. Je öfter eine Kategorie genannt wird, desto wichtiger ist sie. Allerdings treten semantische Probleme wie Mehrdeutigkeit und Negation auf, die bei einer konventionellen Inhaltsanalyse durch die Sprachkompetenz der Codierer gelöst werden. Die Unterschiede zwischen konventioneller und computerunterstützter Inhaltsanalyse beschreibt Klein (Klein 2013: 9-11) ausführlich. Ist das Untersuchungsmaterial zusammengestellt, werden die Texte je nach Erkenntnisinteresse analysiert.

Die Arbeitstechniken von computergestützten Textanalysen sind unterschiedlich (Klein 2010). Bei den hypothesenprüfenden *quantitativen* Verfahren wird ein Kategoriensystem erstellt, dessen Suchbegriffe mit dem Ziel der statistischen Analyse ausgezählt

werden, dabei können auch Zusammenhänge zwischen den Variablen des Kategoriensystems und den externen Variablen untersucht werden. Hierzu eignet sich besonders gut die Korrespondenzanalyse, wie an einem Projekt zur Untersuchung von Heirats- und Kontaktanzeigen gezeigt wird (Klein/Giegler 1994).

Bei der *qualitativen* Datenanalyse werden die wichtigen Textstücke am Bildschirm markiert, und so entsteht sukzessive ein Kategoriensystem. Auch hier kann man externe Variablen definieren und mit den Kategorien verknüpfen. Die meisten Softwareprogramme bieten noch erheblich mehr Möglichkeiten der Analyse, auch ist es möglich, Kategorien zu zählen und statistisch weiter zu verarbeiten. Einen aktuellen Überblick über die aktuelle Software bietet <http://www.textanalysis.info>, allerdings fehlt es an einem systematischen Vergleich, wie er noch vor 20 Jahren von Weitzman und Miles (1995) erstellt wurde. In Büchern zu Methoden der empirischen Sozialforschung werden text- und/oder inhaltsanalytische Verfahren bis auf wenige Ausnahmen nur knapp behandelt. Für die Analyse von Zeitungen und Zeitschriften kommen Verfahren der qualitativen Datenanalyse kaum zum Einsatz.

Textmining ist eine Analyseform, mit der ohne ein Kategoriensystem und ohne manuelle Codierung Texte untersucht werden. Dies ist nur mit entsprechender Software möglich, die im Text Schlüsselwörter sucht und diese miteinander in Beziehung setzt. Die Form, in der Schlüsselwörter und deren Beziehungen dargestellt werden, reicht von statistischen Daten bis hin zu graphischen Darstellungen.

Die inhaltsanalytische Methode hat durch den Zuwachs qualitativer Verfahren in den Sozialwissenschaften an Bedeutung gewonnen. Dazu hat insbesondere die Entwicklung entsprechender Software beigetragen.

In den Standardwerken zu den Methoden der empirischen Sozialforschung kommen inhaltsanalytische Methoden entweder gar nicht oder nur sehr verkürzt vor. Für die Inhaltsanalyse selbst stehen gute Lehrbücher zur Verfügung. Lehrveranstaltungen an sozialwissenschaftlichen Instituten decken inhaltsanalytische Verfahren oft nicht ab, auch ist entsprechende Software an den Instituten nicht so verbreitet wie beispielsweise Statistiksoftware.

## Literatur

- Atteslander, Peter (2010): Methoden der empirischen Sozialforschung, 13. Aufl. Berlin: Schmidt Verlag
- Früh, Werner (2001): Inhaltsanalyse. Theorie und Praxis. 5., überarb. Auflage, Konstanz: UVK
- Klein, Harald/Giegler, Helmut (1994): Correspondence Analysis of Text Data with INTEXT/PC. In: Greenacre, Michael/Jörg Blasius (Hg.): 283-301
- Greenacre, Michael/Blasius, Jörg (Hg.) (1994): Correspondence Analysis in the Social Sciences. London: Academic Press
- Klein, Harald (2010): Inhaltsanalyse. In: Atteslander, Peter (Hg.): 195-224
- Klein, Harald (2013): Computerunterstützte Textanalysen mit TextQuest. Eine Einführung in Methoden und Arbeitstechniken. München und Mering: Rainer Hampp Verlag
- Krippendorff, Klaus (2004): Content Analysis. An Introduction to Its Methodology. 2. Aufl. Thousand Oaks: Sage
- Merten, Klaus (1995): Inhaltsanalyse. Einführung in Theorie, Methode und Praxis, 2. Aufl. Opladen: Westdeutscher Verlag
- Pürer, Heinz/Raabe, Johannes (1996): Medien in Deutschland. Band 1: Presse, 2. Aufl. Konstanz: UVK
- Schütz, Walter J.: (2009): Redaktionelle und verlegerische Struktur der deutschen Tagespresse. In: Mediaperspektiven 2009: 484-493
- Stamm, Willy (2013): STAMM Leitfaden durch Presse und Werbung 2013: Presse- und Medienhandbuch. Essen: STAMM Verlag
- Weitzman, Eben/Miles, Matthew B. (1995): Computer Program for Qualitative Data Analysis. A Software Sourcebook. Thousand Oaks: Sage

## Weblinks (aufgerufen am 7.7.2014)

<http://www.textanalysis.info> gibt seit 1999 einen ständig aktualisierten Überblick über Textanalysesoftware. Dort werden die Programme kurz beschrieben, es gibt Links zu den Handbüchern und zu kostenlosen Test- oder Demoversionen, so dass man ausprobieren kann, welche Software am besten für das eigene Projekt geeignet ist.

[http://de.wikipedia.org/wiki/Publizistische\\_Einheit](http://de.wikipedia.org/wiki/Publizistische_Einheit)

Andreas Schmitz und Olga Yanenko

---

## 65.1 Was sind Logfiles?

Mit Logfiles werden automatisch erstellte Dateien bezeichnet, in denen bestimmte Ereignisse elektronisch aufgezeichnet werden. Im sogenannten *Web Usage Mining* (Srivastava et al. 2000) sind vor allem *Web Server Logs* von Interesse und werden daher oft synonym zum Begriff „Logfiles“ verwendet. Diese dokumentieren die Zugriffe auf Webseiten, um darauf basierend Webseitenstatistiken erstellen und auswerten zu können. In der Regel speichern sie Informationen über Dateien, auf die zugegriffen wurde, den „zugreifenden“ Nutzer sowie Statusinformationen. Von inhaltlichem Interesse in den Sozialwissenschaften ist die Aufzeichnung von Interaktionen zwischen einem System und dessen menschlichen Nutzern, sowie insbesondere zwischen verschiedenen Nutzern, die über ein solches System kommunizieren. Für die empirische Sozialforschung sind solche technisch-menschlichen Interaktionssysteme eine neuartige Methode der Datengewinnung und Datenanalyse.

Die für die Sozialforschung besonders interessanten Logfiles von Anwendungen des Social Web umfassen die Aufzeichnung sämtlicher vor der Untersuchung als relevant definierter Handlungen, etwa Klickaktionen aller Nutzer auf einem bestimmten Web-Angebot. Sie können damit als „elektronisches Äquivalent zu physischen (Verhaltens-)Spuren“ (Jansen et al. 2009) beziehungsweise „web-generierte Prozessdaten“ (Schmitz et al. 2009) verstanden werden. *Wir definieren Logfiles wie folgt: Logfiles sind ein digitales, automatisiert aufgezeichnetes, chronologisches Protokoll von zuvor als relevant definierten Ereignissen in einem computerbasierten System.*

Sowohl die Wahl der Datenstruktur und des Datenformats als auch die Definition der aufzuzeichnenden Ereignisse obliegt dem Administrator des Systems beziehungsweise dem Wissenschaftler, der eine bestimmte Forschungsfrage untersucht. Je nachdem, welche sonstigen Ereignisse als relevant definiert und protokolliert werden, lassen sich aus Logfiles nicht nur Webseitenstatistiken erstellen und aggregieren, sondern auch

Profilinformationen der Nutzer sowie inhaltlich relevante (Inter-) Aktionsereignisse und Netzwerkstrukturen rekonstruieren.

Tabelle 65.1 zeigt einen Datenauszug aus einer Interaktionsdatenbank einer Onlinekontaktbörse, der aus Entitäten der in Abb. 65.2 dargestellten Tabellen abgeleitet und in ein „Flat File“ überführt wurde. Dabei wurden insbesondere die Sender- und Nachrichteninformationen zusammengeführt und zu Analysezwecken angereichert (etwa Abbildung von Sender-Empfänger-Relationen und Identifikation von E-Mail-Adressen über reguläre Ausdrücke).

Man sieht, dass ein Sender mit der ID 16172 einen Receiver mit der ID 9239 am 01.07.2012 um 13:15 und 26 Sekunden kontaktiert hat. Mit diesem Ereignis wurde die 5430 Dyade auf der Kontaktbörse konstituiert. Die angespielten Profildaten erlauben eine Beschreibung dieser Dyade. Beim Sender handelt es sich um einen Mann im Alter von 45 Jahren mit Abitur, der eine 42-jährige Frau, ebenfalls mit Abitur, angeschrieben hat. Diese hat am gleichen Tag um 18:42 geantwortet. Am nächsten Tag erfolgte wiederum eine Nachricht durch den Mann. Das siebte und letzte Kontakt Ereignis, in welchem der Mann seine E-Mail-Adresse (wahrscheinlich) zum Zweck eines Medienwechsels eingetragen hat, erfolgte am Abend desselben Tages. Es zeigt sich, dass mit diesen Daten der zeitliche Interaktionsverlauf bis zum Verlassen des Beobachtungsfensters rekonstruiert werden kann. Weitere Informationen lassen sich aus der Datenbank ableiten und hinzu spielen, insbesondere Aggregationen der Prozessdaten auf Nutzerebene. Basierend auf diesen Daten können nun inhaltlich motivierte Analysen gerechnet werden.

## 65.2 Datenmanagement

Da Logfiles als elektronische Ereignisdaten grundsätzlich in einer beliebigen Form vorliegen können, besteht die erste Herausforderung der Arbeit mit Logfiles in der Extraktion der für spätere statistische Untersuchungen relevanten Daten. Die Datenaufbereitung hängt dabei stark davon ab, welche Datenstruktur und welches Datenformat für die Speicherung verwendet wurden.

**Tab. 65.1** Beispiele für Daten-Auszug

ID	ID	Dyad	Contact	Mail	Time		Sex		Age		Edu	
					S	R	S	R	S	R	S	E
16172	9239	5430	1	0	01.07.2012	13:15.26	M	F	45	42	Abitur	Abitur
9239	16172	5430	2	0	01.07.2012	18:42.19	F	M	42	45	Abitur	Abitur
16172	9239	5430	3	0	02.07.2012	08:33.01	M	F	45	42	Abitur	Abitur
...												
16172	9239	5430	7	1	02.07.2012	22:17.49	M	F	45	42	Abitur	Abitur

Die einfachste Form von Logfiles sind einfache Textdateien, in denen für jedes Ereignis eine Zeile aufgezeichnet wurde. Die Struktur dieser Datenzeilen beziehungsweise die relevanten Inhalte können dabei stark variieren. Mittlerweile gibt es für einige Anwendungen (zum Beispiel Web Server Logs) zwar standardisierte Datenformate (zum Beispiel Common Log Format), jedoch sind diese nur für einen bestimmten Zweck ausgerichtet und daher nicht für jede Fragestellung geeignet (Kohavi 2001). In jedem Fall gilt, dass reine Textdateien lediglich eine Folge von Zeichen darstellen, welche von einem Computer nicht ohne weiteres interpretiert werden können. Das übliche Vorgehen ist daher die Untersuchung der Datei nach Regelmäßigkeiten ihrer Zeichenfolgen (zum Beispiel erst Datum, dann Leerzeichen, dann Name) und eine anschließende Formulierung dieser Regelmäßigkeiten mit Hilfe von regulären Ausdrücken. Unter einem regulären Ausdruck versteht man eine Zeichenfolge, die mit Hilfe syntaktischer Regeln Zeichenketten aus Texten filtern kann. Beispielweise kann man eine syntaktisch korrekte E-Mail-Adresse mit Hilfe des folgenden regulären Ausdrucks identifizieren:

```
^ [a-zA-Z0-9] [\w\.-]* @ (? : [a-zA-Z0-9] [a-zA-Z0-9_-]+\.) + [A-Z,a-z] {2,5} $
```

Die wichtigsten Elemente eines regulären Ausdrucks sind Zeichenklassen ([a-zA-Z0-9]) steht beispielsweise für eine Zeichenmenge bestehend aus Klein- und Großbuchstaben sowie Ziffern) und Quantoren, die angeben, wie oft ein Zeichen vorkommen darf (\* steht beispielsweise für „beliebig oft“ und {2,5} für „zwischen zwei- und fünfmal“) (Friedl 2007). So lassen sich die relevanten Daten aus der Textdatei extrahieren und in eine für eine statistische Analyse besser geeignete Datenstruktur – wie zum Beispiel eine Tabelle – überführen. Abbildung 65.1 zeigt einen beispielhaften Auszug aus einem Web Server Log sowie die dazugehörige Tabelle nach der Vorverarbeitung.

Über eine gezielte Aufzeichnung von Computerereignissen zum Zweck der Performance und der Analyse von Fehlern hinaus existieren auch elektronische Prozessdaten,

```
182.12.175.3 18/05/2012:08:04:22 "GET logo.gif" 512 "http://www.xy.org" 200 ...
174.165.26.36 20/05/2012:10:36:18 "GET banner.jpg" 1024 "http://www.abc.com" 404 ...
```



Wer? (IP-Adresse)	Wann? (Zugriffszeitpunkt)	Was? (Anforderung)	Wie viel? (Datenmenge in Byte)	Woher? (Internetseite)	Status (HTTP- Code)	...
182.12.175.3	18/05/2012:08:04:22	GET logo.gif	512	http://www.xy.org	200	
174.165.26.36	20/05/2012:10:36:18	GET banner.jpg	1024	http://www.abc.com	404	
173.102.144.7	23/05/2012:04:03:45	GET bild.jpg	512	http://www.abc.com	200	

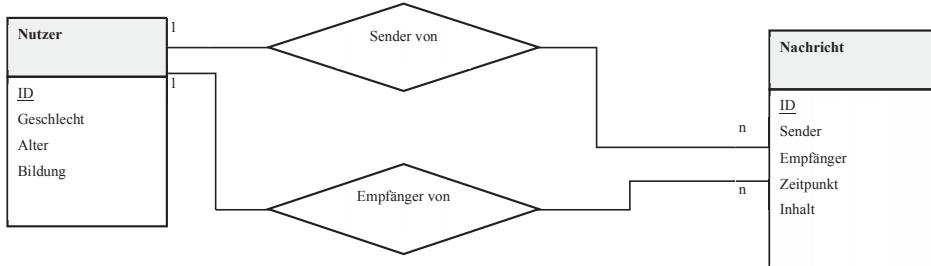
**Abb. 65.1** Vorverarbeitung von Web Server Logs

die ursprünglich keinem inhaltlichen Analysezweck dienten. Diese findet man etwa bei Social-Web-Plattformen, bei denen Ereignisse zwischen Nutzern und dem virtuellen Netzwerk wiederum Inhalte für weitere Nutzer bereitstellen (zum Beispiel Pinnwandeinträge bei Facebook, private Nachrichten bei Onlinekontaktbörsen etc.). Die Umsetzung solcher Plattformen erfolgt in der Regel mit Hilfe einer serverbasierten Skriptsprache (zum Beispiel PHP), die für einen Nutzer relevante Daten aus einer relationalen Datenbank (zum Beispiel MySQL) dynamisch in den Webbrower lädt (Maurice 2009). Da Interaktionen zwischen Nutzern selbst als Webseiteninhalte behandelt werden, werden diese ebenfalls in Tabellen der Datenbank abgelegt. Abbildung 65.2 veranschaulicht die Struktur von Tabellen einer solchen Datenbank am Beispiel von privaten Nachrichten in einem sozialen Netzwerk.

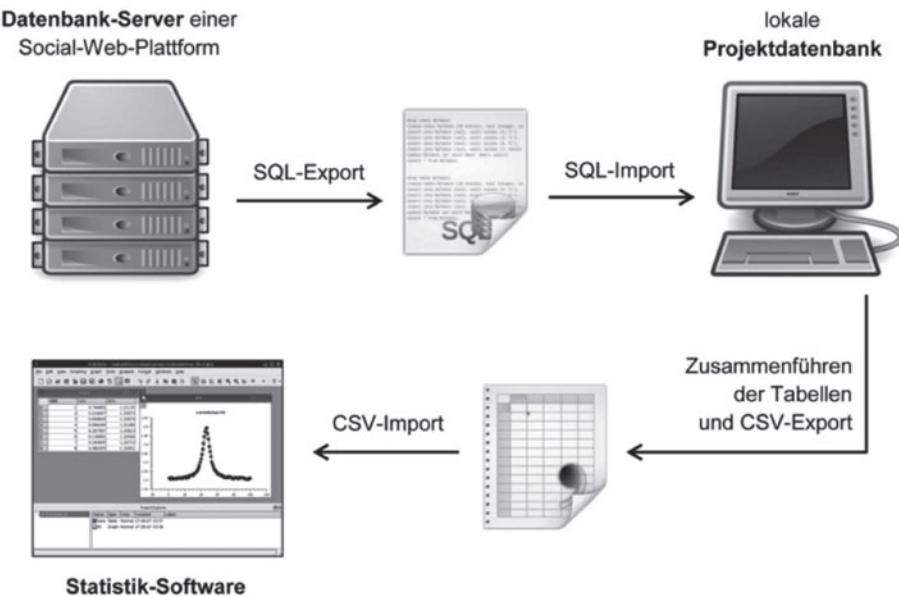
Hierbei werden in der Tabelle „Nutzer“ Profilinformationen der angemeldeten Nutzer einer Plattform gespeichert. Die Tabelle „Nachricht“ beinhaltet Nachrichten, die sich Nutzer gegenseitig schickten.

Durch die automatisierte Erhebung der Daten und eine Haltung in einer zuvor definierten Struktur lassen sich diese Daten leicht in eine für Analysen mit Statistik-Software übliche Datenstruktur („Flat File“) überführen. Zu diesem Zweck übergibt der Betreiber einer Plattform einen anonymisierten Export etwa als SQL-Datei (Structured Query Language) (Throll/Bartosch 2004). Der Export muss zunächst in eine lokale Datenbank reimportiert werden, wobei die Tabellen in derselben Struktur wie in der Ursprungsdatenbank angelegt werden. Mit Hilfe von SQL-Befehlen lassen sich die Tabellen zusammenführen, so dass alle für einen jeweiligen Anwendungsfall relevanten Daten als eine Tabelle ausgegeben werden. Diese Tabelle lässt sich als CSV-Datei exportieren und in eine Statistiksoftware laden. Der Weg von der Anbieterdatenbank bis zum Flat File (s.o.) ist in Abb. 65.3 veranschaulicht.

Die größte Herausforderung in diesem Vorverarbeitungsschritt ist die Aufbereitung und Aggregation der Daten auf Datenbankebene. Herkömmliche Statistiksoftware stößt bei derartigen relationalen Daten sehr schnell an ihre Grenzen. Daher ist es besonders wichtig, bereits im Vorfeld die aufzuzeichnenden Ereignisse sowie den Modus ihrer Vorrhaltung genau zu definieren. Da dies oft nicht möglich ist – vor allem, wenn die Daten von einem privatwirtschaftlichen Anbieter stammen und somit dieser die Struktur der aufzuzeichnenden Daten bestimmt – sollte zumindest der Datenaufzeichnungsmodus des



**Abb. 65.2** Struktur einer relationalen Datenbank mit zwei Tabellen



**Abb. 65.3** Umwandlung der Daten einer relationalen Datenbank

Anbieters exakt rekonstruierbar sein. Beispielsweise ist es wichtig zu wissen, ob es während des Aufzeichnungszeitraums System-Relaunches (verbesserte neue Versionen eines Services, die eventuell veränderte oder neue Funktionen bieten) gab, was zu Strukturbrüchen in den Daten führen kann. Bei der Datenübergabe im Rahmen einer Kooperation mit einem privatwirtschaftlichen Anbieter ist zudem stets zu beachten, dass die Daten in anonymisierter Form transferiert werden und die Nutzer des Systems eingewilligt haben müssen, dass ihre Informationen analysiert werden (Mühlchen, Kapitel 4 in diesem Band).

Besteht keine Kooperation mit einem Anbieter, der Daten zur Verfügung stellt, so hat man dennoch die Möglichkeit online auf Interaktionsdaten zuzugreifen. Im Gegensatz zu einer Kooperation lassen sich hier jedoch nur Daten auswerten, die öffentlich zugänglich sind. Hierzu zählen Diskussionsforen, Blogs mit Kommentarfunktion oder Wikis. Um die Daten automatisiert auslesen zu können, benötigt man einen sogenannten „Crawler“ – ein Programm, das bestimmte Webseiten durchforstet und relevante Daten abspeichert. Beispiele für solche Crawler sind unter anderem der SocSciBot (<http://socscibot.wlv.ac.uk/>), welcher eine oder mehrere Webseiten nach ausgehenden Verweisen durchsucht und in Form von Netzwerkdiagrammen ausgibt, oder der UbiCrawler (Boldi et al. 2004) (<http://law.dsi.unimi.it/software.php>). Findet man keinen passenden Crawler für die vorliegende Fragestellung, ist es oftmals mit geringem Aufwand möglich, ein eigenes individualisiertes Programm zu schreiben. Vor dem Einsatz eines Crawlers sollte man sich jedoch vergewissern, ob der Anbieter die Daten nicht von sich aus kostenfrei zum Download oder

auf Anfrage bereitstellt. Wikipedia beispielsweise stellt einen solchen Datenexport allgemein zur Verfügung, um zu verhindern, dass sich durch intensives Crawlen die Zugriffsgeschwindigkeit auf die Webseite verlangsamt (Wikipedia 2012). Zudem stellen bekannte Social-Web-Plattformen sogenannte Application Programming Interfaces (APIs) bereit, die als Schnittstelle für den Zugriff auf Daten von einem eigenen Computerprogramm gedacht sind.

---

### 65.3 Methodologische Stärken und Schwächen von Logfiledaten

Während durch den Befragungsprozess die Antworten eines Befragten zum Teil stark beeinflusst werden können, zeichnet sich die automatische Aufzeichnung von elektronischen Beobachtungsdaten durch ihre „Nicht-Reaktivität“ beziehungsweise „Nicht-Obtrusivität“ aus (Webb et al. 2000). In der Erhebung von Logfile-Daten stellen Unit- oder Item-Nonresponse (Engel/Schmidt, Kapitel 23 in diesem Band) kein Problem dar. Aufgrund der starken Automatisierbarkeit des Datenerhebungsprozesses ist die *Reliabilität* der Daten im Vergleich zu Befragungen höher einzustufen. Auch die Erhebungskosten sind vergleichsweise gering, da bei minimalem Personal- und Zeitaufwand massenhaft Daten aufgezeichnet werden können. Im Vergleich zu Befragungsdaten sind Logfile-Daten zudem vollständig und folgen als Vollerhebung aller Akteure im digitalen Interaktionskontext nicht der Stichprobenlogik der Befragungsforschung. Gibt es am Ende des Datenerhebungsprozesses informationelle Lücken, sind diese auf die mangelhafte vorherige Definition der aufzuzeichnenden Ereignisse beziehungsweise den Datenhaltungsmodus zurückzuführen. Der Datenbestand kann effizient erzeugt werden, ohne umständliche und fehleranfällige Transkriptionen durchführen zu müssen. Dennoch sind auch Logfile-Daten gründlichen Qualitäts- und Plausibilitätskontrollen zu unterziehen.

Üblicherweise leitet man in der Sozialforschung aus theoretischen Konstrukten zu erhebende Variablen ab. Dies ist mit Logfile-Daten nur dann möglich, wenn vor der Erhebung die aufzuzeichnenden Informationselemente definiert und programmiert wurden. Prozessproduzierte Daten einer privatwirtschaftlichen Social-Web-Plattform entstammen allerdings nicht-operationalisierten Konstrukten und decken damit nicht automatisch diejenigen Bereiche ab, die von theoretischem Interesse sind (Janetzko 2008; Jansen et al. 2009). Eine den Soziologen interessierende Rekonstruktion des jeweils subjektiv gemeinten Sinns etwa ist anhand objektiver Ereignisse schlicht nicht möglich, diese können bestenfalls „behavioristisch“ interpretiert werden. Unter Umständen kann dies bedeuten, dass die Validität eines aufgezeichneten Konstrukts im Vergleich zu einer Befragung gering ist.

Zwar decken Logfiles sämtliche (zuvor als relevant definierte) Akteure und Handlungen ab und unterscheiden sich damit wesentlich von meist stichprobenbasierten Befragungen. Gleichzeitig ist das Problem der Repräsentativität damit keineswegs zu vernachlässigen. Es stellt sich nämlich erstens die Frage *wofür* die vollständig erhobenen Daten repräsentativ sein sollen. Erhebt man beispielsweise per Crawler in Wikipedia sämtliche

Artikel, Autoren sowie die Relationen („Kanten“) zwischen diesen, ist noch ungeklärt, welche die Aussagegesamtheit dieser Daten über Wikipedia hinaus darstellen soll. Ist das deutsche Wikipedia repräsentativ für alle freien Wissensplattformen? Oder für alle Wikis? Sind die Nutzer/innen und/oder deren Verhalten auf einer Kontaktbörsen repräsentativ für alle anderen Dating-Plattformen oder sogar für die Offline-Population der Partnersuchenden? Es gilt also zu berücksichtigen, dass es sich bei derartigen Kontexten um mitunter hochselektive Segmente aus der Gesamtbevölkerung handelt, die sich in Hinblick auf ihre Soziodemographie, aber auch ihr Handeln, systematisch von der Offlinepopulation unterscheiden können.

Die Stärken und Schwächen von web-generierten Prozessdaten lassen sich auch in Abgrenzung zum klassischen sozialwissenschaftlichen Experiment formulieren. Mit Hilfe eines Experiments wird ebenfalls versucht, anhand von Beobachtungen objektiven Verhaltens inhaltliche Rückschlüsse zu ziehen. Das experimentelle Design versucht dabei über die Definition von Kontrollgruppen und Randomisierung valide und reliable Aussagen mit Kausalitätsanspruch zu generieren (Eifler, Kühnel/Dingelstedt, Kapitel 11 und 80 in diesem Band). Die automatisierte Aufzeichnung von Logfiles hingegen bezieht sich – wenn nicht anders programmiert – auf alle Nutzer eines Systems, die sich zudem wechselseitig beeinflussen können. Die so gewonnenen Erkenntnisse können kaum unter ceteris paribus-Bedingungen reproduziert werden, sondern stützen sich auf die Beobachtung von Ereignissen in nicht-künstlichen, dynamischen Situationen. Andererseits stellen Logfile-Daten Informationen eines „tatsächlichen“ sozialen Feldes mit einer deutlich geringeren Obtrusivität dar als experimentelle Daten, die aus hochgradig künstlichen Interaktionssituationen entstammen. Zusätzlich ermöglicht die elektronische Aufzeichnung von Handlungen im Unterschied zu zeitlich begrenzen Experimenten das Erzeugen von Massendaten ohne den Aufzeichnungsprozess zu beenden. Gleichwohl ist es möglich, die Stärken des experimentellen Ansatzes und der elektronischen Beobachtung zu kombinieren, indem man etwa Online-Experimente konzipiert, die Logfile-Daten produzieren (Dandurand et al. 2008). So ist es vielversprechend die Nutzer eines Social-Web-Angebots zu randomisieren und die vergleichbaren Subpopulationen einer systematischen Bedingungsvariation zu unterziehen.

## Literatur

- Boldi, Paolo/Codenotti, Bruno/Santini, Massimo/Vigna, Sebastino (2004): UbiCrawler: A Scalable Fully Distributed Web Crawler. In: Software: Practice/Experience 34: 711-726
- Dandurand, Frédéric/Shultz, Thomas/Onishi, Kristine (2008): Comparing online and lab methods in a problem-solving experiment. In: Behavior Research Methods 40: 428-434
- Fielding, Nigel/Lee, Raymond M./Black Grant (Hg.) (2009): The Sage Handbook of Online Research Method. Thousand Oaks, CA: Sage
- Friedl, Jeffrey (2007): Reguläre Ausdrücke (3. Auflage). Köln: O'Reilly
- Janetzko, Dietmar (2008): Nonreactive Data Collection on the Internet. In: Fielding et. al. (Hg.)
- Jansen, Bernard/Taksa, Isak/Spink, Amanda (2009): Resarch and Methodological Foundations of Transaction Log Analysis. In: Jansen et. al. (Hg.): 1-16
- Jansen, Bernard/Taksa, Isak/Spink, Amanda (Hg.) (2009): Handbook of Research on Web Log Analysis. Hershey, PA: IGI Global
- Kohavi, Ron (2001): Mining e-commerce data: the good, the bad, and the ugly. In: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '01). New York, NY: AMC 8-13
- Maurice, Florence (2009): PHP 5.3 + MySQL 5.1: Der Einstieg in die Programmierung dynamischer Websites. München: Addison-Wesley
- Schmitz, Andreas/Skopek, Jan/Schulz, Florian/Klein, Doreen/Blossfeld, Hans Peter (2009): Indicating mate preferences by mixing survey and process-generated data. The case of attitudes and behavior in online mate search. In: Historical Social Research 34: 77-93
- Srivastava, Jaideep/Cooley, Robert/Deshpande, Mukund/Tan, Pang- Ning (2000): Web usage mining: discovery and applications of usage patterns from Web data. In: SIGKDD Explor. Newslett. 1. 2. 12-23
- Throll, Marcus/Bartosch, Oliver (2004): Einstieg in SQL: SQL-Syntax von MySQL, Access, SQL Server, Oracle, MaxDB/SAPDB, DB2 und Firebird. Bonn: Gallileo Computing
- Webb, Eugene, Campbell, Donald, Schwartz, Richard D., Sechrest, L. (2000): Unobtrusive Measures. Revised Edition. Thousand Oaks, CA: Sage
- Wikipedia (2012): Database download. Quelle: [http://en.wikipedia.org/wiki/Wikipedia:Database\\_download](http://en.wikipedia.org/wiki/Wikipedia:Database_download). Am 16.05.2012

Cornelia Thierbach und Grit Petschick

---

## 66.1 Einleitung

Mittels wissenschaftlicher Beobachtung können viele Themenfelder erforscht werden. Typisch sind Beobachtungen von Menschen im öffentlichen Raum wie etwa in Stadtquartieren, von Randgruppen (wie z.B. Obdachlosen oder Prostituierten), von Experten (wie z.B. Wissenschaftlern), von Abläufen in Unternehmen, oder auch des Geschehens bei Events. Daher kommt die wissenschaftliche Beobachtung in vielen verschiedenen soziologischen Forschungsrichtungen zur Anwendung, z.B. in der Stadt- und Regionalsoziologie, der Organisationssoziologie oder auch in der Wissenschafts- und Technikforschung.

Es handelt sich bei der Beobachtung um eine Datenerhebungsmethode, bei der sich die Beobachterin bewusst (mit den ihr zur Verfügung stehenden fünf Sinnen) an die Orte des Geschehens begibt, an denen sie ihre Daten erheben möchte. Daten mittels Beobachtung zu erheben ist vor allem dann sinnvoll, wenn es darum geht, Prozesse, Organisationen, Beziehungen, Handlungsabläufe oder Interaktionsmuster zu verstehen. Es geht also primär um die Beantwortung von Forschungsfragen, die das „Wie?“ erkunden möchten oder das „Was geschieht hier?“. Vieles von dem, was Personen nicht verbalisieren können oder was nicht abgefragt werden kann, ist dennoch erfahrbar und damit auch erfassbar, da man als Beobachterin Zeuge davon wird. Würde man stattdessen z.B. Personen befragen (Helfferich und Küsters, Kapitel 39 und 40 in diesem Band) wie sie kochen oder Auto fahren, könnten sie natürlich davon berichten. Jedoch könnten dabei unbewusst oder auch bewusst für die Forscherin wichtige Aspekte ausgelassen werden oder gar nicht artikuliert werden, da manchmal einfach nicht die passenden Worte gefunden werden können.

Das Beobachten an und für sich ist für die meisten Forschenden eine gewohnte Tätigkeit, denn wir tun es ständig, z.B. wenn wir im Supermarkt an der Kasse stehen und schauen, wie ein Produkt nach dem nächsten über den Scanner gezogen wird, wenn wir auf einer Parkbank sitzen und unseren Blick über die Landschaft schweifen lassen oder auf einer Straße anderen Personen beim Laufen ausweichen. Im Gegensatz zu diesen alltäglichen

Beobachtungen sind wissenschaftliche Beobachtungen systematisch geplant, sie finden also nicht zufällig statt, sie sind an einer Forschungsfrage ausgerichtet (Reichert, Kapitel 2 in diesem Band), werden dokumentiert, aufbereitet, ausgewertet, und die Erkenntnisse werden in einem wissenschaftlichen Bericht zusammengefasst.

So universell einsetzbar wie diese Datenerhebungsmethode zunächst erscheint, ist sie dennoch nicht für die Beantwortung aller Forschungsfragen geeignet: Nur mit den fünf Sinnen Wahrnehmbares kann auf diese Weise erhoben werden, und auch nur dann, wenn man einen Zugang zu einem *Untersuchungsfeld* (kurz: *Feld*) hat.

Zu unterscheiden sind zunächst verschiedene *Formen der Beobachtung*. So unterteilt Friedrichs (1990 [1973]) die Formen der wissenschaftlichen Beobachtung anhand von fünf Dimensionen:

- *Verdeckt/offen*: Bei einer verdeckten Beobachtung gibt sich die Beobachterin als solche nicht zu erkennen und die Tatsache, dass beobachtet wird, wird verheimlicht. Bei einer offenen Beobachtung hingegen ist den Beobachteten klar, dass sie beobachtet werden, der Zweck muss ihnen aber nicht bis ins kleinste Detail bewusst sein. Aus ethischen Gründen (Friedrichs, Kapitel 3 in diesem Band) empfehlen wir immer eine offene Beobachtung anstatt einer verdeckten. Es mag zwar sein, dass es vorläufig zu Reaktivitäten auf Grund der Anwesenheit einer Forscherin im Feld kommt, aber diese nehmen nach etwas Eingewöhnungszeit erfahrungsgemäß ab.
- *Nicht-teilnehmend* (passive Teilnahme)/*teilnehmend* (aktive Teilnahme): Übernimmt die Forscherin eine Rolle im Feld und nimmt sie an den Interaktionen teil oder befindet sie sich außerhalb des Feldes und beobachtet tatsächlich als Forscherin hauptsächlich das Geschehen? Gold (1958) unterteilt die Beobachterrolle noch etwas differenzierter: „vollständig Beobachtender“, „Beobachtender als Teilnehmender“, „Teilnehmender als Beobachtender“ und „vollständig Partizipierender“. All diese Beobachterrollen können innerhalb eines Forschungsvorhabens in unterschiedlichen Phasen auftreten. Je mehr eine Beobachterin im Feld involviert ist und je stärker sie sich mit den Beforschten identifiziert, desto mehr kann sie die wissenschaftliche Distanz verlieren. Dieser Prozess, auch als „going native“ bezeichnet, wird unterschiedlich bewertet: zum einen als Gefahr, weil einem der Abstand zum Forschungsgegenstand fehlt; zum anderen als Chance, um zu verstehen, was im Feld vor sich geht und zwar durch die eigene Erfahrung und Involviertheit. Ziel sollte es generell sein, eine Balance zwischen Nähe und Distanz zu finden, wobei in der Praxis phasenweise das eine oder das andere dominieren kann. Wichtig ist, dass man seine Rolle im Feld regelmäßig reflektiert (zur Nähe-Distanz-Problematik siehe auch Knoblauch, Kapitel 36 in diesem Band).
- *Systematisch (stark strukturiert)/unsystematisch (schwach strukturiert)*: Hier geht es um die Frage, ob die Beobachtung – ähnlich wie bei einer standardisierten Befragung – nach einem vorab standardisierten Schema erfolgt und man im Extremfall zählt, wie oft eine Kategorie in der Empirie vorkommt, oder ob sie nicht standardisiert, den Interessen der Forscherin folgend abläuft.

- *Natürlich/künstlich*: Wird die Beobachtung im natürlichen Umfeld der Beobachteten durchgeführt oder handelt es sich hierbei um eine Labor- bzw. experimentelle Situation (Eifler, Kapitel 11 in diesem Band) außerhalb des alltäglichen Umfeldes der Beobachteten?
- *Selbstbeobachtung/Fremdbeobachtung*: Der Beobachter kann sich selbst oder andere Personen beobachten (Knoblauch, Kapitel 36 in diesem Band).

Im Folgenden konzentrieren wir uns auf die schwach strukturierte teilnehmende Beobachtung, die eng mit der Ethnographie (Knoblauch, Kapitel 36 in diesem Band) verbunden ist, wo sie die wichtigste Methode der Datenerhebung darstellt (Fetterman 1997).

---

## 66.2 Planung des Forschungsprozesses

Wie bei jedem Forschungsvorhaben müssen auch bei der wissenschaftlichen Beobachtung vor der Datenerhebung eine Reihe von Vorüberlegungen getroffen werden. Ergänzend zu Przyborski und Wohlrab-Sahr (Kapitel 6 in diesem Band) sind bei der Beobachtung insbesondere folgende Faktoren wichtig:

- *Klärung der Forschungsfrage*: Hat man eine Forschungsfrage fokussiert, sollte zunächst geprüft werden, ob man mittels Beobachtung die „relevanten“ Daten auch erheben kann. Im Laufe eines qualitativ-orientierten Projekts kann sich die genaue Forschungsfrage immer wieder leicht ändern (Przyborski/Wohlrab-Sahr und Strübing, Kapitel 6 und 32 in diesem Band). Häufige Reflexionen über das Erlebte und sich selbst als Beobachterin sind während des Forschungsprozesses unabdingbar.
- *Theoretisches Interesse*: Das eigene theoretische Interesse färbt auch schon die Datenerhebung und nicht nur die Interpretation und Auswertung. So wird je nach theoretischem Interesse und infolge selektiver Wahrnehmung jede Beobachterin eine Situation im Feld etwas anders wahrnehmen, da man einigen Dingen unbewusst mehr Aufmerksamkeit und Bedeutung schenkt als anderen, was sich in der Regel auch in den Feldnotizen und Feldprotokollen widerspiegelt. Auch kann es sein, dass man durch das Erleben im Feld auf bestimmte theoretische Probleme gestoßen wird, die ebenfalls berücksichtigt werden müssen.
- *Abstecken des Gegenstands*: Es muss geklärt werden, was das Forschungsfeld des Projekts ist und wie man dazu einen Zugang (siehe Abschnitt 3) erhält. Oft ist in der Praxis nicht ganz eindeutig zu erkennen, was bzw. auch wer zum Feld dazugehört (Akremi und Strübing, Kapitel 17 und 32 in diesem Band).
- *Fallauswahl*: Da es sich bei der Beobachtung um ein sehr zeitintensives und aufwändiges Datenerhebungsverfahren handelt, können nur wenige Fälle beobachtet werden, weshalb entsprechende Stichprobenverfahren für kleine Fallzahlen angewendet werden sollten (Akremi und Hering/Schmidt, Kapitel 17 und 37 in diesem Band).

- *Vorbereitung der Feldaufenthalte:* Da im Feld die unterschiedlichsten Strukturen und Interessen beachtet werden müssen, ist es notwendig, sich über das Feld, in dem man beobachten möchte, zu informieren. Von Interesse sind z.B. die hierarchische Ordnung (formal, informell), räumliche Strukturen (z.B. wie sieht es vor Ort aus und wer sitzt wo), wichtige Personen, organisationale Abläufe, Handlungs- und Kommunikationsmuster und wiederkehrende typische Situationen.
- *Forschungsdesign:* Hat man in etwa eine Vorstellung über Theorie, Fragestellung und Gegenstand, kann man ein Forschungsdesign (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) erstellen. Dazu gehört, genau zu planen, wann und wo (teilweise auch was genau) man beobachten möchte. Auch die konkrete Planung der Feldaufenthalte (einschließlich Absprachen mit Personen aus dem Feld) muss organisiert werden.
- *Verhalten im Feld:* Da man bei der Beobachtung direkten Kontakt zu Menschen hat, ist es unabdingbar, dass man sich bei der Kontaktaufnahme und im Feld „anständig“ verhält (siehe Abschnitte 3 und 4).

---

## 66.3 Feldzugang

Bei Beobachtungen im öffentlichen Raum hat man meist keine Probleme, einen Zugang zu erhalten. Anders sieht es jedoch bei nicht-öffentlichen Räumen aus, wie z.B. bei Betrieben, Institutionen oder geschlossen Veranstaltungen. Hier benötigt man als wissenschaftliche Beobachterin die Erlaubnis der Personen im Feld, dass man dieses erforschen darf, den sogenannten Feldzugang. Da eine teilnehmende Beobachtung nur im Einverständnis mit allen beteiligten Personen erfolgen kann, betrachten wir bei den folgenden Überlegungen nicht nur den formalen Zugang, sondern darüber hinaus auch die Zustimmung der Personen, die konkret beobachtet werden.

### 66.3.1 Top-Down-Prinzip vs. Bottom-Up-Prinzip

Der Feldzugang kann auf unterschiedliche Weise organisiert werden. Beim Top-Down-Prinzip orientiert sich die Beobachterin an formalen Zugängen und Hierarchien in einem Feld und kontaktiert insbesondere Personen auf einer höheren oder der höchsten Leitungsebene, die über die entsprechenden Entscheidungsbefugnisse verfügen. Der Kontakt zu den anderen Personen im Feld wird im Regelfall über diese Person eingeleitet. In einem Unternehmen kann dies z.B. dadurch erfolgen, dass die Chefin die Beobachterin zu einer Mitarbeiterbesprechung einlädt oder durch das Unternehmen führt und den Mitarbeitern vorstellt. Der Nachteil des Top-Down-Prinzips ist, dass bei den Personen auf den unteren Hierarchieebenen das Gefühl entstehen kann, dass sie einer Erforschung zustimmen müssen, obwohl sie es eigentlich gar nicht wollen.

Beim Bottom-Up-Prinzip werden dagegen direkt die Personen angesprochen, die begleitet werden sollen, auch wenn sie sich auf unteren Hierarchieebenen befinden und

formal nicht die Entscheidungsmacht (für die Zustimmung zu einer Erforschung oder für den formalen Zugang zur Institution) haben. Der Nachteil des Bottom-Up-Prinzips ist, dass die Person mit Weisungsbefugnis sich in sehr formal und hierarchisch organisierten Zusammenhängen übergegangen fühlen kann.

Sowohl beim Top-Down- als auch beim Bottom-Up-Prinzip müssen am Ende beide Seiten informiert werden und einer Erforschung durch die Beobachterin zustimmen (Friedrichs und Mühlichen, Kapitel 3 und 4 in diesem Band). Nur in stark hierarchisch organisierten Feldern (z.B. bei staatlichen Behörden) empfehlen wir den Zugang nach dem Top-Down-Prinzip zu realisieren, in den anderen Fällen sollten die Personen angesprochen werden, die beobachtet werden sollen.

Hat die Forscherin bereits bestehende Kontakte zum Feld, sollten diese genutzt werden. Diese Personen können als sogenannte „*Gatekeeper*“ (Garfinkel 1967) fungieren, d.h. als Personen, die die Beobachterin unterstützen und die im Feld über Respekt und Anerkennung verfügen; oft können sie auch der Beobachterin Zugang zu Dingen (Güter, Informationen oder Dienstleistungen) verschaffen, die ohne diese Unterstützung für die Beobachterin nicht oder nur schwer zu erreichen wären.

### 66.3.2 Der erste Kontakt

Der erste Kontakt ist für die gesamte Beobachtung von besonderer Bedeutung, da hiervon der Feldzugang abhängt. Daher ist es wichtig, den Feldzugang möglichst frühzeitig zu klären und nicht erst in dem Moment, in dem die Datenerhebung beginnen soll.

Der erste Kontakt kann schriftlich (E-Mail oder Brief) oder mündlich (Telefon oder persönlicher Besuch) erfolgen. Welcher Weg der Beste ist, ist von Feld zu Feld und Fall zu Fall zu entscheiden. Dabei sollte die Beobachterin bereits gut über das Feld informiert sein sowie das eigene Forschungsvorhaben Dritten erklären können.

Es geht hierbei insbesondere darum, den Personen im Feld zu erklären, wie die beabsichtigte teilnehmende Beobachtung aussieht und was dies für die beforschten Personen konkret bedeutet. Die Beobachterin sollte immer daran denken, dass sie auf die Zusammenarbeit des Feldes angewiesen ist, Vieles über Vertrauen ihr gegenüber läuft und mit persönlichen Sympathien verbunden ist. Ebenso sollte dargestellt werden können, warum es gerade dieses Feld sein sollte und welchen Vorteil das Feld selber von der Forschung hat, denn einige Felder haben ein sehr hohes Interesse daran, dass sie „auch etwas von der Beobachtung haben“. Die Beobachterin sollte sich überlegen in welchem Umfang und in welcher Form sie die Ergebnisse später ins Feld zurückgeben will und kann.

Wichtig ist es auch, sich bewusst zu sein, welche Probleme im Feld durch eine Beobachtung entstehen können. Hierzu gehören zum einen Fragen zu dem Umgang mit den Daten, ihrer Anonymisierung und Verwendung (Friedrichs, Mühlichen und Meyer und Meier, Kapitel 3, 4 und 15 in diesem Band) und zum anderen formale Bestimmungen, die für einen Feldzugang geregelt werden müssen, wie z.B. der Versicherungsschutz im Feld.

Insbesondere bei Beobachtungen in nicht-öffentlichen Räumen sollte in diesem Zusammenhang auch geklärt werden, von wem was erwartet wird – verpflichtet sich das Feld mit der Zustimmung zu bestimmten Dingen, dann sollten diese explizit dargestellt werden: Wann soll die Forschung beginnen? Wer soll begleitet werden? In welchen Zeitintervallen und für welchen Zeitraum soll die Beobachtung erfolgen? Zu welchen Orten braucht die Beobachterin Zugang? Welche Bedingungen müssen erfüllt werden, damit der Zugang möglich ist? Wird ein Arbeitsplatz im Feld benötigt? Und wenn ja, wo sollte dieser sein? Sollen Untersuchungen in einem Labor oder einer Fabrik, in dem beispielsweise mit Maschinen oder gefährlichen Substanzen gearbeitet wird, gemacht werden, so muss gegebenenfalls eine Sicherheitsbelehrung erfolgen oder Schutzkleidung organisiert werden, um den (Versicherungs)Schutz der Beobachterin zu gewährleisten. Um zu vermeiden, dass sich die Personen im Feld bei der Arbeit gestört fühlen, kann vereinbart werden, dass die Zustimmung zur teilnehmenden Beobachtung immer wieder situativ erneuert werden muss und dass in diesem Rahmen auch zu einem späteren Zeitpunkt entweder in Bezug auf bestimmte Situationen oder ganz allgemein eine Ablehnung der Erforschung erfolgen kann.

---

## 66.4 Beobachtungsstrategien und Verhalten im Feld

Hat man Zugang zum Untersuchungsfeld erhalten oder ist dieser prinzipiell gegeben, kann die tatsächliche Datenerhebung beginnen, und auch diese sollte sorgfältig geplant werden.

### 66.4.1 Informationsquellen

Beobachterin im Feld zu sein, bedeutet nicht, dass man sich tatsächlich nur auf seine fünf Sinne stützt – man kann auch Fragen stellen (Helfferich, Kapitel 39 in diesem Band). Umgekehrt kann es auch passieren, dass man als Beobachterin von Personen im Feld angesprochen wird, was man denn da mit welchem Ziel mache, oder was man gerade aufschreibe, wenn man sich Notizen macht. In solchen Fällen sollte man, wenn es die Forschungsergebnisse nicht beeinflusst, auf die Fragen entsprechend eingehen und diese ehrlich beantworten.

Leider ist es oft so, dass man während der Beobachtung eine Neigung hat, sich auf das Gesagte zu konzentrieren. Dabei sollte man gerade als Beobachterin nicht vergessen, dass man sich auch das Setting anschaut, auf Gegenstände (Schubert, Kapitel 70 in diesem Band), Räumlichkeiten (Dangschat, Kapitel 76 in diesem Band) und Geräusche achtet, Gerüche und Atmosphären wahrnimmt.

Während die Interviewerin sich „nur“ auf das Gespräch konzentrieren muss, passiert bei der Beobachtung im Feld Vieles zur gleichen Zeit. So muss man sich nicht nur auf das Gesagte einer Person konzentrieren, sondern man muss Vieles auf einmal im Blick haben

und teilweise auch noch nebenbei Notizen machen. Konzentriert man sich zu sehr auf das eine, verpasst man ggf. das andere. Da hilft es nur, wenn man vorab bewusst einen Fokus festlegt.

Auch die gegenteilige Situation kann von der Beobachterin als problematisch gesehen werden, nämlich dann, wenn sie das Gefühl hat, dass gerade nichts passiert. In diesem Fall sollte man sich selbst fragen, warum man so empfindet. Liegt es daran, dass einem die Vorgänge im Feld bereits bekannt sind und man ähnliche Situationen schon erlebt und protokolliert hat? Vielleicht sieht man auch einfach nichts mehr, weil das sogenannte „going native“ eingetroffen ist. Abstand zum Feld könnte da ganz gut tun.

#### **66.4.2 Dauer der Feldaufenthalte**

Wann und wie oft man im Feld ist, kann in unterschiedlichen Projekten, aber auch phasenweise sehr verschieden sein. Das hängt zum einen damit zusammen, dass das Feld zusagen muss, wann man beobachten darf. Einige Zugänge (z.B. zu bestimmten Räumen wie einem Labor, zu Meetings oder Events) erhält man erst, wenn man das Vertrauen der Personen im Feld gewonnen hat. Zum anderen kommt es auch auf das Forschungsdesign an und wie die Feldaufenthalte geplant sind.

Da das Beobachten teilweise eine sehr ermüdende Aufgabe ist, sollte man die Zeit des konzentrierten Beobachtens nach Möglichkeit begrenzen, u.a. kann man erfahrungsgemäß nicht mehr als drei bis vier Stunden am Stück beobachten: Die Konzentration lässt nach, man wird müde und kann sich nicht mehr an alles erinnern. Nach dieser Zeit sollte zumindest eine Pause eingelegt werden. Bei wiederholten Feldaufenthalten kann man sich zunächst auf die Routinen und Alltaglichkeiten konzentrieren und diese festhalten. Zu späteren Zeitpunkten kann man sich nur auf besondere Ereignisse fokussieren.

#### **66.4.3 Anforderungen an die Beobachterrolle**

Abgesehen von fachlichen Kompetenzen muss generell eine Beobachterin vielen Anforderungen gerecht werden. Zunächst ist die Beobachterin eine Art Visitenkarte für ein Projekt, ein Institut oder gar für die gesamte Fachrichtung. Sie sollte sich zum einen selbst als Forscherin ernst nehmen und selbstbewusst auftreten. Zum anderem muss sie ein Feingefühl dafür entwickeln, was wann erlaubt ist, wann eine Intervention durch die Beobachterin oder gar ihre Anwesenheit stört.

Insgesamt ist die Rolle der Beobachterin zum Teil sehr schwierig, sie folgt ja einer Tätigkeit, die allgemein als störend angesehen wird: andere beobachten. Das fordert viel Feingefühl, aber auch eine gewisse Standhaftigkeit, weil sich diese Rolle immer wieder sehr unangenehm anfühlen kann und es unter Umständen Überwindung kostet sich dieser Situation auszusetzen. Hat die Beobachterin das Gefühl, dass es für die Beobachteten

unangenehm oder schwierig ist, so können diese Probleme auch offen angesprochen werden, man kann sich aus Situationen zurückziehen, wenn sie heikel erscheinen oder man kann sich naiv stellen, um nähere Informationen zu erhalten.

Von der Grundeinstellung her sollte man kooperieren und informieren. Schließlich ist die Forscherin auf Informationen aus dem Feld angewiesen, da kann sie auch Informationen über sich und ihre Forschung preisgeben. Dabei ist Ehrlichkeit sehr wichtig, um allein das Vertrauen zu gewinnen. Definitiv sollte man Interesse bekunden, nicht nur an bestimmten Tätigkeiten, sondern auch an jeder Person im Feld, mit der man öfter zu tun hat.

Für jede Beobachtung gilt, dass man das Feld so wenig wie möglich verändert. Wie stark man sich am Feld beteiligt, sollte daher regelmäßig reflektiert werden. Das kann phasenweise, aber auch den Umständen entsprechend schwanken. Eine zu starke Beteiligung am Feld ist aus zwei Gründen problematisch: Zum einen kann es passieren, dass die Forscherin nicht richtig beobachten kann, wenn sie zeitgleich handeln muss. Zum anderen kann es passieren, dass sie bei Konflikten involviert wird und sich dadurch nicht neutral verhalten kann. Auch sollte man vermeiden, vertrauliche Informationen (Friedrichs und Mühllichen, Kapitel 3 und 4 in diesem Band) auszuplaudern. Dass die Personen einem vertrauen können, ist unabdingbar für die Forschung und darf nicht missbraucht werden. Selbstbeobachtung und ständige Reflexionen bringen einen weiterhin ans Ziel. Vorurteile und Meinungen sollte man sich zudem bewusst machen.

Da auch das eigene Befinden Auswirkungen auf die Wahrnehmung haben kann – z.B. wenn die Beobachterin mit privaten Problemen belastet ist – kann das die Aufmerksamkeit minimieren, oder sie ist nur für bestimmte Themen im Feld sensibel.

---

## 66.5 Feldnotizen und Feldprotokoll

Nicht nur die Feldaufenthalte müssen geplant sein, auch sollte genug Zeit für die Verschriftung des Erlebten und die Reflexion eingeplant werden (Emerson et al. 2001, 2011). Zudem muss beachtet werden, dass zu jeder empirischen Forschung eine sorgfältige Dokumentation gehört, damit man später Aussagen, die man in Publikationen trifft, auch belegen kann. Sie dient weiterhin dazu die Nachvollziehbarkeit der Ergebnisse (Flick, Kapitel 29 in diesem Band) sicherzustellen und eröffnet die Möglichkeit, dass die Daten ggf. reanalyisiert werden können. Darum wird es immer üblicher auch qualitative Forschungsdaten zu archivieren (Medjedović, Kapitel 13 in diesem Band).

### 66.5.1 Feldnotizen

Bereits im Feld sollte die Beobachterin sich Notizen machen. Das direkte Notieren von Feldnotizen hat den Vorteil, dass wichtige Informationen sofort festgehalten werden. Das kann allerdings die Aufmerksamkeit der beobachteten Personen auf sie lenken. Auch kann

es sein, dass sie Dinge nicht mitbekommt, weil sie gerade mit dem Schreiben der Notizen beschäftigt ist. Macht man dagegen erst im Anschluss an die Beobachtungssituation Notizen, kann es sein, dass einige Details bereits vergessen sind. Für eine weiterführende Lektüre zum Schreiben von Feldnotizen und Feldprotokollen empfehlen wir Emerson et al. (2011). Zusätzlich zu dem schriftlichen Festhalten von Informationen können Notizen auch auf ein Diktiergerät gesprochen werden. In Absprache mit den beobachteten Personen können Mitschnitte oder Videoaufnahmen (Tuma/Schnettler, Kapitel 68 in diesem Band), z.B. von Meetings, gemacht werden.

### 66.5.2 Feldprotokolle

Zu jedem Feldaufenthalt sollte so bald wie möglich, am besten direkt im Anschluss, ein Feldprotokoll geschrieben werden. Ein Feldprotokoll ist eine detaillierte Beschreibung des Feldaufenthaltes. Festgehalten wird alles, was für die Forschungsfrage relevant ist: z.B. was habe ich erlebt, was wurde gesagt, wie sahen die Räume aus, wer war beteiligt?

Die Protokollierung sollte möglichst zeitnah (idealerweise noch am selben Tag, sonst innerhalb von 24 Stunden) nach der Beobachtung erfolgen. Ansonsten steigt die Gefahr, dass man das Erlebte (vor allem zahlreiche Details) vergisst oder sich gar falsch erinnert.

Als Faustregel gilt, dass das Schreiben des Protokolls mindestens so lange braucht, wie Zeit im Feld verbracht wurde, wenn nicht sogar noch länger. Die Länge und Frequenz der Feldaufenthalte sollte daher so gestaltet sein, dass zeitnah genügend Zeit für das Schreiben der Protokolle eingeplant wird. Wird zum Beispiel eine Person in ihrem Arbeitsalltag eine gesamte Woche lang begleitet, bedeutet dies für die Beobachterin unter Umständen, dass sie erst nach einem Arbeitstag bzw. Feldtag von über zehn Stunden dazu kommt, die Feldprotokolle zu schreiben. In einem solchen Fall ist es kaum möglich, das Protokoll noch am selben Tag zu schreiben. Soweit dies möglich ist, sollte man daher immer wieder (z.B. tageweise) Feldpausen einlegen.

Wir empfehlen, die Notizen im Präsens zu schreiben. Für den Leser ist es einfacher, sich in das Beschriebene hinein zu versetzen. Formen der Vergangenheit schaffen u.U. eine Distanz zu dem Geschehenen.

Des Weiteren empfehlen wir, die Wiedergabe gesprochener Sprache als direkte Zitate zu notieren und nicht in Form indirekter Rede – auch wenn sich die Beobachterin nicht immer zu 100% an den exakten Wortlaut erinnert. Nicht nur die Wiedergabe des Gesagten, sondern allgemein sind Feldprotokolle immer das Abbild der eigenen, subjektiven Erinnerung und Sichtweise einer Situation, da es schwierig oder sogar unmöglich ist, bei der Beschreibung „neutral“ zu bleiben und nicht zu werten.

Ein weiteres Problem bei der Verschriftlichung von Beobachtungen ist die Umwandlung von Sozialem in Sprache (Hirschauer 2001): Es wird nicht nur Gesprochenes wiedergegeben, sondern auch andere Eindrücke, die über Sprache nur eingeschränkt wiedergegeben werden können, wie z.B. die Reduktion einer komplexen Realität oder die Gleichzeitigkeit von Ereignissen.

Das Beschriebene sollte alle relevanten Informationen enthalten und dabei möglichst detailliert sein. Eigene Gefühle, Meinungen und Einschätzungen sollten als solche gekennzeichnet werden, z.B. durch entsprechende Formulierungen wie „mir kam es so vor“, „ich habe das Gefühl“ oder in Form von Memos (Strübing, Kapitel 32 in diesem Band), Kommentaren oder in kursiver Schrift. Des Weiteren können auch erste Interpretationen (Kurt/Herbrik, Kapitel 33 in diesem Band) festgehalten werden, da sich die Einschätzungen von Situationen im Verlauf der Feldforschung noch ändern können und damit wichtig für die späteren Interpretationen und allgemein für die Reflexion der eigenen Rolle im Feld sind. Sie sind auch ein guter Indikator für eine Entwicklung, die die Beobachterin im Feld gemacht hat (z.B. im Sinne des „going native“). Stellt man eine Änderung in der eigenen Wahrnehmung oder in der Interpretation von beobachteten Situationen fest, können folgende Fragen hilfreich sein: Was hat sich geändert, dass ich die Situation jetzt anders einschätze? Habe ich die Situation anders eingeschätzt, weil ich bestimmte Informationen zu dem Feld noch nicht hatte? Solche Veränderungen in der Wahrnehmung können sehr aufschlussreich für die Interpretation der Beobachtungen sein, da sie die Differenz zwischen einer „Innen-“ und einer „Außen-Perspektive“ (Knoblauch, Kapitel 36 in diesem Band) aufzeigen und verdeutlichen, was zum Feld und seiner spezifischen Logik gehört.

Auch das eigene Befinden und persönliche Ereignisse, die nicht in direktem Bezug zu dem Forschungsvorhaben stehen, können mit aufgenommen werden, da die eigene Situation die selektive Wahrnehmung verändern und zudem ein Erinnerungsanker sein kann.

### **66.5.3 Anonymisierung**

Die Anonymisierung der Feldnotizen ist aus Datenschutzgründen zum Schutz der beobachteten Personen sehr wichtig. Sie kann jedoch sehr schwierig und in gewissen Zusammenhängen sogar unmöglich sein, da es sich oft nicht nur um die Anonymisierung der Beobachteten, sondern auch um dritte Personen handelt, die nur zufällig oder als Randpersonen im Feld erscheinen. Insbesondere detaillierte Daten bergen viele Informationen, die ggf. nicht genannt werden können. So kann zum Beispiel die Nennung eines Hobbies oder einer bestimmten Aufgabe, die einer Person zugeteilt ist, genügen, um diese für andere erkennbar zu machen. Auch Zitate können dazu führen, dass Personen über die Wiedergabe gesprochener Sprache zu erkennen sind. Dies ist vor allem für Personen mit markanten Akzenten einleuchtend, aber auch subtilere Merkmale eines Sprachstils, wie Ausdrucksform, Grammatik oder bestimmte Wörter können ausreichen, damit eine Person für andere erkennbar wird. Zur weiteren Lektüre zu diesem Thema empfehlen wir Hopf (2004) und Häder (2009).

## 66.6 Auswertung von Beobachtungsdaten

Da die Beobachtungsdaten durch das Schreiben der Feldprotokolle in schriftlicher Form vorliegen, können diese nach textbasierten Auswertungsverfahren analysiert werden. Es empfehlen sich z.B. die sozialwissenschaftliche Hermeneutik (Kurt/Hebrik, Kapitel 33 in diesem Band), die Inhaltsanalyse nach Mayring (Mayring/Fenzl, Kapitel 38 in diesem Band) oder eine Auswertung nach der Grounded Theory (Strübing, Kapitel 32 in diesem Band). Zudem empfiehlt sich (gerade bei großen Datenmengen) die Verwendung eines QDA-Programms (Kuckartz/Rädiker, Kapitel 27 in diesem Band).

## Literatur

- Emerson, Robert M./Fretz, Rachel I./Shaw, Linda L. (2001): Participant Observation and Fieldnotes. In: Atkinson, Paul A./Delamont, Sara/Coffey/Amanda/Lofland, John/ Lofland, Lyn H. (Hg.): *Handbook of Ethnography*. London: Sage. 339-351
- Emerson, Robert M./Fretz, Rachel I./Shaw, Linda L. (2011): *Writing Ethnographic Fieldnotes*. Chicago: The University of Chicago Press
- Fetterman, David M. (1997): Ethnography. In: Bickman, Leonard/Rog, Debre J. (Hg.): *Handbook of applied social research methods*. London: Sage. 473-505
- Friedrichs, Jürgen (1990[1973]): Methoden empirischer Sozialforschung. Opladen: Westdeutscher Verlag
- Garfinkel, Harold (1967): Studies in ethnomethodology. Englewood Cliffs NJ: Prentice-Hall
- Gold, Raymond L. (1958): Roles in sociological field observations. In: Social Forces 36 (3): 217-223
- Häder, Michael (2009): Der Datenschutz in den Sozialwissenschaften. RatSWD, Working Paper 90. URL: [http://www.ratswd.de/download/RatSWD\\_WP\\_2009/RatSWD\\_WP\\_90.pdf](http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf), zuletzt besucht 16.04.2014
- Hirschauer, Stefan (2001): Ethnografisches Schreiben und die Schweigsamkeit des Sozialen. In: Zeitschrift für Soziologie 30 (6): 429-451
- Hopf, Christel (2004): Forschungsethik und qualitative Forschung. In: Flick, Uwe/von Kardoff, Ernst/Steinke, Ines (Hg.): *Qualitative Forschung*. Reinbek: Rowohlt. 589-600.

Ralf Bohnsack

---

## 67.1 Einleitung

Die sozialwissenschaftlich-empirische Bildanalyse ist eine Domäne qualitativer Sozialforschung (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band). Letztere sieht sich allerdings vor das Problem gestellt, dass die wesentlichen Fortschritte qualitativer Methoden seit Ende der 1970er Jahre zunächst eng mit der Entwicklung der Verfahren der Textinterpretation verbunden waren, welche wiederum im Zusammenhang mit dem sogenannten *linguistic turn* (verbunden mit Namen wie Richard Rorty, Paul Ricoeur und Jürgen Habermas), also der sprachwissenschaftlichen Wende in den Geistes- und Sozialwissenschaften, zu sehen ist. Dies hat in den qualitativen Methoden zu enormen Fortschritten der Interpretationsverfahren geführt. Damit verbunden war allerdings eine Marginalisierung des Bildes. Die hochentwickelten qualitativen Verfahren der Textinterpretation vermögen sich nur schwer von der Bindung an die Logik von Sprache und Text und vom textlichen Vorwissen zu lösen. Dies wäre Voraussetzung, um zur Eigenlogik des Bildes vorzudringen, wie es in dem u.a. von Mitchell (1994) postulierten *pictorial turn* bzw. dem *iconic turn* im Sinne von Boehm (1994: 13) gefordert wird. Diese Forderung nach einer bildwissenschaftlichen Wende hat schließlich zur Etablierung der sogen. *Bildwissenschaft* geführt. Deren Vertreter Belting (2001:15) hat dies mit Bezug auf Mitchell (1994) dahingehend formuliert, dass es darum gehe, „Bilder nicht mehr mit Texten zu erklären, sondern von Texten zu unterscheiden“.

Die Bildwissenschaft ist zu einem großen Teil aus der Kunstgeschichte hervorgegangen, versteht sich aber als transdisziplinär, bleibt allerdings weitgehend geisteswissenschaftlich-philosophisch orientiert. Erst in den letzten zehn Jahren konnte der Eigenlogik des (unbewegten) Bildes ansatzweise sozialwissenschaftlich-empirische Relevanz verliehen werden. Die derart entfaltete Interpretationsmethodik versteht das Bild als ein „selbstreferentielles System“. Wie ein derartiger Zugang zur Eigenlogik des Bildes forschungspraktisch

realisiert wird, ist in anschaulicher Weise an Forschungsbeispielen demonstriert worden (Bohnssack 2001, 2009: Kap. 4.2 und 4.3).

---

## 67.2 Die Stellung der Bildinterpretation in der sozialwissenschaftlichen Empirie

Der linguistic turn hatte nicht zuletzt deshalb für die empirischen Methoden so tiefgreifende Konsequenzen, weil in allen sozialwissenschaftlichen Methodologien jene Prämissen Gültigkeit hat, die zuerst wohl von Karl Popper in prägnanter Form vorgetragen worden ist: Wirklichkeit muss, wenn sie wissenschaftliche Relevanz gewinnen will, in Form von Beobachtungssätzen oder „Protokollsätzen“ (auch: „Basissätzen“; Popper 1971), also in Form von Texten, vorliegen. Die qualitative oder rekonstruktive Sozialforschung ist dem nicht nur gefolgt, sondern sie hat aus dieser Prämissen noch weitergehende Konsequenzen gezogen: Nur dort, wo das sprachliche, das verbale Handeln der Erforschten, also die von ihnen selbst produzierten Texte, die Ursprungsdaten darstellen, brauchen diese nicht noch einmal durch die Forscher oder Beobachter in Texte, also Protokoll- oder Beobachtungssätze transformiert zu werden. Letzteres ist aber im Bereich der Bildinterpretation in besonderem Maße der Fall und stellt diese – wie auch beispielsweise die teilnehmende Beobachtung (Thierbach/Petschik, Kapitel 66 in diesem Band) – sogleich unter den Verdacht mangelnder Validität.

Die qualitativen Methoden der Bildanalyse unterscheiden sich untereinander ganz wesentlich dahingehend, in welchem Ausmaß sie am Text als generellem „Modell der Sozialwissenschaft“ (Ricoeur 1978) festhalten oder aber sich der Eigenlogik des Bildes vor allem auf der Grundlage seiner formalen Gestaltung und in Ausklammerung des textlichen Vor-Wissen sowie in Abgrenzung von der Logik der Sprache und Textinterpretation (Foucault 1971) zuwenden. An dem erstgenannten Pol ist die Bildinterpretation im Rahmen der Methodologie der „Objektiven Hermeneutik“ (Loer 1994) und der daran angelehnten „Kultursoziologischen Bildhermeneutik“ (Müller-Doohm 1997) zu verorten, an dem anderen die *dokumentarische Bildinterpretation* auf der Grundlage der „Dokumentarischen Methode“ (Bohnssack 2009). Zwischen diesen Polen finden wir aktuell u.a. die in der Tradition der Sozialphänomenologie stehenden Bildinterpretationen im Bereich der „Hermeneutischen Wissenssoziologie“ (Reichertz 1992, Kurt/Herbrik, Kapitel 33 in diesem Band) und der „Visuellen Wissenssoziologie“ (Raab 2008) ebenso wie die Fotoanalyse von Goffman (1979), und die *ethnografisch orientierten Ansätze der Fotoanalyse* (Prosser 1998) sowie die unmittelbar an Panofsky orientierten Bildinterpretationen der „Pädagogischen Hermeneutik“ (Mollenhauer 1983) und der „Seriell-ikonografischen Fotoanalyse“ (Pilarcyk/Mietzner 2005). Anschauliche Beispiele für die Vorgehensweise und die Arbeitsschritte bei der Bildinterpretation auf der Grundlage der Dokumentarischen Methode finden sich am exemplarischen Fall von Werbe- und Familienfotos u.a. in Bohnssack (2001, 2009: Kap. 4.2 und 4.3).

## 67.3 Die Konstitution der Wirklichkeit durch Bilder

Der Eigenlogik des Bildes im Unterschied zu derjenigen des Textes gerecht zu werden, bedeutet auch, dass wir einen Zugang zu jenen Bereichen unserer alltäglichen Kommunikation finden, in denen wir uns, wie ich es genannt habe (Bohsack 2003), *durch* Bilder verständigen – im Unterschied zu einer Verständigung *über* Bilder, die sich immer im Medium von Sprache und Text vollzieht. Dass wir uns im Alltag *durch* Bilder verständigen, bedeutet, dass unsere Welt, unsere gesellschaftliche Wirklichkeit durch Bilder nicht nur *repräsentiert*, sondern auch *konstituiert* wird (Mitchell 1994: 41). Dabei ist die Herstellung der Welt durch Bilder nicht lediglich in dem Sinne zu verstehen, dass die *Deutung* der Welt sich wesentlich im Medium der Ikonizität vollzieht, vielmehr sind Bilder *handlungsleitend*. Sie sind als mentale Bilder von sozialen Szenerien, von Gebärden, Gestik und Mimik eingelassen in die vorreflexiven, impliziten oder „atheoretischen“ Wissensbestände (Mannheim 1964), in ein implizites oder inkorporiertes Wissen, welches vor allem auch das habituelle Handeln strukturiert.

## 67.4 Übereinstimmungen zwischen den wichtigsten Traditionen der Bildinterpretation

Der Komplexität derartiger handlungs-, zeichen- und wissenstheoretischer Voraussetzungen werden einige Theorietraditionen gerecht, so vor allem die semiotischen bzw. semiologischen Theorien und Methodologien von Barthes (1990) und Eco (1994) einerseits, die ikonologische Methodologie des Kunsthistorikers Panofsky (1975) in ihrer Weiterentwicklung durch die Ikonik von Imdahl (1994) andererseits sowie drittens die Wissenssoziologie Karl Mannheims mit der von ihm ursprünglich am Beispiel der Kunstinterpretation (1964) entworfenen Dokumentarischen Methode. Zwischen diesen Theorietraditionen und Methodologien werden deutliche Korrespondenzen sichtbar (Bohsack 2003, 2009: Kap. 3). Übereinstimmungen zeigen sich u.a. zwischen der für die Semiotik bestimmten Differenzierung zwischen denotierender und konnotierender Sinnebene auf der einen Seite und der ikonologischen Methode von Panofsky auf der anderen Seite, bei der zwischen vor-ikonografischer und ikonografischer Sinnebene differenziert wird.

Um diese Differenzierung an einem Beispiel von Eco (1994: 243) zu erläutern (der den konnotativen Code an dieser Stelle explizit als ikonografischen Code bezeichnet), so kann ich auf der *denotativen* bzw. *vor-ikonografischen* Ebene auf einem Bild „eine halbnackte Frau mit einem Männerkopf auf einem Teller“ identifizieren. Erst auf der *konnotativen* bzw. *ikonografischen* Ebene erscheint mir dieses Bild als die Darstellung von „Salomé“. Die Besonderheit und Eigensinnigkeit des Bildes im Unterschied zum Text, d.h. die Besonderheit der bildhaften, der ikonischen Zeichen, entscheidet sich auf der denotativen Ebene. Die Entschlüsselung der denotativen Botschaft bzw. des „ikonischen Codes“, wie Eco (1994: 67) diese auch nennt, also jener Botschaft, die nur durch das Bild zu vermitteln ist, geht allerdings immer durch den ikonografischen oder konnotativen Code hindurch,

„entledigt“ sich aber der (ikonografischen) Konnotationen und ist somit „eine Restbotschaft, die aus dem besteht, was vom Bild übrig bleibt, wenn man (geistig) die Konnotationenzeichen ausgelöscht hat“ (Barthes 1990: 37). In Foucaults (1971: 36) Bildinterpretation am Beispiel von Velasquez: Las Meninas heißt dies (1971: 36): „Man muß also so tun, als wisse man nicht“. Und im Sinne von Imdahl (1996: 435) kann eine derartige Analyse „prinzipiell von der Wahrnehmung des literarischen oder szenischen Bildinhalts absehen, ja sie ist oft besonders erfolgreich gerade dann, wenn die Kenntnis des dargestellten Sujets sozusagen methodisch verdrängt wird“.

Eine derartige Sinninterpretation, welche die konnotative oder ikonografische Ebene in einer noch genauer zu bestimmenden Analyseeinstellung ausklammert oder suspendiert, hat Roland Barthes (1990) vorgelegt und als „*stumpfen Sinn*“ („*sens obtue*“) bezeichnet, welcher sich ganz wesentlich von dem durch konnotative bzw. ikonografische Wissensbestände geprägten „*entgegenkommenden*“ (oder: „*offensichtlichen*“) Sinn („*sens obvie*“) unterscheidet. Der „*stumpfe Sinn*“ lässt sich in sprachlich-textlicher Form nur schwer fassen, und wenn, dann lediglich in Gegensätzlichkeiten. Auch Imdahl (1994: 300) versteht „das Bild als eine solche Vermittlung von Sinn, die durch nichts anderes zu ersetzen ist“, und sieht ebenfalls das Spezifische dieses Sinnes in einer „*Sinnkomplexität des Übergegensätzlichen*“ (Imdahl 1996: 107).

Während Imdahl diese Sinnkomplexität durchaus für beschreibbar hält, beharrt Barthes (1990: 63) darauf, dass man den „*stumpfen Sinn*“ „theoretisch situieren, aber nicht beschreiben kann“. „Der *stumpfe Sinn* ist nicht in der Sprache (nicht einmal in der der Symbole)“ (Barthes 1990: 58). Bei Barthes transzendierte der *stumpfe Sinn* den *entgegenkommenden Sinn* in ähnlicher Weise wie bei Panofsky die ikonologische Interpretation den (auf der Ebene von Common Sense-Typisierungen angesiedelten) ikonografischen Sinngehalt. Indem Panofsky (1975: 38) seine Methode auch am Beispiel des „*Alltagslebens*“ und nicht allein der Kunst erläutert, wird darin deren handlungs- und sozialwissenschaftliche Relevanz sichtbar: Die Gebärde eines Mannes, die auf der vor-ikonografischen Ebene als ein ‚Hutziehen‘ identifizierbar ist, kann – auf der ikonografischen Ebene – einerseits als ein ‚Grüßen‘ interpretiert werden. Andererseits wird diese Gebärde – in einer gegenüber dem Common Sense veränderten Analyseeinstellung – aber auch zum Ausdruck, zum Dokument für die „*Wesensart*“, den „*Wesenssinn*“ dieses Menschen (Panofsky 1932: 115), seinen „*Habitus*“ (Panofsky 1989) – sei dieser nun individueller (z.B.: ‚linkisches Wesen‘) oder „*kollektiver Art*“ (Ausdruck eines Milieus, einer zeitgeschichtlichen Phase oder einer Epoche).

Der Unterschied zwischen der ikonografischen und der ikonologischen Interpretation bei Panofsky lässt sich als Wechsel von der Frage nach dem *Was (geschieht hier)* zur Frage nach dem *Wie (der Herstellung)* dieser Gebärde charakterisieren. Panofsky ist mit dieser veränderten Analyseeinstellung der „*Dokumentarischen Methode*“ von Mannheim (1964) gefolgt, auf den er sich explizit bezieht (Panofsky 1932: 115). Auch Goffman (1979: 24) setzt mit der von ihm vorgelegten Fotointerpretation ganz wesentlich auf der Ebene von Gebärden auf der vor-ikonografischen Sinnebene an, die er – unterhalb der Ebene einer Interpretation von *Handlungen* (wie z.B. „*Grüßen*“) – als „*small behaviors*“ bezeichnet.

Im Bereich der *Fotointerpretation* ist es von wesentlicher Bedeutung, die Rekonstruktion der Gestaltungsleistungen der *abgebildeten* BildproduzentInnen (der AkteurInnen *vor* der Kamera) von denen der *abbildenden* BildproduzentInnen (der AkteurInnen *hinter* Kamera) zu unterscheiden (dazu genauer: Bohnsack 2009: 31)

---

## 67.5 Das Bild als selbstreferentielles System

Max Imdahl hat in der Tradition von Panofsky dessen Methode kritisch erweitert. Die besondere Leistung von Panofsky, den Habitus (einer Epoche) gerade aus den Analogien oder Homologien unterschiedlicher Medien oder Darstellungs- und Kunstgattungen (von der Literatur über die Malerei und Architektur bis zur Musik) hervortreten zu lassen, ist für Imdahl (1994, 1996) Ausgangspunkt seiner kritischen Frage danach, wo dann (noch) das Besondere des Mediums Bild zu suchen sei. Er kritisiert die reduzierte Bedeutung der formalen Komposition des Bildes bei Panofsky. Denn die Formalstruktur stellt die Grundlage der Vermittlung jenes Sinnes dar, der durch den Text und unser textförmiges (ikonografisches) Vor-Wissen nicht zu vermitteln ist. Imdahl hat – in der Rekonstruktion eigener umfangreicher forschungspraktischer Erfahrungen – seine Methode entwickelt, die „Ikonik“, welche die formale Komposition zum Ausgangspunkt nimmt. Die strikte Fundierung der Interpretation in der Formalstruktur hat ganz wesentlich den Sinn, sich von einer an der Textförmigkeit orientierten Sequentialität und Narrativität zu lösen, wie Imdahl (1996: 137) sie an den konventionellen Interpretationsverfahren kritisiert, und sich der für das Bild konstitutiven „Simultanstruktur“ zu nähern (Imdahl 1996: 23).

Die Rekonstruktion der formalen Komposition und der Simultanstruktur des Bildes (exemplarisch dargelegt in Bohnsack 2001, 2009: Kap. 42. u. 4.3) ist wesentliche Voraussetzung für den Zugang zum Bild als selbstreferentiellem System, um damit auch in der sozialwissenschaftlich-empirischen Analyse dem in der philosophischen Bildwissenschaft von Belting (2001:15) formulierten Anspruch gerecht zu werden, „Bilder nicht mehr mit Texten zu erklären, sondern von Texten zu unterscheiden“.

## Literatur

- Barthes, Roland (1990) [1982]: Das Bild. In: ibidem (Hg.): Der entgegenkommende und der stumpfe Sinn. Frankfurt a.M.: Suhrkamp. 11-66
- Belting, Hans (2001): Bild-Anthropologie. München: Fink
- Boehm, Gottfried (1994): Die Wiederkehr der Bilder. In: ibidem (Hg.): Was ist ein Bild? München: Fink. 12-38
- Bohnack, Ralf (2001): „Heidi“: Eine exemplarische Bildinterpretation auf der Basis der dokumentarischen Methode. In: Bohnack, Ralf/Nentwig-Gesemann, Iris/Nohl, Arnd-Michael (Hg.): Die dokumentarische Methode und ihre Forschungspraxis. Opladen: Leske + Budrich. 225-252
- Bohnack, Ralf (2003): Qualitative Methoden der Bildinterpretation. In: Zeitschrift für Erziehungswissenschaft 6 (2): 239-256
- Bohnack, Ralf (2009): Qualitative Methoden der Bild- und Videointerpretation. Opladen & Farmington Hills: Barbara Budrich
- Eco, Umberto (1994): Einführung in die Semiotik. München: Fink
- Foucault, Michel (1971) [1966]: Die Ordnung der Dinge. Frankfurt a. M.: Suhrkamp
- Goffman, Erving (1979). Gender Advertisements. New York et al.: Harper & Row
- Imdahl, Max (1994): Bilder und ihre Anschauung. In: Boehm, Gottfried (Hg.): Was ist ein Bild? München: Fink. 300-324
- Imdahl, Max (1996): Giotto – Arenafresken. Ikonographie – Ikonologie – Ikonik. München: Fink
- Loer, Thomas (1994): Werkgestalt und Erfahrungskonstitution. In: Garz, Detlef/Kraimer, Klaus (Hg.): Die Welt als Text. Frankfurt a. M.: Suhrkamp. 341-382
- Mannheim, Karl (1964): Beiträge zur Theorie der Weltanschauungsinterpretation. In: ibidem (Hg.): Wissenssoziologie. Neuwied/Berlin: Luchterhand: 91-154 (Original: 1921-1922 in: Jahrbuch für Kunstgeschichte XV, 4)
- Mitchell, William J.T. (1994): Picture Theory. Chicago/London: University of Chicago Press
- Mollenhauer, Klaus (1983): Streifzug durch fremdes Terrain. In: Zeitschrift für Pädagogik 30 (2): 173-194
- Müller-Doohm, Stefan (1997): Bildinterpretation als struktural-hermeneutische Bildanalyse. In: Hitzler, Ronald/Honer, Anne (Hg.): Sozialwissenschaftliche Hermeneutik. Opladen: Leske+Budrich. 81-109
- Panofsky, Erwin (1932): Zum Problem der Beschreibung und Inhaltsdeutung von Werken der Bildenden Kunst. In: Logos XXI: 103-119
- Panofsky, Erwin (1975) [1955]: Ikonographie und Ikonologie. In: ibidem (Hg.): Sinn und Deutung in der bildenden Kunst. Köln: DuMont. 36-67
- Panofsky, Erwin (1989): Gotische Architektur und Scholastik. Köln: DuMont
- Pilarcyk, Ulrike/Mietzner, Ulrike (2005): Das reflektierte Bild. Bad Heilbrunn
- Popper, Karl (1971): Logik der Forschung. Tübingen: Mohr

- Prosser, Jon (Hg.) (1998): *Image-based Research: A Sourcebook for Qualitative Researchers.* London: Falmer Press
- Raab, Jürgen (2008): *Visuelle Wissenssoziologie.* Konstanz: UVK
- Reichertz, Jo (1992): Der Morgen danach. Hermeneutische Auslegung einer Werbefotografie in zwölf Einstellungen. In: Hartmann, Hans A./Haubl, Rolf (Hg.): *Bilderflut und Sprachmagie.* Opladen: Westdeutscher Verlag. 141-163
- Ricoeur, Paul (1978): Der Text als Modell. In: Gadamer, Hans-Georg/Boehm, Gottfried (Hg.): *Die Hermeneutik und die Wissenschaften.* Frankfurt a. M.: Suhrkamp. 83-117

René Tuma und Bernt Schnettler

## 68.1 Was ist Videoanalyse?

Videoanalyse bezeichnet eine Reihe von methodischen Verfahren, bei denen audiovisuelle Daten für sozialwissenschaftliche Untersuchungen verwendet werden. Videographien basieren in der Regel auf Aufzeichnungen, die von den Forschenden selbst zum Zwecke wissenschaftlicher Analyse angefertigt werden. Technische und methodische Entwicklungen haben in jüngerer Zeit solchen Verfahren Aufschwung verliehen und mit der Verbreitung erschwinglicher digitaler Kameratechnik ist der Weg zu ihrer intensiveren Verwendung in der sozialwissenschaftlichen Forschung geebnet worden.

Grundsätzlich ist zwischen standardisierenden und interpretativen Videoanalysen zu unterscheiden. Mit *standardisierenden Videoanalysen*, wie sie etwa in der Experimentalforschung (Eifler, Kapitel 11 in diesem Band) Einsatz finden, werden die Aufzeichnungen einer Auswertung unterzogen, die sich an vorab festgelegten Kodierplänen orientiert (Mayring/Fenzl, Züll/Menold, Kapitel 38 und 52 in diesem Band) und das Material nach dem Auftreten bestimmter Ereignisse durchsucht (wie etwa „prosoziales Verhalten“ oder „Sprecherwechsel“). Standardisierende Videoanalysen sind vor allem in der Psychologie und den Erziehungswissenschaften verbreitet, wohingegen in der Soziologie interpretative Verfahren (Kurt/Herbrik, Kapitel 33 in diesem Band) dominieren.

Mit *interpretativen Videoanalysen* wird demgegenüber induktiv (Reichert, Kapitel 2 in diesem Band) gearbeitet und darauf gezielt, entweder die Videos als sinnhafte Produkte oder das aufgezeichnete Miteinander-Handeln von Akteuren zu verstehen. Obwohl sich inzwischen bereits mehrere methodologisch voneinander unterscheidbare Varianten interpretativer Videoanalyse etabliert haben, darf der gegenwärtige Stand der Methodenentwicklung als noch keineswegs abgeschlossen betrachtet werden. Einige Verfahren der interpretativen Videoanalyse analysieren (a) medial vorliegende und verbreitete audiovisuelle Formate, wie beispielsweise Hochzeitsvideos, Fernsehsendungen oder Internet-Videoclips, um sie auf ihre Sinngehalte hin auszuwerten; (b) andere verwenden von den

Forschenden selbst erhobene Daten. Während Erstere eher der Medienanalyse (Akremi, Kapitel 69 in diesem Band) zuzurechnen sind, gehören Letztere in den weiten Bereich ethnographischer Forschung (Knoblauch, Kapitel 36 in diesem Band).

Die *Videographie* ist demzufolge ein interpretativer methodischer Ansatz, der sich – wie der Name bereits indiziert – durch eine enge Einbettung in ethnographische Feldforschung (Knoblauch, Thierbach/Petschick, Kapitel 69 und 66 in diesem Band) auszeichnet. Sie zielt auf die Untersuchung sozialer Situationen, von denen Videoaufnahmen angefertigt werden. Videographische Forschung fokussiert dabei die Aufzeichnung und Analyse sozialer Interaktionen in den „natürlichen“ Kontexten, in denen sie vorkommen. Der methodische Vorzug besteht darin, dass die erhobenen Videodaten den *Ablaufcharakter der aufgezeichneten Interaktionen bewahren* und eine anschließende, sehr *minutiöse Detailauswertung von Geschehensabläufen in ihrem Zusammenhang* erlauben.

Gegenüber den üblicherweise in der klassischen Ethnographie verwendeten *rekonstruktiven Daten* wie Feldnotizen (Thierbach/Petschick, Kapitel 66 in diesem Band) oder Interviewdaten (Helfferich, Küsters, Kapitel 39 und 40 in diesem Band) gestatten Videodaten also eine „Konservierung“ von Geschehensaspekten in ihrem zeitlichen Ablauf. Im Gegensatz zu rekonstruktiven Daten geben Videoaufzeichnungen Forschenden damit einen wiederholten Zugriff auf vergangene Geschehensabläufe in der Gestalt, wie sie sich original entfaltet haben und lassen sich strikt sequenziell in der komplexen Ordnung ihres Ablaufs analysieren. Damit eröffnen sich vor allem für die Kommunikations- und Interaktionsforschung, aber auch für die Exploration von Milieus und sozialen Welten, vollkommen neue Forschungsperspektiven.

---

## 68.2 Verbreitung und Varianten der Videoanalyse

Das Feld der verschiedenen videoanalytischen Verfahren befindet sich in sehr dynamischer Entwicklung, was mit einer gewissen Unübersichtlichkeit einhergeht. Hinzu kommt, dass die bereits erkennbar etablierten Varianten der Videoanalyse auf voneinander mehr oder weniger abweichenden Grundlagentheorien aus einem breiter werdenden Spektrum interpretativer Forschungsansätze basieren. Sie sind außerdem mit verschiedenen materialen Forschungsgebieten eng verzahnt, was ihre je eigenen Ausprägungen erklärt. Dazu zählen vor allem die Interaktionsforschung, die Arbeitsforschung, die Technikforschung und die Bildungsforschung, aus deren Forschungserfordernissen sich die jeweiligen methodologischen Akzentuierungen ableiten. Innerhalb der Sozialwissenschaften sind Videoanalysen vor allem im Bereich erziehungswissenschaftlicher Unterrichtsforschung, der Analyse von technisierten Arbeitskontexten (Technographie, Workplace Studies) und der Kommunikations- und Interaktionsforschung sehr verbreitet. In theoretischer Hinsicht ist die videographische Forschung insbesondere mit interaktions- und kommunikationsssoziologischen Theoriekonzepten oder mit Praxistheorien verbunden. Die derzeit vorhandenen Ansätze lassen sich typologisch drei verschiedenen videoanalytischen Strängen zuordnen:

### 68.2.1 Hermeneutische Verfahren

Die *wissenssoziologische Hermeneutik* (Kurt/Herbrik, Kapitel 33 in diesem Band) hat ein Auslegungsverfahren entwickelt, bei dem die hermeneutische *Sequenzanalyse* als methodische Kunstlehre der Deutung eine tragende Rolle spielt. Die methodologischen Grundlagen der wissenssoziologischen Hermeneutik beziehen sich ausdrücklich auf die Tradition einer verstehenden Soziologie in der Nachfolge Max Webers und verbinden diese mit philologischen Mitteln der Textauslegung, wie sie in den Geisteswissenschaften entfaltet worden sind (Soeffner 2004, Reichertz 2007). Gemeinsam ist allen hermeneutischen Ansätzen die eingehende *Ausdeutung einzelner Materialabschnitte* mit dem Ziel einer erschöpfenden Entfaltung aller möglichen soziologisch denkbaren *Lesarten*. Dabei werden in einer Interpretationsgruppe zunächst alle erdenklichen Bedeutungen der in den Fokus genommenen einzelnen Bildausschnitte oder kurzen Videosequenzen entfaltet. In einem darauf folgenden zweiten systematischen Schritt werden diese Sinnpotenziale anhand nachfolgender Bild- bzw. Videosequenzen sukzessive reduziert und schließlich in eine Strukturhypothese kondensiert (Reichertz/Englert 2010).

Raab (2008: 210ff.) zeigt beispielsweise in seiner Studie zu Hochzeitsvideos anhand der genauen Analyse ausgewählter Einzelbilder die spezifischen Inszenierungsweisen dieser audiovisuell repräsentierten Zeremonie. Die Darstellung eines Straußes roter Rosen wird genauestens ausgelegt und auf die damit verknüpften, auch impliziten oder nur konnotierten Sinngehalte hin untersucht. Nach einer erschöpfenden Beschreibung wird herausgearbeitet, wie sie „ins rechte Licht gerückt und zusätzlich so arrangiert [werden], dass sich die Objekte sowohl in ihrer identischen Erscheinung, in der Wiederholung ihrer Farben und Formen, ergänzen und gegenseitig bestätigen (Rosen-Rosen, Rot-Rot, Ringe-Ringe, Gold-Gold)“ (Raab 2008). Diese Darstellung wird nun verglichen mit nachfolgenden Bildern, u.a. der Abbildung eines sich küssenden Hochzeitspaars, das in seiner symmetrischen Darstellung die für dieses Video spezifischen Grundelemente der *wechselseitigen Verweisung* und der *Gliederung der Grundgestalt* offenbart. So werden schrittweise alle möglichen Lesarten des audiovisuellen Dokuments entfaltet, um auch an die nicht offenkundigen und im Material verborgenen Sinngehalte zu gelangen.

Die Erstellung dieser Interpretationen wird in *Interpretationsgruppen* durchgeführt und folgt dem methodischen Prinzip hermeneutischer Sequenzanalyse (Maiwald 2005), nach dem strikt in der Reihe der aufeinander folgenden *Sinneinheiten* (wie etwa Sequenzen einer bestimmten Dauer oder Einzelbilder) (Akremi, Kapitel 17 in diesem Band) gedeutet wird. Die Videohermeneutik erweist sich als besonders leistungsfähig bei der Analyse edierter Videomaterialien, lässt sich also materialunspezifisch auch für medial vermittelte audiovisuelle Ausdrucksformen verwenden. Dabei berücksichtigt sie die im Material eingearbeiteten Dimensionen ästhetisierter Sinnformationen (Raab 2008). Dem liegt die feste Auffassung zu Grunde, dass alle kulturellen Ausdrücke, gleich welcher Art, immer vieldeutig sind und einer Deutung bedürfen. Das Interpretationsverfahren greift also letztlich auf basale, allgemeine menschliche Deutungsfähigkeiten zurück und nutzt sie, methodisch

kontrolliert und elaboriert, als Grundlage sozialwissenschaftlicher Analyse. In jüngerer Zeit sind außerdem Ansätze entwickelt worden, um die Hermeneutik für weitere Bereiche videografischer Forschung fruchtbar zu machen (Kissmann 2014).

### **68.2.2 Videoanalysen nach der Dokumentarischen Methode**

Ebenso wie die sozialwissenschaftliche Hermeneutik betont auch die dokumentarische Methode die grundsätzliche Interpretativität der Sozialwelt und hat ein dementsprechendes konzeptuelles Programm mit breit rezipierten Methoden etabliert. Das von Bohnsack (2009 und Kapitel 67 in diesem Band) entwickelte, an Karl Mannheim und Erwin Panowski anknüpfende Verfahren der *dokumentarischen Bild- und Videointerpretation* wird vor allem im deutschsprachigen Raum intensiv praktiziert. Ziel des kunsthistorisch informierten Verfahrens ist die Rekonstruktion des in Kulturprodukten verborgenen Sinnes, der als *Dokumentsinn* bezeichnet wird. Die Grundannahme lautet, dass jeder Handlung oder Aussage und jedem Handlungsprodukt ein Dokumentsinn zugrunde liegt. Dieser im Material enthaltene Dokumentsinn geht über den immanenten Sinngehalt hinaus. Er wird von den (vor der Kamera) abgebildeten Handelnden und den (hinter der Kamera) abbildenden Bildproduzenten reproduziert, ohne dass er beiden überhaupt „bewusst“ werden muss. Er ist in diesem Sinne „atheoretisch“.

Die Rekonstruktion des dokumentarischen Sinns läuft in mehreren Schritten ab:

1. Sie beginnt mit einer „*formulierenden Interpretation*“,
2. der eine „*reflektierende Interpretation*“ (der *formalen Bildkomposition*) folgt und
3. mündet schließlich in eine „*Typisierung und Generalisierung*“.

Am Beispiel eines von Schülern produzierten Kurzfilms über Schule schildern Bohnsack und Baltruschat (2010: 39) zunächst eine Lehrerin, die „hinter aufgetürmten Papierstößen [sitzt] und (...) mit beiden Händen abwechselnd weitere Papierstapel nach vorne [transportiert]“. Methodisch wird also zunächst beschrieben, was in den Videos – analysiert häufig an Einzelbildern – zu sehen ist, bevor im nächsten Schritt die Darstellungsweise interpretiert wird: In diesem Fall als nach vorne greifende „Hand eines Ertrinkenden, der sich ans rettende Ufer ziehen möchte“. Bei der Analyse von einzelnen Standbildern werden häufig Fotogramme zur genaueren Perspektivenrekonstruktion angefertigt, um etwa perspektivische Zentren, Fluchtpunkte oder Bildachsen zu identifizieren. Im Vergleich mit weiteren Ausschnitten aus dem Film werden im oben erwähnten Beispiel die den Produzenten des Videos (vor und hinter der Kamera) impliziten (atheoretischen) Wissensbestände rekonstruiert und auf ein habitualisiertes Wechselspiel von „Identifikation und Distanzierung“ mit der Schule zurückgeführt. Die dokumentarische Methode wurde bislang ebenfalls vornehmlich auf produzierte Videos (z.B. von Schülern erstellte Filme, TV Sendungen) angewendet. Vor allem in der Schul- und Unterrichtsforschung wird sie mitunter jedoch ebenfalls zur Analyse der von Forschenden selbst erhobenen Videodaten eingesetzt (Fritzsche/Wagner-Willi 2013).

### 68.2.3 Videointeraktionsanalyse

Der dritte Strang der Entwicklung interpretativer Videoanalyseverfahren (Heath 1986, Heath et. al. 2010) beruht auf Theorieannahmen der *Ethnomethodologie* (Garfinkel 1967) und beruft sich in seinen Grundlagen auf die *Konversationsanalyse* (Sacks 1992). Die Ethnomethodologie geht davon aus, dass die Vertrautheit, Geordnetheit und Faktizität unserer Alltagswelt Resultate einer Leistung der miteinander Interagierenden bzw. der dabei von ihnen verwendeten "Methoden" sind. In der ethnomethodologisch fundierten Videoanalyse geht es methodologisch um zwei Kernpunkte:

1. Um die Bestimmung der *Ressourcen, des Wissens und der praktischen Überlegungen*, die von den Interagierenden selbst bei der Hervorbringung ihrer *in situ* stattfindenden sozialen Handlungen und Aktivitäten genutzt bzw. verfolgt werden.
2. Um die Erforschung der *sequenziellen Ordnung der Interaktionen*, deren minutiöse Rekonstruktion dazu dient herauszufinden, wie sich die Handelnden aneinander orientieren und ihre Interaktionen miteinander koordinieren.

Diese Vorgehensweise ist von einem ethnomethodologischen Verständnis der Sequenzanalyse geprägt (Bergmann 1981, ten Have 1999). Um das an einem jüngeren Beispiel zu erläutern: Heath (2012) analysiert mit Hilfe von Videoaufzeichnungen nicht nur das feingliederige Zusammenspiel von Geboten, Zeigegesten und dargebotenen Gegenständen, die das Geschehen und die „Dynamik von Auktionen“ auszeichnen und nachvollziehbar werden lassen. Er kann dabei darüber hinaus diejenigen elementaren Interaktionsprozesse rekonstruieren, die in der Summe für die Preisbildung auf dem weltumspannenden Kunstmarkt verantwortlich sind.

---

## 68.3 Methodisches Vorgehen bei der Videographie

Die Videographie, der wir uns im Folgenden vertiefend zuwenden, ist aus dem letzten der vorgenannten Ansätze entstanden und zusätzlich stark von der *soziologischen Theorie kommunikativer Gattungen* (Luckmann 1986, Knoblauch 1995) beeinflusst. Letztere stellt die empirische wie konzeptuelle Fortsetzung des wissenssoziologischen Programms der gesellschaftlichen Konstruktion der Wirklichkeit (Berger/Luckmann 1969) dar.

Das besondere Merkmal der Videographie besteht in der Verknüpfung von „fokussierter“ *Feldforschung* (Knoblauch, Thierbach/Petschick, Kapitel 69 und 66 in diesem Band) mit anschließenden extensiven und minutiösen *Materialanalysen* der von den Forschenden angefertigten Videoaufzeichnungen nach den Prinzipien der Videointeraktionsanalyse. Einzelne aufgezeichnete Interaktionssequenzen werden vor dem Hintergrund ethnographisch gewonnenen Wissens analysiert. Die Kombination beider Herangehensweisen im Forschungsprozess erlaubt es, sehr genaue Rekonstruktionen einzelner Situationen vorzunehmen und deren faktischen Kontext bei der Mikroanalyse systematisch mit zu berücksichtigen (Knoblauch 2004).

Die videographische Datenerhebung folgt keinem vorab vollends fixierten Forschungsplan, sondern einer iterativen Verfahrensweise (Przyborski/Wohlrab-Sahr, Strübing, Kapitel 6 und 32 in diesem Band). Ausgehend von einem mehr oder weniger präzisierten Anfangsinteresse führt dies zu einer schrittweisen Engführung der Fragestellung und einer damit parallel laufenden Ausdehnung eines langsam wachsenden Datenkorpus (Akremi, Kapitel 17 in diesem Band). Das entspricht in besonderer Weise dem grundlegenden Postulat der Offenheit und Kontextsensitivität interpretativer Sozialforschung.

### **68.3.1 Zusammenstellung des Datenkorpus**

Dazu werden im ersten Schritt im bereits ethnographisch erschlossenen Feld relevante Ereignisse ausgewählt und mit der Kamera aufgezeichnet. Dies können zum Beispiel Verkaufsgespräche auf einem Straßenmarkt (Tuma et. al. 2013: 85ff.), die Arbeit von Anästhesisten bei Operationen im Krankenhaus (Schubert 2006) oder das Geschehen beim Vortragen von Powerpoint-Präsentationen (Knoblauch 2012) sein. Diese Aufzeichnungen werden einem *Datenkorpus* zugeführt, geordnet und überblicksartig durch Übersichtsprotokolle und *Daten-Logbücher* aufgeschlossen, um sie für die Feinanalyse vorzubereiten.

Anders als bei hermeneutischen Verfahren (Kurt/Herbrik, Kapitel 33 in diesem Band) ist das Vorgehen bei der Videographie also nicht als Fallanalyse (Hering/Schmidt, Kapitel 37 in diesem Band) angelegt, sondern bemüht sich um den systematischen Aufbau von Datensammlungen, um Vergleiche innerhalb des Materials zu bestimmen und die Grenzen des Phänomenbereichs auszuloten – etwa, indem bei der Suche nach wiederkehrenden Strukturmerkmalen von Powerpoint-Präsentationen im Vergleich zu mündlichen Vorträgen nicht allein eine vorab festgelegte Anzahl solcher Präsentationen gefilmt wird, sondern zum Zwecke des Vergleichs und der Abgrenzung ebenfalls rein mündliche Vorträge, Vorträge mit Tafelbildern, mit Overheadprojektoren, Power-Point-Karaoke usw. in den Datenkorpus mit aufgenommen werden.

### **68.3.2 Auswahl der zu analysierende Sequenzausschnitte**

Einem theoretischen Sampling (Strübing, Kapitel 32 in diesem Band) folgend, werden dann einzelne kurze Sequenzausschnitte aus dem Material ausgewählt: zum Beispiel einzelne Geldübergaben auf dem Straßenmarkt, Krisensituationen bei der Operation oder Pannen bei der Präsentation. Die Auswahl von Sequenzen folgt dabei einer doppelten Perspektive. Zum einen wird sie geleitet durch Anfangsvermutungen, für die im Material nach Belegen gesucht wird: Zeigt sich empirisch tatsächlich eine erkennbare und wiederholte typische Ablaufstruktur für das zu untersuchende Phänomen? – Etwas: geht die Frage nach dem Preis dem Bezahlen auf dem Straßenmarkt immer voraus? Lassen sich abweichende Fälle im Material finden oder ist die Struktur womöglich ganz anders? Zum anderen können zuvor unbekannte, aber im Material wiederkehrende Merkmale entdeckt und in den

Mittelpunkt der folgenden Analyseschritte gerückt werden. So etwa die Beobachtung, dass *Zeigen* für die Performanz bei Powerpoint-Präsentationen eine zentrale Rolle spielt und neben den bekannten Instrumenten (Laserpointer, Zeigestock) empirisch auch und sogar vor allem durch die Körperpositionierung des Vortragenden geleistet wird.

### 68.3.3 Feinanalyse

Die systematisch ausgewählten Sequenzen werden im nächsten Schritt einer minutiösen Feinanalyse unterzogen. Zugleich wird die Forschungsfrage schrittweise weiter fokussiert. Jetzt werden die sequentiellen Ablaufstrukturen exakt nachvollzogen, wobei der zeitliche Ablauf und die Orchestrierung der einzelnen Äußerungen der Beteiligten als Handlungszüge detailliert rekonstruiert werden müssen: Wohin richtet zum Beispiel der Vortragende seinen Blick, während er die nächste Folie einblendet? Geht der Klick mit der Maus dem Umdrehen der Manuskriptseiten voraus, erfolgt er gleichzeitig oder anschließend? Diese Beobachtungen am Material konzentrieren sich nicht allein auf die verbalen Äußerungen, sondern nehmen alle weiteren kommunikativen Ausdrucksmodalitäten, wie etwa körperliche und mit technischen Objekten durchgeführte Handlungszüge oder Blicke und Augenbewegungen, mit in den Fokus, *sofern* diese für die von den beobachteten Akteuren ausgeführten Interaktionen *selbst relevant* sind. Es ist also wenig dienlich und methodisch abwegig, bestimmte Modalitäten (gestische, mimische, korporale, lokomotorische) zu transkribieren und bei der Analyse prinzipiell zu berücksichtigen. Vielmehr ist es umgekehrt so, dass zunächst die Relevanz der je einzelnen Ausdrucksmodalitäten für die an der Interaktion Beteiligten nachzuweisen ist. Die Rekonstruktion der je relevanten kommunikativen Äußerungsarten und ihre Koordinierung erfolgt also unter Verzicht auf die Vorfestlegung vermeintlich immer relevanter "kommunikativer Kanäle", also kontextsensitiv und fallgenau. In diesem Sinne folgt die Auswertung keiner vorab festgelegten Schematik, sondern erfordert vor allem eine sehr genaue Beobachtung und deren zunehmende Systematisierung. Das folgende Datenfragment aus unserer Forschung zu technischen Pannen mag der Illustration für das Vorgehen bei der Feinanalyse dienen. Es handelt sich um den Auftakt einer Begrüßungssequenz, in der der Redner seine Ansprache unterbrechen muss (Z. 7–10, Transkript Fragment 1):

Fragment 1 (vereinfachtes Transkript)

- 1        auch ich begrüße Sie zum workshop
- 2        ›bedürfnisse erfüllen, risiken minimieren‹,
- 3        in dem wir die bisherigen erfahrungen
- 4        im aktionsfeld vier des bmbf
- 5        rahmenprogrammsforschung für die
- 6        nachhaltigkeit diskutieren wollen
- 7        (4.0) ähm (1.0)
- 8        [leise zur Assistentin gerichtet:]

- 9 °die vielleicht, jawoll, (startet) schon° ähm,
- 10 als schon ähm,
- 11 als wir 2003 im bmbf
- 12 das neue rahmenprogramm entwickeln

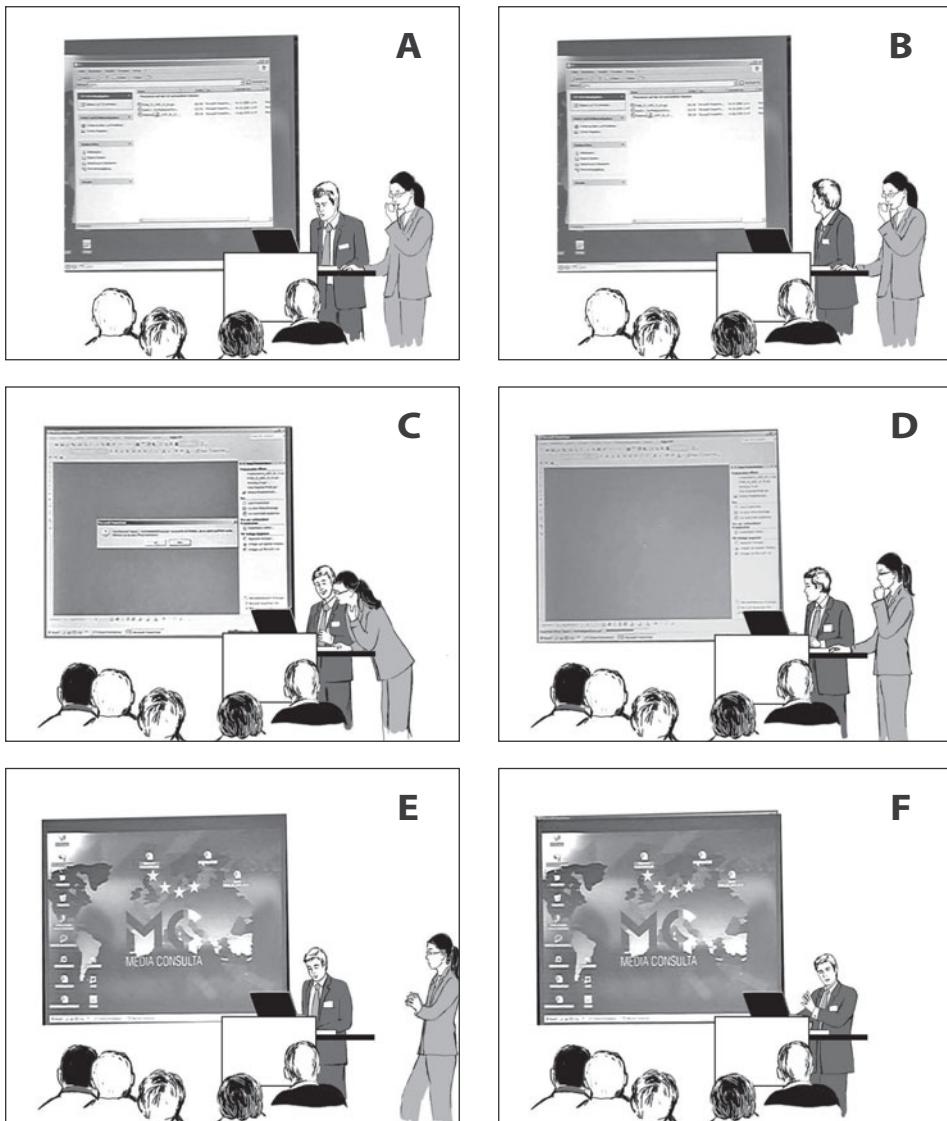
An diesem Materialausschnitt kann die besondere „situative Orchestrierung“ der eingesetzten körperlichen und sprachlichen Mittel aufgedeckt werden, mit denen eine kleine Präsentationspanne im Zusammenspiel von Technik, Redner und Assistentin kollaborativ bewältigt wird: Als der Redner mit seiner Begrüßung ansetzt, stellen sich Probleme beim Starten der Präsentation ein. Während der Redner seinen Vortrag beginnt (Abb. 68.1 A), versucht seine Assistentin das entsprechende Programm zu starten. Das gelingt allerdings nicht und führt dazu, dass der Redner seine einleitenden Worte unterbricht und sich vom Publikum ab und der Leinwand kurz zuwendet (B). Seine Assistentin versucht, über den Computer gebeugt, das technische Problem zu lösen (C). Während der Sprecher seine Ansprache fortzusetzen versucht, sind die damit interferierenden Versuche der Assistentin erkennbar, die Powerpointpräsentation zu starten (D). Nach mehrfachen Anläufen gelingt diese und die Assistentin zieht sich vom Rednerpult zurück (E). Schließlich kann der Sprecher seine Rede wie geplant synchron zu der Projektion fortführen (F).

Im Mittelpunkt der Analyse steht nun die Frage, wie technische Pannen als solche sichtbar (gemacht) werden und ob bzw. wie so etwas wie „Peinlichkeit“ entsteht. Es wird herausgearbeitet, wie dies von den Akteuren durch ihre Handlungszeuge kommunikativ hergestellt wird und welche sozialen Formen der Problemlösung in Pannensituationen zur Anwendung kommen (vgl. Schnettler/Tuma 2007 für den empirischen Fall).

Der abgedruckte Transkriptausschnitt und die Sequenzbilder bieten allerdings eine nur unzureichende Darstellung von Datum und Auswertung. Vor allem gehen der im Video enthaltene Ablaufcharakter und die Verschränkung von Simultaneität und Sequenzialität verloren. Hier wird ein prinzipielles und noch weitgehend ungelöstes Publikationsproblem videographischer Forschung offenkundig. Behelfsmäßige Versuche, die Original-videofragmente online verfügbar zu machen (vgl. für einige Sequenzbeispiele [www.soz.uni-bayreuth.de/de/videoanalysis/index.html](http://www.soz.uni-bayreuth.de/de/videoanalysis/index.html)) stoßen auf Grenzen und unterliegen zudem speziellen Anforderungen an die Datenfreigabe.

## 68.4 Videotranskription

Die Videotranskription ist ein für die Feinanalyse unverzichtbarer Verfahrenzwischenschritt bei der Untersuchung audiovisueller Daten. Entsprechend den Auswertungszielen stehen verschiedene Transkriptionsverfahren zur Verfügung, von denen das *konversationsanalytisch fundierte, zeilenbasierte GAT 2 (Gesprächsanalytisches Transkriptionssystem*, Selting et al. 2009) zu den erprobtesten zählt. Für die Darstellung von Interaktionen werden diese textbasierten Transkripte zumeist mit spezifischen Ergänzungen erweitert, die auch andere „Modalitäten“ wie Gesten, Blicke usw. abbilden können. Ebenso stehen



**Abb. 68.1** Kollaborative Problemlösung einer Präsentationspanne (Quelle: eigene Darstellung)

Partiturschreibweisen und entsprechende Software zur Verfügung (Moritz 2011). Visuelle Transkriptionsweisen befinden sich noch in der Entwicklung.

Zentraler Arbeitsschritt der gängigen videographischen Verfahren ist die gemeinsame Feinauswertung der ausgewählten Sequenzen in Datensitzungen oder Forschungswerkstätten. Hier kann ethnographisches Wissen kontrolliert eingebracht und die Interpretation

entwickelt, verfeinert und überprüft werden. Vor allem in Feldern, in denen den Forschenden die Spezialkenntnisse über die vor ihren Augen vonstatten gehenden Geschehnisse fehlen (wie zum Beispiel in hochtechnisierten Arbeitsumfeldern), kann es analytisch sehr produktiv sein, die Forschungspartner mit einzubinden (Schubert 2006) bzw. mit ihnen gemeinsam das Videomaterial zu betrachten, um dieses Sonderwissen zu elizitieren.

---

## 68.5 Grenzen und Herausforderungen

Die Videographie bleibt trotz maßgeblicher technischer Vereinfachungen ein komplexer und voraussetzungsvoller Forschungsansatz, der im Zuge der wachsenden Durchdringung der Alltagskultur mit Visualisierungen einen erheblichen Aufschwung erfahren hat. Die mimetische Fähigkeit der technischen Apparaturen darf allerdings nicht dazu verleiten, Videodaten als objektive Repräsentationen zu betrachten. Sie stellen Daten dar, in die immer die Konstruktionsleistungen der Forschenden mit eingehen, was bereits mit der *Auswahl des Ortes und Zeitpunktes des Filmens* beginnt und sich in der *Wahl des Bildausschnittes* (Akremi, Kapitel 17 in diesem Band) und der *Kameraführung* sowie in den *Reaktionen auf die Kamera im Feld* (Thierbach/Petschick, Kapitel 66 in diesem Band) manifestiert. Das Problem der *Reaktanz* (Hlawatsch/Krickl, Kapitel 20 in diesem Band) ebenso wie mögliche *forschungsethische Hindernisse* (Friedrichs, Mühlichen, Kapitel 3 und 4 in diesem Band) stellen in dieser Hinsicht bei jeder videographischen Forschung zu beachtende Herausforderungen dar, wenngleich keineswegs prinzipielle Hindernisse. Zu den bislang noch ungelösten Problemen zählt schließlich die Überwindung der offenkundigen Unzulänglichkeiten bei der *Publikation von Ergebnissen* videographischer Forschungen (Meyer/Meier zu Verl, Kapitel 15 in diesem Band), die sich in den bisher gängigen und akzeptierten textlichen Publikationsformaten nur sehr unvollkommen wiedergeben lassen.

Für diese wie für andere Fragen finden sich möglicherweise bislang unausgeschöpfte Potentiale in der Betrachtung spezifischer Formen der Videoanalyse, die außerhalb der Wissenschaft bzw. in Grenzgebieten alltäglich angewendet werden (etwa in der Marktforschung, der Polizeiarbeit oder im Sport), und die als “*Vernacular Video Analysis*” (Tuma 2012) bezeichnet werden können.

## Literatur

- Berger, Peter L./Luckmann, Thomas (1969): Die gesellschaftliche Konstruktion der Wirklichkeit. Frankfurt a.M.: Fischer
- Bergmann, Jörg (1981): Ethnomethodologische Konversationsanalyse. In: Schröder, Peter/Steger, Hugo (Hg.): Dialogforschung. Düsseldorf: Schwann. 9-51
- Bohnsack, Ralf (2009): Qualitative Bild- und Videointerpretation. Stuttgart: UTB
- Bohnsack, Ralf/Baltruschat, Astrid (2010): Die dokumentarische Methode. In: Enzyklopädie Erziehungswissenschaften Online (EEO). Weinheim/München: Juventa. URL: erzissenonline.de; DOI 10.3262/EEO07100072
- Fritzsche, Bettina/Wagner-Willi, Monika (2013): Ethnografie und Videografie in praxeologischer Perspektive. In: Bohnsack, Ralf/Baltruschat, Astrid/Fritzsche, Bettina/Wagner-Willi, Monika (Hg.): Dokumentarische Video- und Filminterpretation. Farmington: Barbara Budrich. 268-283
- Garfinkel, Harold (1967): Studies in Ethnomethodology. Englewood Cliffs: Prentice Hall
- Heath, Christian (1986): Body Movement and Speech in Medical Interaction. Cambridge: Cambridge University Press
- Heath, Christian/Hindmarsh, Jon/Luff, Paul (2010): Video in Qualitative Research. London: Sage
- Kissmann, Ulrike Tikvah (2014): Die Sozialität des Visuellen. Weilerswist: Velbrück Wissenschaft (im Erscheinen)
- Knoblauch, Hubert (1995): Kommunikationskultur. Berlin: De Gruyter
- Knoblauch, Hubert (2004): Die Video-Interaktions-Analyse. In: Sozialer Sinn 1: 123-138
- Knoblauch, Hubert (2012): Powerpoint, Communication, and the Knowledge Society. Cambridge: University Press
- Luckmann, Thomas (1986): Grundformen der gesellschaftlichen Vermittlung des Wissens: Kommunikative Gattungen. In: Sonderheft 27 der Kölner Zeitschrift für Soziologie und Sozialpsychologie: 191-211
- Maiwald, Kai-Olaf (2005): Competence and Praxis. In: Forum: Qualitative Social Research 6(3): Art. 31. URL: <http://nbn-resolving.de/urn:nbn:de:0114-fqs0503310>
- Moritz, Christine (2011): Die Feldpartitur. Wiesbaden: VS Verlag
- Raab, Jürgen (2008): Visuelle Wissenssoziologie. Konstanz: UVK
- Reichert, Jo (2007): Hermeneutische Wissenssoziologie. In: Schützeichel, Rainer (Hg.) (2007): Handbuch Wissenssoziologie und Wissensforschung. Konstanz: UVK: 171-180
- Reichert, Jo/Englert, Carina J. (2010): Einführung in die qualitative Videoanalyse. Wiesbaden: VS Verlag
- Sacks, Harvey (1992): Lectures on Conversation. Oxford: Blackwell. 1964ff.
- Schnettler, Bernt/Tuma, René (2007): Pannen –Powerpoint– Performanz. In: Schnettler, Bernt/Knoblauch, Hubert (Hg.): Powerpoint-Präsentationen. Konstanz: UVK. 163-188
- Schubert, Cornelius (2006): Die Praxis der Apparatemedizin. Frankfurt a.M.: Campus

- Selting, Margret, et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT2). In: Gesprächsforschung 10: 223-272. URL: <http://www.gespraechsforschung-ozs.de>
- Soeffner, Hans-Georg (2004): Der Alltag der Auslegung. Konstanz: UTB
- Ten Have, Paul (2000): Doing Conversation Analysis. London: Sage
- Tuma, René (2012): The (Re)Construction of Human Conduct: "Vernacular Video Analysis". In: Qualitative Sociology Review 8: 152-163
- Tuma, René/Schnettler, Bernt/Knoblauch, Hubert (2013): Videographie. Wiesbaden: VS Verlag

Leila Akremi

---

## 69.1 Was ist sozialwissenschaftliche Filmanalyse?

Wenn Sozialwissenschaftler sich entscheiden, Filme als Datenquelle zu benutzen, kann dies folgende Gründe haben:

1. Die Fragestellung bezieht sich auf längerfristigen sozialen Wandel und für bestimmte weiter zurückliegende Jahrzehnte lassen sich keine aussagekräftigeren Daten als die filmische Darstellung der Gesellschaft finden.
2. Da sich sowohl Filmemacher als auch Sozialwissenschaftler für die Gesellschaft interessieren, kann auch die filmische Beobachtung von Gesellschaft unterstützend für schwer zugängliche Gesellschaftsbereiche genutzt werden. So ist z.B. fraglich, ob ethnographische Studien (Knoblauch und Thierbach/Petschick, Kapitel 36 und 66 in diesem Band) im Hinblick auf Ghettos in Großstädten wie Paris oder New York mehr Erkenntnisse über die spezifischen gesellschaftlichen Probleme liefern können als Filme wie *La Haine* (1995) oder *The Bronx* (1981) (Kunz/Schäfers 2008) – auch vor dem Hintergrund des erheblichen zeitlichen und finanziellen Aufwands der Datenerhebung bei Ethnographien.
3. Schließlich stehen Filme nicht außerhalb der Gesellschaft, sondern sind ein wichtiger Bestandteil von gesellschaftlichem Alltag, so dass gerade die Wechselwirkung zwischen filmischen Darstellungen und realgesellschaftlichen Ereignissen und Diskursen (Traue et al., Kapitel 34 in diesem Band) von Interesse sein kann. Gegenstand von sozialwissenschaftlichen Filmanalysen sind folglich die „Gesellschaft im Film“ (Schroer 2008) sowie das Verhältnis von Film und Gesellschaft.

Prinzipiell muss bei der Verwendung von Filmen (und anderen Massenmedien) zwischen der (1) Produktion, (2) dem Produkt, (3) der Distribution und (4) der Konsumption bzw. Rezeption unterschieden werden. Alle vier Ebenen dieses Prozesses können einzeln oder verbunden von Interesse sein. Wirtschafts- oder organisationssoziologische

Fragestellungen fokussieren dabei eher auf Produktion und Distribution, z.B. auf die nationale und internationale Verflechtung der Filmindustrie (Thiermeyer 1994) oder auf die projektförmige Arbeitsorganisation bei der Produktion von Fernsehsendungen (Windeler et al. 2001). Die Filmanalyse an sich bezieht sich aber hauptsächlich auf die Interpretation des fertigen Produkts und dessen Thematisierung der sowie die Einbettung in die Gesellschaft. Dabei stellen die anderen Ebenen – Produktion, Distribution und Rezeption – wichtige Kontextinformationen dar.

Es lassen sich verschiedene fiktionale Gattungen audiovisueller Medienformate unterscheiden, z.B. Spiel-, Fernsehfilme und Fernsehserien, die, trotz unterschiedlicher Produktions- und Rezeptionsbedingungen, in ähnlicher Weise untersucht werden können, während Dokumentarfilme, wissenschaftliche Filme (Reichert 2007), Reality-TV-Sendungen, audiovisuelle Produkte im Internet wie YouTube-Videos (Tuma/Schnettler, Kapitel 68 in diesem Band) nicht-fiktionale bzw. eine Mischung aus fiktionalen und nicht-fiktionalen Gattungen darstellen und daher methodologisch anders zu behandeln sind. Die folgenden Ausführungen zur Filmanalyse fokussieren sich aus Platzgründen auf die genannten fiktionalen Gattungen und beschreiben die Besonderheiten des Forschungsprozesses (Przyborowski/Wohlrab-Sahr und Stein, Kapitel 6 und 7 in diesem Band) bei der Filmanalyse.

---

## 69.2 Filmauswahl: Grundgesamtheit und Stichprobe

Bei der Bestimmung der Grundgesamtheit und der Stichprobenziehung geht es darum festzulegen, welche Filme generell für das Forschungsinteresse relevant sind, welche Filme aus der definierten Gesamtheit aus welchen Gründen für die Analyse ausgewählt werden und was schließlich bei den ausgewählten Filmen der Stichprobe fein, grob oder gar nicht analysiert wird – was also letztlich den Datenkorpus, die Erhebungs- und Analyseeinheiten, ausmacht, auf den sich die Ergebnisse beziehen. Die Klärung dieser Aspekte ist zwingend erforderlich, um Aussagen über die Verallgemeinerbarkeit der Ergebnisse machen zu können.

Die große Schwierigkeit bei der theoretischen Absteckung der Grundgesamtheit der relevanten Filme besteht darin, dass es keine vollständige Auflistung aller jemals produzierten Filme gibt. Hinzu kommt, dass die Sortierung von Filmen oder Serien zumeist über Genres funktioniert und dieser Begriff eher ungeeignet ist, um eine Grundgesamtheit zu definieren, weil er sich auf verschiedensten Dimensionen wie dem inhaltlichen Thema (z.B. Kriegsfilm, Sportfilm), der Ausstattung (z.B. Kostümfilm) oder den spezifischen emotionalen Reaktionen des Publikums (z.B. Komödie, Thriller) bewegt und sich ständig weiter ausdifferenziert. Zudem folgen sozialwissenschaftliche Forschungsinteressen nicht den Genreeinteilungen, d.h. nicht für jedes Forschungsthema steht ein filmisches Genre parat. Umgekehrt kommen für bestimmte Fragestellungen nahezu alle Genres in Frage. So hat man bei keinem der Themen im Sammelband „Gesellschaft im Film“ (Schroer 2008) – „Architektur und Stadt im Film“, „Familie im Film“, „Jugend und Jugendlichkeit im Film“,

„Medien im Film“, „Wohnen im Film“ – spontan eine konkrete Gruppe von Filmen vor Augen, die den Kern der Grundgesamtheit bilden könnten.

### 69.2.1 Kriteriengeleitetes Vorgehen

Je nachdem, ob die Fragestellung sehr weit oder eher eng gefasst ist und das Vorgehen eher deduktiv oder induktiv (Reichertz, Kapitel 2 in diesem Band), gibt es dennoch verschiedene Strategien zur Bestimmung der Grundgesamtheit, die im Wesentlichen denen entsprechen, die auch bei anderen Definitionen der Grundgesamtheit angewandt werden:

1. zeitlich: Oft sind nur Filme eines bestimmten Jahres, bestimmter Jahre oder Jahrzehnte von Interesse. Allerdings ergeben sich hierbei oft auch große Fallzahlen. Für die Jahre 1980, 1990 oder 2000 sind in der Internet Movie Database (imdb) jeweils zwischen 3.000 und 3.500 Titel nur für die Kategorie „Spielfilm“ gelistet. 2010 sind es bereits ca. 7.000 Spielfilme.
2. räumlich: Fokussierungen können etwa auf dem Hollywood-Kino, dem europäischen Film, anderen lokalen Filmindustrien sowie im Vergleich zwischen diesen liegen.
3. sachlich und inhaltlich: Sachliche Kriterien beziehen sich z.B. auf die Unterscheidung zwischen Mainstream versus Kunstkino, großen Studioproduktionen versus Independent-Film, während inhaltliche Kriterien eine mehr oder weniger konkrete Vorgabe sind, was im Film auf welche Weise enthalten sein soll.

Aus dieser Grundgesamtheit wird dann die Stichprobe nach den in den Sozialwissenschaften üblichen Kriterien (Akremi und Häder/Häder, Kapitel 17 und 18 in diesem Band) gezogen. So verwenden etwa Flicker und Zehenthaler (2012: 224-227) in ihrer Untersuchung zur filmischen Verhandlung von Geschlechterverhältnissen alle genannten Kriterien. Ihr Fokus liegt auf der österreichischen Gesellschaft und dem jüngeren österreichischen Film. Den Startpunkt markiert der Film „Nordrand“ (1999), „der als erste österreichische Produktion seit 1948 für den Hauptpreis der Filmfestspiele von Venedig nominiert wurde“ (Flicker/Zehenthaler 2012:224). Der zeitliche Horizont wurde von 1999 bis 2010 eingegrenzt, räumlich gesehen geht es um den jüngeren österreichischen Spielfilm, wobei dies auch als sachliches Kriterium gelten kann, da dieser Art von Filmen ein „hoher künstlerischer Anspruch“ bescheinigt wird. Inhaltliche Einschränkungen ergaben sich durch die Vorgabe, dass die Handlung größtenteils im gegenwärtigen Österreich stattfinden sollte. Insgesamt erfüllten 35 Filme die genannten Kriterien. Für die Stichprobengestaltung wurde als weiteres inhaltliches Kriterium festgelegt, dass die Thematisierung von „Geschlechternarrationen im Kontext sozialer Ungleichheit“ tragendes Handlungselement sein sollte. Daraus ergab sich ein Datenkorpus von fünfzehn Filmen, wobei unklar ist, ob dies eine Stichprobe ist oder ob es sich nicht um die eigentliche Grundgesamtheit handelt, die Untersuchung folglich eine Vollerhebung darstellt.

### 69.2.2 Offenes Vorgehen

Wählt man ein offenes Vorgehen mit einer noch vagen Fragestellung, welches den Vorteil hat, dass man aus dem Material heraus die Grenzen ziehen kann, statt sie von außen aufzusetzen, so gestaltet sich dieser Prozess deutlich schwieriger. Wenn man sich z.B. dafür interessiert, welche Rolle Arbeit im Film spielt, so ist dies etwas so Alltägliches, dass es praktisch in allen Filmen auftauchen kann. Konkretisierungen – wie die Eingrenzung der Branche oder Berufsgruppe – können hier Abhilfe schaffen, z.B. weiß man aus dem Alltag, dass in bestimmten Filmen oder Serien der Fokus auf medizinischen Berufen liegt, etwa bei Krankenhaus- und Arztserien. Gleichzeitig finden sich natürlich auch Mediziner in anderen Filmen und Serien, so dass es sich anbietet, Darstellungen in Filmen mit spezifischem Fokus auf dieses Thema mit anderen, die diesen Fokus nicht haben, zu kontrastieren. Weiter lässt sich die Frage stellen, was genau beim Thema Arbeit untersucht werden soll, z.B. Arbeitsbeziehungen oder die Organisation von Arbeit? Mittels eigener Filmkenntnisse, Stichwortsuche in der imdb, Filmlexika, Inhaltsangaben, allgemeinen Internet- sowie Literaturrecherchen lassen sich dann weitere Anhaltspunkte für die Grundgesamtheit und die Fokussierung des Forschungsthemas finden. Dies ist ein sehr zeitaufwendiger Suchprozess, in dem man sich von Anhaltspunkt zu Anhaltspunkt hangelt und die Grundgesamtheit häufig unscharf bleibt. Zudem erfordert dies ein iteratives Pendeln zwischen Datenerhebung und Datenauswertung gemäß dem Prinzip des Theoretical Sampling der Grounded Theory (Strübing, Kapitel 32 in diesem Band).

---

## 69.3 Vorbereitung und Organisation der Daten

Zur Vorbereitung der Filmanalyse müssen die Daten beschafft und in einer Form aufbereitet (Kuckartz/Rädiker, Kapitel 27 in diesem Band) werden, die es ermöglicht, eine tiefergehende Analyse durchzuführen. Dies kann durch Fernsehaufzeichnungen, Ausleihe oder Kauf von Filmen erfolgen. Die umfassendste Möglichkeit der Bearbeitung des Materials bietet dabei die audiovisuelle Datei am Computer. Nach einer Grobsichtung des Materials, die im Wesentlichen darin besteht, die ausgewählten Filme anzusehen, muss entschieden werden, ob die Filme in ihrer Gesamtheit analysiert werden sollen oder ob nur bestimmte Teile (z.B. Schlüsselszenen) von Bedeutung sind. Handelt es sich um eine große Filmstichprobe, können zusätzlich Inhaltsbeschreibungen helfen, um die Filme nach relevanten Kriterien zu sortieren und so eine Vorstrukturierung des Materials vorzunehmen.

Die Filme, die für eine tiefergehende Filmanalyse in Frage kommen, müssen in sinnhafte Untereinheiten aufteilt werden, wobei sich diese Einteilung nicht nach medien- oder filmwissenschaftlichen, sondern nach sozialwissenschaftlichen Kriterien richtet. Anders als bei Texten (Salheiser, Ernst und Klein, Kapitel 62, 63 und 64 in diesem Band) gibt es aber bei Filmen keine objektiv festsetzbare kleinste sinntragende Einheit wie z.B. das Wort. Filme lassen sich nur grob in Einzelbilder, Einstellungen, Szenen und Sequenzen unterteilen. Filmtechnisch ist die Einstellung einheitlich definiert: Sie beginnt und endet

mit einem Schnitt. Für die Begriffe Szenen oder Sequenzen gibt es dagegen unterschiedliche Definitionen und sie werden auch synonym verwendet. Als Szene definiert z.B. Mikos (2008: 91) die „Einheit von Ort und Zeit, in der sich eine kontinuierliche Handlung vollzieht“, wobei er noch zwischen Dialogszenen ohne Aktion, Dialogszenen mit Aktion, Aktionsszenen ohne Dialog und deskriptiven Szenen (z.B. atmosphärische Eindrücke von Landschaft, Großstadtleben) unterscheidet. Sequenzen wiederum sind „eine Gruppe von miteinander verbundenen Szenen, die eine Handlungseinheit bilden und sich durch ein Handlungskontinuum von anderen Handlungseinheiten unterscheiden“ (Mikos 2008: 92). Wie nun die Grenzen der sinnhaften Teile gesetzt werden, ob es sich dabei um ein Einzelbild, eine Einstellung, eine Szene oder eine Sequenz handelt, hängt vom jeweiligen Film und vom spezifischen Erkenntnisinteresse ab. Reichertz und Englert (2011: 30) sprechen deshalb von Kalibrierung zwischen Analyseeinheit und Fragestellung. Handlungen müssen dabei keineswegs streng linear über die erzählte Filmzeit abgeschlossen sein. Durch verschiedene Schnitt- und Montagetechniken können sich Bestandteile einzelner Handlungsstränge über den gesamten Film verteilen. Auch dies lässt sich z.B. mit einer Sequenzeinteilung aufspüren.

Es empfiehlt sich in der Vorbereitung einer Filmanalyse, sich neben der Aufarbeitung des thematischen Stands der Forschung auch einen groben Überblick über den Stand der Forschung zum spezifischen Medium Film im Bereich der Filmwissenschaft zu verschaffen, und zwar sowohl in Bezug auf filmtechnische Hintergrundinformationen zu Erzähltechniken, Kamera und Schnitt sowie zur ästhetischen Gestaltung (Monaco 2009; Mikos 2008), als auch in Bezug auf verschiedene Filmtheorien (Stam 2000; Elsaesser/Hagener 2010), damit es zu keinen Fehlinterpretationen oder zu banalen Erkenntnissen kommt.

Zur Verwaltung des Datenmaterials sollte ein selbstgewähltes Katalogisierungssystem (Mayring/Fenzl, Kapitel 38 in diesem Band) entwickelt werden, in dem zwischen den spezifischen Filmen oder Serien und den jeweiligen Einstellungen, Szenen oder Sequenzen unterschieden werden kann. Dies ist nicht nur für das eigene Zurechtfinden während des Forschungsprozesses wichtig, sondern dient gleichzeitig auch der nachvollziehbaren Dokumentation bei der Präsentation der Ergebnisse.

---

## 69.4 Filminterpretation

Eine Zusammenstellung sozialwissenschaftlicher Methoden der Filmanalyse im Sinne einer Anleitung und Gegenüberstellung verschiedener möglicher Ansätze steht noch aus. Bisher gibt es nur Einzelstudien, die verschiedene Methoden anwenden, entweder als Übertragung von Methoden anderer Wissenschaften, wie z.B. die semiotische Filmanalyse (Peltzer 2011), oder als Übertragung von eher sozialwissenschaftlichen Methoden auf den Film, wie die wissenssoziologisch orientierte dokumentarische Methode (Bohsack, Kapitel 67 in diesem Band) oder die hermeneutisch-wissenssoziologische Videoanalyse (Reichertz/Englert 2011, siehe auch Herbrik/Kurt und Schnettler/Tuma, Kapitel 33 und 68 in diesem Band).

Unabhängig davon, welche Methode der Filminterpretation präferiert wird, ist die Unterscheidung zwischen visuellen, audiovisuellen und auditiven Datensorten/Dokumenten zunächst ganz wesentlich. Insbesondere durch das Schlagwort „pictural turn“ bzw. „visual turn“ wurde zwar zurecht die Bedeutung von visuellen Daten für die Sozialwissenschaften herausgestellt, aber auch die Tatsache vernachlässigt, dass viele der untersuchten Datensorten eben nicht nur visuell, sondern vielmehr audiovisuell sind, und dass sie Bewegungen und Handlungen in der Zeit präsentieren. Bildtheoretische Ansätze aus den verschiedensten wissenschaftlichen Bereichen können zwar zur Analyse der Bildinhalte (Müller 2003) herangezogen werden, um ikonographische Topoi (Reichertz/Englert 2011: 29) herauszuarbeiten, sind aber allein nicht ausreichend, um Filme adäquat zu interpretieren.

Insgesamt besteht die Filmanalyse aus dem Wechselspiel zwischen Grob- und Feinanalyse sowie der Einbeziehung von Kontextinformationen. Wesentlich ist sowohl für die grobe Interpretation als auch für (sequenzanalytische) Feininterpretationen, dass Filme verschiedene Informationskanäle besitzen, die zwar getrennt aufgeschlüsselt werden können, deren Sinn sich aber nur aus der Einheit von filmischer Handlung, filmischen Bildern und Bildkompositionen, Dialogen und Interaktionen, Kameraführung, Schnitte und Montage, Filmmusik und anderen Geräuschen sowie visuellen und Spezialeffekten erschließt.

#### **69.4.1 Grobanalyse**

Mit Grobanalysen (Mikos 2008) lassen sich z.B. strukturelle Merkmale von Filmen gut herausarbeiten sowie Vorselektionen für die Feinanalysen treffen. Dazu gehört der Verlauf der Handlung, Beschreibungen zu Figuren bzw. Figurenkonstellationen im Film (wer nimmt welche Rolle ein und steht in welchem Verhältnis zu anderen Figuren?), die Einteilung in Sinnsequenzen oder die Identifikation von Schlüsselszenen.

Schlüsselszenen bzw. -sequenzen sind Situationen, in denen der Film mit dem zuvor etablierten Zusammenhang bricht, z.B. durch den Tod einer für den Film zentralen Figur. Jede einzelne Schlüsselszene könnte für eine Feinanalyse in Frage kommen. Vor der eigentlichen Feinanalyse oder parallel dazu lassen sich diesbezüglich folgende Fragen stellen: Was war das „normale“, das bis zu dieser Situation etabliert wurde? Was genau hat sich durch dieses Ereignis verändert? Wie wird diese Veränderung in Szene gesetzt? Welchen Verlauf nimmt die Geschichte nach diesem Ereignis? In welchem Verhältnis stehen die verschiedenen Schlüsselszenen eines Filmes zueinander? Wie gestaltet sich dies im Vergleich zu anderen Filmen des Datenkorpus? Schlüsselsequenzen können abweichend davon auch nach dem spezifischen Erkenntnisinteresse festgelegt werden, als alle Bestandteile des Filmes, die für die Fragestellung relevant sind.

### 69.4.2 Feinanalyse

Die tiefergehende Beschäftigung mit dem gesamten Film, zumeist aber nur mit bestimmten Teilen des Filmes, erfolgt dann in einer Feinanalyse (Reichertz/Englert 2011: 62-118). Die Feinanalyse steht und fällt mit der Qualität der angelegten Filmpartitur bzw. des Sequenzprotokolls (Reichertz/Englert 2011: 33), das vergleichbar ist mit einem Feldprotokoll bei ethnographischen Beobachtungen (Thierbach/Petschick, Kapitel 66 in diesem Band). In diesem Protokoll müssen alle für die Analyse wichtigen sintragenden Ebenen enthalten sein und so aufgefüllt werden, dass die Einheit des Sinns erhalten bleibt.

Für die Unterteilung der Ebenen gibt es verschiedene Möglichkeiten. Zumeist wird allgemein zwischen Bildinhalten und Kameraaktivitäten, Handlungsbeschreibung inklusive Dialoge sowie Musik und Geräusche unterschieden. So ist etwa die hermeneutisch-wissenssoziologische Videoanalyse an der „Auffindung der sozialen Bedeutung der Handlung der Bildgestaltung plus der durch sie eingefangenen Handlung im Bild interessiert“ (Reichertz/Englert 2011: 29), weshalb das handlungsorientierte Notationssystem (HANOS) zwischen Kamerahandlung (Wie wird etwas gezeigt?) und Handlung vor der Kamera (Was wird gezeigt?) unterscheidet (siehe Tab. 69.1).

Mit für die Filmanalyse geeigneter Software (Reichertz/Englert 2011: 34-36; Moritz 2011: 28-33) – wie z.B. ELAN (<http://tla.mpi.nl/tools/tla-tools/elan/>) – lässt sich das Sequenzprotokoll mit dem Datenmaterial direkt verknüpfen und die Anmerkungen auf der Timeline der Spielfilmzeit fixieren. Für jede Beobachtungskategorie sowie für die Sequenzeinteilung und die transkribierten Dialoge stehen separate Bereiche zur Verfügung und trotzdem wird alles zusammengehalten.

Wie viele Kategorien und welche genau man nun für dieses Feldprotokoll auswählt, hängt davon ab, wie wichtig sie für das eigene Forschungsinteresse sind. Reichertz und Englert (2011) schlagen deshalb vor, im ersten Schritt der Feinanalyse einzelne Abschnitte des Videos oder Filmes offen zu kodieren (Strübing, Kapitel 32 in diesem Band), um herauszufinden, welche Kategorien die Filmpartitur enthalten soll. Wenn dies nicht bereits in einer Grobanalyse geschehen ist, schließt sich daran die Einteilung in zusammenhängende Sinneinheiten an. Für die Notation der Dialoge stehen alle Transkriptionsverfahren zur Verfügung, die auch bei qualitativen Interviews angewandt werden.

In welcher Reihenfolge nun die weiteren Dimensionen der Filmpartitur gefüllt werden, ist Geschmacksache. Reichertz und Englert (2011) interessieren sich primär für die Kamerahandlung, weshalb sie zuerst alle handlungsrelevanten Teile der Kamerahandlung in beschreibender und interpretierender (codierter) Weise vermerken. Eine andere Variante ist es, zunächst mit der Beschreibung und Interpretation der gezeigten Handlungsebene zu beginnen und die anderen Kategorien nach und nach zu ergänzen. In jedem Fall bedeutet die Feinanalyse von filmischen Sequenzen ein mehrfaches Durchlaufen des Materials (Strübing, Kapitel 32 in diesem Band), weil nicht alle Aspekte zugleich beobachtet werden können.

**Tab. 69.1** Notierung der Kamerahandlung mit HANOS (nach Reichertz/Englert 2011: 37f.)

Kamerahandlung	Handelnde Kamera (Was macht die Technik der Kamera während der Produktion mit dem Geschehen vor der Kamera?)	Stand/Handkamera	Fest/bewegt
	Kadrierung	Wie ist das Bild aufgebaut? Vorn/Hinten/Mitte	
	Einstellung	Überblick, amerikanisch, close	
	Schärfentiefe	Was ist scharf, was nicht?	
	Perspektive	Augenhöhe, Frosch-, Vogelperspektive	
	Autonom?	Folgt die Kamera den Bewegungen der Akteure oder geht sie eigene Wege autonom?	
	Tempo	Langsame Bewegungen oder schnelle?	
	Farbe	Farbig oder schwarz/weiß	
Kommentierende Kamera (Wie kommentiert der korporierte Akteur in der Postproduktion das Geschehen?)	Voice over	Was spricht die Kamera wie	
	Stimme aus dem Off	Ist eine Stimme während der Aufnahme aus dem Off zu hören?	
	Was ist mit Verfremdungen	Wird etwas verfremdet?	
	Musik/Geräusch	Gibt es Musik oder Geräusche von der Kamera?	
	Grafik	Ist Grafik eingefügt?	
	Text	Ist Text ins Bild eingefügt?	
Montierende Kamera (Wie werden die takes in der Postproduktion zu einer Sequenz zusammengeführt?)	Schnitt	Harter Schnitt/Überblendung	
	Zeitlupe/Raffer	Sind Zeitlupe oder Zeitraffer vorhanden?	
Handlung vor Kamera	Bühne	Ort	
		Zeit	
	Requisiten	Zum Ort	
		Zur Person: Geschlecht, Alter, Aussehen, Kleidung	
	Akteure	Sozialer Typus: Polizist etc.	
		Handlungstyp: Bedächtig, hektisch	
		Nonverbale Botschaft: verärgert	
	Symbolische Interaktion	Sprechen: Was sagen die Akteure?	
		Gestik: Was bedeuten die Gesten?	
		Mimik: was bedeutet die Mimik?	
	Handeln	Was: Was wird wie mit wem getan?	
		Sinn: Subjektiver Sinn	

Bei der Interpretation wesentlicher Sequenzen empfiehlt es sich, wenn möglich, Datensitzungen im Team (Kurt/Herbrik, Kapitel 33 in diesem Band) durchzuführen. Grundsätzlich gibt es eine unendliche Fülle an möglichen Interpretationen und Aspekten, die bedeutsam sein können, weshalb es sehr nützlich ist, einerseits die Beobachtungsfähigkeit anderer mit einzubeziehen, andererseits aber auch eigene mögliche Deutungen mit denen von anderen abzugleichen. Wenn dies nicht möglich ist, bieten sich ersatzweise Filmrezensionen an, um ein Gefühl dafür zu bekommen, wie unterschiedlich Filme gesehen und interpretiert werden.

Bei allem, was einem zur Handlung, zu den Figuren, zur ästhetischen Gestaltung usw. auffällt, lassen sich die Fragen stellen: Warum ist dies so dargestellt und nicht anders? Wie lässt es sich im Zusammenhang mit dem gesamten Film und anderen Filmen verstehen? Stück für Stück sollten sich die notierten Codes im Verlauf der Analysen dann zu Dimensionen, Typen der Darstellung usw. verdichten.

### 69.4.3 Kalibrierung

Bei einem komplexen Datenmaterial wie dem Film kann es leicht passieren, dass man sich in der Fülle von Details verliert und jedes Pixel interpretieren möchte. Deshalb muss immer der Bezug zur Fragestellung gewahrt werden, was Reichertz und Englert (2011) mit Kalibrierung bezeichnen. Manche Elemente liefern auch keinen Erkenntnisgewinn, die beobachteten filmischen Details haben keine tiefere Bedeutung. Die Einbeziehung von Kontexten kann hier unterstützend wirken. Zum Kontext gehören u.a.:

1. der Gesamtfilm, wenn nur einzelne Sequenzen analysiert werden. Zusätzlich zum Gesamtfilm existieren häufig filmbegleitende Materialien wie Making Of's, Interviews, ausgelassene Szenen oder alternative Enden.
2. Filmvergleiche ermöglichen Darstellungsvergleiche und die thematische Untersuchung über einen größeren Zeitraum hinaus. Filme stehen dabei häufig in Beziehung zu anderen Filmen. Sie können andere Filme zitieren, sie reihen sich ein in Filme, die so ähnlich sind oder brechen bewusst mit gewohnten Konventionen.
3. Filmbeschreibungen und -kritiken können zudem Hinweise auf die Ebene der Rezeption liefern.
4. Realgesellschaftliche Ereignisse und Hintergründe sind ein weiterer wichtiger Kontext, der auch für die Wechselwirkung zwischen Film und Gesellschaft berücksichtigt werden sollte. Die gesamte Reihe an Spielfilmen und Dokumentationen zu den Anschlägen des 11. Septembers 2001 hätte es in dieser Form sicherlich nicht gegeben, wenn dieses Ereignis nicht eingetreten wäre. Gleichzeitig hatte das Ereignis auch Auswirkungen auf andere Filme. Der für Oktober 2001 geplante Kinostart des Films „Collateral Damage“ (2002) wurde aus Pietätsgründen um mehrere Monate verschoben (Bevc 2005: 46). Andererseits ist natürlich die Rückbindung der Ergebnisse der Filmanalyse auf die Gesellschaft und die Beeinflussung der Gesellschaft durch Filme wesentlich.

## 69.5 Präsentation der Ergebnisse

Für die Nachvollziehbarkeit des methodischen Vorgehens ist die Entscheidung über Grundgesamtheit und Stichprobe darzulegen. Zudem ist eine exemplarische Beschreibung der Filmpartitur bzw. der Sequenzprotokolle erforderlich. Die Schwierigkeit der Präsentation (Meyer/Meier zu Verl, Kapitel 15 in diesem Band) von Filmanalysen ist, dass das Datenmaterial in einer anderen medialen Form vorliegt als die anschließende wissenschaftliche Abhandlung darüber. Die Überführung von Filmen in Text ist nicht in der Weise möglich, wie das z.B. bei qualitativen Interviews (Helfferich, Kapitel 39 in diesem Band) der Fall ist. Mikos (2008: 102) schreibt hierzu: „Die Präsentation muss generell darauf abzielen, Lesern oder Zuhörern das audiovisuelle Material, das der Analyse zugrunde lag, in sprachlicher Form so nahezubringen, dass sie die Ergebnisse nachvollziehen können, auch wenn sie den oder die betreffenden Filme nicht gesehen haben“. Praktisch gesprochen, ist das eigentlich unmöglich. Textbeschreibungen, Screenshots, Einzelbildsequenzen usw. sind zwar nützlich und sollten auch verwendet werden, können aber in keiner Weise das Zusammenspiel der verschiedenen Sinnebenen des Films adäquat wiedergeben. Screenshots oder Einzelbildsequenzen lassen sich mit Analyse-Software, wie ELAN oder Dartfish (<http://www.dartfish.com/de/>), oder Grafikprogrammen wie Adobe Photoshop, Indesign (<http://www.adobe.com/de/>) oder GIMP (<http://www.gimp.org/>), erstellen und in den Text einbetten.

Besser ist es allerdings, wenn die Leser zusätzlich die Möglichkeit haben, die im Text beschriebenen Sequenzen, parallel auch ansehen zu können. Filmsequenzen können daher auf CD, DVD oder zum Download bereitgestellt werden. Optimal ist dies aber auch nicht, da man zwischen Buch, eBook oder Datenmaterial hin- und herwechseln muss.

Wünschenswert wäre eine adäquatere Form der Veröffentlichung sozialwissenschaftlicher Filmanalysen z.B. durch interaktive multimediale Formate in denen Texte, unkommentierte und kommentierte Filmausschnitte und optional weitere wichtige Daten vorhanden sind. Dies entspricht allerdings (noch) nicht den Standardformaten der wissenschaftlichen Veröffentlichung und die Herstellung solcher Medien gehört auch nicht zu den Basisfähigkeiten von Sozialwissenschaftlern.

## Literatur

- Bevc, Tobias (2005): Politische Identität im Film. In: Hofmann, Wilhelm/Lesske, Franz (Hg.) (2005): Politische Identität - visuell. Münster: Lit. 45-64
- Elsaesser, Thomas/Hagener, Malte (2010): Film Theory. New York: Routledge
- Flicker, Eva/Zehenthofer, Irene (2012): Geschlechternarrationen im Kontext sozialer Ungleichheit. In: Heinze, Carsten/Moebius, Stephan/Reicher, Dieter (Hg.): Perspektiven der Filmsoziologie. Konstanz/München: UVK. 220-244
- Kunz, Alexa M./Schäfers, Bernhard (2008): Architektur und Stadt im Film. In: Schroer, Markus (Hg.): Gesellschaft im Film. Konstanz: UVK. 14-47
- Mikos, Lothar (2008): Film- und Fernsehanalyse. Konstanz: UVK
- Moritz, Christine (2011): Die Feldpartitur. Wiesbaden: VS Verlag
- Monaco, James (2009): Film verstehen. Reinbek: Rowohlt
- Müller, Marion G. (2003): Grundlagen der visuellen Kommunikation. Konstanz: UVK
- Peltzer, Anja (2011): Identität und Spektakel. Konstanz: UVK
- Reichert, Ramón (2007): Im Kino der Humanwissenschaften. Bielefeld: transcript
- Reichert, Jo/Englert, Carina Jasmin (2011): Einführung in die qualitative Videoanalyse. Wiesbaden: VS Verlag
- Schroer, Markus (Hg.) (2008): Gesellschaft im Film. Konstanz: UVK
- Stam, Robert (2000): Film Theory. Malden, MA: Blackwell
- Thiermeyer, Michael (1994): Internationalisierung von Film und Filmwirtschaft. Köln/Wien et al.: Böhlau
- Windeler, Arnold/Lutz, Anja/Wirth, Carsten (2001): Netzwerksteuerung durch Selektion. In: Montage/av. 1 (10): 91-124

Cornelius Schubert

---

## 70.1 Gegenstände als Datenquelle

Gesellschaftliches Handeln produziert in großem Maße Zahlen, Buchstaben und auch Bilder. Aber es produziert auch Gebrauchsgegenstände, künstlerische und technische Artefakte. Unter letzteren werden im Folgenden alle Dinge, Gegenstände und Gerätschaften verstanden, die als Sachtechnik von Menschen produziert und genutzt werden, etwa Werkzeuge, Kleidung, Transport- oder Kommunikationsmittel, gemeinhin also die Gegenstände des täglichen Gebrauchs. Während die Analyse von Zahlen (Reinecke, Hartmann/Lengerer und Schmitz/Yanenko, Kapitel 44, 71 und 65 in diesem Band), Texten (Salheiser, Ernst und Klein, Kapitel 62, 63 und 64 in diesem Band) und Bildern (Bohnsack, Schnettler/Tuma und Akremi, Kapitel 67, 68 und 69 in diesem Band) zum Standardrepertoire sozialwissenschaftlicher Methoden gehört, bleiben die Gegenstände des täglichen Gebrauchs meist außen vor. Diese Sachvergessenheit der Soziologie (Linde 1972) betrifft neben den Methoden gleichermaßen auch die Theorien, weshalb im Folgenden nicht allein rein methodische, sondern auch methodologische und konzeptuelle Überlegungen dargestellt werden.

Die einfachste Möglichkeit, die Gebrauchsgegenstände wieder in den methodischen und theoretischen Blick zu nehmen, ist, sie als *Bedeutungsträger eines sozialen Sinns* zu verstehen – kurz gesagt als *Dokumente* bzw. als *Texte*. Denn Gebrauchsgegenstände werden von Menschen gemacht und damit sind sie ebenso soziale Erzeugnisse wie Sprache oder Schrift. In diesem Sinne sind Gebrauchsgegenstände gleichzeitig auch Dokumente sozialen Handelns. Eine solch einfache Übertragung greift jedoch an zwei Punkten zu kurz (Woolgar 1991): Erstens werden Gegenstände dann als Produkte sozialen Handelns gedeutet, nicht aber *im Prozess des Handelns* untersucht. Und zweitens gelten die Gegenstände dann als neutrale Mittel menschlichen Tuns bzw. als neutrale Träger sozialen Sinns, denen

ein *konstitutiver Beitrag zum Sozialen* versagt bleibt. Beide Verkürzungen gilt es zu vermeiden, um die Wirkungen von Gegenständen im Gebrauch empirisch untersuchen zu können.

---

## 70.2 Die Sozialität von Gebrauchsgegenständen

Folgt man Linde (1972), so wird der soziale Stellenwert profaner Artefakte in der Soziologie häufig unterschätzt. Dass gerade einfache Dinge bedeutsam an sozialen Beziehungen mitwirken, hat Latour (1996) an einer Vielzahl alltäglicher Gegenstände veranschaulicht, darunter Anschnallgurte, Bodenschwellen und Berliner Schlüssel. Hier sei als Beispiel der Hotelzimmerschlüsselanhänger angeführt (Latour 1996: 53ff.): Ein Hotelier möchte vermeiden, dass seine Gäste versehentlich den Hotelzimmerschlüssel bei ihrer Abreise mit nach Hause nehmen. Dazu stellt er am Ausgang ein Schild mit der Aufschrift „Schlüssel bitte an der Rezeption abgeben“ auf. Aber diese Mitteilung hat nicht die gewünschte Wirkung, da sie leicht übersehen wird. Nun kommt der Hotelier auf die Idee, sein Interesse in ein materielles Artefakt „einzuschreiben“. Er lässt einen klobigen Schlüsselanhänger aus Metall fertigen, der den Gast durch Gewicht und Größe daran erinnern soll, den Schlüssel an der Rezeption abzugeben. Mit dem Schlüsselanhänger wird das Interesse des Hoteliers gewissermaßen in ein Interesse des Gastes „übersetzt“ (s.u.). Das Beispiel macht deutlich, dass eine Analyse profaner Dinge nicht isoliert von deren Gebrauch gelingen kann. Es zeigt auch, dass der Schlüsselanhänger kein neutrales Vehikel der Hotelierinteressen ist, sondern geschickt zwischen den Interessen des Hoteliers und der Gäste vermittelt und diese verändert – auch wenn er selbst vom Hotelier erdacht wurde.

---

## 70.3 Gegenstände als Dokumente?

Bleiben wir zunächst bei der Idee, Gebrauchsgegenstände als Dokumente (Salheiser, Kapitel 62 in diesem Band) zu behandeln. Aber wie und worüber können diese Auskunft geben? Zuerst einmal sind sie *Produkte menschlichen Handelns*.

1. Damit verweisen sie auf einen *Herstellungszusammenhang*, auf mehr oder weniger zielgerichtetes menschliches Tun, auf Vorstellungen und Interessen die mit Hilfe der hergestellten Gegenstände erreicht werden sollen
2. Zum Herstellungszusammenhang kommt der *Verwendungszusammenhang* hinzu. Dieser verweist auf die Mittel und Ziele der Handlung, auf etwas, das unterstützt, kontrolliert, transformiert oder bearbeitet werden soll.

Damit sind Gebrauchsgegenstände eben keine isolierten Dinge oder Dokumente, sondern auf vielfältige Weise in gesellschaftliche Handlungen eingebettet. Gleichzeitig stoßen wir an die Grenze der Dokumentenmetapher. Im klassischen Sinne „dokumentieren“ Gebrauchsgegenstände nämlich keine vorher gegebenen sozialen Tatbestände, sie

sind vielmehr untrennbar mit deren Herstellung verbunden. Solcherlei Gegenstände, insbesondere etwa Werkzeuge, andere Gerätschaften, schaffen neue Beziehungen zwischen Menschen und ihrer Umwelt. Im Gebrauch lassen sich diese Gegenstände nur insoweit als Dokumente verstehen, als dass sie die durch sie erzeugten Verweisungszusammenhänge erkennen lassen. Ohne Kenntnis der Herstellungs- und Verwendungszusammenhänge bleiben die Gegenstände schweigsam. Und wenn wir den Gebrauch der Gegenstände nicht beobachten (Thierbach/Petschick Kapitel 66 in diesem Band) können, so sind wir – wie bei den Erklärungstafeln im Museum – auf zusätzliche Dokumente angewiesen, um den Gebrauch der Gegenstände und dessen soziale Implikationen deuten zu können.

Es existiert eine Reihe von Vorschlägen, wie technische Artefakte für soziologische Studien zugänglich gemacht werden können. Im Folgenden soll einer der prominentesten Ansätze, die sogenannte Akteur-Netzwerk Theorie (ANT), skizziert werden, da sich dort neben konzeptionellen auch starke methodische Bezüge finden.

---

## 70.4 Gegenstände als sozio-materielle Skripte

Die ANT geht zunächst von einer Analogie von Dingen und Dokumenten entlang einer allgemein semiotischen Perspektive aus. Hierfür werden beispielsweise Begriffe wie „script“, „inscription“, „description“, „prescription“, „translation“ und einige mehr eingeführt (Akrich/Latour 1992). Die Idee dahinter ist, dass ähnlich wie etwas auf einen Zettel *aufgeschrieben* werden kann, dieses auch in ein materielles Artefakt *eingeschrieben* werden kann – wie im oben angeführten Beispiel des Schlüsselanhängers. Im Herstellungs-zusammenhang werden den Dingen bestimmte Wirkungsweisen eingeschrieben (*inscription*), z.B. werden die Interessen von Hotelier und Gästen in den Schlüsselanhänger übersetzt (*translation*). Im Verwendungszusammenhang wirken diese Einschreibungen dann ähnlich wie Vorschriften (*prescription*), die von Nutzern übernommen werden oder auch nicht (*subscription*). Die soziologische Analyse (*de-scription*) besteht in der Folge aus der Beobachtung und Beschreibung dieser unterschiedlichen Skripte. Diese allgemein semiotische Herangehensweise reduziert Dinge jedoch nicht auf eine textartige Dokumentationsfunktion, sondern versucht, die semiotisch-materiellen Übergänge zwischen Texten und Dingen zu rekonstruieren (Law 2009). Die schriftliche Bitte, den Schlüssel zurückzugeben, ist eben nicht so wirkungsvoll wie das „moralische Gewicht“ des Schlüsselanhängers (Latour 1996: 53ff.).

---

## 70.5 Datenerhebung: Gegenstände im Gebrauch

Das Problem einer sozialwissenschaftlichen Untersuchung von Gegenständen ist nun gerade, dass Dinge nicht wie Menschen über sich selbst Auskunft geben können. Man kann sie also nicht einfach befragen, sondern muss sich entweder auf die Auskunft von Menschen über die Dinge verlassen oder den Gebrauch der Dinge selbst genau beobachten.

Profane Gegenstände und Gerätschaften entziehen sich einem solchen Zugriff jedoch auf doppelte Weise:

1. da wir im routinierten Gebrauch kaum noch an sie denken und uns so ihrer Wirkungen nur selten bewusst werden. Allein Pannen und Störungen machen uns immer wieder auf die unsichtbare Beteiligung der Dinge am Sozialen aufmerksam.
2. da der routinierte Gebrauch oft schnell und wortlos vollzogen wird und die Feinheiten des Umgangs mit Dingen daher nur schwer zu beobachten sind.

Aus methodischer Sicht ist daher hervorzuheben, dass ein solcher Ansatz zwangsläufig auf die Untersuchung des tatsächlichen Gebrauchs von Dingen angewiesen ist, denn nur dort können die vielfältigen Verweisungszusammenhänge zwischen Menschen und Dingen beobachtet werden. Je nach Fragestellung steht hierfür auch (mehr oder weniger) die Bandbreite sozialwissenschaftlicher Methoden zur Verfügung. So lassen sich beispielsweise Einstellungen zu unterschiedlichen Techniken befragen (Reinecke und Helfferich, Kapitel 44 und 39 in diesem Band) oder gesellschaftliche Diskurse (Traue et al., Kapitel 34 in diesem Band) zu umstrittenen Techniken etwa in Zeitungen (Klein, Kapitel 64 in diesem Band) oder Filmen (Akremi, Kapitel 69 in diesem Band) analysieren. In der Technikforschung wird daher auch ein breites Repertoire der qualitativen Sozialforschung von in-situ Beobachtungen (Thierbach/Petschick, Kapitel 66 in diesem Band) über Interviews (Helfferich, Küsters und Vogl, Kapitel 39, 40 und 41 in diesem Band) bis zu Dokumentenanalysen (Salheiser, Kapitel 62 in diesem Band) genutzt, es werden aber auch Experimente (Eifler, Kapitel 11 in diesem Band) durchgeführt. Je nach Methode lässt sich etwas anderes über die Gebrauchsgegenstände erfahren.

Als (im doppelten Sinne) besonders „gegenstandssensibel“ erweisen sich dabei ethnografisch orientierte Methoden (Knoblauch, Kapitel 36 in diesem Band), mit denen der Umgang mit Dingen und Techniken direkt beobachtet werden kann. In Kombination mit einem konzeptuellen Vokabular, das die Wechselwirkungen zwischen Menschen und Dingen erfassen kann, lassen sich spezifische „technografische“ Studiendesigns (Rammert/Schubert 2006) zur Beantwortung der oben aufgeworfenen Fragen entwickeln. Sie knüpfen an ethnografische Traditionen an, die sich mit der Bedeutung und Wirkung von Dingen beschäftigt haben, etwa der Ethnografie von Arbeit, der materiellen Kultur oder von Wissenschaft und Technik. Technografische Studien nehmen sowohl die Herstellungs- als auch die Nutzungszusammenhänge von Gegenständen und Techniken in den Blick. Sie folgen den Inskriptionen, Präskriptionen und Subskriptionen, den Übersetzungen und Artikulationen, die sich entlang der Entwicklung und Nutzung von Gegenständen und Techniken ausbilden. Insofern folgen technografische Studien sowohl den menschlichen Akteuren als auch den technischen Artefakten und legen größten Wert auf eine detailgenaue in-situ Analyse der Wechselwirkung zwischen ihnen. Denn außerhalb der gelebten sozialen Beziehungen verlieren die Gegenstände ihre Aussagekraft und ganz ohne soziale Bezüge und Verweise bleibt ein Gegenstand für die Soziologie sinnlos.

In den letzten Jahren wurden vermehrt audiovisuelle Aufzeichnungsmethoden (Schnettler/Tuma, Kapitel 68 in diesem Band) verwendet, um diese „Flüchtigkeit“ des alltäglichen Gebrauchs zumindest teilweise festhalten zu können (Hindmarsh/Heath 1998). Über detaillierte Fallstudien hinaus besteht nicht zuletzt ein zunehmendes Interesse an komparativen Designs, mit denen sich empirische Beobachtungen und theoretische Überlegungen systematisch verknüpfen lassen (Hine 2007).

---

## 70.6 Analytischer Fokus: Das konstitutive Mitwirken der Dinge

Neben der Datenerhebung rückt insbesondere die Datenanalyse die Eigenheiten von Gegenständen im Gebrauch in den Fokus. Im tatsächlichen Gebrauch wird deutlich, dass Dinge weder neutrale Mittel menschlicher Handlungen sind, noch dass sie eine autonome Macht gegenüber menschlichen Interessen darstellen. Eine *Analyse dieser konstitutiven Wirkung von Dingen* lässt sich in vier Schritten bzw. vier Bedeutungen des Begriffs der „technischen Vermittlung“ vollziehen (Latour 2000: 211ff.).

### 70.6.1 Verteilung von Entitäten auf Handlungszusammenhänge

Erstens wird damit die *Unterbrechung* eines Handlungsflusses markiert. In gewisser Weise lassen sich Gebrauchsgegenstände als produktive Umwege verstehen, um ein bestimmtes Ziel zu erreichen. Um von A nach B zu kommen benutzt man beispielsweise die Bahn, ein Auto, ein Fahrrad oder zieht sich zumindest Schuhe an. All diese Unterbrechungen des direkten Losmarschierens sind uns so geläufig, dass wir uns kaum noch Gedanken darüber machen, dass erst diese Dinge das Ziel überhaupt erreichbar machen. Die Analyse richtet sich also zuerst auf die Frage, *auf welche Entitäten (Dingen und Menschen) Handlungszusammenhänge verteilt sind*. Radfahren, beispielsweise, verteilt sich auf das Rad und den Fahrenden, Telefonieren auf das Telefonystem und die Sprechenden und Schreiben auf Stift, Papier bzw. Tastatur, Bildschirm, Rechner, Schreibprogramm und den Menschen.

### 70.6.2 Zusammensetzung von Menschen und Dingen

Im Anschluss daran kann zweitens die spezifische *Zusammensetzung* von Menschen und Dingen untersucht werden. Als Analyseeinheit gelten dann nicht mehr entweder Dinge oder Menschen, sondern ihre jeweilige Komposition als Bahn-, Auto- oder Fahrradfahrer bzw. als Fußgänger. Diese Zusammensetzungen sind in modernen Gesellschaften praktisch uferlos, da sich an jeden Gebrauchsgegenstand eine unüberschaubare Kette weiterer Verweise anschließt. An ein Auto etwa die Infrastrukturen von Straßen, Tankstellen und Werkstätten.

### 70.6.3 Störungen

Allerdings wirken diese Zusammensetzungen im täglichen Gebrauch, und das ist der dritte Punkt, meist unscheinbar, da sie im Sinne einer *Black Box* erwartungsgemäß funktionieren. Nur eine Panne macht die vielfältigen Verweisungen offensichtlich. Methodisch stellt uns das vor das Problem, dass die routineartige Nutzung von Gebrauchsgegenständen für die Analyse nur durch *zufällige Störungen, situative Reparaturen, subversive Umnutzungen oder gezielte theoretische Verfremdungen* offensichtlich wird. So untersucht man beispielsweise die Praktiken, mit denen nicht funktionierende Technik wieder in Gang gesetzt wird: vielleicht durch Schütteln oder durch Aus- und wieder Einschalten, durch Hilfesuche über Internetforen oder Telefonhotlines.

### 70.6.4 Artikulation

Den Abschluss der Analyse bildet viertens die Frage nach den Verschiebungen zwischen Zeichen und Dingen, der sogenannten *Artikulation*. Dabei dreht es sich im Kern um die Frage, warum der klobige Schlüsselanhänger wirkungsvoller ist als ein Schild. Mit dem Schlüsselanhänger schafft es der Hotelier, sein Interesse in ein Interesse des Gastes zu übersetzen, der den Schlüssel nicht aus Einsicht, sondern aus Bequemlichkeit wieder abgibt. Eine Analyse der Artikulation untersucht somit die Verschiebungen von Bedeutungen zwischen Zeichen und Dingen, die im Falle des Schlüsselanhängers zwar wie ein externalisierter Zwang erscheinen können, jedoch immer auch durch situative und subversive Nutzungsweise verändert werden können. Beispielsweise kann man die verschiedenen Artikulationen technisch vermittelten Kommunizierens untersuchen. Zu welchen Zwecken wird telefoniert, wann schreibt man eine SMS, wann eine E-Mail oder eine Nachricht im sozialen Netzwerk? Und welche Bedeutung bekommt ein Brief aus Papier dadurch? Je nach Gebrauch nehmen die Gegenstände eine spezifische Position im Akteur-Netzwerk ein und genau diese Verschiebungen stehen im Fokus der Analyse.

## Literatur

- Akrich, Madleine/Latour, Bruno (1992): A summary of a convenient vocabulary for the semiotics of human and nonhuman assemblies. In: Bijker, Wiebe E./Law, John (Hg.): Shaping technology – building society. Studies in sociotechnical change. Cambridge: MIT Press: 259-264
- Hindmarsh, Jon/Heath, Christian (1998): Video and the analysis of objects in action. In: Communication & Cognition 31 (2/3): 111-129
- Hine, Christine (2007): Multi-sited ethnography as a middle range methodology for contemporary STS. In: Science, Technology & Human Values 32 (6): 652-671
- Latour, Bruno (1996): Der Berliner Schlüssel. Erkundungen eines Liebhabers der Wissenschaften. Berlin: Akademie Verlag
- Latour, Bruno (2000): Die Hoffnung der Pandora. Frankfurt a.M.: Suhrkamp
- Law, John (2009): Actor network theory and material semiotics. In: Turner, Bryan S. (Hg.): The new Blackwell companion to social theory. Oxford: Wiley-Blackwell: 141-158
- Linde, Hans (1972): Sachdominanz in Sozialstrukturen. Tübingen: Mohr Siebeck
- Rammert, Werner/Schubert, Cornelius (Hg.) (2006): Technografie. Zur Mikrosoziologie der Technik. Frankfurt a.M.: Campus
- Woolgar, Steve (1991): The turn to technology in social studies of science. In: Science, Technology & Human Values 1 (16): 20-50

Peter H. Hartmann und Andrea Lengerer

---

## 71.1 Einleitung

Ämter und Behörden erfüllen Aufgaben der öffentlichen Verwaltung und erbringen Dienstleistungen für die Bürger. Bei dieser Tätigkeit entstehen Daten. Einerseits entstehen Daten, weil Behörden diese auf individueller oder Firmenebene benötigen, um Entscheidungen über Rechte und Pflichten dieser Einheiten zu treffen. Diese Daten gelten als *prozessproduziert*. Sie werden nicht für Zwecke der Statistik erhoben, können aber trotzdem dafür genutzt werden. Beispiele sind etwa Daten über Einkommen, wie sie im Rahmen der Steuererklärungen gesammelt werden, Daten über Arbeitslose, wie sie im Rahmen der Prüfung der Anspruchsberechtigung auf Arbeitslosengeld ermittelt werden, oder Daten der Rentenversicherung, die aus Meldungen der Arbeitgeber sowie aus Angaben zu gezahlten Leistungen gewonnen werden. Andererseits werden Daten von Politik und öffentlicher Verwaltung als Grundlage für Planungen und Entscheidungen verwendet. Diese Daten werden *eigens* und nur *für Zwecke der statistischen Analyse* erhoben. Beispiele sind Zensen (Volkszählungen) sowie amtliche Stichprobenerhebungen. Die größte amtliche Stichprobenerhebung ist der *Mikrozensus*, der von den Statistischen Landesämtern unter Koordination des Statistischen Bundesamts durchgeführt wird. Wie jede amtliche Erhebung basiert der Mikrozensus auf einer rechtlichen Grundlage. Darin sind sowohl die Inhalte der Erhebung als auch deren Durchführung geregelt.

Neben der amtlichen Nutzung sind sowohl prozessproduzierte Daten als auch Stichprobenerhebungen für die empirische Sozialforschung als Datenbasis nützlich. So erlaubt die Beschäftigtenstichprobe des Instituts für Arbeitsmarkt- und Berufsforschung (IAB), die auf prozessproduzierten Daten der sozialversicherungspflichtig Beschäftigten beruht, Analysen zu Erwerbsverläufen. Der für rein statistische Zwecke durchgeführte Mikrozensus und auch die Einkommens- und Verbrauchsstichprobe sind besonders für demografische und sozioökonomische Analysen relevant. Während der Mikrozensus jährlich aktuelle Eck- und Strukturdaten über die Bevölkerung und den Arbeitsmarkt für die

Zeiträume zwischen den in weiten Abständen durchgeführten Zensen (1950, 1961, 1970, 1987, 2011) liefert, erlauben die in fünfjährigem Abstand durchgeführten *Einkommens- und Verbrauchsstichproben* detaillierte Aussagen zur Einkommenssituation, Haushaltsausstattung und zu Konsumgewohnheiten.

In *aggregierter Form* (Graeff, Kapitel 72 in diesem Band) werden die Daten der amtlichen Statistik regelmäßig veröffentlicht. Sie werden in verschiedenen Fachserien und Statistischen Jahrbüchern als standardisierte Tabellen zur Verfügung gestellt. Für die Forschung sind die Daten der amtlichen Statistik aber vor allem in Form von *Mikrodaten* interessant. Darin sind Angaben zu einzelnen Untersuchungseinheiten enthalten, mit denen spezielle und insbesondere multivariate Analysen möglich sind.

Die Eigenschaften und Besonderheiten amtlicher Mikrodaten sowie ihre Vor- und Nachteile gegenüber sozialwissenschaftlichen Umfragedaten werden im Folgenden am Beispiel des Mikrozensus dargestellt. Für die Sozialforschung ist der Mikrozensus die mit Abstand wichtigste Erhebung der amtlichen Statistik in Deutschland.

---

## 71.2 Amtliche Mikrodaten in Deutschland: Das Beispiel Mikrozensus

### 71.2.1 Erhebungsmethode

Der Mikrozensus wurde 1957 in der alten Bundesrepublik eingeführt und 1991 in den neuen Bundesländern erstmalig erhoben (Emmerling/Riede 1997). Mit einem Stichprobenumfang von einem Prozent der Bevölkerung werden derzeit rund 820.000 Personen in 380.000 Haushalten für die Befragung ausgewählt. Bis einschließlich 2004 wurden die Angaben jährlich für einen bestimmten Zeitpunkt erhoben (feste Berichtswoche), seither findet die Befragung kontinuierlich über das ganze Jahr verteilt statt (gleitende Berichtswoche).

Das Stichprobendesign des Mikrozensus entspricht der einstufigen Klumpenstichprobe (Häder/Häder, Kapitel 18 in diesem Band). Bei den Klumpen handelt es sich um räumlich abgegrenzte Flächen, die mehrere Gebäude, ein ganzes Gebäude oder Teile von Gebäuden umfassen (Meyer 1994). Die durchschnittliche Klumpengröße lag früher bei 23 Wohnungen und wurde mittlerweile auf neun Wohnungen reduziert. Vor der zufälligen Auswahl von einem Prozent dieser Flächen werden diese nach regionalen Merkmalen geschichtet. Gelangt eine Fläche in die Stichprobe, werden Angaben zu allen dort vorhandenen Haushalten und den darin lebenden Personen erhoben. Auch die Bewohner von Institutionen werden erfasst.

Obwohl es sich beim Mikrozensus dem Prinzip nach um eine Querschnitterhebung handelt, wird nicht jedes Jahr eine komplett neue Stichprobe gezogen. Vielmehr verbleiben die ausgewählten Haushalte für vier aufeinander folgende Jahre in der Erhebung (Heidenreich/Herter-Eschweiler 2002). Ein Viertel der Haushalte scheidet jährlich aus und wird durch neue ersetzt. Da es sich um eine Flächenstichprobe handelt, werden fortziehende

Personen jedoch nicht weiter befragt, sondern diejenigen, die im nächsten Jahr in dem betreffenden Haushalt leben.

Durchgeführt wird der Mikrozensus hauptsächlich in Form einer persönlichen Befragung. Nach vorheriger Ankündigung suchen Interviewer die ausgewählten Haushalte auf und erheben die Angaben im CAPI-Verfahren. Proxy-Interviews, in denen eine Person im Haushalt stellvertretend für eine andere antwortet, sind generell zulässig (Wolf/Lüttinger 2009). Außer für Kinder liegen Fremdauskünfte überwiegend für Personen vor, die zum Zeitpunkt der Befragung nicht anwesend sind. Zur Wahrung des Rechts auf informativelle Selbstbestimmung besteht daneben die Möglichkeit der schriftlichen Auskunftserteilung. Wer dies ausdrücklich wünscht, kann seine Angaben selbst in einen Erhebungsbogen eintragen. In Einzelfällen werden Interviews telefonisch realisiert. Im Mikrozensus 2010 wurden 74% aller Haushalte per CAPI befragt, 22% aller Haushalte haben selbst einen Fragebogen ausgefüllt und in 4% der Haushalte wurden Telefoninterviews durchgeführt (Statistisches Bundesamt 2011: 6).

### 71.2.2 Inhalte

Obgleich er ursprünglich dafür konzipiert wurde, Arbeitsmarktdaten in international vergleichbarer Form zur Verfügung zu stellen, enthielt der Mikrozensus von Anfang an weitere Merkmale und hat sich zu einer umfangreichen Mehrthemenerhebung entwickelt (Herberger 1957, Emmerling/Riede 1997).

Der Mikrozensus enthält Angaben zu allen Mitgliedern der für die Befragung ausgewählten Haushalte sowie zur Haushaltsstruktur. Ausführliche Informationen gibt es zunächst zu den zentralen demografischen Variablen wie Alter, Geschlecht und Familienstand. Neben Ehen werden seit 1996 auch nichteheliche Lebensgemeinschaften von Personen gleichen oder verschiedenen Geschlechts erhoben. Die Kinderzahl von Frauen war 2008 erstmalig im Frageprogramm enthalten; sie wird 2012 erneut erhoben. Ab 2005 werden neben der Staatsangehörigkeit auch Angaben zum Migrationshintergrund der Befragten erfasst.

Weiterhin gibt es Daten zu Schulbesuch, allgemeinen und beruflichen Bildungsabschlüssen (erst ab 1976), vor allem aber zu sehr vielen Aspekten des Berufs, der Erwerbstätigkeit und der Erwerbslosigkeit. Die Quelle des Lebensunterhalts wird ermittelt sowie das persönliche und das Haushaltseinkommen. Zusatzerhebungen befassen sich unter anderem mit Gesundheitsfragen, Krankenversicherung, beruflichem Pendeln und der Wohnsituation.

### 71.2.3 Vorteile gegenüber Umfragedaten

Am Beispiel des Mikrozensus lassen sich die Vor- und Nachteile amtlicher Daten für die Sozialforschung gut darstellen. Die erste Stärke des Mikrozensus liegt in der hohen

Fallzahl. Die Stichprobengröße von einem Prozent der Bevölkerung erlaubt besonders genaue Parameterschätzungen, sie ermöglicht Analysen zu sehr speziellen Subgruppen (z.B. zu Angehörigen relativ seltener Berufe) und stellt die Grundlage einer starken Regionalisierung der Ergebnisse auf den Ebenen der Bundesländer, der Regierungsbezirke und auch größerer Städte dar. Eine zweite Stärke des Mikrozensus liegt in der Ausschöpfung der Stichprobe. Der Unit-Nonresponse (Engel/Schmidt, Kapitel 23 in diesem Band) ist außerordentlich gering, denn für den Mikrozensus besteht Teilnahmepflicht. Verbunden damit ist auch eine hohe Zuverlässigkeit und Vollständigkeit einzelner Angaben, für die wiederum in den meisten, aber nicht allen Fällen Auskunftspflicht besteht. Drittens sind die Daten des Mikrozensus durch eine hohe Kontinuität gekennzeichnet. Eine hohe Konsistenz im Variablenprogramm ermöglicht die Konstruktion von Zeitreihen, die teilweise bis 1957 zurückgehen. Da die Individualdaten in maschinenlesbarer Form ab 1962 verfügbar sind, sind auch Kohortenanalysen (Stein, Kapitel 7 in diesem Band) über lange Zeiträume hinweg möglich. Ein vierter Vorteil des Mikrozensus besteht darin, dass er Daten über alle in einem gemeinsamen Haushalt zusammenlebenden Personen erfasst (Koresidenzprinzip). Beziehungen zwischen Personen innerhalb desselben Haushalts sind gut darstellbar. So ist es etwa möglich, Korrelationen zwischen dem Einkommen von Männern und der Art der besuchten Schule ihrer Kinder zu untersuchen oder auch die Abhängigkeit des Erwerbsverhaltens von der Haushaltsform. Soziale Beziehungen, die über die Grenzen des Haushalts hinausreichen, sind dagegen nicht abgebildet (vgl. Abschnitt 2.4). Eine fünfte Stärke des Mikrozensus liegt im Panelcharakter der Erhebung. Da Befragte immer für vier Jahre in der Stichprobe verbleiben, können – mit gewissen Einschränkungen – auch Längsschnittanalysen durchgeführt werden. So untersucht Schimpl-Neimanns (2005) Bildungsverläufe auf Basis des Mikrozensus-Panels, Wolter/Schiener (2009) belegen Einkommenseffekte beruflicher Weiterbildung. Obgleich sich viele amtliche Daten auf ganz spezifische Themen beziehen, liegt sechstens eine Besonderheit des Mikrozensus gerade in seiner thematischen Breite.

#### **71.2.4 Probleme und Nachteile gegenüber Umfragedaten**

Die Nutzung amtlicher Daten durch die empirische Sozialforschung ist zunächst mit den typischen Problemen der Sekundäranalyse behaftet. Da die Daten nicht vom jeweiligen Forscher selbst erhoben wurden, sind sie häufig nur bedingt zur Beantwortung seiner spezifischen Fragestellung geeignet. Bei amtlichen Daten sind diese Restriktionen aber tendenziell stärker ausgeprägt als bei den Daten der empirischen Sozialforschung, da ihre Erhebung rechtlich geregelt ist und administrativen Zwecken dient (Wirth/Müller 2004: 101). Konzepte und Begriffe orientieren sich an legal definierten Sachverhalten. Sie entsprechen häufig nicht den Standards der empirischen Sozialforschung und geben ein bestimmtes Bild der sozialen Realität wieder. Außerdem beschränken sich amtliche Daten auf wenige, harte Fakten und enthalten praktisch keine Angaben über Einstellungen oder Präferenzen.

Dies gilt für Daten der amtlichen Statistik im Allgemeinen und im Besonderen für prozessproduzierte Daten. Eine methodenkritische Nutzung solcher Daten muss ihre Entstehungsbedingungen genau kennen und die jeweiligen Handlungslogiken der Verwaltungen einerseits und der „verwalten“ Personen andererseits reflektieren. Diese Logiken spiegeln sich in der Erfassung einer „Behördewirklichkeit“, manchmal aber auch in eigentümlichen Verzerrungen wieder. Im Kontext der historischen Sozialforschung wurde in diesem Zusammenhang analog zum Begriff der Quellenkunde der Begriff der Datenkunde geprägt (Bick/Müller 1984).

Durch ihre Bindung an rechtliche Grundlagen kann die amtliche Statistik außerdem nur zeitlich verzögert auf gesellschaftliche Entwicklungen reagieren. Bis ein neues Merkmal im Gesetz festgeschrieben wird, muss es nicht nur als soziale Realität anerkannt werden, sondern es muss auch politisch gewollt sein, dass es erhoben wird. Ein Beispiel hierfür sind nichteheliche Lebensgemeinschaften, die sich seit den 1970er-Jahren kontinuierlich ausbreiten, im Mikrozensus aber erst seit 1996 mit einer Frage erfasst werden. Andererseits garantiert das Prinzip der Legalität, dass sich Inhalte und Methoden einer Erhebung nur selten verändern und es somit wenige Brüche gibt, die den Vergleich über die Zeit beeinträchtigen.

Obwohl der Mikrozensus auch Längsschnittanalysen erlaubt (vgl. Abschnitt 2.3), können nur momentane Zustände ohne jede Einschränkung untersucht und miteinander verglichen werden. Ob und wie sich diese Zustände auf der Ebene des Individuums verändern, ist nur bedingt und über einen Zeitraum von maximal vier Jahren nachvollziehbar. Zur Untersuchung des sozialen Wandels auf der Ebene des Aggregats ist der Mikrozensus aber wiederum besonders gut geeignet. Anders als die Wellen eines Panels sind die einzelnen Querschnitte für jeden Zeitpunkt repräsentativ, zu dem sie erhoben werden.

Ein weiteres Problem des Mikrozensus und anderer amtlicher Haushaltsbefragungen ist das Koresidenzprinzip (vgl. Abschnitt 2.3). Es werden nur Personen erfasst, die in einem gemeinsamen Haushalt zusammenleben. Welche Beziehungen es zu Personen außerhalb des Haushalts gibt, ist nicht bekannt. So liegen weder Angaben zu nicht (mehr) im elterlichen Haushalt lebenden Kindern vor noch zu Partnerschaften mit getrennten Haushalten (Lengerer et al. 2007: 196f.). Auch die soziale Herkunft der Befragten wird im Mikrozensus (mit Ausnahme einer Zusatzuntersuchung aus dem Jahr 1971) nicht erfragt, was seine Nutzung für Zwecke der Mobilitätsforschung begrenzt.

---

### 71.3 Amtliche Mikrodaten in Europa

Die international vergleichende Sozialforschung kann Daten der europäischen amtlichen Statistik nutzen. Zu den wichtigsten Daten, die auf Mikroebene vorliegen, zählen der European Union Labour Force Survey (EU-LFS) sowie die European Union Statistics on Income and Living Conditions (EU-SILC). Der EU-LFS beinhaltet vergleichbare Angaben zur Erwerbstätigkeit in den Mitgliedsländern der Europäischen Union und einigen weiteren europäischen Ländern. Er wird jährlich von den nationalen statistischen Ämtern

erhoben und zentral bei Eurostat aufbereitet. Seit 1968 ist der deutsche Teil des EU-LFS in den Mikrozensus integriert. Die EU-SILC dient der Messung von Armut und sozialer Ausgrenzung in der Europäischen Union. Sie umfasst Quer- und Längsschnittdaten zu Einkommen, materiellen Lebensbedingungen, Gesundheit und Bildung, die es erlauben, individuelle Veränderungen dieser Merkmale über einen Zeitraum von maximal vier Jahren zu verfolgen.

Eine zentrale Schwierigkeit europäischer Daten betrifft ihre Vergleichbarkeit. Sowohl die Daten des EU-LFS als auch der EU-SILC sind outputharmonisiert (Körner/Meyer 2005, Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band). Dies bedeutet, dass für ihre Erhebung Rahmenbedingungen vorgegeben werden, die Spielraum für nationale Besonderheiten lassen. Die einzelnen Merkmale werden also nicht in exakt derselben Weise erhoben, sondern so, wie es den Verhältnissen im jeweiligen Land am ehesten entspricht. In vergleichbare Form werden die Merkmale nachträglich gebracht; d.h. auf der Basis mehr oder weniger unterschiedlicher Ausgangsvariablen werden harmonisierte Zielvariablen gebildet. Damit sind Zuordnungsprobleme und Vergrößerungen verbunden.

---

## 71.4 Datenzugang

Für die Wissenschaft stellt sich beim Zugang zu amtlichen Mikrodaten das Problem einer möglichen Identifizierung konkreter Personen. Dies würde gegen die Regeln des Datenschutzes verstößen. Daher werden die Daten nur in anonymisierter Form weitergegeben. Für die Öffentlichkeit frei verfügbar sind *absolut anonymisierte* Public Use Files sowie Campus Files, die für Studium und Lehre entwickelt wurden und im Internet heruntergeladen werden können. Sie enthalten nur einen kleinen Teil der Fälle des ursprünglichen Datensatzes und eine reduzierte Anzahl an Merkmalen. Für wissenschaftliche Einrichtungen werden zum Zweck der Forschung *faktisch anonymisierte* Scientific Use Files bereitgestellt, bei denen die einzelnen Angaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft konkreten Personen zugeordnet werden können (§16 (6) Bundesstatistikgesetz). Im Fall des Mikrozensus handelt es sich dabei um 70%-Unterstichproben der jeweiligen Originaldaten, die als systematische Zufallsauswahl auf Haushaltsebene gezogen werden und in denen einzelne Merkmale vergröbert sind. Sie liegen für ausgewählte Jahre ab 1973 und für alle Jahre ab 1995 vor. Auch die Daten des EU-LFS und der EU-SILC stehen in Form von Scientific Use Files zur Verfügung. Auf lediglich *formal anonymisierte* Daten, die noch weniger vergröbert sind und alle Fälle umfassen, können Wissenschaftler mittels kontrollierter Datenfernverarbeitung oder während eines Gastaufenthalts in einem Forschungsdatenzentrum zugreifen.

## 71.5 Bezugsquellen der Daten

Die *Campus-Files* sind zugänglich über: [www.forschungsdatenzentrum.de/campus-file.asp](http://www.forschungsdatenzentrum.de/campus-file.asp)

Einen umfassenden Service zu den *Scientific Use Files* und Hinweise zu ihrer Bestellung werden vom German Microdata Lab der GESIS zur Verfügung gestellt: [www.gesis.org/gml](http://www.gesis.org/gml).

Die *Scientific Use Files* des *EU-LFS* und der *EU-SILC* können über Eurostat bezogen werden: [epp.eurostat.ec.europa.eu/portal/page/portal/microdata/introduction](http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/introduction). Beim German Microdata Lab der GESIS ist auch zu diesen Daten ein Serviceangebot abrufbar: [www.gesis.org/gml](http://www.gesis.org/gml).

Über ein *Forschungszentrum* verfügen u.a. die Statistischen Ämter des Bundes und der Länder: [www.forschungsdatenzentrum.de](http://www.forschungsdatenzentrum.de), das Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit: [fdz.iab.de](http://fdz.iab.de) sowie die Deutsche Rentenversicherung: [www.fdz-rv.de](http://www.fdz-rv.de).

Eine Übersicht über alle (akkreditierten) Forschungsdaten- und Datenservicezentren bietet der Rat für Sozial- und Wirtschaftsdaten (RatSWD): [www.ratswd.de/forschungsdaten/fdz](http://www.ratswd.de/forschungsdaten/fdz)

## Literatur

- Bick, Wolfgang/Mann, Reinhard/Müller, Paul J. (Hg.) (1994): Sozialforschung und Verwaltungsdaten. Stuttgart: Klett-Cotta
- Bick, Wolfgang/Müller, Paul J. (1994): Sozialwissenschaftliche Datenkunde für prozessproduzierte Daten. Entstehungsbedingungen und Indikatorenqualität. In: Bick et al. (Hg.): 123-159
- Diekmann, Andreas (Hg.) (2006): Methoden der Sozialforschung. Wiesbaden: VS Verlag
- Emmerling, Dieter/Riede, Thomas (1997): 40 Jahre Mikrozensus. In: Wirtschaft und Statistik 3/97: 160-174
- Gabler, Siegfried/Hoffmeyer-Zlotnik, Jürgen/Krebs, Dagmar (Hg.) (1994): Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag
- Heidenreich, Hans-Joachim/Herter-Eschweiler, Robert (2002): Längsschnittdaten aus dem Mikrozensus. Basis für neue Analysemöglichkeiten. In: Wirtschaft und Statistik 8/02: 669-680
- Hoffmeyer-Zlotnik, Jürgen/Harkness, Janet A. (Hg.) (2005): Methodological Aspects in Cross-National Research. ZUMA-Nachrichten Spezial, Band 11. Mannheim
- Körner, Thomas/Meyer, Iris (2005): Harmonising Socio-Demographic Information in Household Surveys of Official Statistics. In: Hoffmeyer-Zlotnik /Harkness (Hg.): 149-162
- Lengerer, Andrea/Janßen, Andrea/Bohr, Jeanette (2007): Familiensoziologische Analysepotenziale des Mikrozensus. In: Zeitschrift für Familienforschung 19: 186-209
- Meyer, Kurt (1994): Zum Auswahlplan des Mikrozensus ab 1990. In: Gabler et al. (Hg.): 106-111
- Schimpl-Neimanns, Bernhard (2005): Bildungsverläufe im Mikrozensuspanel 1996-1999. Besuch der gymnasialen Oberstufe bis zum Abitur. ZUMA-Arbeitsbericht 2005/02. Mannheim
- Statistisches Bundesamt (2011): Qualitätsbericht Mikrozensus 2010. Wiesbaden. [www.destatis.de/DE/Publikationen/Qualitaetsberichte/Bevoelkerung/Mikrozensus2010.pdf](http://www.destatis.de/DE/Publikationen/Qualitaetsberichte/Bevoelkerung/Mikrozensus2010.pdf) (Stand: Juni 2013)
- Weichbold, Martin/Bacher, Johann/Wolf, Christof (Hg.) (2009): Herausforderungen und Grenzen der Umfrageforschung. Wiesbaden: VS Verlag
- Wirth, Heike/Müller, Walter (2006): Mikrodaten der amtlichen Statistik – ihr Potenzial in der empirischen Sozialforschung. In: Diekmann (Hg.): 93-127
- Wolf, Christof/Lüttinger, Paul (2009): Proxy-Angaben im deutschen Mikrozensus. In: Weichbold et al. (Hg.): 395-412
- Wolter, Felix/Schiener, Jürgen (2009): Einkommenseffekte beruflicher Weiterbildung. Empirische Analysen auf Basis des Mikrozensus-Panels. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 61: 90-117

Peter Graeff

---

## 72.1 Was sind Aggregatdaten?

Als *Aggregatdatenanalysen* bezeichnet man Studien, deren Daten aus aggregierten Einheiten bestehen, die dann mit Hilfe statistischer Verfahren ausgewertet werden. *Aggregate*, also die meistens nach mathematischen oder statistischen Regeln angefertigten Zusammenfassungen von Daten, beziehen sich beispielsweise auf Organisationen, Regionen oder Nationen. Entsprechend stehen bei Aggregatdatenanalysen Fragestellungen über Zusammenhänge und Wirkungen von Länder- oder Gesellschaftscharakteristika im Vordergrund.

---

## 72.2 Die Erhebung von Aggregatdaten

Daten, die in den Sozialwissenschaften Verwendung finden, existieren auf unterschiedlichen Ebenen, typischerweise unterschieden nach Mikro-, Meso- und Makroebene. *Individualdaten*, also Ausprägungen über Merkmale von Individuen, sind oftmals die basale Informationsbasis auf der Mikroebene, mit der Aggregatdaten hergestellt werden können. Individualdaten werden in den Sozialwissenschaften beispielsweise in Befragungen (Reinecke, Kapitel 44 in diesem Band), Experimenten (Eifler, Kapitel 11 in diesem Band) oder Beobachtungen (Thierbach/Petschick, Kapitel 66 in diesem Band) erhoben und können mit Hilfe eines Aggregierungsprozesses zu Aggregatwerten zusammengefasst (Lück/Landrock, Kapitel 28 in diesem Band) werden. Nimmt man als Beispiel die Note von Schülern in einem Mathematiktest, dann lassen sich die Noten einer Klasse etwa in einem Mittelwert zusammenfassen, der darüber Auskunft gibt, wie die durchschnittliche Schülerleistung in dieser Klasse ist. Fasst man die Daten zu größeren (Aggregat-)Einheiten zusammen, etwa durchschnittliche Mathematikleistungen in Schulen oder in Bundesländern (Mesoebene), erhöht sich die Aggregatstufe weiter. Daten auf Länderebene stellen eine der höchsten Aggregatstufen dar, die in der Soziologie und Politikwissenschaft von

besonderem Interesse ist. Hier wird die durchschnittliche Mathematikleistung der Schüler eines Landes auf der Makroebene abgebildet, wie das beispielsweise bei den PISA-Studien der Fall ist, bei denen die Aggregatwerte der Länderleistungen in die Diskussion um den (relativen) nationalen Bildungsstandard eingehen.

Neben diesen mikrofundierten Aggregatinformationen existieren auch *Kollektivmerkmale*, also quantifizierte Eigenschaften von Aggregaten bzw. Kollektiven, die bei ihrer Entstehung nicht auf niedrigere Ebenen zurückgeführt werden können. So ist beispielsweise die Ausprägung politischer Rechte in einem Land eine Information, die sich in ihrer Entstehung auf das Aggregat als Ganzes bezieht.

Aufbauend auf Einzelindikatoren werden häufig *Mischindizes* erstellt, bei denen verschiedene Inhalte kombiniert werden. Ein prominentes Beispiel ist der „(Un-)Happy Planet Index“, der für jedes Land einen Wert aus der gemittelten Lebenszufriedenheit und Lebenserwartung und dem sogenannten ökologischen Fußabdruck errechnet. Auf der Messebene werden durch die gleichzeitige Verwendung von Einstellungs- und Bevölkerungsdaten sowie von geographischen Informationen unterschiedliche theoretische Konzepte miteinander vermischt. Der ökologische Fußabdruck ist ein Maß für die biologisch produktive Land- und Wasserfläche, die von den Landesbewohnern genutzt wird, relativ zur verfügbaren Land- und Wasserfläche, die nötig wäre, um Produktions- und Entsorgungsbedingungen sicherzustellen (Kitzes/Wackernagel 2009: 813). Die Lebenszufriedenheit wird über ein Item (Porst, Latechva/Davidov, Kapitel 50 und 55 in diesem Band) bei einer Bevölkerungsbefragung gemessen (z.B.: „Wenn man alles mit einbezieht, wie zufrieden sind Sie zurzeit mit ihrem Leben als Ganzem?“), das dann als Aggregatwert für ein Land in die Berechnung eingeht. Die Lebenserwartung wird über die Jahre erfasst, die eine Person im Durchschnitt bei ihrer Geburt zu erwarten hat. Andere Indizes mischen zwar ebenfalls verschiedene Datenquellen, beziehen sich aber auf ein gemeinsames Konstrukt (Burzan, Kapitel 81 in diesem Band) wie beispielsweise der Index der Ökonomischen Freiheit, der über unterschiedliche Subgruppen von Variablen die Bereiche abbildet, in denen sich wirtschaftliche Freiheit ausdrücken kann.

---

## 72.3 Verwendungsmöglichkeiten

Wenn Aggregatdaten zur Überprüfung von theoretischen Aussagen herangezogen werden, können Sie in zwei *rivalisierenden Erklärungswegen* Verwendung finden. Eine der klassischen Studien, die mit Aggregatdaten arbeitet und anhand der die verschiedenen sozialwissenschaftlichen Erklärungsweisen bei unterschiedlichen Datenniveaus exemplarisch aufgezeigt werden können, ist die Dissertation von Émile Durkheim ([1897] 1973). Dort geht er unter anderem den Fragen nach, wie und warum sich länderspezifische Selbstmordraten voneinander unterscheiden. Sein Anliegen war, aggregierte Einflüsse der Gesellschaft als Ursache für die höchst individuelle Handlung des Selbstmordes nachzuweisen. Dabei ging es Durkheim nicht um die Gründe, warum eine bestimmte Person Selbstmord

begangen hatte, vielmehr versuchte er die Frage zu beantworten, warum sich die Selbstmordhäufigkeiten, die bestimmte Nationen im Durchschnitt aufweisen, voneinander unterscheiden und wie diese sich über die Zeit verändern. Zur Beantwortung dieser Fragen wollte er nicht auf mikrosoziologische bzw. psychologische Erklärungen zurückgreifen, wie z.B. auf die Tatsache, ob ein Selbstmörder vor seinem Tod depressiv oder einem kritischen Lebensereignis ausgesetzt war. Vielmehr verwendete er gesellschaftliche Variablen wie beispielsweise das Konstrukt der Anomie für die Erklärung von Selbstmordraten. Durkheim ([1897] 1973) bezeichnet mit dem Begriff „Anomie“ einen gesellschaftlichen Zustand, in dem Werte und Normen ihren regulierenden und integrierenden Einfluss verlieren. Das Beispiel dieser klassischen Studie verdeutlicht die unterschiedlichen Möglichkeiten, bei Erklärungen in den Sozialwissenschaften auf unterschiedliche Datenniveaus zurückzugreifen. Aggregierte Phänomene wie Selbstmordraten können auf Erklärungsvariablen auf aggregiertem Betrachtungsniveau (wie die Anomie) zurückgeführt werden.

Ein zweiter Erklärungsweg betont die Möglichkeit, *aggregierte Phänomene auf Einflüsse der Mikroebene zurückzuführen*. Die insbesondere von Coleman (1990) vorgeschlagene Vorgehensweise zur Erklärung sozialer Phänomene impliziert, dass die Erklärung von Merkmalsausprägungen auf einem Aggregatniveau (etwa der wirtschaftlichen Leistungsfähigkeit einer Gesellschaft) nur über einen Rückgriff auf die Mikroebene möglich ist. In diesem Erklärungsmodell wird ausgehend von einer Aggregatebene (den Makrobedingungen) auf eine Mikroebene, etwa der von Individuen, zurückgeschlossen, um dann kollektive Effekte mit Hilfe der Aggregation von Einzeleffekten auf der Mikroebene zu erklären (Raub et al. 2011). Zur Illustrierung dieses Erklärungsschemas werden oft Beispiele über kollektives Verhalten angeführt, die dann ein soziales Phänomen auf der Makroebene hervorbringen. Ein solches Beispiel ist etwa die Panik, die durch unachtsamen Gebrauch von Feuerwerkskörpern in einem ausverkauften Fußballstadion ausbricht. Die Erklärung des kollektiven Panikverhaltens beginnt mit einer Verknüpfung der Annahmen (auf der Makroebene) über die relevanten Elemente der Situation in diesem Stadion (wie etwa der Möglichkeit, Feuerwerkskörper mitzunehmen oder dem Auslastungsgrad des Stadions) und der Annahmen (auf der Mikroebene) darüber, wie Stadionbesucher auf diese Situationselemente reagieren. Jedes Individuum hat beispielsweise die Möglichkeit, ruhig den Ausgang zu erreichen oder ohne Rücksicht auf andere loszustürmen. Auf der Mikroebene macht ein solcher Erklärungsweg Aussagen darüber, unter welchen Bedingungen Individuen sich in einer solchen Situation zwischen ruhig gehen und losstürmen entscheiden. Wenn sich nun viele Stadionbesucher auf der Mikroebene dafür entscheiden loszustürmen, entsteht auf der Makroebene das kollektive Verhalten der Panik. Ähnlich kann beispielsweise die wirtschaftliche Leistungsfähigkeit einer Gesellschaft auf die kollektive Spar- und Konsumneigung ihrer einzelnen Wirtschaftsakteure zurückgeführt werden.

## 72.4 Datenquellen

Seit einigen Jahren sind im Internet viele Aggregatdaten kostenlos verfügbar. Nahezu alle großen Datenproduzenten, wissenschaftliche Forschungseinrichtungen oder Behörden stellen Daten und Statistiken zur eigenen Nutzung auf Anfrage oder zum direkten Download zur Verfügung.

Möchte man sich zunächst einen Überblick über ökonomische wie sozialwissenschaftliche Größen für bestimmte Länder verschaffen, so kann man auf das jährlich aktualisierte „CIA World Fact Book“ zurückgreifen. Es umreißt knapp und verständlich die sozialen, politischen und wirtschaftlichen Verhältnisse in allen Ländern der Erde.

*Wirtschaftliche Daten* für Länder und Regionen werden von der Weltbank und anderen Wirtschaftsorganisationen (wie dem „International Monetary Fund“) zur Verfügung gestellt. *Einstellungs-, Zugehörigkeits- oder Verhaltensdaten*, wie sie etwa im „World Value Survey“ verwendet werden, werden durch Zusammenschlüsse von Forschungsorganisationen und -instituten hergestellt und sind zum großen Teil (nach Registrierung) zugänglich. Das gilt auch für die *politikwissenschaftlichen Daten* über demokratische Verhältnisse (beispielsweise durch das „Polity Projekt“), Menschenrechte (wie das „CIRI Human Rights“) oder Konflikte („COW Projekt“). Darüber hinaus lassen sich über Gesundheitsorganisationen (wie die WHO) aggregierte Daten über gesundheitsrelevante Einflussfaktoren wie Drogenkonsum oder die Verbreitung von Krankheiten beziehen.

Im Zuge der Verbreitung des Internets fällt es auch immer leichter, nach bestimmten Daten (z.B. nach Informationen über länderspezifische Einkommensungleichheit) zu suchen. Einige Websites (wie der „Macro Data Guide“, [www.nsd.uib.no/macrodataloguide](http://www.nsd.uib.no/macrodataloguide)) haben sich darauf spezialisiert, einen Überblick über aggregierte Daten zu geben und erleichtern zusätzlich das Auffinden von Datensätzen und Informationen.

---

## 72.5 Methodologische Probleme

Aggregatdaten, die in der vergleichenden Sozialforschung verwendet werden, bringen häufig Messfehlerproblematiken in die Analyse ein. Da die Daten immer das Ergebnis eines Aggregierungsprozesses sind, gehören *Unterschiede bei der Berechnungsgrundlage* zu diesen typischen Schwierigkeiten. Wird beispielsweise die Arbeitslosenrate im Ländervergleich als Indikator für das Ausmaß der Erwerbslosigkeit verwendet, dann bringen bereits länderspezifisch unterschiedliche Maßstäbe für die Klassifikation von Beschäftigungen und Nicht-Beschäftigungen verschiedene Raten hervor. Das gilt auch bei Messungen innerhalb eines Landes über die Zeit. So wurde z.B. die Arbeitslosenquote in Deutschland ab 2005 mit den erwerbsfähigen Sozialhilfeempfängern berechnet, vorher nicht.

Zu den Besonderheiten von Aggregatdaten gehört es, dass sie nur in seltenen Fällen von Forschern selbst erhoben werden, wie das beispielsweise bei Individualdaten durch eigene Befragungen (Reinecke, Kapitel 44 in diesem Band) geschieht. Vielmehr werden Aggregatdaten oftmals als Sekundärdaten (Ohly/Weber, Mochmann, Schupp, Medjedović, Kapitel 5, 14, 73 und 13 in diesem Band) verwendet, die ursprünglich von Unternehmen

(wie Risikobewertungsfirmen) oder Organisationen (wie der Weltbank) gesammelt oder hergestellt werden. Zu diesen Daten gehören ökonometrische bzw. ökonomische Informationen (wie das „Bruttonsozialprodukt pro Kopf“), die von den nationalen Behörden oder Wirtschaftsunternehmen bzw. Wirtschaftsorganisationen (Liebig et al., Kapitel 75 in diesem Band) regelmäßig veröffentlicht werden. Eine weitere häufig verwendete Gruppe sind politische Daten (etwa die „Ausprägung der Demokratie eines Landes“), die in den überwiegenden Fällen von wissenschaftlichen Forschungsinstituten hergestellt werden, oder Ereignisdaten (wie die „Anzahl der Kriege pro Land“) (Mika/Stegemann, Pötter/Prein, Kapitel 85 und 86 in diesem Band), die meist über Auszählungsprozesse aus den Medien (z.B. einer Tageszeitung) (Klein, Kapitel 64 in diesem Band) produziert werden. Darüber hinaus gehören auch Einschätzungsdaten (wie der „Corruption Perception Index“) zu den Datengruppen, die häufig Eingang in Aggregatdatenanalysen finden. Im Falle des „Corruption Perception Index“ treffen Landsexperten eine Aussage über das Ausmaß der dort herrschenden Korruption, was als Ausgangsbasis für eine Einschätzung des tatsächlichen Korruptionsniveaus genommen wird.

Die Datengruppen sind von ihrer *Herkunft* somit sehr *heterogen*. Sie stammen aus verschiedenen Quellen mit unterschiedlichen Interessenlagen und Zielsetzungen. Bei Daten von Landesbehörden kann eine politische Absicht ihrer Publikation zum Tragen kommen, wie das beispielsweise der Chinesischen Regierung für ökonomische Statistiken bis etwa 2009 unterstellt wurde (Liu/Yang 2009). Da die statistischen Ämter in China nicht unabhängig von der Regierung arbeiteten und die Beförderung von Beamten direkt mit dem durch ökonomische Kennzahlen dokumentierten Wirtschaftserfolg (wie dem Wirtschaftswachstum) zusammenhingen, greifen Gesetze gegen Datenfälschung, sofern überhaupt vorhanden, nur unzureichend.

Bei Daten aus privat finanzierten Instituten kann ein kommerzielles Interesse der Datenproduzenten auftreten, das möglicherweise in einem Selektionseffekt von Variablen und betrachteten Fällen mündet. Flächenmäßig kleine, ökonomisch unbedeutende oder gering besiedelte Länder können beispielsweise für Indexproduzenten, deren Kunden die Daten für Investmententscheidungen verwenden, weniger relevant sein, so dass das Interesse an einer hohe Qualität dieser Länderdaten gering sein kann.

Viele Aggregatdaten, die von statistischen Landesbehörden herausgegeben werden und die dann in einen ländervergleichenden Datensatz einfließen, stammen aus (meist nur selten durchgeföhrten) Bevölkerungsstichproben oder Volkszählungen (Häder/Häder, Kapitel 18 in diesem Band). Ein Beispiel ist die Erstellung von Indikatoren der Einkommens- oder Vermögensverteilung (Burzan, Kapitel 81 in diesem Band), wie z.B. der Gini-Index (Huinkink, Kapitel 82 in diesem Band). Die unterschiedliche Stichprobenzusammensetzung, die verschiedenen technischen Möglichkeiten der Erhebung und die Gelegenheiten, potentielle Befragte auch tatsächlich zu erreichen bzw. ihre finanzielle Situation zuverlässig zu erfassen, variieren stark über Länder hinweg.

Andere soziale Phänomene besitzen schon durch die Prozesse ihrer Erhebung Messfehler (Krebs/Menold, Faulbaum, Kapitel 30 und 31 in diesem Band), die zu Abweichungen von den wahren Raten führen können. Dazu gehören beispielsweise kulturelle Verzerrungen von landesspezifisch abgelehnten oder besonders bevorzugten Untersuchungsobjekten.

So sind offizielle Selbstmordraten oder Scheidungsraten in einigen Ländern, in denen diese Phänomene verpönt sind und zu sozialer Missbilligung der Angehörigen führen, im Gegensatz zu anderen Ländern ungewöhnlich niedrig. Für Umfragedaten kann erschwerend hinzukommen, dass die Befragten eine kulturell geprägte und daher unterschiedliche Deutung der Fragen vornehmen und ihre Antworten entsprechend auswählen.

Aggregierte Ereignisdaten wie die „Anzahl von militärischen Konflikten eines Landes“ beruhen oft auf Rohdaten, die durch Auszählung der Ereignisse in der medialen Berichterstattung etwa in Zeitungen zustande kommen. Dabei spielt bereits die Selektion der Medien eine Rolle, weil diese einen spezifischen Fokus der Berichterstattung haben.

Solche und ähnliche Probleme haben bereits in der Mitte der 1950er-Jahre zu einer Diskussion um die Frage geführt, ob aggregierte Daten und hier insbesondere Länderdaten überhaupt verglichen werden dürfen. Die Gegner der *Vergleichbarkeit* argumentieren, dass zwei Objekte, die miteinander verglichen werden, der gleichen Klasse angehören müssen, d.h. sie haben entweder ein bestimmtes Attribut oder nicht. Wenn sie eines besitzen – und nur dann – kann verglichen werden, wo es weniger stark ausgeprägt ist. Klassifikation ist ihres Erachtens ein qualitativer Prozess, der sich einer „natürlichen Sprache“ bedient, während Vergleichsordnungsbildung und Quantifikation beinhaltet. DeFelice (1980) hat dagegen eingewendet, dass „Vergleichbarkeit“ eine doppelte Bedeutung haben kann:

1. In seiner ersten Bedeutung kann Vergleichbarkeit Ähnlichkeit meinen. In diesem Sinne sind die föderalen Parlamentssysteme von Australien und Kanada miteinander vergleichbar, während die föderale Legislative von Kanada und den USA es nicht ist.
2. In der zweiten Bedeutung schließt Vergleichbarkeit die Ähnlichkeiten und Unterschiede ein, so dass die föderale Legislative von Kanada und den USA vergleichbar sind. Es ist nicht möglich, „Äpfel“ und „Birnen“ im Sinne der ersten Bedeutung miteinander zu vergleichen. Aber man kann ihre Attribute in der zweiten Bedeutung (z.B. hinsichtlich ihrer Farbe, ihrer Frische oder ihres Säuregehaltes) miteinander vergleichen.

DeFelice (1980) zeigt, dass eine vergleichende Analyse nicht notwendigerweise voraussetzt, dass die Vergleichsobjekte der gleichen Klasse angehören. Eine solche Restriktion verringert unnötigerweise die Aussagebreite. Es ist vielmehr nötig, dass sie potenziell das gleiche Attribut tragen können.

Weitere Kritikpunkte, die gegen Aggregatdatenanalysen vorgebracht wurden, betrafen beispielsweise das Multikollinearitätsproblem, das bei Ländervergleichen oft auftritt. Von *Multikollinearität* wird gesprochen, wenn mehrere erklärende Variablen hoch miteinander korrelieren, so dass sie die gleiche Erklärungsvarianz in ein statistisches Modell einbringen. Typischerweise korrelieren ökonomische und politische Variablen (wie etwa das „Wohlstands niveau“ und die „Ausprägung der Demokratie in einem Land“) je nach verwendetem Indikator sehr hoch miteinander.

Darüber hinaus kann ein *Validitäts- und Operationalisierungsproblem* (Krebs/Menold, Kapitel 30 in diesem Band) der Untersuchungskonstrukte auftreten, insbesondere dann, wenn keine direkte Messung der Zielvariablen existiert. So muss etwa die „Anomie in einem Land“ über Indikatoren (wie die „Anzahl der Scheidungen“) abgebildet werden, obwohl diese auch andere als anomische Bedingungen messen.

Des Weiteren enthalten reine Querschnittsuntersuchungen (Mochmann, Kapitel 14 in diesem Band) auf Länderebene nur relativ *wenige und nicht immer zufällig ausgewählte Fälle*, was auch inferenzstatistische Schlussfolgerungen insb. in der Mehrebenenanalyse (Pötschke, Kapitel 87 in diesem Band) gefährden kann.

Das bekannteste Problem bei der Schlussfolgerung aufgrund von Aggregatdaten ist der *ökologische Fehlschluss*. Ein solcher Fehlschluss kann auftreten, wenn man Schlussfolgerungen von empirischen Ergebnissen auf der Makroebene auf die Mikroebene überträgt. Robinson (1950) demonstrierte diesen Schlussfolgerungsfehler anhand von ökologischen Korrelationen, also Zusammenhangsberechnungen, bei denen Personengruppen verglichen werden. Er korrelierte für die USA den Anteil der schwarzen Bevölkerung mit dem Anteil der Analphabeten auf verschiedenen Aggregatebenen: auf der Ebene von neun geographischen Teilen der USA, für 48 Bundesstaaten und auf Individuumsebene. Auf der höchsten Aggregatebene der geographischen Teile der USA findet er eine Korrelation zwischen den Merkmalen von  $r = 0,946$ , auf der Ebene der (damaligen) 48 Bundesstaaten ist sie mit  $r = 0,773$  immer noch hoch, auf der Individuumsebene ist sie mit  $r = 0,203$  dann wesentlich kleiner. Ökologische Zusammenhänge fallen nach Robinson (1950) deshalb höher aus, weil die ökologischen Einheiten wie Bundesstaaten homogener als die weniger aggregierten Einheiten sind. Ökologische Zusammenhänge können also nicht ohne weitere Überprüfung als Ausgangsbasis für die Schlussfolgerung über die Zusammenhänge auf Individualniveau genommen werden.

Nach der Entdeckung des ökologischen Fehlschlusses wurde lebhaft über die Frage diskutiert, wie man ihn vermeiden und unter bestimmten Umständen doch von höheren Aggregatebenen auf die niedrigeren Ebenen schließen könne (z.B. Firebaugh 1978). Die Möglichkeiten der ökologischen statistischen Schlussfolgerungen wurden schließlich dadurch verbessert, dass über Mehrebenenmodelle eine gleichzeitige Modellschätzung mit Hilfe von Aggregat- und Individualdaten möglich wurde (Iverson 1991, Pötschke, Kapitel 87 in diesem Band).

---

## 72.6 Auswertung von Aggregatdaten

Die klassische Form der Aggregatdatenanalyse in den Sozialwissenschaften ist nicht über mehrere Auswertungsebenen angelegt, sondern eine einfache Querschnittsuntersuchung, bei der nur die unterschiedlichen Ausprägungen eines Merkmals über Länder oder andere Aggregateinheiten hinweg erklärt werden sollen. Ein typisches Beispiel ist die Erklärung des „Demokratieniveaus“ über Charakteristika von Ländern, wie deren „Wohlstands niveau“, die „Verteilung des Einkommens“ oder die „religiöse Prägung“. Die unabhängigen Variablen werden in einem regressionsanalytischen Ansatz (Kühnel/Dingelstedt, Kapitel 80 in diesem Band) dahingehend überprüft, welchen Beitrag sie zur Vorhersage der abhängigen Variablen leisten. Idealerweise werden die Operationalisierungen aus theoretischen Vorgaben abgeleitet, so dass die Aggregatdatenanalyse eine theoriegestützte Fragestellung überprüft.

Allerdings finden sich für theoretische Vorgaben nicht immer die entsprechenden Variablen auf Aggregatniveau, so dass man auf Proxies zugreifen muss. *Proxies* sind inhaltlich ähnliche Variablen, die anstatt der eigentlich benötigten Variablen in die Rechnungen aufgenommen werden. Die zuvor genannten Schwierigkeiten von Aggregatdaten führen dazu, dass für wohlstandsschwache Länder oft keine Werte für die Variablen verfügbar sind, die in reichen Ländern sehr einfach erhoben werden können (Dogan 1994), weil deren schwache Infrastruktur eine landesweite Datenerhebung nicht erlaubt. Es ist daher nicht verwunderlich, dass man in Aggregatdatenanalysen auch von *Imputationstechniken* Gebrauch macht (Cielebak/Rässler, Kapitel 26 in diesem Band), mit denen fehlende Werte (Engel/Schmidt, Kapitel 23 in diesem Band) ergänzt werden können (Horton/Lipsitz 2001).

Da es lediglich ungefähr 240 Länder gibt, sind den Modellschätzungen bei reinen Querschnittsuntersuchungen bereits durch die relativ geringe Fallzahl Grenzen gesetzt. Daher wurden Verfahren, bei denen neben der Querschnittsinformation auch Zeitinformationen berücksichtigt werden, als eine Möglichkeit gesehen, das Problem der kleinen Fallzahlen bei Aggregatdatenanalysen zu überwinden. Die Aggregateinheiten weisen bei diesen „*Time-Series Cross-Section*“-Untersuchungen (Metz/Thome, Kapitel 84 in diesem Band) eine Ausprägung der Merkmale über die Einheiten und über die Zeit (üblicherweise „Jahre“) auf. Außer den größeren Fallzahlen stehen den Vorteilen zusätzlicher Modellierungsmöglichkeiten solcher Modelle auch Nachteile entgegen, insbesondere durch die damit verbundenen notwendigen Modellspezifikationen. Schließlich bringt der Einbezug der Zeidimension eine neue Qualität in die Aggregatdatenanalysen, weil sich die Anwender nun auch Gedanken über die zeitlichen Dynamiken ihrer Erklärungsvariablen machen müssen, was die Möglichkeit einschließt, dass die Fehler über die Zeit hinweg korreliert sind oder die Einflüsse sich (endogen) aufschaukeln können. Dadurch kann es auch leicht zu Fehlspezifikationen der Modelle kommen (Kittel 2005).

In den letzten Jahrzehnten sind *Mehrerebenenmodelle* (Pötschke, Kapitel 87 in diesem Band) entstanden, bei denen Aggregatdaten und Individualdaten miteinander verbunden werden. Solche Modelle erklären die abhängige Variable auf der Mikroebene über Individual- und Kontexteinflüsse. Die Aggregatdaten werden als Charakteristika des Kontextes verwendet. Ein besonderer Vorteil dieser Modelle ist die Möglichkeit, Variablenwirkungen auf der Mikroebene mit den Kontexteffekten interagieren zu lassen (sogenannte Cross-Level-Interaktion). Damit können, je nach Kontextausprägung, die Einflüsse auf der Mikroebene spezifisch abgebildet werden. Ähnlich wie „*Time Series Cross-Section*“-Analysen gehören Mehrebenenmodelle zu den komplexen statistischen Verfahren, bei denen die am besten an die Daten angepassten Modelle sehr spezifisch ausfallen können. Ergebnisse mit diesen Modellen sind daher eher selten generalisierbar. Generell entziehen sich komplexe Modellierungen, bei denen Aggregatdaten eine Rolle spielen, der Idee der Sparsamkeit von verwendeten Variablen. Die Ergebnisse von klassischen Querschnittsanalysen insbesondere mit Länderdaten wurden vor allem nach dem Kriterium beurteilt, mit einer möglichst sparsamen Anzahl an Variablen (bzw. Parametern) eine möglichst große Erklärungskraft zu erreichen. Mehrebenenanalysen vermeiden allerdings den Nachteil von reinen Querschnittsanalysen mit Aggregatdaten, der darin liegt, dass Unterschiede auf niedrigeren Aussagenebenen unbeachtet bleiben.

## 72.7 Vorteile von Aggregatdatenanalysen

Wie zuvor erwähnt, treten Probleme bei Aggregatdatenanalysen sowohl auf der Daten- wie auf der Auswertungsebene auf. Bei der Masse der Schwierigkeiten ist man leicht geneigt, solche Verfahren als wissenschaftliche Methode gänzlich zu verwerfen. Dabei besitzen Aggregatanalysen jedoch eine Reihe von Vorteilen, die sie für die wissenschaftliche Analyse bestimmter Fragestellungen prädestinieren (Jahn 2009: 191). Als Analyseinstrument eignen sie sich insbesondere für die *sozialwissenschaftlichen Theorien, die von einem Makroansatz ausgehen* und in ihren Erklärungen kaum einen oder keinen Bezug auf niedriger gelagerte Ebenen nehmen. Dazu gehören beispielsweise bestimmte Theorien über Wohlfahrtsstaaten oder Klassifikationsansätze über demokratische Systeme. Wenn eine hinreichende Datenqualität vorliegt, können Aggregatdatenanalysen als ein quantitativer empirischer Prüfansatz dieser Theoriebildung verwendet werden.

Ein anderer, nicht unwesentlicher Vorteil (insbesondere gegenüber Analysen mit nicht frei zugänglichen Mikrodaten) erwächst aus der Tatsache, dass die meisten Aggregatdatenanalysen auf Sekundärdaten zurückgreifen und damit eine *Replizierbarkeit der Ergebnisse* (Krebs/Menold, Kapitel 30 in diesem Band) möglich wird. Das erhöht nicht nur die Glaubwürdigkeit der Resultate, sondern ermöglicht und motiviert auch Anschlussforschungen.

Aggregatanalysen entsprechen darüber hinaus auch dem Bedürfnis, deskriptiven Darstellungen von Kollektivphänomenen eine *vertiefende statistische Analyse* hinzuzufügen. Zu nahezu jedem sozial-wissenschaftlichen Makrophänomen gibt es inzwischen einen relativ leicht verfügbaren Kennwert oder Index. Viele dieser Indizes werden von den Printmedien oder den politischen Institutionen aufgegriffen und für Ihre Zwecke eingesetzt, so dass sie gleichsam in die Tagespolitik oder in die Nachrichtenberichterstattung eingehen und dort auch kommentiert werden.

Ein Beispiel, das in der Tagespresse nahezu regelmäßig vorkommt, mag diesen Punkt illustrieren. Soziale Ungleichheit (Huinkink, Kapitel 82 in diesem Band) hinsichtlich der Einkommens- oder Vermögensverhältnisse gehört zu den häufig diskutierten Themen in den Medien. Die Darstellung solcher Ungleichheiten wird über statistische Kennwerte (wie den bereits erwähnten Gini-Index) vorgenommen und publiziert. Wenn solche nationalen Indikatoren eine steigende Einkommensungleichheit nahe legen, dann erfolgt die wissenschaftliche wie journalistische Berichterstattung mit Blick auf die negativen Konsequenzen, die sich aus einer möglicherweise „auseinanderklaffenden Schere zwischen Reich und Arm“ ergeben können. Aggregatdatenanalysen können die gesellschaftliche Bedeutung von solchen Befürchtungen durch die gleichzeitige Betrachtung konkurrierender Erklärungsvariablen auf eine umfassendere Beurteilungsgrundlage stellen und auch wissenschaftlich prüfen, ob eine hohe soziale Ungleichheit tatsächlich zu negativen Konsequenzen führt.

## Literatur

- Coleman, James. (1990): Foundations of Social Theory. Cambridge (MA): The Belknap Press of Harvard University Press
- DeFelice, E. Gene (1980): Comparison Misconceived: Common Nonsense in Comparative Politics. In: Comparative Politics 13: 119-124
- Dogan, Mattei/ Kzancilgil, Ali (Hg.) (1994): Comparing Nations. Oxford: Blackwell
- Dogan, Mattei (1994): Use and misuse of statistics in comparative research. In: Dogan/ Kzancilgil (Hg.): 35-70
- Durkheim, Emile (1897/1973): Der Selbstmord. Neuwied: Luchterhand
- Firebaugh, Glenn (1978): A rule for inferring individual-level relationships from aggregate data. In: American Sociological Review 43: 557-572
- Horton, Nicholas J./Lipsitz, Stuart R. (2001): Multiple Imputation in Practice: Comparison of Software Packages for Regression Models With Missing Variables. In: The American Statistician 55: 244-254
- Iversen, Gudmund R. (1991): Contextual Analysis. Sage: Thousand Oaks
- Jahn, D. (2009): Die Aggregatdatenanalyse in der vergleichenden Politikwissenschaft. In: Pickel et al. (Hg.): 173-196
- Kittel, B. (2005): Pooled Analysis in der ländervergleichenden Forschung. Probleme und Potenziale. In: Kropp/Minkenberg (Hg.): 96-115
- Kitzes, Justin/Wackernagel, Mathis (2009): Answers to common questions in ecological footprint accounting. Ecological Indicators 9: 812-817
- Kropp, Sabine/Minkenberg, Michael (Hg.) (2005): Vergleichen in der Politikwissenschaft. Wiesbaden: VS Verlag
- Liu, Junguo/Yang, HHong (2009): China Fights against Statistical Corruption. In: Science 325: 675-676
- Pickel, Susanne/Pickel, Gert/Lauth, Hans-Joachim/Jahn, Detlef (Hg.)(2009): Methoden der vergleichenden Politik- und Sozialwissenschaft. Neue Entwicklungen und Anwendungen. Berlin: Springer
- Raub, Werner/Buskens, Vincent/van Assen, Marcel (2011): Micro-Macro Links and Microfoundations in Sociology. In: Journal of Mathematical Sociology 35: 1-25
- Robinson, William S. (1950): Ecological Correlations and the Behavior of Individuals. In: American Sociological Review 15: 351-357

Jürgen Schupp

---

## 73.1 Was sind Paneldaten?

Unter Paneldaten versteht man im Bereich der empirischen Sozialforschung Daten auf Basis einer Untersuchungsanlage, bei der gleichzeitig drei Merkmale erfüllt sind. Es werden bei *denselben Untersuchungseinheiten* dieselben oder zumindest *die gleichen Inhalte* erhoben und die Daten werden *mehrfach*, also mindestens zweimal ermittelt.

Paneldaten werden dabei in der Regel als standardisierte Umfragen (Reinecke, Kapitel 44 in diesem Band) erhoben. Wenn wir in der sozialwissenschaftlichen Survey-Geschichte zurückblicken, so wurden Paneldaten erstmals in den 1930er Jahren im Bereich der Forschung zu Wahlpräferenzen *vor* einer Wahl und den tatsächlichen Wahlentscheidung *kurz nach* einer Wahl erhoben (Lazarsfeld/Fiske 1938). Paneldaten weisen demnach in ihrem Design einige Gemeinsamkeiten zum experimentellen Design auf (Eifler, Kapitel 11 in diesem Band).

Paneldaten dienen vor allem der empirischen Analyse von Veränderungen im Zeitverlauf, da mit Hilfe dieses Forschungsdesigns Fluktuationen wie Stabilitäten nicht allein im historische Zeitverlauf (Trend) (Mochmann, Metz/Thome, Kapitel 14 und 84 in diesem Band), sondern zudem im Lebensverlauf der jeweiligen Untersuchungseinheiten analysiert werden können. Das besondere Analysepotenzial von Paneldaten beruht also in der wiederholten Messung sowohl der abhängigen wie auch der unabhängigen Variablen. Das Untersuchungsdesign ermöglicht deshalb im Grundsatz die Identifizierung der kausalen Reihenfolge von Variablen im Sinne möglicher Analysen von Ursache-Wirkungs-Mechanismen (Kühnel/Dingelstedt, Kapitel 80 in diesem Band). Zwar muss man einschränkend einräumen, dass ein Paneldesign kein echtes experimentelles Design mit einer randomisierten Variation der unabhängigen Variable ist (Eifler, Kapitel 11 in diesem Band), aber gleichwohl zählt es mittlerweile zu den wenigen sozialwissenschaftlichen Untersuchungsdesigns (Stein, Kapitel 7 in diesem Band), mit deren Hilfe man die kausalanalytische Methodik empirisch annähern kann. Wenn „natürliche Experimente“ (Eifler, Kapitel 11

in diesem Band) während der Laufzeit eines Panels stattfinden, also die Änderung von Rahmenbedingungen für eine zufällig bestimmte Teilpopulation (z.B. Kurzschuljahre für einige Geburtsjahrgänge), sind Kausalanalysen ebenso möglich wie durch die Verbindung eines Panels mit einem Verhaltensexperiment bei zufallsbasierter Zuordnung in Kontroll- sowie Treatment-Gruppe.

In der überwiegenden Zahl der heutzutage für Forschungszwecke der Sekundäranalyse bereitstehenden Paneldaten sind (Befragungs-)Personen die jeweiligen *Untersuchungseinheiten* (Akremi, Kapitel 17 in diesem Band). Aber je nach Zielsetzung eines Panels kommen auch größere Einheiten, wie soziale Beziehungen und Aggregate (Graeff, Kapitel 72 in diesem Band), in Frage, also z.B. Haushalte, Familien oder Organisationseinheiten wie Firmen, Vereine, Verbände (Liebig/Meyermann/Gebel, Kapitel 75 in diesem Band) und regionale Einheiten, wie Wohnungen, Gemeinden, Länder (Lakes, Dangschat, Kapitel 77 und 78 in diesem Band).

Während sich bei einem Panel von Personen leicht sicherstellen lässt, dass *dieselben* Personen über die Zeit untersucht werden, ist die Fallabgrenzung bei größeren Einheiten über die Zeit keineswegs mehr so eindeutig sichergestellt, da bspw. Haushalte oder Betriebe vielfach ihre Größe und Zusammensetzung ändern und dies bei der Modellbildung, Analyse wie Interpretation entsprechend berücksichtigt werden muss. Die Analyse repräsentativer Panelstudien privater Haushalte erlaubt, menschliches Verhalten und die dem sozialen Wandel zu Grunde liegenden Mechanismen besser zu verstehen.

Die wiederholte Datenermittlung wird bei Paneldaten „*Erhebungswelle*“ genannt, wobei man bei mehr als zwei Datenerhebungen üblicherweise bestrebt ist, regelmäßige Zeitabstände zwischen den Erhebungswellen zu wählen. Paneldaten unterscheiden sich von einfachen wie auch von replikativ zu mehreren Zeitpunkten (d.h. im Zeitverlauf mit denselben Survey-Fragen) erhobenen Querschnittsdaten (Mochmann, Kapitel 14 in diesem Band) grundlegend, da bei Querschnittserhebungen für die Untersuchungseinheiten jeweils nur eine einzige Messung durchgeführt wird.

---

## 73.2 Sozialwissenschaftlich bedeutsame laufende Panelstudien

### 73.2.1 Sozio-oekonomisches Panel (SOEP)

Das Sozio-oekonomische Panel (SOEP) ([www.soep.de](http://www.soep.de)) ist die größte und am längsten laufende multidisziplinäre Langzeitstudie in Deutschland. Das SOEP ist eine repräsentative Wiederholungsbefragung privater Haushalte in Deutschland, die im jährlichen Rhythmus seit 1984 bei denselben Personen und Familien in der Bundesrepublik durchgeführt wird. Bereits im Juni 1990, also noch vor der Währungs-, Wirtschafts- und Sozialunion, wurde die Studie auf das Gebiet der ehemaligen DDR ausgeweitet, um die historisch einmalige Chance zu nutzen, in ausgewählten Lebensbereichen Längsschnittdaten zur Transformation einer Gesellschaft zu gewinnen. In den Jahren 1994/95 sowie 2013 wurden spezielle

„Zuwanderer-Stichproben“ eingeführt und 2002 erfolgte eine Ergänzung mit einer Stichprobe von Hocheinkommensbeziehern. Weitere zusätzliche repräsentative Stichproben-auffrischungen erfolgten in den Jahren 1998, 2000, 2002, 2006, 2011, 2012 und 2013.

Für die verschiedenen Stichproben des SOEP befragen jedes Jahr etwa 600 Interviewerinnen und Interviewer vom Umfrageinstitut TNS Infratest Sozialforschung etwa 30.000 Menschen in rund 15.000 Haushalten. Die so erhobenen Daten geben unter anderem Auskunft über Einkommen, Erwerbstätigkeit, Bildung und Gesundheit.

Insgesamt stehen im Kern-SOEP für den Zeitraum von 1984 bis 2012 bereits 255.978 Haushaltsinterviews und 484.125 Personeninterviews für (nicht nur) soziologisch ausgerichtete Sekundäranalysen zur Verfügung. Die anonymisierten Mikro-Datensätze des SOEP werden im SOEP-Forschungsdatenzentrum am Deutschen Institut für Wirtschaftsforschung (DIW Berlin) nutzerfreundlich aufbereitet, dokumentiert und Wissenschaftlern an Universitäten und anderen Forschungseinrichtungen zu Zwecken der Forschung und Lehre gebührenfrei zur Verfügung gestellt.

Die Daten des SOEP sind in zweifacher Hinsicht von sozialwissenschaftlichem Interesse.

1. Sie erlauben eine jährlich wiederholbare *Situationsanalyse*, die Stabilitäts- wie Veränderungsprozesse deutlich macht, und somit mit Hilfe sozialer Indikatoren einen Beitrag zur Beschreibung des sozialen Wandels in Deutschland leisten.
2. Diese Daten bieten die Basis, auf der theoretische Analysen in den Sozial- und Wirtschaftswissenschaften weiter vorangetrieben werden können, um Erklärungsansätze für das Verhalten von Individuen, von Haushalten, von Gruppen und Gesellschaften zu replizieren und/oder auch neu zu entwickeln. Die Motivation bei der Entwicklung des SOEP lag primär darin, Längsschnittdaten zur Prüfung kausaler Hypothesen zu gewinnen (Krupp 2008, Kühnel/Dingelstedt, Kapitel 80 in diesem Band), aber das SOEP zählt auch seit Jahren zum festen Bestandteil der Sozialberichterstattung zur Lebenssituation in Deutschland.

Zum SOEP vergleichbare *Haushaltspanelstudien* existieren weltweit in mehreren Ländern, so z.B. in den USA (PSID), Großbritannien (BHPS) und Australien (HILDA). Für einen eingeschränkten Satz an Variablen wurde ein international vergleichbares „*Cross National Equivalent File (CNEF)*“ geschaffen, mit dessen Hilfe vor allem im Bereich Arbeitsmarkt und Einkommen komparative Analysen für acht Länder ermöglicht werden (Lillard 2013). Dabei ist nicht zu erkennen, dass Haushaltspanels weltweit vorwiegend entwickelt wurden, um die Dynamik der Einkommensverteilung und -armut abzubilden. Die erste Panelstudie, die vor fast 50 Jahren in den USA begonnene PSID (<http://psidonline.isr.umich.edu/>), führt diesen Aspekt sogar im Namen: „Panel Study of Income Dynamics“, sie war auch bei der Konzeption und Realisierung des SOEP die zentrale Vorbildstudie. Entsprechend liegen sehr viele Veröffentlichungen auf Basis von Haushaltspanels vor, die primär auf Stabilität wie Veränderungen in Aggregatindikatoren der Einkommensungleichheit einer Gesamtbevölkerung abzielen.

Allerdings beschritt das SOEP sowohl von der Themenauswahl als auch von der Erhebungsmethode her im Vergleich zur Vorbildstudie PSID neue Wege. So wird im SOEP jeder Erwachsene um ein persönliches Interview zu seiner objektiven wie subjektiven Lebenssituation gebeten – im Gegensatz zur PSID, wo für Personen jenseits der Auskunftsperson, die in einem Mehrpersonenhaushalt leben, ein Proxyinterview die persönliche Situation charakterisiert. Zudem enthält das Befragungsprogramm des SOEP von Beginn an eine ganze Fülle an subjektiven Indikatoren zur Bewertung der eigenen Lebenssituation.

Mit der Verfügbarkeit von inzwischen fast 30 Erhebungswellen des SOEP wuchs einerseits das Potenzial für Analysen über den *Lebensverlauf* hinweg, aber vor allem nahmen auch – bedingt durch die erhöhte Zahl an Mehrgenerationenhaushalten sowie das spezifische Weiterverfolgungskonzept aller Personen von Stichprobenhaushalten – die Analysen zu intergenerationalen Zusammenhängen zu. Somit ist das SOEP seit geraumer Zeit bereits über eine „bloße“ Haushaltspanelstudie hinausgewachsen und hat sich zu einer prospektiv angelegten *Mehr-Kohorten-Studie* entwickelt. In einem Haushaltspanel tritt zwar auch der Effekt auf, dass Befragungspersonen, die an der ersten Welle teilnahmen, über die Zeit immer älter werden, jedoch wachsen im Verlauf der Langzeitstudie auch Kinder in das Befragungsalter, so dass rein demographisch eine Haushaltspanelstudie nicht „ausstirbt“.

In den letzten Jahren erfolgten im SOEP vorrangig zwei konzeptionelle Weiterentwicklungen (ausführlich, Schupp et al. 2008):

1. Das SOEP liefert einen originären Beitrag zum Wunsch aller sozial- und verhaltenswissenschaftlichen Teildisziplinen nach einer besseren Erklärung des menschlichen Verhaltens auf Basis einer interdisziplinären Datenerhebung (für einen Überblick Butz/Torrey 2006). Dabei fand eine forcierte Öffnung zu psychologischen Konzepten statt. Seit 2002 wurde eine Reihe von psychologischen Konzepten eingebracht, die sich auf Persönlichkeitsmerkmale beziehen, welche konzeptionell als weitgehend stabil angesehen werden können bzw. deren Stabilität und Handlungsrelevanz noch empirisch zu überprüfen bleiben.
2. Das SOEP folgt der Erkenntnis, dass für bessere Erklärungen des Verhaltens im Erwachsenenalter mehr Informationen über die (früheste) Kindheit von Individuen nötig sind. So werden seit dem Jahr 2000 im SOEP spezielle altersspezifische Erhebungsinstrumente für Kinder und Jugendliche eingeführt.

### **73.2.2 Survey of Health, Ageing and Retirement in Europe (SHARE)**

Mittlerweile gibt es in Deutschland eine wachsende Zahl an weiteren großen sozialwissenschaftlichen Panelstudien. So wurde 2004 erstmals der Survey of Health, Ageing and Retirement in Europe (SHARE, <http://www.share-project.org>) als repräsentative Befragung der über 49jährigen Bevölkerung in Deutschland und in zunächst 10 – inzwischen 19 – europäischen Ländern sowie in Israel bei zuletzt insgesamt 45.000 Personen erhoben. Im Abstand von zwei Jahren erfolgt seitdem jeweils die nächste Erhebungswelle

(Börsch-Supan 2013). Zusätzlich zur Fortführung der Längsschnittuntersuchung beinhaltet jede Befragungswelle ein neues thematisches Modul, 2011/12 z.B. Fragen zu sozialen Netzwerken.

### 73.2.3 Beziehungs- und Familienpanel (pairfam)

Im Jahr 2008 wurde das Beziehungs- und Familienpanel pairfam („Panel Analysis of Intimate Relationships and Family Dynamics“) (<http://www.pairfam.de>) gestartet, das eine multidisziplinäre Längsschnittstudie zur Erforschung der partnerschaftlichen und familialen Lebensformen in Deutschland ist. Die einmal jährlich erhobenen Befragungsdaten von über 12.000 bundesweit zufällig ausgewählten Personen der Geburtsjahrgänge 1971-73, 1981-83 und 1991-93 sowie von deren Partnern, Eltern und Kindern erlauben die differenzierte Analyse der Entwicklung von Partnerschafts- und Generationenbeziehungen in unterschiedlichen Lebensphasen (Huinink et al. 2011).

### 73.2.4 Nationales Bildungspanel (NEPS)

Das im Jahr 2009 begonnene Nationale Bildungspanel (NEPS) (<https://www.neps-data.de>) untersucht Bildungsprozesse und Kompetenzentwicklung in Deutschland, beginnend von früher Kindheit bis ins hohe Erwachsenenalter. In die erste Erhebungswelle waren ca. 60.000 zufällig gezogene Personen aus sechs Kohortenstichproben einbezogen sowie weitere 40.000 Personen des sozialen bzw. familialen Umfelds der Startkohorte.

---

## 73.3 Datenstruktur

Die Datenstruktur bei Paneldaten besteht aus drei Dimensionen. Zunächst gibt es ganz analog zu einem querschnittlichen Design (Stein, Mochmann, Kapitel 7 und 14 in diesem Band) die jeweiligen Untersuchungseinheiten (n) und eine Matrix aus abhängigen wie unabhängigen Variablen (y, x). Als weitere Ebene kommt die Dimension der Zeit (t) hinzu, wobei man beim praktischen Aufbau von Paneldatenstrukturen für Statistik-Analyseprogramme (z.B. SPSS, STATA oder R) zwischen zwei Datenformaten – „wide“ oder „long“ – unterscheidet (beim *wideformat* wird die Variablenmatrix und beim *longformat* die jeweilige Untersuchungseinheit mit der Dimension der Zeit indexiert). So kann man der ersten Zeile in Abb. 73.1 (long-Format) entnehmen, dass die befragte Person Nr. 10001 (n) im Jahr 2011 (t) 39 Stunden pro Woche (x) arbeitete und ein Monatseinkommen von 2.010 € (y) erzielte. In Zeile 2 sieht man, dass dieselbe Person im Jahr 2012 immer noch 39 Stunden pro Woche arbeitete, aber ein Monatseinkommen von 2.050 € erzielte.

Abbildung: Datenstruktur bei Paneldaten				
long - Format				
n (Personennummer)	t( Erhebungsjahr)	x (Arbeitszeit)	y (Einkommen)	
10001	2011	39	2010	
10001	2012	39	2050	
10001	2013	42	3000	
50002	2011	38	4000	
50002	2012	38	4020	
80001	2012	39	2000	
80001	2013	25	1500	

**Abb. 73.1** Datenstruktur bei Paneldaten

Unabhängig vom gewählten Datenformat sind bei der Nutzung von Paneldaten mit mehreren Erhebungswellen die Datenmatrizen vielfach nicht vollständig mit Informationen versehen (missing values) (Engel/Schmidt, Kapitel 23 in diesem Band), sei es aufgrund der Panelmortalität (siehe unten) bei einzelnen Untersuchungseinheiten oder weil Daten von neuen Panelteilnehmern erst zu einem späteren Zeitpunkt erhoben werden. In beiden Fällen spricht man von „*unbalanced panel data*“. Die klassische Paneldatenstruktur ist hingegen „*balanced*“, d.h. für sämtliche Untersuchungseinheiten liegen – idealerweise ohne jegliche einzelne „missing values“ – so viele Beobachtungen von abhängigen und unabhängigen Variablen vor, wie Erhebungswellen durchgeführt wurden.

Die Datenstruktur sozialwissenschaftlicher Paneldaten zeichnet sich häufig durch vergleichsweise viele Untersuchungseinheiten (large n) sowie im Verhältnis dazu wenigen Wellen und damit Messzeitpunkten (small t) aus. Sie unterscheidet sich damit von den in der Ökonomie oder auch in den Politikwissenschaften häufig anzutreffenden Zeitreihen (Metz/ Thome, Kapitel 84 in diesem Band). Dort ist die Situation in der Regel genau umgekehrt: Man verfügt über viele Messungen (large t), aber über vergleichsweise wenige Untersuchungseinheiten (small n), zudem handelt es sich meistens um Aggregatdaten (z.B. Arbeitslosenquoten, Wirtschaftswachstum und Äquivalenzeinkommen). Gleichwohl sind die (panel-)ökonometrischen Verfahren der Datenanalyse vielfach ähnlich, wenngleich bei den Schätzannahmen auch Unterschiede bleiben (Kittel 2006).

## 73.4 Besondere Probleme bei der Datenerhebung: Operationalisierung über die Zeit

### 73.4.1 Prospektive und retrospektive längsschnittliche Untersuchungsanlagen

In der methodischen Literatur werden Paneldaten vielfach auch als *prospektives längsschnittliches Forschungsdesign* im Unterschied zu retrospektiven Untersuchungsanlagen bezeichnet (Featherman 1980). Während bei einem prospektiven Design, bei dem peu à peu Daten erhoben werden, die Längsschnittinformationen erst nach mehreren Erhebungswellen, d.h. erst nach einiger Zeit vorliegen, benötigt das *retrospektive Design* lediglich eine einzige Datenerhebung (Querschnitt), bei der dann aber wesentlich mehr Informationen, nämlich auch solche, die sich auf die individuelle Vergangenheit beziehen, bei den Untersuchungseinheiten eingeholt werden. Vielfach werden aber auch in einer ersten oder zweiten Welle einer Panelstudie ergänzend zu Angaben zum Befragungszeitpunkt retrospektiv zumindest zentrale biographische Hintergrundfragen erhoben, indem bspw. auch monatliche Daten der letzten Beschäftigungssituation oder familiärer Ereignisse erhoben werden. In den Sozialwissenschaften etablierte diese retrospektive Untersuchungsanlage eine spezielle Form der Lebensverlaufsforschung, welche die Lebensläufe mehrerer Geburts- oder sozialer Kohorten (bspw. Schulabgänger- oder Scheidungskohorten) nachträglich erfragt, so dass der Prozess der individuellen Entwicklung und des sozialen Wandels gleichzeitig untersucht werden können.

Neuere Mehr-Kohorten-Studien benutzen den retrospektiven Erhebungsansatz nur noch als Ergänzung prospektiver Ansätze. Sobald nämlich Forschungsfragen auf die Operationalisierung von Variablen angewiesen sind, die bei retrospektiver Abfrage entweder für die Vergangenheit *nicht exakt erinnert werden* oder auch *systematisch verzerrt* sind (z.B. Einkommen, Präferenzen oder Lebenszufriedenheit), wird die Validität der Daten aus einem rein retrospektiven Längsschnittdesign stark eingeschränkt.

Deshalb sind prospektive Panelstudien die einzige Möglichkeit, um Längsschnittdaten für die Sozialwissenschaften zu gewinnen. Prospektive wie retrospektive Untersuchungen können, sofern sie die Zeit genau, also etwa auf Monatsebene, registrieren, mit statistischen Verfahren analysiert werden, die den zeitlichen Prozess modellieren, z.B. mit der Ereignisanalyse (Pötter/Prein, Kapitel 86 in diesem Band) oder auch mit der Sequenzmusteranalyse (Mika/Stegemann, Kapitel 85 in diesem Band).

### 73.4.2 Konstanz und Wandel der Inhalte

Doch auch für Panelstudien ist die Operationalisierung über die Zeit eine methodische Herausforderung, weshalb sich mittlerweile der methodische Einsatz der jeweiligen Erhebungs- und Messverfahren über die Zeit sowie Fragen der Selektivität einer dauerhaften

Teilnahme an Panelstudien zu einem eigenständigen Forschungsfeld etabliert haben (Lynn 2009). Zu den bekannten Problemen gehört zunächst eine Konstanz der Inhalte, die aber nicht unbedingt eine Persistenz der Fragen beinhaltet:

1. Auf der inhaltlichen Ebene *verändern sich im Zeitverlauf oft Inhalte und Begriffe*, die vielfach auch Teil der Forschungsfragen einer Panelstudie sind. So wurde bspw. vor wenigen Jahren der Wehr- und Zivildienst abgeschafft, so dass auch die entsprechende Kategorie der derzeit ausgeübten Tätigkeit angepasst werden musste. Ähnliches gilt für institutionelle Änderungen wie die Einführung von Arbeitslosengeld II im Jahr 2005 als neue Transferleistung sowie der Wegfall der Arbeitslosenhilfe.
2. Zum zweiten ist es bei Panelbefragungen günstig, *bestimmte Fragen nicht allen Teilnehmern zu stellen*, sondern lediglich denjenigen mit Veränderungen seit der letzten Welle. So wäre es sicherlich ermüdend, die Frage zum höchsten schulischen Bildungsabschluss jährlich erneut bei allen Panelteilnehmern abzufragen. Deshalb sind Panelstudien ab der zweiten Erhebungswelle durch direkte *Veränderungsfragen* geprägt. Das dabei entstehende Risiko, dass vor allem Statusereignisse nicht logisch konsistent in den Folgewellen beantwortet werden, kann in Panelstudien vor allem bei technologiegestützten Erhebungsverfahren wie CATI (Hüfken, Kapitel 46 in diesem Band) oder CAPI (Stocké, Kapitel 45 in diesem Band) durch die Rückerinnerung von Vorjahresangaben (*dependent interviewing*) in einer späteren Welle gemildert werden (Jäckle 2009).

Zudem sind die meisten vor allem der langlaufenden Panelstudien durch eine *Mischung von Wiederholungsfragen der letzten Welle, Fragen aus länger zurückliegenden Wellen und neuen Fragen* geprägt. Auf diese Weise können einerseits die längsschnittlichen Befragungsinhalte ausgeweitet und neue Inhalte berücksichtigt werden; vor allem aber steigert eine solche Mischung das Interesse zur Beantwortung der Fragen und mindert das Risiko der Panelmortalität. Aufgrund der Kumulation von Informationen über einzelne Befragte nach einigen Erhebungswellen kommt der strikten Anonymisierung der Befragungsdaten sowie dem Datenschutz bei Panelbefragungen eine besondere Rolle zu.

### **73.4.3 Konstanz und Wandel der Erhebungsmethoden**

Neben der Konstanz der Inhalte ist bei Panelerhebungen anzustreben, dass auch die Erhebungsmethoden von Welle zu Welle gleich bleiben. Es hat sich jedoch herausgestellt, dass sich zum Beispiel die Standards der Erhebungstechnik wandeln können (Ablösung von PAPI durch CAPI (Stocké, Kapitel 45 in diesem Band)). Ebenso kann es passieren, dass Studienteilnehmer zwar weiterhin bereit sind an einer Studie teilzunehmen, aber nicht länger von einem Interviewer (Glantz/Michael, Kapitel 21 in diesem Band) befragt werden möchten, sondern lieber selbst ihre Antworten eintragen, also einen schriftlich-postalischen Fragebogen (Reuband, Kapitel 47 in diesem Band) oder einen Online-Fragebogen (Wagner/Hering, Kapitel 48 in diesem Band) ausfüllen. Deshalb ist es bei Panelstudien

nicht ungewöhnlich, dass mehrere Erhebungsmethoden gleichzeitig zum Einsatz kommen (Mixed Mode-Befragung). In solchen Fällen kann die spezifisch gewählte Erhebungs methode selbst eine Ursache einer beobachteten Veränderung darstellen, und dieses Metho denartefakt muss bei der inhaltlichen Analyse entsprechend berücksichtigt werden.

### 73.4.4 Paneleffekt (Panel-Conditioning)

Ein weiteres methodisches Problem, das ausschließlich bei prospektiven Panelstudien auf tritt, stellt die mögliche *Veränderung des Antwortverhaltens der Studienteilnehmer* unmittel bar durch die wiederholte Erhebung dar. Dies in der Literatur als Paneeffekt oder „Panel Conditioning“ (Schnell 2012) diskutierte Problem kann sich zum Beispiel in Verfestigungen von bestimmten Einstellungen oder aber auch in Lerneffekten äußern. Diese führen dann dazu, dass man bspw. Skalen bei wiederholten Befragungen sorgsamer differenziert als bei einer Querschnitterhebung, oder auf Fragen, auf die man in einer Ersterhebung noch keine Antwort wusste („weiß nicht“), in späteren Wellen inhaltlich präziser antwortet. Es ist somit unklar, wie stark die Datenqualität durch Paneeffekte eingeschränkt wird.

---

## 73.5 Besondere Stichprobenprobleme: Panelmortalität

Die Analyse von Paneldaten (prospektive Längsschnittdaten) beruht auf der Grundlage einer *identischen Dauer-Stichprobe*. Damit ist gemeint, dass eine Stichprobe derselben Untersuchungseinheit längere Zeit bereit ist, an einer solchen Längsschnittstudie teilzu nehmen. Dies hat für die logistische Durchführung von Panelstudien weitreichende Konsequenzen: Die Schwierigkeit besteht darin, dieselben Untersuchungseinheiten für Zwecke der Datenerhebung nicht nur wiederholt anzutreffen, sondern auch für längere Zeit an eine Panelstudie zu binden.

### 73.5.1 Mobilität von Untersuchungseinheiten

Bei der Planung einer Panelstudie müssen also zum ersten hinsichtlich der *räumlichen Mobilität von Untersuchungseinheiten* festgelegt werden, auf welche Weise auch im Anschluss an Umzüge des gesamten Haushalts oder an Auszüge einzelner Personen eines Ursprungshaushalts die jeweiligen Personen weiterhin befragt werden sollen. In der Regel endet die regionale Weiterverfolgung an den Grenzen eines Untersuchungsgebiets (Weiterverfolgungsregeln, Verfahren der Adressrecherche).

### 73.5.2 Panelpflege

Zum zweiten erfordern Panelstudien in der Regel spezielle Maßnahmen der Teilnahmemotivation wie spezielle Broschüren und Informationsmaterial zu bisherigen Ergebnissen der Studie (Panelpflege), um eine Bindung der Untersuchungseinheiten für die Laufzeit der Studie aufzubauen und mögliche Verweigerungen möglichst gering zu halten (zu Best-Practice-Beispielen für Mobilität wie Panelpflege vgl. Schoeni et al. 2013). Vor allem müssen aus forschungsethischen wie datenschutzrechtlichen Gründen (Friedrichs, Mühlichen, Kapitel 3 und 4 in diesem Band) Panelteilnehmer in einer ersten Welle über die geplante Wiederbefragung informiert werden und hierzu auch explizit ihre Bereitschaft zur wiederholten Teilnahme an der Untersuchung erklären, wobei in der Regel jedoch auf eine schriftliche Einverständniserklärung verzichtet wird.

### 73.5.3 Panelmortalität

Die *Ausfallraten* (Engel/Schmidt, Kapitel 23 in diesem Band) aufgrund von Befragungsteilnehmern, die trotz grundsätzlicher Zusage zur wiederholten Teilnahme im Laufe der Zeit aus dem Panel ausscheiden, werden vielfach auch als *Panelmortalität* bezeichnet, wobei es hier zwei Typen zu trennen gilt:

1. Zum einen gibt es die *demografisch bedingten – und in der Regel stichprobentechnisch neutralen – Ausfälle* (Engel/Schmidt, Kapitel 23 in diesem Band). Dazu zählen z.B. Todesfälle von Studienteilnehmern, aber auch bestimmte Formen von Umzügen (bspw. ins Ausland), wenn es sich um eine regional bzw. national beschränkte Panelstudie handelt.
2. Zum anderen kann es aufgrund *gesunkener Teilnahmebereitschaft, Krankheit oder Nickerreichbarkeit der Studienteilnehmer* (Engel/Schmidt, Kapitel 23 in diesem Band) zu temporärer oder dauerhafter Panelmortalität kommen. Während demografisch bedingte Ausfälle zwar zu sinkenden Fallzahlen bei Panelanalysen führen, aber zur Aufrechterhaltung der Repräsentativität des Längsschnitts notwendig und damit erwünscht sind, erfolgen die übrigen Ausfälle in der Regel nicht zufällig, sondern weisen soziale Regelmäßigkeiten auf, die eine *Verzerrung der Panelstichprobe* zur Folge hat. Dies schränkt die Aussagekraft von Panelanalysen vor allem dann ein, wenn sich die „ausgefallenen“ Untersuchungseinheiten im Hinblick auf die relevanten Variablen der inhaltlichen Fragestellung einer Studie von den weiteren Teilnehmenden unterscheiden. Solche Selektivität wird im Fall von Haushaltspanelstudien (wie dem SOEP) durch die Einbeziehung von neuen Familienmitgliedern teilweise kompensiert.

Solange die Fallzahl von Panelstichproben groß genug bleibt, können die durch Panelmortalität hervorgerufenen Herausforderungen hingenommen und durch *Korrekturverfahren (sogenannte Längsschnittgewichtung)* angemessen berücksichtigt werden (Rendtel/Harms 2009). Es ist auch möglich, selektive Ausfallprozesse bei der Spezifikation und

Modellierung der statistischen Analysen explizit zu berücksichtigen und so zu (weitgehend) verzerrungsfreien Analyseergebnissen zu kommen. Die Umgewichtung und Modellierung von Ausfällen ist im Gegensatz zu Ausfällen bei Querschnitt-Stichproben bei Paneldaten sehr effektiv, da ja aus der Erstwelle (und ggf. weiteren Wellen) reichhaltige Informationen zu den Merkmalen der unerwünschten Ausfälle („Verweigerer“) vorliegen.

Die Höhe der Panelmortalität hängt auch ab vom Erfolg der Maßnahmen zur Teilnahmemotivation, der Zeitdauer zwischen den einzelnen Erhebungswellen und der Anzahl der Wellen, an denen teilgenommen wurde, wobei es sich bei der überwiegenden Zahl der in Deutschland durchgeführten Panelstudien um jährliche Wiederbefragungen handelt.

Für Verweigerungen spielen insbesondere die Befragungsdauer, das Interesse der Befragten an den Forschungsfragen sowie feldbedingte Faktoren (bspw. derselbe Interviewer wie im Vorjahr) eine Rolle. Dementsprechend schwanken auch die Ausfallquoten sehr stark und betragen z.B. im SOEP zwischen 5% bei Panelteilnehmern, die bereits seit mehreren Wellen an der Studie teilnehmen und immer wieder von denselben Interviewern befragt werden, und 30% bei Personen, die „erst“ zum zweiten Mal die gleichen Fragen beantworten sollen. Diesen Stichprobenschwund gilt es bereits bei der Konzeption und Planung einer Panelstudie zu berücksichtigen. Werden kommerzielle Umfrageinstitute mit der Durchführung einer Panelstudie beauftragt, wird eine Abschätzung der Panelmortalität in der Angebotserstellung in Abhängigkeit von den konkreten Maßnahmen der Panelpflege explizit berücksichtigt. Wenn die Selektionsprobleme aufgrund der Panelmortalität so gewichtig sind, dass die Forschungsfrage aufgrund zu geringer Fallzahlen nicht länger beantwortet werden kann, wird dem in einigen Panelstudien durch das spätere Hinzufügen spezieller Aufwuchsstichproben begegnet. Solche zeit- und sachgleich durchgeführten Aufwuchsstichproben liefern nicht sofort, sondern erst nach einigen Jahren eine Vergrößerung der Fallzahl für Längsschnittanalysen zu liefern. Jedoch erlaubt eine Aufwuchsstichprobe auch die Einbeziehung neuer Populationen (wie bspw. Neuzuwanderer), die bei der ursprünglichen Stichprobenziehung einer Panelstudie noch keine Chance hatten, in die Langzeitstudie einzbezogen zu werden.

### 73.5.4 Spezielle Paneldesigns zum Umgang mit Panelmortalität

#### 73.5.4.1 Rotierendes Panel

Vor allem zur Kontrolle des Einflusses von Panelmortalität existieren weitere Varianten spezieller Paneldesigns, die in der Regel Paneldaten nur für einen kürzeren Zeitraum – bspw. von fünf Jahren – bereitstellen. Die mit am häufigsten verwendete Form stellt ein *rotierendes Panel* dar.

Die in den Mitgliedstaaten der Europäischen Union im Auftrag der jeweiligen statistischen Ämter durchgeführte „European Union Statistics on Income and Living Conditions“ (EU-SILC) ist eine europäische Erhebung der Bevölkerung in Privathaushalten mit dem Ziel, zeitlich und international vergleichbare mehrdimensionale sowie komparative Paneldaten zu mehreren Themen zu erheben. Es werden dabei sowohl monetäre als auch

nichtmonetäre Aspekte nach objektiven und subjektiven Gesichtspunkten erhoben. Pro Jahr werden in Deutschland ca. 14.000 Haushalte mit 28.000 Personen befragt, wobei jährlich ca. ein Viertel der Stichprobe ausscheidet sowie ein Viertel komplett neu rekrutiert werden muss ([https://www.destatis.de/DE/Meta/AbisZ/EU\\_Silc.html](https://www.destatis.de/DE/Meta/AbisZ/EU_Silc.html)).

#### 73.5.4.2 Access Panel

Schließlich wurden in den letzten Jahren bei Befragungsinstituten zunehmend Stichproben von befragungsbereiten Personen rekrutiert, die ihre Bereitschaft erklären, anschließend wiederholt an Befragungen teilzunehmen. Solche *access panels* bestehen sowohl für schriftliche Befragungen sowie zunehmend als Online Access Panels (Wagner/Hering, Kapitel 48 in diesem Band). Das Ziel dieses Untersuchungsdesigns liegt vor allem in einer Senkung der Untersuchungskosten und weniger in der Gewinnung von Längsschnittinformationen der einzelnen Panelteilnehmer. Die Teilnehmer werden in der Regel für ihre jeweilige Teilnahme in besonderer Weise belohnt (bspw. durch kumulierbare Punkte, die in Waren eingetauscht werden können). Aufgrund der Vielzahl erhebungsmethodischer Probleme, insbesondere der schwer kontrollierbaren Selektivität der Panelteilnehmer, ist die Verwendung von Access Panels für wissenschaftliche sowie für verallgemeinerungsfähige Forschungsfragen nicht ohne weiteres geeignet.

---

### 73.6 Auswertung

Bei Vorliegen von Daten einer Panel-Studie sind vielfach selbst deskriptive Formen der Datenanalyse bereits interessant, da gerade bei verallgemeinerbaren Stichproben schon die Ausweisung von Veränderungen einer Variablen im Zeitverlauf und die entsprechende Trennung zwischen interindividuellen und intraindividuellen Veränderungen wichtige soziale Tatbestände darstellen können. Damit ist die Unterscheidung in der aggregierten Betrachtung einer ganzen Gesellschaft im Gegensatz zu einer auch im individuellen Lebenslauf aufgetretenen Veränderung gemeint. So ist es durchaus von sozialwissenschaftlichem und gesellschaftspolitischem Interesse, ob ein über die Zeit konstanter 15%iger Anteil von Menschen, deren Einkommen unterhalb der Armutsriskoschwelle liegt, wiederholt bei jeweils denselben Personen festgestellt wird oder ob es einen ausgeglichenen Saldo von Auf- wie Abstiegen von Armutsriskiken gab und lediglich die Hälfte der Betroffenen dem Risiko dauerhaft ausgesetzt waren.

Die Wahl komplexer Analyseverfahren von Paneldaten hängt zunächst vom jeweiligen Messniveau der abhängigen wie unabhängigen Variablen ab (Blasius/Baur, Kapitel 79 in diesem Band), aber auch davon, ob es sich um *zeitkonstante Variablen* (wie Geschlecht oder Migrationshintergrund) oder um *zeitinvariante Variablen* (wie Erwerbstatus oder Einkommen) handelt (für einen Überblick siehe Andreß et al. 2013). Die statistischen Analysemodelle von Paneldaten reichen von Strukturgleichungsmodellen (Finkel 1995), diversen Regressionsmodellen (Giesselmann/Windzio 2012), Ereignisanalysen (Blossfeld 2010, Pötter/Prein, Kapitel 86 in diesem Band), Sequenzdatenanalysen (Brüderl/Scherer 2005,

Mika/Stegemann, Kapitel 85 in diesem Band), latenten Wachstumsmodellen (Schmiedeck/Wolff 2010) bis hin zu Kausalanalysen mithilfe von Matchingverfahren (Gangl 2010). Als besonderer Vorteil von Paneldaten erweist sich dabei, dass sich die zeitliche Abfolge von Veränderungen modellieren und berechnen lässt und das in den Sozialwissenschaften vielfach anzutreffende Problem unbeobachteter Heterogenität, also nicht im Modell spezifizierte Unterschiede, sich zumindest im Vergleich zu Querschnittsdaten deutlich abmildern lässt (Brüderl 2010).

## Literatur

- Andreß, Hans-Jürgen/Golsch, Katrin/Schmidt, Alexander W. (2013): *Applied Panel Data Analysis for Economic and Social Surveys*. Berlin - Heidelberg: Springer
- Behnke, Joachim/Gschwend, Thomas/Schindler, Delia/Schnapp, Kai-Uwe (Hg.) (2006): Methoden der Politikwissenschaft. Baden-Baden: Nomos
- Blossfeld, Hans-Peter (2010): Survival- und Ereignisanalyse. In: Wolf/Best (Hg.): 995-1016
- Börsch-Supan, Axel (2013): The Survey of Health, Ageing and Retirement in Europe, SHARE. In: Kleiner et al. (Hg.): 123-134
- Brüderl, Josef (2010): Kausalanalyse mit Paneldaten. In: Wolf/Best (2010): 963-994
- Brüderl, Josef/Scherer, Stefani (2005): Methoden zur Analyse von Sequenzdaten. In: Diekmann (Hg.): 330-347
- Butz, William P./Torrey, Barbara B. (2006): Some Frontiers in Social Science. In: Science 312: 1898-1900
- Diekmann, Andreas (2005): Methoden der Sozialforschung. Wiesbaden: VS Verlag
- Featherman, David L. (1980): Retrospective Longitudinal Research: Methodological Considerations. In: Journal of Economics and Business 32: 152-169
- Finkel, Steven E. (1995): Causal Analysis with Panel Data. Beverly Hills: Sage
- Gangl, Markus (2010): Causal Inference in Sociological Research. In: Annual Review of Sociology 36: 21-48
- Giessmann, Marco/Windzio, Michael (2012): Regressionsmodelle zur Analyse von Paneldaten. Wiesbaden: VS Verlag
- Huinink, Johannes/Brüderl, Josef/Nauk, Bernhard/Walper, Sabine/Castiglioni, Laura/Feldhaus, Michael (2011): Panel Analysis of Intimate Relationships and Family Dynamics (pairfam) – Conceptual Framework and Design. In: Zeitschrift für Familienforschung 23: 77-101
- Jäckle, Annette (2009): Dependent Interviewing: A Framework and Application to Current Research. In: Lynn (Hg.): 93-111
- Kittel, Bernhard (2006): Panelanalyse. In: Behnke et al. (Hg.): 239-249
- Kleiner, Brian/Renschler, Isabelle/Wernli, Boris/Farago, Peter/Joye, Dominique (Hg.) (2013): Understanding Research Infrastructure in the Social Sciences. Zürich: Seismo Press
- Krupp, Hans-Jürgen (2008): Die Anfänge: Zur Entstehungsgeschichte des SOEP. In: *Vierteljahrshefte zur Wirtschaftsforschung* 77: 15-26
- Lazarsfeld, Paul/Fiske, Marjorie (1938): The “Panel” as a New Tool for Measuring Opinion. In: Public Opinion Quarterly 2: 596-612
- Lillard, Dean R. (2013): Cross-National Harmonization of Longitudinal Data: The Example of National Household Panels. In: Kleiner et al. (Hg.), 80-88
- Lynn, Peter (Hg.) (2009): Methodology of Longitudinal Surveys. Chichester: Wiley
- Rendtel, Ulrich/Harms, Thorsten (2009): Weighting and Calibration for Household Surveys. In: Lynn (Hg.): 265-286

- Schmiedek, Florian/Wolff, Julia K. (2010): Latente Wachstumskurvenmodelle. In: Wolf/Best (Hg.): 1017-1029
- Schnell, Rainer (2012): *Survey-Interviews - Methoden standardisierter Befragung*. Wiesbaden: VS Verlag
- Schoeni, Robert F/Stafford, Frank/McGonagle, Katherine A./Andreski, Patricia (2013): Response Rates in National Panel Surveys. In: *The Annals of the American Academy of Political and Social Science* 645: 60-87
- Schupp, Jürgen/Spiess, C. Katherina/Wagner, Gert G. (2008): Die verhaltenswissenschaftliche Weiterentwicklung des Erhebungsprogramms des SOEP. In: *Vierteljahrsshefte zur Wirtschaftsforschung* 77: 63-76
- Wolf, Christof/Best, Henning (Hg.) (2010): *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS Verlag

Nina Baur

---

## 74.1 Was sind Netzwerkdaten und was ist Netzwerkanalyse?

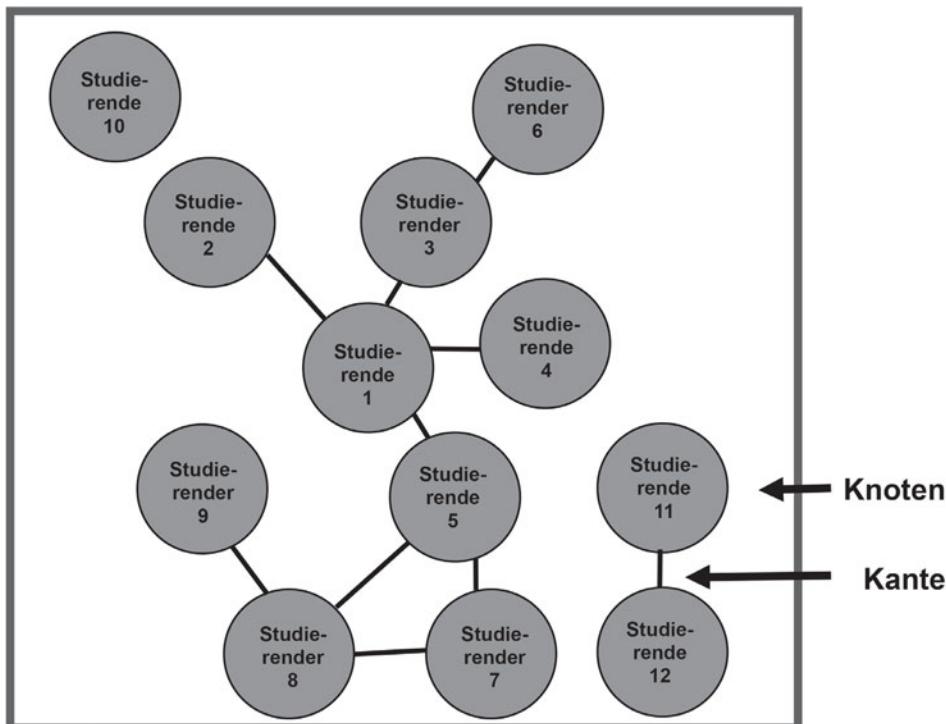
In der herkömmlichen Datenerhebung – etwa bei standardisierten Befragungen (Reinecke, Kapitel 44 in diesem Band) – erhebt man Daten über ein einziges Individuum, z.B. eine einzelne Person oder Organisation. So erheben etwa Lang und Neyer (2004) in einer Studie über Karrierewege in der Wissenschaft bei einer Stichprobe von 579 Psychologen Merkmale wie das Geschlecht, das Promotionsjahr und (als Indikator für die wissenschaftliche Produktivität) die Zahl der Publikationen, die in Einzel- bzw. als Ko-Autorenschaft verfasst wurden. Im Gegensatz zu diesen sogenannten *absoluten Variablen* erfassen sogenannte *relationale Variablen* die Beziehungen der Individuen zueinander. Entscheidend ist hier, dass man die Individuen nicht mehr einzeln behandeln kann, sondern ihre Vernetzung zueinander erfasst, weshalb solche Daten auch „*Netzwerkdaten*“ heißen. So erheben Lang und Neyer (2004) ebenfalls Informationen über die Ko-Autoren und Gutachter (also z.B. Dissertationsbetreuer) der befragten Wissenschaftler – wie viele sind es? Wie produktiv sind deren Netzwerkpartner? Welche Beziehungen haben die Netzwerkpartner untereinander? Auf Basis dieser Netzwerkdaten zeigen Lang und Neyer (2004), dass Forscher, die sehr produktive Betreuer haben sowie nicht nur alleine, sondern zusammen mit anderen publizieren, deutlich höhere Chancen auf eine Professur haben.

Ein älteres Beispiel stammt von Moreno (1954), der die sozialen Beziehungen und Gruppendynamiken von Kleingruppen (etwa in einer Schulkasse, einem studentischen Seminar oder in einem Unternehmen) untersuchte und u.a. wissen wollte, wer in einer Gruppe wen sympathisch fand und welche Gruppenmitglieder zusammenarbeiteten. Will man etwa – in Anlehnung an Moreno – wissen, wer in einem studentischen Seminar mit wem zusammenarbeitet, könnte man folgendermaßen vorgehen:

1. *Definition der Netzwerkgrenzen*. Man legt zunächst die Grundgesamtheit fest. Jedes Netzwerk besteht aus Netzwerkmitgliedern, den sogenannten *Knoten* (auch: *Fälle*, *Elemente*, *Akteure*, *Nodes*) – das sind in diesem Beispiel die Seminarteilnehmer. Die

Knoten könnten neben Personen auch Organisationen, Unternehmen, Gruppen, Städte usw. sein. Um Informationen über das Netzwerk zu erfassen, muss das Netzwerk klar (zeitlich, räumlich und sachlich) abgegrenzt sein, d.h. man muss genau wissen, wer z.B. zu dem Seminar gehört und wer nicht. Nehmen wir an, es handelt sich um ein fortgeschrittenes Seminar im Master, das im WS 2013/2014 an der Universität X gehalten wurde und an dem 12 Studierende teilgenommen haben.

2. *Definition der Knoten (Akteure):* Alle Netzwerkmitglieder müssen namentlich bekannt sein, wozu man schon vor der Datenerhebung eine sog. *Namensliste* braucht. Diese Namensliste könnte z.B. die Teilnehmerliste des Seminars oder bei Unternehmen die Liste der Firmennamen sein.
3. *Definition der Kanten (Beziehungen):* Nun muss man festlegen, welche Informationen man über das Netzwerk haben will, d.h. welche *Beziehungen* (auch: *Kanten, Ties, edges, relationalen Variablen*) zwischen den Knoten man erfassen will. Typische Beispiele bei Personen sind Freundschaften, Arbeitsbeziehungen und gegenseitige Unterstützung; bei Unternehmen sind es z.B. Kapitalverflechtungen oder Konkurrenzbeziehungen. Nehmen wir für das erste Beispiel an, dass die Forscherin von jedem Seminarteilnehmer wissen will, welche der anderen Seminarteilnehmer er kennt und mit welchen er schon einmal zusammengearbeitet hat, z.B. um ein Referat vorzubereiten.
4. *Entwicklung der Erhebungsinstrumente und Vollerhebung der Knoten:* Weiß man, welche Informationen man erfassen will, muss man die Erhebungsinstrumente entwickeln. Nehmen wir an, dass sich die Forscherin entscheidet, die Studierenden standardisiert (Reinecke, Kapitel 44 in diesem Band) mit Hilfe einer Online-Befragung (Wagner/Hering, Kapitel 48 in diesem Band) zu befragen. Sie würde nun für jeden Akteur des Netzwerks im Rahmen der Datenerhebung ermitteln, welche Beziehungen er zu allen anderen Akteuren im Netzwerk hat. Dafür könnte sie jedem Befragten eine Liste mit den Namen aller Seminarteilnehmer vorlegen und ihn bitten anzukreuzen, wen er kennt und mit wem er schon einmal zusammengearbeitet hat. Um die Befragten nicht zu demotivieren, sollte ein Netzwerk (und die damit verbundene Namensliste) „nicht zu groß“ sein, wobei die maximale Größe des Netzwerkes von der Datenerhebungsform abhängt. Mit unseren 12 Seminarteilnehmern und zwei Netzwerkfragen sollten wir im Bereich des Zumutbaren sein.
5. *Auswertung und/oder grafische Darstellung:* Man kann nun die Beziehungen zwischen Netzwerkmitgliedern entweder mit Hilfe entsprechender multivariater Verfahren (Blasius/Baur, Kapitel 79 in diesem Band) statistisch auswerten oder grafisch in einem sogenannten Soziogramm darstellen. Diese Auswertung von Netzwerkdaten nannte Moreno „*Soziometrie*“, in der modernen Methodenforschung heißt sie „*Netzwerkanalyse*“. Abb. 74.1 zeigt ein solches Soziogramm, das die Informationen aus der Beispielerhebung zusammenstellt und grafisch darstellt, welche Seminarteilnehmer (Knoten) schon einmal miteinander gearbeitet haben (Kanten). Man sieht, dass die Studierenden 1 bis 9 ein loses Arbeitsnetzwerk bilden, in dem die Studierenden 1, 5 und 8 relativ zentrale Positionen einnehmen: Studierende 1 hat vier direkte Arbeitspartner (die aber untereinander nicht zusammenarbeiten) und ist über Studierenden 3 indirekt mit



**Abb. 74.1** Soziogramm eines studentischen Seminars (Vollerhebung des Gesamtnetzwerkes)

Studierendem 6 verbunden. Die Studierenden 5, 7 und 8 arbeiten zu dritt zusammen. Studierende 5 arbeitet darüber hinaus mit Studierender 1, Studierender 8 mit Studierendem 9 zusammen. Studierender 10 hat dagegen noch nie mit einem anderen Seminar teilnehmer zusammengearbeitet – vielleicht ist er ein Einzelgänger, oder er kommt aus einem anderen Fach und kennt niemand –, und Studierende 11 und 12 bilden eine Zweiergruppe.

Im Folgenden gebe ich einen kurzen Überblick über die *Besonderheiten von Netzwerkdaten und der Netzwerkanalyse*. Dabei ist zu berücksichtigen, dass sich die Netzwerkanalyse in den vergangenen 25 Jahren aus sehr unterschiedlichen Forschungsfeldern und -traditionen parallel entwickelt hat, weshalb es (noch) keine einheitliche Terminologie gibt. Weiterhin ist zu unterscheiden zwischen „Netzwerkanalyse“ als Theorie und „Netzwerkanalyse“ als Methode. Mit anderen Worten: Man kann netzwerktheoretische Fragen auch mit anderen Methoden als jenen der Netzwerkanalyse bearbeiten, und man kann mit den Methoden der Netzwerkanalyse auch andere theoretische Fragen als netzwerktheoretische Fragen beantworten.

In der *soziologischen Theorie* findet man eine relationale Perspektive, die die Beziehungen zwischen Akteuren in den Analysefokus nimmt, schon sehr lange und bei so unterschiedlichen Theoretikern wie Georg Simmel und George C. Homans, und unterschiedlichen Theorientraditionen, wie die Gestalttheorie, die Feldtheorie und die Soziometrie.

Auch die *Methoden der Netzwerkanalyse* umfassen ein sehr breites Feld von Verfahren, das auch *qualitative* (Hollstein/Straus 2014) und *Mixed Methods-Ansätze* (Dominguez/Hollstein 2014) beinhaltet. In der Forschungspraxis dominieren momentan die quantitativen Ansätze der Netzwerkanalyse (Hannemann/Riddle 2005), die auch *strukturelle oder soziale Netzwerkanalyse (SNA)* genannt werden und die u.a. durch die Arbeiten von Jacob Levy Moreno (1954), Ronald S. Burt (1992), Mark Granovetter (1985) und Harrison C. White (2002) inspiriert wurden (zur Geschichte der SNA siehe: Jansen 2003: 38). Aus Platzgründen befasst sich die folgende Darstellung allein mit der strukturellen Netzwerkanalyse (SNA), und ich verwende im Folgenden die Begriffe „Netzwerkanalyse“ und „strukturelle Netzwerkanalyse“ synonym.

## 74.2 Datenstruktur

Netzwerddaten können grundsätzlich auf zwei verschiedene Arten abgebildet werden: Entweder als *Datenmatrix* (also als Datensatz) oder grafisch als *Soziogramm* (Hanneman/Riddle 2005: Kapitel 1-5). Dabei gibt es drei Varianten, die Beziehungen zwischen Knoten und Kanten abzubilden (Scott 2000: 45), wobei es vom spezifischen Forschungsineresse abhängt, welche Darstellungsform zu bevorzugen ist:

1. Die *Ereignismatrix* gleicht einem herkömmlichen Datensatz insofern, als dass die Knoten (also Fälle) in den Zeilen und die Kanten (also Beziehungen bzw. relationalen Variablen) in den Spalten abgebildet werden. So stellt etwa Abb. 74.2 dar, wie vier fiktive Unternehmen über Aufsichtsratsposten miteinander verbunden sind. Diese Arten von

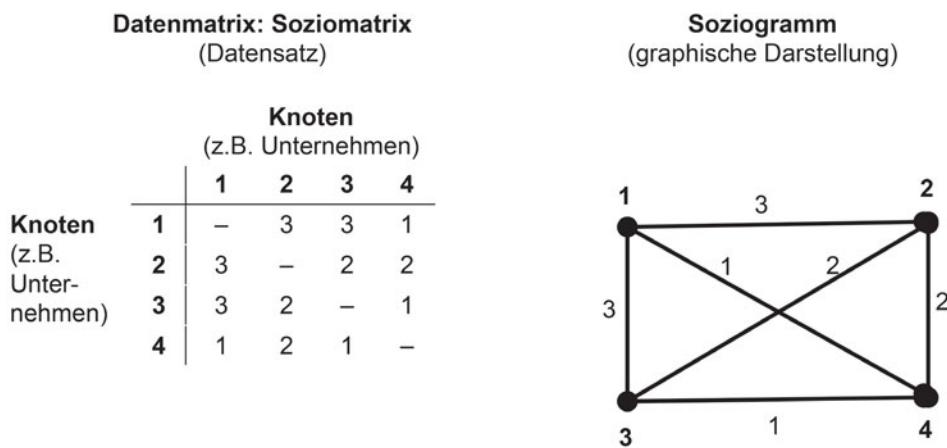
**Abb. 74.2** Aufbau des Datensatzes 1: Ereignismatrix (nach Scott 2000: 45)

**Datenmatrix: Ereignismatrix  
(Datensatz)**

		Kanten (z.B. Aufsichtsratsposten)				
		A	B	C	D	E
Knoten (z.B. Unter- nehmen)	1	1	1	1	1	0
	2	1	1	1	0	1
	3	0	1	1	1	0
	4	0	0	1	0	1

Beziehungen werden oft in der Organisations- und Wirtschaftssoziologie erhoben (z.B. Ziegler 1984). Liest man die Ereignismatrix *zeilenweise*, so kann man erkennen, dass in diesem Beispiel Unternehmen 1 vier Aufsichtsratsposten besetzt, nämlich in den Aufsichtsräten der Unternehmen A, B, C und D. Unternehmen 2 stellt ebenfalls vier Aufsichtsratsposten – in den Aufsichtsräten A, B, C und E usw. Liest man die Ereignismatrix *spaltenweise*, so erkennt man, dass sich im Aufsichtsrat A Vertreter der zwei Unternehmen 1 und 2 treffen, im Aufsichtsrat B treffen sich Vertreter der Unternehmen 1, 2 und 3, und in Aufsichtsrat C treffen sich sogar Vertreter aller vier Unternehmen. Für die Ereignismatrix gibt es keine übliche grafische Darstellung, was bei der Soziomatrix anders ist.

2. Als Datenmatrix ist die *Soziomatrix* symmetrisch aufgebaut und stellt die Zahl der Beziehungen zwischen den Knoten dar, d.h. sowohl die Zeilen, als auch die Spalten stellen die Knoten dar, in der Matrix ist die Zahl der Beziehungen abgebildet. Anhand Abb. 74.3 ist zu erkennen dass sich die Vertreter der Unternehmen 1 und 2 in drei Aufsichtsräten treffen (in den Aufsichtsräten A, B und C, vgl. Abb. 74.2). Die Vertreter der Unternehmen 1 und 4 treffen sich nur im Aufsichtsrat C. Stellt man diese Informationen statt als Datenmatrix grafisch als Soziogramm dar, dann wird jedes Unternehmen (jeder Knoten) als Punkt dargestellt. Wenn zwei Knoten eine Beziehung zueinander haben (hier also: eine Kontaktmöglichkeit über einen Aufsichtsrat haben), wird eine Linie zwischen den beiden Knoten gezogen, wobei die Stärke der Beziehung u.a. durch Farben, Liniendicke oder – wie in diesem Beispiel – durch eine Zahl dargestellt werden kann.
3. Alternativ kann man die Datenmatrix auch als *Affiliationsmatrix* aufbauen (Abb. 74.4). Auch diese ist symmetrisch, allerdings stehen hier in den Zeilen und Spalten die Beziehungen, in der Matrix ist die Zahl der Fälle abgebildet, die diese Beziehung gemeinsam haben – für die grafische Darstellung gelten dieselben Regeln wie bei der Soziomatrix.

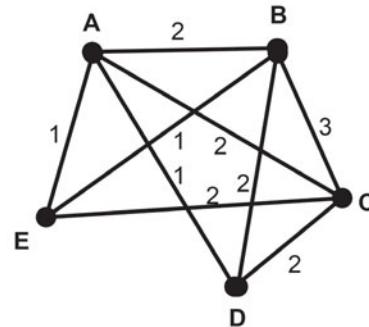


**Abb. 74.3** Aufbau des Datensatzes 2: Soziomatrix (nach Scott 2000: 45)

**Datenmatrix: Affiliationsmatrix**  
(Datensatz)

		Kanten (z.B. Aufsichtsratsposten)				
		A	B	C	D	E
Kanten (z.B. Aufsichtsrats- rats- posten)	A	-	2	2	1	1
	B	2	-	3	2	1
	C	2	3	-	2	2
	D	1	2	2	-	0
	E	1	1	2	0	-

**Soziogramm**  
(graphische Darstellung)



**Abb. 74.4** Aufbau des Datensatzes 3: Affiliationsmatrix (nach Scott 2000: 45)

So werden in die Aufsichtsräte A und B Vertreter von zwei derselben Unternehmen entsandt, nämlich von den Unternehmen 1 und 2 (siehe Abb. 74.2). Aufsichtsräte D und E haben dagegen keine solchen Beziehungen gemeinsam, da Aufsichtsrat D von Unternehmen 1 und 3, Aufsichtsrat E von Unternehmen 2 und 4 kontrolliert werden. Die Spalte im Datensatz bleibt leer, und im Soziogramm wird keine Verbindungsleitung gezeigt.

Es gibt zahlreiche andere Informationen, die in den Daten festgehalten werden können. Unter anderem können die Beziehungen etwa *nominal*, *ordinal* oder *metrisch skaliert* sein (Blasius/Baur, Kapitel 79 in diesem Band), und man kann neben der *Stärke* auch die *Richtung einer Beziehung* festhalten. So kann etwa ein Unternehmen das andere über den Aufsichtsrat einseitig kontrollieren (*asymmetrische* bzw. *gerichtete Beziehung*) oder die Unternehmen können Aufsichtsratsposten im jeweils anderen Unternehmen besetzen und sich so wechselseitig kontrollieren (*symmetrische* oder *ungerichtete Beziehung*).

Mit Hilfe eines geeigneten *Softwarepaketes* (z.B. Ucinet, [www.analytictech.com](http://www.analytictech.com)) können Datenmatrix und Soziogramme leicht ineinander überführt und die Netzwerkdaten können ausgewertet werden (Hanneman/Riddle 2005: Kapitel 6. Für einen vollständigen Überblick über verschiedene Softwarepaketes für Netzwerkdaten siehe Jansen 2003).

## 74.3 Stichprobenstrategien

### 74.3.1 Vollerhebung

Es existieren grundsätzlich drei Datenauswahlstrategien, um Netzwerkdaten zu erheben. Die erste Variante der Datenauswahl für Netzwerkdaten ist die Vollerhebung des gesamten Netzwerkes, die oben in Abschnitt 74.1 beschrieben wurde, d.h. hier befragt man *jeden* Netzwerkakteur zu seiner Beziehung zu *jedem* anderen Akteur. Voraussetzungen hierfür sind, dass die Netzwerkgrenzen sowie alle Netzwerkakteure bekannt sind, dass das Netzwerk „nicht zu groß“ ist und dass eine Namensliste existiert, die man bei der Datenerhebung den Netzwerkakteuren vorlegen kann.

Hat man die Wahl, ist die Vollerhebung des Netzwerks immer zu bevorzugen, weil nur sie potenziell die Möglichkeit bietet *alle* Auswertungsmöglichkeiten (siehe Abschnitt 74.5) zu nutzen, die die Netzwerkanalyse potenziell bietet. Allerdings ist sie in der Praxis oft nur schwer oder sogar gar nicht umzusetzen, denn oft stellt bereits die *Identifikation aller Elemente der Grundgesamtheit* vor der Datenerhebung ein Problem dar. So kennt der Dozent bei einem Seminar mit Anwesenheitspflicht zwar vermutlich alle Teilnehmer, aber bei einer Vorlesung mit über 100 Teilnehmern ist das anders. Selbst wenn der Dozent hier ungefähr weiß, wie viele Teilnehmer er hat, kennt er nicht unbedingt deren Namen und hat auch nicht von allen Kontaktdataen.

Weiterhin ist das Netzwerk oft schlicht *zu groß*, um (zumindest in Befragungen) den Netzwerkmitgliedern zuzumuten, ihre Beziehung zu wirklich *jedem* anderen Netzwerkmitglied zu erläutern. So ist es eine Sache, Seminarteilnehmer zu ihren Beziehungen zu 12 Kommilitonen zu befragen – was aber ist, wenn es sich um eine Vorlesung mit 100, 200 oder noch mehr Teilnehmern handelt? Würde man dies versuchen, würde man vermutlich zahlreiche fehlende Werte oder gar Abbrüche der Befragung provozieren. Gerade *Ausfälle* sind aber bei Vollerhebungen von Netzwerkdaten ein noch stärkeres Problem als sonst (Engel/Schmidt, Kapitel 23 in diesem Band), weil sie die Vorteile der Vollerhebung zunichte machen: Die netzwerkanalytischen Verfahren zur Analyse von Gesamtnetzwerken und Teilgruppen (Abschnitt 74.5.2 und 74.5.3) setzen voraus, dass vollständige Informationen über *alle* Beziehungen zwischen *allen* Netzwerkpartnern vorliegen – erhebt man nur eine Stichprobe des Netzwerkes oder fehlen Werte, fehlt eben auch ein Teil der für viele Maße erforderlichen Informationen, so dass diese Maße nicht oder nur verzerrt berechnet werden können.

### 74.3.2 Zufallsstichprobe

Handelt es sich um ein sehr großes Netzwerk mit sehr vielen potenziellen Beziehungen zwischen den Netzwerkpartnern, bietet es sich an, statt einer Vollerhebung eine Zufallsstichprobe (Häder/Häder, Kapitel 18 in diesem Band) zu ziehen. Nehmen wir z.B. an, eine Forscherin hat zufällig 1.000 Studierende der Universität X ausgewählt, die sie nun zu

ihren Beziehungen zu ihren Kommilitonen befragen möchte. Daher erhebt sie nun von diesen ausgewählten Personen die sogenannten *egozentrierten Netzwerke*. Hierzu existieren verschiedene Varianten (siehe ausführlich: Wolf 2006). Weil die Forscherin – anders als bei der Vollerhebung – die Namen der Netzwerkmitglieder nicht kennt, haben diese Varianten gemeinsam, dass man in einem ersten Schritt einen *Namensgenerator* benötigt:

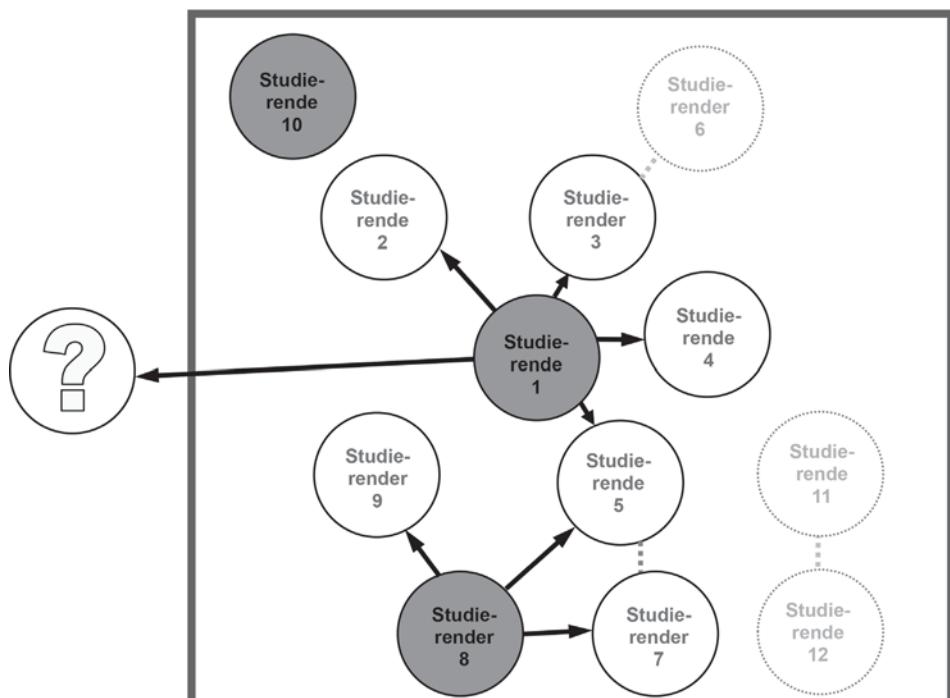
Beim sogenannten „*Burt-Generator*“ (Burt 1984) stellt man hierzu genau eine Frage z.B.: „Hin und wieder besprechen die meisten Leute wichtige Angelegenheiten mit anderen. Wenn Sie an die letzten sechs Monate zurückdenken: Mit wem [aus dem Seminar] haben Sie über Dinge gesprochen, die Ihnen wichtig waren?“ Durch den *Burt-Generator* erfasst man die Netzwerkpartner, zu denen eine starke emotionale Beziehung besteht. Wenn man stärker auf Interaktionen abheben möchte, könnte man (abweichend von Burt) die Frage auch abwandeln und z.B. die Studierenden bitten, die (z.B. 5) wichtigsten Kommilitonen zu nennen, mit denen sie im letzten Semester zusammengearbeitet haben.

Der *Burt-Generator* hat den Vorteil, dass er kurz ist und man ihn deshalb relativ gut in Umfragen integrieren kann, bei denen die Netzwerkbeziehungen einer Person ein Nebenaspekt der Studie sind und bei denen man auch Fragen zu anderen Themen stellt. Allerdings hat er den Nachteil, dass nur nach einer einzigen (meistens emotionalen) Beziehung gefragt wird. Dadurch werden nur sehr wenige Netzwerkpartner generiert, und es wird auch nur eine einzige Beziehungsform zu diesen ermittelt. Die meisten Menschen haben aber wesentlich größere Beziehungsnetzwerke und auch die verschiedensten Arten von Beziehungen zu unterschiedlichen Personen. Liegen diese Beziehungen in ihrer Komplexität im Fokus des Forschungsinteresses, bietet sich der sogenannte „*Fischer-Generator*“ (McCallister/Fischer 1978) an. Dieser stellt Fragen zu Interaktionen in den verschiedensten Kontexten, u.a.: Mit wem hat man schon zusammen gearbeitet? Mit wem spricht man über Privates? Mit wem hat man schon einmal etwas in der Freizeit unternommen? Da die Liste der abgefragten Interaktionen bewusst sehr lang ist, um möglichst alle relevanten Interaktionspartner der Befragten zu erfassen, werden sehr viel mehr Netzwerkpartner und -beziehungen ermittelt. Der Nachteil ist, dass die Erhebung der Netzwerkinformationen relativ lange dauert, so dass man nur wenige andere Fragen stellen kann.

Egal ob man den *Burt-* oder den *Fischer-Generator* verwendet, als nächstes stellt sich das Problem, dass man – da man nicht alle Netzwerkteilnehmer befragt – nun zwar weiß, dass der Befragte eine soziale Beziehung zu den genannten Personen hat, aber sonst weiß man nichts über die andere Person (außer ihre Namen bzw. Pseudonyme). Da aber i.d.R. nicht nur interessant ist, zu wie vielen Personen ein Befragter eine Beziehung hat, sondern auch welche Eigenschaften diese Personen haben, fragt man Basisinformationen zu jeder im Namensgenerator genannten Person ab. So könnte man die Studierenden z.B. fragen, welches Geschlecht die Arbeitspartner haben, welches Fach sie studieren und woher man sie kennt. Derartige *Namensinterpretatoren* sind für die Befragten potenziell sehr ermüdend, weil sie zu verschiedenen Personen immer wieder dieselben Fragen beantworten müssen.

Egozentrierte Netzwerke können auch bei großen Gesamtheiten ermittelt werden, aber der Preis der Zufallsstichprobe ist, dass nicht alle Netzwerkbeziehungen vollständig ermittelt werden können: Nehmen wir an, in der o.g. Beispieldumfrage wurden zufällig

auch die Studierenden 1, 8 und 10 aus dem studentischen Seminar ausgewählt, das in Abschnitt 74.1 als Beispiel eingeführt wurde und in dem eine Vollerhebung durchgeführt worden war, weshalb wir sämtliche Netzwerkbeziehungen kennen (Abb. 74.1). Erfasst man die egozentrierten Netzwerke der ausgewählten Studierenden 1, 8 und 10, so ergibt sich das Soziogramm in Abb. 74.5. Vergleicht man Abb. 74.5 mit Abb. 74.1, so wird deutlich, dass man erstens nur einseitige (gerichtete) Beziehungen ermitteln kann, d.h. man weiß zwar, dass z.B. Studierende 1 Studierenden 2 als wichtigen Beziehungspartner nennt – man weiß aber nicht ob Studierender 2 das auch so sieht, dass Studierende 1 eine wichtige Person ist. Es gehen zweitens drei Beziehungen (gestrichelte Linien) komplett verloren, weil es sich um Beziehungen zwischen nicht befragten Personen handelt. Drittens hat man über drei Personen aus dem Seminar (gestrichelte Kreise) gar keine Informationen, weil keine der befragten Personen mit ihnen befreundet ist. Da man den Befragten keine Liste aller Personen des Netzwerkes vorlegt, sondern offen nach den Beziehungen fragt, können viertens auch Beziehungen zu Personen genannt werden, die gar nicht Teil der Gesamtheit sind – etwa weil diese Studierenden gar nicht Teil des Seminars an der Universität X sind, sondern an einer anderen Universität studieren. Insgesamt bedeutet dies für die Auswertung (Abschnitt 74.5), dass man zwar Aussagen über die Ego-Netzwerke der befragten Personen, nicht aber über das Netzwerk als Ganzes machen kann.



**Abb. 74.5** Soziogramm eines studentischen Seminars (Zufallsstichprobe und Erhebung von egozentrierten Netzwerken)

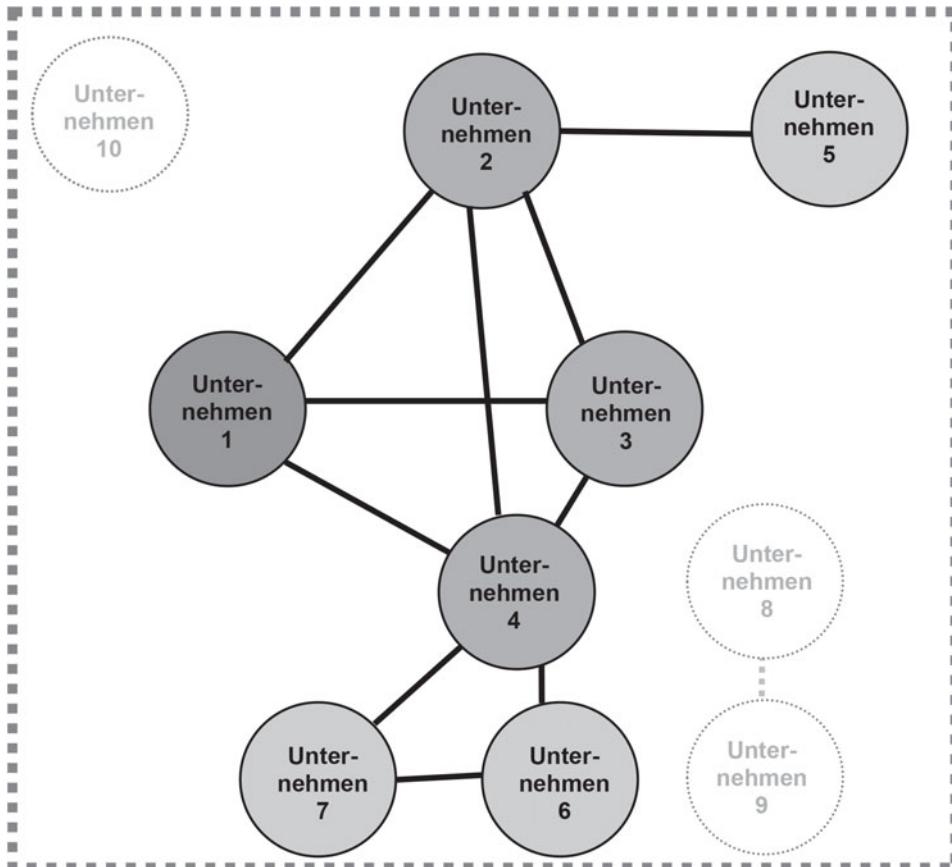
So verschieden Vollerhebung und Zufallsstichprobe sind, so haben sie doch gemeinsam, dass sie voraussetzen, dass die *Größe der Grundgesamtheit* bekannt ist. Für die Vollerhebung müssen zusätzlich die Namen *aller* potenziellen Netzwerkpartner, für die Zufallsstichprobe zumindest die Namen der zu befragenden Personen schon vor der Datenerhebung bekannt bzw. ermittelbar sein. Bei den obigen Beispielen von Studierenden einer Universität ist dies zwar relativ einfach und eindeutig. Anders ist das aber, wenn man etwa die Konkurrenz von Unternehmen auf Märkten (Baur 2013) ermitteln will, weil hier oft unklar ist, welche Unternehmen überhaupt miteinander konkurrieren, welche Produkte zueinander in Konkurrenz stehen und wie der Markt räumlich abgegrenzt wird, ob die Konkurrenz also auf lokaler, nationaler oder globaler Ebene stattfindet.

Die klare *Abgrenzung des Netzwerkes* ist insofern zentral, als dass viele statistische Verfahren zur Analyse der Makroebene (Abschnitt 74.5) die Gesamtgröße des Netzwerks für die Berechnung von Maßzahlen erfordern – kennt man diese nicht, kann man diese Verfahren nicht anwenden. In der Forschungspraxis werden daher die Netzwerkgrenzen oft pragmatisch gezogen, d.h. der Forscher wählt z.B. alle deutschen Dax-Unternehmen oder alle Unternehmen mit einem Firmensitz in Deutschland und einem Jahresumsatz von mehr als 1 Mio. € aus und definiert dies als sein Netzwerk. Dies löst aber das Problem nicht, sondern verschärft es sogar eher noch, weil man damit das Netzwerk umdefiniert und unterstellt, dass das, was man in den Daten hat, auch die sozial relevante Gruppe ist: Ein Stahlhersteller, ein IT-Unternehmen und ein Spielzeughersteller stehen nicht umgekehrt zueinander in Konkurrenz, nur weil sie mehr als 1 Mio. € Umsatz machen. Umgekehrt ist es durchaus wahrscheinlich, dass einige Konkurrenten dieser Unternehmen nicht im Dax gelistet sind, weniger als 1 Mio. € Jahresumsatz machen oder ihren Firmensitz zwar nicht in Deutschland haben, aber Deutschland dennoch ein Absatzmarkt ist.

### 74.3.3 Schneeballverfahren

Wenn man weder weiß, wie groß das Netzwerk ist, noch wo die Netzwerkgrenzen liegen, ist das Schneeballverfahren eine weitere Alternative. Wie bei den egozentrierten Netzwerken zieht man hier eine Stichprobe aus allen Knoten. Diese ausgewählten Knoten nennt man auch *Gatekeeper*. Nehmen wir etwa an, die Forscherin will die Konkurrenzbeziehungen z.B. auf dem Spielemarkt des Landes X erheben. Dann könnte sie einen beliebigen Spielhersteller bitten, an der Studie teilzunehmen, dies wäre das Unternehmen Nr. 1. Anders als bei den egozentrierten Netzwerken befragt man nun aber nicht nur den Gatekeeper, vielmehr iteriert man Datenerhebung und Datenauswahl:

Man beginnt bei Knoten 1 (Gatekeeper) und ermittelt dessen Beziehungen. So ergibt sich etwa bei dem Spielemarkt, dass Unternehmen 1 ein Hersteller von Brettspielen für Kinder ist und mit drei weiteren Unternehmen konkurriert, nämlich Unternehmen 2, 3 und 4.



**Abb. 74.6** Soziogramm der Konkurrenten eines Marktes (Schneeballverfahren und Erhebung von egozentrierten Netzwerken)

Im zweiten Schritt befragt man alle von Knoten 1 genannten Knoten – also im Beispiel die drei Konkurrenten von Unternehmen 1. Dabei ergibt sich, dass alle drei auch untereinander konkurrieren. Unternehmen 4 stellt auch noch Brettspiele für Erwachsene her und konkurriert daher noch mit den Unternehmen 6 und 7 (die nur Erwachsenenspiele produzieren). Unternehmen 2 produziert neben Brett- auch Kartenspiele für Kinder – und steht diesbezüglich in Wettbewerb mit Unternehmen 5.

Auch diese Unternehmen befragt man nun – und stellt fest, dass Unternehmen 6 und 7 sich auch wechselseitig als Konkurrenten wahrnehmen.

Gegenüber den egozentrierten Netzwerken hat das Schneeballverfahren den Vorteil, dass man alle Beziehungen innerhalb der Teilgruppe erfasst (dunkle Kreise in Abb. 74.6). Allerdings ist es mit diesem Vorgehen nicht möglich, die Größe des Gesamtnetzwerkes zu bestimmen. Wie Abb. 74.6 zu entnehmen ist, fehlen nun v.a. die isolierten Akteure bzw. Informationen über Teilgruppen, die man nicht befragt hat. Im hypothetischen

Beispielmarkt gibt es etwa noch zwei Unternehmen (Nr. 8 und 9), die Lernspiele für Schulklassen herstellen, die nicht über die Eltern, sondern über die Lehrer vertrieben werden. Weiterhin dringt gerade ein ausländisches Unternehmen (Nr. 10) in den Markt ein, das iPhone-Apps für kleine Kinder entwickelt und international vertreibt. Alle drei Firmen befinden sich (noch nicht) im Wahrnehmungshorizont der befragten Unternehmen, weshalb sie auch nicht als Konkurrenten genannt werden.

---

## 74.4 Datenerhebung

### 74.4.1 Standardisierte Befragung

Es gibt verschiedene Strategien, Netzwerkdaten zu erheben. Die obigen Beispiele sind etwa davon ausgegangen, dass die Netzwerkdaten mittels standardisierter Befragungen (Reinecke, Kapitel 44 in diesem Band) erhoben wurden. Zusätzlich zu den bei „normalen“ Bevölkerungsumfragen auftretenden Problemen weist aber die Erhebung von relationalen Daten in Umfragen eine Reihe zusätzlicher Probleme auf. Zu nennen ist zunächst die *Datenschutzproblematik* (Mühlichen, Kapitel 4 in diesem Band): Zwar werden (bei datenschutzrechtlich sachgemäß durchgeführten Untersuchungen) Befragte um ihr Einverständnis für die Befragung gebeten, aber sie werden im Falle von Netzwerkfragen ja auch zu anderen Personen befragt, die nie gefragt wurden, ob sie mit der Befragung einverstanden sind. Zweitens ist die Datenerhebung – wie bereits erwähnt – für die Befragten sehr *zeitaufwändig*. Deshalb ist es drittens erforderlich, die Zahl der Beziehungsdimensionen, die man erfragt, auf ein Minimum zu beschränken, was zu einer gewissen *Informationsdünne* führen kann. Bei Gesamtnetzwerken darf deshalb, viertens, das *Netzwerk nicht zu groß* sein (Faustregel: max. 30 Personen), was in der Praxis oft dazu führt, dass die Netzwerkgrenzen willkürlich gesetzt werden und dadurch nicht alle Kanten miteinander vergleichbar sind. Besonders bei egozentrierten Netzwerken sollte, fünftens, der *Fragebogen nicht zu lang sein*, d.h. je mehr Namen man abfragt, desto weniger andere Fragen kann man stellen. Grundsätzlich sollte man sich immer fragen: Würde ich das selbst noch beantworten? Auch ist das Zusammenspielen dieser egozentrierten Netzwerkinformation oft schwer, weil die *Beziehungen i.d.R. nur einseitig erhoben wurden*.

### 74.4.2 Alternative Datenquellen

Eine Alternative ist, die Netzwerkinformationen mit Hilfe anderer Datenquellen zu erfassen. So werden in der Stadtsoziologie, Gruppensoziologie und Ethnologie Netzwerkdaten oft aus *Beobachtungen* (Thierbach/Petschick, Kapitel 66 in diesem Band) gewonnen.

Vor allem in der Organisations- und Wirtschaftssoziologie werden oft *Dokumente und andere prozessproduzierte Daten* (Salheiser, Kapitel 62 in diesem Band) inhaltsanalytisch kodiert (Mayring/Fenzl, Kapitel 38 in diesem Band). So untersucht Ziegler (1984)

wechselseitige Kapitalbeteiligung und Aufsichtsratsverflechtungen zwischen Unternehmen. Er kann dabei zeigen, dass 1976 die deutschen Großunternehmen nicht nur sehr stark miteinander verflochten waren, sondern dass vier große Finanzunternehmen (Allianz, Commerzbank, Deutsche Bank und Dresdner Bank) dieses System dominierten. Als Datenbasis für seine Analyse hat er u.a. Geschäftsberichte, Zeitungsartikel und Nachschlagewerke verwendet. Wie groß das Netzwerk ist, das man dabei analysieren kann, hängt davon ab, welchen Aufwand der Forscher bereit ist, bei der Datenaufbereitung zu betreiben. Durch die zunehmende Bedeutung des Internets, insbesondere von Online-Communities und sozialen Netzwerken im Internet (wie z.B. Facebook), werden auch *Web Server Logs und Logfiles* (Schmitz/Yanenko, Kapitel 65 in diesem Band) eine immer wichtige Datenquelle für Netzwerkdaten.

Auch wenn es – wie die Studie von Ziegler (1984) zeigt – gelungene Beispiele für Netzwerkanalysen mit prozessproduzierten Daten gibt – stellt sich hier oft das Problem, dass für viele sozialwissenschaftlich relevante Fragestellungen keine geeigneten relationalen Daten zur Verfügung stehen und die vorhandenen Daten oft nicht die tatsächlichen Beziehungen widerspiegeln. Wenn man etwa in Facebook mit jemand als „Freund“ verlinkt ist, heißt das noch lange nicht, dass es sich auch im realen Leben um einen Freund handelt. In der Wissenschaftsforschung werden oft Ko-Autorenschaften analysiert, aber diese sagen nur wenig über reale Arbeitsbeziehungen aus. In wirtschafts- und organisationssoziologischen Studien werden oft – wie in der Studie von Ziegler (1984) – Aufsichtsratsbeziehungen als Zeichen für Kapitalverflechtung und wechselseitige Einflussnahme gedeutet, die potenziell Absprachen ermöglicht und so den Konkurrenzmechanismus ausschaltet – was aber noch lange nicht heißt, dass solche Preisabsprachen auch tatsächlich stattfinden. Insgesamt sind also die Informationen, die man über prozessproduzierte Daten zu Netzwerken bekommt, oft sehr dünn. Hinzu kommen dieselben Probleme der Abgrenzung der Grundgesamtheit wie bei standardisierten Befragungen.

Gelingt es, valide Daten zu erheben, sind relationale Daten für viele soziologische Fragestellungen hoch relevant, und es bieten sich zahlreiche Analysemöglichkeiten, die oft weit über die von absoluten Daten hinausgehen.

---

## 74.5 Auswertung

Liegen die Netzwerkdaten vor, kann man sie mit Hilfe einer „Sozialen Netzwerkanalyse“ bzw. „Strukturellen Netzwerkanalyse“ (SNA) statistisch auswerten. Hierbei handelt es sich nicht um ein einzelnes Auswertungsverfahren, sondern um einen Sammelbegriff für alle Verfahren, die zur Auswertung von standardisierten Netzwerkdaten geeignet sind (Hanneman/Riddle 2005, Trappmann et al. 2005, Scott 2000, Faust/Wassermann 2014). Die SNA beinhaltet sowohl Möglichkeiten der grafischen Abbildung und Analyse von Netzwerken in Soziogrammen, als auch der statistischen Analyse. Zu unterscheiden ist dabei zwischen der *Erhebungseinheit* („Unit of Observation“) – diese besteht, wie oben dargestellt, i.d.R. aus den einzelnen Knoten des Netzwerkes (z.B. Akteure) – und der *Analyseeinheit* („Unit

*of Analysis*“), d.h. der Einheit, über die im Rahmen der Auswertung Aussagen gemacht werden. Die Analyseeinheit kann erstens aus den einzelnen Elementen des Netzwerkes bestehen, sich also mit der Erhebungseinheit decken. Je nach Forschungsinteresse kann die Analyseeinheit aber auch zweitens aus dem Gesamtnetzwerk und drittens aus Teilgruppen des Netzwerkes bestehen.

#### **74.5.1 Mikroperspektive: Ego-Netzwerke**

In der Mikroperspektive kann die Analyse Fragen an die persönlichen Netzwerke der Elemente des Netzwerkes stellen (Hanneman/Riddle 2005: Kapitel 9). Typische Fragen an Ego-Netzwerke sind: Wie viele Beziehungspartner hat ein Knoten? Wie stabil sind diese Beziehungen? Haben die Beziehungspartner ähnliche Rollen? Gibt es wichtige und unwichtige Beziehungspartner? Kennen sich die Beziehungspartner untereinander? Welche Gemeinsamkeiten und Unterschiede gibt es zwischen den Beziehungspartnern? So hat im Beispiel in Abschnitt 74.1 (Abb. 74.1) Studierende Nr. 1 vier und Studierende Nr. 10 keinen einzigen Beziehungspartner. Die restlichen Studierenden lassen sich in drei Gruppen einteilen: Sechs Studierende (Nr. 2, 4, 6, 9, 11 und 12) haben einen Beziehungspartner und jeweils zwei Studierende haben zwei Beziehungspartner (Nr. 3 und 7) bzw. drei Beziehungspartner (Nr. 5 und 8).

Viele der Fragen zu Ego-Netzwerken kann man mit uni- und bivariaten Analysen beantworten, und man braucht auch oft keine Spezialsoftware, sondern es genügt Standardsoftware wie SPSS oder Stata. Oft werden auch Informationen über das Netzwerk in multivariaten Verfahren (Blasius/Baur, Kapitel 79 in diesem Band) weiter verarbeitet, um etwa zu analysieren, wie sich Netzwerke über die Zeit entwickeln und welche Ursachen und Folgen diese Netzwerkbeziehungen haben. So arbeiten Lang und Neyer (2004) in dem oben erwähnten Beispiel zu Karrierenetzwerken von Wissenschaftlern mit bivariaten Analysen und einer logistischen Regressionsanalyse, um zu zeigen, dass Forscher, die mit anderen zusammenarbeiten, deutlich höhere Chancen haben, auf eine Professur berufen zu werden.

#### **74.5.2 Analyse des Gesamtnetzwerkes**

Während die Verfahren zur Analyse von Ego-Netzwerken bei allen Stichprobenformen – also Vollerhebungen (Abschnitt 74.3.1), Zufallsstichproben (Abschnitt 74.3.2) und Schneeballverfahren (Abschnitt 74.3.3) – funktionieren, erlauben Vollerhebungen zusätzlich, die Makroperspektive einzunehmen, wobei Daten über mindestens zwei Netzwerke miteinander verglichen werden sollten. So könnte man das in Abschnitt 74.1 (Abb. 74.1) ermittelte Arbeitsnetzwerk in einem studentischen Seminar mit dem Arbeitsnetzwerk eines anderen Seminars vergleichen und untersuchen, in welchem Seminar die Studierenden insgesamt enger untereinander zusammenarbeiten. Ebenso kann man Konkurrenz- und Kooperationsnetzwerke auf Märkten vergleichen. So vergleicht etwa Ziegler (1984) das in Abschnitt

74.4 erläuterte deutsche Wirtschaftssystem mit dem österreichischen Wirtschaftssystem und kann dadurch zeigen, dass in Deutschland Großbanken und Versicherer wesentlich mehr Einfluss auf die Gesamtwirtschaft haben als in Österreich.

Erst bei der Analyse von Gesamtnetzwerken kann die SNA ihre volle Leistungsfähigkeit entfalten, da nun eigene Maße für das Gesamtnettwerk berechnet werden können, was gewissermaßen eine Sonderform der Aggregatdatenanalyse darstellt. Diese Maße können weiterhin mit der Analyse von Ego-Netzwerken verbunden werden, d.h. es können Mikro-Makro-Beziehungen analysiert werden. Grundsätzlich unterscheidet die SNA zwei Perspektiven der Charakterisierung von Gesamtnettwerken:

1. Maße der *Kohäsion, des Zusammenhalts bzw. der Verbundenheit von Netzwerken* (Hanneman/Riddle 2005: Kapitel 7 & 8; Scott 2000: 63-99) sollen die Möglichkeiten des Netzwerkes zur Kooperation und zum kollektiven Handeln kennzeichnen. Ein häufig verwendetes Maß ist die *Dichte* („Density“). Diese gibt an, welcher Anteil aller theoretisch möglichen Zweierbeziehungen empirisch auch realisiert wurde. Theoretisch könnte nämlich jeder Knoten mit jedem anderen verbunden sein, empirisch ist dies aber fast nie der Fall. Vielmehr nimmt i.d.R. mit zunehmender Netzwerkgröße die Dichte ab. Ziegler (1984: 593) zeigt etwa für das deutsche Wirtschaftssystem, dass es 1976 einen Kern von zentralen Unternehmen gab, die 74% aller möglichen Beziehungen untereinander auch realisierten, während andere Unternehmen gar keine Beziehung zu anderen Unternehmen hatten.
2. Alternativ können Gesamtnettwerke über die *Zentralität und das Prestige von Akteuren im Netzwerk* (Hanneman/Riddle 2005: Kapitel 10; Trappmann et al. 2005: 25-70) beschrieben werden: Man geht davon aus, dass Personen, die über viele Kontakte im Netzwerk verfügen, Zugang zu Netzwerkressourcen, Kontrollmöglichkeiten und Informationen (soziales Kapital) haben. Je mehr das Netzwerk an einem einzelnen Akteur ausgerichtet ist, desto mehr Möglichkeit hat dieser zur Einflussausübung auf die übrigen Netzwerkpartner. Gibt es mehrere gleichermaßen wichtige Akteure, sind Machtkämpfe und Konkurrenz dagegen wahrscheinlicher. So zeigt Ziegler (1984: 607), dass in verschiedenen Branchen des deutschen Wirtschaftssystems die Machtressourcen unterschiedlich verteilt waren: Bei den Bergbau- und Energieunternehmen war die Netzwerkdichte relativ hoch, d.h. relativ viele der möglichen Beziehungen untereinander waren auch realisiert und nur jedes zehnte Unternehmen war mit keinem einzigen anderen Unternehmen verflochten („Isolierte“). Dagegen war die Netzwerkdichte in der Verbrauchsgüter- und Nahrungsmittelindustrie sehr gering und mehr als ein Drittel der Unternehmen dieser Branche hatte gar keine Netzwerkbeziehungen. Dabei existieren verschiedene Vorschläge, wie man Zentralität und Zentralisierung des Netzwerks messen könnte. So soll etwa der „Degree“ erfassen, wer auf wen einen direkten Einfluss ausüben kann, die „Closeness“ soll bezeichnen, wie lang die Informationswege im Netzwerk sind, die „Betweenness“ dient zur Identifikation von Gatekeepers, die Informationen kontrollieren können. Der „Eigenvector“ („Popularity“) erfasst Akteure, die besonders gut vernetzt bzw. besonders beliebt sind (Hannemann/Riddle 2005: Kapitel 10).

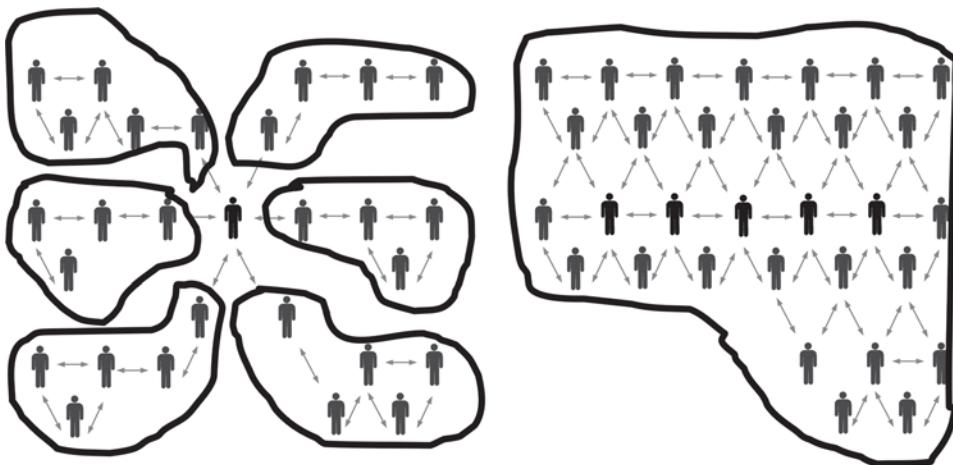
### 74.5.3 Mesoebene: Analyse von Teilgruppen im Netzwerk

Schließlich bietet die SNA auch die Möglichkeit, einzelne Teilgruppen im Netzwerk zu analysieren, wobei hier drei grundsätzliche Verfahrensgruppen zu unterscheiden sind:

Zunächst ist die Analyse von *Dyaden* (Zweierbeziehungen) und *Triaden* (Dreierbeziehungen) ein eigener Zweig der Netzwerkanalyse (Pappi 1987: 129-176, Trappmann et al. 2005: 173-208).

Ein zweiter Analysefokus liegt auf *der relationalen Perspektive zwischen Netzwerkpartnern* (Hanneman/Riddle 2005: Kapitel 11; Trappmann et al. 2005: 71-96; Scott 2000: 100-122): Es sollen Subgruppen des Netzwerkes identifiziert werden, die untereinander besonders stark miteinander verbunden sind und über relativ wenige Beziehungen nach außen verfügen. Man geht davon aus, dass innerhalb einer solchen Subgruppe die Binnenkommunikation erhöht ist und dadurch mehr Kooperationschancen bestehen. So zerfällt etwa das linke Netzwerk in Abb. 74.7 in sechs Subgruppen, wenn der zentrale Akteur entfällt. Da im rechten Netzwerk die Netzwerkpartner stärker untereinander vernetzt sind, gibt es keinen solchen zentralen Gatekeeper. Daher lässt sich das Netzwerk auch nicht in Teilgruppen unterteilen. Es gibt dabei zwei Möglichkeiten, wie man solche Subgruppen identifizieren kann (Hannemann/Riddle 2005: Kapitel 11):

1. Verwendet man sog. *Bottom-up-Ansätze*, nennt man die identifizierten Subgruppen *Cliquen*. Bei der Gruppenbildung richtet man den Blick auf die Verbundenheit zwischen den Akteuren, d.h. wer mit wem in Verbindung steht, wobei sich die verschiedenen Ansätze der Cliquenbildung u.a. darin unterscheiden, ob nur direkte Beziehungen (*Cliquen*) oder auch indirekte Beziehungen zwischen Akteuren berücksichtigt werden



**Abb. 74.7** Relationale vs. positionale Perspektive: Cliquen bzw. Komponenten vs. Blöcke

und über wie viele Akteure diese Beziehung hergestellt werden darf (*n-Cliquen*); ob diese indirekten Beziehungen auch über Akteure außerhalb der Subgruppe oder ausschließlich über die Mitglieder der Subgruppe hergestellt werden dürfen (*n-Clans*); ob Ko-Mitgliedschaften in verschiedenen Subgruppen erlaubt sind (*k-Plexe*); und wie viele andere Mitglieder einer Subgruppe der Akteur mindestens erreichen muss, um noch als Teil der Gruppe zu gelten (*k-Cores*).

2. Im Gegensatz dazu spricht man bei Subgruppen, die mit Hilfe von *Top-Down-Ansätze* gebildet wurden, von *Komponenten*. Hier liegt der Fokus auf Schwachstellen im Netzwerk: Wo würde das Netzwerk zerfallen, wenn ein Akteur („*Cutpoint*“) oder eine Brücke („*Bridge*“) wegfielen?

Während bei der Analyse von Cliques und Komponenten der Analysefokus auf den Beziehungen zwischen Netzwerkpartnern (relationale Perspektive) liegt, kann man auch eine *positionale Perspektive* einnehmen, d.h. man kann fragen, welche Netzwerkpartner ähnliche Rollen, Funktionen bzw. Positionen für das Netz einnehmen. Solche ähnlichen Netzwerkmitglieder nennt man *Blöcke*. Anders als bei Cliques und Komponenten richtet sich der Blick nicht auf die Kommunikation und Kooperation, sondern auf die Konkurrenz zwischen Akteuren: Wenn zwei Netzwerkmitglieder äquivalente Positionen innerhalb des Netzwerks einnehmen, sind sie prinzipiell ersetzbar. So haben die sechs schwarzen Akteure in Abb. 74.7 ähnliche Rollen, weil sie alle gleich viele (nämlich 6) Beziehungen zu anderen Netzwerkpartnern haben, allerdings gibt es im rechten Netzwerk fünf Personen mit derselben Rolle, während der Akteur im linken Netzwerk der einzige ist, der diese Rolle einnimmt – weshalb er weniger ersetzbar ist als sein Pendant im rechten Netzwerk. Mit Verfahren der *Blockmodellanalyse* (Hanneman/Riddle 2005: Kapitel 12-15; Trappmann et al. 2005: 97-172; Scott 2000: 123-145), z.B. der Clusteranalyse (Blasius/Baur, Kapitel 79 in diesem Band), fasst man solche Akteure mit ähnlichen Positionen zu Blöcken zusammen.

## Literatur

- Baur, Nina (2013): Was ist eigentlich ein Markt? In: SozBlog. URL: <http://soziologie.de/blog/?p=997>
- Burt, Ronald S. (1984): Network Items and the General Social Survey. In: Social Networks 6 (4): 293-339
- Burt, Ronald S. (1992): Structural Holes. Cambridge: Harvard University Press
- Dominguez, Silvia/Hollstein, Betina (Hg.) (2014): Mixed-Methods Social Network Research. New York: Cambridge University Press
- Faust, Katherine/Wasserman, Stanley (2014): Social Network Analysis. New York/Cambridge: Cambridge University Press
- Granovetter, Mark (1985): Economic Action and Social Structure. In: American Journal of Sociology 91 (3): 481-510
- Hanneman, Robert A./Riddle, Marc (2005); Introduction to Social Network Methods. Riverside: University of California. URL: <http://faculty.ucr.edu/~hanneman/>
- Hollstein, Betina/Straus, Florian (Hg.) (2014): Qualitative Netzwerkanalyse. Wiesbaden: VS Verlag
- Jansen, Dorothea (2003): Einführung in die Netzwerkanalyse. Wiesbaden: VS Verlag
- Lang, Frieder R./Neyer, Franz J. (2004): Kooperationsnetzwerke und Karrieren an deutschen Hochschulen. Der Weg zur Professur am Beispiel des Faches Psychologie. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 56 (3): 520-538
- McCallister, Lynne/Fischer, Claude S. (1978): A Procedure for Surveying Personal Networks. In: Sociological Methods and Research 7 (2): 131-148
- Moreno, Jacob Levy (1954): Die Grundlagen der Soziometrie. Köln: Springer
- Pappi, Franz Urban (Hg.) (1987): Methoden der Netzwerkanalyse. München: Oldenbourg
- Scott, John (2000): Social Network Analysis. A Handbook. Thousand Oaks/London/New Delhi: Sage
- Trappmann, Mark/Hummell, Hans J./Sodeur, Wolfgang (2005): Strukturanalyse sozialer Netzwerke. Konzepte, Modelle, Methoden. Wiesbaden: VS Verlag
- White, Harrison C. (2002): Markets From Networks. Princeton: Princeton University Press
- Wolf, Christof (2006): Egozentrierte Netzwerke. In: Sonderheft 44 der KZfSS. 244-273
- Ziegler, Rolf (1984): Das Netz der Personen- und Kapitalverflechtungen deutscher und österreichischer Wirtschaftsunternehmen. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 36 (3): 585-614

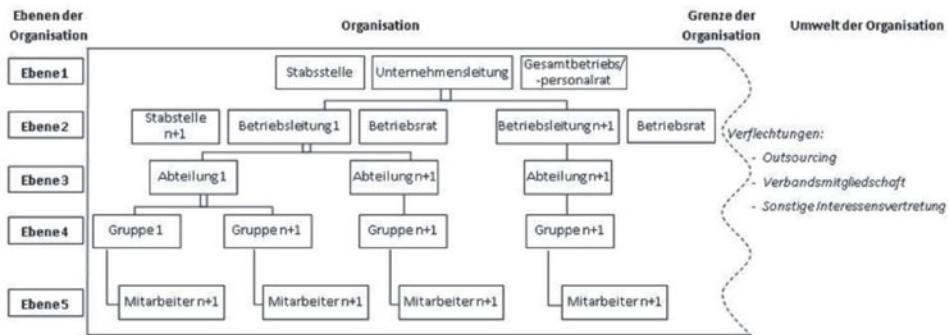
Alexia Meyermann, Tobias Gebel und Stefan Liebig

---

## 75.1 Organisationen als Gegenstand sozial- und wirtschaftswissenschaftlicher Forschung

In den Sozial- und Wirtschaftswissenschaften setzt sich mehr und mehr die Erkenntnis durch, dass die Ebene der einzelnen Organisation für das Verständnis einer Reihe von Forschungsgegenständen entscheidend ist. Neben der Makroebene (Staat und Gesellschaft) und der Mikroebene (Individuum) steht zunehmend die Organisation als Analyseeinheit zur Erklärung sozialer Phänomene im Fokus der Sozialwissenschaften (z.B. Coleman 1986, Jäger/Schimank 2005). Damit einhergehend hat die Nachfrage nach Organisationsdaten in den letzten Jahrzehnten stark zugenommen. Dies geschah nicht nur in der genuinen Organisationsforschung, sondern in einer Vielzahl von Untersuchungsfeldern, in denen Organisationen theoretisch und empirisch eine zunehmend wichtigere Rolle spielen. Neben die klassischen, auf Organisationen fokussierenden Disziplinen wie die Arbeits- und Industriesoziologie, die Organisations- und Wirtschaftssoziologie oder die Ökonomie, treten u.a. die soziologische Ungleichheitsforschung und die Bildungs- und Gesundheitswissenschaften. Dabei fragen Ungleichheitsforscher, inwiefern Organisationen als Kontexte oder Akteure sozialer Ungleichheitsgenese in Erscheinung treten, Bildungsforscher, inwiefern Schulen jenseits von Klassen und Bildungssystemen für Leistungsunterschiede zwischen Schülern verantwortlich sind, und Gesundheitsforscher, inwiefern Strukturen und Praktiken in Krankenhäusern den Erfolg medizinischer Maßnahmen bestimmen.

Organisationen (z.B. Betriebe, Schulen, öffentliche Verwaltungen, Universitäten, Krankenhäuser u.a.) sind komplexe soziale Gebilde, bestehend aus mehreren, mehr oder weniger eigenständigen (Unter-)Einheiten und Ebenen, deren äußere Grenzen zunehmend mit ihrer Umwelt verschwimmen und durchlässiger werden (siehe Abb. 75.1).



**Abb. 75.1** Beispielorganigramm - Ebenen der Organisationsanalyse (eigene Abbildung)

Abbildung 75.1 veranschaulicht anhand eines Beispielorganigramms die potenziell vorhandenen Ebenen innerhalb einer Betriebsorganisation. Die Darstellung beschränkt sich nicht auf interne Strukturen, sondern bezieht die *Organisationsumwelt* als Forschungsgegenstand mit ein. Sie zeigt, dass

1. Organisationen auf die Umwelt wirken,
2. die Umwelt für Strukturen und Prozesse in Organisationen bedeutsam ist und
3. Organisationsgrenzen unschärfer und durchlässiger werden (z.B. durch Auslagerung von Aktivitäten, Netzwerkbildung u.a.).

Je nach Forschungsperspektive können damit Organisationsdaten zur Bearbeitung von Analysen auf verschiedenen Aggregatebenen (Graeff, Kapitel 72 in diesem Band) benutzt werden. So werden Personen und Gruppen in Organisationen als interagierende Personen betrachtet (*Mikroperspektive*) oder auch Organisationen als Ganze mit Blick auf ihre internen Strukturen und Prozesse untersucht (*Mesoperspektive*). In der *Makroperspektive* rücken Organisationen selbst als Teil des wirtschaftlichen bzw. gesellschaftlichen Systems in den Fokus.

### 75.1.1 Organisationen als Akteure und soziale Kontexte

Organisationsdaten können erstens dazu genutzt werden, das *Handeln von Organisationen als „korporative Akteure“* (Coleman 1986) zu beschreiben. Als Akteure treten Organisationen beispielsweise als Abnehmer von Arbeitskraft (*Arbeitsmärkte*), Anbieter von Bildungs- und Weiterbildungsleistungen (*Bildungswesen*) oder Anbieter oder Nachfrager von Gütern und Dienstleistungen (*Güter- und Finanzmärkte*) auf. In dieser Hinsicht interessieren Praktiken und Strategien von Organisationen sowie deren wirtschaftliche und gesellschaftliche Implikationen. Organisationale Akteure vertreten ihre Interessen auch im *politischen System* neben den speziell darauf ausgerichteten Interessensorganisationen wie

Gewerkschaften und Arbeitgeberverbände. Die für entsprechende Analysen notwendigen Daten beziehen sich auf die jeweiligen Organisationen als Analyseeinheiten.

Zweitens dienen Organisationsdaten zur Beschreibung von *Organisationen als Kontexte individuellen und kollektiven Handelns*. In ihrer Eigenschaft als soziale Aggregate (Graeff, Kapitel 72 in diesem Band) begrenzen und ermöglichen sie individuelles Handeln. In diesem Zusammenhang interessieren sowohl die *Strukturen einer Organisation* – in statischer Perspektive – als auch die in ihr *ablaufenden Prozesse* – in dynamischer Perspektive. Die dafür notwendige Datenstruktur umfasst einerseits Informationen über Organisationen als Kontexte und andererseits Informationen über die in diesen Kontexten handelnden Individuen.

Ein Beispiel für die Relevanz von Organisationen als Kontexte individuellen Handelns ist etwa ihre Bedeutung für berufliche Aufstiegsmobilität, da individuelle Aufstiegschancen sowohl von der Verfügbarkeit von Positionen (offene Stellen im Betrieb), als auch von der betrieblichen Personalstrategie (interne oder externe Rekrutierung) abhängen. Die Analyse individueller Aufstiegschancen benötigt dann eine spezifische Datenstruktur, nämlich Informationen über Individuen (z.B. Alter, Geschlecht, Bildung, Beruf) und Organisationen (z.B. Wachstum, Rekrutierungspraxis) zugleich, wie ihn sogenannten *verknüpften Personen-Betriebsdatensätzen bzw. Linked-Employer-Employee Data* enthalten (dazu unten mehr). (dazu unten mehr).

### 75.1.2 Organisationen als zu erklärendes und als erklärendes Phänomen

Daten zu Organisationen werden sowohl für die Erklärung organisationaler Phänomene benötigt, als auch für die Erklärung von Phänomenen, an deren Entstehung Organisationen beteiligt sind.

In der klassischen Organisationsforschung steht die *Organisation als das zu erklärende Phänomen* im Mittelpunkt. Es geht hier um die Ursachen für das Entstehen und die Existenz von Organisationen. Später rückten Phänomene wie Wachstum oder Schrumpfung, die Ausbildung bestimmter Strukturmerkmale (wie Zentralisierung und Differenzierung), die Formen der Arbeitsorganisation und die Gestaltung und Praxis der Personalpolitik in den Blick. Darüber hinaus kann die Wirkung wirtschafts- oder auch bildungspolitischer Maßnahmen auf Existenz und Aussehen von Organisationen evaluiert werden. Die zentralen Fragestellungen hierbei sind, wie Organisationen auf Veränderungen in ihren rechtlichen, ökonomischen oder sozialen Kontextbedingungen reagieren und welche Effekte bestimmte Veränderungen in der Organisationsumwelt auf die Organisation haben.

Demgegenüber kann die *Organisation* auch ein *erklärender Faktor* für Phänomene sowohl der Makroebene (Konjunkturzyklen, Arbeitsmarktdynamiken, Bildungsbeteiligung) als auch der Mikroebene (Arbeitsbedingungen, Karriere- und Bildungschancen, soziale Beziehungen und subjektives Wohlbefinden) sein, weshalb sie oft im Rahmen von Mehrebenen-Designs (Pötschke, Kapitel 87 in diesem Band) analysiert wird.

Zur Erklärung von Phänomenen auf der Makroebene stellen Organisationen die Einheiten auf der Mikroebene dar. Entsprechend werden bei solchen Fragestellungen – etwa in der Wirtschaftsforschung – Organisationsdaten auch als „Mikrodaten“ bezeichnet. Fragestellungen, die sich auf organisationsexterne Folgen organisationalen Handelns oder organisationsinterner Strukturen und deren Veränderungen beziehen, sind beispielsweise die Folgen betrieblicher Beschäftigungsentwicklung, Aus- und Weiterbildung, Produktivität und Investitionen für die entsprechenden, auch wirtschaftspolitisch wichtigen Themenfelder. Themen der organisationsbezogenen Bildungsforschung sind die relevanten organisatorischen Rahmenbedingungen und Regelungen in den einzelnen Organisationen, deren Zusammensetzung in Bezug auf das Personal (Lehrer) und ihre Klienten (Kinder, Schüler, Studenten) und die sich daraus ergebenden Effekte auf Bildungsleistungen, Bildungsbeteiligungen oder andere soziale Ungleichheiten (Gamoran et al. 2006).

---

## **75.2 Sekundäranalytische Nutzung von Organisationsdaten: Bestand und Zugang**

Um die durch diese unterschiedlichen Perspektiven in den Blick geratenden Phänomene empirisch adäquat abbilden (Beschreibung) und ihre jeweiligen Ursachen und Folgen identifizieren zu können (Erklärung), sind Daten über das Handeln von Organisationen und von deren Mitgliedern sowie über Strukturen und Prozesse von Organisationen notwendig. Die in den Sozial- und Wirtschaftswissenschaften bisher genutzten Organisationsdaten lassen sich grob in drei Gruppen unterteilen:

1. Daten, die von der amtlichen Statistik zur Verfügung gestellt werden und aus Erhebungen (verpflichtend oder freiwillig) oder administrativen Aufgaben (prozessgenerierte Daten) resultieren,
2. Organisationsdaten, die von wissenschaftlichen Einrichtungen, Forschungsverbünden und Einzelforschern erhoben werden und
3. Daten kommerzieller Anbieter.

### **75.2.1 Organisationsdaten der amtlichen Statistik**

Organisationsdaten der amtlichen Statistik (Hartmann/Lengerer, Kapitel 71 in diesem Band) sind die Betriebs- und Unternehmensdaten der Statistischen Ämter des Bundes und der Länder, der Bundesbank, des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA) und des Bundesinstituts für Berufsbildung (BIBB). Sie werden im Rahmen des gesetzlichen Auftrags der statistischen Ämter direkt über regelmäßige Betriebs- bzw. Unternehmensbefragungen (Borg, Kapitel 57 in diesem Band) erhoben, oder sie sind das Ergebnis von Verwaltungshandlungen, wie z.B. der Statistik der sozialversicherungspflichtigen Beschäftigten oder der Steuerverwaltung.

Die Daten sind in den *Forschungsdatenzentren (FDZ)* der einzelnen Einrichtungen für die Wissenschaft verfügbar, z.B. FDZ der Statistischen Ämter des Bundes und der Länder ([www.forschungsdatenzentrum.de](http://www.forschungsdatenzentrum.de)) oder FDZ der BA im IAB ([fdz.iab.de](http://fdz.iab.de)).

Für die amtliche Statistik finden die Erhebungen nach Wirtschaftsbereichen getrennt statt, woraus sich ein unterschiedliches Datenangebot ergibt. Es reicht von monatlichen Angaben und Totalerhebungen bis hin zu jährlichen Stichprobenerhebungen (Brandt et al. 2007). Die *Daten der Statistischen Ämter* bestehen daher aus *Daten für einzelne Wirtschaftsbereiche* sowie *bereichsübergreifende Statistiken, Finanz- und Steuerstatistiken*.

Neben den Daten der Statistischen Ämter werden von der Bundesagentur für Arbeit weitere Datensätze zur Verfügung gestellt. Dabei handelt es sich einmal um das *IAB-Betriebspanel*, das auf einer jährlichen, freiwilligen Befragung von ca. 16.000 Betrieben seit dem Jahr 1993 (in Ostdeutschland: 1996) beruht. Das IAB-Betriebspanel stellt von den verfügbaren Datensätzen der amtlichen Statistik den thematisch breitesten Umfang an Betriebsinformationen bereit. Mit seinem längsschnittlichen Forschungsdesign (Schupp, Kapitel 73 in diesem Band) und durch die branchenübergreifende Stichprobe bietet es eine Vielzahl von Analysemöglichkeiten für Forschende und stellt damit den aktuell wahrscheinlich bedeutendsten, quantitativen Organisationsdatensatz dar.

Daneben ist das *Betriebs-Historik-Panel (BHP)* verfügbar. Als prozessgenerierter Organisationsdatensatz beinhaltet es ab 1975 (Ost: ab 1992) aus den Angaben der sozialversicherungspflichtig Beschäftigten aggregierte Informationen zur Personal- und Lohnstruktur von jährlich zwischen ca. 650.000 und 1,3 Millionen Betrieben. Beide Datensätze sind im Forschungsdatenzentrum der BA im IAB verfügbar.

Seit 1989 wird zudem mit der *IAB-Erhebung des Gesamtwirtschaftlichen Stellenangebots (EGS)* eine jährliche Querschnittsstudie durchgeführt, deren Bedeutung in der Gewinnung von Daten über die Arbeitsnachfrage und den Verlauf von Stellenbesetzungsprozessen liegt. Im Angebot des Forschungsdatenzentrums (FDZ) finden sich außerdem die zwei verknüpften Personen- und Betriebsdatensätze „*Linked-Employer-Employee-Daten des IAB*“ (LIAB) und das Panel „*WeLL*“ des Projektes „*Berufliche Weiterbildung als Bestandteil Lebenslangen Lernens*“. Beide Datensätze ermöglichen die simultane Analyse der Angebots- und Nachfrageseite des Arbeitsmarktes und aufgrund ihrer Mehrebenendatenstruktur (Beschäftigte in Betrieben) die Anwendung eines breiten Spektrums statistischer Analyseverfahren.

Für viele der amtlichen Erhebungen besteht eine gesetzliche Auskunftspflicht der Betriebe. Die daraus resultierenden Daten zeichnen sich daher durch *geringe Teilnahmeverweigerungen* (Engel/Schmidt, Kapitel 23 in diesem Band) und *vollständige Informationen auch bei sensiblen Fragen* (Hlawatsch/Krickl, Kapitel 20 in diesem Band) aus. Zugleich weisen die von der amtlichen Statistik zur Verfügung gestellten Datensätze *sehr hohe Fallzahlen* (Häder/Häder Kapitel 18 in diesem Band) auf, so dass differenzierte Analysen z.B. auch in kleineren regionalen Einheiten (Lakes, Kapitel 77 in diesem Band) vorgenommen werden können. Die verfügbaren Mikrodaten der Wirtschaftsstatistik liegen in der Regel auch als *Paneldaten* (Schupp, Kapitel 73 in diesem Band) vor, so dass Wandlungsprozesse

untersucht und Modelle kausaler Erklärungen (Kühnel/Dingelstedt, Kapitel 80 in diesem Band) überprüft werden können (Brandt et al. 2007).

Ein zentrales Problem besteht allerdings darin, dass die Organisationsdaten der amtlichen Statistik jeweils *nur einen Ausschnitt der für die Wissenschaft interessanten Informationen* erfassen und die Erhebungen zum Teil *auf einzelne Wirtschaftsbereiche beschränkt* sind. Eine Möglichkeit, dieses Defizit zu beheben, besteht in der seit 2005 gesetzlich möglichen *Verknüpfung unterschiedlicher Datensätze* (Cielegak/Rässler, Kapitel 26 in diesem Band) der amtlichen Statistik über das *Unternehmensregister* ([www.unternehmensregister.de](http://www.unternehmensregister.de)). In den Projekten „*Kombinierte Firmendaten für Deutschland*“ (KombiFiD) und „*Amtliche Firmendaten für Deutschland*“ (AFiD) wurden solche Verknüpfungen durchgeführt. Für KombiFiD werden zur Generierung eines kombinierten Unternehmensdatensatzes Daten der Statistischen Ämter des Bundes und der Länder, des IAB, der BA sowie der Deutschen Bundesbank genutzt. Ziel des Projekts AFiD ist die Integration aller Mikrodaten der Wirtschafts- und Umweltstatistiken der amtlichen Statistik über das Unternehmensregister. Das Analysepotential der einzelnen Unternehmens- bzw. Firmendaten wird damit deutlich erhöht, weil Informationen aus verschiedenen Statistiken sich ergänzen und gemeinsam genutzt werden können.

### **75.2.2 Durch die Wissenschaft generierte Organisationsdaten**

Organisationsdaten, die im Rahmen von Forschungsprojekten durch Forschungseinrichtungen, Forschungsverbünde oder Einzelforscher generiert wurden, zeichnen sich in der Regel durch *geringere Stichprobengrößen* aus, sind aber *stärker an substanzwissenschaftlichen Fragestellungen orientiert*. Zudem weisen sie ein *großes methodisches Spektrum* quantitativer und qualitativer Erhebungsverfahren auf. In diesem Sinne erweitern sie das amtliche Datenangebot und stellen die Grundlage für die organisationsbezogene Forschung in unterschiedlichen disziplinären Kontexten dar. Es existiert eine Vielzahl durch öffentliche Forschungsmittel (DFG, Ressortforschung) oder Mittel anderer Förderinstitutionen (z.B. Volkswagen-Stiftung, Hans-Böckler-Stiftung, Thyssen-Stiftung) finanzierte quantitative und qualitative Organisationsdaten, die in den meisten Fällen jedoch nur in begrenztem Umfang für wissenschaftliche Sekundäranalysen zur Verfügung stehen.

Die Verwendung qualitativer Verfahren (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) und die Konzentration auf einige wenige Betriebe bzw. Unternehmen ist im internationalen Vergleich lange Zeit ein Spezifikum bundesdeutscher Organisationsforschung gewesen (Grunow 1995). Was die Erhebung und Bereitstellung quantitativer Daten zu Wirtschaftsorganisationen in Deutschland anbelangt, so ist die Situation bis Ende der 1980er Jahre vor allem durch kleinere, auf einzelne Wirtschaftsbereiche, Unternehmensarten und Regionen begrenzte, querschnittliche Studien geprägt. Erst in den 1990er Jahren werden eine Reihe größerer, auch längsschnittlich angelegter Organisationsdatensätze generiert (z.B. NIFA-Panel). Geringe Stichprobenumfänge und fehlende Längsschnittdesigns stellen ein zentrales Defizit wissenschaftlicher Organisationsdaten dar (Wagner

2008). Denn nur Datensätze mit hinreichend großen Stichprobenumfängen ermöglichen differenzierte Analysen auf der Ebene von Wirtschaftszweigen, Teilbranchen oder regionalen Einheiten. Längsschnittliche Daten sind erforderlich, um Wandlungsprozesse im Zeitverlauf adäquat abbilden und kausale Erklärungsmodelle überprüfen zu können.

In den letzten Jahren gab es international (in den USA das *National Organizations Survey NOS* ([www.icpsr.umich.edu/icpsrweb/ICPSR/studies/04074](http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/04074)), in UK die *Workplace Employment Relations Study WERS* ([www.gov.uk/government/collections/workplace-employment-relations-study-wers](http://www.gov.uk/government/collections/workplace-employment-relations-study-wers)) und national einige Projekte, die eine Verknüpfung von Organisations- und Individualdaten vornahmen. Derartige *LEE-Datensätze* (*Linked-Employer-Employee-Datensätze*) eröffnen neue Analysepotentiale, weil man damit beispielsweise in der Arbeitsmarktforschung sowohl die Arbeitsnachfrage als auch die Arbeitsangebotsseite betrachten kann (Kmec 2003, Hamermesh 2008). Weil diese Daten auch außerhalb der klassischen Organisationsforschung zunehmend genutzt werden, erweitert sich auch das entsprechende Angebot. Zu nennen sind hier erstens Daten, die Individual- und Haushaltsinformationen mit Informationen über Arbeitgeber verbinden, sog. *Employee-First-Datensätze*, z.B. *ALLBUS-BB 2008* (DOI:10.7478/s0001.1.v1) oder *SOEP-LEE 2012/2013* (DOI:10.7478/s0549.1.v1). Hiervon zu unterscheiden sind, zweitens, solche Datensätze, in denen eine Vielzahl von Mitarbeitern ausgewählter Betriebe befragt werden – sog. *Employer-First-Datensätze*, wie z.B. *BEATA* (DOI:10.7478/s0002.1.v1) oder *SFB882-B3* (DOI:10.4119/unibi/sfb882.2014.12).

Eine weitere Quelle wissenschaftlicher Organisationsdaten sind die Datenerhebungen von Forschungsinstituten, wie diejenigen des *ifo-Instituts* – z.B. *EBDC Business Panels* ([www.cesifo-group.de](http://www.cesifo-group.de)) – oder des *Zentrum für Europäische Wirtschaftsforschung* (ZEW) – z.B. das *KfW-Mittelstandspanel* ([www.zew.de](http://www.zew.de)). In den letzten Jahren wurden diese Daten über die jeweiligen Datenzentren zur Sekundärnutzung zur Verfügung gestellt. Daneben gibt es aktuell eine wachsende Anzahl von Forschungsinstituten – z.B. das *sofi Göttingen* ([www.sofi-goettingen.de](http://www.sofi-goettingen.de)) –, die Vorbereitungen treffen, ihren – auch qualitativen – Datenbestand für die Wissenschaft zu öffnen.

### 75.2.3 Organisationsdaten kommerzieller Anbieter

Kommerzielle Organisationsdaten werden derzeit vor allem für Unternehmensdaten angeboten. Zwei in Deutschland sehr bekannte Datenbanken sind die *Hoppenstedt-Firmendatenbank* ([www.firmendatenbank.de](http://www.firmendatenbank.de)) und die *Creditreform-Firmenprofile* ([www.firmenwissen.de](http://www.firmenwissen.de)). Für internationale Unternehmensdaten ist beispielsweise der Datenbestand des *Bureau van Dijk* ([www.bvdinfo.com](http://www.bvdinfo.com)) zu nennen. In der Datenbank *Lexis-Nexis* ([www.lexisnexis.de](http://www.lexisnexis.de)) sind Informationen über Unternehmen aus unterschiedlichen Quellen – Hoppenstedt, Bundesanzeigern, Handelsregistern und Pressemeldungen – zusammengetragen.

Unternehmensdatenbanken ist üblicherweise gemeinsam, dass sie nur einen begrenzten Umfang an Angaben enthalten (z.B. Rechtsform, Umsatzzahlen und Mitarbeiterzahlen der vergangenen Jahre, Kontaktdata der obersten Führungsebene) und sich auf

Unternehmen als rechtliche Einheiten beziehen. Im Unterschied zum IAB-Betriebspanel sind somit keine Informationen über Betriebe enthalten, die den Kontext individuellen Arbeitshandelns bilden. Die Informationen beruhen auf Eintragungen in den öffentlichen Registern und auf den Recherchen der Anbieter und Firmenbefragungen. Damit ist der *Informationsbestand für viele sozial- und wirtschaftswissenschaftliche Fragen sehr eingeschränkt*. Hinzu kommen methodische Schwächen, weil der Prozess der Datengenerierung in der Regel kaum dokumentiert und damit nicht nachvollziehbar ist. Die Zugänglichkeit dieser Daten für die Wissenschaft ist jedoch als sehr gut zu bezeichnen, weil die Anbieter entweder spezielle, kostengünstige Angebote für Forschungszwecke offerieren oder die Datenbanken Teil des Standardangebots vieler wissenschaftlicher Hochschulen sind.

Darüber hinaus existiert eine Reihe anderer nationaler oder auch international vergleichender Datensätze über geschäfts- und personalpolitische Strategien von Betrieben und Unternehmen, die von privaten Wirtschafts- oder Unternehmensberatungsfirmen zumeist über Befragungen generiert sind. Sie weisen einen breiteren Informationsbestand auf, sind jedoch für wissenschaftliche Zwecke nicht bzw. nur über entsprechende bilaterale Kooperationen von Wissenschaftlern und dem einzelnen Datenproduzenten zugänglich.

#### **75.2.4 Dateninfrastruktur**

In den letzten Jahren haben sich die Nutzungsmöglichkeiten für Organisationsdaten durch den Aufbau einer differenzierten Forschungsdateninfrastruktur ([www.ratswd.de](http://www.ratswd.de)) erheblich verbessert. Mittlerweile existiert eine Reihe von Forschungsdatenzentren (FDZ), die Organisationdaten zur wissenschaftlichen Sekundäranalyse bereitstellen. Neben den bereits genannten FDZs sind dies:

- das FDZ *Bildung des Deutschen Institut für Internationale Pädagogische Forschung* ([www.fachportal-paedagogik.de/forschungsdaten\\_bildung](http://www.fachportal-paedagogik.de/forschungsdaten_bildung)),
- das *Economics and Business Data Center des ifo-Instituts (EBDC)* ([www.cesifo-group.de/de/ifoHome/facts/EBDC](http://www.cesifo-group.de/de/ifoHome/facts/EBDC)) an der LMU München,
- das FDZ *des Zentrums für Europäische Wirtschaftsforschung (ZEW-FDZ)* ([www.zew.de/de/forschung/zew-forschungsdatenzentrum.php](http://www.zew.de/de/forschung/zew-forschungsdatenzentrum.php)),
- das *Internationale Datenservicezentrum des Forschungsinstituts zur Zukunft der Arbeit (IDSC am IZA)* ([idsc.iza.org](http://idsc.iza.org)) und
- das FDZ *Ruhr am Rheinisch-Westfälisches Institut für Wirtschaftsforschung* (Stand 11/2013; [fdz.rwi-essen.de](http://fdz.rwi-essen.de)).

Das auf Organisationsdaten spezialisierte *Datenservicezentrum Betriebs- und Organisationsdaten der Universität Bielefeld (DSZ-BO)* ([www.uni-bielefeld.de/dsz-bo](http://www.uni-bielefeld.de/dsz-bo)) bietet ein umfassendes Informations- und Dienstleistungsangebot für Datenproduzenten und Datennutzer aus der Wissenschaft. Über sein webbasiertes Studienportal vermittelt es einen breiten Überblick über quantitative und qualitative Studien der sozial- und wirtschaftswissenschaftlichen Organisationsforschung, das Datenangebot umfasst quantitative und qualitative Studien und wird fortlaufend erweitert (Edler et al. 2012).

## 75.3 Primärerhebung von Organisationsdaten

Wenn der vorhandene Bestand an Sekundärdaten für die Bearbeitung der eigenen Forschungsfragen nicht geeignet ist, sind eigene Erhebungen durchzuführen. Hierzu kann auf das gesamte Spektrum der qualitativen und quantitativen Methoden der empirischen Sozialforschung zurückgegriffen werden. Gleichwohl ergeben sich sowohl bei der Stichprobenziehung als auch bei der Datenerhebung spezifische Besonderheiten und Probleme, die es zu beachten gilt (Snijkers et al. 2013).

### 75.3.1 Stichprobenziehung

#### 75.3.1.1 Bestimmung der Grundgesamtheit

Für die *Auswahl der zu untersuchenden Einheiten* (Betriebe, Unternehmen, Organisationen) und zur Bestimmung der Grundgesamtheit kann in Deutschland auf verschiedene *Adressbestände* zurückgegriffen werden. Zu unterscheiden sind Adressbestände

- der Bundesagentur für Arbeit,
- des Unternehmensregisters,
- kommerzieller Anbieter,
- von Verbänden und Kammern sowie diejenigen in
- Branchentelefonbüchern.

Die *Grundgesamtheit der Bundesagentur für Arbeit* enthält ausschließlich Betriebe innerhalb Deutschlands mit mindestens einem sozialversicherungspflichtig Beschäftigten. Eine Adressziehung aus diesem Datenbestand ist gegen Gebühr und unter Einhaltung der rechtlichen Voraussetzungen (§75 SGB10) möglich. Aufgrund gesetzlicher Bestimmungen sind im *Unternehmensregister* alle Organisationen aufgelistet, die juristische Personen sind. Dieser Adressbestand ist kostenfrei verfügbar. Im Gegensatz zu den Datenbeständen der BA und des Unternehmensregisters sind diejenigen der *kommerziellen Anbieter* unvollständig und selektiv, insbesondere wenn sie Unternehmenstypen ausschließen und auf freiwilligen Einträgen beruhen. Je nach Forschungsinteresse können auch Adressbestände von *Wirtschaftsverbänden und Kammern* oder der Rückgriff auf *Branchentelefonbücher* geeignet sein.

Eine Möglichkeit der Adressgenerierung besteht in der sog. *Employee-First-Methode*, die zur Generierung von LEE-Daten genutzt wird (Kmec 2003). Hier werden bei den Arbeitnehmern die Kontaktdaten ihrer Arbeitgeber erhoben, um diese dann für eine gesonderte Betriebs- oder Unternehmensstudie zu nutzen (so etwa SOEP-LEE 2012/2013).

Die genannten Adressbestände unterscheiden sich nicht nur hinsichtlich *Vollständigkeit* und der zugrundeliegenden *Definition von Organisationen* (z.B. juristische Person vs. Arbeitgeber), sondern auch hinsichtlich der *Aktualität* und des *Umfangs an zusätzlichen Informationen*, die sie enthalten. Die Aktualität des Adressbestands ist wichtig, da Organisationen aufgrund möglicher Umfirmierungen, Namensänderungen, Übernahmen, Auflösungen oder Umzüge sehr fluide Objekte sind. Der Bestand an Zusatzinformationen ist

sowohl relevant für die Auswahl der zu untersuchenden Organisationen (75.3.1.2.) als auch für die Auswahl von Organisationsmitgliedern aus der Organisation heraus (75.3.1.3.).

### 75.3.1.2 Auswahl der zu untersuchenden Organisationen

Neben der *Ziehung von Zufallsstichproben* (Häder/Häder, Kapitel 18 in diesem Band) kann die Auswahl der zu untersuchenden Organisationen durch geschichtete Stichprobenziehung oder bei qualitativen Befragungen durch bewusste Fallauswahl erfolgen. *Geschichtete Stichprobenziehungen* (Häder/Häder, Kapitel 18 in diesem Band) erfolgen üblicherweise nach den Merkmalen Branche und Größe (Beschäftigtenzahl), wofür entsprechende Informationen in den Adressbeständen erforderlich sind. Für eine *bewusste Fallauswahl* (Akremi, Kapitel 17 in diesem Band) durch ein *theoretisches Sampling* (Strübing, Kapitel 32 in diesem Band) können ganz unterschiedliche, weitergehende Informationen über die Organisationen benötigt werden. Man kann sich dabei bestehender Organisationsdatensätze bedienen und auf der Grundlage der dort zur Verfügung gestellten Informationen entsprechende Auswahlen treffen. Zusätzlich ermöglicht ein Rückgriff auf bestehende Organisationsdaten, die bereitgestellten Informationen für eine Anpassung der Erhebungsmaterialien oder Erhebungsschritte zu nutzen (sog. *Tailoring*).

### 75.3.1.3 Auswahl von Organisationsmitgliedern

Das spezifische Problem bei Organisationsstudien besteht darin, dass Organisationen nicht für sich selbst Auskunft geben können. Daher tritt neben den Schritt der Auswahl von Organisationen ein weiterer Auswahlschritt hinzu: Die Auswahl von Mitgliedern der Organisation. Zwei Verfahren sind zu unterscheiden:

1. Die *Auswahl von Proxy-Informanten*, die stellvertretend für die zu untersuchende Organisation Auskunft geben. Auf solche Stellvertreterauskünfte wird beispielsweise beim Einsatz standardisierter Organisationsbefragungen oder Expertenbefragungen zurückgegriffen, die sich an die auskünftsfähigen Repräsentanten der jeweiligen Organisation richten.
2. Die *Auswahl von Mitarbeitern oder sonstigen Mitgliedern der Organisation*, die für sich selbst in ihrer Rolle als Organisationsmitglied Auskunft geben. Mitgliederinformationen sind auf die höhere Ebene der Organisation *aggregierbar* (Graeff, Kapitel 72 in diesem Band). Insofern können Mitarbeiter- und Mitgliederbefragungen (Borg, Kapitel 57 in diesem Band) dazu verwendet werden, Informationen über die Organisation als Ganzes zu erhalten.

Die *Identifizierung eines geeigneten Informanten*, d.h. eines auskünftsfähigen und -bereiten Ansprechpartners in der Organisation ist für quantitative ebenso wie für qualitative Befragungen schwierig. Hürden durch „Gatekeeper“ in Gestalt von Telefenzentralen, Sekretariaten oder Firmenhotlines sind zu überwinden, möglicherweise schließen Firmenpolitiken die Teilnahme an Umfragen generell aus, und die Auskünftsfähigkeit des Ansprechpartners schwankt mit der Ausdifferenzierung von Informationen innerhalb der Organisation. Beobachtbare Merkmale (z.B. die Stellung in der Positionsstruktur, der

Aufgabenbereich eines Organisationsmitgliedes) helfen bei der Bestimmung von Proxy-Informanten. Sollen mehrere Personen in einer Organisation befragt werden, so kann die Auswahl auf Basis von Personallisten erfolgen, was aber die Kooperation der Organisationsleitung voraussetzt. Oftmals wird eine Auswahl von Befragungspersonen durch die Organisationsleitung vorgegeben, was insofern problematisch ist, weil durch diese bewusste Auswahl bestimmte Perspektiven auf den Organisationsalltag möglicherweise nicht abgebildet werden können.

### 75.3.2 Datenerhebung

Zu den häufig angewandten Erhebungsverfahren der Organisationsforschung zählen Expertenbefragungen (Helfferich, Kapitel 39 in diesem Band), Fokusgruppen-Interviews (Vogl, Kapitel 41 in diesem Band), Dokumentenanalysen (Salheiser, Kapitel 62 in diesem Band) oder Betriebs-Fallstudien (Hering/Schmidt, Kapitel 37 in diesem Band). Daneben werden Befragungen (Surveys) (Reinecke, Kapitel 44 in diesem Band) eingesetzt, wobei Organisationssurveys, also Befragungen von Organisationen als Ganzes, von Mitarbeiter- oder Mitgliederbefragungen (Borg, Kapitel 57 in diesem Band) zu unterscheiden sind.

#### 75.3.2.1 Quantitative Erhebungsmethoden

Die Entscheidung, welche (quantitativen und/oder qualitativen) Methoden zur Gewinnung von Organisationsdaten eingesetzt werden, ist selbstverständlich abhängig von der Fragestellung.

*Standardisierte Organisationsbefragungen* werden eingesetzt, um Aussagen über Organisationen insgesamt (in ihren Akteurseigenschaften) und im Vergleich (über die Population von Organisationen) treffen zu können. Üblicherweise werden mindestens zentrale Eigenschaften wie Branche, Rechtsform, Beschäftigtenzahl oder Geschäftsvolumen erfasst, wenngleich eine einheitliche Definition von „Standardmerkmalen“ von Organisationen, wie es sie in der Individualsurveyforschung gibt (Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band), bislang nicht definiert wurde (Krause 2012). Auch darüber hinausgehende, erprobte und in ihrer methodischen Qualität geprüfte Skalen oder Items (Latechva/Davidov und Blasius, Kapitel 55 und 83 in diesem Band) zur Messung von organisationalen Phänomenen ([www.meadow-project.eu](http://www.meadow-project.eu)) existieren kaum oder sind veraltet (Pugh 2003 [1973]).

Der Einsatz von Organisationsbefragungen hängt davon ab, ob Proxy-Informanten über die gewünschten Informationen hinreichend Auskunft geben können. Allgemein ist die Erfassung betrieblicher Strukturen (z.B. innerbetriebliche Heterogenität, Organisationskultur) über Organisationsbefragungen – auf Stellvertreter-Ebene – nicht so detailliert und genau möglich wie über die Aggregierung (Graeff, Kapitel 72 in diesem Band) von *Mitarbeiterbefragungen* (Borg, Kapitel 57 in diesem Band) oder von *prozessgenerierten Mitarbeiter-Daten* (z.B. Meldedaten der Sozialversicherung). So ergibt die Summe aller Merkmalsausprägungen der Mitglieder einer Organisation die Struktur eines Betriebes im

Hinblick auf dieses Merkmal (z.B. Frauenanteil, Altersstruktur). Anhand der Verteilung eines Merkmals können Betriebe beispielsweise über die Berechnung von Heterogenitäts- oder Diversitätsmaßen (Huinkink, Kapitel 82 in diesem Band) beschrieben und miteinander verglichen werden (z.B.: Wie gleich oder ungleich verteilt ist dieses Merkmal innerhalb der Belegschaft eines Betriebes?).

Wenngleich Mitarbeiterbefragungen aufgrund ihrer Informationstiefe und -genauigkeit methodisch von Vorteil sind, ist die Möglichkeit, sie für vergleichende Aussagen zu nutzen, durch den hohen Aufwand ihrer Erhebung begrenzt. Stärker als bei Bevölkerungsbefragungen stellt sich hier das *Problem des Feldzugangs*:

- Für Betriebe stellen Mitarbeiterbefragungen einen Kostenfaktor dar (organisatorischer Aufwand, Zeitaufwand für Erhebung etc.),
- es verbinden sich damit oftmals Bedenken hinsichtlich der vermuteten innerbetrieblichen Folgen, die sich aus der Thematisierung „heikler Themen“ ergeben und
- Mitarbeiterbefragungen sind mitbestimmungspflichtig, d.h. es müssen nicht nur Management und Betriebsleitung, sondern auch Betriebsrat oder Personalvertretung einverstanden sein.

Die Realisierung von Organisationsstudien im Mehrebenendesign (Pötschke, Kapitel 87 in diesem Band) ist deshalb mit besonders hohen Anforderungen verbunden.

### 75.3.2.2 Qualitative Erhebungsmethoden

Eine in der qualitativen Organisationsforschung am weitesten verbreitete Form der Erhebung ist die *Betriebsfallstudie* (Pongratz/Trinczek 2010). Üblicherweise wird hierbei ein Mix unterschiedlicher Methoden eingesetzt, wie qualitative (Experten-)Interviews, Beobachtungen, Dokumentenanalysen und standardisierte Befragungen. Eine solche Form der Erhebung zielt primär auf die Analyseeinheit Organisation, kann aber durchaus Informationen auf Individualebene enthalten. Fallstudien (Hering/Schmidt, Kapitel 37 in diesem Band) vermitteln ein umfangreiches und detailliertes Bild über die komplexen Prozesse und Strukturen innerhalb und zwischen Organisationen.

Des Weiteren finden die verschiedenen Formen des qualitativen Interviews für vielfältige Erkenntnisinteressen Anwendung (Kühl et al. 2009): *Experteninterviews* (Helfferich, Kapitel 39 in diesem Band) werden zur Gewinnung von Fach-, Dienst- und Geheimwissen für explorative oder hypothesenprüfende Fragestellungen sowie zur Erschließung bislang unbekannter Felder eingesetzt, *narrative Interviews* (Küsters, Kapitel 40 in diesem Band) zur Wiedergabe von Ereignissen und Ereignisketten in Organisationen (Transformationsprozesse, einzelne Projekte, Brüche, Erfolge und Krisen), die zur Sinnkonstitution, zur Bilanzierung und Evaluation anregen, wobei die Organisationsmitglieder eigene Relevanzen und Deutungsmuster zu den Ereignissen entwickeln können. *Gruppendifussionen* (Vogl, Kapitel 41 in diesem Band) dienen der Erfassung von Entscheidungs-zusammenhängen und kollektiven Erfahrungen, der Evaluation eines Status quo und der Maßnahmenplanung.

Neben den Interviewverfahren werden *Dokumenten- und Aktenanalysen* (Salheiser, Kapitel 62 in diesem Band) für Organisationsdiagnosen und zur Untersuchung organisationaler Rahmenbedingungen verwendet.

Die *teilnehmende Beobachtung* (Thierbach/Petschick, Kapitel 66 in diesem Band) wird zur Erhebung impliziter Wissensbestände und der Untersuchung von situativen und natürlichen Verhalten der Organisationsmitglieder genutzt.

---

## 75.4 Ausblick

Insgesamt hat sich die Verfügbarkeit von sozialwissenschaftlich relevanten Informationen über unterschiedliche Typen von Organisationen hinweg in den letzten Jahren bedeutend verbessert. Dies gilt insbesondere für die quantitative Organisationsforschung, die vermehrt auch auf komplexe Datenstrukturen (Mehrebenendesign, Längsschnittdaten, Informationen aus unterschiedlichen Datenbeständen) zurückgreifen kann. Im Bereich der qualitativen Organisationsforschung dominieren zwar immer noch Primärstudien, jedoch werden durch den Aufbau entsprechender Forschungsdateninfrastrukturen zunehmend auch qualitative Daten für die Sekundäranalyse verfügbar sein.

Davon unabhängig ist festzustellen, dass methodische Fragen und Probleme von Organisationstudien bislang wenig bearbeitet wurden. Im Unterschied zur klassischen, auf Individuen und Haushalte bezogenen empirischen Sozialforschung konnte sich im Bereich der Organisationsforschung bislang noch keine explizite Methodenforschung etablieren. Im deutschsprachigen Raum fehlen an methodischen Gütekriterien (Krebs/Menold und Flick, Kapitel 29 und 30 in diesem Band) orientierte Sammlungen von Erhebungsinstrumenten ebenso wie empirisch fundierte Darstellungen von „Best Practices“ für die einzelnen Erhebungsschritte von Organisationsstudien.

## Literatur

- Brandt, Maurice/Oberschalsiek, Dirk/Pohl, Ramona (2007): Neue Datenangebote in den Forschungsdatenzentren Betriebs- und Unternehmensdaten im Längsschnitt. FDZ-Arbeitspapier 23. Wiesbaden: Statistisches Bundesamt. URL: [http://www.forschungsdatenzentrum.de/publikationen/veroeffentlichungen/fdz\\_arbeitspapier-23.pdf](http://www.forschungsdatenzentrum.de/publikationen/veroeffentlichungen/fdz_arbeitspapier-23.pdf) am 21.08.2013
- Coleman, James S. (1986): Die asymmetrische Gesellschaft. Weinheim/Basel: Beltz
- Edler, Susanne/Meyermann, Alexia/Gebel, Tobias/Liebig, Stefan/Diewald, Martin (2012): The German Data Service Center for Business and Organizational Data (DSC-BO). In: Schmollers Jahrbuch 132 (4). Berlin: Duncker & Humblot: 619-634
- Gamoran, Adam/Secada, Walter G./Marret, Cora B. (2006): The Organizational Context of Teaching and Learning. In: Hallinan, Maureen T. (Hg.): Handbook of the Sociology of Education. New York: Springer. 37-64
- Grunow, Dieter (1995): Research Design in Organization Studies. In: Organization Science 6 (1): 93-103
- Hamermesh, Daniel (2008): Fun with matched firm-employee data. In: Labour Economics 15 (4): 663-673
- Jäger, Wieland/Schimank, Uwe (Hg.) (2005): Organisationsgesellschaft. Wiesbaden: VS Verlag
- Kmec, Julie C. (2003): Collecting and Using Employer-Worker Matched Data. In: Sociological Focus 36 (1): 81-96
- Krause, Ina (2012): Beschreibung und vergleichende Diskussion der Indikatoren und Messinstrumente zur Erfassung von betriebsdemografischen Merkmalen in der amtlichen Statistik und in standardisierten Betriebsbefragungen im Rahmen von Forschungsprojekten. DSZ-BO Working Paper 3. URL: am [http://www.uni-bielefeld.de/dsz-bo/pdf/2012-09-12\\_WP3\\_Indikatoren.pdf](http://www.uni-bielefeld.de/dsz-bo/pdf/2012-09-12_WP3_Indikatoren.pdf) 02.05.2014
- Kühl, Stefan/Strodttholz, Petra/Taffertshofer, Andreas (Hg.) (2009): Handbuch Methoden der Organisationsforschung. Wiesbaden: VS Verlag
- Pongratz, Hans J./Trinczek, Rainer (2010): Industriesoziologische Fallstudien. Berlin: edition sigma
- Pugh, D.S. (2003 [1973]): The Measurement of Organization Structures. In: Handel, Michael J. (Hg.): The Sociology of Organizations. Thousand Oaks: Sage. 66-75
- Snijkers, Ger/Haraldsen, Gustav/Jones, Jacqui/Willimack, Diane (2013): Designing and Conducting Business Surveys. Wiley
- Wagner, Joachim (2008): Improvements and Future Challenges for the Research Infrastructure in the Field of Firm Level Data. Lüneburg: Working Paper Series in Economics 88. URL: [www.uni-lueneburg.de/fb2/vwl/papers/wp\\_88\\_Upload.pdf](http://www.uni-lueneburg.de/fb2/vwl/papers/wp_88_Upload.pdf) am 21.08.2013

Jens S. Dangschat

## 76.1 Was sind qualitative raumbezogene Daten?

Der Unterschied zwischen quantitativen räumlichen Daten (Lakes, Graeff, Braun, Kapitel 77, 72 und 56 in diesem Band) und den hier diskutierten qualitativen räumlichen Daten wird am besten deutlich, wenn man „gedankliche Karten“ (mental maps) neben Stadtpläne legt. Stadtpläne geben Entfernung und Richtungen eindeutig wieder und verorten öffentliche Gebäude und andere Einrichtungen an dem jeweils „richtigen“ Ort, während gedankliche Karten genau dieses nicht erfüllen, weil sie auf Erinnerungen, selektiven Wahrnehmungen und Präferenzen beruhen.

*Weltkarten* sind oft Ausdruck eines egozentrierten Weltbildes: In Europa hergestellte Karten rücken Europa in die Mitte, während solche aus den U.S.A. und Südostasien ihren Kontinent als Mittelpunkt der Erde betrachten. Auch *Stadtpläne* geben nicht alle Informationen wieder, sondern diese sind aufgrund der Fülle potentieller Informationen selektiv (oftmals vom Auftraggeber von Anzeigen bestimmt) und die Straßen sind wegen der Lesbarkeit überdimensioniert.

Qualitative räumliche Daten orientieren sich daher nicht am physikalischen Raum und seinen „objektiven“ Gegebenheiten, sondern am *erlebten Raum*, d.h. an der jeweils subjektiven Interpretation (Küsters, Kapitel 40 in diesem Raum) des Wahrgenommenen. *Raumwahrnehmungen* sind daher immer *relational* (Baur, Kapitel 74 in diesem Band), d.h. sie beziehen sich immer auch auf andere Personen im Raum (Thierbach/Petschick, Kapitel 66 in diesem Band) sowie dessen materielle Ausstattung (Schubert, Kapitel 70 in diesem Band). Sie basieren zudem auf dem Erfahrungshintergrund der Sozialisation und sind daher *hochgradig sozial selektiv*. Das qualitative Paradigma (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) richtet sich auf das „Handeln der Menschen in der Herstellung und Interpretation von Bedeutungen in intersubjektiven Interaktionsverhältnissen“ (Rothfuß/Dörfler 2013: 23).

Die qualitative Methodik sollte sich jedoch nicht auf die Beschreibung der Konstruktionen „ersten Grades“ (Schütz 1971) sozialräumlicher Phänomene beschränken, sondern sollte auch die Konstruktionsleistungen selbst rekonstruieren (Konstruktion zweiten Grades) (Dörfler 2013a: 34). Das ist nur möglich, wenn der Raum mehr als nur ein räumliches Arrangement von Personen als Träger einzelner Merkmalsausprägungen ist. Daher gehe ich im folgenden Abschnitt auf den „relationalen Raum“ ein, um dann im zweiten Abschnitt qualitative Methoden der Raumanalyse vorzustellen.

---

## 76.2 Raumbezug: Container-Raum zum relationalen Raum

Das (analytische) Raumverständnis ist über das materialistische Raumkonzept der chonologischen Geographie geprägt. Danach ist der Raum „objektiv“, wodurch es möglich ist, jeden Punkt hinsichtlich seiner Lage auf der Erde, der Richtung und Entfernung zu allen anderen Orten eindeutig zu bestimmen („*absoluter Raum*“). Auch der Vektorraum in der Mathematik ist eindeutig und logisch bestimmt, wie auch der Newtonsche Raum der klassischen Mechanik absolut ist, unveränderlich und unbeeinflusst von den physikalischen Vorgängen, die sich in ihm abspielen; er ist dreidimensional und Distanzen werden (meistens) euklidisch bestimmt.

Dieses „*objektive*“ Raumverständnis (von Einstein als „*Container-Raum*“ bezeichnet), welches die Grundlage für Technik- und Naturwissenschaft, Statistik und die traditionelle Raumplanung ist, unterscheidet sich jedoch fundamental von den sozialpsychologischen Grundlagen menschlichen Handelns, denn Menschen handeln nicht nach objektiven oder logischen Sachverhalten, sondern vor dem Hintergrund komplexer Motivationsbündel, eigener Sozialisationserfahrungen und eingebettet in soziale Prozesse.

Theoretisch beziehen sich diese *mikrosoziologischen Ansätze* zum einen auf Arbeiten zur Raumproduktion (Lefèvre 1975, Bourdieu 1991, Läpple 1992, Dangschat 2007), indem die Instrumentalisierung, die symbolische Aufladung und die Identifikation von Raum in den Mittelpunkt gerückt werden. Danach werden mit der materiellen und ideologischen Gestaltung („*placemaking*“; Healey 2010) Raum-Eigenschaften vermittelt, welche die Interpretation dieser Räume beeinflusst und damit – wenn auch sehr unterschiedlich gefiltert – Bestandteil unterschiedlicher *subjektiver Raumbilder* wird.

Zum anderen bezieht man sich mit den qualitativen Methoden auf die Raumsoziologie von Löw (2001). Diese basiert vor allem auf der Raumproduktion durch die (An) Ordnung von Körpern und Dingen im Raum („*Spacing*“ – was eigentlich mit „*Placing*“ bezeichnet werden sollte), als auch auf der Syntheseleistung der Menschen (das eigentliche „*Spacing*“), welche – im Rahmen der jeweiligen Erfahrungen – den Raum in seinen materiellen, sozialen und symbolischen Aspekten wahrnehmen. Problematisch in ihrem konstruktivistischen Ansatz ist jedoch der Rückfall in a-priori-Typologien, entlang derer eine unterschiedliche Wahrnehmung und Bewertung (Kognition) des sozialen Raumes vorgenommen wird (wie beispielsweise nach Geschlecht oder Alter), anstatt qualitative Typologien entlang gemeinsamer Konstruktionsleistungen relationaler sozialräumlicher

Phänomene zu bilden (vgl. zum Konzept der relationalen Milieutheorie: Dörfler 2013a: 51-55, 2013b: 246-252).

Jedoch sind nicht nur die individualistischen Deutungen von Räumen auf der mikrosoziologischen Ebene bedeutsam, sondern auch Aspekte der *Meso- und der Makroebene*. In Architektur und Städtebau ist vom „Genius Loci“, in der geschichtswissenschaftlichen Betrachtung Fernand Braudels von „Longue Durée“, in der philosophischen Betrachtung von Maurice Halbwachs vom „kollektiven Gedächtnis“ (unterteilt in personengebundenes kommunikatives Gedächtnis und ortsgebundenes kulturelles Gedächtnis) und in der Segregationsforschung vom „Habitus des Ortes“ die Rede (Dangschat 2007).

Auf der Makro-Ebene sind die Phänomene zu verorten, welche die *Produktionen des Raumes* bewirken – das sind neben materiellen Aspekten wie Investitionen, Gesetze und Verordnungen vor allem auch Interpretationen der Herausforderungen von Stadt- und Regionalentwicklung durch Stakeholder innerhalb ihrer institutionellen Framings. In die gleiche Kategorie fallen Entscheidungen von InvestorInnen als Ergebnis der Deutungen der jeweiligen Teil-Märkte. Auf der Meso-Ebene werden *Merkmale des Quartiers* wirksam, die sich nicht in „objektiven“ Infra- und Wohnbaustrukturen resp. Lagequalitäten erschöpfen, sondern auch im kulturellen Klima, in der politischen Kultur und in akteursbasierten Netzwerken (Baur, Kapitel 74 in diesem Band) mit den darin aufgeladenen Wertvorstellungen und Deutungsmustern bestehen.

---

### 76.3 Qualitative Methoden zur Raumanalyse

Mit der Anwendung qualitativer Methoden der Raumanalyse wird angestrebt, die jeweilige subjektive Wahrnehmung sozialräumlicher Artefakte („Spacing“) möglichst als genaue Reproduktion des Alltages abzubilden. Dieses kann unterschiedlich breit und intensiv in gesellschaftliche Kontexte eingebunden sein. Während die Grounded Theory (Strübing, Kapitel 32 in diesem Band) Fall für Fall analysiert und daraus neue theoretische Konstrukte ableitet, bemüht sich die ethnografische Forschung (Knoblauch, Kapitel 36 in diesem Band), die erhobenen Selbstdeutungen der jeweiligen sozialräumlichen Alltags-Phänomene im ersten Schritt als eine „dichte Beschreibung“ (Geertz 1983) zu kondensieren, um sie dann im zweiten Schritt in allgemeine gesellschaftliche Kontexte (analytische Bezugsystem) zu stellen und vor diesem Hintergrund neu zu bewerten.

Prinzipiell haben Daten immer einen Raumbezug (Lakes, Kapitel 77 in diesem Band). Dieser wird in den Sozialwissenschaften jedoch meist ausgeblendet oder – wie im Falle von amtlichen Statistiken (Hartmann/Lengerer, Kapitel 71 in diesem Band) – in einem „banalen“ Container-Raum-Verständnis (Francois Perroux) angewendet und allenfalls unzulänglich und vor allem nicht konsensual operationalisiert.

Wie problematisch eine Sekundäranalyse (Mochmann, Kapitel 14 in diesem Band) amtlicher Statistiken ist, zeigt die Segregationsforschung. Aus dem gleichzeitigen Anwesen-Sein sozialer Gruppen (erfasst lediglich durch die Meldeadresse) werden weitreichende Rückschlüsse hinsichtlich des gesellschaftlichen Zusammenhangs und sozialer

Problematik gezogen, was wiederum einen großen Einfluss auf die Stadtentwicklungs- und Belegungspolitik von Wohnraum hat. Daraus ergibt sich auch wissenschaftlich eine Reihe von Problemen, weil bislang unklar ist, worauf Nachbarschaftseffekte zurückzuführen sind (die Wechselwirkung der sozialen Gruppen, die Ausstattung mit Wohnbau- und Infrastrukturen, die Erreichbarkeit oder das Image). Zudem ist ungeklärt, was bei einer vergleichenden Forschung genau betrachtet wird, weil das Wechselverhältnis sozialer Gruppen, Bedingungen des lokalen Wohnungsmarktes sowie Rückzugstendenzen nicht überall gleich sind. Soziale Bezüge sind in ihrer sozialen Ausdifferenzierung eben nicht raum- und zeitunabhängig, sondern immer an konkrete Personen und sozialräumliche Figurationen gebunden.

### **76.3.1 Beobachtungen**

Beobachtungen (Thierbach/Petschick, Kapitel 66 in diesem Band) gehören zu den klassischen Methoden sozialräumlicher Erhebungen. Sie werden nach Formen der Offenheit und Teilhabe unterschieden und es gibt eine Reihe von Vorschlägen zur Dokumentation des Gesehenen. In der Regel werden dabei Positionen von Personen im Raum und ggf. deren Interaktionen dokumentiert. Dabei bleibt der Ort in seiner sozialen Bedeutung jedoch eher im Hintergrund und bekommt allenfalls Bedeutung über die wahrgenommenen Gelegenheiten (Bänke, Brunnen, Beleuchtung) resp. der Erdgeschossnutzung anliegender Gebäude. Selten wird die Wirksamkeit der Orte reflektiert, denn der Raum wird in der Regel allenfalls auf die Rolle der Bühne und des Bühnenbildes reduziert. So wird zwar das „Placing“ beobachtet, doch bleibt das „Spacing“ – die eigentlich die Handlungen auslösende Syntheseleistung – unberücksichtigt. Die Interpretation von Beobachtungen setzt also eine sehr gute Kenntnis der jeweiligen kulturellen Codes voraus, was jedoch in einer zunehmend vielfältigen Gesellschaft immer schwieriger wird.

### **76.3.2 Mental Maps**

Gedankliche Karten („mental maps“) sind zeichnerische Darstellungen eigener Wahrnehmungen bekannter Räume (von Nachbarschaften, Routinewegen etc.). Sie spiegeln die subjektive Interpretation „objektiver“ sozialräumlicher Tatbestände wider. Mental maps wurden ursprünglich im klinischen Bereich und der Kinder- und Jugendpsychologie eingesetzt, um Entwicklungsstörungen hinsichtlich der Raumwahrnehmung und -verarbeitung analysieren zu können. Heute werden sie u.a. in der Raumplanung und dem Städtebau (Lynch 1960) eingesetzt, um selektive Wahrnehmungen sozialräumlicher Gegebenheiten des Alltages zu erfassen. Dieses wird in der Regel thematisch zugespielt: Welche Orte nutzen Sie in Ihrer Nachbarschaft? Oder: Welche Orte empfinden Sie als unangenehm?

Mental Maps werden auch in aggregierter Form verwendet, um daraus „Stadtpläne der Kriminalität“ oder „Stadtpläne der Gefühle“ zu erzeugen. Zum einen gibt es eine

akademische Auseinandersetzung – allerdings basierend auf fragwürdigen theoretischen Annahmen („broken windows“-Theorie) und mit fragwürdigen Erhebungsmethoden (<http://pulse.media.mit.edu/vision/>; <http://biomapping.net/>; <https://www.google.at/#q=-emotional+mapping>; <http://emotionalcartography.net/EmotionalCartographyLow.pdf>). Diese „emotionale Kartographie“ wird aber auch von Privatpersonen, Versicherungsunternehmen und dem Immobiliensektor genutzt, um „no go“-areas zu beschreiben. Für nahezu jede US-amerikanische Stadt, aber beginnend auch europäische Städte, sind im Internet eine Reihe solcher „Stadtpläne der Kriminalität“ herunterzuladen, in denen nicht nur raum-zeitlich Delikte, sondern auch Kriminalitätsfurcht verortet sind (<http://www.emotionmap.net/>; <http://web.mit.edu/newsoffice/2013/quantifying-urban-perceptions-0724.html>; <http://yochicago.com/a-map-of-chicagos-no-go-zones/28003/>). Die „Stadtpläne der Gefühle“ bringen hoch aggregierte und in der Regel nicht sozial selektiv dargestellte Orte des Wohlfühlens, der Angst, der Verunsicherung und Verwahrlosung zum Ausdruck.

### 76.3.3 Räumliche Narrationen

Für sozialwissenschaftliche Analysen werden die Personen beim Zeichnen der „mental maps“ beobachtet, wobei nur auf Nachfragen kommuniziert wird. Anschließend wird anhand der fertigen Zeichnung ein narratives Interview (Küsters, Kapitel 40 in diesem Raum) geführt, in dem vor allem die Assoziationen über spezifische sozialräumliche Artefakte (Schubert, Kapitel 70 in diesem Band) aufgrund der gewählten Darstellungsweise (auch des Weglassens, der deutlichen Verzerrung, der unterschiedlichen Genauigkeit und Detailliertheit sowie der Reihenfolge) thematisiert werden.

Im Rahmen narrativer Interviews werden auch Fotos (Bohnsack, Kapitel 67 in diesem Band) oder kurze Filme (Akremi, Schnettler/Tuma, Kapitel 69 und 68 in diesem Band) eingesetzt, um die assoziative Wirkung und damit den Erzählfluss hinsichtlich der emotionalen Aufgeladenheit zu verstärken. Die Methoden unterscheiden sich danach, ob die Fotos von den Forschenden oder den ProbandInnen gemacht werden und wie diese in den qualitativen Erhebungen eingesetzt werden (Dirksmeier 2013). Die Begründung für den Einsatz visueller Reize liegt in der Tatsache, dass die visuelle Wahrnehmung in der Evolutionsgeschichte des Menschen früher als die sprachliche entwickelt wurde. Ein Ansprechen dieser tiefer liegenden Assoziationen – so die Annahme – hat eine stärkere Wirkung auf Narrationen als die „kopfgesteuerte“ Sprache (Harper 2002).

Ein Sonderfall der „Photo Elicitation“ besteht darin, den Befragten Aufgaben zu erteilen, zu welchen sie eine bestimmte Zahl von Fotos erstellen und daraus begründete Auswahlen treffen sollen (beispielsweise zur „typischen“ Darstellung des Quartiers). Auch hier werden über Interviews Begründungen für die gewählten Blickwinkel, Ausschnitte und Zeitpunkte erhoben, die zu Narrationen über den Raum führen).

Weitere raumbezogene Narrationen können über Spaziergänge durch den Raum erzeugt werden, die dann im Gehen mittels Tonbandaufzeichnungen dokumentiert werden. Hierbei wird nach besonderes markanten, auffallenden, störenden oder identitätsstiftenden

Orten gefragt und über Details die Syntheseleistung des Ortes analysiert. Die Erhebungsmethode ist die gleiche, wie beim oben kritisierten Erstellen emotionaler Landkarten (hier werden in der Regel aber tracking-Verfahren eingesetzt, um die jeweiligen Orte objektiv zu bestimmen).

Hinsichtlich der Auswertung werden die Unterschiede jedoch deutlich: Bei narrativen Interviews geht es um das Verstehen bestimmter Eindrücke sozialräumlicher Phänomene aus der Sicht einzelner Personen, während es bei der Kartierung um „Objektivierungen über eine große Zahl“ geht, ohne auch nur einen Hinweis über sozialstrukturelle Differenzen oder gar von Empfindungen und Narrationen zu erheben.

## Literatur

- Bourdieu, Pierre (1991): Physischer, sozialer und angeeigneter physischer Raum. In: Wentz, Martin (Hg.): Stadt-Räume. Frankfurt a.M./New York: Campus. 25-34
- Dangschat, Jens S. (2007): Raumkonzept zwischen struktureller Produktion und individueller Konstruktion. In: Ethnoscripts 9: 24-44
- Dirksmeier, Peter (2013): Zur Methodologie qualitativer visueller Methoden – Die Beispiele der Autofotografie und reflexiven Fotografie. In: Rothfuß, Eberhard/Dörfler, Thomas (Hg.): Raumbezogenen qualitative Sozialforschung. Wiesbaden: VS Verlag. 83-101
- Dörfler, Thomas (2013a): Milieu und Raum – Zur relationalen Konzeptionalisierung eines sozio-räumlichen Zusammenhangs. In: Rothfuß, Eberhard/Dörfler, Thomas (Hg.): Raumbezogenen qualitative Sozialforschung. Wiesbaden: VS Verlag. 33-59
- Dörfler, Thomas (2013b): Die Praxis der relationalen Milieuforschung. In: Rothfuß, Eberhard/Dörfler, Thomas (Hg.): Raumbezogenen qualitative Sozialforschung. Wiesbaden: VS Verlag. 245-266
- Geertz, Clifford (1983): Dichte Beschreibung. Frankfurt a.M.: Suhrkamp
- Healey, Patsy (2010): Making Better Places. Hounds mills/London: Palgrave Macmillan
- Läpple, Dieter (1992): Essay über den Raum. In: Häußermann, Hartmut/Ipsen, Detlev/Krämer-Badoni, Thomas/Läpple, Dieter/Rodenstein, Marianne (Hg.): Stadt und Raum. Pfaffenweiler: Centaurus. 157-207
- Lefèvre, Henri (1974): La production de l'espace. Paris: Editions Anthropos
- Löw, Martina (2001): Raumsoziologie. Frankfurt a.M.: Suhrkamp
- Lynch, Kevin (1960): The Image of the City. Boston: MIT Press
- Rothfuß, Eberhard/Dörfler, Thomas (2013): Prolog – Raumbezogenen Qualitative Sozialforschung. In: Rothfuß, Eberhard/Dörfler, Thomas (Hg.): Raumbezogenen qualitative Sozialforschung. Wiesbaden: VS Verlag. 7-31
- Schütz, Alfred (1971): Gesammelte Aufsätze 1. Den Haag: Martinus Nijhoff

Tobia Lakes

---

## 77.1 Was sind Geodaten?

Schätzungsweise 80% aller weltweiten qualitativen und quantitativen Daten weisen einen Raumbezug auf. Die zunehmende Verfügbarkeit von regionalisierten Datensätzen einerseits und von frei verfügbaren computergestützten Tools und GPS-gestützten Kommunikationsinstrumenten andererseits, lassen im Kontext von Diskussionen um den Zensus (Hartmann/ Lengerer, Kapitel 71 in diesem Band) und den Datenschutz (Mühlichen, Kapitel 4 in diesem Band) ein gesteigertes Interesse am Raumbezug in soziologischen Studien verzeichnen.

Als *Geodaten* bezeichnet man die formale Beschreibung von Geoinformationen zur computergerechten Verarbeitung. *Geoinformationen* werden dabei als „Informationen über geographische Phänomene, die direkt oder indirekt mit einer auf die Erde bezogenen Position verbunden sind“ definiert (DIN ISO 19101). Diese *geographischen Phänomene* oder *Geoobjekte*, wie zum Beispiel „die Stadt Berlin“, weisen also nicht nur die gemeinhin verwendeten thematischen und zeitlichen Merkmale auf, sondern darüber hinaus geometrische Merkmale (d.h. eine eindeutige Verortung, z.B. über XY-Koordinaten) und topologische Merkmale (d.h. räumliche Beziehungen zu anderen Objekten, z.B. Nachbarschaftsbeziehungen) auf.

### 77.1.1 Georeferenzierung von Daten

Um die geometrischen und topologischen Merkmale von Geoobjekten zu beschreiben, bedarf es der *Georeferenzierung*, also der eindeutigen Verortung. Diese kann über zwei Ansätze erfolgen:

1. über die *indirekte Georeferenzierung* z.B. über Postleitzahlen oder administrative Einheiten. Solche Raumeinheiten werden häufig verwendet, es müssen jedoch spezifische

Probleme bei der Nutzung berücksichtigt werden. So werden beispielsweise Raum-einheiten wie Postleitzahlbezirke über die Jahre verändert. Darüber hinaus wurde ihre Abgrenzung mit dem Ziel der Postzustellung vorgenommen, so dass die Bezirke beispielsweise aus soziologischer Sicht sehr heterogen sein können. Die in Abb. 77.1 dargestellten Daten des Monitorings für soziale Stadtentwicklung (Sozialer Entwick-lungsindex) werden beispielsweise auf der Ebene der Planungsräume zusammenge-fasst. Ziel dieser in Berlin neu eingeführten Abgrenzung von Raumeinheiten ist „die Abbildung lebensweltlicher Homogenität bei gleichzeitiger Wahrung einer Vergleich-barkeit der Planungsraumeinheiten“ ([http://www.stadtentwicklung.berlin.de/planen/basisdaten\\_stadtentwicklung/lor/](http://www.stadtentwicklung.berlin.de/planen/basisdaten_stadtentwicklung/lor/)).

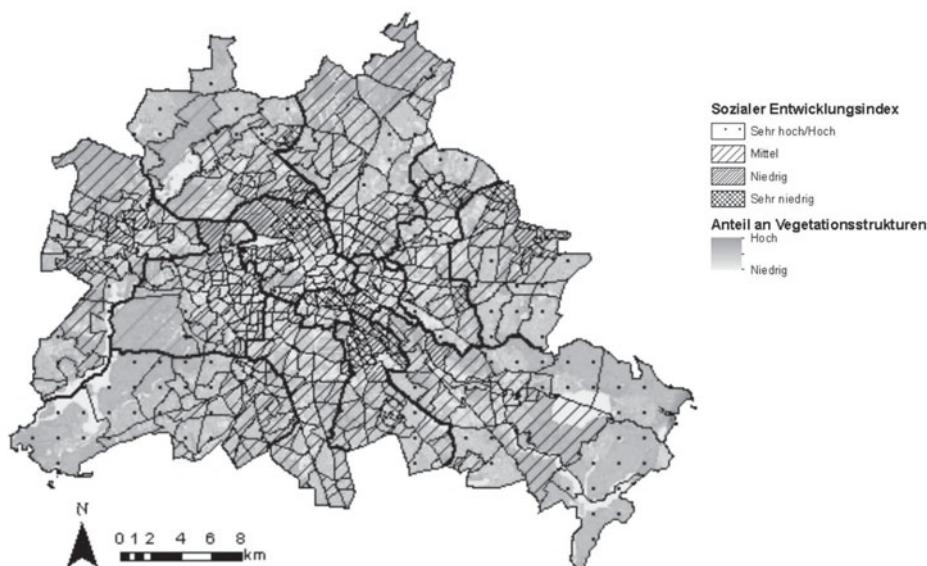
2. über die *direkte Georeferenzierung* über Koordinaten, z.B. die exakte zwei- oder dreidimensionale Verortung über x, y und z- Koordinaten im Raum. Die zunehmende Ver-fügbarkeit von Positionierungssystemen, wie zum Beispiel GPS in Smartphones oder Kameras, führt zu einer *vermehrten Erfassung von direkt georeferenzierten Geodaten*.

*Sekundärdaten*, wie zum Beispiel Umfragedaten (Reinecke, Kapitel 44 in diesem Band) oder Daten der amtlichen Statistik (Hartmann/Lengerer, Kapitel 71 in diesem Band) lassen sich *georeferenzieren*. Ein bei Haushaltsanalysen weit verbreiteter Ansatz ist die *Adress-kodierung*, bei der der Adresse eines Haushalts die entsprechenden Raumbezugs-Koordinaten zugewiesen werden und damit ein digitaler Geodatensatz zur Verfügung steht (Hintze/Lakes 2009). Online verfügbare Tools nutzen beispielsweise Adressdatenbanken von GoogleMaps, wie z.B. Batchgeo (<http://www.batchgeo.com>).

### 77.1.2 Vektordaten und Rasterdaten

Geodaten lassen sich grundsätzlich in zwei unterschiedlichen Datenmodellen beschrei-ben, den Datenmodellen Vektor- und Rasterdaten genannt (Bill 2010):

- *Vektordaten* beschreiben beispielsweise ein Gebäude über die Elemente „Punkt“, „Linie“ und „Polygon“, wie sie in gängiger Grafiksoftware verwendet werden. So sind in Abbildung 1 die Entwicklungsindizes der Stadt Berlin im Format eines Polygon-Vektordatensatzes angegeben. Die Daten werden in vier Klassen gruppiert für die räumliche Ebene der Planungsräume dargestellt.
- *Rasterdaten* hingegen beschreiben ein Gebäude über gleichgroße Rasterzellen, die aus der digitalen Photographie und Geofernerkundung bekannt sind. Jede Rasterzelle weist einen spezifischen Wert auf, der die Landoberfläche beschreibt. So wurden in Abb. 77.1 beispielsweise die Vegetationsstrukturen aus satellitengestützten Fernerkundungssen-soren abgeleitet. Dies ermöglicht eine flächenhafte Darstellung der Vegetationsstruk-turen mit einer durch den Sensor definierten räumlichen Auflösung der Pixelgröße.



**Abb. 77.1** Sozialräumliche Verteilung von Vegetationsstrukturen im Stadtgebiet Berlins (Datenquellen: Monitoring für soziale Stadtentwicklung, Senatsverwaltung für Gesundheit und Soziales; Geographisches Institut, Humboldt-Universität zu Berlin)

### 77.1.3 Geobasisdaten und Geofachdaten

Anhand ihres Beschreibungsgegenstands lassen sich Geodaten in Geobasisdaten und Geofachdaten unterscheiden, die von verschiedenen administrativen und privaten Stellen erfasst und bereitgestellt werden.

*Geobasisdaten* enthalten allgemeine topographische sowie Eigentumsinformationen und bieten damit die Grundlage für viele Studien. Die entsprechenden Datensätze werden über das Geodatenzentrum des Bundesamtes für Kartographie und Geodäsie (BKG) zentral bereitgestellt (<http://www.geodatenzentrum.de>). Die beiden wichtigsten, dort verfügbaren deutschlandweit einheitlichen Geobasisdatensätze sind:

- *ATKIS (Amtliches Topographisch-Kartographischen Informationssystem)* enthält digitale Landschaftsmodelle, Geländemodelle, topographische Karten, Orthophotos, Straßennamen, geografische Namen und Verwaltungsgrenzen (<http://www.bkg.bund.de>).
- *ALKIS (Automatisiertes Liegenschaftskataster Informationssystem)* umfasst die Liegenschaftskarte, das Liegenschaftsbuch und die Hauskoordinaten (<http://www.bkg.bund.de>).

Neben Basisinformationen sind für die soziologische Forschung vorwiegend thematische *Geofachdaten* von Interesse, wie z.B. Umweltdaten, Daten über sozialversicherungspflichtige Beschäftigte oder Unternehmensdaten (Liebig et al., Kapitel 75 in diesem Band). Auf der administrativen Seite werden sie in der Bundesrepublik Deutschland in der föderalen und sektoralen Strukturteilung erfasst und verwaltet. Internationale Daten liegen beispielsweise im Socioeconomic Data and Application Center (SEDAC) in den USA (<http://sedac.ciesin.columbia.edu>).

Darüber hinaus werden administrativ erhobene Geobasis- wie auch Geofachdaten zunehmend durch neue Ansätze ergänzt. Hier ist insbesondere das *OpenStreetMap-Projekt* (OSM) zu nennen (<http://www.openstreetmap.de/>), ein Beispiel für die internetbasierte Erfassung, Verwaltung und Bereitstellung von Informationen im Sinne von Wikipedia. Darüber hinaus ist der Einfluss kommerzieller Geodatenanbieter, wie zum Beispiel im Geomarketing, stark gewachsen.

Über Geodateninfrastrukturen und Metadaten, also Daten über Geodaten, werden Vielfalt und Menge an Geodaten für verschiedene Nutzer zugänglich gemacht, wie zum Beispiel in der *Geodateninfrastruktur GDI-DE* des Bundes, der Länder und der Kommunen ([www.geoportal.de](http://www.geoportal.de)).

---

## 77.2 Analysemethoden

Die Möglichkeiten, die Geodaten und entsprechende räumliche Analysen für bislang weitgehend „nicht räumliche“ Wissenschaften und Anwendungsbereiche liefern, lassen sich in sechs Bereiche kategorisieren:

1. Geodaten beinhalten thematische Daten wie jeder andere Datensatz. Darüber hinaus bieten sie *geometrische Merkmale*, die von zusätzlichem Interesse sein können. So lassen sich die Einwohnerzahlen auf Ebene von statistischen Blöcken mit klassischen statistischen Verfahren analysieren. Geometrische Eigenschaften erlauben dann die Beantwortung folgender Fragen: Welche Flächengröße weisen die entsprechenden Blöcke auf? Wie groß ist die Einwohnerdichte der Blöcke? Hierzu stehen in jeder GIS-Software Funktionalitäten zur Berechnung der entsprechenden Algorithmen zur Verfügung (Bill 2010; de Lange 2005).
2. Geodaten können über ihre *topologischen Merkmale* Analysen der Nähe, der Überlappung, der räumlichen Distanz, Nachbarschaft etc. zwischen verschiedenen Objekten erlauben (Bill, 2010; de Lange 2005). Topologische Analysen können sich dabei auf gleiche Objekte beziehen, wie z.B. die Berechnung der räumlichen Distanz zwischen Hochschulen oder auf unterschiedliche Objekte, wie z.B. die Berechnung der räumlichen Distanz von einem Haushalt zur nächsten Bildungseinrichtung (Spiess/Wrohlich 2008).

3. Geodaten bieten den Vorteil, *verschiedene Datensätze über die räumliche Lage zu integrieren* (Cielebak/Rässler, Kapitel 26 in diesem Band) und mögliche Zusammenhänge zwischen Datensätzen zu analysieren. So wird beispielsweise die Frage nach der sozialräumlich gerechten Verteilung von Umweltressourcen und -belastungen über die integrative Auswertung von sozioökonomischen Daten mit umweltrelevanten Daten beantwortet (Abbildung 1; Lakes et al. 2013; Crowder/Downey 2010). In Abb. 77.1 für Berlin wird beispielsweise deutlich, dass Planungsräume mit einem hohen bzw. sehr hohen sozialen Entwicklungsindex insbesondere im Randbereich der Stadt liegen, in dem gleichzeitig ein hoher Anteil an Vegetationsstrukturen existiert. Im Innenbereich Berlin hingegen, wie zum Beispiel im Norden Neuköllns, überlagern sich geringe und sehr geringe Werte des sozialen Entwicklungsindex mit geringen Anteilen an Vegetationsstrukturen. Weiterführende, detailliertere Analysen ermöglichen beispielsweise die Identifikation von Gebäuden, die einen niedrigen Sozialstatus und eine hohe Umweltbelastung oder einen geringen Vegetationsanteil im direkten Wohnumfeld aufweisen.
4. Deskriptive wie inferentielle statistische Analysen erlauben das *Erkennen von Mustern, Anomalien, Ausreißern* etc. Methoden der explorativen räumlichen Statistik erlauben bspw. die Analyse von Hotspots sozialer Indikatoren in Panelanalysen (Anselin et al. 2007). So lassen sich beispielsweise, basierend auf der kartographischen Darstellung in Abb. 77.1, räumliche Cluster von Planungsräumen in Berlin berechnen, in denen sich geringe und sehr geringe Werte des sozialen Entwicklungsindex konzentrieren.
5. Geodaten bieten den grundsätzlichen Vorteil der *kartographischen Visualisierung von Informationen* (Bill 2010). Visuelle Darstellungen erlauben die intuitive Erfassung von Mustern und Verteilungen und bieten gleichzeitig die Möglichkeit Aussagen über Karten allgemeinverständlich zu kommunizieren. So lässt sich über Monitoringverfahren, wie die des Monitorings für soziale Stadtentwicklung in Berlin ([http://www.stadtentwicklung.berlin.de/planen/basisdaten\\_stadtentwicklung/monitoring](http://www.stadtentwicklung.berlin.de/planen/basisdaten_stadtentwicklung/monitoring)), die räumliche Verteilung der sozialen Entwicklungsindizes über die Zeit hinweg sehr gut nachvollziehen.

Zusammenfassend wird deutlich, dass eine große Vielzahl an Geodaten und Anwendungsmöglichkeiten für soziologische Analysen und Studien zur Verfügung steht, die es nun in Wert zu setzen gilt.

---

### 77.3 Software

Geodaten können wie alle anderen Daten mithilfe statistischer Software analysiert werden, R ([www.r-project.org](http://www.r-project.org)), SPSS ([www.ibm.com](http://www.ibm.com)) oder STATA ([www.stata.com](http://www.stata.com)). Um den zusätzlichen Mehrwert aus der geometrischen und topologischen Dimension der Geodaten zu nutzen, bedarf es spezifischer Techniken der *Geoinformatik* (Bill 2010, de Lange 2005). Die Geoinformatik befasst sich explizit mit der Entwicklung und Anwendung von Methoden der Informatik zur Lösung geographische Fragestellungen unter besonderer

Berücksichtigung des Raumbezugs von Geoinformationen (Bill 2010, de Lange 2005). Die Software *GIS* (*Geographisches Informationssystem*) ist ein EDV-System für die Erfassung, Verwaltung, Analyse und Darstellung von Geodaten, das sowohl proprietäre Produkte, z.B. ArcGIS ([www.esri.de](http://www.esri.de)), als auch OpenSource-Produkte, z.B. Quantum GIS ([www.qgis.org](http://www.qgis.org)) oder GRASS ([grass.osgeo.org](http://grass.osgeo.org)), umfasst. Darüber hinaus existieren zunehmend räumliche Erweiterungen für weit verbreitete Datenbank-Systeme, z.B. Oracle Spatial ([www.oracle.com](http://www.oracle.com)), PostgreSQL ([www.postgresql.org](http://www.postgresql.org), [postgis.net](http://postgis.net)). Von besonderem Interesse sind Web-Services, die grundlegende räumliche Dienste über den Browser im Internet anbieten, ohne die Notwendigkeit einer installierten GIS-Software auf dem PC des Benutzers (OGC, Open Geospatial Consortium, [www.opengeospatial.org](http://www.opengeospatial.org)). Darüber hinaus sind frei verfügbare Internet-Tools mit einfacher räumlicher Funktionalität in weiter Verbreitung, z.B. GoogleEarth ([maps.earth.de](http://maps.earth.de)), GoogleMaps ([maps.google.de](http://maps.google.de)).

## Literatur

- Anselin, Luc/Sridharan Sanjeev/Gholston Susan (2007): Using Exploratory Spatial Data Analysis to Leverage Social Indicator Databases: The Discovery of Interesting Patterns. In: *Social Indicators Research* 82: 287-309
- Bill, Ralf (2010): Grundlagen der Geo-Informationssysteme. 5. Auflage. Berlin/Offenbach: Herbert Wichmann Verlag
- Crowder, Kyle D./Downey, Liam (2010): Interneighborhood Migration, Race, and Environmental Hazards: Modeling Microlevel Processes of Environmental Inequality. In: *American Journal of Sociology* 115 (4): 1110-1149
- De Lange, Norbert (2005): Geoinformatik in Theorie und Praxis. 2. Auflage. Berlin: Springer
- Hintze, Peter/Lakes, Tobia (2009): Geographically Referenced Data in Social Science. A Service Paper for SOEP Data Users. Data Documentation 46. DIW Berlin
- Lakes, Tobia/Brückner, Maria/Krämer, Alexander (2013). Development of an environmental justice index to determine socio-economic disparities of noise pollution and green space in residential areas in Berlin. *Journal for Environmental Planning and Management* 57 (4). DOI: 10.1080/09640568.2012.75546
- Spiess, C. Katharina/Wrohlich, Katharina (2008): Does Distance Determine Who Attends a University in Germany? SOEPpaper 118. DIW Berlin

Monika Pritzel und Hans J. Markowitsch

---

## 78.1 Was sind neurowissenschaftliche Methoden?

Neurowissenschaftliche Methoden (Pritzel et al. 2009; Pritzel 2006) versuchen, mittels noninvasiver Verfahren und durch eine möglichst engmaschige räumlich-zeitliche Auflösung bestimmte als relevant angesehene Veränderungen im Gehirn in Relation zum zeitlichen Verlauf bestimmter Verhaltensweisen (z.B. zwischen Reizbeginn und Antwort) abzubilden. Dadurch sollen Aussagen möglich werden, wie das Gehirn Reize verarbeitet, Verhalten vorbereitet und steuert, Informationen einspeichert und abruft, Emotionen, Motivationen und Bewusstsein erzeugt.

### 78.1.1 Morphologisch-histologische Erfassung der Ansammlung von Nervenzellkörpern

Unter den Sammelbegriff der neurowissenschaftlichen Methoden (Pritzel et al. 2009; Pritzel 2006) fällt klassischerweise zunächst die statisch zu nennende morphologisch-histologische Erfassung der Ansammlung von Nervenzellkörpern und deren Fortleitungen, den Axonen, sowie von deren informationsaufnehmenden Teilen, den Synapsen. Dazu gehören auch die Ermittlung von Typen von Überträgersubstanzen, welche für den Datenaustausch zwischen Nervenzellen und für die Erkundung von makroskopisch, also weiträumig angelegten „Transmittersystemen“ im Gehirn sorgen. Auch die Durchführung von Läsionsstudien, welche u.a. auf der Kartographierung von Nervenzelltypen, ihrer Packungsdichte in bestimmten Hirnarealen und ihren Verbindungswegen beruht, zählt zu diesen klassischen Methoden der Neurowissenschaften. Deren Ziel ist bzw. war es, ausgewählte neuroanatomische Orte und ihre Verbindungswege mit ausgewählten Verhaltensweisen in Beziehung zu setzen, sprich, das Denken und Handeln eines Menschen auch „materialisieren“ zu können.

Auch wenn heute kaum noch eine dieser Möglichkeiten als Methode der Wahl für die Sozialwissenschaft infrage kommt, so basieren die nachfolgenden, besser für die Sozialwissenschaften geeigneten Verfahren doch auf diesen Methoden, und es besteht ein Anwendungspotenzial für die Zukunft. So erhofft man sich durch neurowissenschaftliche Methoden z.B. Antworten auf die Frage, wo und wie im Gehirn so etwas Komplexes wie Empathie zum Ausdruck gebracht werden kann (de Vignemont/Singer 2006).

### **78.1.2 Elektrophysiologische Methoden**

Elektrophysiologische Methoden kommen immer dann ins Spiel, wenn sich gedankliche Entwürfe über die funktionale Bedeutung neuronaler Verbindungen auf sogenannte Ordnungsübergänge von Funktionszuständen des Gehirns konzentrieren (Schiepek 2003), also die Frage im Raum steht, unter welchen Bedingungen ein Neuron oder ein Nervenzellverband seine Signale ändert.

Aufgrund von Aktivitätsprofilen versucht die insbesondere in der klinischen Anwendung eingesetzte Elektroenzephalographie (EEG), Echtzeit-Aussagen über elektrophysiologische Erregungszustände im Gehirn während der Durchführung eines bestimmten Verhaltens zu erlangen. So genannte ereigniskorrelierte Potentiale (ERPs) stellen eine Anwendungsvariante dar, bei der es gelingt, den Erregungszustand von ausgesuchten Nervenzellpopulationen der Großhirnrinde mit Reizbeginn und Reaktion oder Reaktionsdauer (gemessen im Bereich von Sekundenbruchteilen) des Individuums in Beziehung zu setzen und so „automatisch“ ablaufende, nicht bewusst kontrollierbare Reaktionen von solchen zu trennen, die kognitive Entscheidungen verlangen, welche ihrerseits aber nicht gleichzeitig auch bewusst gemacht werden können. In den Sozialwissenschaften hat man dadurch ein relativ einfaches, weil wenig aufwändiges und gleichzeitig unbestechliches, weil bewusst nicht manipulierbares Verfahren zur Verfügung, um eine Entscheidung innerhalb eines Zeitfensters von etwa einer halben Sekunde zu untersuchen.

### **78.1.3 Bildgebende Verfahren**

Mit bildgebenden Verfahren können dynamische Vorgänge im Gehirn in Beziehung zu bestimmten Verhaltensweisen gesetzt werden und Verhaltensänderungen, die durch eine definierte Aufgabenstellung vorgegeben werden, mit Messwerten des neuronalen Stoffwechsels oder bestimmten Blutwerten korreliert werden. Man kann sich nun z.B. fragen, was im Gehirn geschieht, wenn man sich vor bzw. nach einer Therapie ein traumatisches Erlebnis nochmals vor Augen führt. Lässt sich „Erfolg“ oder „Misserfolg“ der jeweiligen Intervention tatsächlich neuronal verorten? Gelingt es dynamischen Verfahren z.B., sich als ein zusätzliches „Evaluationsinstrument“ (Döhring, Kapitel 9 in diesem Band) auf Dauer zu etablieren, so könnte man dies als Gewinn für die Sozialwissenschaft verbuchen.

Unter dynamischen Darstellungstechniken versteht man im Unterschied zu statischen jene, die im Gehirn die Konzentration bestimmter Isotope (radioaktiv markierter Moleküle) oder den Eigendrehimpuls (den Spin) bestimmter Atome erfassen. So lassen sich Stoffwechseländerungen in Hirnstrukturen mit Verhaltensänderungen korrelieren, wodurch man letztendlich zu Aussagen über Struktur-Funktionsbeziehungen kommen kann. Die unter dem Begriff der funktionellen bildgebenden Methoden zusammengefassten Verfahren fMRT (funktionelle Kernspintomographie) und PET (Positronen-Emissions-Tomographie) nutzen z.B. gemeinsam den Umstand, dass aktivierte Neurone vermehrt Sauerstoff bzw. Glukose benötigen und so einen Hinweis auf die Inanspruchnahme bestimmter Gehirnregionen für bestimmte Verhaltensweisen geben. So ist z.B. „normalerweise“ ein Bereich im vorderen Teil des Gehirns aktiv, wenn man eine Handlung bei anderen Menschen beobachtet, die man auch selbst ausführen könnte. Man schließt daraus, dass dort die Neurone, sogenannte „Spiegelneurone“ („mirror neurons“), besonders aktiv sein müssen, welche zur Transformation von beobachtetem in selbstinduziertes Handeln beitragen (Thill et al. 2003).

## 78.2 Wann und warum sind neurowissenschaftliche Methoden für die Sozialwissenschaft geeignet?

Unter welchen Umständen können Sozialwissenschaftler Gewinn aus der Anwendung neurowissenschaftlicher Methoden ziehen, oder grundsätzlicher formuliert, auf Grund welcher beiden Disziplinen gemeinsamen Grundüberzeugungen sollte das überhaupt der Fall sein?

Anders als in den Sozialwissenschaften herrscht in der Neurowissenschaft die Meinung vor, jedes menschliche Verhalten, sei es nun „einfach“ oder „komplex“, jede Handlung, geschehe sie nun bewusst oder unbewusst, ja, schlachtweg alles, was wir je denken oder empfinden, basiere letztlich auf hirnphysiologischen Prozessen. Es kann bei der Suche nach einer für die Sozialwissenschaft geeigneten neurowissenschaftlichen Methode also nicht darum gehen, darüber Auskunft zu geben, ob oder inwieweit Teilespekte unserer Persönlichkeit Produkte physiologischer Prozesse sind bzw. inwieweit geistige von materiellen Vorgängen zu trennen sind. Es kann nur um die Frage gehen, welche der physiologischen Mechanismen, die als Korrelat eines bestimmten Verhaltens infrage kommen, am verlässlichsten messbar sind.

Die Bedeutung neurowissenschaftlicher Methoden für die Sozialwissenschaft ergibt sich somit im Rückgriff auf die gemeinsame empirische Ausrichtung im Denken beider Disziplinen, aber unter Zurückstellung ungelöster erkenntnistheoretischer Grundprobleme. Letzteres wird deutlich, indem man in der Neurowissenschaft etwa danach fragt, wie man sich mittels naturwissenschaftlicher Erkenntnisse die Entstehung eines bewusst durchdachten Handlungsplanes vorstellen kann, nicht aber danach, ob dieser Plan auch Ausdruck eines „freien Willens“ sein könnte, der sich einer naturwissenschaftlichen Messmethode eventuell entzieht. Die damit verbundene Vorstellung, nicht „wirklich“ Herr oder

Herrin des eigenen Handelns zu sein, sondern Ausführende/r einer bestimmten, eventuell auch messbaren neuronalen Netzwerkkonfiguration, gilt es zu akzeptieren, auch wenn man sich intuitiv eher einer dualistischen Sichtweise zugetan fühlen mag.

Was aber gewinnt man als Sozialwissenschaftler in *puncto* Relevanz neurowissenschaftlicher Messmethoden, wenn o.g. Leitgedanken der Neurowissenschaft folgend jede Gewissheit darüber, eine selbständig entscheidende, die Umwelt angemessen miteinbeziehende Person zu sein, ohnehin zu einer dem Überleben dienenden nützlichen Illusion zu schrumpfen scheint?

1. Prinzipiell gewinnt man dadurch zunächst die Möglichkeit, bestimmte explizite, sich in bestimmten Verhaltensweisen manifestierende und ebenso bestimmte implizite, im Zentralnervensystem nachweisbare Aktivitätsmuster miteinander in Beziehung zu setzen und auf diese Weise sozialwissenschaftlich gewonnene Ergebnisse neurowissenschaftlich zu stützen. Die o.g. These akzeptierend, dass es letztlich immer vermittels unseres Gehirns geschieht, wie wir den Blick auf uns und die Welt um uns richten, bereichert eine solche Absicherung auch diejenigen Aussagen, welche zuvor – anders als etwa solche zu „Stress“, Übermüdung etc. – nicht auf neurophysiologische Begleiterscheinungen schließen ließen. Die o.g. These akzeptierend, dass wir immer nur vermittels unseres Gehirns den Blick auf uns und die Welt um uns herum richten, kann man durch eine Einbeziehung neurowissenschaftlicher Aussagen auch nur gewinnen. Dies gilt insbesondere, da die Neurowissenschaft ihrerseits zu fast allen Teilbereichen der Sozialwissenschaft Aussagen macht (Cacioppo 2002).
2. Unser alltägliches situationsgemäßes Handeln kann man aus naturwissenschaftlicher Sicht auch als eine Fähigkeit verstehen, die sich im Laufe der Evolution als vorteilhaft herauskristallisiert hat, *weil* sie sich in ständiger Interaktion mit der Umgebung zu wandeln vermochte, ohne einen Umbruch übergeordneter Persönlichkeitscharakteristika nach sich zu ziehen. D.h. wir selbst erleben uns zwar als etwas Kontinuierliches, Unbezweifelbares und Selbstverständliches, auch wenn wir nicht in der Lage sind, uns die neuronalen Vorgänge zu vergegenwärtigen, die das erzeugen. Wir *erfahren* aber gleichwohl die Veränderungen, die vor dem Hintergrund dieser „stabilen Illusion des Selbst“ möglich sind. Diese „Erfahrbarkeit von Veränderungen“ in uns, bzw. durch oder um uns begründet ein gemeinsames Interesse von Hirnforschung und Sozialwissenschaft: Beide versuchen Grundkonstanten menschlichen Seins in empirisch fassbare Begrifflichkeiten der Wahrnehmung, der Sprache, des Gedächtnisses, der Aufmerksamkeit, der sie begleitenden Emotionen etc. aufzuschlüsseln und die Regeln des Zusammenwirkens umweltbedingter und dem Menschen innenwohnender Konstituenten mit ihren jeweiligen Methoden zu erforschen. Sozialwissenschaftlich zu untersuchen, in welchem Umfang ein Mensch auf Erfahrungen mit der Außenwelt angewiesen ist bzw. geprägt wird, um ein „angemessenes“ Verhalten zu zeigen, heißt übertragen auf die Neurowissenschaft nichts anderes als Versuchsbedingungen (Eifler, Kapitel 11 in diesem Band) zu schaffen, die plastische, also veränderungsbereite Kräfte im Gehirn zutage treten lassen. In Analogie zum weiter oben bereits Ausgeführten ist es somit letztlich die Dynamik der Gehirnorganisation, welche die Grenzen möglicher Verhaltensänderungen bestimmt. Ein Wissen um die Aussagekraft derjenigen

neurowissenschaftlichen Methoden, die darüber Aufschluss geben könnten, inwieweit und wie überdauernd das zentrale Nervensystem von seinem Aufbau und seiner Funktionsweise durch Erfahrungen veränderbar ist, ist deshalb auch für Sozialwissenschaftler sinnvoll: Es erhöht die Kenntnis um die Spezifität eines aufeinander bezogenen veränderungsrelevanten Einflusses von Gehirn- und Verhaltensvariablen und damit auch die Vorhersagemöglichkeit im Bereich der Sozialwissenschaft.

---

## 78.3 Anwendungsbeispiele

### 78.3.1 Neurowissenschaftliche Methoden und Biographieforschung

Brand und Mitarbeiter untersuchten 2009 die Gehirne von 14 Patienten, die auf Grund von Umwelteinflüssen amnestisch geworden waren, mittels funktioneller Hirnbildgebung. Alle 14 Patienten konnten sich nicht mehr an ihre persönliche Vergangenheit erinnern, wussten aber noch allgemeine Fakten („Weltwissen“: z.B. wer Harald Juhnke oder wer Goethe war) und konnten weiterhin lesen, schreiben, rechnen etc. Da sie für ihre Biographie (Rosenthal, Kapitel 35 in diesem Band) amnestisch waren, erinnerten sie auch nicht ihre frühere Identität und damit auch nicht, wie sie sich beispielsweise ihren Lebenspartnern oder -partnerinnen gegenüber benommen hatten. Alle 14 Patienten und meist auch nahe Verwandte oder Bekannte von ihnen wurden zuvor zur Ermittlung der Krankengeschicht-eintensiv befragt und neuropsychologisch umfassend getestet. Somit war der Auslöser der Amnesie bekannt, aber auch die bei allen Patienten von Stresserlebnissen (meist schon in der Kindheit) gezeichnete Lebensgeschichte. Durch neurowissenschaftliche Methoden (hier u.a. Untersuchungen zum Glukose-Stoffwechsel) ließ sich also nachweisen, dass psychisch (und damit durch die Umwelt) induzierte Einflüsse sich im Gehirn niederschlagen und somit die Hirnveränderungen den Nachweis für die Amnesie bilden. Dies bedeutet, dass die Patienten nicht gelogen haben, sondern tatsächlich an einer psychischen Krankheit litten. Außerdem ließ sich nachweisen, dass auch psychiatrische Krankheitsbilder („dissoziative Störungen“) eine neurale Grundlage haben.

### 78.3.2 Neurowissenschaftliche Methoden und Beobachtung

In den letzten Jahren wurden in den Neurowissenschaften eine Reihe von Spielparadigmen entwickelt, wie das „ultimatum game“ und das „dictator game“. Darüber hinaus gibt es Erkenntnisse, dass die Verabreichung eines Bindungshormons (z.B. Oxytocin, nasal gesprührt) Menschen vertrauensseliger Geschäfte abschließen lässt (und auch solche, die eher zu ihren Ungunsten sind). In Kombination mit der Beobachtung sozialer Verhaltensweisen (Thierbach/Petschick, Kapitel 66 in diesem Band) erlaubt dies die Aufdeckung der Bedeutung von Hormonen für Sozialverhalten und die Aufdeckung der diesem Verhalten zugrunde liegenden Hirnaktivitäten und Hirnaktivitätsveränderungen über die Zeit und in Abhängigkeit von der aktuellen sozialen Interaktion.

## Literatur

- Brand, Matthias/Eggers, Carsten/Reinhold, Nadine/Fujiwara, Esther/Kessler, Josef/Heiss, Wolf-Dieter/Markowitsch, Hans J. (2009). Functional Brain Imaging in Fourteen Patients with Dissociative Amnesia Reveals Right Inferolateral Prefrontal Hypometabolism. In: Psychiatry Research: Neuroimaging Sections 174: 32-39
- Cacioppo, John T. (Hg.) (2002): Foundations in Social Neuroscience. Cambridge: MIT Press
- de Vignemont, Frederique/Singer, Tanja (2006): The Empathic Brain. In: Trends in Cognitive Science 10: 435-441
- Pritzel, Monika (2006). Biopsychologische Grundlagen und Methoden. In: Pawlik, Kurt (Hg.): Handbuch der Psychologie. Berlin: Springer: 37-60
- Pritzel, Monika/Brand, Matthias/Markowitsch, Hans J. (2009). Gehirn und Verhalten. Heidelberg: Spektrum
- Schiepek, Günter (Hg.) (2003): Neurobiologie der Psychotherapie. Stuttgart: Schattauer
- Thill, Serge/Caligiore, Daniele/Borghi, Anna M./Ziemke, Tom/Baldassarre, Gianlucca (2013): Theories and Computational Models of Affordance and Mirror Systems. In: Neuroscience and Biobehavioral Reviews 37: 491-521

---

## **Teil 7**

# **Datenauswertung in der quantitativen Sozialforschung**

Jörg Blasius und Nina Baur

---

## 79.1 Statistik und quantitative Sozialforschung

### 79.1.1 Datenerhebung und Auswertung

In der quantitativen Sozialforschung (Stein, Kapitel 7 in diesem Band) liegen *Daten* entweder bereits in standardisierter Form vor – z.B. in Form von Daten der amtlichen Statistik (Hartmann/Lengerer, Kapitel 71 in diesem Band), als Logfiles (Schmitz/Yanenko, Kapitel 65 in diesem Band) oder Kundendatenbanken (Liebig/Meyermann/Gebel, Kapitel 75 in diesem Band) –, oder sie werden im Rahmen von Primärerhebungen – z.B. in Form von Befragungen (Reinecke, Kapitel 44 in diesem Band) – standardisiert erhoben, oder schwach strukturierte Daten werden z.B. mittels inhaltsanalytischer Verfahren analysiert und kodiert (Mayring/Fenzl, Kuckartz/Rädiker, Kapitel 38 und 27 in diesem Band).

Anders als in der qualitativen Sozialforschung werden diese Daten in der quantitativen Sozialforschung auf der Basis einer Vielzahl von Fällen erhoben – und zwar nach Möglichkeit als Zufallsstichprobe (Häder/Häder, Kapitel 18 in diesem Band) mit möglichst wenigen Ausfällen (Engel/Schmidt, Kapitel 23 in diesem Band). So ist beim ARD-DeutschlandTREND für standardisierte Befragungen der deutschen Bevölkerung eine Fallzahl von  $n = 1.000$  Befragten der typische Standard (<http://www.infratest-dimap.de/umfragen-analysen/bundesweit/ard-deutschlandtrend/>).

Nach der Datenerhebung liegt somit typischerweise über eine Vielzahl von *Fällen* (z.B. Personen, Organisationen, Nationen) eine Vielzahl von Informationen in Form von *Variablen* (z.B. Fragen in einem Fragebogen) vor. Für jede Variable ist jedem Befragten eine spezifische *Ausprägung* (z.B. eine Antwort im Fragebogen) zugeordnet (Baur/Lamnek 2007). Man weiß dann z.B., dass Befragter Nr. 1 bei der Variable „Geschlecht“ die Ausprägung „männlich“, bei der Variable „Alter“ die Ausprägung „39 Jahre“ und bei der „letzten Bundestagswahl“ die „CDU“ gewählt hat, während Befragte Nr. 2 bei der Variable „Geschlecht“ die Ausprägung „weiblich“, bei der Variable „Alter“ die Ausprägung „27 Jahre“ und bei der „letzten Bundestagswahl“ die „Linke“ gewählt hat. Im Rahmen der Datenaufbereitung

(Akremi et al. 2011: 22-144, Lück/Landrock, Kapitel 28 in diesem Band) werden diese verbalen Informationen in maschinenlesbare Zahlen umgewandelt und in Form von Datensätzen gespeichert, d.h. im *Datensatz* würde statt „weiblich“ z.B. „0“ stehen und statt „männlich“ die „1“. Welche inhaltlichen Informationen in welche Zahlen umzuwandeln sind, wird im *Codeplan* bzw. *Codebuch* festgehalten.

### 79.1.2 Deskriptive und induktive Statistik

Der Datensatz ist eine für den Forscher unübersichtliche Tabelle aus Zahlen, aus der keine inhaltlichen Schlüsse gezogen werden können, da schon angesichts der reinen Fülle an Daten keine Systematiken zwischen den Befragten bzw. zwischen den Variablen erkannt werden können. Mit Hilfe der *deskriptiven Statistik* (oder *beschreibenden Statistik*) können die Informationen im Datensatz so komprimiert werden, dass man mit ihnen schon erste Forschungsfragen beantworten kann. So werden z.B. die Informationen im Datensatz zu einer *Tabelle*, einer *Grafik* oder zu aussagekräftigen *Maßzahlen* verdichtet.

Mit Hilfe der *induktiven Statistik* (oder *schließenden Statistik*) (Dubben/Beck-Bornholdt 2011), die eine Zufallsstichprobe (Häder/Häder, Kapitel 18 in diesem Band) voraussetzt, wird untersucht, inwiefern sich die Ergebnisse der Stichprobe auf die Grundgesamtheit verallgemeinern lassen, wobei meistens Hypothesen getestet (*statistische Tests*) werden. So kann z.B. untersucht werden, ob bei der letzten Bundestagswahl signifikant mehr Frauen als Männer die CDU gewählt haben.

### 79.1.3 Univariate, bivariate und multivariate Statistik

Da eine sehr große Bandbreite statistischer Verfahren existiert, gibt es verschiedene Systematisierungsversuche. Klassisch ist etwa die Differenzierung nach der Zahl der Variablen, die in die Analyse einbezogen werden:

- Die *univariate Statistik* befasst sich mit *einzelnen Variablen*, die in Form von Häufigkeitsverteilungen in Tabellen oder Grafiken dargestellt werden können und für die sogenannte *Lagemaße* (dazugehören der Median und der Mittelwert) und *Streuungsmaße* (dazu gehören Varianz und Standardabweichung) berechnet werden. So kann z.B. auf Basis einer univariaten Analyse festgestellt werden, dass bei der Bundestagswahl 2013 die CDU/CSU 41,5% der Zweitstimmen bekam, die SPD 25,7%, die Linke 8,6% und die Grünen 8,4%, während alle anderen Parteien an der 5%-Hürde scheiterten, darunter auch die FDP, die mit einem Anteil von 4,8% erstmals seit 1949 nicht in den Bundestag gewählt wurde ([http://www.forschungsgruppe.de/Aktuelles/NewsL\\_BTW\\_2013.pdf](http://www.forschungsgruppe.de/Aktuelles/NewsL_BTW_2013.pdf), Zugriff vom 22.1.2014).

- Bei den *bivariaten Verfahren* werden immer *genau zwei Variablen* in Abhängigkeit zueinander gesetzt, z.B. die Noten in zwei Schulfächern oder die geschlechtsspezifische Verteilung der Zweitstimmen bei der Bundestagswahl 2013. Während im ersten Fall nicht gesagt werden kann, welche Variable kausal abhängig und welche unabhängig ist (Kühnel/Dingelstedt, Kapitel 80 in diesem Band), so ist im zweiten Fall die Parteipräferenz abhängig vom Geschlecht. Wäre es andersherum, müsste behauptet werden, dass die gewählte Partei beeinflusst, welches Geschlecht jemand hat, was ersichtlich nicht sein kann. Wie bei der univariaten Statistik können die Ergebnisse grafisch bzw. numerisch darstellt werden. So haben bei der Bundestagswahl 2013 z.B. 44% der Frauen, aber nur 39% der Männer die CDU/CSU gewählt, d.h. die Wahl war geschlechtsspezifisch ([http://www.forschungsgruppe.de/Aktuelles/News1\\_BTW\\_2013.pdf](http://www.forschungsgruppe.de/Aktuelles/News1_BTW_2013.pdf), Zugriff vom 22.1.2014).
- Da in den Sozialwissenschaften meistens viele Faktoren zusammenwirken, beinhalten die meisten Forschungsfragen wesentlich *mehr als zwei Variablen*. So ist die Parteipräferenz nicht nur vom Geschlecht abhängig, sondern z.B. auch von der Bildung, von einer Vielzahl von Einstellungen gegenüber politischen und aktuellen Themen sowie von der Einschätzung der wichtigsten Politiker der Parteien. Sollen Fragen wie: „Warum wurde die CDU/CSU von wem gewählt?“ beantwortet werden, wird die *multivariate Statistik* benötigt, also statistische Verfahren, mit denen *mehr als zwei Variablen* auf einmal verarbeitet werden. Einführungen in die verschiedenen Verfahren werden etwa im Sammelband von Wolf und Best (2010) gegeben. Im Unterschied zu univariaten und bivariaten Analysen sind mit multivariaten Auswertungen wesentlich weitergehende Interpretationen von sozialwissenschaftlichen Zusammenhängen möglich. Werden z.B. nur die zwei Variablen Einkommen und Geschlecht betrachtet, so gibt es sehr deutliche Einkommensunterschiede zwischen Männern und Frauen, wobei letztere deutlich weniger verdienen. Werden entsprechende Ergebnisse publiziert, so kann nur von einer massiven Benachteiligung der Frauen gesprochen werden. Wird als dritte Variable die wöchentliche Arbeitszeit berücksichtigt – die meisten Halbtagsstellen sind in Deutschland von Frauen besetzt – so ist die Einkommensdifferenz zwischen den Geschlechtern bereits deutlich geringer. Eine weitere sehr wichtige Variable zur Interpretation der Einkommensungleichheit ist der Beruf – Frauen arbeiten wesentlich häufiger in intrinsisch motivierten und schlecht bezahlten Berufen (z.B. als Arzthelferin, als Krankenschwester oder als Erzieherin) als Männer, die meistens extrinsisch motivierte und gut bezahlte Berufe bevorzugen (z.B. als Investmentbanker, Chefarzt eines großen Krankenhauses oder als Professor). Eine weitere Variable, mit der ebenfalls Einkommensunterschiede zwischen den Geschlechtern erklärt werden können, ist die Dauer der Zugehörigkeit zum Betrieb – Elternzeiten werden immer noch überwiegend von Frauen genommen, was zwangsläufig zu kürzeren Zeiten bei der Betriebszugehörigkeit sowie bei der beruflichen Erfahrung führt. Je nach theoretischem Modell kann die Liste der Variablen nahezu beliebig erweitert werden.

### 79.1.4 Messniveau der Variablen

Neben der Zahl der zu verarbeitenden Variablen erfolgt eine weitere wichtige Unterscheidung nach dem sogenannten *Messniveau der Variablen*, also der Ordnungsfähigkeit der Ausprägungen der Variablen, wobei in den Sozialwissenschaften klassischerweise drei Arten unterschieden werden:

- Bei *nominalen* (oder *ungeordnet kategorialen*) *Variablen* handelt es sich um eine reine qualitative Zuordnung von Merkmalsausprägungen zu einer Variablen, wie dies z.B. bei den Indikatoren „Parteipräferenz“ oder „Familienstand“ der Fall ist. Typisch für diese Merkmale ist, dass es kein „mehr“ und kein „weniger“ gibt, sondern nur eine eindeutige Zuordnung und einen Vergleich der Häufigkeiten. So kann lediglich qualitativ zwischen „ledig“, „verheiratet“ und „verwitwet“ unterschieden werden, und es kann gesagt werden, von welcher Gruppe (Ledige, Verheiratete, ...) es wie viele Personen in der Stichprobe gibt, aber mehr auch nicht. Eine Person kann auch nicht zur Hälfte verheiratet und zur Hälfte geschieden sein, bei Bedarf ist es aber gegebenenfalls möglich, eine weitere qualitative Kategorie einzufügen, z.B. „getrennt lebend“. Hat eine nominale Variable nur zwei Ausprägungen (z.B. „männlich“ und „weiblich“ beim „Geschlecht“), wird sie auch als *dichotome, binäre* oder *Dummy-Variable* bezeichnet.
- Bei *ordinalen* (*geordnet kategorialen*) *Variablen* gibt es ein „mehr“ und ein „weniger“, anhand dessen die Ausprägungen verglichen werden können, wobei die Abstände zwischen einander nachfolgenden Kategorien zwar meistens numerisch, aber nicht inhaltlich (sozialwissenschaftlich) identisch sind. Ein typisches Beispiel hierfür sind Schulnoten, der Abstand zwischen „1“ und „2“ ist zwar rein numerisch identisch mit jenen zwischen „4“ und „5“, inhaltlich ist er aber vollkommen anders.
- Ein *metrisches (kontinuierliches)* *Messniveau* setzt zusätzlich gleiche Abstände zwischen nachfolgenden Ausprägungen voraus. Eine der typischen Variablen, der metrisches Messniveau unterstellt wird, ist „Einkommen“. Aber ist der Abstand zwischen 1.000 und 1.100 Euro wirklich der gleiche wie zwischen 4.000 und 4.100 Euro?

Die Skalenniveaus sind dabei in zweierlei Hinsicht für den Forschungsprozess relevant:

- *Homomorphie der Messung*: Bei der Umwandlung der inhaltlich-verbalen Aussagen in Zahlen im Rahmen der Datenaufbereitung muss darauf geachtet werden, dass die Zahlen dieselben Verhältnisse widerspiegeln wie die Worte, d.h. die Ordnungsinformation muss erhalten bleiben. Bei nominalskalierten Variablen ist es weniger wichtig, welche Zahlen für welche Inhalte vergeben werden, zentral ist lediglich, dass alle Personen mit den gleichen Ausprägungen denselben Zahlenwert erhalten, also dass z.B. bei der Variable „Geschlecht“ alle Frauen mit „0“ und alle Männer mit „1“ kodiert werden. Bei ordinal skalierten Variablen muss die Reihenfolge der Zahlen der Reihenfolge der Werte entsprechen. Bei Noten kann z.B. folgende Kodierung gewählt werden: Sehr gut = 1; Gut = 2; Befriedigend = 3; Ausreichend = 4; Nicht ausreichend = 5. Bei metrischen Werten wird meistens der Wert als Code verwendet, z.B. 37 für „37 Jahre“.

- *Wahl des richtigen statistischen Verfahrens:* Entsprechend dem Messniveau sind unterschiedliche mathematische Operationen erlaubt, wovon wiederum die Wahl des statistischen Verfahrens abhängt. Grundsätzlich gilt: Je höher das Messniveau ist, desto mehr Operationen sind erlaubt. So darf beim Alter (metrisch) der Mittelwert berechnet werden, während beim Familienstand (nominal) lediglich der häufigste Wert (der Modus) angegeben werden kann, wo sollte ein „durchschnittlicher Familienstand“ auch liegen?

Die Messniveaus sind hierarchisch geordnet, d.h. für die Variablen mit höherem Skalenniveau sind auch die statistischen Operationen der niedrigeren Skalenniveaus erlaubt (aber nicht umgekehrt). So kann bei der metrischen Variable Einkommen nicht nur der Mittelwert angegeben werden, sondern auch der Median (der mittlere Wert) und der häufigste Wert (Modus). In diesem Fall ist der Median meistens sogar der bessere Indikator, da Extremwerte (hier extrem hohe Einkünfte) kaum einen Einfluss auf das Ergebnis haben, was z.B. beim Äquivalenzeinkommen (Burzan, Huinink, Kapitel 81 und 82 in diesem Band) sehr wichtig ist.

---

## 79.2 Typische Verfahrensschritte

Unabhängig vom Skalenniveau und davon, ob es sich um ein uni-, bi- oder multivariates Verfahren handelt, bestehen alle statistischen Verfahren aus denselben typischen Verfahrensschritten, die im Folgenden kurz dargelegt werden.

### 79.2.1 Mathematische Modellierung der Forschungsfrage

Die Forschungsfrage muss so formuliert und operationalisiert werden, dass sie formal (also: statistisch) modelliert werden kann. Dies ist einer der wichtigsten Schritte im quantitativen Forschungsprozess, weil hier sozial- und gesellschaftstheoretische Annahmen getroffen werden müssen – sind diese falsch, sind alle weiteren Ergebnisse hinfällig. Eine Schwierigkeit bei interpretativen Theorien (Reichert, Kapitel 2 in diesem Band) ist, dass diese von einer unscharfen Struktur der Wirklichkeit ausgehen, die meisten mathematischen Modelle aber Exaktheit voraussetzen. Man kann dies entweder dadurch lösen, dass man mit mathematischen Modellen arbeitet, die Unschärfe handhaben können (die dann aber meistens sehr komplex sind) oder dass man später bei der Interpretation der Daten berücksichtigt, dass die statistischen Ergebnisse eine größere Exaktheit vorspiegeln, als dies der Wirklichkeit entspricht.

### 79.2.2 Wahl eines geeigneten statistischen Verfahrens

Die Wahl eines geeigneten statistischen Verfahrens ist eine der wichtigsten Entscheidungen im Rahmen der Operationalisierung, da hierbei die theoretischen (sozialwissenschaftlichen) Annahmen des Forschers formalisiert und damit für die Analyse festgelegt werden. Dabei ist zu beachten, dass mit den verschiedenen multivariaten Verfahren jeweils nur bestimmte Strukturen (oder Muster) in den Daten entdeckt bzw. beschrieben werden können, womit die Interpretation auf einen Bereich festgelegt ist. Wird ein der Fragestellung (den postulierten Annahmen) unangemessenes statistisches Verfahren verwendet, so werden die erwarteten inhaltlichen Zusammenhänge zwischen den Variablen allenfalls per Zufall erkannt, sofern sie überhaupt vorhanden sind. Des Weiteren können irrtümlich Zusammenhänge erkannt werden, die es gar nicht gibt. Bei allen quantitativen Verfahren ist die Art der Interpretation abhängig vom gewählten Analyseverfahren. So können mit der Clusteranalyse Gruppen von Personen (z.B. Milieus) identifiziert werden, aber es ist nicht möglich zu sagen, welche Variablen einen wie starken Effekt auf die zu erklärende Variable haben, wie dies z.B. mit einem Regressionsverfahren möglich ist. Es sei angemerkt, dass mit verschiedenen Verfahren z.T. die gleichen Zusammenhänge auf Basis unterschiedlicher Theorien auch unterschiedlich interpretiert werden können; in diesem Fall ist dann auch das zugrundeliegende statistische Modell unterschiedlich. So können sowohl Regressions- als auch Faktorenanalysen (siehe unten) auf der Basis der gleichen Zusammenhänge zwischen Variablen berechnet werden, bei der Regressionsanalyse werden diese Zusammenhänge aber oft als ein Zeichen einer möglichen Kausalität, bei der Faktorenanalyse als Zeichen für das Vorliegen einer gemeinsamen latenten Hintergrundvariable interpretiert.

### 79.2.3 Überprüfung der Anwendungsvoraussetzungen

Jedes statistische Verfahren hat Anwendungsvoraussetzungen, die überprüft werden müssen:

1. Es soll sich um valide, ggf. standardisierte Daten handeln, wobei die Messfehler so gering wie möglich sein sollten (Krebs/Menold, Faulbaum, Kapitel 30 und 31 in diesem Band), d.h., wenn die *Qualität der Daten* (Blasius/Thiessen 2012) zweifelhaft ist, sollten sie nicht ausgewertet werden.
2. Die Stichprobe muss eine Mindestgröße haben; wobei die *Mindeststichprobe* von der Forschungsfrage und vom gewählten statistischen Modell abhängt.
3. Sollen die Ergebnisse auf die Grundgesamtheit verallgemeinert werden, dann ist eine *Zufallsstichprobe* (Häder/Häder, Kapitel 18 in diesem Band) erforderlich, wobei die Stichprobe *keine systematischen Ausfälle* enthalten darf. Letzteres gilt sowohl auf der Ebene der Zielpersonen, als auch auf der Ebene der im Modell verwendeten Variablen (Engel/Schmidt, Cielebak/Rässler, Kapitel 23 und 26 in diesem Band).

4. Das *Skalenniveau* der Daten (ungeordnet kategorial, geordnet kategorial, kontinuierlich) muss dem gewünschten Auswertungsverfahren angemessen sein.
5. Viele Verfahren setzen eine bestimmte *Verteilung der Daten in der Grundgesamtheit* (z.B. Normalverteilung) voraus. Werden diese Voraussetzungen missachtet, so kann dies zu falschen Ergebnissen führen.
6. Alle statistischen Verfahren sind darauf ausgerichtet, Gemeinsamkeiten und Unterschiede zwischen Fällen bzw. Variablen zu erkennen. Hierzu müssen aber diese Unterschiede auch vorhanden sein, weshalb statistische Verfahren eine *Mindeststreuung der Variablen* voraussetzen, da ohne Varianz nichts erklärt werden kann. So macht es wenig Sinn, geschlechtsspezifische Differenzen zu beschreiben und nur Frauen (oder nur Männer) in der Stichprobe zu haben.
7. Je nach gewählten statistischen Verfahren können noch *weitere Voraussetzungen an die Daten* hinzukommen, die entsprechenden Details können den entsprechenden Lehrbüchern entnommen werden (Wolf/Best 2010, Backhaus et al. 2010, 2012).

Ist bereits eine der benötigten Voraussetzungen nicht erfüllt, so kann dies zu fehlerhaften Ergebnissen führen, wobei die einzelnen Verfahren unterschiedlich *robust* gegenüber entsprechenden Verletzungen sind. Das Problem ist, dass in der Forschungspraxis sehr oft eine oder sogar mehrere der Anwendungsvoraussetzungen nicht erfüllt sind. So ist beispielsweise aus den oben genannten Gründen die Anforderung an das Skalenniveau sehr oft nicht erfüllt. Der Forscher muss dann entscheiden, ob er – mit entsprechenden Einschränkungen hinsichtlich der internen und externen Validität (Krebs/Menold, Kapitel 30 in diesem Band) und damit auch der Gültigkeit seiner Ergebnisse – trotzdem das Verfahren anwenden möchte oder ob es sinnvoller ist, ein alternatives Verfahren anzuwenden. Fehlende Anwendungsvoraussetzungen, z.B. hinsichtlich der benötigten Stichprobengröße, können auch zu dem Schluss führen, dass statt eines quantitativen ein qualitatives Verfahren gewählt wird.

#### 79.2.4 Datenaufbereitung

Im Zuge der Datenaufbereitung (Lück/Landrock, Cielebak/Rässler, Kapitel 28 und 26 in diesem Band) wird jedem Individuum für jeden beobachteten Zeitpunkt und für jede Variable ein Wert zugeordnet. Für das Auswertungsverfahren ist es dabei egal, ob es sich bei den Individuen um Situationen, einzelne Personen oder um soziale Kollektive handelt. Des Weiteren können dieselben Daten mit Hilfe verschiedener statistischer Verfahren ausgewertet werden, die Daten müssen allerdings die Anwendungsbedingungen für das jeweilige Verfahren erfüllen.

### 79.2.5 Deskriptive Statistik

Nach der Datenaufbereitung wird das statistische Modell mit Hilfe einer geeigneten Software (z.B. STATA, SPSS, R) auf die vorhandenen Daten angewendet. Dabei werden die Daten i.d.R. zunächst mit Hilfe der *deskriptiven Statistik* (auch: *beschreibende Statistik*) verdichtet. Die Ergebnisse der deskriptiven Statistik sind i.d.R. *Tabellen, Grafiken oder bestimmte Parameter und Maßzahlen*. Soll z.B. die Geschlechterungleichheit im deutschen Hochschulwesen untersucht werden und gibt es dazu einen Datensatz, der aus einer Befragung von 100 Informatikstudierenden der Universität XY aus dem Jahr 2014 kommt, dann könnte das Ergebnis der beschreibenden Analyse sein: „In dem Datensatz sind 20% Frauen“, d.h. in diesem Datensatz sind 20 Frauen und 80 Männer.

### 79.2.6 Induktive Statistik

Sofern eine Zufallsstichprobe vorliegt, können die Ergebnisse mit Hilfe der *induktiven Statistik* (auch: *schließende Statistik*) auf die zuvor definierte Grundgesamtheit verallgemeinert werden. Dabei muss zunächst festgelegt werden, was die Grundgesamtheit ist (Häder/Häder, Kapitel 18 in diesem Band). Sind es in diesem Beispiel etwa die Informatiker der Universität XY aus dem Jahr 2014? Oder sind es die Informatiker der Universität XY aus den letzten 20 Jahrgängen? Oder sind es alle Informatiker in Deutschland? An diesem Beispiel wird ersichtlich, dass es viele theoretische Annahmen über das Verhältnis von Stichprobe und Grundgesamtheit geben kann (Behnke/Behnke 2006).

Angenommen es soll nur etwas über die Informatiker der Universität XY aus dem Jahr 2014 gesagt werden, und es ist bekannt, dass an dieser Universität insgesamt  $N = 1.000$  Personen Informatik studieren (Grundgesamtheit), dann ist der nächste Schritt, von der Stichprobe ( $n = 100$ ) auf die Grundgesamtheit ( $N = 1.000$ ) zu schließen. Da nur eine (Zufalls-)Stichprobe und nicht die Gesamtheit befragt wurde, kann die bereits implizit erfolgte Schlussfolgerung, dass an der Universität XY der Frauenanteil der Informatikstudierenden bei 20% liegt, falsch sein. Denn es ist theoretisch auch möglich, dass lediglich eine (sehr) ungünstige Stichprobe gezogen wurde. Vielleicht studieren an der Universität XY insgesamt nur 20 Frauen Informatik (der Frauenanteil läge also bei 2%), aber diese sind zufällig alle in die Stichprobe geraten. Vielleicht studieren auch 500 Frauen Informatik (der Frauenanteil läge also bei 50%). In beiden Fällen würden wir auf Basis unserer Stichprobe den Frauenanteil falsch einschätzen. Das Problem ist, dass unbekannt ist, wie hoch der Frauenanteil tatsächlich ist, und solange keine Vollerhebung vorliegt, kann auch mit der schließenden Statistik nicht *sicher* angegeben werden, ob die Daten richtig oder falsch sind. Um dennoch eine konkrete Aussage machen zu können, kann die Wahrscheinlichkeit berechnet werden, mit der ein Fehler gemacht wurde (Dubben/Beck-Bornholdt 2011). Typische Varianten des statistischen Schließens sind *statistische Tests* (auch: *Hypothesentests*), mit deren Hilfe ein *Signifikanzniveau* berechnet wird, und *Konfidenzintervalle*, mit deren Hilfe auf einem bestimmten *Konfidenzniveau Sicherheitsbereiche* geschätzt werden

(Behnke/Behnke 2006). Wie diese konkret berechnet werden, hängt ab von dem gewählten statistischen Verfahren, welches wiederum vom Messniveau der verwendeten Variablen abhängt.

### 79.2.7 Interpretation der Daten

Sind die relevanten Maßzahlen berechnet, dann müssen sie sozialwissenschaftlich *interpretiert* werden, d.h. es muss überlegt werden, was dieser Wert für die Forschungsfrage bzw. die soziologische Theorie bedeutet. Was bedeutet es z.B., wenn der Frauenanteil in der Informatik bei 20% liegt?

Er liegt sicherlich deutlich unter 50%, und damit (vermutlich) unter dem durchschnittlichen Anteil der weiblichen Studierenden an der Universität XY, aber diese Aussage ist nicht gesichert. Es kann sich um eine technisch ausgerichtete Universität handeln, an welcher der gesamte Frauenanteil unter 20% liegt, der Wert somit schon relativ hoch ist. Des Weiteren muss dieser Wert mit dem Frauenanteil an Informatikstudierenden im gesamten Land verglichen werden, um eine inhaltliche Aussage machen zu können.

Je nachdem, was man als Vergleichsmaßstab heranzieht, könnte man in diesem Beispiel sagen, dass 20% sehr wenig ist, weil bei vollkommener Gleichberechtigung der Frauenanteil bei 50% liegen sollte, oder man könnte dies auch als hervorragendes Ergebnis im Sinne der Geschlechtergerechtigkeit deuten, weil an anderen Universitäten der Frauenanteil in der Informatik deutlich niedriger liegt. Und wie wäre es zu interpretieren, wenn der Frauenanteil bei 45% läge – sind diese fünf fehlenden Prozentpunkte zur Gleichverteilung so wichtig, dass man von Geschlechtergerechtigkeit sprechen kann?

Kurz: Um die Relevanz von statistischen Ergebnissen beurteilen zu können, müssen sie interpretiert werden, und für diese Interpretationen werden die Sozialwissenschaften benötigt. Vor allem sollte aber deutlich geworden sein, dass quantitative Sozialforschung nicht theoriefrei ist, sondern dass im Gegenteil an zahlreichen Stellen theoretische Annahmen getroffen werden müssen, um inhaltlich sinnvolle Analysen überhaupt durchführen zu können. In den meisten Fällen ist die Fragestellung wesentlich komplexer, so dass eine Vielzahl von Variablen gleichzeitig in die Analysen einzubinden sind – dies geht mit Hilfe der multivariaten Statistik.

---

## 79.3 Varianten multivariater Verfahren

Es gibt nahezu unzählige Varianten multivariater Verfahren und auch unterschiedliche Möglichkeiten, diese zu klassifizieren. Im Folgenden stellen wir eine Klassifikationsmöglichkeit für Analysen vor, die mehrere Variablen über Fälle (z.B. Personen) zu einem Zeitpunkt (Querschnittsanalyse) erfassen. Für diese sog. multivariaten Querschnittsanalysen werden die unterschiedlichen Verfahren der multivariaten Datenanalyse zunächst auf drei Ebenen unterschieden, diese sind in Abbildung 1 dargestellt.

### 79.3.1 Funktion und Struktur: Verfahren für manifeste und latente Variablen

Auf der ersten Ebene besteht die Unterscheidung zwischen *Funktion*, also dass nur mit manifesten Variablen gearbeitet wird, und *Struktur*, welches die Einbindung von latenten Variablen anzeigt. Dabei sind *manifeste Variablen* direkt beobachtbare Merkmale. Dazu gehören z.B. das Alter, das Geschlecht, der Bildungsabschluss, ob jemand an der letzten Bundestagswahl teilgenommen hat oder nicht, welche Partei er/sie gewählt hat sowie auch die einzelnen Fragen zur Messung von Einstellungen. Ein Beispiel für eine solche Frage ist: „Wie stark stimmen sie der folgenden Aussage zu: Für eine Frau ist es wichtiger, ihrem Mann bei seiner Karriere zu helfen, als selbst Karriere zu machen“, die Antwort kann auf einer vier-, fünf- oder auch höherstufigen Skala (Franzen, Kapitel 51 in diesem Band) vorgegeben werden, die von „stimme voll und ganz zu“ bis „stimme überhaupt nicht zu“ reicht. Alle diese Fragen werden direkt gestellt, sie sind direkt „beobachtbar“.

Im Gegensatz zu manifesten Variablen sind *latente Variablen* nicht direkt beobachtbar, sie werden aus einer Verknüpfung von manifesten Variablen ermittelt (Latcheva/Davidov, Blasius, Kapitel 55 und 83 in diesem Band). Ein Beispiel ist der Intelligenzquotient (IQ), oder besser, verschiedene Arten des Intelligenzquotienten (mathematische Intelligenz, sprachliche Intelligenz, ...), der eine normierte Kenngröße zur Bewertung der intellektuellen Leistung ist; als Mittelwert wird gewöhnlich 100 und als Standardabweichung 15 gewählt. Da nicht gefragt werden kann „wie intelligent sind sie?“ (oder irgendeine ähnliche Formulierung) wird mittels der Antworten auf Fragen eines umfangreichen Tests statistisch geschätzt, welche intellektuelle Leistungen der Proband erbracht hat.

Funktion	
Kontinuierlich	Kategorial
<i>Regression</i> $y \approx f(x)$ Regressionsanalyse Varianzanalyse	<i>Klassifikation</i> $g \approx f(x)$ Logistische Regression Logitanalyse
<i>Skalierung</i> $x \rightarrow y_1, y_2, \dots$ Faktorenanalyse Korrespondenzanalyse	<i>Clustertechniken</i> $g \rightarrow y_1, y_2, \dots$ Clusteranalyse Analyse latenter Klassen

### Struktur

**Abb. 79.1** Übersicht über multivariate Analyseverfahren

### 79.3.2 Verfahren für kontinuierliche und kategoriale Variablen

Auf der zweiten Ebene wird unterschieden, ob die *abhängige Variable* kategorial (einschließlich dichotom) oder kontinuierlich skaliert ist. Auf beiden Stufen kann zudem nach *geordnet* und *ungeordnet kategorial* unterschieden werden, die den *geordnet kategorial* zugeordneten Verfahren werden in den Sozialwissenschaften jedoch relativ selten verwendet, so dass hier lediglich vier Gruppen von Verfahren unterschieden werden:

- Die abhängige Variable ist kontinuierlich skaliert und wird manifest erhoben.
- Die abhängige Variable ist kategorial skaliert und wird manifest erhoben.
- Die abhängige Variable ist kontinuierlich skaliert und ein latentes Merkmal.
- Die abhängige Variable ist kategorial skaliert und ein latentes Merkmal.

In jedem dieser insgesamt vier Bereiche wird dann auf der Ebene der *unabhängigen Variablen* ebenfalls danach unterschieden, ob diese kategorial (oder auch geordnet kategorial) oder kontinuierlich skaliert sind. Mischformen von kategorial und kontinuierlich skalierten Variablen werden meist innerhalb der Verfahren diskutiert.

### 79.3.3 Kausalbeziehungen

Eine weitere Ebene der Unterscheidung der oben genannten Bereiche ist die Frage der Kausalität, also haben die unabhängigen Merkmale die abhängigen beeinflusst? Kühnel und Dingelstedt (Kapitel 80 in diesem Band) nennen als Beispiel für eine kausale Wirkung, dass der Schaden nach einem Brand u.a. von der Stärke dieses Brandes abhängt. Bei den meisten sozialwissenschaftlichen Fragestellungen ist diese Kausalität nur sehr schwer zu zeigen, denn die Ursache muss zeitlich immer vor der Wirkung liegen, so wie dies z.B. beim Experiment im Untersuchungsdesign angelegt ist (Eifler, Kapitel 11 in diesem Band). Auch Paneldaten (Schupp, Kapitel 73 in diesem Band) und Ereignisdaten (Mika/Stegemann, Kapitel 85 in diesem Band) eignen sich zur Bestimmung kausaler Wirkungen.

---

## 79.4 Verfahren mit ausschließlich manifesten Variablen

### 79.4.1 Regressionsansätze

Wenn auf der ersten Ebene entschieden wurde, dass es im statistischen Modell ausschließlich manifeste Variablen gibt (linker und rechter oberer Quadrant in Abb. 79.1), und wenn die abhängige Variable (zweite Ebene) kontinuierlich skaliert ist, dann wird von Regressionsansätzen gesprochen (linker oberer Quadrant). Sind die unabhängigen Variablen zumindest nicht ungeordnet kategorial skaliert, dann wird oft die *multiple Regressionsanalyse* verwendet.

Das zugrunde liegende Modell der multiplen Regression ist:

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_m x_{mi} + u_i.$$

Dabei ist  $y$  die *abhängige Variable*. Dies kann z.B. das Einkommen sein. Die Variablen  $x_1$  bis  $x_m$  sind die *unabhängigen Merkmale*, mit denen die abhängige Variable erklärt werden soll, z.B. Berufserfahrung (in Jahren gemessen) und höchster Bildungsabschluss.  $b_1$  bis  $b_m$  sind die *Regressionskoeffizienten* (dies sind Indikatoren für die Stärke und Richtung des Zusammenhangs zwischen der abhängigen und den unabhängigen Variablen), der Index „ $i$ “ verweist auf die  $i$ te Person und  $u_i$  ist der *Fehlerterm*, der sich aus der Differenz aus dem erklärenden Teil der Gleichung ( $b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_m x_{mi}$ ) und dem Wert der zu erklärenden Variable ( $y_i$ ) ergibt.

Dabei muss die abhängige Variable kontinuierlich skaliert sein, bei den unabhängigen Variablen können auch (wenige) dichotome (z.B. mit den Ausprägungen ja/nein) und ordinal skalierte Merkmale einbezogen werden. Wäre die abhängige Variable lediglich fünfstufig skaliert, dann läge die Gleichung  $b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_m x_{mi}$  nur relativ selten dicht an den Antwortvorgaben (1, 2, 3, 4, 5), d.h. der Fehlerterm  $u_i$  wäre für die meisten Befragten schon aus modellspezifischen Gründen relativ groß. In der Forschungspraxis werden dennoch relativ häufig Regressionsanalysen mit einer ordinal skalierten abhängigen Variable gerechnet, vielfach vermutlich schon, weil die den Daten angemesseneren Alternativen (z.B. die ordinale Regression) relativ schwer anzuwenden und recht aufwändig zu interpretieren sind (Long 1997).

Ist die abhängige Variable kontinuierlich skaliert (z.B. Einkommen) und sind die unabhängigen Variablen kategorial skaliert (z.B. Berufspositionen), werden oft Methoden der *Varianzanalyse* verwendet.

Der Regressions- und Varianzanalyse ist gemeinsam, dass die Interpretation sowohl über Parameter der *Signifikanz* (häufig mit  $\alpha$  abgekürzt), z.B. bezüglich der einzelnen Variablen, als auch über Parameter der *Teststärke* (in der Regressionsanalyse meistens R-Quadrat, in der Varianzanalyse meistens Eta-Quadrat) verläuft (Beck-Bornholdt/Dubben 2006). D.h., man würde bei der Regressionsanalyse sowohl den Regressionskoeffizienten, als auch das Ergebnis eines statistischen Tests aus der induktiven Statistik für die Interpretation heranziehen. Dabei ist zu beachten dass bei großen Fallzahlen die Parameter fast immer statistisch signifikant sind, die Teststärke (die Erklärungskraft) jedoch oft nur sehr gering ist. Bei kleinen Fallzahlen kann die Erklärungskraft relativ hoch sein, ohne dass die Parameter statistisch signifikant sind, ergo, auch hier ist die sozialwissenschaftliche Aussagekraft meistens gering. Für alle Regressionsansätze gilt daher, dass gehaltvolle Aussagen nur im Zusammenspiel von Teststärke und statistischer Signifikanz gemacht werden können (Akremi et al. 2011: 245-298).

### 79.4.2 Klassifikationsansätze

Ist die abhängige Variable manifest und kategorial skaliert, wird von Klassifikationsansätzen gesprochen (Tabelle 1, rechter oberer Quadrant). Ist die abhängige Variable dichotom, z.B. bei der Frage nach der Wahlbeteiligung bei der letzten Bundestagswahl (mit den Antwortmöglichkeiten „ja“ und „nein“), und sind die unabhängigen Variablen kontinuierlich skaliert (wobei, wie auch bei der Regressionsanalyse, ein paar geordnet kategorial skalierte bzw. dichotom skalierte Merkmale zugelassen werden können), dann wird sehr oft die *binäre logistische Regressionsanalyse* verwendet. Die *multinomiale logistische Regressionsanalyse* werden verwendet, wenn die abhängige Variable mehr als zwei kategorial skalierte Ausprägungen hat, wobei in den Sozialwissenschaften eigentlich nie mehr als drei Ausprägungen verwendet werden.

Klassifikationsverfahren unterscheiden sich von den Regressionsverfahren dadurch, dass nicht die Ausprägung der abhängigen Variable geschätzt wird, sondern die Wahrscheinlichkeit ihres Auftretens. In Wahlstudien kann z.B. geschätzt werden, mit welcher Wahrscheinlichkeit eine Person an der nächsten Bundestagswahl teilnimmt oder nicht. Die Schätzung basiert auf dem Verlauf einer *Verteilungsfunktion* (die mit einer sehr geringen Wahrscheinlichkeit beginnt und mit einer sehr hohen aufhört), anhand derer die entsprechende Wahrscheinlichkeit bestimmt wird. Diese ist durch die Ausprägungen der unabhängigen Merkmale geprägt, z.B. steigt die Wahrscheinlichkeit, an der nächsten Bundestagswahl teilzunehmen, mit zunehmender Bildung bzw. mit zunehmendem politischem Interesse kontinuierlich.

Bei der logistischen Regression wird die *logistische Verteilung* angenommen, für diese gilt:

$$P = \frac{e^z}{1 + e^z}, \text{ mit } z = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_m x_{mi}$$

Die Parameter sind hier die gleichen wie im multiplen Regressionsmodell, nur dass sie im Exponenten der Eulerschen Zahl stehen. Ist  $P$  die Wahrscheinlichkeit, dass die Person  $i$  an der Wahl teilnimmt, so ist  $(1 - P)$  die Wahrscheinlichkeit, dass diese Person *nicht* an der Wahl teilnimmt. Liegt die geschätzte Wahrscheinlichkeit für einen Befragten bei über 50%, so ist die Vorhersage der Wahlteilnahme „ja“. Wird anstelle der logistischen Verteilung für die Schätzung der Parameter die Normalverteilung zugrunde gelegt, so wird von *Probitmodellen* gesprochen.

Bei zwei möglichen Ausprägungen sind die Berechnung und insbesondere die Interpretation von Wahrscheinlichkeiten und Komplementärwahrscheinlichkeiten relativ einfach, bei drei Ausprägungen ( $a, b, c$ ) wird es schon relativ komplex. Ist  $P(a)$  berechnet, so ist  $P(b) = 1 - P(a) - P(c)$ , d.h. es gibt keine einfachen Komplementärwahrscheinlichkeiten, was die Interpretation der Parameter deutlich erschwert. Je mehr Ausprägungen zugelassen werden, desto komplizierter wird die Berechnung und desto verwirrender die Interpretation. Ist nicht nur die abhängige Variable kategorial skaliert, sondern sind auch die unabhängigen Variablen kategorial skaliert, dann werden oft Ansätze der *Logitanalyse* verwendet (Vermunt 1997).

## 79.5 Verfahren mit latenten Variablen

### 79.5.1 Clusteranalyse

Latente Variablen sind nicht direkt beobachtete Variablen, die auf der Basis eines Satzes von manifesten Variablen bestimmt werden. Ist die latente Variable ungeordnet kategorial skaliert und sind die manifesten Variablen kontinuierlich skaliert, so handelt es sich meistens um Verfahren der Clusteranalyse (Bacher et al. 2010). Die Clusteranalyse ordnet jedes Objekt (bzw. jede Person) genau einem Cluster zu, und zwar jenem, dem dieses Objekt (diese Person) am ähnlichsten ist. Dabei gilt zum einen, dass die einzelnen Cluster so homogen wie möglich sein sollen, d.h. die Streuung in den Clustern soll minimal sein. Zum anderen sollen sich die Cluster so deutlich wie möglich unterscheiden, d.h. die Streuung zwischen den Clustern soll maximal sein. Anders ausgedrückt: Mit Hilfe der Clusteranalyse wird eine Vielzahl von Objekten (Personen, Länder etc.) anhand von Merkmalen derart zusammengefasst, dass Personen mit ähnlichen Merkmalen zum gleichen Cluster (zur gleichen Gruppe) gehören. Ebenso wie Personen können Variablen zu Clustern zusammengefasst werden, wobei jene eine Cluster bilden, die von den Befragten einander ähnliche Bewertungen bekamen.

Als typisches Beispiel für die Verwendung der Clusteranalyse können die Sinus-Milieus oder die Mediennutzertypologien genannt werden. Ein weiteres Beispiel ist Esping-Andersen's (1990) Klassifikation der europäischen Wohlfahrstaaten, in der er drei Typen von Wohlfahrstaaten identifizierte (liberale, konservative und sozialdemokratische Wohlfahrtsstaaten) und jedes europäische Land einem Typus (Cluster) zuordnete. Deutschland und Österreich gehören z.B. zu den konservativen, Großbritannien zu den liberalen und Schweden zu den sozialdemokratischen Wohlfahrtregimen.

Das Problem bei der Berechnung von Clusteranalysen ist zum einen die Bestimmung der Ähnlichkeit (wie wird diese definiert?) und zum anderen die Art der Zusammenfassung der Objekte (Personen, Länder) zu einem Cluster (nach welchem Algorithmus erfolgt die Zuordnung zu der einen oder anderen Gruppe?). So wurde etwa Esping-Andersen's Klassifikation der Wohlfahrstaaten später um weitere Cluster erweitert (südeuropäisches und postsozialistisches Modell), und einige Staaten wurden anderen Typen zugeordnet.

Aufgrund der Vielzahl von Maßen, mit denen die Ähnlichkeit zweier Merkmalsträger beschrieben werden kann, und aufgrund der Vielzahl von Algorithmen, nach denen die Cluster gebildet werden können, gibt es nicht *die* Clusteranalyse, sondern verschiedene *Verfahren der Clusteranalyse*. In der Regel führt die Verwendung verschiedener Ähnlichkeitsmaße, die es auch erlauben, sowohl kategorial skalierte als auch dichotome Merkmale zu berücksichtigen, bzw. verschiedener Algorithmen der Clusterung auch zu unterschiedlichen Ergebnissen, was häufig den Eindruck der Beliebigkeit von Ergebnissen hinterlässt.

Sind die unabhängigen Variablen kategorial skaliert und gibt es eine Theorie oder Annahmen, welche Variablenkategorien zu welchen Clustern gehören, dann werden oft Techniken der *latenten Klassenanalyse* (*Latent Class Analysis*, kurz *LCA*) verwendet. Das

Verfahren kann den Strukturgleichungsmodellen (s.u.) zugeordnet werden, auch hier muss vor der Analyse die Struktur der Verknüpfungen der Variablenausprägungen (bei den Strukturgleichungsmodellen sind es die Variablen) spezifiziert werden.

### 79.5.2 Skalierung

Soll(en) die latente(n) Variable(n) kontinuierlich skaliert sein (meistens mit Mittelwert Null und Standardabweichung Eins), dann werden in der Regel Skalierungsverfahren verwendet. Sind die manifesten Variablen ebenfalls kontinuierlich skaliert, so kann die Faktorenanalyse bzw. die Hauptkomponentenanalyse für die Schätzung benutzt werden. Als klassisches Beispiel kann die Bestimmung von Intelligenzquotienten genannt werden. Ein anderes Beispiel ist Otte's (2005) Analyse von Lebensstilen, in der er zeigt, dass die meisten lebensstilbezogenen Alltagspraktiken und Einstellung auf zwei Dimensionen reduziert werden können: Modernität und Ausstattungsniveau. Als wichtige Indikatoren für Modernität nennt Otte eine geringe Religiosität, die Abkehr von Familientraditionen, dass man viel ausgeht und es gerne hat, wenn ständig etwas los ist, sowie dass man das Leben in vollen Zügen genießt.

Die *Hauptkomponentenanalyse* (*Principal Component Analysis*, kurz *PCA*) ist ein mathematisches Verfahren, mit dem ein Satz von untereinander korrelierten Variablen – z.B. die Variablen einer Fragebatterie zum Thema „politische Entfremdung“ – in einen Satz von linear unkorrelierten Variablen, den sogenannten *Hauptkomponenten* (oder *Faktoren*, *Dimensionen*), überführt wird. Die Anzahl der Hauptkomponenten ist gleich der Anzahl der ursprünglichen Variablen; in der sozialwissenschaftlichen Praxis werden aber immer nur die Faktoren verwendet, welche die meiste Varianz erklären. Dabei erklärt die erste Hauptkomponente die meiste Varianz, jeder nachfolgende Faktor wiederum die maximale Varianz der jeweils verbleibenden Varianz. Die Faktoren sind voneinander unabhängig, sie sind unkorreliert (oder als oft verwendeter Fachausdruck „orthogonal“).

Die Hauptkomponentenanalyse wird im deutschsprachigen Raum sehr oft mit der *Faktorenanalyse* verwechselt, obwohl sich die beiden Methoden grundlegend unterscheiden. Während es bei der Hauptkomponentenanalyse keinen Fehlerterm gibt, d.h. die zugrunde liegenden manifesten Variablen können durch die gleiche Anzahl von latenten Variablen vollständig erklärt werden, hat die Faktorenanalyse für jede manifeste Variable einen spezifischen Messfehler. Die Schätzung dieser Fehler kann auf sehr unterschiedliche Art erfolgen, dementsprechend sollte hier der Plural verwendet werden, es gibt viele Formen der Faktorenanalyse.

Werden diese Korrelationen zwischen den latenten Variablen in einem theoretischen Modell zuvor postuliert, wird gegebenenfalls auch eine kausale Wirkung der Faktoren aufeinander erlaubt, wird zudem angenommen, welche Variablen auf welchen Faktoren laden und wie ggf. die Fehlerterme miteinander korrelieren, dann wird von *Strukturgleichungsmodellen* gesprochen. Diese wurden in den 1970er Jahren insbesondere von Karl Jöreskog

(Reinecke 2005) formuliert, und sie erfreuen sich insbesondere aufgrund des möglichen Einbezugs einer kausalen Wirkung (Kühnel/Dingelstedt, Kapitel 80 in diesem Band) einer relativ großen Beliebtheit. Einschränkend sollte gesagt werden, dass die Annahmen, die an die Struktur der Daten gemacht werden (Reinecke 2005), häufig nicht eingehalten werden können, da die Qualität sozialwissenschaftlicher Daten oft wesentlich schlechter ist, als dies allgemein angenommen wird (Blasius/Thiessen 2012).

Sind die zugrundeliegenden manifesten Variablen ordinal skaliert, z.B. Fragen zur Einstellung bezüglich der politischen Entfremdung, die jeweils vier oder fünf Ausprägungen haben, z.B. von „stimme voll und ganz zu“ bis „stimme überhaupt nicht zu“, und soll diese Ordinalität der Daten berücksichtigt werden, dann kann die *kategoriale (oder nicht-lineare) Hauptkomponentenanalyse* (Gifi 1990) verwendet werden. Bei diesem Verfahren, welches in den meisten sozialwissenschaftlichen Anwendungen den Daten angemessener sein dürfte als die einfache Hauptkomponentenanalyse, werden die Abstände zwischen den Kategorien iterativ neu geschätzt. Auf der Basis dieser iterativen Schätzungen wird dann die klassische Hauptkomponentenanalyse gerechnet, die Resultate können analog zu dieser interpretiert werden.

In vielen Fällen kann bei sozialwissenschaftlichen Daten nicht einmal ordinales Daten-niveau vorausgesetzt werden, z.B. bei Merkmalen des Lebensstils oder bei der Parteipräfe-renz; in diesen Fällen ist die *multiple Korrespondenzanalyse* (auch: *Homogenitätsanalyse*) das geeignete Instrument zur Skalierung der Daten (Blasius 2001). Ein bekanntes Beispiel sind Bourdieus (1982) Analysen in den „feinen Unterschieden“, wo er untersuchte welche Berufsgruppen welchen Freizeitaktivitäten nachgehen und über welche Kapitalsorten sie verfügen. Anders als bei der Hauptkomponentenanalyse steht bei der multiplen Korre-spondenzanalyse die Interpretation der grafischen Darstellung im Vordergrund der Inter-pretation, die numerischen Werte werden, obwohl vorhanden, nur selten interpretiert.

---

## 79.6 Analysen von Zeit, Raum und sozialen Beziehungen

Abbildung 79.1 gibt eine grundlegende Einteilung der unterschiedlichen multivariaten Verfahren wieder, um Standardauswertungen durchzuführen. Beziehungen Verfahren, mit denen die zeitliche Dimension oder der sozial-räumliche Kontext abgebildet werden kann. Für diese Daten bestehen eigene Techniken, wobei es sich oft um Fortentwicklungen der oben genannten Verfahren handelt. Unabhängig davon gilt die in Abb. 79.1 wiederge-gabene Einteilung in die Bereiche „Regression“, „Klassifikation“, „Clusterung“ und „Skalie-rung“ auch hier.

### 79.6.1 Verfahren zur Analyse sozialer Prozesse

Ein weiteres häufig vorkommendes Ziel quantitativer Sozialforschung ist die Analyse sozialer Prozesse (Baur 2015), bei denen Daten über einen zeitlichen Verlauf beschrieben werden sollen. Dabei können sowohl Veränderungen in den Lebensverläufen als auch Veränderungen auf der Meso- und Makroebene (z.B. Industrialisierung, Modernisierung etc.) beschrieben werden. Um solche Prozesse analysieren zu können, sind Längsschnittdaten (Stein, Kapitel 7 in diesem Band) erforderlich, also entweder Daten im Trenddesign (Mochmann, Kapitel 14 in diesem Band) oder im Paneldesign (Schupp, Kapitel 73 in diesem Band). Typische Verfahren zur Analyse sozialer Prozesse sind die *Kohortenanalyse* (Stein, Kapitel 7 in diesem Band), die *Zeitreihenanalyse* (Metz/Thome, Kapitel 84 in diesem Band), die *Verlaufsdatenanalyse* (Pötter/Prein, Kapitel 86 in diesem Band) sowie die *Ereignisdatenanalyse* und die *Sequenzmusteranalyse* (Mika/Stegmann, Kapitel 85 in diesem Band).

### 79.6.2 Verfahren zur Analyse räumlicher Prozesse

Um räumliche Unterschiede mit in die Analyse einzubeziehen, werden oft Geoinformationssysteme und entsprechende Verfahren der räumlichen Analyse (Lakes, Kapitel 77 in diesem Band) verwendet. Des Weiteren werden in der kulturvergleichenden Forschung oft unterschiedliche Länder miteinander verglichen (Graeff, Kapitel 72 in diesem Band).

### 79.6.3 Verfahren zur Analyse von Mikro-Makro-Problemen

Eng verwandt mit der räumlichen Analyse ist das Interesse, Mikro-Makro-Beziehungen aufzuklären, also zu untersuchen, wie soziale Kontexte bzw. Aggregate (z.B. Nationen, Organisationen, Städte) auf Individuen (z.B. Personen) wirken. Hierzu eignet sich die *Mehrebenenanalyse* (Pötschke, Kapitel 87 in diesem Band). Dieses Verfahren setzt voraus, dass in den Daten nicht nur Informationen über die Mikroebene, sondern auch über das Aggregat (Graeff, Kapitel 72 in diesem Band) vorhanden sind.

### 79.6.4 Verfahren zur Analyse von sozialen Beziehungen

Um die Beziehungsstrukturen zwischen Akteuren, z.B. zwischen Politkern, aber auch zwischen Freunden, Organisationen, oder den Mitgliedern einer Fußballmannschaft beschreiben zu können, werden *netzwerkanalytische Verfahren* verwendet (Baur, Kapitel 74 in diesem Band). Hier kann von einer Person ausgegangen werden, die ihr Verhältnis zu ihren besten Freunden beschreibt, das sind die sogenannten *Ego-Netzwerke*, oder es werden *Teilgruppen* oder die *Gesamtstruktur eines Netzwerkes* beschrieben.

## 79.7 Deduktive und induktive Logik: Strukturprüfende und strukturentdeckende Verfahren

Unabhängig von der Art des Skalenniveaus wird in den Sozialwissenschaften oft zwischen strukturprüfenden und strukturentdeckenden Verfahren unterschieden. Bei *strukturprüfenden Verfahren* wird gemäß dem deduktiv-nomologischen Modell (Reichertz, Stein, Kapitel 2 und 7 in diesem Band) statistisch geprüft, ob eine theoretisch vorgegebene, also inhaltlich abgeleitete, Struktur der Daten auch empirisch nachgewiesen werden kann. Im einfachsten Fall wird dabei von einer linearen Regression (Abschnitt 79.4.1) ausgegangen, in welcher postuliert wird, dass die Variablen  $x_1$  bis  $x_m$  unter Konstanthaltung der jeweils anderen Variablen, einen signifikanten, linearen Effekt auf die zu erklärende Variable  $y$  haben. In komplizierteren (oder komplexeren) Fällen kann es eine Vielzahl von abhängigen und unabhängigen Merkmalen geben, bei denen auch latente Merkmale einbezogen werden, d.h. es gibt komplexe Strukturen von Zusammenhängen, die möglichst simultan getestet werden sollen. Dies ist mit Hilfe der oben bereits erwähnten Strukturgleichungsmodelle möglich.

Bei *strukturentdeckenden Verfahren* werden gemäß der Logik des induktiv-statistischen Syllogismus (Reichertz, Kapitel 2 in diesem Band) – etwa bei der Hauptkomponentenanalyse, der Korrespondenzanalyse oder der schrittweisen Regressionsanalyse – im Vorfeld der Analysen keine expliziten Annahmen über die Struktur der Daten gemacht, diese Strukturen (oder Zusammenhänge) zwischen den Variablen bzw. den Variablenausprägungen sollen erst im Rahmen der Analysen entdeckt werden.

Quantitative Sozialforschung kann folglich sowohl einer deduktiven, als auch einer induktiven Logik folgen. Dabei muss beachtet werden, dass – wie Reichertz (Kapitel 2 in diesem Band) erläutert – keine der Logiken theoriefrei ist. Vielmehr basiert auch die Anwendung strukturentdeckender Techniken auf sozialwissenschaftlicher Theorie. So muss im Fall der Hauptkomponenten- und der Korrespondenzanalyse zuvor inhaltlich postuliert werden, welche Variablen (Variablenausprägungen) Gemeinsamkeiten haben, z.B. welche Merkmale auf den gleichen Achsen laden und ob diese Merkmale positiv oder negativ miteinander assoziiert sind. Als Beispiel können hier die bereits erwähnten Arbeiten von Bourdieu genannt werden, z.B. seine „feinen Unterschiede“ (Bourdieu 1982), bei der die latenten Variablen als „ökonomisches Kapital“ und „kulturelles Kapital“ bzw. als „Zusammensetzung von ökonomischem und kulturellem Kapital“ und „Kapitalvolumen“ bezeichnet werden.

Es sei abschließend noch angemerkt, dass in der Forschungspraxis der Unterschied zwischen strukturprüfenden und strukturentdeckenden Verfahren deutlich geringer ist, als dies in Publikationen den Anschein hat. Es hat den einfachen Grund, dass postulierte Zusammenhänge, die statistisch *nicht* belegt werden können, so gut wie nie publiziert werden (Friedrichs, Kapitel 16 in diesem Band). Vielmehr wird zumindest in der sozialwissenschaftlichen Forschungspraxis solange nach Zusammenhängen gesucht, bis ein zufriedenstellendes Ergebnis gefunden wurde. „Zufriedenstellend“ meint hierbei, dass die gefundenen Zusammenhänge statistisch signifikant sind, dass die Erklärungskraft

ausreichend hoch ist (was manchmal nur behauptet wird, inhaltlich aber nicht zutreffend ist) und dass es eine gute inhaltliche Erklärung gibt, die in Übereinstimmung mit der Theorie (welcher auch immer) steht.

Des Weiteren werden die zur Verfügung stehenden Statistikprogramme, mit denen derartige Strukturzusammenhänge geprüft werden können, immer benutzerfreundlicher; inzwischen gibt es Algorithmen, die Vorschläge machen, welche Parameter wie gesetzt werden sollten, damit der Modellfit optimal ist. Werden dementsprechend die statistischen Modelle immer wieder verändert – und damit auch die zugrundeliegende sozialwissenschaftliche Theorie, so sollte dieses Vorgehen eigentlich nicht als „strukturprüfend“, sondern als „strukturentdeckend“ bezeichnet werden. In Publikationen werden dann aber nur die signifikanten Ergebnisse berichtet und nicht jene, mit denen gezeigt wird, dass die Annahmen, die für die ersten Analysen gemacht wurden, falsch waren.

## Literatur

- Akremi, Leila/Baur, Nina/Fromm, Sabine (Hg.) (2011): Datenanalyse mit SPSS für Fortgeschrittene 1: Datenaufbereitung und uni- und bivariate Statistik. Wiesbaden: VS Verlag
- Bacher, Johann/Pöge, Andreas/Enzig, Knuw (2010): Clusteranalyse: Anwendungsorientierte Einführung in Klassifikationsverfahren. München: Oldenbourg
- Backhaus, Klaus/Bernd Erichson/Rolf Weiber (2012): Fortgeschrittene Multivariate Analysemethoden. Berlin: Springer
- Backhaus, Klaus/Bernd Erichson/Wulff Plinke/Rolf Weiber (2010): Multivariate Analysemethoden. Berlin: Springer
- Baur, Nina (2015): Verlaufsmusteranalyse. Wiesbaden: VS Verlag
- Baur, Nina/Lamnek, Siegfried (2007): Variables. In: Ritzer, George (Hg.): The Blackwell Encyclopedia of Sociology. Blackwell Publishing Ltd.: 3120-3123
- Beck-Bornholdt, Hans-Peter/Dubben, Hans-Hermann (2006): Die Bedeutung der statistischen Signifikanz. In: Diekmann, Andreas (Hg.): Methoden der Sozialforschung. Wiesbaden: VS Verlag: 61-74
- Behnke, Joachim/Behnke, Nathalie (2006): Grundlagen der statistischen Datenanalyse. Eine Einführung für Politikwissenschaftler. Wiesbaden: VS Verlag
- Blasius, Jörg (2001): Korrespondenzanalyse. München: Oldenbourg
- Blasius, Jörg/Thiessen, Victor (2012): Assessing the Quality of Survey Data. London: Sage
- Bourdieu, Pierre (1982): Die feinen Unterschiede. Frankfurt a. M.: Suhrkamp
- Diekmann, Andreas (2006) (Hg.): Methoden der Sozialforschung. Wiesbaden: VS Verlag
- Dubben, Hans-Hermann/Beck-Bornholdt, Hans-Peter (2011): Der Hund, der Eier legt. Reinbek: Rowohlt
- Esping-Andersen, Gøsta (1990): The Three Worlds of Welfare Capitalism. Cambridge/Oxford: Polity/Blackwell
- Gifi, Albert (1990): Nonlinear Multivariate Methods. Chichester, UK: Wiley
- Long, Scott (1997): Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage
- Otte, Gunnar (2005): Entwicklung und Test einer integrativen Typologie der Lebensführung für die Bundesrepublik Deutschland. In: Zeitschrift für Soziologie 34: 442-467
- Reinecke, Jost (2005): Strukturgleichungsmodelle in den Sozialwissenschaften. München: Oldenbourg
- Vermunt, Jereon (1997): Log-Linear Models for Event Histories. Thousand Oaks, CA: Sage
- Wolf, Christof/Best, Henning (Hg.) (2010): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS Verlag

Steffen Kühnel und André Dingelstedt

Kausalvorstellungen, also Vermutungen darüber, welche Wirkungen oder welche Effekte eine Ursache hat, spielen im Alltag wie auch in der Wissenschaft eine große Rolle. In den empirischen Wissenschaften ist es ein zentrales Ziel, kausale Zusammenhänge aufzudecken und tatsächlich bestehende, *kausale Effekte* von nur vermeintlichen Effekten (*Scheinkausalitäten*) zu unterscheiden. So gibt es einen positiven Zusammenhang zwischen der Größe eines Feuers und der Anzahl der Feuerwehrleute, die den Brand bekämpfen. Zudem ist damit zu rechnen, dass die Höhe des durch ein Feuer verursachten Schadens mit der Größe des Feuers ansteigt. Daraus folgt, dass empirisch auch ein positiver Zusammenhang zwischen der Zahl der Feuerwehrleute und der Größe des Schadens besteht: Je mehr Feuerwehrleute eingesetzt werden, desto größer ist der Schaden. Diese positive Beziehung ist aber *keine* kausale Beziehung, sondern eine Scheinkausalität. Die Konsequenz, möglichst wenige Feuerwehrleute bei einem Feuer einzusetzen, um dadurch den Brandschaden zu reduzieren, wäre ausgesprochen kontraproduktiv.

Wenn nun – um das Beispiel des positiven empirischen Zusammenhangs zwischen dem Schaden bei einem Feuer und der Zahl der zur Brandbekämpfung eingesetzten Feuerwehrleute fortzuführen – aufgrund der Kenntnis des positiven Zusammenhangs bei großen Bränden weniger Feuerwehrleute eingesetzt würden, dann würde die empirische positive Beziehung zwischen der Zahl der Feuerwehrleute und der Höhe des Schadens nicht mehr zu beobachten sein. Stattdessen würden die wenigen Feuerwehrleute nicht mehr in der Lage sein, ein großes Feuer schnell zu löschen, so dass die Wahrscheinlichkeit noch größerer Schäden stiege. Der empirische Zusammenhang würde sich ändern, vermutlich sogar umkehren und negativ werden: Je weniger Feuerwehrleute eingesetzt werden, desto größer ist der Brandschaden.

Nach diesem Prinzip funktioniert die klassische *experimentelle Methode der Prüfung von Kausalität* (Eifler, Kapitel 11 in diesem Band): Durch eine systematische Manipulation der vermuteten kausal wirkenden Größe (hier: der Zahl der Feuerwehrleute) wird untersucht, ob ein vermuteter kausaler Zusammenhang tatsächlich besteht. Wenn im Beispiel

diese Manipulation unabhängig von der Größe des Feuers erfolgt, würde sich sehr schnell zeigen, dass die Zahl der Feuerwehrleute nicht oder allenfalls marginal die Schadenshöhe beeinflusst.

In den Sozialwissenschaften sind experimentelle Methoden allerdings nicht immer einsetzbar. Aus technischen, aber auch aus ethischen Gründen (Friedrichs, Kapitel 3 in diesem Band) lassen sich bestimmte Experimente nicht durchführen. So würden in dem gegebenen Beispiel Hausbesitzer und deren Versicherungen wenig Verständnis dafür haben, wenn ein wissenschaftliches Experiment mit der Zahl der bei einem Brand eingesetzten Feuerwehrleute stattfände. Aus gutem Grund ist es daher auch in der medizinischen Forschung Standard, geplante empirische Untersuchungen zunächst einer Ethik-Kommission vorzulegen. Neben sogenannten echten oder randomisierten Experimenten werden daher auch nicht-experimentelle Methoden verwendet, um kausale Beziehungen zu überprüfen (Stein, Przyborski/Wohlrab-Sahr, Strübing, Kapitel 7, 8 und 32 in diesem Band). Bevor wir auf verschiedene Ansätze zur Untersuchung von kausalen Beziehungen in der quantitativen Sozialforschung eingehen, soll der Begriff der „Kausalität“ diskutiert werden (zu Kausalität in der qualitativen Sozialforschung siehe Strübing, Kapitel 32 in diesem Band).

---

## 80.1 Was ist Kausalität?

Bereits die philosophischen Denker des Altertums beschäftigten sich mit der Frage nach den kausalen Ursachen von Ereignissen. So entwickelte Aristoteles im 4. Jahrhundert v. Chr. in Griechenland die Idee der „Causae“. Die Ideen über Ursache und Wirkung wurden im Zeitverlauf vielfältiger (hierbei sehr oft unter theologischen Aspekten), dabei nicht nur im europäisch-christlichen Raum, sondern auch im arabisch-islamischen Raum (z.B. al-Ghazali im 12. Jh. n. Chr. in Persien). Die moderne philosophisch-wissenschaftstheoretische Diskussion über Kausalität geht zurück auf David Hume (1711 - 1776) und John Stuart Mill (1806 - 1873). Diese Diskussion wird auch heute noch geführt, so z.B. von Wolfgang Stegmüller (1983).

In der *philosophischen Diskussion* wird Kausalität oft als eine spezifische Beziehung zwischen Ereignissen betrachtet. Dabei wird debattiert, wann die Aussage gerechtfertigt ist, ein Ereignis A (im obigen Beispiel das Ereignis „Einsatz von x Feuerwehrleuten“) sei kausale Ursache eines anderen Ereignisses B (im Beispiel das Ereignis „Schaden von y Euro“) oder kürzer: „A bewirkt B“. Wie muss die Beziehung zwischen A und B beschaffen sein, damit die Aussage „A bewirkt B“ zutrifft?

In den *empirischen Wissenschaften* geht es eher um die damit eng zusammenhängende Frage, warum ist B aufgetreten? Die Antwort „B ist aufgetreten, weil A aufgetreten ist“ wird hier allerdings als eine unzureichende Antwort für die Erklärung eines Ereignisses B angesehen. Die korrekte Antwort lautet: Es gibt ein zutreffendes Kausalgesetz, nach dem für alle Ereignisse des Typs A und des Typs B gilt: Wenn A auftritt, dann tritt auch B auf. Außerdem ist ein Ereignis des Typs A aufgetreten. Daher ist das Ereignis B aufgetreten.

Diese Argumentation folgt dem sog. *Hempel-Oppenheim-Schema* einer wissenschaftlichen Erklärung (Hempel/Oppenheim 1948).

### 80.1.1 Nicht-Singularität und raumzeitliche Ungebundenheit von Kausalgesetzen

Damit A Ursache von B sein kann, wird also die Gültigkeit eines Kausalgesetzes verlangt. Die Postulierung eines Kausalgesetzes setzt aber bereits eine Definition von Kausalität voraus, da ansonsten diese Forderung sinnlos (nichtssagend) wäre. Immerhin wird durch dieses Postulat klar, dass ein kausaler Zusammenhang über das Auftreten zweier einzelner, *singulärer* Ereignisse hinausgehen muss. Es ist ausgeschlossen, dass unter gleichen Randbedingungen ein kausaler Zusammenhang nur einmal und dann nie wieder auftritt. Wenn A in einer Situation auf B wirkt, dann muss dies auch in allen vergleichbaren Situationen so sein. Der Ausdruck „alle vergleichbaren Situationen“ bezieht sich auf alle Zeiten und alle Orte des Universums. Ein Kausalgesetz ist *raum-zeitlich ungebunden* oder *nicht-singulär*. Nicht ausgeschlossen ist allerdings, dass es sehr schwer sein kann, *vergleichbare* Situationen zu finden, bei der Ereignisse des Typs A und des Typs B auftreten.

### 80.1.2 Kausalität als Beziehung zwischen Variablen

Darüber hinaus wird in den empirischen Wissenschaften Kausalität eher nicht als eine Beziehung zwischen Ereignissen bzw. genauer zwischen Typen, Klassen oder Mengen von Ereignissen betrachtet, sondern als eine *Beziehung zwischen Variablen*: Die Variable X wirkt kausal auf die Variable Y. Der Wechsel von einer Ereignisklasse zu einer Variablen scheint trivial, wenn etwa das Ereignis A durch die Ausprägung X=1 der Variable X und das gegenteilige, *komplementäre* Ereignis nicht-A durch X=0 symbolisiert wird und entsprechend das Ereignis B durch Y=1 und nicht-B durch Y=0.

Unter einer „Variablen“ (Blasius/Baur, Kapitel 79 in diesem Band) versteht man in der Datenanalyse das Vorhandensein oder Nichtvorhandensein bzw. das Ausmaß des Vorhandenseins einer definierten *Eigenschaft* bei einem eindeutig festgelegten *Merkmalsträger*, technisch gesprochen bei einer *Untersuchungseinheit* (Häder/Häder, Kapitel 18 in diesem Band). Wird von Variablen gesprochen, muss also klar sein, auf welche Merkmalsträger sich die Eigenschaft bezieht. Im Beispiel der Zahl der Feuerwehrleute könnten sich die Variablen „Anzahl der Feuerleute“, „Schadenshöhe“ und „Größe des Feuers“ auf die Einheit „Brandfall“ beziehen. Der Bezug auf Variablen statt auf Ereignisse (Mika/Stegmann, Pötter/Prein, Kapitel 85 und 86 in diesem Band) kann daher als eine Präzisierung aufgefasst werden, der für die empirische Untersuchung von kausalen Zusammenhängen notwendig ist.

Durch den *eindeutigen Bezug auf Merkmalsträger* wird auch ein weiteres Problem gelöst, das in der philosophischen Diskussion von Kausalität angesprochen wird, dass nämlich die

Ereignisse, zwischen denen eine kausale Beziehung besteht, irgendwie miteinander verbunden sein müssen. Diese *Verbundenheit* kann z.B. darin bestehen, dass die Ereignisse A und B bzw. die Variablenausprägungen X=1 und Y=1 verschiedene Eigenschaften eines gemeinsamen Merkmalsträgers sein können. Die Forderung, dass sich die Variablen auf die gleichen Untersuchungseinheiten beziehen, ist allerdings zu eng. So wird etwa in der Mehrebenenanalyse (Pötschke, Kapitel 87 in diesem Band) davon ausgegangen, dass es Beziehungen zwischen den Einheiten unterschiedlicher Ebenen geben kann, also z.B. dass das Leistungsniveau in einer Schulklassie (X) einen Einfluss auf den Lernerfolg der einzelnen Schülerinnen und Schüler der Klasse (Y) hat. Wenn sich X und Y auf unterschiedliche Merkmalsträger beziehen, muss in der Datenanalyse deutlich werden, wieso die unterschiedlichen Merkmalsträger miteinander interagieren. Genau dies kann als Präzisierung der Verbundenheit von Ereignissen aufgefasst werden.

Neben dem klaren Bezug auf Merkmalsträger besteht ein weiterer Vorteil des Wechsels von kausalen Beziehungen zwischen Ereignissen zu kausalen Beziehungen zwischen Variablen darin, dass *funktionale Formen von Kausalbeziehungen* unterschieden werden können. Wenn X und Y metrische Variablen (Blasius/Baur, Kapitel 79 in diesem Band) sind, also jeweils das quantifizierbare Ausmaß einer Eigenschaft erfassen, dann kann zwischen linearen und nicht-linearen Kausalbeziehungen unterschieden werden. Bei einer *linearen Kausalbeziehung* ist das Ausmaß der durch X bewirkten Änderung bei Y stets proportional zur Änderung des Ausmaßes bei X. Bei einer *nicht-linearen Kausalbeziehung* wirkt sich eine gleich große Veränderung bei X dagegen in Abhängigkeit vom Ausgangswert unterschiedlich auf Y aus. Wenn X und Y ordinale Messniveau (Blasius/Baur, Kapitel 79 in diesem Band) aufweisen, kann danach unterschieden werden, ob die kausale Beziehung *positiv, negativ oder nichtmonoton* ist. Und selbst bei nur nominalskalierten Variablen (Blasius/Baur, Kapitel 79 in diesem Band) X und Y lassen sich Aussagen darüber formulieren, ob Y ausschließlich durch X oder auch durch andere und möglicherweise nur zusammen mit anderen Variablen kausal beeinflusst wird.

Dabei darf nicht der Fehler gemacht werden, etwas durch sich selbst erklären zu wollen. *Die Variable Y kann nicht ihre eigenen Ausprägungen erklären.* Wenn Rauch definitiver Bestandteil eines Feuers ist, dann kann das Vorhandensein von Rauch nicht durch das Vorhandensein von Feuer erklärt werden. Wenn jedoch die Zusammensetzung oder das Ausmaß von Rauch eine eigene Variable ist, die unabhängig von anderen Variablen gemessen werden kann, die Feuer kennzeichnen, dann können diese anderen Variablen (z.B. der Feuchtigkeitsanteil des verbrennenden Materials) Rauch kausal erklären.

### **80.1.3 Art der Kausalbeziehung: deterministisch oder statistisch?**

Uneinigkeit besteht in der Wissenschaft darin, ob eine kausale Beziehung immer *deterministisch* sein muss oder ob es auch *statistische Kausalbeziehungen* geben kann, so dass eine Veränderung bei einer kausal wirkenden Variable X nicht mit Sicherheit, sondern nur mit einer gewissen Wahrscheinlichkeit Änderungen bei einer kausal beeinflussten Variablen Y

hervorruft (Suppes, 1970). Selbst wenn man davon ausgeht, dass ein *beobachteter kausaler Effekt* nur mit einer gewissen Wahrscheinlichkeit auftritt, wie dies in der empirischen Sozialforschung nahezu immer unterstellt wird, kann gleichwohl angenommen werden, dass jede Kausalbeziehung deterministisch ist. So kann ein nichtdeterministischer Zusammenhang Folge davon sein, dass die Variable Y nicht nur durch die betrachtete Variable X, sondern durch weitere Variablen W, Z, ... kausal beeinflusst wird. Auch wenn alle kausal wirkenden Variablen zusammen die Ausprägungen der von ihnen beeinflussten Variable Y vollkommen determinieren, ist es möglich, dass eine Veränderung von X alleine einen (scheinbar) nichtdeterministischen Effekt aufweist. Alternativ kann die kausal beeinflusste Variable Y als eine echte Zufallsvariable betrachtet werden, deren Realisierungen durch einen zufallsgesteuerten Prozess generiert werden. Der deterministische Einfluss einer kausal wirkenden Variable X bezieht sich dann nicht auf die einzelnen Realisierungen von Y, sondern auf die Parameter ihrer Verteilung, also z.B. auf ihren Erwartungswert oder ihre Varianz.

Anders als bei der Frage, ob ein kausaler Effekt deterministisch sein muss oder statistisch sein kann, besteht in der Wissenschaft Einigkeit darüber, dass bei Kausalbeziehungen eine *zeitliche Reihenfolge* bestehen muss, nach der die Wirkung nicht vor der Ursache auftreten kann. Ob und wie lang der zeitliche Abstand ist, der zwischen Ursache und Wirkung auftritt, wird dagegen eher als eine empirische Frage betrachtet, die jeweils zu untersuchen ist. Um den zeitlichen Abstand zu berücksichtigen, können die in einer Kausalanalyse betrachteten Variablen mit einem Zeitindex versehen werden, der angibt, wann die Realisierung einer Variable beobachtet wird.

#### 80.1.4 Definition von Kausalität

In der *statistischen Ereignisanalyse* (Mika/Stegemann, Pötter/Prein, Kapitel 85 und 86 in diesem Band) wird Ereignis als eine Zustandsänderung definiert. Dieses Definitionsmerkmal scheint uns auch für den Begriff „Kausalität“ wesentlich. Von Kausalität sollte unserer Ansicht nach nur dann gesprochen werden, wenn auf eine Änderung der Ausprägung einer kausal wirkenden Variable X eine Änderung der Ausprägung der kausal beeinflussten Variablen Y oder der Verteilungsparameter der Verteilung von Y folgt.

Aus der Ereignisanalyse sollte aber auch die Vorstellung eines *absorbierenden Zustands* übernommen werden. Eine Veränderung in X kann dazu führen, dass bei Y eine finale Ausprägung hervorgerufen wird, die nach ihrer Realisierung bei dem betroffenen Merkmalsträger nicht mehr veränderbar ist. So kann das Auslösen eines Pistolschusses für den Getroffenen zum Tode führen.

Irrelevant für das Bestehen eines kausalen Effektes von X auf Y ist unserer Ansicht nach dagegen, dass die Ausprägungsveränderung bei X durch einen Akteur willentlich hervorgerufen sein muss. Ein tödlicher Pistolschuss könnte also auch zufällig ausgelöst worden sein.

Das Vorliegen einer kausalen Beziehung zwischen X und Y bedeutet nicht, dass X bzw. Y beobachtet werden oder auch nur beobachtet werden können. In der Datenanalyse wird zwischen beobachteten Variablen und latenten Variablen unterschieden (Blasius, Kapitel 83 in diesem Band). Wenn etwa der kausale Effekt variierender Lehrmethoden auf die Bildung einer Person untersucht wird, dann wird die Bildung nicht direkt beobachtet, sondern indirekt über Indikatoren (Burzan, Latcheva/Davidov, Kapitel 81 und 55 in diesem Band), z.B. über Wissensfragen erfasst. Ob eine latente Variable prinzipiell unbeobachtbar ist oder nur aktuell nicht beobachtet wurde, ist im Hinblick auf kausale Zusammenhänge irrelevant, kann aber für die Prüfung von kausalen Zusammenhängen von Bedeutung sein.

Zusammenfassend halten wir es für sinnvoll von Kausalität bzw. einer kausalen Beziehung zu sprechen, wenn auf Ausprägungsänderungen einer *erklärenden Variablen* X immer entweder Ausprägungsänderungen einer *abhängigen Variablen* Y oder aber immer Änderungen der Verteilung von Y folgen. Dabei spielt es keine Rolle, ob die Veränderungen bei X bzw. Y beobachtet werden.

---

## 80.2 Prüfung von kausalen Beziehungen und Schätzung kausaler Effekte

Eine klare Vorstellung darüber, was Kausalität ist, ist eine notwendige aber nicht hinreichende Bedingung dafür, um empirisch untersuchen zu können, ob Kausalität vorliegt und wenn ja, welche Art und Stärke eine kausale Beziehung aufweist. Das Beispiel der Zahl der bei einem Brand eingesetzten Feuerwehrleute und der Schadenshöhe zeigt, dass ein empirisch beobachteter Zusammenhang nicht notwendigerweise mit dem zugrundeliegenden kausalen Zusammenhang korrespondiert. Dies wird auch durch den Satz ausgedrückt, dass zwischen *Korrelation* und *Kausalität* Unterschiede werden muss.

Bezogen auf die für die Messung von kausalen Effekten oft verwendeten Regressionsmodelle wird dann auch davon gesprochen, dass die geschätzten Effekte „konfundiert“ seien. In einem Regressionsmodell (Blasius/Baur, Kapitel 79 in diesem Band) wird untersucht, ob und wie die Mittelwerte einer abhängigen Variable von den Werten einer oder mehrerer unabhängigen Variablen abhängen. *Konfundierung* liegt dann vor, wenn sich die Art und Höhe der Kausalbeziehung von der Art oder Höhe der empirisch-statistischen Regressionsbeziehung unterscheidet. Im Beispiel ergibt die bivariate Regression der Schadenshöhe auf die Zahl der Feuerwehrleute ein positives Regressionsgewicht. Nur wenn die Größe des Feuers als zusätzliche erklärende Variable im Regressionsmodell mit berücksichtigt wird, kann der kausale Effekt durch ein Regressionsmodellkorrekt geschätzt werden.

In den Sozialwissenschaften lassen sich gegenwärtig zwei unterschiedliche Hauptansätze unterscheiden, die beide versuchen, „echte“ kausale Beziehungen aufzudecken und von bloßen korrelativen Zusammenhängen zu unterscheiden. Der Ansatz des Vergleichs potentieller Ergebnisse basiert letztlich auf der Logik des Experiments (Eifler, Kapitel 11 in diesem Band). Der Ansatz der Kausaldiagramme entspricht dagegen eher der Tradition der Beobachtungsstudien (Ex-post-facto-Designs) (Stein, Kapitel 7 in diesem Band).

Tatsächlich werden aber beide Ansätze sowohl im Kontext von experimentellen wie nicht-experimentellen Studien eingesetzt.

### 80.2.1 Kausale Effekte als Vergleich potentieller Ergebnisse

Mit dem Aufkommen der modernen Naturwissenschaften hat sich das wissenschaftliche Experiment (Eifler, Kapitel 11 in diesem Band) als die Methode etabliert, mit der kausale Zusammenhänge zwischen einer erklärenden Variablen X und einer abhängigen Variablen Y empirisch untersucht werden können.

In einem Experiment werden bei den für eine empirische Untersuchung zur Verfügung stehenden Untersuchungseinheiten gezielt Ausprägungen der Variable festgesetzt, deren kausaler Effekt auf eine abhängige Variable untersucht werden soll. Im Kontext eines Experiments wird von einem *Treatment* gesprochen, dem die Untersuchungseinheiten ausgesetzt werden. Um einen kausalen Effekt zu erfassen, müssen mindestens zwei unterschiedliche Ausprägungen von X festgesetzt werden. Die erklärende Variable eines Experiments wird auch als *experimenteller Faktor* bezeichnet. Da die Ausprägungen eines experimentellen Faktors in der Untersuchung festgesetzt werden, werden durch das Experiment die beteiligten Untersuchungseinheiten unterschiedlichen Gruppen zugeordnet. Oft ist ein Treatment eine bestimmte Maßnahme, z.B. die Anwendung einer neuen Lehrmethode in einem Unterrichtsexperiment. Man spricht daher oft auch von einer *Treatment-* oder *Experimentalgruppe*, deren Angehörige einem (neuen) Treatment ausgesetzt sind, und von einer *Kontrollgruppe*, deren Angehörige dem (neuen) Treatment nicht ausgesetzt sind.

Die Idee der *potentiellen Ergebnisse definiert den kausalen Effekt* des Faktors X als über alle Merkmalsträger einer Population definiertes Maß der Unterschiedlichkeit der Ausprägungen von Y, wenn die Merkmalsträger der Treatmentgruppe oder aber der Kontrollgruppe zugeordnet worden wären. Da ein Merkmalsträger aber nicht *gleichzeitig* sowohl der Treatmentgruppe als auch der Kontrollgruppe angehören kann, ist es prinzipiell unmöglich, einen (so definierten) kausalen Effekt zu beobachten. Paul W. Holland (1986: 947) bezeichnet dies als *Fundamentalproblem von Kausalschlüssen*. Aufgrund der unmöglichen Beobachtung von Unterschieden zwischen realisierten und fiktiven Werten, wird auch von einem *kontrafaktischen Kausalitätsansatz* gesprochen. Auch wenn der Vergleich der beobachteten Ausprägungen mit den potentiellen Ausprägungen bei einer anderen Festsetzung einer Faktorausprägung nicht möglich ist, kann doch versucht werden, kausale Effekte „ersatzweise“ über einen Vergleich der Ausprägungen von Y in der Experimental- und der Kontrollgruppe zu bestimmen. Ein formaler Ansatz dazu besteht darin, anstelle einer abhängigen Variable Y (mindestens) zwei abhängige Variablen  $Y_c$  und  $Y_t$  („c“ für *control* und „t“ für *treatment*) zu definieren, die jeweils für alle Untersuchungseinheiten definiert ist, wobei  $Y_c$  der Wert der kausal beeinflussten Variable ist, wenn eine Untersuchungseinheit in der Kontrollgruppe ist oder wäre, und  $Y_t$  der entsprechende Wert, wenn eine Untersuchungseinheit in der Treatmentgruppe ist oder wäre. Der kausale Effekt kann dann z.B. als Differenz der Populationsmittelwerte dieser beiden Variablen geschätzt

werden. Eine konsistente Schätzung ergibt sich allerdings nur, wenn die fehlenden Werte der Ausprägungen von  $Y_c$  in der Treatmentgruppe und der Ausprägungen von  $Y_t$  in der Kontrollgruppe statistisch gesehen *ignorierbar* sind.

Da diese Sichtweise nach Holland (1986: 946) erstmals ausführlich bei Donald B. Rubin (1974) formuliert wurde, spricht Holland vom *Rubin-Model*. Rubin (2005: 324) selbst bezeichnet die Dissertation von Jerzy Neyman (1923) als erste Quelle, in der die Idee des Vergleichs potentieller Realisierungen explizit und formal betrachtet wurde. Auch wenn Neyman somit als Begründer dieser Forschungstradition gelten kann, wird die Formulierung der für die Schätzung kontrafaktischer Vergleiche notwendigen Anforderungen an ein experimentelles Design, insbesondere die Randomisierung (Eifler, Kapitel 11 in diesem Band), d.h. die zufällige Zuordnung der Faktorausprägungen auf die Untersuchungseinheiten, meist Ronald A. Fisher (1925, 1935) zugeschrieben.

Rubin (1974) diskutiert auch die Schätzung kausaler Effekte in Untersuchungen ohne randomisierte Kontroll- und Experimentalgruppen. Um sich dabei der Logik von Kontroll- und Experimentalgruppe maximal anzunähern, werden Fälle verglichen, die bei möglichst vielen *Kontrollvariablen* (*Kovariaten*) gleiche oder möglichst ähnliche Werte aufweisen. Die Ausprägungen der Kontrollvariablen müssen dabei im Zeitverlauf *vor* den Werten des zu analysierenden Einflussfaktors ihren jeweiligen Wert angenommen haben. Diese Idee wird auch bei der Schätzung von *propensity-scores* verfolgt (Rosenbaum/Rubin, 1983).

### 80.2.2 Kausale Effekte in Kausaldiagrammen

Der zweite Ansatz zur Prüfung und Schätzung von kausalen Zusammenhängen lässt sich auf die *Pfadanalyse* zurückführen (Wright, 1921). Die Logik der Pfadanalyse wurde zur Methodik *linearer Strukturgleichungsmodelle* weiterentwickelt (Blasius/Baur, Kapitel 79 in diesem Band), die lineare (Kausal-) Beziehungen modelliert und schätzt, wobei neben beobachteten Variablen auch latente Variablen (Blasius, Latcheva/Davidov, Kapitel 83 und 55 in diesem Band) berücksichtigt werden können (Kaplan, 2000). Auch beim Vorliegen von nominal skalierten Variablen können Strukturgleichungsmodelle mit und ohne latente Variablen formuliert werden (Hagenaars, 1990). Mit der Theorie *gerichteter azyklischer Diagramme* (*DAG: Directed Acyclic Graph*), die auch als *Bayessche Netzwerke* bezeichnet werden, wird die Logik der Pfadanalyse auf kausale Beziehungen angewendet, ohne dabei die Verteilungen der betroffenen Variablen spezifizieren zu müssen (Pearl, 2010).

Die Analyse von Kausaldiagrammen kann auf Beobachtungsstudien (Ex-post-facto-Designs) oder Experimente angewendet werden. Ausgangspunkt sind immer *theoretische Vermutungen über die (kausale) Beziehungsstruktur* zwischen Variablen, für die Messungen vorliegen, wobei die Gegebenheiten des Erhebungsdesigns bzw. des datengenerierenden Prozesses berücksichtigt werden. Die vermutete Kausalstruktur wird meist *grafisch* dargestellt, wobei ein gerichteter Pfeil die direkte kausale Beziehung von einer kausal wirkenden Variablen auf eine kausal beeinflusste Variable symbolisiert. In einem solchen Diagramm

können auch Ketten von Beziehungen entstehen, die als indirekte Kausalbeziehungen interpretiert werden. Eine in einer solchen Kette zwischen einer kausal beeinflussenden Variable und einer kausal beeinflussten Variable liegende Variable wird als intervenierende Variable oder *Mediatorvariable* bezeichnet. Wird von einer Beziehung zwischen zwei Variablen ausgegangen, deren Kausalrichtung unklar ist bzw. die durch unbeobachtete dritte Variablen ausgelöst wird, die beide Variablen kausal beeinflusst, kann dies durch einen ungerichteten (oder auch an *beiden* Enden gerichteten) Pfeil symbolisiert werden. Das gemeinsame Wirken mehrerer Variablen kann dadurch dargestellt werden, dass sich die Pfeile dieser nur interaktiv wirkenden Variablen zu einem Pfeil vereinigen, der auf die kausal abhängige Variable gerichtet ist.

Im zweiten Analyseschritt werden die *Konsequenzen der postulierten Kausalstruktur* analysiert. Dabei gibt es Unterschiede zwischen der parametrischen Analyse von Strukturgleichungsmodellen und der nichtparametrischen Analyse von Bayesschen Netzwerken. Bei Strukturgleichungsmodellen wird zunächst geprüft, ob die Parameter des Modells identifiziert sind. Ist dies der Fall, werden sie mit vorliegenden Stichprobendaten geschätzt. Dabei wird auch geprüft, ob das postulierte Modell mit den Stichprobendaten vereinbar ist. Ist dies nicht der Fall, wird das Modell verworfen. Die postulierten Kausalbeziehungen gelten dann als falsifiziert. Sind Kausalmodell und Daten dagegen vereinbar, können Modellparameter als kausale Effekte interpretiert werden. In linearen Kausalmodellen ist es zudem möglich, den totalen Effekt einer kausal wirkenden Variablen in direkte und indirekte Effekte aufzuteilen. So kann z.B. das Vermögen von Eltern direkte und indirekte Effekte auf das Vermögen ihrer Kinder haben: Der Effekt ist direkt, wenn das elterliche Vermögen etwa durch Erbschaft übertragen wird; der Effekt ist indirekt, wenn das Vermögen der Eltern in eine gute Bildung der Kinder investiert wird und die gute Bildung der Kinder diesen zum Aufbau eigenen Vermögens verhilft. Eine vollständige Zerlegung einer empirischen Beziehung in direkte und indirekte Effekte ist allerdings nicht möglich, wenn in einem Modell ungerichtete Pfade vorkommen. Dann gibt es neben direkten und indirekten Effekten auch korrelierte Effekte, die streng genommen keine Effekte, sondern rein korrelative Beziehungen sind. Korrelierte Effekte können auch Folge von Scheinkausalität sein, wenn eine Variable X sowohl Y als auch W kausal beeinflusst. Im Unterschied zu korrelierten Effekten als Folge von ungerichteten Pfaden ist diese Art von Scheinkausalität aber kausal aufgelöst.

Da Bayessche Netzwerke einen nichtparametrischen Ansatz verfolgen, geht es hier zunächst darum festzustellen, ob und welche Kausalbeziehungen in dem Sinne identifiziert sind, dass sie empirisch geschätzt werden können, wenn auf vorgegebene Werte der kausal wirkenden Variablen konditioniert wird bzw. diese Werte von außen festgesetzt werden. Schätzbare Effekte können dann geschätzt und dadurch gegebenenfalls falsifiziert werden. Außerdem kann untersucht werden, ob bei Kontrolle von Modellvariablen Unabhängigkeit zwischen zwei Variablen bestehen muss. Auch dies kann dann überprüft werden. Bei nicht schätzbaren Effekten ist zu überlegen, ob sich durch Modifikation des Untersuchungsdesigns ein datengenerierendes Kausalmodell ergibt, dass zu identifizierbaren Kausalrelationen führt. Allerdings muss darauf hingewiesen werden, dass oft die

Möglichkeit besteht, dass zwei oder mehrere *konkurrierende* Kausalstrukturen modelliert werden, welche ähnlich gut mit den empirischen Daten vereinbar sind. Wenn verschiedene Kausalmodelle in allen denkbaren Stichproben stets in gleicher Weise mit empirischen Daten vereinbar sind, dann sind die Modelle *beobachtungäquivalent*. Dann kann die Wahl der zu bevorzugenden Kausalstruktur oft nur nach rein theoretischen Gesichtspunkten entschieden werden.

---

### 80.3 Diskussion

Die Durchführung eines Experiments (Eifler, Kapitel 11 in diesem Band) wird in allen empirischen Wissenschaften als die Methode verstanden, mit deren Hilfe kausale Zusammenhänge am sichersten untersucht werden können. Bisweilen wird sogar anderen Untersuchungsdesigns abgesprochen, Kausalbeziehungen prüfen zu können. Aber auch Experimente können kausale Beziehungen nicht beweisen. Der kontrafaktische Ansatz zeigt vielmehr, dass jedes Experiment mit Annahmen verbunden ist, die falsch sein können (Rubin 2005: 324). Wenn sich die Unterschiede zwischen Experimental- und Kontrollgruppe nicht auf die potentiellen Ergebnisse verallgemeinern lassen, die sich einstellen würden, wäre die Gruppenzuteilung der Untersuchungseinheiten eine andere, dann kann auch ein randomisiertes Experiment einen kausalen Effekt nicht konsistent erfassen.

Hier stellt sich allerdings die Frage, ob die kontrafaktische Definition von kausalen Effekten nicht möglicherweise unangemessen streng ist. In unserer Diskussion über Kausalität haben wir es vermieden, Kausalität kontrafaktisch zu definieren. Stattdessen greifen wir auf die Idee von Zustandsänderungen zurück. Angewendet auf ein Experiment würde dies allerdings bedeuten, dass während eines Experiments die Ausprägungen des untersuchten Faktors nicht nur einmal, sondern mindestens zweimal gesetzt werden und die Ausprägungen der abhängigen Variable auch mindestens zweimal gemessen werden müssten. Tatsächlich wird die Logik des Vergleichs von Veränderungen bei erklärenden wie abhängigen Variablen bei einigen Modellen der *Panelanalyse* (Schupp, Kapitel 73 in diesem Band) eingesetzt, wenngleich dort in der Regel die Realisierungen der erklärenden Variablen nicht experimentell gesetzt, sondern nur beobachtet werden. Das mehrfache Messen oder gar Festsetzen der Werte kann aber auch problematisch sein, als es den zu untersuchenden Kausalprozess beeinflussen kann. Die *Reaktivität* der Messung (Krebs/Menold, Kapitel 30 in diesem Band) kann dadurch erhöht werden und zu einer Verzerrung der Messung eines kausalen Effektes führen. Der kontrafaktische Vergleich von potentiellen mit aktuellen Messungen kann daher als Methode verstanden werden, die Reaktivität zu minimieren. Kontrafaktizität sollte unserer Ansicht aber nicht Bestandteil der Definition von Kausalität sein.

Randomisierten Experimenten wird in den Sozialwissenschaften zwar eine hohe *interne Validität* zugesprochen, aber oft zulasten der *externen Validität* (Krebs/Menold, Kapitel 30 in diesem Band). Gemeint ist hiermit, dass die experimentell gewonnenen Ergebnisse zu einem kausalen Zusammenhang zwar für die in einem Experiment beobachteten

Untersuchungseinheiten Geltung beanspruchen können, intern valide sind, aber nicht auf beliebige andere Merkmalsträger verallgemeinert werden können und somit nicht extern valide sind (Cook/Campbell, 1979). Tatsächlich wird hier ein Problem angesprochen, dass alle Untersuchungen über kausale Effekte betrifft und nicht nur Experimente. In unserer Diskussion des Begriffs Kausalität haben wir gefordert, dass kausale Beziehungen raumzeitlich ungebunden sind. Daher kann eine kausale Beziehung auch nicht bewiesen, sondern höchstens widerlegt werden. Auf der anderen Seite sollte es irrelevant sein, welche Merkmalsträger für eine Untersuchung ausgewählt werden, da Kausalität ja nur vorhanden oder aber nicht vorhanden sein kann, aber nicht nur in bestimmten Situationen auftreten kann.

Tatsächlich sollte das Problem der externen Validität unserer Ansicht nach nicht auf die Auswahl der Untersuchungseinheiten bezogen werden, sondern auf die Situation, in der ein kausaler Effekt auftritt. Wenn ein kausaler Effekt in einem Experiment mit Studierenden andere Ergebnisse aufweist, als in einem Experiment mit Personen ohne Schulabschluss, dann ist die unterstellte Kausalbeziehung streng genommen falsifiziert. Tatsächlich dürfte in solchen Fällen meist ein Interaktionseffekt des experimentell manipulierten Faktors mit anderen Faktoren vorliegen, die im Experiment nicht erfasst wurden. Der Ansatz der Kausaldiagramme, zunächst alle als relevant vermuteten Kausalbeziehungen zu formulieren, kann hier hilfreich sein. Werden vor der Durchführung einer Untersuchung Interaktionseffekte vermutet, sollte ein Untersuchungsdesign in der Lage sein, Daten zu generieren, mit deren Hilfe die postulierten Interaktionseffekte schätzbar sind.

Bei unserer Diskussion der Bedeutung von Kausalität haben wir darauf hingewiesen, dass es eine empirische Frage ist, welcher zeitliche Abstand zwischen der Änderung bei X und der Änderung bei Y notwendig ist, um den Effekt von X auf Y zu erfassen. Leider wird unserer Ansicht nach unabhängig vom verfolgten Ansatz der Untersuchung kausaler Effekte dieser Frage zu wenig Aufmerksamkeit geschenkt. So kann in einem Experiment der zeitliche Abstand zwischen dem Setzen des Treatments und der Messung des Effekts zu klein sein, um einen tatsächlich bestehenden kausalen Effekt zu erfassen. Ist ein kausaler Effekt nur von kurzer Dauer, kann der Abstand auch zu groß sein. In Beobachtungsstudien werden selbst bei korrekter Spezifikation der Kausalstruktur kausale Effekte mittels Regressions- oder Strukturgleichungsmodellen verzerrt geschätzt, wenn die Daten nicht im *Gleichgewichtszustand* sind, d.h. alle Effekte einer erfolgten Änderung in X auf Y abgeschlossen sind.

Letztlich gibt es unserer Ansicht nach keinen Königsweg bei der Untersuchung kausaler Effekte. Das heißt aber nicht, dass die Untersuchung kausaler Beziehungen ein sinnloses Unterfangen ist. Im Gegenteil: ohne den Versuch, kausale Beziehungen aufzudecken, kann es keinen Erkenntnisfortschritt geben. Eine Forscherin oder ein Forscher sollte sich nur darüber klar sein, dass Bemühungen stets scheitern können.

## Literatur

- Al-Ghazālī, Abū Hāmid Muhammad ibn Muhammad (1997): The Incoherence of the Philosophers. Provo: Brigham Young University Press
- Aristoteles (1995): Physik: Vorlesung über die Natur. Philosophische Schriften 6. Hamburg: Felix Meiner Verlag
- Cook, Thomas D./Campbell, Donald T. (1979): Quasi-Experimentation. Design & Analysis Issues for Field Settings. Chicago: Rand McNally College Publishing Company
- Fisher, Ronald A. (1925): Statistical Methods for Research Workers. London: Oliver & Boyd
- Fisher, Ronald A. (1935): Design of Experiments. Edinburgh: Oliver & Boyd
- Hagenaars, Jaques A. (1990): Categorical longitudinal data: log-linear panel, trend and cohort analysis. Newbury Park: Sage
- Hempel, Carl G./Oppenheim, Paul (1948): Studies in the Logic of Explanation. In: Philosophy of Science 15: 135-175
- Holland, Paul W. (1986): Statistics and Causal Inference. In: Journal of the American Statistical Association 81: 945-970
- Hume, David (2007): Eine Untersuchung über den menschlichen Verstand. Frankfurt am Main: Suhrkamp
- Kaplan, David (2000): Structural Equation Modeling: Foundations and Extensions. Sage: Newbury Park
- Mill, John S. (1974): A System of Logic. Ratiocinative and Inductive. Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation. Toronto/Buffalo: University of Toronto Press
- Neyman, Jerzy (1923): On the Application of the Probability Theory to Agricultural Experiments. Essays on Principles, Section 9. In: Statistical Science 5: 465-480
- Pearl, Judea (2010): Causality; Models reasoning, and inference. Cambridge/New York/Melbourne/Madrid/Cape Town/Singapore/Sao Paulo/Delhi/Dubai/Tokyo: Cambridge University press
- Rosenbaum, Paul R./Rubin, Donald B. (1983): The central role of the propensity score in observational studies for causal effects. In: Biometrika 70: 41-55
- Rubin, Donald B. (1974): Estimating causal effects of Treatments in Randomized and Nonrandomized Studies. In: Journal of Educational Psychology 66: 688-701
- Rubin, Donald B. (2005): Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. In: Journal of the American Statistical Association 100: 322-331
- Stegmüller, Wolfgang (1983): Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band I: Erklärung, Begründung, Kausalität. Zweite, erweiterte und verbesserte Auflage. Berlin/Heidelberg/New York: Springer
- Suppes, Patrick (1970): A Probabilistic Theory of Causality. Amsterdam: North-Holland Publishing Company
- Wright, Sewall (1921): Correlation and causation. In: Journal of Agricultural Research 20: 557-585

Nicole Burzan

Oft sind im Rahmen der quantitativen empirischen Forschung Sachverhalte aus Forschungsfragen und Hypothesen so allgemein bzw. mehrdeutig, dass sie nicht direkt messbar sind (z.B. Sachverhalte wie „gesellschaftliche Teilhabe“ oder „Mobbing“). Sie müssen erst operationalisiert werden, das heißt, es sind genaue Angaben dazu erforderlich, wie ein Sachverhalt gemessen werden soll. *Indikatoren*, kurz gesagt sind dies *Anzeiger für Sachverhalte*, spielen hierbei eine zentrale Rolle. Dies soll zunächst an einem Beispiel gezeigt werden, bevor zentrale Begriffe erläutert und die Einordnung in den Forschungsprozess sowie methodische Herausforderungen systematisiert werden.

---

## 81.1 Ein Beispiel: Indikatoren zur Messung von Armut

Bezogen auf das *Beispiel Armut* könnte ein (quantitativ vorgehender) Forscher sich beispielsweise nicht einfach ins Feld begeben und Personen fragen: „Sind Sie arm?“ oder „Wie arm sind Sie?“ oder ihnen intuitiv ihre Armut „ansehen“, ohne zuvor definiert zu haben, was er unter Armut verstehen will. Man unterscheidet in der Armutsforschung etwa ein physisches Existenzminimum (*absolute Armut*) und ein soziokulturelles Minimum (*relative Armut*). Während sich die absolute Armut auf unzureichend befriedigte existentielle Bedürfnisse in Bezug auf Nahrung, Kleidung und Wohnen richtet, bezieht sich die relative Armut auf die deutliche Unterschreitung des mittleren Lebensstandards in dem Land oder der Region, in der man lebt (Hradil 2010). Eine solche *Definition* konkretisiert den Sachverhalt zwar, sagt aber noch nicht zwingend etwas darüber aus, wie man ihn ganz genau misst.

Dabei ziehen unterschiedliche Armutdefinitions unterschiedliche *Messkonzepte* nach sich: Der Versuch, sich der absoluten Armut über den Indikator anzunähern, dass jemand weniger als zwei Dollar pro Tag zur Verfügung hat, bildet einen großen Teil der z.B. für westliche Industrieländer relevanteren relativen Armut noch nicht ab. Diese misst

man meist entweder über das Ressourcen- oder das Lebenslagenkonzept. Ein *Indikator für relative Armut* nach dem *Ressourcenkonzept* könnte etwa die Unterschreitung eines bestimmten Einkommens oder das Fehlen von weiteren Ressourcen wie finanzieller Rücklagen oder bestimmter Wohnungsausstattungen (z.B. einer Tiefkühltruhe) sein. Ein üblicher *Indikator für die Einkommensarmut* ist der Anteil der Haushalte, die weniger als 60% des (gemessen am Median) mittleren Nettoäquivalenzeinkommens aller Haushalte zur Verfügung haben. Das Äquivalenzeinkommen, ein gewichtetes Pro-Kopf-Einkommen, berücksichtigt die Anzahl und die Altersstruktur (unter/ab 14 Jahre) der Personen im Haushalt, da man von Einspareffekten bei Mehrpersonenhaushalten ausgeht, z.B. brauchen drei Personen in einem Haushalt nicht drei Kühlschränke.

Diese kurzen Hinweise zeigen, dass auch ein prinzipiell eindeutiger Indikator wie „Unterschreitung einer Einkommensgrenze“ noch weiterer Festlegungen bedarf (z.B. die Gewichtung von Personen in den Haushalten und die Bestimmung der Grenze von „60% des mittleren Einkommens“). Es wird deutlich, dass der ohnehin stark normativ besetzte Armutsbegriff längst nicht so eindeutig ist, wie man im Alltag glauben könnte.

Im anderen Messkonzept, dem *Lebenslagenansatz*, spielen grob gesagt zusätzlich Teilhabeaspekte eine Rolle, neben der Teilhabe am Konsum z.B. die Teilhabe in den Bereichen der Erwerbsarbeit, der Bildung und der Gesundheit. Die „Teilhabe im Gesundheitsbereich“ beispielsweise wäre eine Dimension von Armut nach dem Lebenslageansatz, die sich nicht direkt messen lässt. Dafür braucht man Indikatoren, die direkt messbar sind, z.B. den Indikator „Teilnahme an ärztlichen Vorsorgeuntersuchungen in den letzten zwölf Monaten“.

Die Beispiele verdeutlichen, dass es durchaus ganz deutlich vom Indikator und seiner genauen Konkretisierung abhängt, welches Ergebnis man erzielt, wie viele Menschen also als arm gelten. Der Anteil von Haushalten unterhalb einer bestimmten Einkommensgrenze kann etwa ein ganz anderer sein als der Anteil der Haushalte ohne finanzielle Rücklagen und ohne moderne Heizungsanlage (Burzan 2007).

Dass schon die Festlegung der Einkommensarmut je nach Konkretisierung des Indikators unterschiedlich ausfallen kann, zeigt die Sozialberichterstattung Nordrhein-Westfalens (Sozialberichte NRW Online 2013). Bereits die *Operationalisierung* von Armut als Einkommensarmut stellt die erste methodisch-konzeptionelle Entscheidung dar. Die Darstellung unterscheidet weiterhin ein altes und ein neues Verfahren, durch das man sich ab 2012 dem EU-Standard angepasst hat. Der Indikator, die Armutsrisikoquote, lautet nach der älteren Berechnung: Anteil der Haushalte, deren Einkommen weniger als 50% des arithmetischen Mittels des Äquivalenzeinkommens (nach alter OECD-Skala) der NRW-Bevölkerung beträgt. Die neue Berechnung dagegen misst den Anteil der Haushalte, die über ein Einkommen von weniger als 60% des Medians des Äquivalenzeinkommens (nach neuer OECD-Skala) der NRW-Bevölkerung verfügen.

Ein vergleichbarer Indikator, die Armutsrisikoquote, wird also „nur“ durch zwei verschiedene Berechnungsarten als jeweilige Konkretisierung dieses Indikators ermittelt, wobei sich die Einkommensgrenze (50% oder 60%), die Berechnung des mittleren Einkommens (nach dem arithmetischen Mittel oder dem Median) und schließlich die Berechnung des Äquivalenzeinkommens unterscheiden. In der Folge zeigen sich z.B. für 2011

deutliche Differenzen: Danach sind bei gleicher Datenbasis nach dem alten Verfahren 14,8% der NRW-Bevölkerung einkommensarm, nach dem neuen sind es 15,8%. Noch gravierender ist der Unterschied z.B. für die unter 18-Jährigen: Deren Armutsriskoquote ist zwar jeweils überdurchschnittlich hoch, beträgt nach alter Berechnungsart jedoch 25,7% und nach der neuen „nur“ 21,6%.

Der Skeptiker könnte aus diesen unterschiedlichen Befunden folgern, er oder sie habe immer schon gewusst, dass empirische Befunde manipulationsanfällig sind – je nachdem, ob man Armut als politisches Problem herausstellen oder die Gefahr abschwächen möchte, bedient man sich eben anderer Methoden oder anderer Indikatoren. Die Kritik greift jedoch zu kurz, da Forschende nicht die Möglichkeit haben, einen „objektiven“ goldenen Mittelweg zu gehen, der Armut weder überproblematisiert noch verharmlost, sondern sie müssen sich für einen Indikator bzw. mehrere Indikatoren entscheiden. *Wichtig ist es, die Indikatoren möglichst gut theoretisch bzw. inhaltlich zu begründen, seine Entscheidungen transparent zu machen und so herauszustellen, unter welchen Bedingungen die empirischen Befunde entstanden sind.* Der Hinweis auf die internationale Vergleichbarkeit im Sozialbericht des Landes NRW ist dabei zwar nicht zwingend eine hinreichende Begründung, aber dennoch ein wichtiger Aspekt. Auf der Basis transparenter Indikatoren kann dann sachlich-konstruktive Kritik zugunsten der Optimierung der Gültigkeit und damit weiterer Erkenntnisgewinne ansetzen.

---

## 81.2 Begriffe

Kurz gefasst kann man sagen: *Indikatoren* sind Anzeiger für Sachverhalte. Sie haben die Eigenschaft, direkt beobachtet oder gemessen werden zu können und weisen auf Sachverhalte hin, die nicht direkt beobachtbar sind. Dabei kann der mit Hilfe einer Fragebogenfrage erhebbare Indikator „Haushaltsnettoeinkommen des letzten Monats“ zur Messung von Armut verwendet werden. Diese Direktheit macht den Indikator aus und unterscheidet ihn von den allgemeineren Dimensionen bzw. Aspekten einer Forschungsfrage (wie „Armut“ im obigen Beispiel).

Indikatoren sind durch *Korrespondenzregeln* mit der Dimension bzw. dem Begriff oder Sachverhalt, den sie anzeigen sollen, verbunden. Während der Indikator bestimmt, *was* den Sachverhalt anzeigt (im Beispiel das Haushaltseinkommen), legt die Korrespondenzregel fest, *wie* der Indikator mit dem Sachverhalt verbunden ist. Dies geschieht oft durch wenn-dann- oder je-desto-Verknüpfungen. Im genannten Beispiel könnte die Regel lauten: „Wenn jemand ein Haushaltseinkommen unter X Euro hat, dann ist er arm“. Diese Korrespondenzregeln sind Festlegungen, die den Forschungsprozess transparent machen und zeigen, dass die Messungen sinnvoll mit der Forschungsfrage verknüpft sind. Diese Regeln sind aber nicht selbst Gegenstand der empirischen Überprüfung, es wird z.B. nicht geprüft, ob ein bestimmtes Haushaltseinkommen „wirklich“ zu Armut führt – für diese Frage werden andere bzw. weitere Indikatoren für Armut benötigt. Bei der Festlegung der Indikatoren und Korrespondenzregeln handelt es sich damit nicht allein um die

Umsetzung methodischer Vorgehensweisen, sondern um wichtige theoretisch-konzeptuelle Entscheidungen der Forscherinnen und Forscher.

Werden verschiedene Indikatoren zu einem Sachverhalt zusammengefügt, spricht man von einem *Index* (Latcheva/Davidov und Blasius, Kapitel 55 und 83 in diesem Band). So könnte etwa Ressourcenarmut durch verschiedene Indikatoren gemessen werden, z.B. durch das Einkommen, die finanziellen Rücklagen und die Ausstattung der Wohnung mit moderner Dämmung und Heizung, wobei diese Teilindikatoren unterschiedlich gewichtet werden können.

Der Begriff *Indikator* wird auch im Ausdruck *soziale Indikatoren* benutzt. Soziale Indikatoren sind statistische Maßzahlen, die gesellschaftlich bzw. gesellschaftspolitisch relevante Sachverhalte – oft im Rahmen der Wohlfahrts- und Lebensqualitätsforschung – anzeigen. Beispiel für einen solchen Indikator sind die Säuglingssterblichkeit oder die Armutsrisikoquote. „Soziale Indikatoren“ sind auch Titel und Thema einer Sektion in der Deutschen Gesellschaft für Soziologie (2013; vgl. zum Zugang zu Daten auch GESIS (2013)).

---

### **81.3 Einordnung in den Forschungsablauf**

Die Entwicklung von Indikatoren erfolgt im Rahmen einer empirischen Untersuchung in der Phase der Operationalisierung (Stein, Kapitel 7 in diesem Band). Ein wichtiges Kennzeichen der quantitativen Forschung besteht darin, dass Forschende vor der Datenerhebung genau festlegen, was sie und wie sie ihren Gegenstand erforschen wollen. Dies geschieht oft in Form von *Hypothesen* oder *konkreten Forschungsfragen*, die dann in der *Operationalisierungsphase* in konkrete *Erhebungsprozeduren* übersetzt werden müssen. Dazu gehört etwa die Festlegung des Erhebungsinstruments, z.B. einer Befragung (Reinecke, Kapitel 44 in diesem Band) oder Inhaltsanalyse (Mayring/Fenzl, Kapitel 38 in diesem Band). Bevor man sich jedoch genaue Fragebogenformulierungen, Reihenfolgen oder auch Beobachtungskategorien etc. überlegt, ist es entscheidend, den Hypothesen und Forschungsfragen Indikatoren und Korrespondenzregeln zuzuordnen. Es besteht ansonsten die Gefahr, nicht mehr hinreichend systematisch vorzugehen, so dass man z.B. bei manchen Fragen und ihren Antwortmöglichkeiten nicht mehr weiß, wozu genau man das noch einmal wissen wollte: Was bedeutet es beispielsweise für die Frage nach der Ressourcenarmut, wenn 60% der Befragten angeben, dass sie gern öfter Kleidung kaufen würden?

Teilweise scheint der Indikator recht banal bzw. der Zwischenschritt der Indikatorenbildung zwischen Sachverhalten und Fragen kaum notwendig zu sein (so dass das Thema „Indikatoren“ auch in einigen methodischen Lehrbüchern recht kurz kommt), wenn z.B. der Sachverhalt „Geschlecht der befragten Person“ mit den üblichen binären (aber auch hier: hinterfragbaren) Antwortkategorien „weiblich“ und „männlich“ gemessen wird. In vielen Fällen ist es jedoch zugunsten der Systematik des Vorgehens hilfreich, Indikatoren und Korrespondenzregeln *vor* der detaillierten Entwicklung des Erhebungsinstruments festzulegen, am besten in systematischer tabellarischer Form mit der Zuordnung zu den

entsprechenden Sachverhalten. Dies ist zudem nützlich für die Datenauswertung und Interpretation, weil dadurch die Zuordnung der statistischen Ergebnisse zur inhaltlichen Ebene sowie die Reflexion des eigenen methodischen Vorgehens ermöglicht bzw. erleichtert werden.

In der qualitativen Forschung (Przyborski/Wohlrab-Sahr, Kapitel 6 in diesem Band) spielt die Bildung von Indikatoren keine wesentliche Rolle, weil dort vor der Datenerhebung die theoretischen Konzepte nicht konkret festzulegen und zu operationalisieren sind.

---

## 81.4 Herausforderungen bei der Entwicklung (gültiger) Indikatoren

Die Entwicklung von Indikatoren berührt ein zentrales Gütekriterium empirischer Forschung und zwar die *Gültigkeit* (oder: *Validität*; vgl. Krebs/Menold, Kapitel 30 in diesem Band). Ergebnisse sind gültig, wenn der Forscher das gemessen hat, was er messen wollte. Ein zentrales Erfordernis hierfür besteht neben anderen darin, *angemessene Indikatoren* zu nutzen.

Ein Gegenbeispiel wäre etwa, wenn man in westlichen Gesellschaften für den Sachverhalt „Armut“ den Indikator „Besitz eines Fernsehers“ festlegen würde. Weder zeigt allein der Besitz eines (möglicherweise alten) Fernsehers an, dass die Personen im Haushalt insgesamt über Ressourcen verfügen, die einen Lebensstandard oberhalb einer soziokulturellen Armutsgrenze ermöglichen, noch weist der Nicht-Besitz eines Fernsehers (z.B. aus dezidiert mangelndem Interesse an diesem Medium) zwingend auf eine Armutslage hin. Der Fernseherbesitz wäre also allein kein angemessener Indikator, um den Sachverhalt „Armut“ anzuzeigen, damit wäre die Gültigkeit der Messung nicht gegeben. Das Beispiel zeigt wiederum, in welch hohem Maße methodische Entscheidungen kein mechanisches Anwenden erprobter Verfahren darstellen, sondern mit inhaltlich-konzeptionellen Aspekten des zu erforschenden Themas eng verbunden sind. Auch beim Indikator „Einkommen“ für den Sachverhalt „Armut“ ließe sich hinterfragen, ob das Einkommen allein ohne beispielsweise die Berücksichtigung der Ausgabenseite (z.B. Kosten für eine chronisch kranke Person im Haushalt, Notwendigkeit von Mietzahlungen gegenüber dem Wohnen in einem geerbten Haus) einen gültigen und hinreichenden Indikator für die empirische Erfassung von Armut darstellt.

Um gültige Indikatoren festzulegen, müssen Forschende sich also gut im Themenfeld der Forschung auskennen. Für einige Sachverhalte, für die häufig Indikatoren benötigt werden (z.B. das Merkmal „Bildung“), sind zudem gewisse Standards als Orientierungshilfe entwickelt worden, so dass empirische Befunde zumindest zum Teil vergleichbar sind (vgl. für demographische Standards Statistisches Bundesamt 2010, Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band). Solche Standards sind andererseits weder unveränderlich noch für jeden Kontext in gleicher Weise geeignet. Über diese allgemeine Anforderung hinaus sollen an dieser Stelle beispielhaft drei weitere Herausforderungen dafür behandelt werden, angemessene Indikatoren zu finden.

Eine Herausforderung besteht darin, *dass der Indikator mehr oder weniger direkt mit dem Sachverhalt, den er anzeigen soll, verbunden sein kann*. Je enger der Indikator mit der Definition des Sachverhalts verknüpft ist, desto eher sind gültige Ergebnisse erwartbar. Wenn etwa Armut als Einkommensarmut definiert wird, liegt es nahe, das Einkommen als Indikator für Armut zu setzen; es müsste dann zusätzlich noch bestimmt werden, dass es etwa um das monatliche Nettoäquivalenzeinkommen des Haushalts geht, das eine bestimmte Grenze unterschreitet (wobei noch darüber gestritten werden könnte, ob Einkommensarmut eine hinreichende Definition für Armut im Allgemeinen darstellt). Wenn man Armut hingegen z.B. daran mäße, wie gesund oder ungesund im Sinne bestimmter Nahrungsbestandteile sich jemand ernährt – etwa aufgrund der Annahme, dass sich ärmere Haushalte gesundes Essen nicht leisten können –, wäre der Indikator nur indirekt mit dem Sachverhalt „Armut“ verbunden und die Gültigkeit gefährdet. Kromrey (2009: Kap. 4.1) spricht im Zusammenhang dieser abgestuften Verknüpfung von Indikator und Sachverhalt auch von *defitorischen, korrelativen und schlussfolgernden Indikatoren*.

Nun ist es allerdings nicht immer möglich, eine möglichst direkte Verbindung zwischen Indikator und Sachverhalt herzustellen und dadurch indirekte Indikatoren zu vermeiden. Dies ist etwa bei Aspekten sozialer Erwünschtheit (Hlawatsch/Krickl, Kapitel 20 in diesem Band) der Fall. Soll beispielsweise gemessen werden, wie viel Aufmerksamkeit Eltern ihren Kindern im Vorschulalter entgegenbringen, kann man u.a. direkte Fragen stellen, z.B. „Wie oft lesen Sie ihrem Kind etwas vor?“ (mit der Korrespondenzregel: „Je häufiger die Eltern dem Kind etwas vorlesen, desto mehr Aufmerksamkeit bringen sie dem Kind entgegen“). Allerdings werden Eltern die Häufigkeit solcher Handlungen in ihrer Erinnerung eher über- als unterschätzen, um nicht als „schlechte Eltern“ zu gelten. Um dieses Problem sozialer Erwünschtheit zu umgehen, könnte der/die Forschende alternativ z.B. auf einem Spielplatz Eltern unaufdringlich beobachten. Jedoch hätte er damit einen indirekten Indikator (Vater spricht X-mal mit dem Kind, putzt ihm die Nase etc.) festgelegt, der die Aufmerksamkeit über die Spielplatzsituation hinaus nicht direkt misst – und spätestens eine verdeckte Beobachtung in nicht-öffentlichen Situationen wie der eigenen Wohnung wäre nicht nur praktisch schwierig, sondern forschungsethisch auch nicht statthaft. Das Beispiel zeigt an, dass es oft nicht möglich ist, ideale Lösungen für methodische Probleme zu finden, ohne dass sich im Gegenzug andere methodische Probleme vergrößern: In diesem Fall hätte man das Problem sozialer Erwünschtheit gegen die Probleme indirekter Indikatoren und ggf. forschungsethischer Bedenken eingetauscht.

Eine weitere methodische Herausforderung ergibt sich dadurch, *dass möglicherweise das Vorliegen des Indikators einen Sachverhalt anzeigen, sein Fehlen jedoch noch nicht das Gegenteil bedeuten muss*. Angenommen, die Teilnahme an einer Demonstration gegen Fremdenfeindlichkeit soll als Indikator für „nicht fremdenfeindliche Einstellungen“ gelten, dann zeigt die Nicht-Teilnahme keinesfalls eine „fremdenfeindliche Einstellung“ an, da es viele Gründe dafür geben kann, dass jemand sich nicht an dieser Veranstaltung beteiligt hat. Der Sachverhalt „fremdenfeindliche Einstellung“ ist ein typisches Beispiel dafür, dass ein Indikator allein oft nicht ausreicht, um den Sachverhalt zu messen, sondern dass mehrere Fragen zu (handlungsrelevanten) Einstellungen und Verhalten in bestimmten

Situationen gestellt werden müssen (vgl. z.B. den Themenschwerpunkt „Einstellungen gegenüber ethnischen Gruppen“ in der allgemeinen Bevölkerungsumfrage ALLBUS (2013)). Diese können zu einem Index oder einer Skala zusammengefasst werden, die wiederum eigene Anforderungen mit sich bringt (Latcheva/Davidov und Blasius, Kapitel 55 und 83 in diesem Band).

Ein letztes Beispiel für methodische Herausforderungen bei der Indikatorenbildung besteht darin, die *Differenziertheit von Ausprägungen* (z.B. Antwortmöglichkeiten im Fragebogen, Ausprägungen von Kategorien in Codierbögen) so zu gestalten, dass sie zu den Korrespondenzregeln und dem Indikator passen. Bei metrischen Skalen ist das zum Teil vergleichsweise einfach, z.B. kann der Indikator „Anzahl der Fachsemester“ für den Sachverhalt „Studienerfahrung“ verwendet werden. Über die Korrespondenzregel „je mehr Fachsemester, desto mehr Studienerfahrung“ hinaus ist es nicht zwingend nötig, vorab eine Semesteranzahl festzulegen, ab der jemand als „studienerfahren“ gilt, sondern bei der Auswertung kann sich erweisen, ob z.B. die Anzahl der Fachsemester mit einem anderen forschungsrelevanten Merkmal, z.B. Zufriedenheit mit dem Studium, statistisch zusammenhängt. In anderen Fällen dagegen ist eine Zuordnung schon in der Operationalisierungsphase notwendig. Wenn in einer Befragung von Museumsbesuchern etwa interessiert, ob sie auch ansonsten hoch- bzw. populärkulturelle Veranstaltungen besuchen, ist es zentral, erfragte Veranstaltungen diesen Sparten der Hoch- und Populärkultur zuzuordnen (was nicht immer ganz einfach ist – der kulturelle Anspruch an Kinofilme ist z.B. höchst unterschiedlich) und zu entscheiden, ob z.B. ein Besuch in den letzten zwölf Monaten ausreicht oder eine bestimmte Anzahl erreicht sein muss, um jemanden als „Theaterbesucher“ oder „Kinogänger“ zu klassifizieren. Das Beispiel zeigt, dass bei der „Übersetzungsleistung“ von Sachverhalten in Indikatoren im Rahmen der Operationalisierung durchaus oft viele Details zu berücksichtigen sind.

Insgesamt handelt es sich bei der Bestimmung von Indikatoren als direkten Anzeigern zu erforschender Sachverhalte um ein Kernstück der Operationalisierung innerhalb der quantitativen Forschung, welches das theoretische Konzept und die Messung miteinander verknüpft. Dadurch treffen Herausforderungen konzeptioneller und verfahrenstechnischer Art zusammen, wie das Beispiel der Armutsmessung zeigt. Nichtsdestoweniger liefert die Indikatorenbildung wichtige Orientierungen für ein empirisches Vorgehen, das nicht zuletzt die Interpretation der empirischen Befunde im Lichte der theoretisch angeleiteten Forschungsfrage erleichtert.

## Literatur

- ALLBUS (2013): Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. URL: <http://www.gesis.org/allbus/allbus-home/>
- Burzan, Nicole (2007): Quantitative Forschung in der Sozialstrukturanalyse. Wiesbaden: VS Verlag
- Deutsche Gesellschaft für Soziologie (2013): Willkommensseite. URL: <http://www.sozioologie.de/de/sektionen/sektionen/soziale-indikatoren/portrait.html>
- GESIS (2013): Soziale Indikatoren. URL. <http://www.gesis.org/unser-angebot/daten-analysieren/soziale-indikatoren/>
- Hradil, Stefan (2010): Der deutsche Armutsdiskurs. In: APuZ 2010 (51-52): 3-8
- Kromrey, Helmut (2009): Empirische Sozialforschung. Stuttgart: Lucius & Lucius
- Sozialberichte NRW Online (2013): Einkommensarmut. URL: [http://www.mais.nrw.de/sozialberichte/sozialindikatoren\\_nrw/indikatoren/7\\_einkommensarmut/index.php](http://www.mais.nrw.de/sozialberichte/sozialindikatoren_nrw/indikatoren/7_einkommensarmut/index.php)
- Statistisches Bundesamt (2010): Statistik und Wissenschaft. Demographische Standards. Ausgabe 2010. Band 17. Wiesbaden: Statistisches Bundesamt

Johannes Huinink

---

## 82.1 Soziale Ungleichheit und Ungleichheitsmerkmale

Unter sozialer Ungleichheit verstehen wir strukturell verankerte Disparitäten in den Lebens- und Handlungsbedingungen von Individuen, „die ihnen in unterschiedlichem Ausmaß erlauben, in der Gesellschaft allgemein anerkannte Lebensziele zu verwirklichen.“ (Huinink/ Schröder 2008: 99). Man unterscheidet verschiedene *Dimensionen sozialer Ungleichheit*:

- ökonomische Dimensionen (Bildung, Einkommen, Vermögen);
- wohlfahrtsstaatliche Dimensionen (Indikatoren des Zugangs zu Leistungen des Wohlfahrtsstaates und zur Infrastruktur moderner Wohlstandsgesellschaften),
- soziale Dimensionen (Umfang des sozialen Kapitals, der Entscheidungs- bzw. Machtbefugnisse, das soziale Prestige von Individuen) und
- emanzipatorische Dimensionen (Grad individueller Handlungsautonomie und Partizipationschancen an gesellschaftlichen Willensbildungsprozessen).

Alle Dimensionen sozialer Ungleichheit drücken ein Mehr oder Weniger von etwas aus, das für die Verfolgung relevanter Lebensziele wertvoll ist (*Ungleichheitsmerkmale*). Sie werden daher qua definitionem mit rational-, intervall- oder ordinal skalierten Variablen gemessen. Somit unterscheiden sie sich von nominal skalierten Variablen, wie etwa das Geschlecht, die nur Heterogenität zwischen Individuen bezeichnen (Blasius/Baur, Kapitel 79 in diesem Band).

## 82.2 Beispiel: Bildung, Einkommen und Beruf

In der Ungleichheitsforschung spielen seit je her die Ungleichheitsmerkmale *Bildung*, *Einkommen* und *beruflicher Status* eine zentrale Rolle (Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band). Sie liegen auch intervallskalierten Indizes des *sozioökonomischen Status* von Individuen zugrunde, die ebenfalls als Maß für deren Position im Ungleichheitsgefüge einer Bevölkerung dienen (Burzan, Kapitel 81 in diesem Band). In der Forschung werden verschiedene Versionen eines Socioeconomic Index (SEI), einer „Socioeconomic Status Scale“ (SES), vorgeschlagen (Duncan 1961).

Für den internationalen Vergleich ist der „International Socio-Economic Index of Occupational Status“ (ISEI) von Ganzeboom et al. (1992) entwickelt worden. Die ISEI-Skala ordnet beruflichen Tätigkeiten von Individuen einen mehr oder weniger hohen sozio-ökonomischen Statuswert zu. Die Skalenwerte für die beruflichen Tätigkeiten werden so bestimmt, dass die partielle Korrelation zwischen dem Ausbildungsniveau einer Person und dessen Einkommen bei Kontrolle des sozio-ökonomischen Status als Drittvariable minimal ist. Die inhaltliche Idee dahinter ist, dass ein Individuum die mit der Ausbildung erworbene Qualifikation nur über eine bestimmte berufliche Tätigkeit in Einkommen ummünzen kann, die berufliche Tätigkeit also den Zusammenhang zwischen Ausbildung und Einkommen vermittelt (Duncan 1961).

In Deutschland findet auch die 1971 erstmals erhobene „Stellung im Beruf“ als Maß für den beruflichen Status Anwendung (Hoffmeyer-Zlotnik/Geis 2003; Hoffmeyer-Zlotnik/Warner, Kapitel 54 in diesem Band). Erwähnt seien auch nominale Klassenschemata, wie sie von Wright (1985) und Erikson et al. (1979) vorgeschlagen worden sind. Sie bilden, wie die berufliche Stellung, keine strikt *vertikale Ungleichheitsordnung* ab. In der empirischen Praxis wurden jedoch auch daraus hierarchische Statusskalen abgeleitet.

Indizes zur Messung des *Berufsprestige*, die häufig in der Sozialstrukturforschung verwendet werden, sind die „Standard International Occupational Prestige Scale“ (SIOPS) von Treiman (1977) oder die Magnitude-Prestigeskala von Wegener (1985). Die Grundlage für diese Maße bilden Befragungen, in denen die Befragten für verschiedene Berufe angeben, wie angesehen oder geachtet sie ihrer Ansicht nach in der Gesellschaft sind. Die Prestigeskalen beziehen sich in der Regel auf die „International Standard Classification of Occupations“ (ISCO) (Hoffmeyer-Zlotnik/Geis 2003).

---

## 82.3 Ein- und mehrdimensionale Ansätze sozialer Ungleichheit

In der ökonomischen und sozialwissenschaftlichen Ungleichheits- und Armutsforschung gewinnen mehrdimensionale Messkonzepte sozialer Ungleichheit an Bedeutung. Während in einem *eindimensionalen Ansatz* lediglich eine Dimension sozialer Ungleichheit, wie das Einkommen oder der Berufsstatus, betrachtet wird, werden in *mehrdimensionalen Ansätzen* verschiedene Ungleichheitsdimensionen gleichzeitig berücksichtigt. Das Ausmaß sozialer Ungleichheit wird etwa durch Kombinationen des Bildungsstatus, des

Einkommensniveaus, der Güte der Wohnbedingungen und des Gesundheitszustandes gemessen, ohne dass dabei (immer) eine Indexbildung vorgenommen wird. Man geht davon aus, dass die Lebenslage von Individuen nur in dieser differenzierten Weise adäquat charakterisiert werden kann (Huinkink/Schröder 2008). Eine normative Theorie, die zu der stärkeren Beachtung einer mehrdimensionalen Analyse sozialer Ungleichheit beigetragen hat, wurde von Sen (1985) mit dem Capability Approach entwickelt.

---

## 82.4 Strukturmaße sozialer Ungleichheit

Man kann strukturelle und dynamische Maße sozialer Ungleichheit unterscheiden. Erstere lassen sich in zwei Typen aufgliedern:

1. *Maße sozialer Ungleichheit innerhalb einer Bevölkerung* messen, wie ungleich die Status der Mitglieder einer Bevölkerung im Hinblick auf ein Ungleichheitsmerkmal verteilt sind. Ein Beispiel ist die Einkommensungleichheit unter den Mitgliedern einer Bevölkerung, gemessen etwa mit dem *Gini-Koeffizienten* (siehe Abschnitt 6).
2. *Maße sozialer Ungleichheit zwischen Bevölkerungen oder Bevölkerungsgruppen* messen, wie verschieden die Verteilung eines Ungleichheitsmerkmals zwischen verschiedenen Bevölkerungen oder Teilgruppen einer Bevölkerung ist. Ein Beispiel sind die Unterschiede in der Verteilung des Schulabschlusses bei Personen mit und ohne Migrationshintergrund, etwa berechnet als Differenz des Anteils der Hauptschulabsolventen.

---

## 82.5 Dynamische Maße sozialer Ungleichheit

Dynamische Maße sozialer Ungleichheit messen, wie verschieden die (relative) Chance einzelner Mitglieder in einer Bevölkerung ist, einen bestimmten Status in einer Dimension sozialer Ungleichheit zu erreichen oder von einem Status zum anderen zu wechseln (Aufstiege, Abstiege). Sie erfassen also das unterschiedliche Ausmaß von Statusmobilität und -stabilität in einer Bevölkerung. Dazu gehören Übergangswahrscheinlichkeiten oder -raten (Pötter/Prein diesem Band) zwischen Statuspositionen, die zwischen verschiedenen Teilgruppen der Bevölkerung verglichen werden. Zum Beispiel ist die Wahrscheinlichkeit, nach der Grundschule auf das Gymnasium zu wechseln, unter Kindern unterschiedlicher sozialer Herkunft verschieden hoch. Ein anderes Beispiel sind Aufstiegs- und Abstiegsraten von Individuen in unterschiedlichen beruflichen Statuspositionen.

Strukturmaße und dynamische Maße der Ungleichheit hängen miteinander zusammen. Dynamische Maße bilden ab, wie sich das strukturelle Ausmaß sozialer Ungleichheit in der Zeit verändert. Dennoch können einander entsprechende strukturelle und dynamische Maße sozialer Ungleichheit unterschiedliche Einschätzungen des Wandels von sozialer Ungleichheit begründen. So sind die Übergangsichten von Arbeiterkindern von der Grundschule zum Gymnasium im Zuge der letzten Jahrzehnte deutlich stärker gestiegen

als die entsprechenden Raten von Kindern höherer Statusgruppen – aber von einem sehr viel niedrigeren Niveau des absoluten Anteils der Kinder ausgehend, die das Gymnasium besucht haben. Der Anteil der Arbeiterkinder, die zum Gymnasium gehen, ist gleichzeitig weniger gewachsen als der entsprechende Anteil der Kinder von leitenden Angestellten oder Beamten (Schimpl-Neymanns 2000).

## 82.6 Maße sozialer Ungleichheit innerhalb einer Bevölkerung

Im diesem Abschnitt werden Maße sozialer Ungleichheit vorgestellt. Sie verlangen Intervallskalenniveau der entsprechenden Ungleichheitsmerkmale. Zur Illustration dafür wird das Einkommen verwendet. Anschließend werden Verfahren angesprochen, mit denen ein Vergleich der Verteilung von Ungleichheitsmerkmalen zwischen verschiedenen Teilbevölkerungen vorgenommen werden kann. Ein schon erwähntes Beispiel ist der Vergleich von sozialstrukturellen Gruppen nach dem Schulbildungs- oder Ausbildungsniveau.

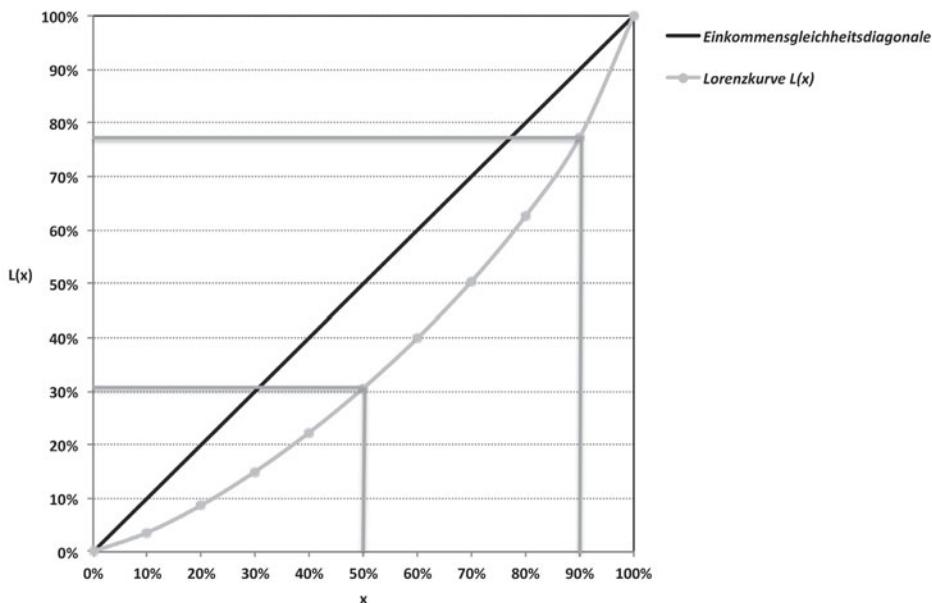
### 82.6.1 Ein einfaches Beispiel: Die Lorenzkurve einer Einkommensverteilung

Eine prominente graphische Darstellung für das Ausmaß der Ungleichheit des Einkommens unter den Mitgliedern einer Bevölkerung ist die Lorenzkurve. Daraus lassen sich weitere Ungleichheitsmaße ableiten, die noch vorgestellt werden (Hartmann 1985, Cowell 2011).

Die Lorenzkurve ist der Graph der Lorenzfunktion  $L(x)$ , deren Werte wie folgt bestimmt werden:  $x$  bezeichnet einen Prozentwert zwischen 0 und 100.  $L(x)$  ist dann gleich dem Anteil des kumulierten Einkommens der einkommensschwächsten  $x$  Prozent der Bevölkerung  $P$  am kumulierten Gesamteinkommen von  $P$ .

In Abb. 82.1 sind in Zehnerschritten von  $x$  die Werte der Lorenzfunktion der Verteilung des Einkommens in Deutschland abgetragen (hellgraue Kurve). Die Daten stammen aus der Einkommens- und Verbrauchsstichprobe 2008 des Statistischen Bundesamts (Hartmann/Lengerer, Kapitel 71 in diesem Band). Als Einkommensgröße wird das monatliche Nettoäquivalenzeinkommen der Haushalte verwendet, in dem die Mitglieder der Bevölkerung leben (Statistisches Bundesamt 2012: 59).

In Abb. 82.1 wurde auch die Lorenzkurve eingezeichnet, die den Fall repräsentiert, in dem alle Mitglieder von  $P$  dasselbe Einkommen haben (Einkommensgleichheitsdiagonale; schwarze Gerade). Man erkennt, dass die Lorenzkurve des Nettoäquivalenzeinkommens für Deutschland deutlich von der Einkommensgleichheitsdiagonalen abweicht. Das ist ein Hinweise darauf, dass das Einkommen im Jahr 2008 unter den Einwohnern Deutschland ungleich verteilt war. An der Lorenzkurve lässt sich zum Beispiel auch ablesen, dass die einkommensschwächeren 50% der Bevölkerung nur etwas mehr als 30% des gesamten Nettoäquivalenzeinkommens verdienen. Die einkommensstärksten 10% der Bevölkerung erwirtschaften dagegen ungefähr 23%.



**Abb. 82.1** Lorenzkurve des Nettoäquivalenzeinkommens für Deutschland 2008 (Datenquelle: Statistisches Bundesamt 2012: 59; eigene grafische Darstellung.)

### 82.6.2 Allgemeine Charakterisierung von Maßen sozialer Ungleichheit

Das Ausmaß der Ungleichheit in den Ausprägungen  $u_1, \dots, u_n$  eines Ungleichheitsmerkmals  $U$ , z.B. dem Einkommen, unter den Mitgliedern  $i$  (mit  $i=1, 2, \dots, n$ ) einer Bevölkerung  $P$  kann intuitiv als Funktion des „Abstandes“ von dem Zustand verstanden werden, in dem  $U$  für alle Mitglieder von  $P$  denselben Wert  $u_1 = u_2 = u_3 = \dots = u_n = \mu_U$  aufweist, wobei  $\mu_U$  der Erwartungswert der Häufigkeitsverteilung von  $U$  ist. Ungleichheit in  $P$  äußert sich in einem mehr oder weniger großen Unterschied der  $u_i$ , welche die Mitglieder  $i$  in  $P$  haben. Zusätzlich ist von Bedeutung, wie stark unterschiedlich große Werte von  $U$  auf Teile der Bevölkerung konzentriert sind. So können beispielsweise sehr hohe Einkommen einer sehr kleinen Teilgruppe in  $P$  vorbehalten sein, während der überwiegende Teil der Bevölkerung über ein vergleichsweise geringes Einkommen verfügt.

Ein Ungleichheitsmaß  $I(V(U))$  bezogen auf  $U$  bestimmt eine Ordnung auf der Menge möglicher „Verteilungen“  $V(U) = (u_1, \dots, u_n)$  der Gesamtmenge von  $U$  unter den Mitgliedern von  $P$ . Ein zentrales explizites Prinzip, gemäß dessen vergleichend beurteilt werden kann, ob eine Verteilung  $V(U)$  ungleicher ist als eine andere mögliche Verteilung  $V'(U) = (u'_1, \dots, u'_n)$ , ist das Transferprinzip nach Pigou und Dalton (Hartmann 1985). Ihm liegt die plausible Annahme zugrunde, dass der Transfer eines Betrages  $\tau$  von  $U$  (z. B. ein Einkommensbetrag) von einem Mitglied  $i$  mit einem großen Wert  $u_i$  zu einem Mitglied  $i'$

mit einem kleineren Wert  $u_i$ , die Ungleichheit von  $V(U)$  verringert, solange  $u_i$  nach dem Transfer nicht größer als  $u_i$  geworden ist. Geht  $V'(U)$  durch eine solche Transformation aus  $V(U)$  hervor, dann soll  $I(V(U))$  größer sein als  $I(V'(U))$ .

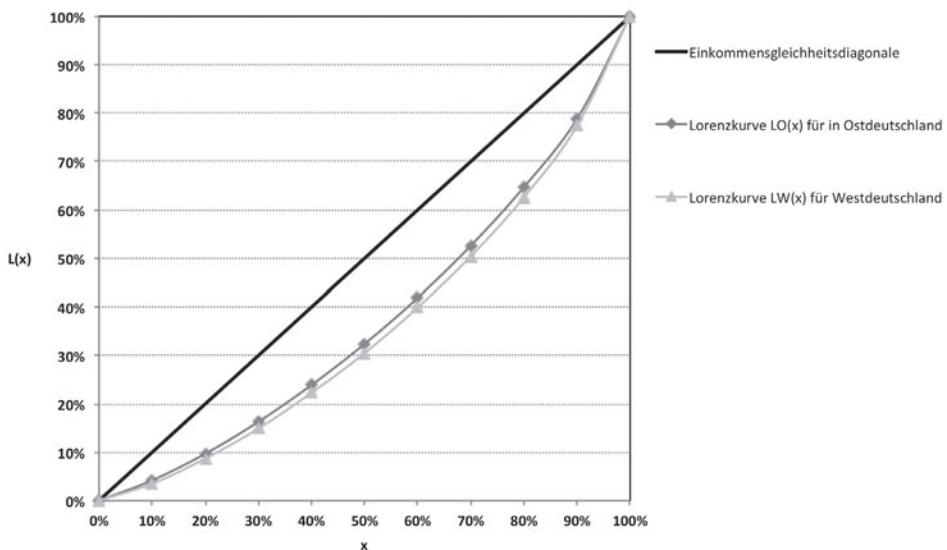
Vielfach werden drei weitere Prinzipien berücksichtigt. Die Skaleninvarianz von  $I$  impliziert, dass der Wert von  $I$  für eine Verteilung von  $U$  genauso groß ist wie von einem Vielfachen von  $U$ . Sie bedeutet beispielsweise, dass die Veränderung der absoluten Differenz zwischen zwei Einkommensgrößen, wenn diese mit einem Faktor multipliziert werden, für das Ungleichheitsmaß bedeutungslos bleibt (relative Einkommensungleichheit). Gilt das Prinzip der Populationsinvarianz, ist  $I$  unabhängig von der Größe der Bevölkerung.  $I$  ändert sich also zum Beispiel nicht, wenn man die Einkommensverteilung (und die Zahl der Einkommensbezieher) verdoppeln würde. Das Prinzip der Anonymität oder Symmetrie besagt, dass  $I$  unabhängig davon ist, welche Mitglieder von  $P$  welches Einkommen erhalten. Eine veränderte Zuordnung der Einkommen zu den Einkommensbeziehern ändert also an dem Wert von  $I$  nichts. Diese vier Prinzipien erlauben keine vollständige Ordnung auf der Menge der Verteilungen von  $U$  bzgl. des Ausmaßes ihrer Ungleichheit. Es müssen daher weitere Annahmen bei der Definition von  $I$  gemacht werden, die aber oft nur implizit bleiben (Hartmann 1985).

Ausgehend von der Lorenzfunktion kann zum Beispiel immer nur dann eindeutig ein Vergleich der Ungleichheit von zwei Verteilungen  $V(U)$  und  $V'(U)$  für den Fall vorgenommen werden (Lorenz-Dominanz), dass sich die Lorenzkurven zu  $V(U)$  und  $V'(U)$  nicht schneiden.  $V(U)$  ist dann ungleicher als  $V'(U)$ , wenn alle Werte der Lorenzfunktion zu  $V(U)$  kleiner als die entsprechenden Werte der Lorenzfunktion zu  $V'(U)$  sind. Verteilungen von  $U$  sind immer genau dann nach dem Lorenzkriterium vergleichbar, wenn sie gemäß den oben genannten vier Prinzipien vergleichbar sind. Wie sie definiert das Lorenzkriterium also keine vollständige Ordnung auf der Menge der Verteilungen.

In Abb. 82.2 werden die Lorenzkurven der Verteilung der Nettoäquivalenzeinkommen in der Bevölkerung Westdeutschlands und Ostdeutschlands,  $L_W(x)$  (hellgraue Kurve) und  $L_O(x)$  (dunkelgraue Kurve), für das Jahr 2008 miteinander verglichen. Die Datenbasis ist wieder die EVS des Jahres 2008.  $L_W(x)$  und  $L_O(x)$  liegen zwar nahe beieinander,  $L_O(x)$  liegt aber für jedes  $x$  näher an der Einkommensungleichheitsdiagonalen als  $L_W(x)$ . Die Einkommensverteilung ist im Westen also ungleicher als in Ostdeutschland.

### 82.6.3 Maße der Einkommensungleichheit

Im Folgenden werden einige bekannte Maße sozialer Ungleichheit am Beispiel der Verteilung des Einkommens unter den Mitgliedern einer Bevölkerung  $P$  vorgestellt (Cowell 2011). Die einfachste Art und Weise, sich einen Eindruck über Einkommensungleichheit zu verschaffen, ist die *Häufigkeitsverteilung der Einkommenshöhen in P* zu betrachten. Sie zeigt, welcher Anteil der Bevölkerung welche Einkommenshöhe realisiert. Die Differenz der  $x\%$  höchsten und der  $x\%$  niedrigsten gemessenen Einkommen gibt die *Spannweite* der Verteilung an. So genannte *Perzentilabstände* und -*relationen* messen ebenfalls den



**Abb. 82.2** Lorenzkurven des Nettoäquivalenzeinkommens für Westdeutschland (einschließlich Berlin-West) und Ostdeutschland (einschließlich Berlin-Ost) im Jahr 2008 (Datenquelle: Statistisches Bundesamt 2012: 60/61; eigene grafische Darstellung.)

Abstand oder das Verhältnis der Einkommensgrößen. Ein Perzentil  $P_x$  ist gleich dem Einkommensbetrag, den x Prozent der Mitglieder von P unterschreiten oder gerade erreichen. Das Perzentil  $P_{50}$  ist der Median der Verteilung der Einkommensgrößen. Bei der Berechnung der relativen Einkommensarmut (Burzan, Kapitel 81 in diesem Band) wird dieser Wert anstelle des arithmetischen Mittels als mittleres Einkommensmaß verwendet, weil er gegenüber Ausreißern nach oben, also extrem hohen Einkommen, robust ist.

Ist x ein Vielfaches von 10, sprechen wir von Dezilen. Als Ungleichheitsmaße werden dann *Dezil-Verhältnisse* bestimmt:  $P_{90}/P_{10}$  ist das Verhältnis des neunten Dezils zum ersten Dezil. Je größer dieses Verhältnis ist, desto größer ist die Spreizung der Einkommen in einer Bevölkerung. Die Dezil-Verhältnisse  $P_{50}/P_{10}$  und  $P_{90}/P_{50}$  zeigen Einkommensungleichheit spezifisch im unteren bzw. oberen Teil der Verteilung an. Sie geben Hinweise darauf, aus welchem Einkommensbereich sich die Gesamtungleichheit speist. Will man nicht relative, sondern absolute Einkommensungleichheit messen, muss man die Differenz der Perzentile statt deren Verhältnis betrachten.

Ausgehend von der *Lorenzfunktion* können kumulierte Einkommensanteile von Teilstichproben in der Bevölkerung P am Gesamteinkommen in P als skaleninvariante Ungleichheitsmaße bestimmt werden. Beispielsweise wird der Anteil des kumulierten Einkommens der einkommensstärksten 20% am Gesamteinkommen der Bevölkerung zu dem entsprechenden Anteil der einkommensschwächsten 20% in Beziehung gesetzt. Dieser Quotient

wird als Quintilsverhältnis oder S80/S20-Rate bezeichnet (Statistisches Bundesamt 2012: 68). Man bezeichnet diese Bevölkerungsteile als das oberste (fünfte) oder unterste (erste) Einkommensquintil (Fünftel) der Bevölkerung. Das Quintilsverhältnis betrug 2008 in Westdeutschland 4,3 und in Ostdeutschland 3,7.

Ein in der Literatur und öffentlichen Debatte um Einkommensungleichheit oft benutztes Maß, das den Abstand einer empirischen Verteilung von Einkommensgrößen zum Fall vollkommener Einkommensgleichheit in einer Bevölkerung misst, ist der *Gini-Koeffizient* G. Er ist gleich dem doppelten Betrag der Fläche zwischen der Lorenzkurve L(x) und der Einkommensgleichheitsgeraden. Er beträgt 0 bei vollständiger Einkommensgleichheit und wird gleich  $(n-1)/n$  für den Extremfall, dass nur ein Mitglied der Bevölkerung das gesamte Einkommen besitzt. Der Gini-Koeffizient wird berechnet gemäß:

$$G = \frac{1}{2} \times \frac{1}{n^2 \times \mu_U} \sum_{i=1}^n \sum_{j=1}^n |u_i - u_j|.$$

Er ist also gleich dem Verhältnis des Mittelwerts aller möglichen, absoluten Einkommensdifferenzen zwischen zwei Mitgliedern i und j in der Bevölkerung P zum Erwartungswert  $\mu_U$  der Häufigkeitsverteilung des Einkommens in P, dividiert durch 2. Der Gini-Koeffizient der Einkommensverteilung in Ostdeutschland lag im Jahr 2008 bei 0,26 und in Westdeutschland bei 0,29. Auch anhand dieses Koeffizienten zeigt sich, dass die Einkommensungleichheit in Ostdeutschland geringer als in Westdeutschland war.

Ein Ungleichheitsmaß, das mit dem Gini-Koeffizient eng zusammenhängt, ist der *Variationskoeffizient* des Einkommens in P, der in der Statistik als relatives Streuungsmaß bekannt ist. Er berechnet sich als das Verhältnis der Standardabweichung  $\sigma_U$  zum Erwartungswert  $\mu_U$ . Aktuelle Zahlen dazu sind für Deutschland leider nicht veröffentlicht.

Eine Gruppe von ähnlich interpretierbaren Maßen bilden so genannte *Entropie-basierte Maße* der Einkommensungleichheit. Die Entropie der Wahrscheinlichkeitsverteilung über eine Menge von Ereignissen ist umso größer, je ähnlicher diese Verteilung einer Gleichverteilung ist, gemäß der alle Ereignisse mit derselben Wahrscheinlichkeit auftreten können. Die Entropie ist gleich Null, wenn ein Ereignis mit der Wahrscheinlichkeit 1 auftritt – und damit die Ordnung maximal ist. Die Ereignisse werden in den Entropie-basierten Maßen sozialer Ungleichheit durch die Mitglieder der Bevölkerung P und die Wahrscheinlichkeiten durch den Anteil ihres Einkommens am Gesamteinkommen ersetzt (Cowell 2011: 53). Der Gleichverteilung oben entspricht die Einkommensgleichheit. Das bekannteste Entropie-basierte Maß stammt von dem Ökonomen Theil (1965). Der Theil-Index T ist gleich der Differenz zwischen der Entropie der so umdefinierten Gleichverteilung und der Entropie der Einkommensverteilung auf den Mitgliedern der Bevölkerung, die sich aus deren Einkommensanteilen ergibt. Es wird berechnet als:

$$T = \frac{1}{n} \sum_{i=1}^n \frac{u_i}{\mu_U} \times \ln \left( \frac{u_i}{\mu_U} \right).$$

Bei Einkommensgleichheit ist der Theil-Index Null. Er wird mit steigender Einkommensungleichheit größer.

In der Ungleichheitsliteratur wird eine weitere Klasse von Maßen sozialer Ungleichheit vorgeschlagen, die wohlfahrtstheoretisch fundiert sind. Ausgangspunkt ist ein für die Bevölkerung P bestimmtes, *soziales Wohlfahrtsmaß*  $W$ , das als Funktion der individuellen Wohlfahrt  $B_i$  der Mitglieder i der Bevölkerung von P bestimmt wird. Die  $B_i$  sind ihrerseits eine Funktion des Einkommens der Mitglieder  $i=1, \dots, n$ . Dann wird angenommen, dass die Wohlfahrtfunktion das Pareto-Prinzip erfüllt, also ihr Wert zunimmt oder mindestens gleich bleibt, wenn das Einkommen eines Mitglieds i ansteigt, ohne dass das Einkommen eines anderen Mitglieds i' sinkt. Des Weiteren soll die Wohlfahrtfunktion symmetrisch, additiv bzgl. der individuellen Nutzen  $B_i$  und konkav sein, d.h. steigende Einkommen sollen mit einem abnehmenden Grenznutzen einhergehen. Diese Annahmen begründen eine Präferenz für Einkommensgleichheit in P, da das Transferkriterium von Pigou und Dalton (Hartmann 1985) dann als Kriterium zur Steigung der sozialen Wohlfahrt W betrachtet werden kann. W ist also maximal, wenn alle Mitglieder der Bevölkerung P dasselbe Einkommen haben, das dann gleich dem Mittelwert  $\mu_U$  ist. In Abhängigkeit davon, wie stark Einkommensungleichheit in einer Bevölkerung als wohlfahrtsmindernd angesehen wird (*Ungleichheitsaversion*), drückt Abweichen von dieser Konstellation wieder eine mehr oder weniger große soziale Ungleichheit aus.

Das bekannteste Beispiel für diese Klasse von Maßen sozialer Ungleichheit ist das *Atkinson-Maß* (Cowell 2011: 51). Zu dessen Berechnung wird zunächst in Abhängigkeit von der (angenommenen) Stärke der Ungleichheitsaversion der Einkommensbetrag  $u_e$  bestimmt, der, würden ihn alle Mitglieder von P als Einkommen erhalten, dem Wert der sozialen Wohlfahrt W entsprechen würde, wie er für die faktische(ungleiche) Einkommensverteilung berechnet wird. Da die soziale Wohlfahrt W bei Vorherrschen von Einkommensungleichheit nicht maximal ist, ist dieses hypothetische Einkommen  $u_e$  kleiner  $\mu_U$ , was jedes Mitglied in P ja bei faktischer Einkommensgleichheit verdiente. Das Atkinson-Maß als Maß sozialer Ungleichheit ist dann gleich  $1 - (u_e/\mu_U)$  und ist damit umso größer, je kleiner  $u_e$  ist.

## 82.6.4 Dekomposition von Maßen der Einkommensungleichheit

Mit Verfahren der Dekomposition von Maßen der Einkommensungleichheit kann man untersuchen, wie stark das Ausmaß sozialer Ungleichheit in einer Bevölkerung durch die Zusammensetzung der Bevölkerung nach sozialstrukturellen Merkmalen bedingt ist (Jenkins/van Kerm 2009). Für Teilgruppen der Bevölkerung P, deren Mitglieder bestimmte Merkmale (z.B. den Beruf oder die Region des Wohnorts) gemeinsam haben, lässt sich feststellen, ob die Einkommensungleichheit in P eher auf der Einkommensungleichheit innerhalb dieser Teilgruppen beruht oder ob sie eher durch die Ungleichheit zwischen diesen Teilgruppen (Abschnitt 4) begründet ist. Analog kann die Frage beantwortet werden, ob die Veränderung der Ungleichheit der Verteilung von U in P über die Zeit eher das

Ergebnis einer veränderten Komposition von P nach ihren sozialstrukturellen Teilgruppen ist, in denen sich selbst die Ungleichheit bezogen auf U kaum verändert hat, oder ob sie eher durch die Veränderung der Ungleichheit innerhalb der Teilgruppen bewirkt wurde.

---

## 82.7 Maße sozialer Ungleichheit zwischen (Teil-) Bevölkerungen

Ein Vergleich der (Häufigkeits-)Verteilung eines Ungleichheitsmerkmals U zur Analyse der sozialen Ungleichheit nicht innerhalb, sondern zwischen (Teil-)Bevölkerungen, ist als zweiter Typ der strukturbezogenen Messung sozialer Ungleichheit angeführt worden. Zum Beispiel wird untersucht, wie sich Teilgruppen der Bevölkerung eines Landes oder auch Bevölkerungen unterschiedlicher Länder im Hinblick auf das Bildungsniveau (ordinalskaliertes Merkmal), die Einkommenshöhe (intervallskaliertes Merkmal) oder andere Ungleichheitsmerkmale unterscheiden.

Dazu greift man bei intervallskalierten Merkmalen auf varianzanalytische Verfahren zurück, die Mittelwertvergleiche vornehmen und testen. Gruppenspezifische Unterschiede ordinalskalierter Ungleichheitsmerkmale lassen sich z.B. mit dem Mann-Withney-U-Test oder Kruskal-Wallis-Test testen. Verfahren der Kontingenztafelanalyse, angefangen mit der Berechnung von Anteilsdifferenzen bezogen auf bestimmte Kategorien (z.B. Anteil der Hauptschüler in verschiedenen Teilbevölkerungen) können Anwendung finden, soweit sie für ordinalskalierte Merkmale sinnvoll sind. Maße oder Tests der reinen Verschiedenheit von Verteilungen zwischen Teilpopulationen, wie der Chi<sup>2</sup>-Test oder der Dissimilaritäts- bzw. Segregationsindex, gehören im engeren Sinne nicht dazu, weil sie auch bei nominalskalierten Merkmalen einsetzbar sind und keinen direkten Hinweis auf soziale Ungleichheit geben.

Mit regressionsanalytischen Verfahren (Mehrebenenmodelle) werden Niveauunterschiede von Ungleichheitsmerkmälern zwischen unterschiedlichen Bevölkerungsgruppen unter Einbezug individueller oder kontextbezogener Merkmale (Prädiktoren) geschätzt (Pötschke, Kapitel 87 in diesem Band). Mit solchen Modellen kann auch eine Dekomposition der Niveauunterschiede in U (Mittelwertdifferenzen) zwischen Teilgruppen in einen Anteil, der auf die Zusammensetzung der Teilgruppen nach einzelnen Merkmalen, und in einen Anteil, der auf Teilgruppen spezifische Effekte der Prädiktoren auf U zurückgeht, vorgenommen werden (Oaxaca/Ramson 1994).

---

## 82.8 Weiterführende Literatur

In diesem Beitrag wurden nur wenige Einblicke in die Messung sozialer Ungleichheit gegeben. Ausführliche Einführungen, die in die technischen Details der hier angeführten Maße und in zahlreiche weitere Ansätze einführen, sind in schon zitierten Arbeiten von Hartmann (1985), Jenkins/van Kerm (2009) und vor allem Cowell (2011) zu finden. Cowell gibt den umfangreichsten Überblick und ergänzt diesen mit Übungsaufgaben und zahlreichen

praktischen Anleitungen. Jenkins/van Kerm (2009) und Cowell (2011) führen auch in Dekompositions- und Regressionsmethoden ein und behandeln Einkommensmobilität.

In diesem Beitrag wurde die mehrdimensionale Messung sozialer Ungleichheit nicht behandelt, die als immer wichtiger angesehen wird. Cowell (2000) gibt einen guten Überblick über den bisherigen Stand zu Messverfahren der Ungleichheit mehrdimensionaler Verteilungen. Konzepte der Messung von sozialer Ungleichheit und Armut bei mehrdimensionalen Verteilungen stellen auch Atkinson (2003) und Alkire/Foster (2011) vor.

Zu dynamischen Maßen sozialer Ungleichheit sei auf den Aufsatz von Mare (1981) verwiesen, auf den eine umfangreiche Forschungstradition aufbaut, sowie auf den Überblick von Jenkins/van Kerm (2009).

## Literatur

- Alkire, Sabina/Foster, Foster (2011): Counting and Multidimensional Poverty Measurement. In: Journal of Public Economics 95: 476-487
- Atkinson, Anthony B./Bourguignon, Francois (2000) (Hg.): Handbook of Income Distribution, North Holland, Amsterdam. 87-166
- Atkinson, Anthony B. (2003): Multidimensional deprivation: Contrasting social welfare and counting approaches. In: Journal of Economic Inequality. 1: 51-65
- Cowell, Frank A. (2000): Measurement of inequality. In: Atkinson, Anthony B./ Bourguignon, Francois (Hg.): 87-166
- Cowell, Frank A. (2011): Measuring inequality. New York: Oxford University Press
- Duncan, Otis D. (1961): A Socioeconomic Index for all Occupations. In: Reiss Jr./Albert B. (Hg.): 109-138
- Reiss Jr./Albert B. (1961) (Hg.), Occupations and Social Status Glencoe: The Free Press
- Erikson, Robert/Goldthorpe, John H./Portocarero, Lucienne (1979): Intergenerational Class Mobility in Three Western Societies: England, France and Sweden. In: British Journal of Sociology 30: 415-441
- Fahrmeir, Ludwig/Künstler, Rita/Pigeot, Iris/Tutz, Gerhard (2009): Statistik. 7. Auflage, Berlin Heidelberg: Springer
- Ganzeboom, Harry B.G./De Graaf, Paul M./Treiman, Donald J. (1992): A Standard International Socio-Economic Index of Occupational Status. In: Social Science Research 21: 1-56
- Hartmann, Peter H. (1985): Die Messung sozialer Ungleichheit. Pfaffenweiler : Centaurus-Verlagsgesellschaft
- Hoffmeyer-Zlotnik, Jürgen H.P./Geis, Alfons (2003): Berufsklassifikation und Messung des beruflichen Status/Prestige. In: ZUMA-Nachrichten 52: 125-138
- Huinink, Johannes/Schröder, Torsten (2008): Sozialstruktur Deutschlands. Konstanz: Universitätsverlag Konstanz
- Jenkins, Stephen P./Kerm, Philippe van (2009): The Measurement of Economic Inequality. In: Salverda, Weimer/Nolan, Brian/Smeeding/Timothy M. (Hg.): 40-67
- Mare, Robert D. (1981): Change and Stability in Educational Stratification. In: American Sociological Review 46: 72-87
- Oaxaca, R. L./Ransom, M. R. (1994). On Discrimination and the Decomposition of Wage Differentials. In: Journal of Econometrics 61: 5-21
- Salverda, Weimer/Nolan, Brian/Smeeding/Timothy M. (2009) (Hg.): The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford
- Schimpel-Neymanns, Bernhard (2000): Soziale Herkunft und Bildungsbeteiligung. Empirische Analysen zu herkunftspezifischen Bildungsungleichheiten zwischen 1950 und 1989. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 52: 636-669
- Sen, Amartya (1985): Commodities and Capabilities. Amsterdam: North-Holland

- Theil, Henri (1965): The Information Approach to Demand Analysis. In: *Econometrica* 33: 67-87
- Treiman, Donald J. (1977): Occupational Prestige in Comparative Perspective. New York: Academic Press
- Wegener, Bernd (1985): Gibt es Sozialprestige? In: *Zeitschrift für Soziologie* 14: 209-235
- Wright E. Olin (1985): Classes. London: Verso

Jörg Blasius

---

## 83.1 Was ist Skalierung?

Unter Skalierung versteht man, wie ein oder mehrere Beobachtungsmerkmal(e) einem oder mehreren Skalenwert(en) zugeschrieben wird (werden). Dabei sind die Beobachtungsmerkmale *manifest*, sie werden direkt erhoben, z.B. indem sie in einer Umfrage abgefragt werden. Dies können z.B. Merkmale des christlichen Glaubens sein, wie die Fragen „Glauben Sie an Gott?“, „Glauben Sie an die Hölle?“ und „Glauben Sie an den Teufel?“. Mit Hilfe eines *Skalierungsverfahrens* werden diese Merkmale für jeden Befragten zu einem individuellen Skalenwert zusammengefasst, sie werden *skaliert*. Diese resultierende *Skala* ist eine *latente Variable*, sie wird auch als *Dimension* oder als *Faktor* bezeichnet; in dem genannten Beispiel kann sie als „christliche Religiosität“ interpretiert werden. Die resultierende Variable ist metrisch skaliert (Blasius/Baur, Kapitel 79 in diesem Band) und meistens standardisiert, üblicherweise mit dem Mittelwert „Null“ und der Standardabweichung „Eins“. Die Interpretation dieser Skala wäre dann: „je positiver der Wert ist, desto religiöser ist die entsprechende Person“.

Wichtig ist hierbei, dass alle manifesten Merkmale die gleiche Dimension beschreiben, z.B. dass mit allen berücksichtigten Fragen die „Religiosität“ gemessen wird. Dies setzt sowohl eine *Messtheorie* voraus (Latcheva/Davidov, Kapitel 55 in diesem Band), als auch ein sehr sorgfältiges Formulieren der Fragen (im Fall von Skalierung wird oft der Begriff *Items* verwendet) (Porst, Kapitel 50 in diesem Band) und der Antwortvorgaben (Franzen, Kapitel 51 in diesem Band). Die Beschränkung auf eine einzige Dimension ist dabei keinesfalls zwingend, es gibt eine Vielzahl von multivariaten Verfahren, mit denen mehrere Dimensionen unterschieden werden können (Blasius/Baur, Kapitel 79 in diesem Band); in dem gegebenen Beispiel könnte es eine Unterteilung nach „religiösem Wissen“ und „religiösen Praktiken“ geben, die dann mit entsprechenden Fragen beschrieben werden.

Ziel dieses Kapitels ist es, die grundlegenden Ideen der Skalierung anhand von eindimensionalen Skalen darzustellen (für mehrdimensionale Skalen, Blasius/Baur, Kapitel 79

in diesem Band). Ein ganz besonderes Augenmerk wird dabei auf die Struktur der zugrundeliegenden Daten gelegt (Coombs 1964), was heutzutage in den Sozialwissenschaften leider oft vernachlässigt wird und was zu fehlerhaften Interpretationen führt. So ist die Datenstruktur eine vollkommen andere, wenn die Antwortalternativen dichotom sind, z.B. in dem nur „ja“ und „nein“ zugelassen werden, oder wenn sie in geordneter kategorialer Form vorgegeben sind, z.B. auf einer Fünfer-Skala, die von „stimme voll und ganz zu“ bis „stimme überhaupt nicht zu“ verläuft.

---

### 83.2 Warum werden Fragen skaliert?

Als Pioniere der Skalierung werden meistens Pearson (1901) und Spearman (1904) genannt, die als erste Beziehungen zwischen Variablen formal beschrieben haben. Pearson entwickelte einen Korrelationskoeffizienten für metrische Daten, Spearman führte den nach ihm benannten Rangkorrelationskoeffizienten ein. Des Weiteren entwickelte er einen Algorithmus, um die Existenz eines allgemeinen Faktors in der Intelligenz-Forschung zu dokumentieren. Seitdem wurden diese Ideen, die auf eine (theoretisch) unbegrenzte Anzahl von Faktoren erweitert werden können, kontinuierlich weiterentwickelt. Die erste wissenschaftliche Anwendung geht vermutlich auf MacDonell (1902) zurück, der von 3.000 männlichen Gefängnisinsassen über 21 Jahren sieben physische Merkmale (u.a. Körpergröße, Länge des linken Mittelfingers, Größe des linken Fußes) zu einer Skala zusammenfasste.

In den Naturwissenschaften ist die Skalierung von Daten in der Regel sehr präzise, wobei oft ein direkt messbarer Indikator, also eine *manifeste Variable*, ausreicht, um eine Dimension, also eine *latente Variable*, zu beschreiben. Als Beispiele können die Messung von Temperatur (in Celsius), das Gewicht eines Gegenstandes (in Kilogramm) und Distanzen (z.B. zwischen Städten, in Kilometern) genannt werden. In diesen Fällen ist die Zuordnung von manifesten Messwerten zu Skalenwerten sehr einfach: Distanzen im Bereich von Zentimetern werden mit dem Lineal gemessen, die Temperatur mit dem Thermometer und das Gewicht mit Hilfe einer Waage. Aber auch für die Naturwissenschaften gilt: Je besser das Messinstrument geeicht ist, desto genauer ist die Messung. Das Blutdruckmessgerät, das für wenige Euro im Supermarkt erhältlich ist, liefert im Vergleich zu dem geeichten Blutdruckmessgerät, welches in jeder Arztpraxis vorhanden ist, allenfalls näherungsweise richtige Werte.

In den Sozial- und Wirtschaftswissenschaften sind die Messungen i.d.R. wesentlich ungenauer als in den Naturwissenschaften, und es ist meistens viel schwieriger, Messungen durchzuführen, ein Indikator (eine Frage) ist eigentlich nie ausreichend, um ein latentes Merkmal adäquat zu erfassen. Des Weiteren kann in vielen Fällen nicht davon ausgegangen werden, dass ein latentes Merkmal überall (in allen Ländern oder auch nur in allen Regionen eines Landes) mit den gleichen Items gemessen werden kann, was aber eigentlich die Voraussetzung für einen internationalen Vergleich ist (Braun, Kapitel 56 in diesem Band).

Die Schwierigkeit dieser einheitlichen Messung soll an einem kleinen Beispiel verdeutlicht werden, bei dem auf den ersten Blick ein recht ungewöhnlicher Index berechnet wird.

Für das internationale Finanzwesen ist u.a. die Kaufkraftparität ein sehr wichtiger Indikator zur Beschreibung des Wohlstandes in einem Land. Die Gültigkeit der entsprechenden Berechnungen (Krebs/Menold, Kapitel 30 in diesem Band) hängt von der Wahl eines geeigneten „Warenkorbs“ an Konsumgütern ab – und dieser ist in allen Ländern unterschiedlich zusammengestellt. Vollkommen ungeeignet ist eine in Deutschland zentrale Größe wie die Heizkosten, diese dürften in den meisten afrikanischen Ländern bei Null liegen, zumindest wenn die Klimaanlage unberücksichtigt bleibt (die sich zudem nur wenige Bewohner leisten können). Dieser Indikator ist schon innerhalb vieler Länder kaum zu verwenden: Während z.B. der Norden von Italien über bekannte Ski-Gebiete verfügt, sind auf Sizilien Heizungen in privaten Wohnungen selten.

Um dennoch einen validen internationalen Vergleich zu ermöglichen, führte die Zeitschrift „The Economist“ im Jahr 1986 den sogenannten „Burgernomics“ ein, der darin besteht, dass seitdem alljährlich die Preise von McDonald's „Big Mac“ in verschiedenen Ländern veröffentlicht werden (<http://www.economist.com/content/big-mac-index>, Zugriff vom 6. Februar 2014). Der Vorteil dieses „Big Mac Index“ ist, dass er auf der ganzen Welt produziert und konsumiert wird; der „Warenkorb“ für die Herstellung eines Big Macs enthält die Preise für Rindfleisch, Käse, Salat, Zwiebeln und Brot. Da das Rezept weltweit identisch ist, gibt es keine Unterschiede für die benötigten Anteile dieser fünf Zutaten zwischen den Ländern. Der *Big Mac Index* ist eindimensional und wird aus genannten Indikatoren errechnet; diese haben verschiedene „Gewichte“, zum Beispiel ist der Anteil von Rindfleisch höher als der von Zwiebeln.

In den Sozialwissenschaften werden die benötigten Indikatoren oft mit Hilfe von Umfragen (Reinecke, Kapitel 44 in diesem Band) erhoben, mit denen die Einstellungen der Bevölkerung gemessen werden. Die Erhebung fängt an mit der Wahl von verständlich formulierten Fragen (Porst, Kapitel 50 in diesem Band) und Antwortvorgaben (Franzen, Kapitel 51 in diesem Band), sie beinhaltet eine geeignete Stichprobenziehung (Häder/ Häder, Kapitel 18 in diesem Band) und ggf. auch ein angemessenes Verhalten der Interviewer (Glantz/Michael, Reinecke, Blasius, Kapitel 21, 44 und 22 in diesem Band). Die Gültigkeit der Messungen (Krebs/Menold, Kapitel 30 in diesem Band) hängt von der Qualität des Messinstrumentes ab, mit einem schlechten Fragebogen können Einstellungen nicht valide gemessen werden. Während in der Ökonomie z.B. die Kaufkraftparität im Zentrum des Interesses steht, ist es in den Sozialwissenschaften so etwas wie Anomie, politische Entfremdung, (christliche) Religiosität oder nationale Identität. Im Folgenden sollen drei klassische Methoden der Skalierung vorgestellt werden, diese sind mit den Namen Likert, Guttman und Thurstone verbunden.

### 83.3 Likert-Skalierung

Wenn in den Sozialwissenschaften von Likert-Skalierung (Likert 1932) gesprochen wird, sind meistens Fragen mit fünf Antwortalternativen gemeint, mit denen eine bestimmte Dimension von Einstellungen gemessen werden soll. Die Antwortvorgaben reichen von starker Zustimmung über einen neutralen Mittelwert bis zu starker Ablehnung, so z.B. fünf Fragen im deutschsprachigen Teil des ISSP 2003 (International Social Survey Program, <http://www.issp.org/>; Mochmann, Kapitel 14 in diesem Band), die mit folgendem Statement eingeleitet wurden: „Nun einige Fragen zu den Beziehungen zwischen Deutschland und anderen Ländern: Inwieweit stimmen Sie den folgenden Aussagen zu oder nicht?“ Als Antwortalternativen wurden vorgegeben: 1 = stimme voll und ganz zu, 2 = stimme eher zu, 3 = weder noch, 4 = stimme eher nicht zu, 5 = stimme überhaupt nicht zu.

- Deutschland sollte die Einfuhr ausländischer Produkte beschränken, um seine eigene Wirtschaft zu schützen.
- Bei bestimmten Problemen wie der Umweltverschmutzung sollten internationale Institutionen das Recht haben, Lösungen durchzusetzen.
- Deutschland sollte seine eigenen Interessen verfolgen, selbst wenn dies zu Konflikten mit anderen Ländern führt.
- Ausländern sollte es nicht erlaubt sein, in Deutschland Grund und Boden zu erwerben.
- Das deutsche Fernsehen sollte deutschen Filmen und Programmen den Vorzug geben.

Ein erster Ansatz, um diese Art von Daten zu analysieren, wurde von Likert (1932) formuliert, seine Ideen sind unter den Begriffen „Likert-Skalierung“ und „Technik der summierten Ratings“ bekannt. Die Methode kann in drei Schritten zusammengefasst werden:

Erstens, das Sammeln einer großen Menge (ca. 100) von positiv und negativ formulierten Fragen, die mögliche *Indikatoren* für die Dimension des Interesses sind. In dem oben gegebenen Beispiel kann diese Dimension als „Nationalbewusstsein“ bezeichnet werden, wobei die zweite Frage in gegensätzlicher Richtung zu den vier anderen formuliert wurde.

Zweitens, die Bewertung der gesammelten Fragen: Im Rahmen einer *Stichprobe*, die repräsentativ für die Zielpopulation ist, sollen Befragte angeben, wie stark sie den einzelnen Einstellungen zustimmen. Dann werden die Fragen ausgeschlossen, auf die sehr unterschiedliche Befragte sehr ähnliche Antworten gaben, und jene Fragen, auf die sehr ähnliche Befragte sehr unterschiedliche Antworten gaben.

Im dritten Schritt wird die *Trennschärfe* geprüft, wofür die Befragten auf der Basis ihrer Angaben in Quartile aufgeteilt werden, d.h. in vier gleich große Gruppen. Unterscheiden sich die beiden Extremgruppen (die 25% der Befragten mit den niedrigsten Werten von jenen 25% der Befragten mit den höchsten Werten) statistisch nicht signifikant voneinander, dann wird das entsprechende Item ausgeschlossen. Für die anschließende Hauptuntersuchung werden nur jene 10 bis 20 Fragen verwendet, die die Extremgruppen am besten differenzieren. Damit ist gewährleistet, dass nur Variablen verwendet werden, mit denen die Befragten unterschieden werden können; Fragen, denen alle Befragten zustimmen bzw. die alle Befragten ablehnen, werden ausgeschlossen.

**Tab. 83.1** Summenindex

Fall	F1	F2	F3	F4	F5	Summe	Index
1	1	1	1	1	2	6	1,2
2	5	5	4	5	5	24	4,8
3	3	3	3	3	3	15	3,0
4	1	5	1	5	3	15	3,0
5	3	1	5	1	5	15	3,0
6	2	2	4	4	3	15	3,0
7	4	2	1	3	5	15	3,0
8	5	4	3	2	1	15	3,0
9	1	5	5	1	k.A.	12	3,0
10	3	k.A.	k.A.	k.A.	3	6	3,0
...							

Eine sehr einfache Art der Auswertung dieser Daten ist die Bildung eines *additiven Indexes*; dafür werden die Werte der Antwortkategorien zuerst addiert, und anschließend wird der so ermittelte Wert durch die Anzahl der gültigen Angaben dividiert. Um diese Berechnung sinnvoll durchführen zu können, müssen zuvor die Fragen so rekodiert (oder *gepolt*) werden, dass sie alle die gleiche Richtung haben, z.B. dass der positivste Werte immer mit „1“ kodiert wird und der „negativste“ immer mit „5“. Dies soll an einem kleinen Beispiel mit den fünf oben angegebenen Fragen zum Nationalbewusstsein (abgekürzt mit F1 bis F5) verdeutlicht werden (Tab. 83.1).

Die Interpretation der resultierenden Indexwerte ist an den Endbereichen der Skala relativ eindeutig, in Tabelle 1 gilt dies für die Fälle 1 und 2. In beiden Fällen sind die ursprünglichen Angaben der Befragten relativ eindeutig zu rekonstruieren, in beiden Fällen kann es nur fünf verschiedene Antwortkombinationen geben – und diese unterscheiden sich auch nur dadurch, bei welcher Frage die „2“ bzw. die „4“ angegeben wurde. Problematisch ist jedoch der gesamte mittlere Bereich der Skala, in Tabelle 1 exemplarisch durch die Fälle 3 bis 8 dargestellt. Hier wird deutlich, dass sehr unterschiedliche Angaben zu exakt dem gleichen Indexwert führen. Es bleibt damit unklar, ob eine Person abwechselnd eine schwache Zustimmung und eine schwache Ablehnung angab – und bei welchen Fragen dies jeweils der Fall war – ob sie alle Fragen mit der mittleren Kategorie beantwortete oder ob sie gar abwechselnd eine starke Zustimmung und eine starke Ablehnung äußerte (was ein Zeichen einer eher willkürlichen Beantwortung der Fragen wäre, die allerdings häufiger vorkommt als allgemein angenommen wird, dazu ausführlich Blasius/Thiessen 2012).

Die exakte Übereinstimmung der Werte bei dieser Art von Indexbildung ist auch eine Folge davon, dass alle Fragen das gleiche Gewicht (die gleiche Wichtigkeit) haben. In dem gegebenen Beispiel wird damit implizit angenommen, dass alle fünf Fragen die zugrundeliegende Dimension „Nationalbewusstsein“ in dem gleichen Maß erklären, was inhaltlich

eine sehr starke Vereinfachung der Realität ist. Nur angedeutet wird in Tab. 83.1 das Problem der fehlenden Werte (Engel/Schmidt, Kapitel 23 in diesem Band), wo die gleichen Indexwerte auf der Basis eines verkürzten Satzes von Fragen – und damit auf einer anderen und ggf. kaum noch vergleichbaren Basis – errechnet werden (zu diesem Problem und möglichen Lösungen, siehe Cielebak/Rässler, Kapitel 26 in diesem Band).

Ein Ansatz für eine unterschiedliche Gewichtung der Fragen ist mittels der *Hauptkomponentenanalyse* möglich (Hotelling 1933), die allerdings ein metrisches Datenniveau voraussetzt. Obwohl die Abstände zwischen den Ausprägungen numerisch gleich sind (in dem Beispiel jeweils „1“, z.B.  $4-3 = 3-2 = 2-1 = 1$ ; vgl. Tab. 83.1), sind sie im latenten Bereich ungleich. Dies kann relativ einfach daran veranschaulicht werden, dass wenn die zweite Kategorie mit „2 = stimme zu“ statt mit „2 = stimme eher zu“ bezeichnet wird (und die vierte Kategorie entsprechend mit „4 = stimme nicht zu“ statt mit „4 = stimme eher nicht zu“), sich die Randverteilungen der Ausprägungen ändern (Franzen, Kapitel 51 in diesem Band). In dem gegebenen Beispiel werden die Extremwerte im zweiten Fall seltener verwendet: Wenn der „wahre Wert“ bei „stimme zu“ liegt, wird im ersten Fall vermutlich immer dann die Antwort „stimme vollkommen zu“ gewählt, wenn der „wahre“ Wert dichter an dieser Ausprägung als an „stimme eher zu“ liegt. Daraus folgt, dass auch die einzelnen Variablenausprägungen ein unterschiedliches Gewicht erhalten sollten bzw. dass die „wahren“ Abstände zwischen den Ausprägungen für alle Variablen in einem vorzuziehenden Schritt neu geschätzt werden sollten; dies kann z.B. mit Hilfe der kategorialen (oder nicht-linearen) Hauptkomponentenanalyse (Gifi 1990) erfolgen.

Aufbauend auf diesen und ähnlichen Ideen gibt es heutzutage eine Vielzahl von Techniken, mit denen derartige Fragebatterien ausgewertet werden können, die zudem weder die Eindimensionalität der Fragen noch der gesamten Fragenbatterie voraussetzen. Eine Übersicht über diese Methoden und weitere Literaturhinweise geben Blasius/Baur (Kapitel 79 in diesem Band).

### 83.4 Guttman-Skalierung

Eine andere Möglichkeit Einstellungen zu messen, ist mittels eines Satzes von dichotomen Fragen, die einen unterschiedlichen „Schwierigkeitsgrad“ haben, entsprechend dessen sie geordnet werden können. Im Bereich des Sports ist dies sehr leicht nachzuvollziehen, z.B. beim Hochsprung, wo mit jedem höheren „Skalenwert“ der nächstniedrige sogar direkt als „erreicht“ bewertet wird: Wenn jemand die 1,80 überspringt, so hat er auch die 1,75 übersprungen und wer 1,85 übersprungen hat, der bekommt auch die 1,80 (und ebenso die 1,84) gutgeschrieben. In den Sozialwissenschaften ist es nicht ganz so einfach, eine analoge Skala zu entwickeln, hier ein kleines hypothetisches Beispiel mit vier Fragen, die alle mit „ja“ oder „nein“ beantwortet werden sollen:

- Mich stört es nicht, wenn Ausländer in meiner Nachbarschaft wohnen.
- Mich stört es nicht, wenn Ausländer in dem gleichen Haus wie ich wohnen.
- Mich stört es nicht, wenn Ausländer auf der gleichen Etage wie ich wohnen.
- Ich würde meine Wohnung jederzeit auch an Ausländer vermieten.

In diesem Beispiel wird angenommen, dass, wer seine Wohnung an Ausländer vermietet, auch mit Ausländern auf der gleichen Etage wohnen würde, und dass, wer dazu bereit ist, auch mit Ausländern im gleichen Haus wohnen würde. Die „schwierigste“ Frage ist demnach, die eigene Wohnung an Ausländer zu vermieten, die „leichteste“, mit Ausländern in der gleichen Nachbarschaft zu leben. Ein weiteres Beispiel, welches z.B. im World Value Survey 1996 verwendet wurde, basiert auf dem christlichen Glauben. Die entsprechenden Fragen beinhalten u.a. den Glauben an Gott, daran, dass Menschen eine Seele haben, an ein Leben nach dem Tod, den Himmel, den Teufel und die Hölle. Auch hier gibt es eine inhärente Ordnung, so ist der Glaube an die Seele einfacher als der Glaube an den Himmel, und dieser ist einfacher als der Glaube an die Hölle (so das Ergebnis für die Niederlande auf Basis von Daten des World Values von 1990, van Schuur/Blasius 2006).

Guttman (1944, 1950) entwickelte eine Methode, die auf einer Reihe von derart dichotomen Variablen basiert; diese Methode ist als *Skalogrammanalyse*, *kumulative Skalierung* oder *Guttman-Skalierung* bekannt, die Daten werden als *Dominanzdaten* bezeichnet. Wenn die Reihenfolge der Elemente perfekt ist, dann sollte es wie beim Hochsprung eine Reihenfolge der Variablen geben, bei der jeder Befragte bis zu einem gewissen Punkt alle Fragen mit „ja“ beantwortet, alle folgenden dann mit „nein“. In einer perfekten Guttman-Skala würde niemand an die Hölle glauben, der nicht auch an den Himmel glaubt, und alle, die auch an den Himmel glauben, würden auch an die Seele glauben. Eine derart perfekte Guttman-Skala mit 10 Personen und neun Fragen (A bis I) ist in Tab. 83.2 wiedergegeben.

**Tab. 83.2** Skalogramm (Dominanzdaten)

Entsprechend Tab. 83.2 beantwortet die erste Person alle Fragen mit „nein“, die zweite Person beantwortet nur die erste Frage mit „ja“ und alle anderen mit „nein“, und die letzte Person beantwortet alle Fragen mit „ja“. In dieser Skala gibt es keinen einzigen Fehler. Als solcher würde jede „Null“ bewertet, die zwischen zwei Einsen und jede „Eins“, die zwischen zwei Nullen steht.

In der empirischen Umsetzung kann der unterschiedliche Schwierigkeitsgrad der Fragen bestimmt werden, indem die Zeilen und Spalten solange vertauscht werden, bis die Anzahl der Nullen zwischen den Einsen und die der Einsen zwischen den Nullen minimal ist. Anders als in dem hypothetischen Beispiel aus Tab. 83.2 werden bei empirischen Erhebungen immer „Fehler“ bleiben; es gibt eben Personen, die an die Hölle, aber nicht an den Himmel glauben. Verursacht eine Frage (gegebenenfalls auch eine Person) zu viele Fehler, so passt sie nicht in die erwartete eindimensionale Skala (Ausmaß der christlichen Religiosität) und kann aus der Analyse herausgenommen werden. Ob dies inhaltlich sinnvoll ist, ist eine andere Frage und muss im konkreten Anwendungsfall diskutiert werden.

In diesem *deterministischen Modell* ist die Abfolge zwischen den einzelnen Stufen immer identisch, es folgt einer Treppenfunktion, die in Tab. 83.2 durch die Linien ange deutet wird. Nun wird der Schwierigkeitsgrad, mit der eine weitere Variable als „erreichtes Ziel“ aufgenommen wird, nicht linear ansteigen, sondern er wird einer Verteilungsfunktion folgen. Eine entsprechende Erweiterung der *deterministischen Guttman-Skalierung* ist die *probabilistische Rasch-Skalierung*, wobei angenommen wird, dass die Wahrscheinlichkeit für das Erreichen der nächsten Stufe auf der Basis der logistischen Verteilung berechnet wird (Rost 1996).

### 83.5 Thurstone-Skalierung

Thurstone (1928) war vermutlich der erste Autor, der eine formale Methode zur Messung von Einstellungen vorschlug. Sein Ausgangspunkt ist, dass Einstellungen zwar mehrdimensional sind, dass aber jeweils nur eine Dimension betrachtet wird. Als Beispiele für derartige Einstellungen diskutiert Thurstone (1928: 544) die drei (latenten) Merkmale „Pazifismus-Militarismus“, „Prohibition“ und „Einstellungen gegenüber der Kirche“.

In einem ersten Schritt sollen verschiedene Experten kurze Statements zu den genannten Dimensionen aufschreiben, bezogen auf die erste der oben genannten Skalen gehören dazu z.B. Angehörige des Militärs und Pazifisten. Des Weiteren soll die Literatur bearbeitet werden, so dass das gesamte Spektrum der gesuchten Einstellungsdimension mit insgesamt 100 bis 150 Statements abgedeckt wird.

An diese erste Liste von Statements werden Kriterien gelegt, so z.B. dass sie so kurz wie möglich sein sollten und dass sie die späteren Befragten nicht ermüden und deshalb davon abhalten, die gesamte Liste zu lesen. Die Liste der Statements sollte dadurch auf 80 bis 100 reduziert werden (Thurstone 1928: 545).

Im nächsten Schritt werden 200 bis 300 Beurteiler gebeten, die einzelnen Statements nach ihrer *Günstigkeit* zu beurteilen und dabei in elf gleich große Intervalle einzuteilen. In

dem gegeben Beispiel soll das erste Intervall die Statements beinhalten, welche die stärkste pro-militärische Einstellung wiedergeben, das zweite Intervall jene Statements, die eine etwas schwächere pro-militärische Einstellung widerspiegeln, das mittlere Intervall soll einer neutralen Einstellung entsprechen und das letzte Intervall der stärksten pro-pazifistischen Einstellung. Die eigene Meinung soll hierbei nicht berücksichtigt werden, es geht ausschließlich um die Zuordnung der Statements zu den 11 Intervallen.

Wurden die Statements von allen Beurteilern auf dieser *Günstigkeitsskala* eingeteilt, dann werden zuerst jene aussortiert, die eine zu große Varianz haben, also Statements, die sehr verschiedenen Intervallen zugeordnet wurden. Des Weiteren muss die Zuordnung zu den Intervallen unabhängig von der Einstellung zu dem zu messenden Objekt erfolgen, als Kriterium schlägt Thurstone (1928: 548) die relative Platzierung auf der Skala vor: „Wenn z.B. 85% der Angehörigen des Militärs angeben, dass Item A ein militaristisches Statement als Item B ist, dann müsste auch ungefähr der gleiche Anteil der Pazifisten zu dieser Bewertung kommen“ (eigene Übersetzung).

Von den nach dieser Prozedur verbleibenden Variablen werden die Mittelwerte berechnet, das sind die sogenannten *Itemkennwerte*. Werden fünf Beurteiler und das Statement (x) mit den Beurteilungen  $x_1 = 3$ ,  $x_2 = 4$ ,  $x_3 = 2$ ,  $x_4 = 5$ ,  $x_5 = 4$  betrachtet, dann ist der *Itemkenntwert* dieses Statements

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{5}(3 + 4 + 2 + 5 + 4) = \frac{18}{5} = 3,6$$

Zu Thurstones Zeiten, als in den Vereinigten Staaten die Prohibition (von Alkohol) herrschte, könnte das Item „Alle Hersteller von Alkohol gehören ins Gefängnis“ einen Mittelwert von -4,5 gehabt haben, das Item „Ich habe gute Freunde, die Alkohol selbst brennen“ einen Wert von +4,2 und „Ich trinke gerne ein Bier“ einen Wert von +0,5.

In der eigentlichen Erhebung werden den Befragten die final ausgewählten Statements vorgelegt die dann mit „ja“ oder „nein“ beantwortet werden sollen. Die *Itemkennwerte* der „ja“-Antworten werden für jeden Befragten addiert und durch die Anzahl der positiven Antworten dividiert, das Ergebnis ist der *Personenkennwert*. Da die Befragten idealerweise entweder den positiven, den negativen oder den neutralen Statements zustimmen werden, nicht aber positiven und negativen Statements, kann die optimale Struktur der Antworten für die Befragten ( $N=7$ ; Tab. 83.3) und die neun Statements (A bis I) in Form eines *Parallelogramms* abgebildet werden.

In Tab. 83.3 sind die neun Statements in sukzessiver Reihenfolge dargestellt. Dabei ist A das negativste und I das positivste Statement; der Itemkenntwert bleibt hierbei unberücksichtigt. Entsprechend Tab. 83.3 hat Person 1 die drei negativsten Statements mit „ja“ beantwortet, alle anderen mit „nein“, Person 7 hat die drei positivsten Statements mit „ja“ beantwortet und die verbleibenden mit „nein“, während Person 4 die drei mittleren Werte mit „ja“ beantwortet hat, die anderen mit „nein“.

In diesem idealtypischen Beispiel gibt es keine „Null“ zwischen zwei Einsen, und keine „Eins“ zwischen zwei Nullen, in der empirischen Anwendung gilt dies nicht. Nullen zwischen den Einsen kommen dadurch zustande, dass die Befragten eine

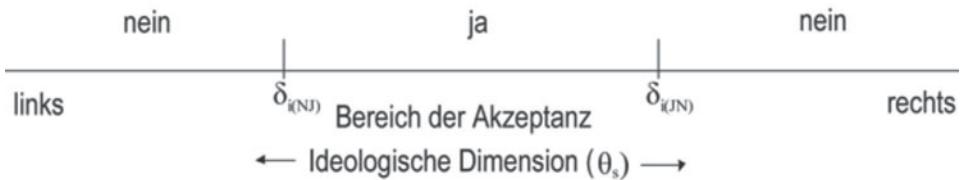
**Tab. 83.3** Parallelogrammstruktur

	A	B	C	D	E	F	G	H	I
Person 1	1	1	1	0	0	0	0	0	0
Person 2	0	1	1	1	0	0	0	0	0
Person 3	0	0	1	1	1	0	0	0	0
Person 4	0	0	0	1	1	1	0	0	0
Person 5	0	0	0	0	1	1	1	0	0
Person 6	0	0	0	0	0	1	1	1	0
Person 7	0	0	0	0	0	0	1	1	1

unterschiedliche Anzahl von Fragen bejahen. Dabei muss das fehlende „ja“ nicht auf die vorhandene Einstellung zurückzuführen sein, es kann auch sein, dass Befragte dies spezielle Statement inhaltlich nicht verstanden haben und ihm deshalb nicht zustimmten. Kommt es bei einem Statement zu relativ vielen Nullen, die an beiden Seiten von Einsen umgeben sind, so kann es nachträglich aussortiert werden. Damit wird auch gewährleistet, dass von zwei Statements, deren Itemkennwerte einander sehr ähnlich sind, in der Regel (mindestens) eins nicht weiter berücksichtigt wird.

Die implizite Ordnung der Statements setzt auch voraus dass die Personen transitiv antworten (können). In dem gegebenen Beispiel ist A eine negativere Aussage als B und B ist eine negativere Aussage als C und wenn, wie vorausgesetzt, die Skala eindimensional ist, dann folgt daraus, dass A negativer als C ist. Auf ähnliche Weise wie bei den Statements, wo jene, die zu viele „Fehler“ verursachen (Nullen zwischen den Einsen), aussortiert werden, könnten auch jene Personen aussortiert werden, die zu oft nicht-transitiv antworten, d.h. die zu viele „Fehler“ machten.

Das hier dargestellte Modell von Thurstone ist deterministisch, die Itemkennwerte werden bei der Berechnung der „Fehler“ nicht berücksichtigt. Statt von einer deterministischen Ordnung, in Tab. 83.3 durch die „Treppenfunktion“ angedeutet, wird heutzutage meistens von einem probabilistischen Modell ausgegangen, demzufolge es einen Wertebereich gibt, in dem der Befragte den Statements mit einer hohen Wahrscheinlichkeit zustimmt. Diese Modelle werden als Unfolding-Modelle (Rost 1996) bezeichnet. In Abb. 83.1 ist ein derartiger „Bereich der Akzeptanz“ auf einer ideologischen Dimension (links-rechts) wiedergegeben, bei der die beiden Schnittpunkte, wo es von einer Ablehnung zu einer Zustimmung und umgekehrt kommt ( $\delta_i$ ), eingezeichnet sind.



**Abb. 83.1** Ideologische Dimension, Bereich der Zustimmung

## 83.6 Fazit

Kenntnisse über die Struktur der Daten sind heute ebenso wichtig wie zu den Zeiten, als die ersten Skalierungsmodelle entwickelt wurden. Nur wenn die Struktur bekannt ist, können sinnvolle Aussagen über inhaltliche (sozialwissenschaftliche) Zusammenhänge gemacht werden. Mit den heutigen Computern und Statistikprogrammen sind sehr viele neue Möglichkeiten vorhanden, diese Strukturen zu untersuchen, in vielen Fällen ist schon SPSS oder STATA ausreichend, in anderen kann auf frei verfügbare Programme in R (<http://www.r-project.org/>) zurückgegriffen werden.

Einfache Mittelwertindizes werden zwar immer noch gerechnet, aber es ist heutzutage eigentlich ebenso einfach, die Werte der einzelnen Variablen, gegebenenfalls auch der einzelnen Variablenausprägungen, zu gewichten, also z.B. eine (kategoriale) Hauptkomponentenanalyse zu rechnen, und die resultierenden Faktorwerte zu verwenden. Dasselbe gilt für das klassische Verfahren der Guttman-Skalierung, welche durch die Rasch-Skalierung abgelöst wurde, und für die Thurstone-Skalierung, die durch das Unfolding abgelöst wurde. Rasch-Skalierung und Unfolding sowie deren Erweiterungen (Rost 1996) sind Bestandteil der neueren Testtheorie und als solcher in der Literatur beschrieben. Die von den Klassikern vorgeschlagenen Methoden zur Fragengewinnung sind sehr aufwändig und kostspielig und werden in dieser Form nicht (mehr) praktiziert, die entsprechenden aktuellen Praktiken werden insbesondere in den Beiträgen von Porst und Franzen (Kapitel 50 und 51 in diesem Band) beschrieben.

## Literatur

- Blasius, Jörg/Thiessen, Victor (2012): Assessing the Quality of Survey Data. London: Sage
- Coombs, Clyde H. (1964). A Theory of Data. New York: Wiley
- Greenacre, Michael/Blasius, Jörg (Hg.) (2006): Multiple Correspondence Analysis and Related Methods. Boca Raton, Florida: Chapman & Hall
- Guttman, Louis (1944): A Basis for Scaling Qualitative Data. In: American Sociological Review 9: 139-150
- Guttman, Louis (1950): The Basis for Scalogram Analysis. In: Stouffer et al. (Hg.): 60-90
- Likert, Rensis (1932): A Technique for the Measurement of Attitudes. In: Archives of Psychology 140: 1-55
- Hotelling, Harold (1933): Analysis of a Complex of Statistical Variables into Principal Components. In: Journal of Educational Psychology 24: 417-441 und 498-520
- MacDonell, W.R (1902): On Criminal Anthropometry and the Identification of Criminals. In: Biometrika 1: 177-227
- Pearson, Karl P. (1901): On Lines and Planes of Closest Fit to Systems of Points in Space. In: The London, Edinburgh and Dublin Philosophical Magazine and Journal 6: 559-572
- Rost, Jürgen (1996): Testtheorie und Testkonstruktion. Bern: Hans Huber
- Spearman, Charles E. (1904): 'General Intelligence' Objectively Determined and Measured. In: American Journal of Psychology 15: 201-293
- Stouffer, Samuel A./Guttman, Louis/Suchman, Edward A./Lazarsfeld, Paul F./Star, Shirley A./Clausen, John A. (eds.) (1950): Measurement and Prediction. Princeton: Princeton University Press
- Thurstone, Louis L. (1928): Attitudes can be Measured. In: The American Journal of Sociology 33: 529-554
- Van Schuur, Wijbrandt H./Blasius, Jörg (2006): Visualizing Structures of Dichotomous Unidimensional Item Response Theory Data by Multiple Correspondence Analysis. In: Greenacre/ Blasius (Hg.): 237-258

Rainer Metz und Helmut Thome

---

## 84.1 Einleitung

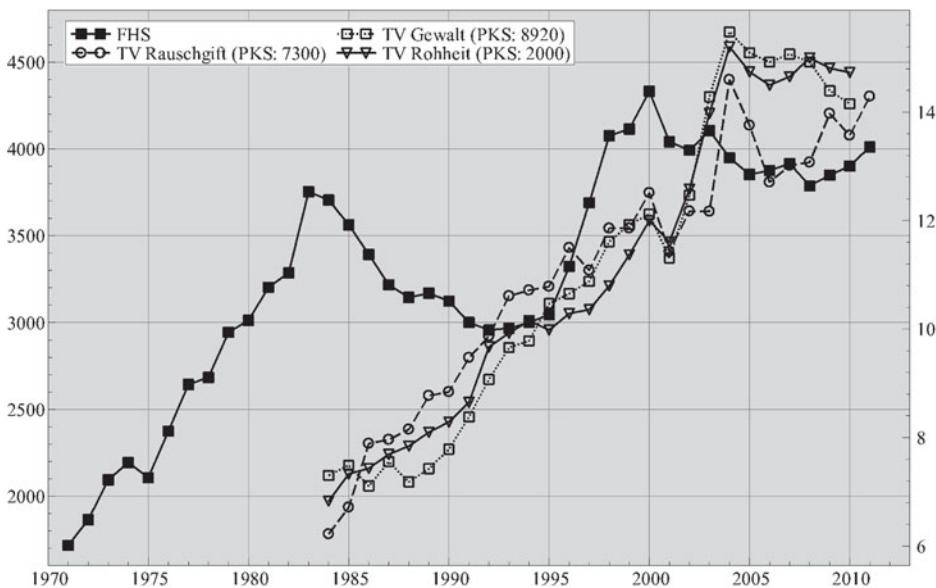
*Zeitreihenanalyse* ist ein Sammelbegriff für statistische Verfahren zur Analyse von Zeitreihen. Unter einer *Zeitreihe* versteht man eine Serie von zeitlich geordneten Messergebnissen, die aus Erhebungen stammen, die relativ häufig (mindestens etwa 30 bis 40 Mal) in gleichbleibenden Abständen am gleichen Objekt zur gleichen Merkmalsdimension vorgenommen wurden.

Abbildung 84.1 zeigt als Beispiel vier Zeitreihen. Dabei handelt es sich um die Anzahl der Strafgefangenen (SG) im Vollzug von Freiheitsstrafe (FHS) (linke Achse) sowie um die in der Polizeilichen Kriminalstatistik ausgewiesenen Tatverdächtigen (TV) für Gewalt-, Rohheits- und Rauschgiftdelikte (in Tausend, rechte Achse) in Hessen.

Formal lässt sich eine Zeitreihe als eine chronologisch geordnete Folge  $Y_1, Y_2, \dots, Y_T$  numerischer Größen definieren, wobei  $T$  die Länge einer Zeitreihe bezeichnet. Die einzelnen „Fälle“ einer Zeitreihenanalyse sind also nicht wie in der Querschnittsanalyse (Lück/Landrock, Kapitel 28 in diesem Band) durch die verschiedenen Erhebungs- oder Untersuchungseinheiten konstituiert (für die jeweils nur eine zeitgleiche Messung vorliegt), sondern durch die Vielzahl der *Messzeitpunkte*.

*Untersuchungseinheiten* soziologischer Zeitreihenanalysen sind häufig Länder oder andere regionale Einheiten, für die z.B. monatliche Arbeitslosendaten oder jährliche Kriminalitätsraten erhoben werden, aber auch Organisationen (z.B. politische Parteien mit der Entwicklung ihrer Mitgliederzahlen) oder (eher in der psychologischen Forschung) einzelne Personen, die über eine längere Zeitstrecke z.B. aufzeichnen, wie viel Zeit sie wöchentlich mit Freunden gemeinsam verbringen.

Zeitreihen sind somit eine *spezielle Form von Längsschnittdaten* (dazu Schupp, Kapitel 73 in diesem Band). In diesem Artikel können wir nur einige ausgewählte Analysetechniken und Erkenntnismöglichkeiten darstellen, die die Zeitreihenanalyse eröffnet.



**Abb. 84.1** Strafgefangene und Tatverdächtige in Hessen

## 84.2 Deskription und statistisches Modell

Jede Zeitreihenanalyse sollte mit der *Deskription der Reihe(n)* beginnen. Dabei versucht man, erste Anhaltspunkte über spezifische kurz- und langfristige Entwicklungsmuster in und zwischen Zeitreihen, aber auch über historische Besonderheiten, wie z.B. Strukturbrüche oder Ausreißer zu entdecken. Unverzichtbar ist eine graphische Inspektion des Verlaufs der Reihe(n).

Nach der *deskriptiven Inspektion* wird man versuchen, interessierende Entwicklungsmuster, historische Besonderheiten bzw. strukturelle Zusammenhänge zu identifizieren und mit bestimmten *Kennzahlen (Parametern)* zu schätzen. Dies geschieht mit Hilfe *statistischer Modelle*. Sie sind Grundlage für die Beschreibung der kurz- und langfristigen Dynamik und für die Zerlegung von Zeitreihen in spezifische Komponenten (wie Trend, Zyklus und Saison), für die Identifizierung von Ausreißern und Strukturbüchen, aber auch für die Analyse spezifischer Einflussgrößen (bspw. einschneidende Ereignisse) sowie die Zusammenhangs- bzw. Kausalanalyse (Kühnel/Dingelstedt, Kapitel 80 in diesem Band) auf der Basis mehrerer Zeitreihen:

- Unter „Trend“ versteht man eine langfristig ansteigende oder abfallende Entwicklungstendenz (z.B. den Anstieg der Körperverletzungsdelikte in der BRD in der zweiten Hälfte des 20. Jahrhunderts),

- unter „*Zyklus*“ ein sich periodisch wiederholendes (z.B. wellenförmiges) Ablaufmuster (z.B. Auf- und Abschwung der allgemeinen Wirtschaftslage),
- unter „*Saison*“ jahreszeitlich bedingte Schwankungen (wie sie z.B. bei Arbeitslosendaten auftreten).
- Als „*Ausreißer*“ bezeichnet man einzelne Werte, die (aus bekannten oder unbekannten Gründen) besonders stark (positiv oder negativ) vom zu diesem Zeitpunkt erwartbaren mittleren Niveau der Zeitreihe abweichen.
- Unter „*Strukturbrüchen*“ versteht man eine (mehr oder minder rasche) Veränderung in der Entwicklungsdynamik der Zeitreihe, z.B. hinsichtlich der Einflussstärke, mit der vorangegangene Zeitreihen-Werte die Realisierung der nachfolgenden Werte mit bestimmen.

Solche statistischen Modelle werden nicht nur für die Komponentenzerlegung oder die Kausalanalyse, sondern auch für die Prognose zukünftiger Werte benötigt. Das Problem der Modellierung besteht darin, aus der Vielzahl von verfügbaren Modellen, die i.d.R. ganz unterschiedliche Eigenschaften aufweisen, ein Modell auszuwählen, das sich für die Analyse als „geeignet“ erweist und die Zeitreihen, bzw. den Zusammenhang zwischen ihnen, „angemessen“ beschreibt. Ob ein Modell eine angemessene Beschreibung liefert, hängt u.a. davon ab, ob sich die Ergebnisse auch substantiell interpretieren lassen.

In der *univariaten Analyse einzelner Zeitreihen* geht es vor allem um die *Unterscheidung verschiedener Komponenten*, die aus unterschiedlichen kausalen Dynamiken erwachsen können, also aus unterschiedlichen Konstellationen von Einflussgrößen, die kurz- oder längerfristig wirken. Dies betrifft insbesondere die Identifikation linearer oder nicht-linearer *Trendkomponenten* („lokal“ für kürzere aufeinanderfolge Zeitabschnitte oder „global“ für den gesamten Zeitraum) sowie von *Zyklen* und, bei „unterjährigen“ Reihen (für die vierteljährliche, monatliche oder gar wöchentliche Daten vorliegen), von saisonalen Komponenten. Solche längerfristig beobachtbaren Verlaufsmuster, die für sich alleine oder in Kombination miteinander auftreten können, resultieren aus Einflussgrößen, die über Zeit relativ stabil wirksam bleiben. Sie werden aber in aller Regel von einer sog. „*Restkomponente*“ überlagert, also von Effekten, die einer ständigen Abfolge kurzfristig auftretender, inhaltlich oft unbekannter und diverser („zufälliger“) Einflussgrößen entstammen. Schließlich kann noch eine weitere Komponente hinzutreten, die für außergewöhnliche Einflüsse steht, die selten und unregelmäßig auftreten und bspw. mehr oder weniger abrupte Niveauverschiebungen hervorrufen.

Bei der Bestimmung einzelner Komponenten einer Zeitreihe lassen sich grundsätzlich zwei unterschiedliche *Herangehensweisen* unterscheiden, und zwar der *Filter-Design-Ansatz* und der *modellbasierte Ansatz*.

Beim *Filter-Design-Ansatz* werden die Komponenten über ihre Schwingungsdauer definiert; so wird z.B. der Trend als eine Schwingung mit unendlicher Dauer aufgefasst und dann versucht, einen Algorithmus zu konstruieren, der genau diese Schwingung(en) aus der Zeitreihe herausfiltert. Häufig verwendete Filter sind der Hodrick-Prescott-Filter

(1997), der Baxter-King-Filter (1999) und der Christiano-Fitzgerald-Filter (2003). Diesen Ansatz können wir hier aus Platzgründen nicht weiter erläutern.

Hinsichtlich verschiedener Varianten des *modellbasierten Ansatzes* sei einführend der einfache Fall eines Trendmodells vorgestellt, bei dem angenommen wird, dass der Trend der Reihe über den gesamten Beobachtungszeitraum linear deterministisch verläuft, so dass er im Rahmen der *einfachen linearen Regressionsanalyse* (Blasius/Baur, Kapitel 79 in diesem Band) modelliert werden kann. Dazu schätzt man folgende Gleichung:

$$(1) \quad Y_t = \mu + \beta \cdot t + \varepsilon_t$$

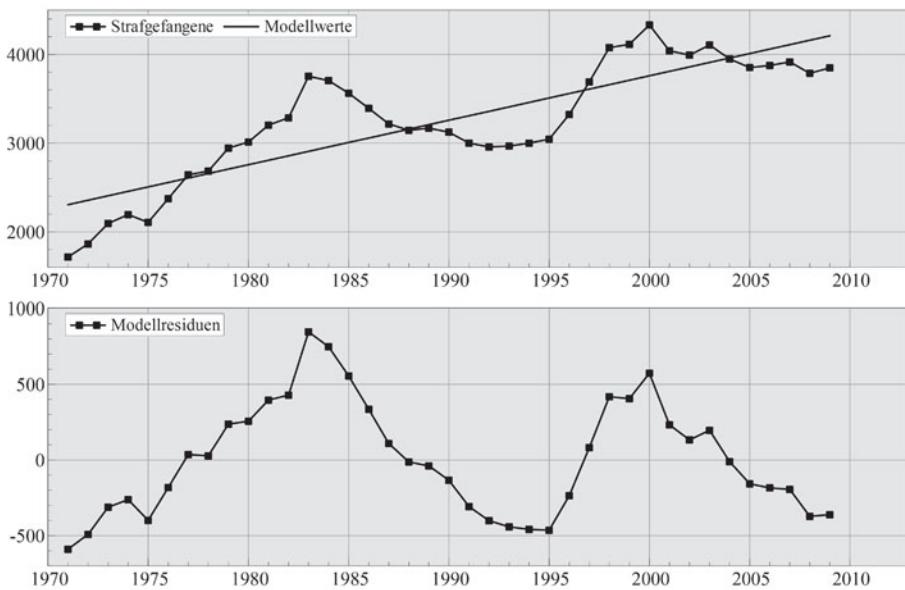
Wenn die Konstante „ $\mu$ “ und der Steigungskoeffizient „ $\beta$ “ ermittelt sind, lassen sich nicht nur die Trendwerte für alle Messzeitpunkte  $t$  als Prognosewerte der Gleichung (1) errechnen, sondern auch die Residuen ( $\varepsilon_t$ ), die sich als Differenz zwischen (beobachtetem) Reihen- und (geschätztem) Trendwert ergeben.

Liegen unterjährige Werte, also z.B. Monats- oder Quartalszahlen vor, lässt sich Modell (1) leicht um eine konstante *Saisonkomponente* erweitern, indem man zusätzlich ( $s-1$ ) Dummy-Variablen einführt. Dabei steht „ $s$ “ für die Saisonspanne, also die Zahl der Messzeitpunkte, über die sich ein saisonaler Zyklus jeweils vollzieht (bei Monatsdaten  $s = 12$ , also  $(s-1) = 11$  Dummy Variablen).

Betrachten wir dazu ein Beispiel. Abbildung 84.2 (obere Graphik) zeigt die Anzahl der Strafgefangenen (SG) im Vollzug von Freiheitsstrafe (FHS) von 1971 bis 2009 in Hessen, also die Y-Werte geordnet anhand des Zeitindex  $t = (1, 2, \dots, T)$ . Da es sich hier um eine Reihe mit Jahreswerten handelt, können wir den Trend der Reihe als lineare Funktion dieses Zeitindex gemäß Gleichung (1) schätzen und erhalten folgendes Ergebnis (Standardfehler in Klammern, der doppelte Standardfehler markiert das fünfprozentige Signifikanzniveau):

$$\text{SG}_t = +2256 + 50.12 \cdot t + \varepsilon_t \\ (123) (5.35)$$

In unserem Beispiel wird die Konstante mit  $\mu = 2256$  und der Steigungskoeffizient mit  $\beta = 50.12$  geschätzt. Der in Abb. 84.2 (obere Graphik) eingezeichnete Trend hat also eine positive Steigung, was bedeutet, dass sich der Bestand der Strafgefangenen in Hessen im Beobachtungszeitraum jahresdurchschnittlich um etwa 50 Personen erhöht hat. Die Residualkomponente dieses Modells (vgl. Abb. 84.2, untere Graphik) zeigt deutlicher als die Originalreihe ein langfristiges Schwankungsmuster, wobei die Werte in einer Größenordnung von maximal +844 im Jahr 1983 und minimal -540 im Jahr 1971 schwanken, also vom linearen Trend abweichen. Mit diesen Schwankungen und dem Trendverlauf werden wir uns weiter unten noch beschäftigen.



**Abb. 84.2** Strafgefangene in Hessen mit Trend und Trendresiduen

### 84.3 Kausalanalyse mit Zeitreihen

Die Identifikation der Komponenten einer Zeitreihe, also z.B. Trend-, Zyklus- und Saisonkomponente, kann schon für sich genommen von Interesse sein, bspw. um Prognosen für die weitere Entwicklung anzustellen. Die univariate Analyse wird aber auch benötigt, um ein (dynamisches oder statisches) Modell für Zwecke der *Kausalanalyse* angemessen spezifizieren zu können. Das soll im Folgenden anhand eines einfachen Regressionsmodells kurz erläutert werden. Der einfachste Fall einer *nicht-dynamischen bivariaten Beziehung* lässt sich regressionsanalytisch wie folgt darstellen:

$$(2) \quad Y_t = \mu + \beta X_t + \varepsilon_t$$

Man könnte z.B. überprüfen, ob der zahlenmäßige Anstieg der Strafgefangenen linear mit einem Anstieg der erfassten Straftaten verbunden ist. Solche einfachen Erklärungsmodelle sind natürlich für die wissenschaftliche Forschung in der Regel unbefriedigend. Es lassen sich aber einige zentrale methodische Aspekte am ehesten anhand solch einfacher Modelle einführend erläutern.

Bei Modell (2) besteht der einzige Unterschied zur querschnittlichen Analyse darin, dass sich der Fallindex  $t$  nicht auf unterschiedliche Probanden (z.B. eine Stichprobe von einzelnen Personen), sondern auf unterschiedliche Messzeitpunkte bei ein und derselben

Untersuchungseinheit bezieht. Daraus folgt u.a. das Problem, dass die Fehlergrößen – anders als bei der üblichen OLS-Regression vorausgesetzt – in der Regel nicht unabhängig voneinander sind, da die (im Modell nicht erfassten) Einflussgrößen in der Regel auto-korriert sind, ein hoher Wert zum Zeitpunkt  $t$  also (bei positiver Korrelation) auch einen hohen Wert zum Zeitpunkt ( $t+1$ ) – und evtl. auch noch über diesen Zeitpunkt hinaus – nach sich zieht.

Es müssen deshalb alternative Schätzverfahren eingesetzt werden, in deren Rahmen auch die zeitliche Abhängigkeit der Residuen untereinander zu modellieren ist. Im einfachsten Falle könnte die Residualreihe  $\varepsilon(t)$  als lineare Funktion von  $\varepsilon(t-1)$  dargestellt werden. In einem ersten Schritt kann das Modell (2) mit dem normalen OLS-Verfahren der Regressionsanalyse geschätzt und sodann geprüft werden, ob die Residuen unabhängig voneinander sind oder eine bestimmte Abhängigkeitsstruktur aufweisen. Ein gängiges Testverfahren, das in der Standardsoftware statistischer Analyseprogramme implementiert ist, ist der „*Durbin-Watson-Test*“ (*DW-Test*).

## 84.4 Dynamische Modelle

Gleichung (2) ist ein „statisches“ Modell, das voraussetzt, dass ein Input  $X(t)$  zum Zeitpunkt  $t$  seine Wirkung auf die Output-Variable  $Y(t)$  sofort und vollständig entfaltet. Es könnte jedoch sein, dass die *Wirkung verzögert*, also erst zu einem späteren Zeitpunkt (mit einem „lag“) eintritt und/oder sich über mehrere Zeitintervalle entfaltet (*dynamisches Modell mit „distributed lags“*).

Wenn in einer solchen Situation ein statisches Modell gemäß Gleichung (2) geschätzt wird, führt das in der Regel zu *autokorrelierten Residuen* – selbst dann, wenn sie bei korrekter „dynamischer“ Modellspezifikation unabhängig voneinander blieben. (Allerdings ist es durchaus auch möglich, dass die Residuen sich als abhängig voneinander erweisen, obwohl ein dynamisches Modell in korrekter Weise spezifiziert worden ist)

Die Spezifikation dynamischer Modelle ist natürlich nicht nur (und auch nicht vorrangig) ein technisches Hilfsmittel, um die Autokorrelation der Residuen in der Nähe von Null zu halten, sondern sie ist für sich genommen substanzwissenschaftlich von hohem Interesse: Wann tritt eine Wirkung ein, wie rasch und nach welchem Muster entfaltet sie sich? Mit reinen Querschnittsdaten lassen sich solche Fragen nicht beantworten.

Bevor wir kurz erläutern, wie dynamische Modelle spezifiziert werden können, sei noch auf ein Problem hingewiesen, das beachtet werden muss: die Daten müssen sich „*im Äquilibrium*“ befinden, d.h., die Prädiktorvariablen müssen ihre Wirkung während des Beobachtungszeitraumes voll entfaltet haben. Ein dynamisches Modell mit nur einer Input-Variablen, das weiter unten am Beispiel der SG noch erläutert wird, könnte wie folgt aussehen:

$$(3) \quad Y_t = \mu + \beta_0 X_t + \beta_1 X_{t-1} + \dots + \beta_k X_{t-m} + \varepsilon_t$$

Der Idee nach stellen die  $m$  Steigungskoeffizienten dar, in welcher Weise sich die Wirkung über  $m+1$  Zeitintervalle entfaltet, während die Summe dieser Koeffizienten den Gesamteffekt ergibt. Zu fragen ist allerdings: Wie soll die Verzögerungslänge  $m$  festgelegt werden, wie lange dauert es, bis sich eine Änderung in  $X$  voll ausgewirkt hat? Anhaltspunkte hierzu liefert die empirische „Kreuzkorrelationsfunktion“ (KKF). Hierbei werden die beiden Zeitreihen,  $Y(t)$  und  $X(t)$ , sukzessive mehrfach miteinander korreliert. Im ersten Schritt werden die Datenpaare,  $x(t)$  und  $y(t)$ ,  $t = (1, 2, \dots, T)$ , gebildet (also eine Korrelation ohne „lag“), im zweiten Schritt wird eine Verzögerung um ein Zeitintervall eingeführt, kombiniert werden also  $x(t-1)$  und  $y(t)$ , sodann  $x(t-2)$  und  $y(t)$  usw. Es lässt sich dann beobachten, in welchem Muster sich die Kreuzkorrelationskoeffizienten entwickeln; ob sie sich überhaupt über einen längeren Zeitraum ansteigend oder abfallend entwickeln („distributed lags“) und wann sie sich bei Null einpendeln.

Ein weiteres mögliches Problem ist, dass sich die eventuell zahlreichen  $\beta$ -Parameter nicht unabhängig voneinander schätzen lassen (wodurch das Wirkungsmuster evtl. nicht adäquat abgebildet wird) und die Fallzahlen (Messzeitpunkte) für eine effiziente Schätzung zu gering sind. Für die üblichen Wirkungsmuster lässt sich Gleichung (3) allerdings so umformen, dass man auch bei langer (im Prinzip unendlich langer) Wirkungsdauer mit wenigen Parametern auskommt. Ein Beispiel liefert die sog. *Koyck-Transformation*, die zu folgendem „*Autoregressive-Distributed-Lag-Model*“, kurz *ARDL-Modell* führt:

$$(4) \quad Y_t = \mu^* + \delta Y_{t-1} + \beta_0 X_t + \varepsilon_t^*$$

Die um ein Zeitintervall zurück versetzte *abhängige* Variable wird nun als (technische) Prädiktorvariable eingeführt, substanzialer Einflussfaktor ist alleine  $X(t)$ . Dieses Modell formalisiert einen Wirkungsverlauf, der sofort einsetzt (andernfalls müsste als Regressor nicht  $X(t)$ , sondern eine um  $k$  „lags“ verschobene  $X$ -Reihe,  $X(t-k)$ , herangezogen werden). Der von einer Änderung in  $X$  ausgehende Wirkungsimpuls ist anfänglich am stärksten (indiziert durch  $\beta_0$ ); die nachfolgenden zusätzlichen Wirkungsimpulse fallen exponentiell gegen Null ab, umso rascher, je kleiner  $\delta$ . Der Gesamteffekt einer dauerhaften Einheitsänderung in  $X(t)$  ergibt sich aus dem Quotienten  $\beta_0/(1-\delta)$ , der Summenformel für eine geometrische Reihe. Auch ein solches Modell lässt sich mit dem OLS-Algorithmus in der Regel nicht adäquat schätzen. Die Parameterschätzungen werden noch komplexer, wenn eine um zwei Zeitintervalle verzögerte abhängige Variable  $Y_{t-2}$  als zusätzliche Regressor-Variable eingeführt wird. Dadurch lässt sich ein Effektmuster abbilden, demzufolge die sukzessiven Wirkungsimpulse zunächst ansteigen und dann erst gegen Null abfallen. Außerdem können zusätzliche Input-Variablen mit oder ohne zeitlich verteilte Wirkungsmuster in die Modellgleichung eingeführt werden.

Als Beispiel für ein einfaches ARDL-Modell betrachten wir wieder die Reihe der Strafgefangenen in Hessen (vgl. Abb. 84.1). Da sowohl bei den bisher dargestellten statischen wie auch bei den dynamischen Regressionsmodellen vorausgesetzt wird, dass die Zeitreihen „stationär“ sind, dass sich also weder die Mittelwerte noch die Varianzen über die Zeit verändern, verwenden wir für das folgende Beispiel zunächst nicht die Originalreihe,

sondern die oben ermittelten Trendabweichungen, von denen wir annehmen, dass sie stationär sind. Im Folgenden geht es um die Frage, welche Faktoren den Verlauf dieser Reihe beeinflusst haben könnten.

Als „theoretische“ Antwort darauf bietet die Kriminologie ein weit gespanntes Faktorenbündel an, das von demographischen über sozio-ökonomische Faktoren bis hin zu Praktiken der Strafverfolgung reicht. In unserem Beispiel wollen wir uns auf die Strafverfolgung beschränken, da sie den unmittelbaren Input des Strafvollzugs darstellt. Wie eine neuere Untersuchung (Sohn/Metz 2011) gezeigt hat, weist die Reihe der zu langjährigen Freiheitsstrafen Abgeurteilten ein ähnliches Entwicklungsmuster auf, wie die Reihe aller Strafgefangenen, obwohl sie nur eine Teilmenge von etwa zwei Prozent der insgesamt Verurteilten darstellen. Man kann also vermuten, dass die Verfolgungs- und Verurteilungspraxis, die zunächst für gravierende Vergehen eingesetzt wird, auf die Behandlung weniger schwerwiegender Vergehen übertragen wird.

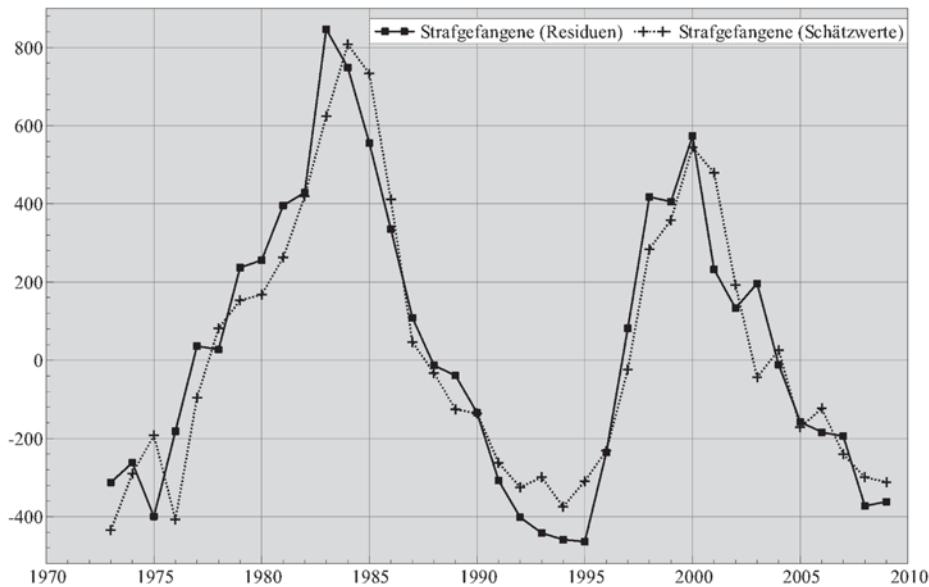
Diese Hypothese soll im Folgenden überprüft werden. Für die Spezifikation des entsprechenden ARDL-Modells schätzen wir daher für die Reihe der langjährig Verurteilten zunächst ebenfalls einen linear-deterministischen Trend, woraus sich wieder die Trendresiduen ergeben (wie oben gezeigt). Für das ARDL-Modell verwenden wir dann die Trendresiduen der Strafgefangenen insgesamt (Res\_SG) als abhängige bzw. zu erklärende Variable und die Trendresiduen der zu langfristigen Freiheitsstrafen Abgeurteilten (Res\_LFHS) als erklärende bzw. unabhängige Variable. Somit schätzen wir folgendes ARDL-Modell (Standardfehler in Klammern):

$$\begin{aligned} \text{Res\_SG}_t &= +0,4552 * \text{Res\_SG}_{t-1} + 4,049 * \text{Res\_LFHS}_t + e_t \\ (0,125) (0,794) \\ R^2 &= 0,71 \end{aligned}$$

Wie man in Abb. 84.3 sieht, liefert das Modell gute Prognosewerte für die Trendresiduen der SG und überdies weisen die Residuen des Modells ( $e_t$ ) keine signifikante serielle Abhängigkeit auf (DW = 1,89).

Nach diesem Modell ist der Wert der Trendresiduen der SG in einem bestimmten Jahr von zwei Größen ablesbar und zwar von dem Wert der Trendresiduen der SG im Vorjahr (gewichtet mit  $\delta = 0,45$ ) sowie vom Wert der Trendresiduen der zu langfristigen FHS Verurteilten im selben Jahr (gewichtet mit  $\beta = 4,0$ ). Eine Erhöhung der Trendresiduen der langfristig Verurteilten führt also noch im selben Jahr zu einer Erhöhung der Trendresiduen der SG um den Faktor 4, im nächsten Jahr zu einer Erhöhung um den Faktor  $4 \times 0,45 = 1,8$ . Sollte die Erhöhung der LFHS-Residuen länger anhalten, müssten die entsprechenden Effekte addiert werden; der Effekt würde sich somit dem Grenzwert von  $4/(1-0,45) = 7,3$  annähern.

Unter kriminologischen Gesichtspunkten ließen sich also nach diesem Modell die langfristigen Trendabweichungen bei der Zahl der Strafgefangenen insgesamt zu einem erheblichen Teil aus der Zahl der zu langfristigen FHS Verurteilten erklären. Interpretiert man die zu langfristigen FHS Verurteilten als Indikator für die „schwere Kriminalität“, könnte man auch sagen, dass die langfristigen Schwankungen im hessischen Strafvollzug



**Abb. 84.3** Trendresiduen der Strafgefangenen und Schätzwerte des ARDL-Modells

primär eine Folge der Sanktionierung der „schweren Kriminalität“ darstellen, die dann auch die Sanktionierungspraxis bei minder schweren Kriminalfällen bestimmen.

Allerdings haben wir mit diesem Modell den Kausalzusammenhang zwischen den beiden Reihen nur für die Trendresiduen, nicht aber für die Reihen selbst festgestellt, was darin begründet ist, dass die oben beschriebenen Kausalmodelle eben nur für stationäre Reihen valide Ergebnisse liefern. Für Reihen, die einem Trend folgen, führen kausaltheoretische Interpretationen mit diesen Modellen dagegen zu erheblichen Problemen. (Das Problem ungleicher Varianzen lassen wir hier einmal beiseite; oft lässt es sich schon mit Hilfe von Datentransformationen wie Logarithmus- oder Wurzelbildung lösen.) Wenn nämlich Zeitreihen einem Trendverlauf folgen, werden sie auch dann miteinander korrelieren, wenn sie nicht kausal aufeinander einwirken – dies nennt man *Scheinkausalität* (Kühnel/Dingelstedt, Kapitel 80 in diesem Band).

Oder man versucht, wie in unseren bisherigen Beispielen, den Trend zu eliminieren, wofür es aber, wie oben erwähnt, ganz unterschiedliche Möglichkeiten gibt, so dass natürlich auch die Ergebnisse von der jeweils gewählten Trendbereinigung abhängen. So kann man neben der hier vorgeführten linearen Trendbereinigung bspw. anstelle der Originaldaten „differenzierte Reihen“ (z.B. „einfach“ differenziert:  $x(t)-x(t-1)$ ) für die Analyse heranziehen (u. U. muss die differenzierte Reihe nochmals oder mehrfach differenziert werden, bevor sie stationär wird). Durch die, wie auch immer, durchgeführte Trendbereinigung handelt man sich aber möglicherweise ein neues Problem ein: das der *scheinbaren Nicht-Kausalität*. Es könnte ja sein, dass systematische Niveauveränderungen in der

X-Reihe entsprechende Anpassungsreaktionen, rasch oder eher verzögert, in der Y-Reihe hervorrufen. Auf jeden Fall scheint es naheliegend und sinnvoll, für die Kausalanalyse nicht die trendbereinigten Reihen, sondern die Originalreihen zu verwenden. Im Folgenden wollen wir uns deshalb der Frage zuwenden, wie man trendbehaftete Zeitreihen auf bestehende Kausalzusammenhänge untersuchen kann.

Ein kausaler Zusammenhang (eine *langfristige Gleichgewichtsbeziehung*) in der Niveauentwicklung der Reihen kann dann überprüft und modelliert werden, wenn der jeweilige Trend nicht „*deterministisch*“ verläuft (eine stabile Funktion der Zeit ist), sondern „*stochastisch*“, wenn es also „zufällige“ Veränderungen in der Trendrichtung gibt. Im einfachsten Fall handelt es sich hierbei um einen *nichtstationären stochastischen Prozess*, der durch die *Kumulation* („*Integration*“) von Zufallseinflüssen bzw. ganz allgemein durch die *Integration* eines *stationären stochastischen Prozesses* entsteht. Solche *nichtstationären stochastischen Prozesse* werden auch als „*integrierte Prozesse erster (u.U. auch höherer) Ordnung*“ oder, in Kurzform, als *I(1)-Prozesse* bezeichnet. Im Gegensatz dazu wird ein *stationärer stochastischer Prozess* als *I(0)-Prozess* bezeichnet.

Die moderne Zeitreihenanalyse hat eine Reihe von Verfahren entwickelt, mit denen man testen kann, ob eine Reihe einem stochastischen oder, alternativ, einem deterministischen Trend folgt. Die in diesem Zusammenhang wichtigsten Verfahren werden als „*Unit-Root-Tests*“ bezeichnet, auf die wir hier jedoch nicht eingehen können (zur Einführung siehe Thome 2005, 155-162).

Im Falle *nichtstationärer stochastischer Prozesse* lässt sich empirisch prüfen, ob die stochastischen Trendverläufe unterschiedlicher Reihen miteinander korrespondieren, ob die Niveauänderungen in der X-Reihe früher oder später gleichgerichtete oder inverse Niveauänderungen in der Y-Reihe hervorrufen. Die entsprechenden Analysestrategien werden in der Fachliteratur unter dem Etikett der „*Kointegrations- und Fehlerkorrekturmodelle*“ behandelt, denen in der empirischen Forschung eine grundlegende Bedeutung zukommt.

Das *Kointegrationskonzept* (Granger 1981, 1986; Engle/Granger 1987), beschreibt ein *Zusammenhangssystem* zwischen zwei oder mehr Variablen, die gemeinsamen stochastischen Trends („*common trends*“) folgen. Jede der beteiligten Zeitreihen ist zwar *nicht-stationär*; aufgrund der Tatsache, dass sie gemeinsamen stochastischen Trends folgen, können sie sich jedoch langfristig nicht beliebig weit voneinander entfernen, sondern werden sich immer wieder aufeinander zu bewegen, d.h. die Abweichungen der Reihen voneinander sind *stationär*, was man auch als *langfristige Gleichgewichtsbeziehung* interpretiert. Diese *Gleichgewichtsbeziehung* bedeutet, dass eine *stationäre Linearkombination* zwischen den Reihen existiert. Das heißt, im Falle gemeinsamer stochastischer Trends korrespondieren die Trendbewegungen der einzelnen Reihen miteinander, so dass z.B. ein Anstieg des lokalen Trends in X mit entsprechenden Anpassungen des lokalen Trends in Y einhergeht.

Ob für konkrete Zeitreihen Kointegration vorliegt, und ob es bei mehr als zwei Zeitreihen gegebenenfalls mehrere Kointegrationsbeziehungen gibt, ist a priori nicht bekannt. Deshalb sind spezielle Tests erforderlich, die eine Entscheidung darüber erlauben, ob überhaupt Kointegration vorliegt und mit wie vielen Kointegrationsbeziehungen zu rechnen

ist. Am einfachsten gestalten sich Kointegrationstests, wenn man annehmen kann, dass nur eine Kointegrationsbeziehung besteht.

In diesem Zusammenhang haben Engle und Granger (1987) ein Verfahren vorgeschlagen, das zwei Stufen umfasst und auf den schon erwähnten „Unit Root-Tests“ beruht. Während in der ersten Stufe geprüft wird, ob die Zeitreihen zum gleichen Grad integriert sind (s. oben), wird, falls dies der Fall ist, im zweiten Schritt geprüft, ob die Residualreihe (nach erfolgter Regression) stationär ist, denn nur wenn die Zeitreihen kointegriert sind, ist dies der Fall. Seien  $X_{1t}$  und  $X_{2t}$   $I(1)$  mit der Kointegrationsbeziehung  $X_{1t} = \gamma X_{2t}$ , dann sind die Abweichungen  $Z_{1t} := X_{1t} - \gamma X_{2t}$  stationär und der Kointegrationsvektor ist  $(1, -\gamma)$  mit unbekanntem Koeffizienten  $\gamma$ . Man schätzt also zunächst den unbekannten Kointegrationsvektor, hier den Parameter  $\gamma$  in der Regression (sog. Kointegrationsgleichung)

$$(5) \quad X_{1t} = \gamma X_{2t} + e_t$$

mit Hilfe von OLS und bildet dann die Residuen  $e_t := X_{1t} - \gamma X_{2t}$ . Wenn  $X_{1t}$  und  $X_{2t}$  nicht kointegriert sind, dann kann  $e_t$  nicht  $I(0)$  sein. Wenn  $X_{1t}$  und  $X_{2t}$  kointegriert sind, dann sollten die Residuen stationär sein.

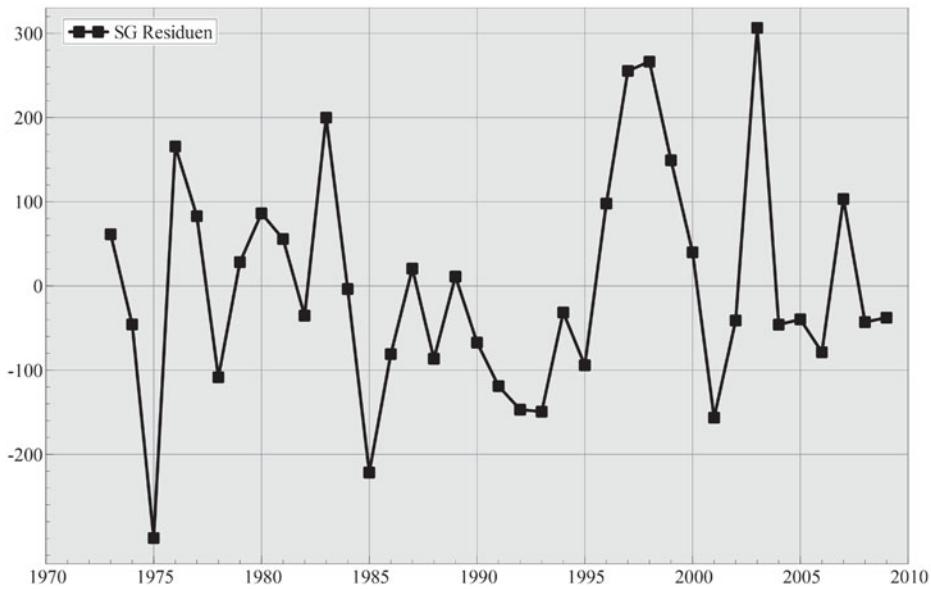
Wenn eine Kointegrationsbeziehung vorliegt, stellt sich die weitere Frage, in welchem zeitlichen Ablauf die eine Reihe,  $Y(t)$ , ihr Niveau an die Niveauentwicklung der anderen Zeitreihe,  $X(t)$ , anpasst. Diese dynamische Modellierung erfolgt in einem zweiten Schritt mit Hilfe der sog. Fehlerkorrekturmodelle („Error Correction Models“, ECM). Diese Modelle ermöglichen es, eine vorhandene langfristige Gleichgewichtsbeziehung zwischen zwei oder mehr Variablen und die kurzfristige Anpassungsdynamik, mit der eine Störung des Gleichgewichts vom System verarbeitet wird, also die sog. Fehler-Korrektur, simultan zu schätzen.

Als Beispiel wollen wir den Zusammenhang zwischen den oben schon vorgestellten Reihen der Strafgefangenen (SG, vgl. Abb. 84.2) und der zu langfristigen FHS Verurteilten (LFHS) auf mögliche Kointegrationsbeziehungen untersuchen. Zunächst ist zu prüfen, ob die beiden Reihen einem stochastischen Trend folgen, der denselben Integrationsgrad aufweist. Für beide Reihen ergeben die „Unit-Root“ Tests jeweils einen  $I(1)$ -Prozess. In einem zweiten Schritt schätzen wir nun die Kointegrationsgleichung (5) mit OLS und erhalten folgendes Ergebnis:

$$\begin{aligned} SG_t &= + 1817 + 9,513 * LFHS_t + e_t \\ &\quad (103) \quad (0,615) \end{aligned}$$

In langfristiger Perspektive betrachtet, führt eine Veränderung im LFHS-Niveau zu einer 9,5-mal so starken (gleichgerichteten) Veränderung im SG-Niveau. Die Residuen des Modells  $e_t$ , im Folgenden als *Residuen* bezeichnet, sind offensichtlich stationär (Abb. 84.4), was durch einen „Unit-Root-Test“ bestätigt wird, signifikante Veränderungen ergeben sich weder im Mittelwert noch hinsichtlich der Varianz der Residualreihe.

Nachdem wir festgestellt haben, dass die beiden Reihen einem gemeinsamen stochastischen Trend folgen, können wir mit Hilfe des *Fehlerkorrekturmodells* die kurz- und



**Abb. 84.4** Residuen der Kointegrationsgleichung

langfristige Dynamik zwischen den beiden Reihen schätzen, wofür wir die ersten Differenzen (also die jährlichen Veränderungen) der SG-Reihe ( $\Delta SG$ ) auf die ersten Differenzen der LFHS-Reihe ( $\Delta LFHS$ ) regredieren. Die folgende Gleichung repräsentiert die dabei geschätzten Parameter

$$\Delta SG_t = +0,2798 * \Delta SG_{t-1} + 5,259 * \Delta LFHS_t - 0,2292 RESIDUEN_{t-1} + e_t$$

(0,0881)            (0,632)            (0,0822)

$$DW = 1,83$$

Es zeigt sich, dass kurzfristig ein Anstieg der langfristig Verurteilten einen 5,26-mal höheren Anstieg der Zahl der Strafgefangenen insgesamt nach sich zieht, während der langfristige Anstieg, wie wir schon sahen, mit dem Faktor 9,5 anzusetzen ist. Die Anpassung an diesen langfristigen Effekt (der sog. Reäquilibrierungsprozess) vollzieht sich mit einer jährlichen Rate von 23 % der jeweils noch verbliebenen Rest-Differenz (s. den für die RESIDUEN geschätzten Koeffizienten von -0,229). Für weitere Details und Beispiele verweisen wir auf Thome (2005, Kap. 5.6.2).

## 84.5 Schlussbemerkung

Die soziologische wie auch die politikwissenschaftliche Forschung in Deutschland bedient sich nur selten formaler Verfahren der Zeitreihenanalyse und der damit verbundenen Modellierungsstrategien. Zeitreihen dienen nicht nur dazu, langfristige gesellschaftliche Entwicklungsprozesse anschaulich darzustellen. Jede Zeitreihe repräsentiert ein Zusammenspiel verschiedener Komponenten, die jeweils unterschiedlichen Entwicklungsdynamiken folgen und deshalb analytisch zu differenzieren sind. Diese Entwicklungsdynamiken können über längere Zeit relativ stabil verlaufen, sich aber auch allmählich oder sehr kurzfristig verändern. Dabei müssen u.a. abrupte oder allmäßliche Niveauverschiebungen, die von externen Ereignissen ausgelöst werden, von kontinuierlichen Trendentwicklungen unterschieden werden. Ohne explizite Modellierungen sind die unterschiedlichen Einflussgrößen und wechselseitigen Zusammenhänge nicht erkennbar und kausale Zuschreibungen nicht überprüfbar. Die oben skizzierten kointegrierten Prozesse liefern hierzu ein besonders relevantes Beispiel, da viele der soziologisch interessanten Zeitreihen Trendverläufe aufweisen, die miteinander zu korrespondieren scheinen, ohne dass entsprechende kausale Zusammenhänge vorliegen. Zeitreihenanalytische Modellierungsstrategien bieten gerade einer gesellschaftstheoretisch und historisch ausgerichteten Soziologie spezifische Erkenntnismöglichkeiten, die ausgiebiger als bisher genutzt werden sollten.

## Literatur

- Baxter, Marianne./King, Robert G. (1999): Measuring Business Cycles. In: The Review of Economics and Statistics 81: 575-593
- Box, George E. P./Jenkins, Gwilym M. (1970): Time Series Analysis. San Francisco: Holden Day
- Christiano, Lawrence J./Fitzgerald, Terry J. (2003): The Band Pass Filter. In: International Economic Review 44: 435-465
- Engle, Robert F./Granger, Clive W.J. (1987): Co-Integration and Error Correction: Representation, Estimation, and Testing. In: Econometrica 55: 251-276
- Granger, Clive W. J. (1981): Some Properties of Time Series Data and their Use in Econometric Model Specification. In: Journal of Econometrics 16: 121-130
- Granger, Clive W. J. (1986): Developments in the Study of Co-integrated Economic Variables. In: Oxford Bulletin of Economics and Statistics 48: 213-228
- Hodrick, Robert J./Prescott, Edward C. (1997): Post-War U.S. Business Cycles. In: Journal of Money, Credit, and Banking 29: 1-16
- Sohn, Werner/Metz, Rainer (2011): Gefangenenzahlen in Hessen. In: Bewährungshilfe. Soziales – Strafrecht – Kriminalpolitik. 58 (1). 24-53
- Thome, Helmut (2005): Zeitreihenanalyse. Eine Einführung für Sozialwissenschaftler und Historiker. München: Oldenbourg

Tatjana Mika und Michael Stegmann

Längsschnittdaten liegen vor, wenn zu einer Befragungseinheit von Personen oder Haushalten, gelegentlich auch von Betrieben, Daten zur Verfügung stehen, die nicht nur die Situation zu einem einzigen Messzeitpunkt, sondern über einen längeren Zeitraum erfassen. Dies können Daten sein, die aus Befragungen, insbesondere aus Panel-Erhebungen entstehen, wie das Sozioökonomische Panel (Schupp, Kapitel 73 in diesem Band). Es kann sich aber auch um Daten handeln, die aus Verwaltungsprozessen abgeleitet werden (Hartmann/Lengerer, Kapitel 71 in diesem Band), wie zum Beispiel aus den Daten der Sozialversicherung. Die Länge dieses Zeitraums sowie die Engmaschigkeit der Beobachtungen innerhalb des Gesamtzeitraums hängen von der Forschungsfrage ab, es kann sich um einzelne Tage, um einzelne Wochen oder auch um Monate handeln. Längsschnittdaten ermöglichen die tatsächliche Beobachtung von sozial bedeutsamen Veränderungen wie etwa dem Absinken der Arbeitslosigkeit durch erfolgreiche Arbeitsuche vormals Arbeitsloser, während mehrere unabhängige Messungen nur Veränderungen im Aggregat messen können, z.B. das Absinken und Steigen der Arbeitslosenquote im Zeitverlauf.

---

## 85.1 Befragungsdatenbasierte und prozessproduzierte Längsschnittdaten

Zu den wesentlichen sozialwissenschaftlichen Datenquellen für Längsschnittanalysen gehören Paneldaten (Schupp, Kapitel 73 in diesem Band), Verwaltungsdaten und Daten der amtlichen Statistik (Hartmann/Lengerer, Kapitel 71 in diesem Band) sowie Logfiles (Schmitz/Yanenko, Kapitel 65 in diesem Band). Die Wahl des Gesamtzeitraums und der Erhebungsdichte hängt dabei maßgeblich von der Fragestellung und den Möglichkeiten der Datenerhebung ab.

Ereignisanalysen, die nachfolgend zuerst dargestellt werden, benötigen weniger Messzeitpunkte als Sequenzdatenanalysen, die im Anschluss erläutert werden. *Ereignisanalysen*

benötigen grundsätzlich nur zwei Zeitinformationen, den Anfang und insbesondere das Ende einer Episode, z.B. Heirat oder die Beendigung von Arbeitslosigkeit. Gemessen wird die Veränderung eines Zustandes, etwa von „ledig“ zu „verheiratet“, wobei besonders der Zeitpunkt der Zustandsänderung (hier: Hochzeit) interessiert. So wird zum Beispiel das Alter bei der ersten Heirat gemessen und untersucht. Ergebnis kann sein, dass in der Regel nach Abschluss der Ausbildung geheiratet wird, weshalb höher Gebildete erst in höherem Alter heiraten. Ereignisse können in retrospektiv erhobenen Daten Erinnerungsfehlern unterliegen, wenn das Ereignis als unbedeutend oder unerfreulich ganz vergessen wurde oder nicht mehr genau datiert werden kann, wenn zum Beispiel die erste Ehe geschieden wurde. Daher sind im Verlauf der Datenerhebung gemessene Ereignisse im Hinblick auf die Datenqualität zuverlässiger als retrospektiv erhobene Daten. Wenn die Erhebung sehr kurz nach dem Ereignis stattfindet, dann ist die Qualität der Erinnerung meistens sehr gut. Je länger das Ereignis zurückliegt, umso weniger genau werden Zeitpunkte erinnert. Dokumente wie Scheidungsurkunden oder Arbeitsbescheinigungen, die Grundlage von prozessproduzierten Daten sind, geben die Ereignisse meist auf den Tag genau an und sind der Erinnerung an weiter zurück liegende Ereignisse daher überlegen.

Für *Sequenzdatenanalysen* werden durchgängig gemessene Zustände benötigt, z.B. alle Monate eines Jahres oder eines anderen Beobachtungsfensters. Beispielsweise kann für jede Person erfasst werden, welchen Zustand sie im Januar, Februar etc. eines Jahres einnimmt. Daher sind regelmäßige Messzeitpunkte von zentraler Bedeutung. Hierfür ist es auch sehr wichtig, dass die Konzepte der Zustände, die gemessen werden, im Verlauf der Erhebung nicht verändert wurden.

---

## 85.2 Auswertungsstrategien und Datenformate

Es gibt zwei wesentliche Methoden der Auswertung sozialwissenschaftlicher Längsschnittdaten, die je ein spezifisches Datenformat bedingen:

1. *Verweildaueranalyse*: Methode, mit der überwiegend die Dauer bestimmter sozialer Zustände oder Übergänge und deren Einbettung im Lebenslauf deskriptiv oder statistisch-modellierend untersucht werden. Der Analysefokus liegt auf der Verweildauer in einem Zustand, die z.B. mit einem Ereignis abgeschlossen wird. Es gilt zu untersuchen, ob und wann ein Zustandswechsel eintritt. Übergänge werden durch *Ereignisdaten* gemessen.
2. *Sequenzdatenanalyse*: deskriptive sowie heuristisch modellierende Verfahren, welche die Biografien als vollständige Stränge aufeinanderfolgender Lebensphasen untersuchen (*Sequenzdaten*).

## 85.3 Ereignisdaten und Verweildaueranalyse

### 85.3.1 Grundmodell

Das grundlegende Modell der Ereignisdatenanalyse ist es, die Verweildauer bis zum Eintreten oder ab dem Eintritt eines bestimmten Zustandes zu untersuchen. Mit dem Wechsel von einem diskreten Zustand in einen anderen (z.B. von erwerbstätig zu arbeitslos) ist ein Ereignis definiert. Man spricht dabei von einem *absorbierenden Zielzustand*, wenn ein erneuter Zustandswechsel unwahrscheinlich bzw. unmöglich ist (z.B. Tod, Geburt des ersten Kindes). Von *rekurrerenden Ereignissen* spricht man, wenn diese mehrmals im Lebenslauf auftreten können (z.B. Arbeitslosigkeit). Die Zeitspanne, die ein Zustand einnimmt, bevor es zum Zustandswechsel kommt, bezeichnet man als *Episode*. Alle Personen, bei denen ein Ereignis noch eintreten kann (alle Ledigen könnten noch heiraten, alle Arbeitslosen eine Beschäftigung finden), bezeichnet man als *risk-set*, und sie werden weiter daraufhin beobachtet, ob und wann ein Zustandswechsel (Hochzeit) eintritt. Tritt das interessierende Ereignis ein, dann scheiden die Personen aus dem *risk-set* aus. Tabelle 85.1 zeigt beispielhaft den Aufbau von Ereignisdaten.

Fall 1 hat einen Übergang von Arbeitslosigkeit in Beschäftigung, der am 1.1.2001 stattfand, Fall 3 war im Januar 2004 arbeitslos, davor war sie in Ausbildung, danach beschäftigt. Wird die Forschungsfrage untersucht, wie lange Personen im Durchschnitt arbeitslos sind, bevor sie eine (erneute) Beschäftigung finden, dann lautet in Fall 1 die Antwort „vom 1.2.2000 bis zum 31.12.2000, also 11 Monate“ und in Fall 3 „1 Monat“. Ein Ergebnis einer Untersuchung könnte sein, dass Arbeitslosigkeit nach einer Ausbildung kürzer andauert als nach einer Phase der Beschäftigung. Fall 2 hat keine Zeiten der Arbeitslosigkeit und wäre daher gar nicht in dem *risk-set* enthalten. Lautet die Forschungsfrage dagegen, wie

**Tab. 85.1** Beispiel für Ereignisdaten (Quelle: eigene Darstellung)

Fall	Beginn	Ende	Status	Geschlecht
1	01.01.2000	31.01.2000	Beschäftigung	Männlich
1	01.02.2000	31.12.2000	Arbeitslosigkeit	Männlich
1	01.01.2001	31.12.2010	Beschäftigung	Männlich
2	01.01.2000	31.12.2004	Schule und Studium	Weiblich
2	01.01.2005	31.12.2010	Beschäftigung	Weiblich
3	01.01.2000	31.12.2003	Berufliche Ausbildung	Männlich
3	01.01.2004	31.01.2004	Arbeitslosigkeit	Männlich
3	01.02.2004	31.12.2010	Beschäftigung	Männlich

lange Personen nach der schulischen oder beruflichen Ausbildung arbeitslos sind, bevor sie eine Beschäftigung finden, dann sind nur die Fälle 2 und 3 im *risk-set*, weil wir nur für diese beiden Fälle die Zeit des „Studiums“ oder der „beruflichen Ausbildung“ in den Daten erfasst haben. Fall 2 hat keine Arbeitslosigkeit erlebt, Fall 3 einen Monat, nämlich den Januar 2004. Werden nur diese Fälle betrachtet, dann könnte das Ergebnis der Forschungsfrage sein, dass Personen nach dem Studium schneller eine Beschäftigung finden als nach einer beruflichen Ausbildung.

### **85.3.2 Zensierung**

Von besonderer Bedeutung im Rahmen der Ereignisdatenanalyse sind die so genannten Zensierungen. Man unterscheidet dabei Links- und Rechtszensierungen. Bei einer *Linkszensierung* ist der Beginn des Zustandes unbekannt, er liegt vor der beobachteten Periode. Bei einer *Rechtszensierung* dauert der Zustand am Ende der Beobachtungszeit noch an. Wird zum Beispiel der Eintritt der ersten Arbeitslosigkeit nach Aufnahme der Erwerbstätigkeit untersucht, dann liegt für all die Beschäftigten eine Rechtszensierung vor, die bis zum letzten Beobachtungszeitpunkt noch nicht arbeitslos geworden sind. Für alle, die zu Beginn der Beobachtung bereits arbeitslos waren, liegt dagegen eine Linkszensierung vor. Grundsätzlich wird beobachtet, wie lange (T) eine Untersuchungsperson in einem Zustand verweilt, bis die Zustandraumvariable (Y) wechselt oder eine Rechtszensierung eintritt. Eine mögliche Frage ist dabei z.B. „Wie alt sind die untersuchten Personen, wenn sie wegen Erwerbsminderung in Rente gehen müssen?“. Bei solchen Forschungsfragen, die seltene und spät im Leben auftretende Ereignisse untersuchen, ist häufig eine Linkszensierung gegeben, weil die Personen erst in bereits fortgeschrittenem Alter von der Erhebung erfasst wurden, der Beginn der Berufstätigkeit liegt außerhalb des Beobachtungszeitraums.

### **85.3.3 Überlebensfunktion, Sterbetafeln und Hazardrate**

Die klassischen beschreibenden Verfahren zur Untersuchung von Ereignisdaten sind Methoden, die sich die *Überlebensfunktion* zu Nutze machen. Diese Funktion gibt die Wahrscheinlichkeit an, dass vor einem Zeitpunkt t kein Ereignis eintritt oder anders ausgedrückt, dass die Person noch bis zu einem Zeitpunkt t dem Risiko des Zustandswechsels ausgesetzt ist. Es wird dann beobachtet, ob das Ereignis in der Zeit der Beobachtung eingetreten ist und gemessen, wann dies passierte. Traditionell heißen Tabellen, in denen die Zeittauern bis zum Eintritt des untersuchten Ereignisses aufgelistet werden, *Sterbetafeln*.

Eine Sterbetafel ordnet die untersuchten Personen anhand der Variable „Alter“, im nachfolgenden Beispiel in Kategorien von jeweils fünf Jahren, und gibt an, wie viele Untersuchungspersonen zu jedem Zeitpunkt, also dem beobachteten Lebensalter, aus der Untersuchungsgruppe ausgeschieden sind. Auf der Grundlage der Sterbetafel kann bestimmt werden, wie hoch zu einem gegebenen Alter die Wahrscheinlichkeit ist, aus der

Untersuchungsgruppe aufgrund des vorher bestimmten Ereignisses auszuscheiden. Diese „Überlebensfunktion“ lässt sich für diverse Gruppen (z.B. Alterskohorten, Geschlechter, regionale Herkunft, berufliche Vorgeschichte etc.) beschreiben und vergleichend analysieren.

An einem Beispiel wird dies deutlicher: Die Wahrscheinlichkeit, wegen Erwerbsminderung verrentet zu werden, nimmt mit dem Alter deutlich zu, da mit steigendem Alter die Wahrscheinlichkeit steigt, dauerhaft zu erkranken. Dieses Wachstum ist zunächst langsam und stetig, nimmt mit steigendem Alter aber deutlich stärker zu. Die Wahrscheinlichkeit des Ereignisses hängt also mit dem Lebensalter als Verweildauer zusammen.

Tabelle 85.2 zeigt in der zweiten Spalte die Anzahl der Personen, die in jedem Zeitintervall dem Risiko (hier: Erwerbsminderung) ausgesetzt waren. Die Anzahl sinkt - wie Spalte 2 zeigt - beständig von 104.105 auf 9.684 Personen. In der dritten Spalte ist die Anzahl der Personen verzeichnet, die in jedem Zeitintervall aus dem Risiko ausgesetzten Gruppe ausgeschieden sind, weil sich das Risiko verwirklicht hat. In der Altersgruppe 35 bis 40 Jahre waren dies nur 194 Personen, im letzten Zeitintervall von 55 bis 60 Jahren waren es 1.104 Personen. Die vierte Spalte zeigt das Verhältnis der Personen, bei denen das Risiko eintrat („Terminierende“), zu allen, die dem Risiko ausgesetzt waren. Daraus ergibt sich auch die Anzahl derer, bei denen das Risiko in der Altersklasse nicht eingetreten ist („Anteil der Überlebenden“). Aus diesen beiden Spalten lässt sich abschätzen, ob das Risiko über die Zeit steigt, fällt oder auch möglicherweise erst steigt und dann wieder fällt. Hier steigt das Risiko mit der Zeit. Oft ist auch von Interesse, wie viele Personen insgesamt über alle bisher beobachteten Altersgruppen hinweg das Risiko überlebt haben. Dies ergibt sich aus den kumulierten Anteilen der Überlebenden. Diese kumulierte Rate der Überlebenden ist niedriger, weil sie diejenige als Überlebende zählt, bei denen das Risiko im Verlauf der gesamten Beobachtung nicht eingetreten ist, die also am Ende noch beobachtet werden können. Hierfür wird bei jeder beobachteten Altersgruppe die neu errechnete Überlebensquote in diesem Zeitabschnitt mit der multipliziert, die bis zu diesem Zeitpunkt erreicht war. 99,7% aller von Anfang an beobachteten Personen überlebten

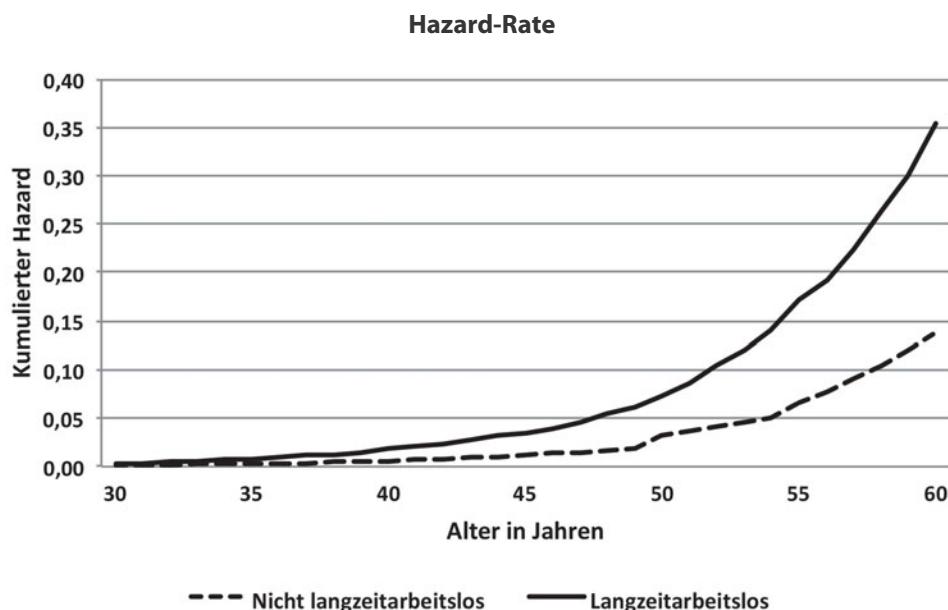
**Tab. 85.2** Sterbetafel – Verrentung wegen Erwerbsminderung nach Alter (Quelle: FDZ-RV – Versicherungskontenstichprobe 2011, nur Männer)

Anfangszeit des Intervalls	Anzahl der Ausgesetzten dem Risiko	Anzahl terminaler Ereignisse	Anteil der Terminierenden	Anteil der Überlebenden	Kumulierter Anteil der Überlebenden am Intervallende
35	104.105	194	0,2%	99,8%	99,7%
40	73.871	361	0,5%	99,5%	99,2%
45	47.288	522	1,1%	98,9%	98,1%
50	25.744	970	3,8%	96,2%	94,4%
55	9.684	1.104	11,4%	88,6%	83,7%

bis zur Altersklasse 35 bis 39. Sie hatten also weder vor dem Alter 35, noch zwischen 35 und 39 erstmals eine Rente wegen Erwerbsminderung bezogen. In der Altersgruppe von 55 bis 60 ist das kumulierte Risiko dann bei 83,7% noch nicht eingetreten. Zum Ende der Untersuchung weicht die kumulierte Überlebenswahrscheinlichkeit deutlich vom Anteil der in dieser Altersgruppe beobachtbaren Überlebenden ab. Dies bedeutet, dass das Risiko bei vielen Personen schon bis zum Alter 55 eingetreten ist.

Die *Hazardrate* bezeichnet die Wahrscheinlichkeit, dass ein Ereignis im beobachteten Intervall stattfindet. Sie kann ohne statistische Modellannahmen mit so genannten nicht parametrischen Verfahren explorativ untersucht bzw. beschrieben werden. Üblich ist auch die grafische Darstellung. Sie erleichtert die Einschätzung des Verlaufs des Risikos über die Zeit. Für unterschiedliche Gruppen kann sie getrennt geschätzt werden.

Die kumulierte Hazardrate in Abb. 85.1 zeigt als Linie für jedes Alter auf der X-Achse den Anteil der Langzeitarbeitslosen und sonstigen Beschäftigten, bei denen das beobachtete Ereignis, hier die Verrentung, stattgefunden hat. Jede Linie folgt dem Verlauf des geschätzten Risikos für die beiden definierten Gruppen. Die Grafik zeigt, dass die Wahrscheinlichkeit, wegen Erwerbsminderung vorzeitig in Rente zu gehen, für Langzeitarbeitslose, definiert als Personen, die 12 Monate oder länger arbeitslos waren, deutlich höher ist. Die Wahrscheinlichkeit beginnt in dieser Gruppe in jüngerem Alter zu steigen. Im Alter ab 50 Jahren verstärkt sich zudem der zu Grunde liegende Prozess wie die größer werdenden Stufen in jedem weiteren Lebensjahr. Die Steigerung ist also nicht stetig. Am Ende des hier dargestellten Beobachtungszeitraums beim Alter von 60 Jahren ist das Ereignis für 35% der



**Abb. 85.1** Hazard-Rate der Verrentung wegen Erwerbsminderung nach Langzeitarbeitslosigkeit  
(Quelle Versicherungskontenstichprobe 2011, nur Männer, eigene Auswertungen)

Langzeitarbeitslosen, aber nur für 13% der sonstigen Versicherten der Rentenversicherung eingetreten. Die Kurven zeigen das Risiko im jeweiligen Alter an, in Frührente zu gehen. Der Verlauf der Kurve ermöglicht eine Abschätzung, ob das Risiko stetig ist oder sich mit dem Alter steigert oder vermindert, hier im Vergleich für zwei Untersuchungsgruppen. Die Untersuchung ließe sich auch mit mehr als zwei Gruppen, etwa mit gruppierten Berufen, durchführen (Mika, 2013).

## 85.4 Sequenzdaten und Sequenzdatenanalyse

### 85.4.1 Datenstruktur

Sequenzdaten sind so aufgebaut, dass der betrachtete Zeitraum, z.B. eine Biografie, durch aufeinanderfolgende Zustände beschrieben wird. Die zeitliche Dichte, für die die Zustände abgebildet werden, ist dabei grundsätzlich variabel. So kann eine monatliche Betrachtung stattfinden, die sich über einen gewissen kalendarischen oder biografischen zeitlichen Abschnitt erstreckt. Nimmt man beispielsweise an, dass für einen Zustandsraum, der zehn mögliche Zustände enthält, ein Jahr lang monatlich dokumentiert wird, welchen Zustand eine Person in jedem Monat einnimmt, ergibt sich folgende Datensatzstruktur.

Aus Tab. 85.3 wird ersichtlich dass die erste Person im beobachteten Jahr in den ersten beiden Monaten in Ausbildung ist, daran schließen sich acht Monate mit einer beruflichen Tätigkeit an. Die letzten beiden Monate ist die Person arbeitslos.

Für die Sequenzmusteranalyse eignen sich vor allem Daten, die für einen langen Zeitraum eine durchgängige Beobachtung zulassen. Ein Beispiel dafür sind unter anderem die Daten der Rentenversicherung, die über das Forschungsdatenzentrum der Rentenversicherung der wissenschaftlichen Öffentlichkeit angeboten werden.

### 85.4.2 Sequenzmusteranalyse und Optimal Matching

Die *Sequenzmusteranalyse* untersucht mit der anschließenden *Optimal Matching Analyse* Ähnlichkeiten und Muster bei Lebensverläufen. Dabei wird nicht die Dauer bestimmter Zustände als Bilanz einer Biografie fokussiert, sondern die Struktur der Biografie steht im

**Tab. 85.3** Schematisches Beispiel für Sequenzdaten

Fall	Monat 1	Monat 2	Monat 3	Monat 4	Monat 5	Monat 6	Monat 7	Monat 8	Monat 9	Monat 10	Monat 11	Monat 12
1	A	A	B	B	B	B	B	B	B	B	C	C
2	A	B	B	B	C	C	C	B	B	B	B	B
3	A	A	A	A	A	A	C	B	B	B	B	B

A= Ausbildung, B= berufstätig, C= arbeitslos.

Mittelpunkt. Untersucht wird die Ähnlichkeit von Lebensverläufen als Abfolge von biografischen Zuständen. Sie empfiehlt sich im Gegensatz zur Ereignisanalyse, wenn sowohl mehrfache Wechsel als auch unterschiedliche Übergänge untersucht werden sollen.

Die methodische Kernfrage der Sequenzmusteranalyse lässt sich wie folgt beschreiben: Wie groß ist der Modifikations- bzw. Manipulationsaufwand, um die Sequenzen von zwei Personen (z.B. Biografieabschnitte) in Übereinstimmung zu bringen? Dafür werden drei Arten von *Transformationen in den Sequenzen* durchgeführt: Einfügen, Löschen und Austauschen. Man bedient sich dieser Operationen, um die Sequenzen zweier Personen in Übereinstimmung zu bringen. Je geringer der zu betreibende Aufwand ist, um diese hypothetische Gleichheit zu erreichen, umso ähnlicher sind sich die Fälle. Diese Operationen werden in rechnerische Werte übersetzt, die einen numerischen Distanzwert für die Unähnlichkeit von Verläufen ausweisen.

Die Optimal Matching Technik geht davon aus, dass die drei verschiedenen Arten der Transformation mit unterschiedlichen *Transformationskosten* verbunden sein können. Grundlegend wird dabei nach den Kosten, die durch ein Austauschen entstehen, den *Substitutionskosten* (oder *Subkosten*) unterschieden. Durch die Substitution wird erreicht, dass zwei untersuchte Verläufe zu dem Zeitpunkt der Einsetzung gleich werden (Stegmann/Werner/Müller 2013). So wird z.B. der Zustand „arbeitslos“ durch den Zustand „erwerbstätig“ ersetzt. Neben der Substitution bestehen in den Vorgängen *Löschen* und *Ersetzen* weitere Möglichkeiten der Transformation. Es handelt sich dann um eine zeitliche Verschiebung der auf die Transformation folgenden Zustände. Für die Berechnung der Distanzen ist die häufigste Vorgehensweise, dem Vorgang des Löschens und Ersetzens niedrigere Kosten zuzuweisen als dem Austauschen (Substitution), wodurch das Programm „das Löschen“ und an anderer Stelle „das Einfügen“ jenem „der Ersetzung“ vorzieht. Bei gegeneinander verschobenen Verläufen ist die Aktion des Verschiebens des Zielverlaufs somit billiger als das Austauschen von Ereignissen (Erzberger/Prein 1997:66).

Die Methode reagiert sehr stark auf den dauerhaften oder häufigen Wechsel von einem gemessenen Zustand in einen anderen. Dieser Effekt kann dadurch verringert werden, dass die Kosten der Transformation zwischen zwei inhaltlich naheliegenden Zuständen herabgesetzt werden. Ein Beispiel könnten verschiedene Formen der Arbeitslosigkeit wie mit oder ohne Bezug von Arbeitslosengeld I sein, die inhaltlich als ähnlicher zueinander definiert werden können, als beide gegenüber dem Zustand der Erwerbstätigkeit. Das Programm würde, bei entsprechender Festlegung der herabgesetzten Distanz, dann weniger Kosten für den Fall kalkulieren, dass die Verläufe von zwei Arbeitslosen in verschiedenen Zuständen (ALG I und ALG II) dadurch angeglichen würden, dass es in einem von beiden ALG I durch ALG II ersetzt hätte. Eine Alternative bei einer großen Ähnlichkeit von sozialen Zuständen ist es auch, diese vor Beginn der Untersuchung zu einem Zustand zusammenzufassen. Umgekehrt können auch zwei Zustände, welche die untersuchten Personen zur gleichen Zeit innehaben (z.B. arbeitslos und ledig), kombiniert werden, wodurch mehrere Dimensionen im Verlauf gleichzeitig untersucht werden können.

Nur in wenigen Fällen kann man davon ausgehen, dass alle Sequenzen, die für die Personenpaare vorliegen, die gleiche Länge haben, meistens sind die Daten in unterschiedlicher Weise links und rechts zensiert. Um eine Interpretation der Transformationskosten zu ermöglichen, muss dies berücksichtigt werden. Einerseits müssen Sequenzpaare von unterschiedlicher Länge miteinander vergleichbar sein, andererseits muss auch die unterschiedliche Länge zweier Sequenzen innerhalb eines Sequenzpaars berücksichtigt werden. Die Transformationskosten können nicht unabhängig von der Länge der Sequenzen betrachtet werden, sie werden in einer Distanzmatrix gespeichert, die durch eine Clusteranalyse zu aussagefähigen Gruppen ähnlicher Biografien zusammengefasst werden.

### 85.4.3 Clusteranalyse der Distanzmatrix

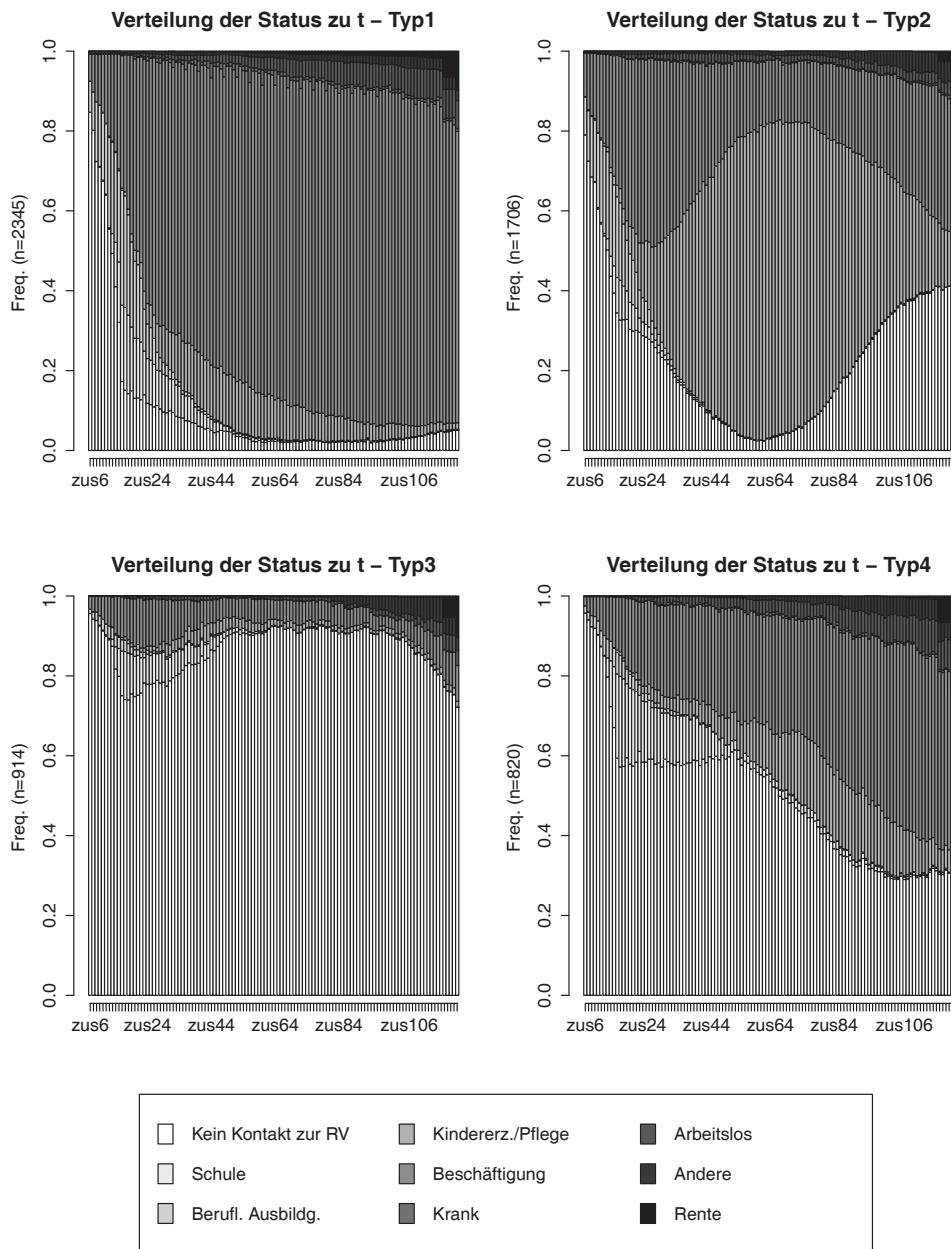
Mit Hilfe der Clusteranalyse kann eine Vielzahl von Variablen auf wenige Cluster reduziert werden (Blasius/Baur, Kapitel 79 in diesem Band). Als Eingabeinformation einer Clusteranalyse kann auch die Distanzmatrix einer Sequenzmusteranalyse verwendet werden.

Dies wird nun mit Hilfe der Versicherungskontenstichprobe, die bereits für das Beispiel oben verwendet wurde, erläutert. Dabei wurde auf eine Substichprobe zurückgegriffen, um die Biografien deutscher Frauen für deren Alterszeitraum 16 bis 43 Jahre zu untersuchen (dazu Stegmann 2008). In Abb. 85.2 werden die Profile der unterschiedlichen Typen von weiblichen Versicherungsbiografien deutlich, die durch eine Clusteranalyse bestimmt wurden.

Bei den vier Typen zeigt sich ein Biografietyper, den man als Drei-Phasen-Modell-Frauen bezeichnen kann (Typ 3). Das sind Frauen, die nach einer beruflichen Tätigkeit eine Pause einlegen und im Anschluss an die Kindererziehung wieder berufstätig werden.

Andere finden nach dem Eintritt in die Erziehungsphase kaum wieder zurück in den Beruf (Typ 1), und ein dritter Biografietyper zeichnet sich durch eine vorrangig durchgängige Erwerbsbiografie aus (Typ 2). Ferner ergibt sich ein Typus von Frauen, die zu einem großen Teil und über weite Strecken des Lebenslaufs keinen Kontakt zur Rentenversicherung haben. Im Rahmen der weiteren Analyse gilt es im Folgenden die Unterschiede zwischen den Typen vergleichend zu untersuchen.

Die Methode hat damit geholfen zu zeigen, dass sowohl die Reihenfolge der Ereignisse im Lebenslauf der Frauen, als auch die Dauer sich sehr deutlich unterscheiden. Beide Aspekte, sowohl das Vorhandensein unterschiedlicher Zustände im gleichen Lebensalter, als auch die unterschiedliche Dauer der Abfolge ähnlicher Zustände, wurden gleichzeitig zur Charakterisierung der Cluster untersucht.



**Abb. 85.2** Sequenzmusteranalyse – Clusterprofile (Frauen, geboren 1936 bis 1964)  
 (Quelle: Versicherungskontenstichproben 2002, 2005, 2007)

## Literatur

- Erzberger, Christian/Prein, Gerald (1997): Optimal Matching Technik Ein Analyseverfahren zur Vergleichbarkeit und Ordnung individuell differenter Lebensverläufe. In ZUMA-Nachrichten 40: 52-80
- Mika, Tatjana (2013): Risiken für eine Erwerbsminderung bei unterschiedlichen Berufsgruppen. In: Bundesgesundheitsblatt 56: 391-397
- Stegmann, Michael (2008): Typisierung von Versicherungsbiografien: Eine empirische Analyse auf Basis der Versicherungskontenstichprobe. In Deutsche Rentenversicherung 63: 221-239
- Stegmann, Michael/Werner, Julia/Müller, Heiko (2013): Sequenzmusteranalyse. Einführung in Theorie und Praxis. Rainer Hampp Verlag, München und Mering

Ulrich Pötter und Gerald Prein

Verlaufsdaten beziehen sich auf Angaben über Ereignisse und Folgen von Ereignissen. Zu denken ist etwa an die Arbeitsmarktbeteiligung von Befragten, wobei Ereignisse Arbeitsplatzwechsel, Übergänge in die Arbeitslosigkeit oder aus der Arbeitslosigkeit, Übergänge in die Rente etc. sein können. Weitere Beispiele sind Heiraten, Schulwechsel und Schulabschlüsse, Beginn des Bezugs von Hartz-IV-Leistungen, berufliche Auf- oder Abstiege, Umzüge, Arztbesuche, Unfälle, Einkäufe bestimmter Waren, Insolvenzen oder Patentanmeldungen von Firmen etc. Die Verlaufsdatenanalyse beantwortet Fragen nach der Dauer bis zu bestimmten Ereignissen, der Häufigkeit von Folgen von Ereignissen, deren Abhängigkeit von Kovariablen sowie nach Abhängigkeiten zwischen Verläufen.

In den letzten 30 Jahren hat sich in der statistischen Literatur eine dynamische Betrachtungsweise durchgesetzt, in der das Eintreten eines Ereignisses in Abhängigkeit von dessen Vorgeschichte modelliert wird. Veränderungen werden durchweg im Kontext ihrer zeitlichen Vorbedingungen analysiert. Diese Perspektive entspricht unmittelbar vielen Fragestellungen der Sozialwissenschaften. Insbesondere in der Lebensverlaufsforschung, der Analyse von Arbeitsmarktfragen, in der Demografie, aber auch in betriebswirtschaftlichen Modellen des Firmenverhaltens werden Forschungsfragen oft dynamisch formuliert, indem nach den Vorbedingungen spezifischer Ereignisse oder Veränderungen gefragt wird. Oder andersherum: Was sind die wahrscheinlichen zukünftigen Entwicklungen unter gegebenen Bedingungen? In beide Richtungen bietet die Verlaufsdatenanalyse direkte Anknüpfungspunkte, die sich als äußerst produktive Analyseperspektive erwiesen haben. Zeitlich lokale, dynamische Modelle für Verlaufsdaten sind in den letzten Jahren auch als Bausteine einer statistischen Analyse kausaler Zusammenhänge verwandt worden (Kühnel/Dingelstedt, Kapitel 80 in diesem Band). Diese Ansätze der Kausalanalyse sind gerade in den Sozialwissenschaften hilfreiche Ergänzungen zu heute schon klassischen statistischen Ansätzen, die sich weitgehend an randomisierten Experimenten orientieren (Eifler, Kapitel 11 in diesem Band).

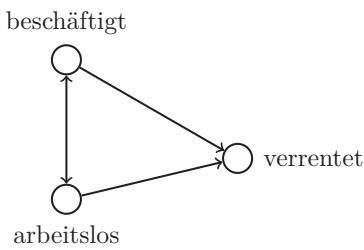
## 86.1 Ereignisse und Übergänge

Es gibt allerdings einige Beschränkungen, deren man sich bei der Verwendung von Methoden der Verlaufsdatenanalyse bewusst sein sollte. Die Verlaufsdatenanalyse setzt Ereignisse voraus, deren Typ festgestellt und von anderen Ereignisarten unterschieden werden kann. Da man Statistik mit Daten über Ereignisse betreiben möchte, muss der Ereignisbegriff zudem auf solche Ereignisse beschränkt werden, die mehrfach auftreten können. Man spricht dann von Ereignissen eines *Ereignistyps*. Denn nur für Ereignisse spezifischer Ereignistypen kann man sinnvoll Häufigkeiten ihres Eintretens und andere Statistiken berechnen. „Die deutsche Wiedervereinigung“, „das Erdbeben von Lissabon“ oder „die Kündigung des Arbeitsvertrages von Hans Müller zum Jahresende“ sind sicherlich Ereignisse. Aber solche singulären Ereignisse erfordern keine statistische Beschreibung. Auf der anderen Seite können zumindest die letzten beiden Beispiele auch unter die Ereignistypen „Erbeben“ und „Kündigung“ subsumiert werden, so dass häufig singuläre Ereignisse unter anderen Beschreibungen als Ereignisse eines Ereignistyps aufgefasst werden können.

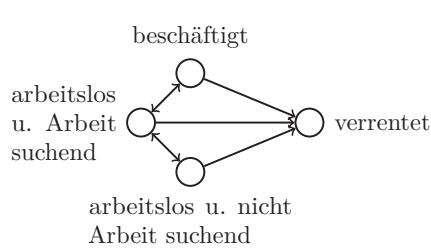
Die weitaus meisten Anwendungen in den Sozialwissenschaften beschränken sich zudem auf Ereignisse, die sich einzelnen Individuen (oder Organisationen, Firmen, Gruppen etc.) zuordnen lassen. In vielen Fällen ist die Zuordnung einfach: Man denke etwa an Geburten oder Verrentungen. In anderen Fällen sind mehrere Zuordnungen denkbar (Kündigungen, Heiraten etc.). Bei Erdbeben, Streiks, Demonstrationen etc. ist dagegen keine Zuordnung sinnvoll.

Hat man sich für eine Zuordnung entschieden, kann man Ereignisse häufig auch als Zustandsänderungen des jeweiligen Individuums auffassen. Die Beschreibung von Ereignissen eines Typs ergibt sich dann aus der Angabe der interessierenden Zustände des Individuums sowie den möglichen Übergängen zwischen diesen Zuständen. In einer Lebensverlaufsperspektive spricht man von einem *Biographieschema*, das wie in Abb. 86.1 (a) aussehen könnte.

Die Abbildung 86.1 (a) beschreibt die Arbeitsmarktbeteiligung einer Person durch die Unterscheidung der drei Zustände „beschäftigt“, „arbeitslos“ und „verrentet“. Ereignisse sind daher Übergänge zwischen diesen Zuständen. Man kann also aus einer Beschäftigung heraus arbeitslos werden, aus der Arbeitslosigkeit heraus eine Beschäftigung aufnehmen



(a) Ein einfaches Biographieschema.



(b) Ein verfeinertes Biographieschema.

**Abb. 86.1** Biographieschemata

oder verrentet werden. Die fehlenden, vom Zustand „verrentet“ ausgehenden Pfeile deuten an, dass Übergänge aus der Rente in eine Beschäftigung oder Arbeitslosigkeit nicht berücksichtigt werden sollen.

Ein Biographieschema kann weiter verfeinert werden, etwa indem zwischen „arbeitslos gemeldet und Arbeit suchend“ und „arbeitslos und nicht Arbeit suchend“ unterschieden wird. Das ist in Abb. 86.1 (b) angedeutet. Andersherum kann man ein Biographieschema auch vergrößern, indem z.B. Übergänge in die Rente außer Betracht gelassen werden. Das Biographieschema legt fest, welche Übergänge in einer statistischen Analyse interessieren sollen. Es ist nicht notwendigerweise eine eindeutige Abbildung sozialrechtlicher Regelungen sondern orientiert sich an sozialwissenschaftlichen Fragestellungen.

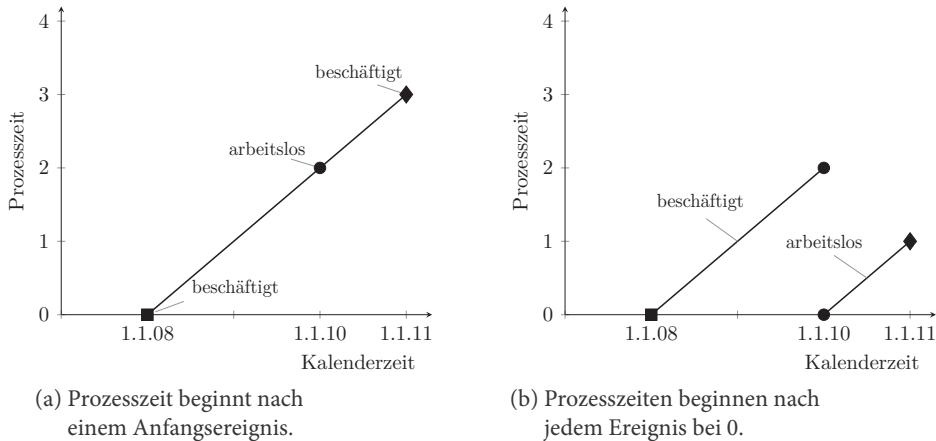
---

## 86.2 Zeitskalen

Mit einem Biographieschema hat man die interessierenden Zustände und damit die interessierenden Ereignistypen abgegrenzt. Man kann aber allein im Rahmen eines Biographieschemas noch nicht von Verläufen sprechen, weil ein Bezug auf die Zeit, auf das Aufeinanderfolgen von Ereignissen und damit eine Unterscheidung in Vergangenheit und Zukunft noch fehlt. Ein Biographieschema muss daher noch durch die Angabe einer *Zeitskala* ergänzt werden, um einen vollständigen Bezugsrahmen für die Modellierung von Verläufen zu schaffen. Eine Möglichkeit besteht in der Wahl einer vorgegebenen *Kalenderzeit* als allgemeinen Bezugsrahmen für die zeitliche Einordnung von Ereignissen. Ein Verlauf im Biographieschema (Abb. 86.1 (a)) lässt sich dann durch folgende Angaben darstellen: (erster) Beschäftigungsbeginn am 1.1.2008, (erster) Arbeitslosigkeitsbeginn am 1.1.2010, (zweiter) Beschäftigungsbeginn am 1.1.2011, ... .

In einer statistischen Perspektive ist man oft gar nicht an der exakten Lage eines individuellen Verlaufs in der historischen Zeit interessiert. Stattdessen sollen Aussagen etwa über die durchschnittliche Dauer der Arbeitslosigkeit abgeleitet werden. Dann ist es einfacher, anstelle der *Kalenderzeit* eine *Prozesszeit* zu definieren. In der Prozesszeit wird ein Ereigniszeitpunkt, etwa das Ausbildungsende oder der Beginn der ersten Arbeitslosigkeit, als Nullpunkt der Zeitskala ausgezeichnet und die Dauer seit diesem Ereignis dient als Zeitmaß. Insbesondere kann für jede Person eine je individuelle Zeitskala festgelegt werden.

Es ist auch möglich, die Zeitskala zu jedem Ereigniszeitpunkt wieder bei 0 beginnen zu lassen. In diesem Fall betrachtet man nur die Dauer zwischen zwei aufeinanderfolgenden Ereignissen. Man spricht auch von *Episoden*, die durch den Ausgangszustand, den End- oder Zielzustand, in den übergegangen wird, sowie durch die Dauer zwischen diesen Übergängen charakterisiert sind. Fügt man einen Verweis auf die Kalenderzeit bei Beginn der Episode oder auf die Anzahl von vorherigen Episoden seit einem ersten Ereigniszeitpunkt hinzu, dann kann man die Ereignisse einer Episode in der Kalenderzeit oder zumindest in der Prozesszeit seit einem ersten Ereignis rekonstruieren. Die Beziehung zwischen Prozess- und Kalenderzeit wird für beide Fälle in Abb. 86.2 durch *Lexis-Diagramme* illustriert. In Abb. 86.2 (a) ist die Prozesszeit durch ein einziges Anfangsereignis (etwa Eintritt in den



**Abb. 86.2** Lexis-Diagramme: Kalenderzeit und Prozesszeit. Die Symbole stellen Zeitpunkte für den Übergang in den folgenden Zustand dar.

Arbeitsmarkt) definiert, in Abb. 86.2 (b) beginnt jede Episode wieder zur Prozesszeit 0. Unterschiedliche Typen von Ereignissen sind durch unterschiedliche Symbole angedeutet.

Hat man sich auf ein Biographieschema festgelegt, dann reicht in der Prozessdarstellung (Abb. 86.2 (a)) offenbar die Angabe des jeweils beginnenden Zustands, um die Art der Übergänge festzulegen. In der Darstellungsform (Abb. 86.2 (b)) gibt es zu jeder Episode nur die Angabe ihres Typs. Aus welchem Zustand heraus jemand in diese Episode gelangt oder in welcher Form die Episode beendet wird, muss zusätzlich notiert werden.

### 86.3 Zensuren und unvollständige Informationen

Bevor wir uns einzelnen statistischen Verfahren zur Beschreibung von Verläufen zuwenden können, müssen wir noch auf ein Problem eingehen, das in vielen anderen Zusammenhängen unbeachtet bleiben kann. Weil Ereignisse und deren Beobachtung in der gleichen Kalenderzeit stattfinden, wird es i.d.R. Episoden geben, die bis zum Beobachtungsende noch nicht abgeschlossen sind. Ein Arbeitsloser muss ja nicht schon eine neue Arbeit gefunden haben, wenn der Beobachtungszeitraum beendet ist. Dann kennt man zum Ende des *Beobachtungsfensters* nicht die Dauer der Arbeitslosigkeitsepisode. Man weiß nur, dass die später einmal abgeschlossene Episode eine Dauer haben wird, die länger sein muss als die Zeit vom Beginn der Arbeitslosigkeit bis zum Beobachtungsende. Man spricht dann von *rechts zensierten* Beobachtungen. Zwar kennt man bei zensierten Beobachtungen nicht die Dauer der Episode, man weiß aber mehr als nichts, denn man kennt eine untere Schranke der Dauer, die Zeit bis zum Beobachtungsende. Diese Information sollte immer in die statistische Analyse einfließen. Denn würde man nur unzensierte Beobachtungen als Basis für die Berechnung von Statistiken benutzen oder, andersherum, die Zeit bis zum

Minimum von Beobachtungsende oder Übergang betrachten, dann ist klar, dass die Zeit bis zum Übergang in Beschäftigung immer unterschätzt wird.

Eine ganz ähnliche Form unvollständiger Daten aufgrund eines beschränkten Beobachtungsfensters ergibt sich, wenn Informationen über einen individuellen Verlauf erst nach einem bestimmten Datum oder Ereignis bekannt werden. Dann spricht man von *links abgeschnittenen* Beobachtungen. Auch in diesem Fall muss jede statistische Analyse die Möglichkeit links abgeschnittener Beobachtungen berücksichtigen.

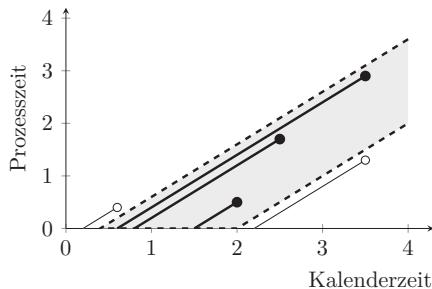
## 86.4 Erhebungsdesign, Beobachtungsfenster und unvollständige Informationen

Je nach dem Erhebungsdesign einer Studie und den erhobenen Informationen kann es zu weiteren Formen unvollständiger, aber partiell informativer Daten kommen. Eine hilfreiche Darstellungsweise bieten wieder Lexis-Diagramme, in die zusätzlich zu den Übergängen auch die Beobachtungsfenster eingetragen sind. Abbildung 86.3 illustriert einige Möglichkeiten des Zusammenspiels von Erhebungsdesign, Beobachtungsfenster und partieller Information für den Fall eines einzigen Episodentyps. Die Beobachtungsfenster sind grau unterlegt. Fette Linien und ausgefüllte Kreise deuten die mit dem jeweiligen Beobachtungsschema erzielte Information an, normale Linien und unausgefüllte Kreise hingegen Informationen, die nicht beobachtet werden.

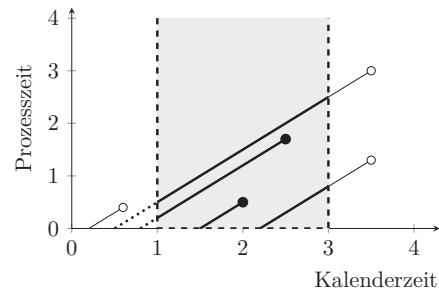
Bleibt man beim Beispiel von Arbeitslosigkeitsepisoden, dann entspricht Abb. 86.3 (a) einer Kohortenerhebung, bei der alle Arbeitslosigkeitsepisoden, die während eines Kalenderzeitintervalls beginnen, bis zu ihrem Ende weiterverfolgt werden. Über die vorher oder nachher beginnenden Arbeitslosigkeitsepisoden weiß man nichts. In diesem Erhebungsdesign gibt es keine links abgeschnittenen Beobachtungen. Rechts zensierte Beobachtungen entstehen nur, wenn die Beobachtung vor dem Ende der letzten Arbeitslosigkeitsepisode abgebrochen wird.

Abbildung 86.3 (b) zeigt ein einfaches, wenn auch eher selten gewähltes Erhebungsdesign, das das Zusammenspiel zwischen Erhebungsdesign und erhaltener Information besonders anschaulich illustriert. In diesem Design werden alle während eines Zeitraums beginnenden, endenden oder durchgehend andauernden Arbeitslosigkeitsepisoden erfasst. Man erhält dann allerdings nur partielle Informationen über die Dauern von Arbeitslosigkeit: Die erste, ganz links dargestellte Episode fällt nicht in das Beobachtungsfenster. Die zweite ist rechts zensiert, weil die Dauer der Episode am Beobachtungsende noch unbekannt ist. Sie ist aber auch links abgeschnitten, weil die Voraussetzung für die Aufnahme in die Erhebung ja gerade darin besteht, arbeitslos zu sein. Also entspricht diese Beobachtung jenen Episoden, die zumindest länger als die Zeit vom Beginn der Episode zum Beginn des Beobachtungsfensters sind. Die dritte Beobachtung ist ebenfalls links abgeschnitten, die vierte ist vollständig beobachtet und die letzte ist rechts zensiert.

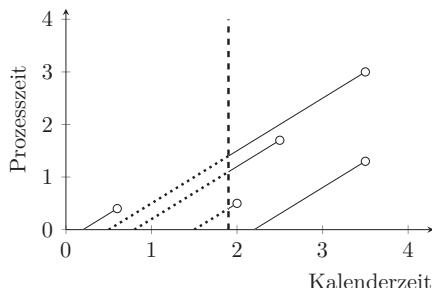
Dieses Erhebungsdesign wird selten verwandt, weil es weniger Informationen als eine Kohortenerhebung liefert (Stein, Kapitel 7 in diesem Band), aber mindestens den gleichen



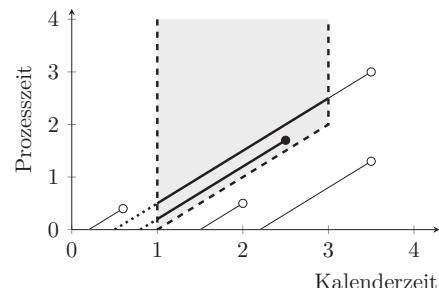
(a) Kohortenerhebung.



(b) Beobachtungen in festem Zeitfenster.



(c) Querschnitterhebung.

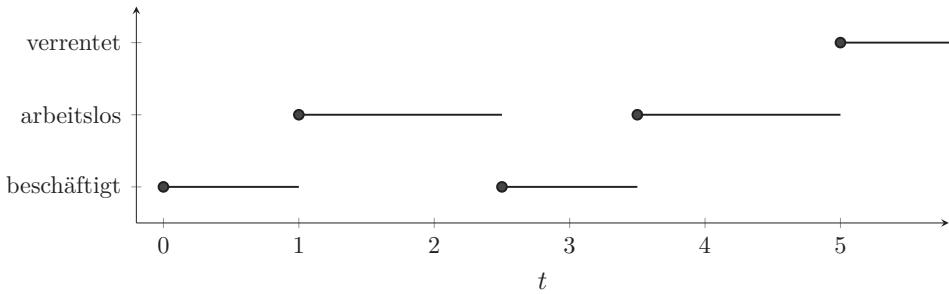


(d) Panelerhebung.

**Abb. 86.3** Erhebungsdesign, Beobachtungsfenster und Information in Lexis-Diagrammen. Die Beobachtungsfenster sind grau unterlegt, die erhaltene Information ist durch fette Linien — bzw. durch volle Kreise ● angedeutet. Gepunktete Linien ··· verweisen auf links abgeschnittene Beobachtungen, leere Kreise ○ auf unbeobachtete Ereignisse.

Erhebungsaufwand erfordert. Ein Grenzfall, bei dem die Länge des Beobachtungsintervalls zu einem Zeitpunkt zusammenschrumpft, entsteht oft im Rahmen von allgemeinen Querschnittbefragungen (Abb. 86.3 (c)). Dann kennt man nur die bisherige Dauer der Arbeitslosigkeit aller Befragten, die zum Befragungszeitpunkt arbeitslos waren. Der Informationsgehalt dieses Designs ist offenbar sehr gering, weil Angaben über die Dauer abgeschlossener Arbeitslosigkeitsepisoden nicht zustande kommen können. Die Verteilung der bisherigen Dauern liefert ohne zusätzliche Daten oder Annahmen nur sehr grobe Schranken für die Verteilung der Arbeitslosigkeitsdauern.

Daher werden oft Querschnittserhebungen zu Panelerhebungen (Schupp, Kapitel 73 in diesem Band) erweitert. Das ist schematisch in Abb. 86.3 (d) angedeutet. Die Auswahl der Arbeitslosigkeitsepisoden erfolgt wie in der Querschnitterhebung zu einem Zeitpunkt. Dann verfolgt man aber wie in einer Kohortenerhebung alle Prozesse für eine gewisse Zeit weiter. Das Erhebungsverfahren ist oft einfacher zu realisieren als eine Kohortenerhebung und es ist informativer als Querschnittserhebungen. Allerdings sieht man auch, dass wie bei Beobachtungsverfahren mit festem Zeitfenster sowohl vollständige Informationen über Episoden als auch rechts zensierte oder links abgeschnittene unvollständige Informationen zu erwarten sind.



**Abb. 86.4** Ein Verlauf in Prozesszeit ab einem Anfangsereignis.

## 86.5 Datenstrukturen

Praktisch entstehen Beobachtungsergebnisse unter sehr viel komplexeren Bedingungen als den angedeuteten Erhebungsdesigns. Zudem interessiert oft das Zusammenspiel mehrerer Prozesse. Eine angemessene Analyse von Verlaufsdaten erfordert daher als Ausgangspunkt eine möglichst einfache und flexible Form, um Beobachtungen von Verläufen auf Rechnern repräsentieren zu können. Insbesondere sollte es möglich sein, unterschiedliche Formen unvollständiger Daten ebenso wie Entwicklungen in anderen, parallel stattfindenden Prozessen in einem einheitlichen Schema angeben zu können. Beides gelingt mit einer Struktur, die die meisten statistischen Analyseprogramme interpretieren und aus Dateien oder Datenbanken einlesen können. Die Datenstrukturen sollen hier angedeutet werden, weil sie auch einen Zugang zu statistischen Methoden der Verlaufsdatenanalyse eröffnen. Betrachten wir wieder das Biographieschema Abb. 86.1 (a) mit den Zuständen „beschäftigt“, „arbeitslos“ und „verrentet“, dann kann man sich einen Verlauf wie in Abb. 86.4 vorstellen.

Eine sehr nützliche Datenstruktur notiert diesen Verlauf, indem Informationen zu Episoden jeweils in einer Zeile gespeichert werden. Zeiten beziehen sich auf eine Prozesszeitachse wie in Abb. 86.2 (b), in der jede Episode wieder bei 0 beginnt. Zusätzlich muss für jede Episode deren Typ angeben werden. Diese Angaben reichen noch nicht aus, um auch die Reihenfolge der Episoden bzw. ihre Lage bzgl. eines vorgegebenen Ereignisses zu bestimmen. Dazu wird jeder Episode ihre Episodennummer – ihre Folgenummer in der zeitlich geordneten Liste der Episoden der Person  $u$  – zugewiesen. Damit lässt sich auch die Prozesszeit seit einem Anfangsereignis wie in Abb. 86.2 (a) rekonstruieren, etwa die Zeit seit dem Anfangsereignis der ersten Episode der Person  $u$ . Ergänzt man diese Datenstruktur um die Angabe des Zustands, in den der Prozess am Ende einer Episode übergeht, dann hat der Verlauf in Abb. 86.4 die Repräsentation:

Person	Episodennummer	Episodentyp	Zielzustand	$t_{\text{Anf}}$	$t_{\text{Ende}}$
$u$	1	beschäftigt	arbeitslos	0	1
$u$	2	arbeitslos	beschäftigt	0	1.5
$u$	3	beschäftigt	arbeitslos	0	1
$u$	4	arbeitslos	verrentet	0	1.5
$u$	5	verrentet	verrentet	0	0.8

Rechts zensierte Beobachtungen sind dadurch charakterisiert, dass der Zielzustand der Episode gleich dem Episodentyp ist. Denn dann ist bis zum Zeitpunkt  $t_{\text{Ende}}$  noch kein Zustandswechsel und somit auch noch kein Ereignis eingetreten. Das ist in der Episode 5 der Tabelle illustriert und motiviert die Aufnahme des Zielzustands in die Datenstruktur.

---

## 86.6 Parallele Prozesse, zeitabhängige Kovariablen und Episodensplitting

In unserem Beispiel bleibt noch unklar, warum die Spalte  $t_{\text{Anf}}$  in die Datenstruktur aufgenommen werden sollte. Sie hat in dem Beispiel immer den Wert 0. Eine offensichtliche Nutzungsmöglichkeit besteht in der Repräsentation links abgeschnittener Beobachtungen: Hätte die Beobachtung des Prozesses erst nach einem halben Jahr begonnen, dann könnte man das durch die Angabe  $t_{\text{Anf}} = 0.5$  in der ersten Episode festhalten.

Für die Analyse von Verläufen erweist es sich aber als viel wichtiger, dass diese Angabe auch die Repräsentation zeitlich parallel verlaufender Prozesse erlaubt. Ein zentrales Interesse der Verlaufsdatenanalyse ist es ja gerade, die Abhängigkeit eines Verlaufs, etwa der Arbeitsmarktbeteiligung, von weiteren Prozessen wie regionalen Arbeitsmarktindikatoren oder Fortbildungen etc. zu beschreiben. Solche sich parallel zum interessierenden Verlauf entwickelnden Prozesse nennt man in der Statistik oft *zeitabhängige Kovariablen*.

Diese zusätzliche Information kann leicht in die Datenstruktur integriert werden, indem kürzere, durch die Ereignisse (Zustandswechsel) der zeitabhängigen Kovariablen definierte Episoden zusätzlich zu den Episoden des interessierenden Verlaufs eingeführt werden. Man spricht dann von *Episodensplitting*. Besucht eine Person  $u$  etwa zwischen Zeit 1.5 und 2 eine Fortbildung, dann würde man zwischen Episode 2 und 3 zwei zusätzliche Episoden einfügen, die den Besuch der Fortbildung festhalten.

Dabei ist die neue Episode 2a censiert, sie dauert ja auch noch an, als die Person  $u$  eine Fortbildung beginnt. Ihr Zielzustand „arbeitslos“ ist auch der Episodentyp. Gleiches gilt für die Episode 2b. Diese Episode entspricht der Fortbildungsepisode, entsprechend hat die Spalte „Fortbildung“ hier eine 1. Dagegen ist Episode 2c nicht censiert, denn zurzeit 2.5 findet ein Übergang in Beschäftigung statt. Der Beginn dieser Teilepisode  $t_{\text{Anf}}$  ist nun 1, weil zu diesem Zeitpunkt die Fortbildung begann.

Person	Episodennummer	Episodentyp	Zielzustand	$t_{\text{Anf}}$	$t_{\text{Ende}}$	Fortbildung
$u$	1	beschäftigt	arbeitslos	0	1	0
$u$	2a	arbeitslos	arbeitslos	0	0.5	0
$u$	2b	arbeitslos	arbeitslos	0.5	1	1
$u$	2c	arbeitslos	beschäftigt	0	1.5	0
$u$	3	beschäftigt	arbeitslos	0	1	0
$u$	4	arbeitslos	verrentet	0	1.5	0
$u$	5	verrentet	verrentet	0	0.8	0

Das Episodensplitting ist in allen gängigen Programmen zur Verlaufsdatenanalyse implementiert. Fast alle Statistikprogramme enthalten zudem Routinen, die die hier vorgestellte Verlaufsdatenstruktur aus diversen anderen Datenrepräsentationen erstellen.

## 86.7 Die Ratenfunktion

Im einfachsten Fall enthält das Biographieschema nur einen einzigen Übergang, etwa von der Arbeitslosigkeit in eine Beschäftigung. Lässt man vorläufig alle weiteren Aspekte außer Betracht, dann wird im Rahmen eines statistischen Modells alles Wesentliche durch die Verteilungsfunktion der Dauer  $T$  der Arbeitslosigkeit  $\Pr(T \leq t)$  ausgedrückt. Die Verteilungsfunktion ist der Anteil der Personen, die höchstens  $t$  Monate arbeitslos waren, aufgefasst als Funktion von  $t$ . Nun ist die Verteilungsfunktion ungeachtet ihrer fundamentalen Bedeutung in der Statistik sicherlich kein dynamisches Konzept, denn sie bezieht sich immer auf abgeschlossene Perioden der Arbeitslosigkeit. In der Verlaufsdatenanalyse benutzt man daher anstelle der Verteilungsfunktion die *Rate*, eine zeitlich lokale Beschreibung der Dauerverteilung, die die Dynamik von Übergängen in Beschäftigung in den Vordergrund rückt. Die Rate ist definiert als die Wahrscheinlichkeit eines Übergangs zum Zeitpunkt  $t$  unter der Bedingung, dass sich bisher noch kein Übergang ereignet hat. Sie beschreibt, was als nächstes unter gegebenen Umständen passiert. Formal wird das in der Formel

$$r(t) = \Pr(T = t \mid T > t-1)$$

ausgedrückt, in der  $\Pr(T = t \mid T > t-1)$  die bedingte Wahrscheinlichkeit für eine Arbeitslosigkeitsdauer von genau  $t$  Monaten angibt, wenn man mindestens  $t - 1$  Monate arbeitslos war.

In unserem Beispiel mit nur einem Übergang im Biographieschema ergibt sich noch keine inhaltliche Differenz zwischen Beschreibungen durch Verteilungsfunktionen oder durch Raten. Beide sind in diesem Fall äquivalent. Die Verteilungsfunktion lässt sich aus der Ratenfunktion berechnen und andersherum kann die Ratenfunktion aus der Verteilungsfunktion abgeleitet werden. Das ändert sich, sobald man die Abhängigkeit von Übergängen von weiteren Ereignissen und zeitlich veränderlichen Bedingungen

berücksichtigen möchte, wenn also etwa die Dauer der Arbeitslosigkeit in Abhängigkeit von Veränderungen auf dem Arbeitsmarkt sowie von Fortbildungen und anderen Aktivitäten der Betroffenen beschrieben werden soll. Dann kann immer noch eine Ratenfunktion definiert werden, die nun von der gesamten Vorgeschiede, also bspw. von der bisherigen Arbeitslosigkeitserfahrung, abhängig sein kann.

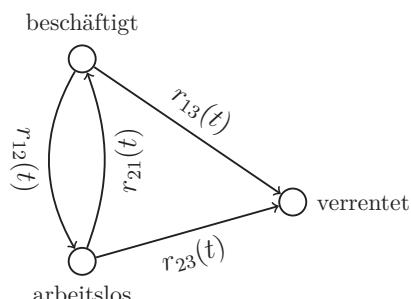
Wichtig ist dabei, dass man die Ratenfunktion an der Zeitstelle  $t$  unter Bedingungen berechnet, die bis zur Zeitstelle  $t - 1$  festliegen, also aus der Kenntnis der strikten Vergangenheit (vor  $t$ ) bestimmt werden können. Insbesondere gehört wieder zu den Bedingungen, dass der interessierende Übergang bis  $t - 1$  einschließlich noch nicht eingetreten ist. Dieser verallgemeinerte Ratenbegriff unterscheidet explizit zwischen einer für die Akteure abgeschlossenen Vergangenheit und möglichen zukünftigen Ereignissen. So entstehen Beschreibungen zukünftiger Ereignisse in Abhängigkeit von vergangenen Ereignissen (fast) beliebiger Art. Damit ist es möglich, Konzepte der Kausalanalyse wie Reichenbachs „Common Cause Principle“, das eine ausgezeichnete Zeitrichtung voraussetzt, in einem statistischen Kontext durch verallgemeinerte Ratenfunktionen auszudrücken. Zudem lassen sich viele analytische Fragestellungen in allen Bereichen der Sozialwissenschaft als Fragen nach (vergangenen) Vorbedingungen zukünftiger Ereignisse formulieren, so dass es nahe liegt, bei ihrer statistischen Modellierung verallgemeinerte Ratenfunktionen zu benutzen.

Enthält das Biographieschema mehrere Übergänge, dann kann man für jeden Episodenotyp und jeden Zielzustand getrennt Ratenfunktionen definieren. Die Situation ist in Abb. 86.5 skizziert. Es kann gezeigt werden, dass diese getrennt berechenbaren Ratenfunktionen alle statistisch relevanten Informationen über einen Verlauf enthalten. Insbesondere kann man aus ihnen etwa die durchschnittlichen Verweildauern in allen Episodenarten, die Häufigkeit von Sequenzmustern und weitere Charakteristika rekonstruieren.

## 86.8 Regressionsmodelle

Wie in anderen Bereichen der Statistik auch ist das wohl wichtigste Hilfsmittel zur Beschreibung von Abhängigkeiten zwischen Verläufen ein Regressionsmodell. In der Verlaufsdatenanalyse haben ein spezielles Regressionsmodell und ein dazu passendes Schätzverfahren

**Abb. 86.5** Ratenfunktionen zwischen den Zuständen des Biographieschemas.



wegen ihrer im Allgemeinen sehr günstigen Eigenschaften andere Ansätze fast vollständig verdrängt. Die Modellannahme lässt sich wieder am Einfachsten durch die Ratenfunktion angeben: Die Ratenfunktion zum Zeitpunkt  $t$ , bedingt auf die Werte einer zeitabhängigen Kovariablen  $X(t-1)$ , soll die Form

$$r(t | X(t-1)) = r_0(t) \exp(X(t-1)\beta)$$

haben. Dabei ist  $r_0(\cdot)$  eine nicht näher spezifizierte Ratenfunktion, die sogenannte *Basisrate*, die für alle Fälle als gleich angenommen wird. Der Einfluss der (zeitabhängigen) Kovariablen besteht nun darin, diese Rate an jeder Zeitstelle mit dem Wert  $\exp(X(t-1)\beta)$  zu multiplizieren. Dieses Modell nennt man ein *proportionales Ratenmodell I*. Ein Wert von  $\exp(X(t-1)\beta) > 1$  führt zu einer höheren Rate und damit zu kürzeren Dauern, während ein Wert  $\exp(X(t-1)\beta) < 1$  zu kleineren Raten und damit zu längeren Dauern führt. Das Modell gehört zur Klasse der *semiparametrischen* Modelle, die sowohl durch einen endlich-dimensionalen Vektor von Parametern  $\beta$  als auch durch eine beliebige (nicht negative) Funktion  $r_0(\cdot)$  bestimmt wird. David Cox (1972) hat ein Schätzverfahren vorgeschlagen, dass keine Annahmen über die Form der Basisrate benötigt und gleichzeitig Effekte von (zeitabhängigen) Kovariablen auf die Ratenfunktion effizient schätzen kann. Das Schätzverfahren, die *partielle Likelihood*, beruht auf der Idee, dass zwar die Verteilung der jeweiligen Dauer wesentlich von dem zugrundeliegenden  $r_0(\cdot)$  bestimmt wird, dass aber die Verteilung der Ränge der jeweiligen Dauer nur von den Kovariablenfunktionen  $\exp(X(t-1)\beta)$  abhängt. Es hat sich gezeigt, dass dieses Schätzverfahren, das manchmal auch etwas irreführend als *Cox' proportionales Ratenmodell* bezeichnet wird, in der Lage ist, die Information über Effekte der Kovariablen fast ebenso gut auszunutzen wie klassische statistische Verfahren. Außerdem ist es robust gegenüber leichten Abweichungen von der Proportionalitätsannahme. Aus diesen Gründen ist das proportionale Ratenmodell zusammen mit der partiellen Likelihood als Schätzverfahren der wesentliche Baustein für die Analyse auch komplexer Verlaufsdaten geworden.

---

## 86.9 Weiterführende Literatur

Der klassische Ansatz der Verlaufsdatenanalyse wird wohl am besten durch die kurze, aber sehr lesenswerte Darstellung von Cox/Oakes (1984) wiedergegeben. Die dynamische Betrachtungsweise wird im Lehrbuch von Aalen et al. (2008) betont. Die Bücher von Blossfeld et al. (2007) und von Cleves et al. (2013) sind gut lesbare Einführungen in die Verlaufsdatenanalyse, die die praktische Umsetzung der Analysekonzepte mit *Stata* demonstrieren. Einführungen in die Verlaufsdatenanalyse, die sich bei der praktischen Umsetzung auf die Software *R* beziehen, sind u.a. von Mills (2011) sowie von Therneau/Grambsch (2000) vorgelegt worden.

## Literatur

- Aalen, Odd O./Borgan, Ørnulf/Gjessing, Håkon (2008): Survival and Event History Analysis. A Process Point of View. Berlin: Springer
- Blossfeld, Hans-Peter/Rohwer, Götz/Golsch, Katrin (2007): Event History Analysis with Stata. Mahwah: Lawrence Erlbaum
- Cleves, Mario/Gould, William W./Gutierrez, Roberto/Marchenko, Yulia (2010): An Introduction to Survival Analysis Using Stata. College Station: Stata Press
- Cox, David/Oakes, David (1984): Analysis of Survival Data. London: Chapman & Hall
- Mills, Melinda (2011): Introducing Survival and Event History Analysis. Thousand Oaks: Sage
- Therneau, Terry M./Grambsch, Patricia M. (2000): Modeling Survival Data. Extending the Cox Model. Berlin: Springer

Manuela Pötschke

---

## 87.1 Begriff und Einordnung

In der modernen Gesellschaft gewinnen Bildungskarrieren und Bildungsabschlüsse immer mehr an Bedeutung. Neben dem höheren Schulabschluss an sich spielt dabei die Abschlussnote als Indikator für die Leistungsqualität eine Rolle. Wovon hängen nun aber gute oder weniger gute Schulleistungen ab? Naheliegend ist, beispielsweise besondere Talente, das Vorwissen, die Interessen der Schüler oder die zur Verfügung stehende Lernzeit als Erklärungen heranzuziehen. Seitdem die Ergebnisse ländervergleichender Schulstudien wie PISA oder TIMSS veröffentlicht wurden, wird darüber hinaus auch das Lernumfeld als Einflussfaktor breit diskutiert und z.B. die Wirkung der sozialen Herkunft oder von Klasseneigenschaften empirisch gezeigt (Neumann et al. 2010). Erste Erfahrungen mit entsprechenden Analysestrategien gehen auf Coleman et al. (1966, 1981) zurück, die die Schulleistungen von Schülern in Abhängigkeit von der sozialen Zusammensetzung der Schülerschaft einer Schule (Coleman et al. 1966) bzw. des Schultyps (Coleman et al. 1981) betrachteten.

Die Idee der Integration von individuellen Merkmalen und Kontextinformationen in ein Erklärungsmodell des Schulerfolgs kann in der statistischen Analyse durch ein Mehrebenenmodell umgesetzt werden. Die Schulforschung ist dabei zwar nach wie vor eines der zentralen Anwendungsgebiete und ein Beispiel wird deshalb auch hier zur Verdeutlichung der Grundidee verwendet werden. Aber in der Forschungspraxis sind die untersuchten Fragen nicht thematisch darauf beschränkt, sondern kommen aus verschiedenen Fachwissenschaften und beziehen sich auf vielfältige Handlungsfelder.

Mit Hummels (1972: 13) lässt sich allgemein sagen, dass Mehrebenenanalysen immer dann vorliegen, wenn „Objekte verschiedener Ordnung gleichzeitig zum Gegenstand der Analyse werden“. Das heißt, es liegen Daten auf unterschiedlichen Aggregatebenen bzw. Abstraktionsniveaus vor. Es gibt eine Reihe traditioneller Strategien, diesem Sachverhalt in statistischen Analysen zu begegnen:

1. Eine Möglichkeit, individuelles Handeln zu erklären, besteht zuerst einmal darin, die Kontextabhängigkeit des Handelns zu ignorieren und es lediglich auf individuelle Eigenschaften der Person zurückzuführen. Zum Beispiel könnten Schülerleistungen ausschließlich auf individuelle Merkmale der Schüler bezogen werden und die Schulklass als Umfeld unberücksichtigt bleiben. Wenn die Leistungen als der zu erklärende Sachverhalt und die Wirkungen von individuellen erklärenden Variablen nicht durch das Umfeld beeinflusst werden, wäre diese Strategie angemessen. Die Unabhängigkeit individueller Handlung vom Kontext wird dadurch vorweggenommen und nicht empirisch geprüft, was zu Fehlschlüssen über die Abhängigkeiten der einbezogenen Variablen führen kann. Im Beispiel würden die Leistungsunterschiede zwischen Schülern allein durch die Schülermerkmale zustande kommen. Die Schulforschung hat aber nicht zuletzt in den PISA-Auswertungen hinlänglich zeigen können, dass die Schulleistungen eben auch von den Eigenschaften einer Schulklass, einer Schule oder des Bildungssystems abhängen (Prenzel/Baumert 2008).
2. Eine weitere Strategie des Umgangs mit der Kontextabhängigkeit individuellen Handelns ist in der Aggregierung von Individualdaten (Graeff, Kapitel 72 in diesem Band) zu finden. Dieses Vorgehen unterstellt die Homogenität der Aggregateinheiten und übersieht die Variabilität innerhalb der Aggregateinheiten. Wird beispielweise das Leistungsniveau von Schulklassen auf der Basis individueller Schülerleistungstests gebildet, kann das gleiche Niveau durch unterschiedliche Kompositionen der einzelnen Leistungen zustande kommen. Die Schüler einer Klasse könnten sich in ihren Leistungen sehr ähneln oder sehr unterschiedlich sein und in beiden Fällen den gleichen mittleren Leistungswert für die Klasse generieren. Merkmale wie die Klassengröße wirken möglicherweise in Abhängigkeit davon unterschiedlich, durch welche Konstellation die mittleren Klassenwerte gebildet wurden. Aggregatanalysen können dieser Unterschiedlichkeit nicht gerecht werden. Darüber hinaus kann aus Aggregatanalysen nicht auf Wirkungsmechanismen auf individueller Ebene geschlossen werden, so dass Kausalinterpretationen (Kühnel/Dingelstedt, Kapitel 80 in diesem Band) nicht geleistet werden können.
3. Eine Strategie besteht darin, die Kontextinformationen als individuelle Merkmale einzuführen. Im Schulbeispiel würden den einzelnen Schülern Merkmale wie die Klassengröße oder die Zusammensetzung der Klasse individuell zugeschrieben und dadurch Gruppenunterschiede als individuelle Variationen spezifiziert. Schüler aus unterschiedlichen Klassen mit gleicher Klassengröße oder -zusammensetzung weisen in diesem Fall die gleichen Merkmalsausprägungen auf, obwohl die Klassenkontakte unterschiedlich wirken.
4. Eine weitere Möglichkeit des Umgangs mit kontextbezogenen Daten besteht darin, getrennte Analysen für jeden einzelnen Kontext durchzuführen. Die Schulleistungen von Schülern werden hier auf der Basis der individuellen Merkmale für jede Klasse einzeln geschätzt. Danach werden die erhaltenen Regressionskoeffizienten miteinander verglichen. Dieses Vorgehen greift die zu trennenden Handlungsabhängigkeiten auf individueller und kontextueller Ebene adäquat auf. Die Interpretation ist bei einer

großen Zahl von Kontexten jedoch schnell unübersichtlich und eine systematische Darstellung der Gruppenvergleiche ist kaum möglich. Außerdem ist der Einbezug von Kontextmerkmalen zur Erklärung nicht vorgesehen.

5. Aus der gerade beschriebenen vergleichenden Analyse von Teilgruppen geht eine Strategie hervor, die der Mehrebenenanalyse am nächsten kommt, nämlich die der sogenannten "Slopes as Outcome"-Analysen. Hier werden zuerst Regressionsmodelle mit Individualmerkmalen für jede einzelne Gruppe, z.B. jede Klasse, separat berechnet. Die Regressionskoeffizienten (Intercept und Slope) dieser Modelle werden dann als abhängige Variablen in weitere Regressionsmodelle eingeführt, wobei zu ihrer Erklärung Kontextmerkmale genutzt werden. Kontexteinflüsse und Individualeffekte werden dadurch systematisch getrennt präsentiert; eine gleichzeitige Betrachtung von Individual- und Kontexteffekten ist aber nicht möglich.

Mehrebenenanalysen vereinigen die Wirkungen individueller und kontextbezogener Merkmale in einem Modell. Dadurch wird ein besseres Abbild der vermuteten Wirkungszusammenhänge auf menschliches Handeln erreicht als mit den klassischen Herangehensweisen.

Neben den inhaltlichen Gründen spricht die verbreitete sozialwissenschaftliche Praxis gestufter Stichprobenziehungen (Häder/Häder, Kapitel 18 in diesem Band) für die Verwendung von Mehrebenenmodellen. Die Interpretation von Analyseergebnissen aus klassischen Modellen ohne Ebenenstruktur führt in diesen Fällen zu einer Unterschätzung der Standardfehler. Das Ausmaß der Unterschätzung wird als Designeffekt bezeichnet. In Mehrebenenanalysen kann der gestuften Stichprobenziehung durch die Modellspezifikation direkt Rechnung getragen werden, so dass korrekte Interpretationen von Signifikanzen möglich sind (für eine Beispielberechnung der Auswirkungen des Designeffekts auf die notwendige Stichprobengröße in mehrstufigen Ziehungen im Vergleich zu einfachen Zufallsziehungen siehe Snijders/Bosker2012: 23-24).

Zusammenfassend können als Potentiale für Mehrebenenmodelle also festgehalten werden,

1. dass sich auf ihrer Basis die Kontextabhängigkeit und die individuelle Variabilität eines Sachverhalts gleichzeitig in einem Modell zeigen lassen. Zum Beispiel lässt sich angeben, welcher Anteil der Leistungsunterschiede von Schülern auf klassenspezifische Aspekte und welcher Anteil auf persönliche Eigenschaften rückführbar ist.
2. dass die gleichzeitige Untersuchung von Effekten verschiedener Ebenen und
3. dass der Einbezug von Wechselwirkungen (Interaktionseffekten) über verschiedene Ebenen hinweg möglich wird. Die spezifische Kombination von Merkmalsausprägungen einer individuellen und einer kontextuellen Variable entfaltet dann eine eigenständige Wirkung. Wenn eine umfangreichere Übungszeit also nicht nur für sich wirkt, sondern einen zusätzlichen Effekt entfaltet, je nachdem wie groß die Klasse ist, in der ein Schüler lernt, sind Interaktionseffekte zu beobachten.
4. dass die zufällige Variation von Effekten unterschiedlicher Ebenen gleichzeitig berücksichtigt werden kann und

5. dass die Standardfehler für mehrstufige Zufallsauswahlen (Häder/Häder, Kapitel 18 in diesem Band) korrekt geschätzt werden.

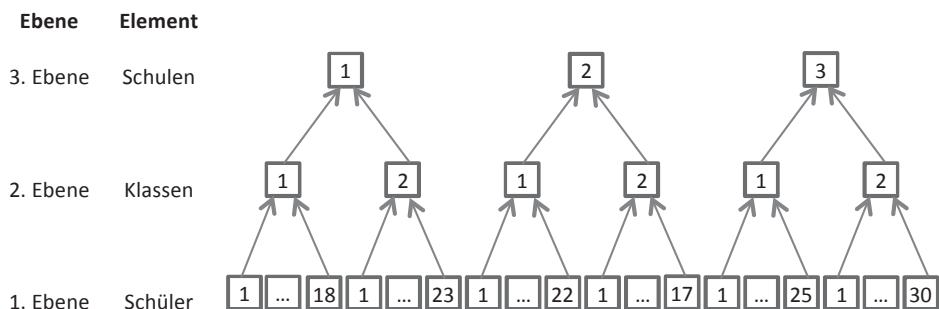
## 87.2 Anforderungen an die Daten

### 87.2.1 2.1 Hierarchische Datenstruktur

Die Untersuchungsobjekte in einer Mehrebenenanalyse sind in der Regel in einer hierarchischen Struktur organisiert. Das heißt, dass die Elemente einer unteren, ersten Ebene dem Inklusionsprinzip folgend jeweils genau einem Element der oberen, zweiten Ebene zuzuordnen sind. So sind in Schulstudien beispielsweise Schüler jeweils genau einer Klasse zugeordnet. Außerdem umfassen die Elemente der oberen Elemente insgesamt alle Elemente der ersten Ebene. Es gibt also keinen Schüler, der keiner Klasse zugeordnet werden könnte. In diese Logik können nun weitere Aggregateinheiten einbezogen werden, im Beispiel wären das Schulen. In Abb. 87.1 sind zur Verdeutlichung des Prinzips sechs Klassen aus drei Schulen aufgeführt, in denen eine unterschiedliche Anzahl von Schülern befragt wurde. In der ersten Klasse liegen individuelle Informationen von insgesamt 18, in der zweiten Klasse von 23 Schülern usw. vor.

Für jede Ebene können nun spezifische Merkmale der jeweiligen Elemente in die Datenanalyse aufgenommen werden. Der Typologie von Lazarsfeld/Menzel (1961) folgend, kann es sich hierbei um absolute, komparative, relationale oder kontextuelle Merkmale für die Individualebene und um analytische, strukturelle oder globale Merkmale für die höheren Ebenen handeln:

- Bei den absoluten Merkmalen handelt es sich um Individualeigenschaften, die ohne Bezug zum Kontext oder zu anderen Personen bestehen (z.B. Geschlecht, Taschengeld).
- Komparative Merkmale drücken die relative Position eines Individuums in einer Verteilung aus (z.B. über- oder unterdurchschnittliche Höhe des Taschengeldes).
- Relationale Merkmale (Baur, Kapitel 74 in diesem Band) beschreiben dagegen die Beziehungen zu anderen Individuen (z.B. Popularität oder Isolation in einer Klasse).



**Abb. 87.1** Hierarchische Struktur eines Mehrebenenmodells

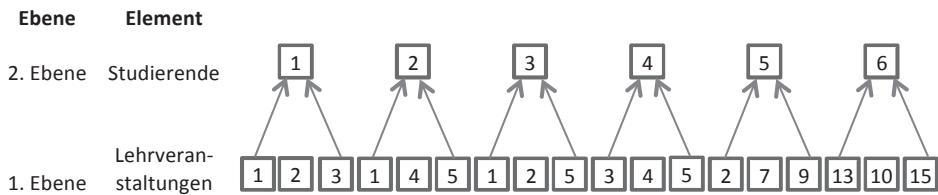
- Kontextuelle Merkmale beschreiben die Individuen als Mitglieder eines spezifischen Kontextes (z.B. Schüler in einer leistungsstarken oder -schwachen Klasse).
- Analytische Merkmale der Kontexte basieren auf Informationen der Individuen (z.B. Anteil der Mädchen in einer Klasse).
- Strukturelle Merkmale beschreiben Beziehungsstrukturen in den Kontexten (z.B. Grad der Fraktionierung von Klassen).
- Globale Merkmale sind solche, die nicht auf individuellen Informationen basieren (z.B. Rituale und Traditionen einer Klasse).

Die Differenzierung der unterschiedlichen Merkmalstypen innerhalb der Ebenen ist lediglich hilfreich für das Verständnis unterschiedlicher Eigenschaften. Die Zuordnung der Daten zu einer spezifischen Ebene muss jedoch erfolgen, weil dadurch die, für die Schätzung eines Mehrebenenmodells notwendige, hierarchische Datenstruktur entsteht.

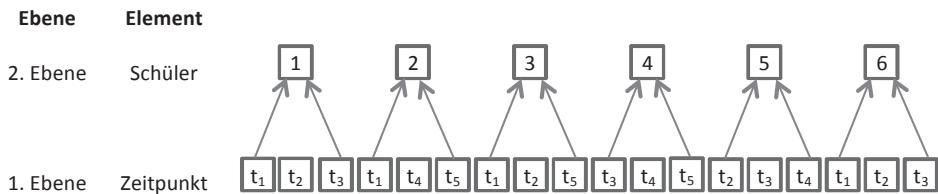
Häufig sind, wie im Schülerbeispiel, Personen auf der ersten Ebene anzutreffen, die dann auf der zweiten Ebene zu Gruppen zusammengefasst sind, die wiederum größere Gruppen bilden können. Es ist jedoch nicht zwingend, dass es sich bei den Elementen der ersten Ebene um Personen handelt. Werden beispielsweise Studierende mehrfach gebeten, unterschiedliche Lehrveranstaltungen zu evaluieren, so lassen sich die einzelnen Beurteilungen als Qualitätsausweis der Veranstaltungen interpretieren und auf der ersten Ebene verorten. Die Varianz in den Einschätzungen auf dieser ersten Ebene ist dann auf die Unterschiede in den Qualitäten der Lehrveranstaltungen zurückzuführen.

Auf der zweiten Ebene sind Studierende zu betrachten, die sich in ihren Bewertungsgewohnheiten, ihren prinzipiellen Bewertungskriterien oder Grundhaltungen aber auch in ihren soziodemographischen Merkmalen unterscheiden. Abbildung 87.2 stellt das Prinzip des Designs dar. Hier sind beispielhaft sechs Studierende auf der zweiten Ebene aufgeführt, die jeweils unterschiedliche Lehrveranstaltungen beurteilt haben. Die Einzelbeurteilungen der Lehrveranstaltungen befinden sich auf der ersten Ebene. Die Eigenschaften der Studierenden werden nun zur Erklärung von Einzelbeurteilungen herangezogen. So sind einige Studierende prinzipiell als optimistisch und positiv denkend zu kennzeichnen und andere verfügen per se über eine eher negative Grundstimmung. Dies wird sich in den Beurteilungen der Lehrveranstaltung niederschlagen, ein Analysebeispiel mit dem beschriebenen Design gibt Engel (1998).

Wenn in Panelanalysen (Schupp, Kapitel 73 in diesem Band) gleichzeitig individueller Wandel und der Vergleich von Personen thematisiert werden sollen, wird die erste Ebene eines Mehrebenenmodells durch die Messzeitpunkte gebildet. Beispielsweise werden Schüler in regelmäßigen Abständen Schulleistungstests unterzogen, die den individuellen Lernfortschritt und gleichzeitig Unterschiede zwischen den Schülern dokumentieren sollen. In Abb. 87.3 ist diese Struktur prinzipiell am Beispiel von sechs Schülern verdeutlicht, die zu maximal fünf Zeitpunkten getestet wurden. Nicht jeder Schüler hat zu jedem Zeitpunkt einen Test absolviert. Bei der Analyse von Paneldaten im Mehrebenendesign ist es also möglich, auch Fälle mit fehlenden Werten (Engel/Schmidt, Kapitel 23 in diesem Band) zu einzelnen Messzeitpunkten einzubeziehen.



**Abb. 87.2** Mehrebenenstruktur mit Personen auf der zweiten Ebene



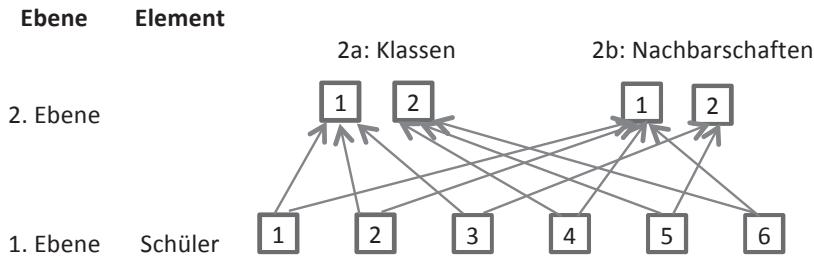
**Abb. 87.3** Mehrebenenstruktur in Paneldaten

### 87.2.2 Nicht hierarchische Datenstrukturen

Mittlerweile können auch Modelle geschätzt werden, in denen die hierarchische Struktur verletzt ist. So genannte Non-Nested Models (Gelman/Hill 2007: 244) oder Cross-Classified Models (ausführlich Hox 2010: 171ff) verwenden Daten, die in geschachtelten Ebenen vorliegen. Betrachten wir z.B. die Schulleistungen von Schülern nicht nur in Abhängigkeit der Klassen und Schulen, in denen ein Schüler lernt, sondern wollen gleichzeitig das soziale Umfeld einer Nachbarschaft mit einbeziehen, wird deutlich, dass weder die eine noch die andere aggregierte Ebene jeweils ein Element der anderen sein kann (vgl. Abb. 87.4). Geschätzt werden diese Modelle durch die Bildung künstlicher Ebenen, die die Spezifik der Kombinationen auf der zweiten Ebene aufgreifen (Rabe-Hesketh/Skrondal 2012).

### 87.2.3 Skalenniveau

Das Skalenniveau der eingehenden Variablen (Blasius/Baur, Kapitel 79 in diesem Band) kann unterschiedlich ausfallen. Hier gelten die Voraussetzungen aus einfachen Regressionsmodellen (Blasius/Baur, Kapitel 79 in diesem Band). Auch für Mehrebenenmodelle gilt, dass sich die Auswahl eines geeigneten Modells grundsätzlich am Niveau der abhängigen Variablen und an ihrer Verteilung orientiert. Liegt eine normalverteilte, metrische Variable vor, kann eine lineare Regression verwendet werden, für eine dichotome oder kategoriale abhängige Variable z.B. ein logistisches Mehrebenenmodell.



**Abb. 87.4** Mehrebenenstruktur mit sich kreuzenden Elementen auf der zweiten Ebene

#### 87.2.4 Mindeststichprobengröße

Eine vielfach diskutierte Frage ist die nach der notwendigen Fallzahl auf den einzelnen Ebenen. Hier kann in der Literatur eine Entwicklung nachgezeichnet werden, die auch den Fortschritt der Modellentwicklung für Mehrebenenmodelle dokumentiert. Ausgangspunkt ist die so genannte 30/30-Regel, die auf Kreft (1996) zurückgeht. Sie stützt sich auf Erfahrungen aus der Schulforschung und geht davon aus, dass in Modellen mit mindestens 30 Elementen der zweiten Ebene (z.B. Anzahl einbezogener Klassen) und jeweils mindestens 30 Elementen der ersten Ebene (z.B. 30 Schüler pro Klasse) unverzerrte Parameter geschätzt werden können.

In der Folge führten vor allem Simulationsstudien zu differenzierteren Ansichten in der Frage der hinreichenden Stichprobengröße. Hox (1998) und Maas und Hox (2004) zeigten in ihren Studien, dass die Anforderungen an die Fallzahlen der zweiten Ebene steigen, wenn Varianzkomponenten und Interaktionen geschätzt werden sollen. Sie schlagen als Mindestanforderung 50 Elemente der zweiten Ebene vor, wobei jedes Element 20 Elemente der ersten Ebene enthalten sollte. Für das Beispiel aus der Schulforschung würde das eine Datenbasis von mindestens 50 Klassen mit jeweils mindestens 20 Schülern bedeuten.

Bei besonderem Interesse an der Variation der Effekte in verschiedenen Kontexten sollte sich das Verhältnis auf 100/10 verändern.

Als Quintessenz aus der Entwicklung der Ansprüche an die Fallzahlen auf den einzelnen Ebenen zieht Hox (2010: 234) den Schluss, dass für eine genaue und erklärendskräftige Schätzung eine große Zahl an Elementen der zweiten Ebene (z.B. Klassen) wichtiger sei als eine große Fallzahl auf der ersten Ebene (z.B. Schüler in Klassen).

Anders argumentieren Gelman und Hill (2002). Liegen nur wenige Fälle auf der zweiten Ebene vor, so sei die Konsequenz lediglich, dass die Schätzungen der Parameter nahe an denen des Einebenenmodells sind (Gelman/Hill 2002: 275). Auf der ersten Ebene sollten sogar zwei Fälle ausreichen (Gelman/Hill 2002: 276). Die Interpretation der Ergebnisse scheint aber umso schwieriger, je weniger Fälle enthalten sind, da die Unsicherheit der Schätzung mit kleinen Fallzahlen wächst.

## 87.3 Beschreibung der Modelle

Mehrebenanalysen beinhalten in der Regel eine Reihe von Einzelmodellen, die aufeinander aufbauend entwickelt werden. Bei der Modellentwicklung hat sich ein systematisches Vorgehen bewährt, das im Kern auf Boyd und Iversen (1979) zurückgeht, von Hox (2010: 56-59) explizit aufgegriffen und von Snijders/Bosker (2012: 102-108) differenzierter, aber der gleichen Logik folgend, vorgeschlagen wird.

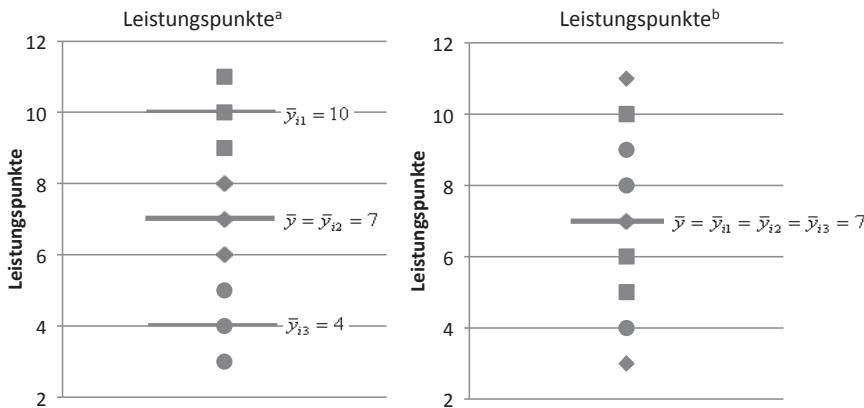
### 87.3.1 Nullmodell

An erster Stelle steht das sogenannte Nullmodell (Empty Model). Dieses Modell stellt die Varianz der abhängigen Variable ohne den Einbezug erklärender Variablen dar, wobei die Gesamtvarianz in erklärbare Varianzanteile auf den einzelnen Ebenen zerlegt wird. In Schulstudien wird hier beispielsweise beobachtet, welcher Anteil an der Streuung von Leistungspunkten auf individuelle Merkmale der einzelnen Schüler zurückgeführt werden kann und welcher Anteil darauf beruht, dass ein Schüler in einer spezifischen Klasse lernt.

Der Varianzanteil, der auf der zweiten Ebene angesiedelt ist, wird als Unterschied zwischen den Gruppenmittelwerten einbezogen und als Intraklassen-Korrelation bezeichnet. Neben der inhaltlichen Interpretation gibt der Wert der Intraklassen-Korrelation auch einen Hinweis darauf, ob die konzipierte Ebenenstruktur den empirischen Beobachtungen angemessen ist. Wenn auf der zweiten Ebene eine zu vernachlässigende erklärbare Varianz vorzufinden ist, das heißt, wenn die Unterschiede zwischen den zu erklärenden Messungen allein auf den Eigenschaften der Elemente der ersten Ebene beruhen, dann kann auf das Mehrebenenmodell unter Umständen verzichtet werden.

Abbildung 87.5 verdeutlicht das Prinzip dieser Überlegungen. In der linken Darstellung sind die Leistungspunkte von neun Schülern aus drei Klassen (jeweils gleiches Symbol) abgetragen. Es wird deutlich, dass alle Schüler sich in ihren Leistungen voneinander unterscheiden und dass auch die drei Klassen jeweils unterschiedliche mittlere Leistungen aufweisen. Die drei Schüler der ersten Klasse erreichen im Durchschnitt zehn, die Schüler der zweiten Klasse sieben und die der dritten Klasse vier Leistungspunkte. Die zweite Klasse verfügt danach über den mittleren Wert für die Leistungspunkte aller Schüler. Es gibt also neben den individuellen Unterschieden auch solche, die sich aus der Klassenzugehörigkeit ergeben. In der rechten Darstellung unterscheiden sich die individuellen Leistungspunkte ebenfalls voneinander, die Gruppenmittelwerte sind dagegen für alle Klassen gleich (bei sieben Leistungspunkten).

Neben der Varianzzerlegung besteht eine zweite wichtige Funktion des Nullmodells darin, dass es mit dem Devianzwert eine Referenz für die Beurteilung der Erklärungsleistung nachfolgender Modelle zur Verfügung stellt. Wir können den Devianzwert eines Modells als Ausdruck der Güte einer Schätzung verstehen. Im Nullmodell basiert die Schätzung lediglich auf der Verteilung der abhängigen Variablen. Werden in die nachfolgenden Modelle Erklärungsvariablen einbezogen, verändert sich der Devianzwert in



<sup>a</sup> mit erklärbare Varianz auf der zweiten Ebene; <sup>b</sup> ohne erklärbare Varianz auf der zweiten Ebene

**Abb. 87.5** Zerlegung erklärbarer Varianz auf die Ebenen

Abhängigkeit der Erklärungsleistungen der unabhängigen Variablen. Eine signifikante Veränderung im Vergleich zum Referenzwert aus dem Nullmodell bedeutet dann, dass das geschätzte Modell insgesamt signifikante Erklärungskraft hat (vgl. auch Abschnitt 3.7 Modellgüte).

### 87.3.2 Fixed-Effect-Model

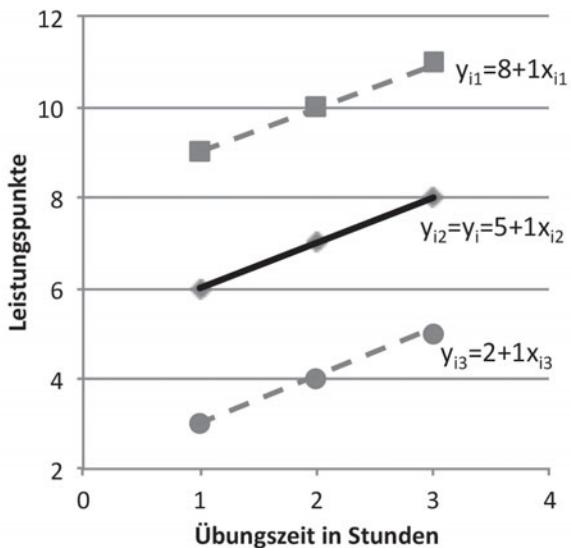
Sind die erklärbaren Varianzanteile auf den Ebenen identifiziert, dann können unabhängige Variablen in das Modell aufgenommen werden. Die Regeln und die Annahmen über diese Variablen sind die gleichen wie in den statistischen Modellen ohne Mehrebenenstruktur. Es sollen also möglichst Variablen gewählt werden, die einen hohen Erklärungsbeitrag leisten, so dass das Modell insgesamt bei guter Anpassung möglichst wenig komplex ausfällt.

Zuerst wird unterstellt, dass die Wirkung dieser Variablen in allen Aggregateinheiten gleich ist. Hierfür wird häufig der Begriff „fixed effect model“ verwendet (Hox 2010; Snijders/Bosker 2012). In Abb. 87.6 wird das Prinzip des „fixed effect models“ deutlich. Für jede Klasse wird ein separates Regressionsmodell zur Erklärung der erreichten Leistungspunkte durch die investierte Übungszeit abgetragen. Dabei unterscheiden sich die geschätzten Gruppenmittelwerte ohne Einfluss der Übungszeit zwischen den Klassen (Konstanten sind ungleich). Die Steigungskoeffizienten als Ausdruck der Wirkung investierter Übungszeit sind jedoch über die Klassen hinweg gleich stark ausgeprägt. Die Regressionsgeraden verlaufen deshalb parallel.

Im Beispiel entspricht das Modell für die zweite Gruppe dem Gesamtmodell für alle Schüler ohne Beachtung der Klassenzugehörigkeit. Die anderen beiden Klassen unterscheiden sich von dieser generellen Schätzung. Die allgemeine Beschreibung dieser Modelle

**Abb. 87.6** Prinzip des Fixed Effect Model

Leistungspunkte in Abhängigkeit der Übungszeit in drei Klassen



erfolgt durch die Angabe des mittleren geschätzten Gruppenmittelwertes (Konstante) und eines Variationsterms für die Konstante (der dann die Unterschiedlichkeit der Gruppen im Ausgangsniveau beinhaltet) sowie des Regressionsparameters, der für alle Gruppen gleich ist. Nach der Modellprüfung verbleiben die signifikanten Effekte im Modell und es wird um Erklärungsvariablen der zweiten Ebene ergänzt. Das könnte im Beispiel die Klassengröße sein.

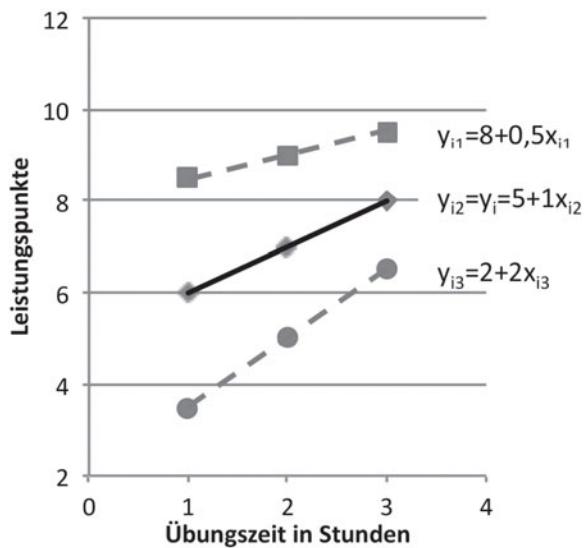
### 87.3.3 Random Coefficient Models

Der nächste Schritt der Modellentwicklung besteht darin, die Annahme der gleichen Wirkung der unabhängigen Variablen in den Aggregateinheiten aufzugeben. Es handelt sich in diesen Fällen um so genannte „random coefficient models“. Abbildung 87.7 zeigt für drei Klassen Regressionen der individuellen Leistungspunkte auf die eingesetzte Übungszeit. Dabei wird nicht nur das unterschiedliche Ausgangsniveau der erreichten Leistungspunkte der Klassen deutlich, sondern auch die differierenden Effekte der eingesetzten Übungszeit.

Beschrieben wird das Modell nun durch die Angabe mittlerer Schätzungen für die Konstanten und die Regressionskoeffizienten jeder Gruppe, die jeweils um eine Varianzkomponente ergänzt werden. Die Varianzkomponente bringt hier die Streuung des Effekts zwischen unterschiedlichen Aggregateinheiten zum Ausdruck. Eine kleine Streuung bedeutet,

**Abb. 87.7** Prinzip des Random Coefficient Model

### Leistungspunkte in Abhängigkeit der Übungszeit in drei Klassen



dass die Effekte in den Gruppen sehr ähnlich sind, die eingesetzte individuelle Übungszeit also in allen Klassen ähnlich starke Wirkung auf die Leistungspunkte entfaltet. Eine große Varianz heißt, dass der Effekt der Einflussvariablen Übungszeit in einigen Klassen deutlich höher ausfällt als in anderen.

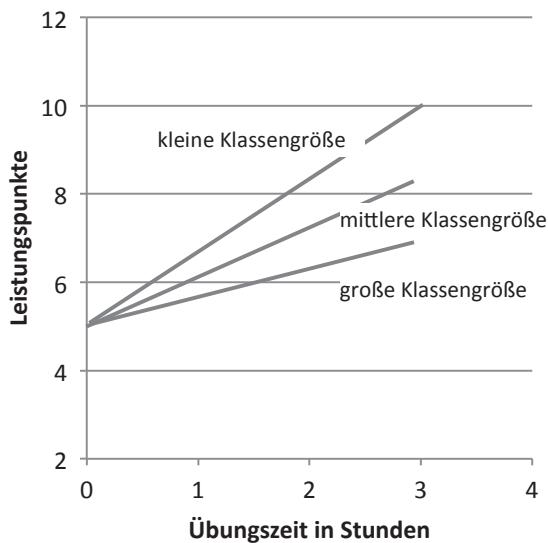
Neben den Varianzkomponenten als Ausdruck der Unterschiedlichkeit in den Regressionsmodellen für die einzelnen Elemente der zweiten Ebene ist der mögliche systematische Zusammenhang zwischen Ausgangsniveau (Konstante) und Wirkmechanismus (Regressionskoeffizient) von besonderer Bedeutung. In unserem Abbildungsbeispiel beobachten wir eine negative Korrelation zwischen dem Ausgangsniveau der Leistungen und dem Effekt der Übungszeit. Je besser die Ausgangsleistungen sind, desto weniger stark wirkt sich die Übungszeit auf die Leistungsverbesserung aus. Ist das Ausgangsniveau gering, so erbringt eine Investition durch Übungszeit deutlich mehr an Zuwachs in den Leistungspunkten als es bei einem hohen Ausgangsniveau der Fall ist.

#### 87.3.4 Cross-Level-Interaction

Im letzten Schritt können in die Modelle Wechselwirkungen (Interaktionseffekte) über die Ebenen mit einbezogen werden (cross-level interaction). Zur Veranschaulichung werden Interaktionseffekte über die Ebenen hinweg häufig im Sinne der Moderation eines Effekts auf der Individualebene durch die spezifische Ausprägung in der zweiten Ebene

**Abb. 87.8** Prinzip des Cross-Level-Interaction-Effekts

Leistungspunkte in Abhängigkeit der Übungszeit für drei Klassengrößen



interpretiert (Hox 2010: 66). Beispielsweise kann die mittlere Wirkung der investierten Übungszeit davon abhängen, wie groß eine Schulklasse ist.

Abbildung 87.8 zeigt unterschiedliche Regressionsmodelle für kleine, mittlere und großen Klassen. Ausgehend von einem einheitlichen Leistungsniveau (die Konstante ist für alle Klassengrößen gleich) wirkt sich die investierte Übungszeit der einzelnen Schüler in den kleinen Klassen stärker auf die erreichten Leistungspunkte aus als in den mittleren und großen Klassen. Die Wirkung der individuellen Übungszeit fällt also unterschiedlich stark aus, je nachdem wie groß die Klasse ist, in der der Schüler lernt.

### 87.3.5 Zentrierung von Variablenwerten

Wenn nicht der Variablenwert einer Person selbst von Interesse ist, sondern die relative Position der Person in einer Verteilung, kann eine Zentrierung der Werte in Frage kommen. Beispielsweise können sich zwei Schüler in ihren Leistungspunkten gleichen, wobei der eine in einer insgesamt sehr starken Gruppe vergleichsweise durchschnittlich zu bewerten ist, während der andere Schüler in einer eher schwachen Gruppe zu den Besten zählt. Aus dieser relativen Position erwachsen dann möglicherweise unterschiedliche Selbstwahrnehmungen, die wiederum den Lernerfolg beeinflussen. Dieser Sachverhalt wird als sogenannter „frog pond effect“ bezeichnet. Er drückt aus, dass die Bedeutung einer Merkmalsausprägung davon abhängt, in welcher relativen Position sie zu allen

anderen Merkmalsausprägungen in der Gruppe steht. Die Zentrierung erfolgt hier am jeweiligen Gruppenmittelwert.

Die individuelle Ausprägung einer Variablen kann darüber hinaus aber auch am generellen Mittelwert (über alle Aggregateinheiten hinweg) zentriert werden. In diesem Fall wird die Konstante als die Schätzung interpretiert, die für eine Person mit mittlerer Ausprägung in der unabhängigen Variable abgegeben werden kann. Hilfreich ist eine solche Zentrierung am Mittelwert, wenn die Merkmalsausprägung Null in der unabhängigen Variablen keine inhaltliche Bedeutung aufweist. Insbesondere für die Interpretationen von Wechselwirkungseffekten kann die Zentrierung hilfreich sein (ausführlich, Hox 2010: 63ff und Snijders/Bosker 2012: 81ff.).

### 87.3.6 Modellannahmen

Prinzipiell finden sich in Mehrebenenmodellen die Modellannahmen der Einfachregression wieder. So wird unterstellt, dass der Erwartungswert der Residuen bei gegebenen unabhängigen Variablen und einem zufälligen Intercept gleich Null ist, dass die Residuen auf der ersten Ebene normalverteilt sind und ihre Varianz homoskedastisch ist, dass die Residuen unabhängig voneinander und von den unabhängigen Variablen sind (Langer 2004: 102; Hesketh/Skrondal 2010: 128-129).

### 87.3.7 Modellgüte

Zur Interpretation der Modellgüte stehen unterschiedliche Möglichkeiten zur Verfügung. Im einfachsten Fall kann ein Devianztest Auskunft darüber geben, inwieweit die einbezogenen Variablen zur Erklärung der Varianz der abhängigen Variable beigetragen haben. Dazu wird die Signifikanz der Differenz aus dem Devianzwert (-2loglikelihood) des Nullmodells und dem Devianzwert des Erklärungsmodells betrachtet. Die Differenz ist Chi-Quadrat-verteilt, die Freiheitsgrade richten sich nach der Anzahl der geschätzten Parameter. Im Nullmodell werden drei Parameter geschätzt (Intercept und seine Varianzen auf erster und zweiter Ebene). Zum Beispiel wird für die Streuung individueller Leistungspunkte von Schülern eine generelle mittlere Schätzung (Intercept) vorgenommen und um die Angaben zur erklärbaren Varianz durch die Schüler und der erklärbaren Varianz durch die Klassen ergänzt. Soll auf der Basis dieses Nullmodells nun die Übungszeit als erklärende Variable eingeführt werden, die in allen Klassen gleich wirkt (fixed effect model), dann sind für das Erklärungsmodell vier Parameter zu schätzen. Die Differenz der Freiheitsgrade beträgt  $4-3=1$ . Eine Verbesserung des Devianzwertes ab 3,84 würde für eine signifikante Modelleistung sprechen.

Bryk und Raudenbush (1992) schlagen für die Beurteilung der Modellgüte ein analoges Vorgehen zur linearen Einfachregression vor (auch Snijders/Bosker 2012:102-114). Sie definieren ein  $R^2$  als Verhältnis zwischen der Differenz aus den Varianzen des Null- und

des Erklärungsmodells und der Varianz des Nullmodells. Dieses Vorgehen kann auf das Modell global oder auf die einzelnen Ebenen bezogen werden, so dass deutlich wird, um welchen Anteil sich die Erklärungskraft durch den Einbezug von Variablen auf den einzelnen Ebenen erhöht. Die Nachteile dieses Ansatzes bestehen in unplausiblen negativen Werten für den Anteil erklärter Varianz, z.B. wenn die Gruppengrößen stark variieren, und darin, dass die Varianzen von der Skalierung der zwischen den Gruppen variierenden erklärenden Variablen abhängen (Hox 2010: 69ff.; Snijders/Bosker 2012: 109ff.; Langer 2004: 152).

Für das letzte Problem gibt es bisher keine Lösung (Hox 2010: 77). Das erste Problem wurde auf verschiedenen Wegen zu lösen versucht. Snijders und Bosker (2012: 111ff.) schlagen eine Korrektur im Rahmen der oben beschriebenen Strategie vor, indem entweder der unterschiedlichen Gruppengröße Rechnung getragen wird, oder der Bezug auf die Gesamtvarianz hergestellt wird, auch wenn Teilvarianzen der Ebenen erklärt werden sollen.

Ein weiterer Vorschlag findet sich bei Langer (2004: 161). Mit dem dort präsentierten Maddala R<sup>2</sup> wird auf den Devianzwert zurückgegriffen, so dass die Anfälligkeiten der R<sup>2</sup>-Maße vermieden werden. Allerdings kann damit nur auf die globale Erklärungskraft des Modells Bezug genommen werden.

Für die Interpretation der Signifikanz der Effekte wird, wie in der linearen Einfachregression, auf den t-Test zurückgegriffen. Die Testung der Varianzen erfolgt auf der Basis des Wald-Tests. Hox (2010: 47ff) argumentiert auf der Basis weiterer Quellen, dass zur Beurteilung der Signifikanz der Varianzen auch auf den Devianztest zurückgegriffen werden sollte.

---

## 87.4 Software

Mittlerweile gibt es eine Vielzahl von Programmen, mit denen Mehrebenenmodelle geschätzt werden können. Mit den beiden Programmen MLwin und HLM, die speziell für Mehrebenenprobleme entwickelt wurden, stehen zwei Tools zur Verfügung, die ausgeweitete Analysen erlauben. Durch die explizite Entwicklung dieser Tools für solche Modelle sind vielfältige Modellspezifikationen und die Verwendung verschiedener Schätzalgorithmen möglich. Ein kleiner Nachteil der Spezialsoftware besteht im Datenhandling, das den Komfort großer Programmpakete zum Teil nicht erreicht.

Zu den großen Programmpaketen, die Mehrebenenmodelle implementiert haben, gehören unter anderem SPSS, STATA, MPLUS und SAS. Für das kostenlose Programm R steht mit lme4 ein umfangreiches und leistungsstarkes Paket zur Verfügung.

## Literatur

- Boyd, Lawrence H./Iversen, Gudmund R. (1979): Contextual Analysis. Belmont: Wadsworth Publishing Company
- Bryk, Anthony/Raudenbush, Stephen W. (1992): Hierarchical Linear Models. Newbury Park/London/New Delhi: Sage
- Coleman, James S./Campbell, Ernest Q./Hobson, Carol J./McPartland, James/Mood, Alexander M./Weinfeld, Frederic D./York, Robert L. (1966): Equality of educational opportunity. Washington (D.C.): U.S. Government Printing Office
- Coleman, James S./Hoffer, Thomas/Kilgore, Sally (1982): Cognitive Outcome in Public and Private Schools. In: Sociology of Education 55 (2-3): 526-545
- Engel, Uwe (1998): Einführung in die Mehrebenenanalyse. Opladen/Wiesbaden: Westdeutscher Verlag
- Gelman, Andrew/Hill, Jennifer (2007): Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge et al.: Cambridge University Press
- Hox, Joop J. (1998): Multilevel Modeling: When and Why. In: Balderjahn, Ingo/Mathar, Rudolf/Schader, Martin (Hg.): Classification, Data Analysis, and Data Highways. New York: Springer. 147-154.
- Hox, Joop J. (2010): Multilevel Analysis. New York/Hove: Routledge
- Kreft, Ita (1996): Are Multilevel Techniques Necessary? Los Angeles: California State University
- Langer, Wolfgang (2004): Mehrebenenanalyse. Wiesbaden: VS Verlag
- Lazarsfeld, Paul F./Menzel, Herbert (1961): On the Relation between Individuals and Collective Properties. In: Etzioni, Amitai (Hg.): A Sociological Reader on Complex Organizations. London: Holt, Rinehart and Winston. 422-440.
- Maas, Cora J.M./Hox, Joop J. (2005): Sufficient Sample Size for Multilevel Modeling. In: Methodology 1/05: 86-92
- Neumann, Marko/Milek, Anne/Maaz, Kai/Gresch, Cornelia (2010): Zum Einfluss der Klassenzusammensetzung auf den Übergang von der Grundschule in die weiterführenden Schulen. In: BMBF (Hg.): Der Übergang von der Grundschule in die weiterführende Schule. Bonn/Berlin: BMBF. 229-252
- Prenzel, Manfred/Baumert, Jürgen (Hg.) (2008): Vertiefende Analysen zu PISA 2006. Sonderheft der Zeitschrift für Erziehungswissenschaften 10/2008
- Rabe-Hesketh, Sophia/Skrondal, Anders (2012): Multilevel and Longitudinal Modeling Using STATA. College Station: STATA Press
- Snijders, Tom/Bosker, Roel (2012): Multilevel Analysis. Newbury Park/London/New Delhi: SAGE

Michael Wagner und Bernd Weiß

---

## 88.1 Meta-Analyse: Begriff und Ziele

Der Begriff „Meta-Analyse“ wurde im Jahr 1976 von Gene V. Glass eingeführt. Er verstand darunter die „analysis of analyses“ (Glass 1976), die neben Primär- und Sekundäranalysen einen dritten Forschungstyp darstellt. Im Gegensatz zu Primär- und Sekundäranalysen werden hier nicht die Originaldaten erstmalig oder wiederholt ausgewertet. Vielmehr richtet sich die Meta-Analyse in der von Glass vorgeschlagenen Form auf die *quantitative Integration publizierter empirischer Befunde aus möglichst allen für die jeweilige Fragestellung relevanten Studien*. Dabei geht es nicht nur darum, empirische Befunde zusammenzufassen, sondern auch deren Unterschiedlichkeit zu beschreiben und zu erklären. In der Regel handelt es sich um in Fachartikeln veröffentlichte Befunde. Empfohlen wird aber, auch „graue“ Literatur, beispielsweise Konferenzbeiträge, Arbeitspapiere, Dissertationen oder Examensarbeiten in eine Meta-Analyse einzubeziehen.

Während die von Glass propagierte „klassische“ Meta-Analyse auch *Aggregate Person Data-Meta-Analysis (APD)* genannt wird, wird ein zweiter Typ von Meta-Analyse, die *Individual Person Data-Meta-Analysis (IPD)* (Stewart/Tierney 2002) immer populärer. Die IPD-Meta-Analysen zielen nicht auf eine Synthese publizierter empirischer Befunde, sondern auf die *Reanalyse der Originaldaten*, die den jeweiligen Studien zugrundeliegen. Diese Datensätze können für sich je einzeln aber in einheitlicher Weise oder in zusammengefasster („pooled“) Form analysiert werden. IPD-Meta-Analysen sind aufwendig und daher heute noch selten, haben aber mehrere Vorteile: die Vergleichbarkeit der Variablen kann erhöht werden, das Spektrum möglicher Analyseverfahren ist größer, es lassen sich Hypothesen untersuchen, die bislang nicht untersucht wurden.

Meta-Analysen sind eine bessere Alternative zu so genannten *narrativen (qualitativen) Reviews*, die immer noch häufig zur Beschreibung eines aktuellen empirischen Forschungsstands eingesetzt werden. Denn bei narrativen Reviews ist das Vorgehen oft wenig standardisiert und damit kaum nachprüfbar. Beispielsweise sind Literaturrecherchen häufig

unvollständig, vorhandene Informationen werden nicht voll genutzt, die Gewichtung einzelner Studien wird nach subjektiven Kriterien vorgenommen und widersprüchliche Forschungsergebnisse können nur schlecht zusammengefasst werden. Ferner kann mit Hilfe narrativer Reviews eine große Anzahl von Forschungsbefunden nicht mehr angemessen bearbeitet und interpretiert werden.

APD- sowie IPD-Meta-Analysen haben beide zum Ziel, den Forschungsstand zu einer empirischen Fragestellung zu bestimmen und zu genaueren und sichereren Schlussfolgerungen zu gelangen, als es mit einer oder nur wenigen Studien oder mit einem einzelnen Datensatz möglich wäre. Solche systematischen Reviews werden immer dringlicher, weil die Menge der Forschungsbefunde in den letzten Jahrzehnten in allen wissenschaftlichen Feldern explosionsartig gestiegen ist (Petticrew/Roberts 2006: 7; Weiß/Wagner 2008).

Meta-Analysen sind auch deshalb geboten, weil empirische Ergebnisse zu demselben Variablenzusammenhang, die aber aus verschiedenen Studien stammen, sehr heterogen und inkonsistent sein können. Das mag an differierenden Operationalisierungen (Stein, Kapitel 7 in diesem Band) oder dem Stichprobendesign (Häder/Häder, Kapitel 18 in diesem Band) liegen oder daran, dass Effektstärken von in der Originalanalyse statistisch nicht berücksichtigten Faktoren abhängen können. Mit Hilfe von Meta-Analysen lässt sich nicht nur bestimmen, *wie groß die Heterogenität der Befunde ist*. Vielmehr können auch diejenigen *Faktoren* identifiziert werden, *die für diese Heterogenität der Befunde verantwortlich sind*.

Meta-Analysen tragen dazu bei, drei bedeutende *wissenschaftliche Ziele* besser zu erreichen:

1. Empirische Forschung soll sich kumulativ entwickeln, Forschung soll Lücken füllen, die andere zurückgelassen haben. Die Erklärungskraft von Theorien und ihr praktischer Nutzen nehmen zu, wenn wir wissen, welche Hypothesen widerlegt wurden und welche nicht.
2. Die praktische Brauchbarkeit von Maßnahmen und Interventionen kann nur dann optimiert werden, wenn zuverlässiges Wissen um ihre Effizienz vorhanden ist. Die Aufgabe von systematischen Reviews ist es, Forschungsergebnisse zu synthetisieren, um die besten Methoden herauszufinden, mit denen bestimmte soziale oder sozialwissenschaftliche Probleme im Praxis- und Anwendungsbereich gelöst werden können.
3. Auf vielen Gebieten besteht ein zentrales Forschungsziel darin, den Informationsgehalt von Theorien durch eine Suche nach universalen Gesetzen zu verbessern. Lassen sich Erkenntnisse zu empirischen Fragestellungen mit mehreren unabhängigen Stichproben replizieren, dann erhöht sich der Grad an Generalisierbarkeit. So lassen sich Populationsparameter präziser bestimmen, wenn sie auf der systematischen Kombination von Ergebnissen mehrerer empirischer Studien basieren (Allen 2009: 399).

Meta-Analysen sind *zentraler Bestandteil quantitativer systematischer Reviews*, die mittlerweile in vielen wissenschaftlichen Disziplinen fest etabliert sind. Dies zeigen nicht zuletzt die zahlreichen Lehr- und Handbücher zur Theorie und praktischen Durchführung von Meta-Analysen (Borenstein et al. 2009; Cooper et al. 2009; Lipsey/Wilson 2001). In

der Medizin und Epidemiologie hat sich die Meta-Analyse bereits in Form der *evidenzbasierten Medizin* mit dem im Jahr 1993 gegründeten UK Cochrane Centre institutionalisiert (Petticrew/Roberts 2006: 20). Die Cochrane Collaboration ist eine internationale Non-Profit-Organisation, die im Feld der medizinischen Versorgung systematische Reviews evaluiert und veröffentlicht. Eine ähnliche Organisation wie die Cochrane Collaboration ist die Campbell Collaboration, die im Jahr 1999 gegründet wurde. Es handelt sich dabei um ein Forschungsnetzwerk, das systematische Reviews über die Effekte sozialer Interventionen erstellt und evaluiert.

---

## 88.2 Typische Forschungsprobleme für Meta-Analysen in den Sozialwissenschaften

Im Folgenden wollen wir anhand von Beispielen darstellen, welcher Typ von Forschungsproblemen mittels der Meta-Analyse angegangen werden kann. Drei prototypische Forschungsprobleme können unterschieden werden. Sie richten sich auf die Deskription und Exploration sozialer Sachverhalte, den Test von Hypothesen sowie auf die Evaluation von Maßnahmen.

### 88.2.1 Deskription und Exploration sozialer Sachverhalte

In den Sozialwissenschaften kommt es häufig vor, dass man die Verbreitung eines sozialen Phänomens in der Gesellschaft nicht genau kennt, weil die amtliche Statistik darüber keine Auskunft gibt oder keine entsprechend große Primäruntersuchung vorliegt. In diesem Fall liegt also eine unvollständige Deskription eines sozialen Tatbestands vor. So ist beispielsweise unbekannt, wie viele Jugendliche in Deutschland die Schule schwänzen. Allerdings gibt es in Deutschland zahlreiche lokal begrenzte Schüler- und Jugendstudien, in denen das Schulschwänzen erhoben wird. Weiß (2008) hat eine individualdatenbasierte Meta-Analyse dieser Studien vorgelegt und eine Hochrechnung zur prozentualen Verbreitung des Schulschwänzens vorgenommen. Mit Hilfe dieser Meta-Analyse konnte hochgerechnet werden, dass der Anteil der Jugendlichen, die überhaupt schon einmal geschwänzt haben, bei 17% liegt, mit einem 95%-Konfidenzintervall (95% CI) von [10%; 27%]; für das häufige Schulschwänzen ergibt sich ein deutlich niedrigerer Anteilswert von knapp 2% (95% CI [0%; 5%]).

### 88.2.2 Test von Hypothesen

Um herauszufinden, ob und warum Scheidungsrisiken in Europa variieren, führten Wagner/Weiß (2006) eine Meta-Analyse auf der Basis von 120 empirischen Studien aus 19 europäischen Ländern durch. Getestet wurden Hypothesen zum Zusammenhang von

Makrovariablen und individuellen Scheidungsrisiken. Es konnte zunächst nachgewiesen werden, dass es Faktoren gibt, die in allen Ländern signifikant mit dem Scheidungsrisiko assoziiert sind. Ferner wurde belegt, dass die Übertragung des Scheidungsrisikos zwischen der Eltern- und der Kindergeneration umso geringer ist, je niedriger die Scheidungsbarrieren in einem Land sind. Zudem verdeutlicht diese Untersuchung, dass Meta-Analysen in der international vergleichenden Forschung (Braun, Kapitel 56 in diesem Band) gewinnbringend eingesetzt werden können.

### **88.2.3 Evaluation von Maßnahmen**

Meta-Analysen werden auch oft zur Evaluation von Maßnahmen eingesetzt, sei es im therapeutisch-medizinischen Bereich, sei es in der Sozialpolitik. So befasst sich eine Meta-Analyse von Cebulla/Greenberg (2006) mit US-amerikanischen „welfare-to-work“-Programmen. Ein Ergebnis dieser Meta-Analyse ist, dass Programme, die sich auf die Arbeitsplatzsuche richten, besonders effektiv sind, wenn es um eine Erhöhung des Einkommens oder den Erwerbsstatus geht.

In der Präventionsforschung zum abweichenden Verhalten Jugendlicher publizierten Wilson et al. (2001) eine Meta-Analyse zur Frage, welche Interventionen von Schulen das abweichende Verhalten von Jugendlichen reduzieren können. Ein Ergebnis der Meta-Analyse ist, dass Interventionen zur Erhöhung der Selbstkontrolle oder der sozialen Kompetenz besonders erfolgreich sind.

---

## **88.3 Durchführung von Meta-Analysen**

Eine Meta-Analyse umfasst alle Elemente eines sozialwissenschaftlichen Forschungsprozesses (Stein, Kapitel 7 in diesem Band): Problemanalyse und Hypothesenbildung, Datenerhebung, Datenaufbereitung, Datenanalyse, Interpretation der Befunde.

Bei *APD-Meta-Analysen* besteht die *Analyseeinheit* meistens aus bivariaten Assoziationsmaßen, beispielsweise Korrelationskoeffizienten, Mittelwertdifferenzen, Odds Ratios, aber auch Regressionskoeffizienten, univariate Anteils- oder Mittelwerte können zum Gegenstand einer Meta-Analyse werden. Bei *IPD-Meta-Analysen* bilden die einzelnen Messwerte, die in den verschiedenen Datensätzen enthalten sind, die Analyseeinheit.

Der erste Schritt, die *Problemanalyse* und gegebenenfalls die *Hypothesenbildung*, verläuft so wie bei anderen empirischen Studien auch. Allerdings setzt eine Meta-Analyse voraus, dass zu einer Forschungsfrage schon publizierte Befunde vorliegen. Bei einer Meta-Analyse, die auf Individualdaten basiert, ist es hingegen auch möglich, auf der Basis vorhandener Datensätze neue Fragestellungen zu untersuchen.

Möglichst vollständige und systematische *Literaturrecherchen* sind eine wichtige Voraussetzung jeder publikationsbasierten APD-Meta-Analyse (White 2009). Meta-Analysen, die graue Literatur ausschließen, berichten stärkere und häufiger signifikante Effekte als

Meta-Analysen, die auch graue Literatur berücksichtigen. So belegen Meta-Analysen aus der Medizin (McAuley et al. 2000; Conn et al. 2003), dass der Behandlungserfolg überschätzt wird, wenn graue Literatur nicht berücksichtigt wird. Der Grund hierfür könnte ein Publikationsbias sein, der darin besteht, dass beispielsweise Studien, die signifikante Befunde berichten, mit größerer Wahrscheinlichkeit publiziert werden als Studien, die keine signifikanten Befunde ermitteln. Bei *IPD-Analysen* ist nicht so sehr die Voll-erhebung der Literatur relevant, sondern die möglichst komplettte *Sammlung relevanter Originaldatensätze*.

Es lassen sich insgesamt drei Quellen ausmachen, um an relevante Literatur zu gelangen: (a) Experten, (b) Datenbanken sowie das Durchsuchen der durch (a) und (b) gefundenen Literatur nach weiteren Studien („Schneeballsystem“). Die materialreichsten Quellen stellen die Fachdatenbanken dar (zum Beispiel Social Science Citation Index, SocINDEX, Sociological Abstracts, SOFIS, SOLIS).

Ähnlich einer Inhaltsanalyse (Mayring/Fenzl, Kapitel 38 in diesem Band) wird nach der Literaturrecherche das gefundene *Material vercodet*. Relevant sind vor allem die eigentlichen Befundstatistiken (etwa die Effektstärken und deren Standardfehler). Um die Heterogenität der Befunde aufklären zu können, sollten aber auch Publikationsmerkmale (Publikationstyp, -jahr etc.), Studien- bzw. Stichprobenmerkmale (Erhebungsjahr, Stichprobendesign etc.) oder auch Qualitätsmerkmale berücksichtigt werden (Lipsey 2009).

Die *Datenanalyse* besteht aus drei Teilen:

1. *Befundintegration*: Bei der Befundintegration geht es darum, die einzelnen Befundstatistiken zu einer Gesamtstatistik zusammenzufassen.
2. *Heterogenitätsanalyse*: Bei der Heterogenitätsanalyse wird die Variation der Befunde (Effektstärken) analysiert und gegebenenfalls erklärt.
3. *Analyse fehlender Werte*: Bei der Analyse fehlender Werte (Engel/Schmidt, Kapitel 23 in diesem Band) wird unter anderem überprüft, ob ein Publikationsbias vorliegt. Ein Publikationsbias liegt dann vor, wenn die Vorbereitung, Einreichung oder Publikation von Forschungsbefunden von Merkmalen dieser Befunde – wie etwa der statistischen Signifikanz oder der Richtung der Effekte – abhängt (Weiß/Wagner 2011).

---

## 88.4 Verfahren der Befundintegration

Je nachdem, wie informationshaltig die Darstellung der Forschungsbefunde ist, existieren mehrere Verfahren der Befundintegration (Borenstein et al. 2009; Cooper et al. 2009).

Liegen lediglich Angaben dazu vor, ob die Ergebnisse statistisch signifikant sind, kann das so genannte *Vote Counting* verwendet werden. Hier werden die Effekte danach geordnet, ob sie signifikant positiv, signifikant negativ oder nicht signifikant sind. Die Integration der Befunde besteht dann darin, den Modalwert dieser Verteilung zu ermitteln (Bushman/Wang 2009). Ein weiteres Verfahren ist die *Zusammenfassung von Irrtumswahrscheinlichkeiten* (Borenstein et al. 2009).

Sofern sowohl Informationen über die Befundstatistiken als auch deren Standardfehler vorliegen, besteht das am häufigsten eingesetzte meta-analytische Verfahren zur Befundintegration darin, statistisch unabhängige Befunde mit einem *gewichteten arithmetischen Mittel* zusammenzufassen. Bei der Berechnung der Gewichte wird der (inverse, quadrierte) Stichprobenfehler berücksichtigt und Befundstatistiken, die auf großen Fallzahlen basieren, erhalten einen entsprechend großen Einfluss bei der Parameterschätzung.

Die Befunde werden bereits auf Grund des Stichprobenfehlers eine gewisse Variation aufweisen. Eine zentrale Frage ist, ob diese Befundstreuung nur auf den Stichprobenfehler zurückzuführen ist (Modell 1), oder ob die Streuung der Effektstärken das allein durch den Stichprobenfehler erwartete Niveau übersteigt und sich diese zusätzliche Heterogenität auf weitere (möglicherweise unbekannte) Einflussgrößen zurückführen lässt (Modell 2). Die Beantwortung dieser Frage hat sowohl Konsequenzen für die Parameterschätzung als auch die statistische Inferenz.

Das hier als Modell 1 eingeführte Verfahren ist in der Literatur als „*Fixed-Effect Model*“ (FEM) bekannt. Hier wird angenommen, dass sämtlich beobachteten Effektstärken *einen gemeinsamen* Effektstärkenparameter (den „*Fixed Effect*“) repräsentieren, da sich die beobachteten Effektstärken nur durch den Stichprobenfehler unterscheiden. Wenn aber die Streuung der Effektstärken das allein durch den Stichprobenfehler erwartete Niveau übersteigt, dann muss angenommen werden, dass es einen zusätzlichen systematischen Einfluss von bekannten wie unbekannten Größen gibt.

Die Entscheidung, inwieweit eine zusätzliche Varianzkomponente bei der Berechnung der mittleren Effektstärke berücksichtigt werden sollte, kann einerseits von den oben dargestellten Homogenitätstests abhängen, andererseits aber auch methodische oder inhaltliche Gründe haben (Borenstein et al. 2009: 105-125). Solche Gründe liegen beispielsweise vor, wenn vermutet wird, dass die Literaturrecherche nicht vollständig war oder die Befunde aus Studien mit einem unterschiedlichen Stichprobendesign stammen. Ein weiteres Argument betrifft den Fall, dass Regressionskoeffizienten aus multiplen Regressionsmodellen zusammengefasst werden sollen, die Modelle in den Studien jedoch unterschiedlich spezifiziert wurden. Es ist dann nicht angebracht, von einem gemeinsamen Effektstärkeparameter auszugehen. Wenn mindestens eine der beiden genannten Bedingungen zutrifft, sollte mit Hilfe des sogenannten „*Random-Effects Model*“ (REM) (oben als Modell 2 eingeführt) diese zusätzliche Befundvariation sowohl im Rahmen der statistischen Inferenz als auch bei der Schätzung einer mittleren Befundstatistik berücksichtigt werden. In die Konstruktion der Gewichte fließt nun neben dem Stichprobenfehler auch die sogenannte Zwischenstudienvarianz ein, welche die über den Stichprobenfehler hinausgehende Heterogenität der Effektstärken beschreibt (Hedges/Olkin 1985).

Mit dem REM wird unterstellt, dass jede beobachtete Effektstärke eine eigene Verteilung mit einem eigenen (Lage-)Parameter hat. Mit dem REM wird nicht mehr eine gemeinsame, sondern eine *mittlere* Effektstärke geschätzt. Es wird vorausgesetzt, dass die beobachtete Effektstärkenverteilung eine zufällige Auswahl aus der Menge aller Studien mit allen möglichen Kombinationen aller möglichen Einflussgrößen darstellt. Durch die Berücksichtigung der Zwischenstudienvarianz wird die statistische Inferenz dergestalt

beeinflusst, dass die geschätzte mittlere Effektstärke einen größeren Standardfehler und ein größeres Konfidenzintervall hat als der entsprechende FEM-Schätzer. Darüber hinaus ändert sich der relative Einfluss der Gewichte im REM. Studien mit großen FEM-Gewichten (also großen Fallzahlen) verlieren bei der Verwendung des REM an Bedeutung, während Studien mit kleineren Gewichten an Bedeutung gewinnen. Damit unterscheiden sich auch die jeweiligen Parameterschätzer zwischen beiden Modellen.

Die bisherigen Ausführungen haben das Vorgehen für statistisch unabhängige Effektstärken beschrieben. Sofern die Effektstärken aber nicht voneinander unabhängig sind – etwa wenn sich mehrere Effektstärken auf denselben Datensatz beziehen – gibt es unterschiedlich anspruchsvolle Verfahren. Dazu gehören die einfache Befundsynthese innerhalb identischer Datensätze sowie die Schätzung robuster Standardfehler (Hedges et al. 2010) und Mehrebenen-Meta-Analysen (Van den Noortgate et al. 2013).

---

## 88.5 Verfahren der Heterogenitätsaufklärung

Neben der Befundsynthese ist im Rahmen einer Meta-Analyse die Heterogenitätsaufklärung sehr bedeutsam. Werden kategoriale Merkmale zur Aufklärung der Heterogenität herangezogen, kommen *varianzanalytische Modelle* zum Einsatz. Im Fall metrischer Merkmale können *Meta-Regressionen* geschätzt werden (Borenstein et al. 2009). Beispielsweise können sich Effektstärken danach unterscheiden, ob sie bivariate Zusammenhänge widerspiegeln oder ob sie im Rahmen von multiplen Regressionsmodellen geschätzt wurden. Die Effektstärken lassen sich also in zwei Subgruppen unterscheiden und es lässt sich statistisch testen, ob die mittleren Effektstärken der beiden Subgruppen signifikant differieren. Analog lässt sich überprüfen, ob die Effektstärken vom Jahr der Datenerhebung, vom Stichprobendesign oder vom geografischen Kontext abhängen.

Die Validität einer publikationsbasierten Meta-Analyse ist gefährdet, wenn die Publikationschancen unter anderem von den Befunden selbst abhängen, also ein Publikationsbias vorliegt. In vielen Disziplinen zeigt sich, dass Studien mit signifikanten Befunden eine größere Publikationswahrscheinlichkeit haben. Dahinter verbirgt sich die irrtümliche Vorstellung, dass statistische Signifikanz (Häder/Häder, Krebs/Menold, Kapitel 18 und 30 in diesem Band) die Gültigkeit einer Theorie „beweist“. Es gibt inzwischen eine Reihe von Techniken, mit denen sich feststellen lässt, ob bei einer Meta-Analyse verzerrte Daten vorliegen (vgl. Rothstein et al. 2005, Weiß/Wagner 2011).

---

## 88.6 Probleme und offene Fragen

Gegen die Meta-Analyse können mehrere Argumente ins Feld geführt werden. Häufig wird behauptet, man vergleiche „Äpfel mit Birnen“, Meta-Analysen würden Unvergleichbares vergleichen. Diese Heterogenität kann sich auf die abhängigen und unabhängigen Variablen oder auf die Stichprobe (z.B. unterschiedliche Untersuchungseinheiten oder

geographische Räume) beziehen. Es werden zwei Lösungen für das Heterogenitätsproblem vorgeschlagen. Die theoretische Lösung richtet sich auf eine angemessene Generalisierung der Kategorien (statt „Äpfel“ und „Birnen“ wird „Obst“ zur Bezugskategorie). Die methodische Lösung besteht darin, im Rahmen regressionsanalytischer Verfahren zu überprüfen, ob unterschiedliche Kategorisierungen zu signifikant unterschiedlichen Effektstärken führen.

Eine weitere Kritik richtet sich auf die *Berücksichtigung qualitativ „schlechter“ Studien („garbage in, garbage out“)*. Es wird die Auffassung vertreten, dass bei einer Kombination methodisch „guter“ und „schlechter“ Studien Meta-Analysen irreführende Ergebnisse erbringen. Gewöhnlich wird diesem Problem durch Rating-Skalen begegnet, welche die Qualität der Studien bewerten. Die Skalenwerte werden als Gewichtungsfaktoren in der Meta-Analyse verwendet (Glass et al. 1981: 222f.).

Kritisiert werden auch *publikationsbasierte Meta-Analysen, die auf Beobachtungsstudien basieren*. Besonders verbreitet sind Meta-Analysen in solchen Disziplinen, in denen viele, vor allem experimentelle Untersuchungen mit kleinen Fallzahlen (häufig weniger als 150 Studienteilnehmer) durchgeführt werden, wie etwa in der Medizin oder der Psychologie. Aufgrund der geringen Stichprobengrößen sind die Untersuchungsergebnisse statistisch sehr unsicher und streuen stark. Hier liegt es nahe, Befunde aus mehreren einander sehr ähnlichen Studien zusammenzufassen, um damit die Stichprobengröße zu erhöhen, was wiederum eine Erhöhung der statistischen Teststärke bewirkt. In der Soziologie dominieren dagegen große, dafür aber auch seltener durchgeführte Umfragestudien. Daher ist es hier unter Umständen weniger dringend, empirische Befunde aus verschiedenen Datenquellen zu integrieren als bei den oben angeführten Disziplinen.

Erschwerend kommt hinzu, dass sich *in nicht-experimentellen Studiendesigns* vergleichbare Gruppen erst nachträglich (ex-post-facto) durch statistische Verfahren – üblicherweise multiple Regressionsmodelle – konstruieren lassen. Es wird argumentiert, dass sich Koeffizienten, die aus unterschiedlich spezifizierten multiplen Regressionsmodellen stammen, nicht sinnvoll zusammenfassen lassen (Brüderl 2004). Abgesehen davon, dass unklar ist, was der „eigentliche“ Effekt jeweils ist, kann man ermitteln, ob die Befunde signifikant variieren, je nachdem, welche Drittvariablen kontrolliert werden. Es sollte sich jedenfalls eine Publikationspraxis durchsetzen, bei der immer auch bivariate Befunde, Standardfehler und Fallzahlen präsentiert werden. Denn dann können Befunde mit und ohne Drittvariablenkontrolle in die Meta-Analyse einfließen und angemessen berücksichtigt werden. IPD-Meta-Analysen haben hier den Vorteil, dass direkt die Originaldaten analysiert werden und vergleichbare Regressionsmodelle geschätzt werden können (Klein et al. 2013; Weiß/Wagner 2008).

## Literatur

- Allen, Mike (2009): Meta-Analysis. In: Communication Monographs 76: 398-407
- Borenstein, Michael/Hedges, Larry V./Higgins, Julian P.T./Rothstein, Hannah R. (2009): Introduction to Meta-Analysis. Chichester: John Wiley
- Brüderl, Josef (2004): Meta-Analyse in der Soziologie: Bilanz der deutschen Scheidungsforschung oder „statistischer Fruchtsalat“? In: Zeitschrift für Soziologie 33: 84-86
- Bushman, Brad J./Wang Morgan C. (2009): Scientific Communication and Literature Retrieval. In: Cooper et al. (Hg.): 207-220
- Cebulla, Andreas/Greenberg, David (2006): The Effects of Welfare-to-work Programs in the United States: Findings from a Meta-Analysis. In: Sozialer Fortschritt 55: 139-145
- Conn, Vicki/Valentine, Jeffrey/Cooper, Harris/Rantz, Marilyn (2003): Grey Literature in Meta-Analyses. In: Nursing Research 52: 256-261
- Cooper, Harris/Hedges, Larry V./Valentine, Jeffrey C. (Hg.) (2009): The Handbook of Research Synthesis. New York: Russel Sage Foundation
- Glass, Gene V. (1976): Primary, Secondary and Meta-Analysis of Research. In: Educational Researcher 5: 3-8
- Glass, Gene V./McGaw, Barry/Smith, Mary Lee (1981): Meta-Analysis in Social Research. Beverly Hills: Sage
- Hedges, Larry V./Olkin, Ingram (1985): Statistical Methods for Meta-Analysis. Orlando: Academic Press
- Hedges, Larry V./Tipton, Elizabeth/Johnson, Matthew C. (2010): Robust variance estimation in meta-regression with dependent effect size estimates. In: Research Synthesis Methods 1: 39-65
- Klein, Thomas/Kopp, Johannes/Rapp, Ingmar (2013): Metaanalyse mit Originaldaten. Ein Vorschlag zur Forschungssynthese in der Soziologie. In: Zeitschrift für Soziologie 42: 222-238.
- Lipsey, Mark W. (2009): Identifying Interesting Variables and Analysis Opportunities. In: Cooper et al. (Hg.): 147-158
- Lipsey, Mark W./Wilson, David B. (2001): Practical Meta-Analysis. Thousand Oaks: Sage
- McAuley, Laura/Pham, Ba'/Tugwell, Peter/Moher, David (2000): Does the Inclusion of Grey Literature Influence Estimates of Intervention Effectiveness Reported in Meta-analyses? In: The Lancet 356: 1228-1231
- Petticrew, M./Roberts, H. (2006): Systematic Reviews in the Social Sciences. A Practical Guide. Malden, MA: Blackwell Publishing
- Rothstein, Hannah R./Sutton, Alexander J./Borenstein, Michael (Hg.) (2005): Publication Bias in Meta-Analysis. Prevention, Assessment and Adjustments. Chichester: John Wiley
- Stewart, Lesley A./Tierney, Jayne F. (2002): To IPD or not to IPD? Advantages and Disadvantages of Systematic Reviews Using Individual Patient Data. In: Evaluation & the Health Professions 25: 76-97

- Van den Noortgate, Wim/López-López, José Antonio/Marín-Martínez, Fulgencio/Sánchez-Meca, Julio (2013): Three-level Meta-analysis of Dependent Effect Sizes. In: Behavior Research Methods 45: 576-594
- Wagner, Michael/Weiß, Bernd (2006): On the Variation of Divorce Risks in Europe: Findings from a Meta-Analysis of European Longitudinal Studies. In: European Sociological Review 22: 483-500
- Weiß, Bernd (2008): Meta-Analyse als Verfahren der Forschungssynthese in der Soziologie. Dargestellt anhand zweier Fallbeispiele zu Schulabsentismus. Dissertation, Wirtschafts- und Sozialwissenschaftliche Fakultät: Universität zu Köln
- Weiß, Bernd/Wagner, Michael (2008): Potentiale und Probleme von Meta-Analysen in der Soziologie. In: Sozialer Fortschritt 57: 250-256
- Weiß, Bernd/Wagner, Michael (2011): The Identification and Prevention of Publication Bias in the Social Sciences and Economics. In: The Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik), 231: 661-684
- White, Howard D. (2009): Scientific Communication and Literature Retrieval. In: Cooper et al. (Hg.): 51-71
- Wilson, David B./Gottfredson, Denise C./Najaka, Stacy S. (2001): School-Based Prevention of Problem Behaviors: A Meta-Analysis. In: Journal of Quantitative Criminology 17: 247-272