

Andreas Daniel / Ute Hoffstätter / Björn Huß / Percy Scheller

DZHW-Studienberechtigten-panel 2008

Daten- und Methodenbericht zu den Erhebungen des Studienberechtigtenjahrgangs 2008 (1. bis 3. Befragungswelle)

Version 1.0.0

Daten- und Methodenbericht

Juni 2017

Dieses Werk steht unter der Creative Commons Namensnennung – Nicht kommerziell – Weitergabe unter gleichen Bedingungen 3.0 Deutschland Lizenz (CC-BY-NC-SA)
<https://creativecommons.org/licenses/by-nc-sa/3.0/de>



Autor(inn)en:
Andreas Daniel
Ute Hoffstätter
Björn Huß
Percy Scheller

Unter Mitarbeit von:
Dr. Anne Weber (Endredaktion)

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH
Forschungsdatenzentrum
Lange Laube 12 | 30159 Hannover | <https://fdz.dzhw.eu>
Juni 2017

Inhaltsverzeichnis

| | |
|---|----|
| Abbildungsverzeichnis | 1 |
| Tabellenverzeichnis | 1 |
| Vorwort | 1 |
| I Einleitung | 2 |
| II Übersicht zum DZHW-Studienberechtigtenpanel 2008 | 3 |
| III Datennutzungshinweise | 5 |
| 1 Inhalt und Anlage der Studie | 8 |
| 2 Erhebungsinstrumente | 12 |
| 2.1 Inhalte der Erhebungsinstrumente | 12 |
| 2.2 Pretest | 13 |
| 3 Grundgesamtheit und Stichprobenverfahren | 15 |
| 4 Durchführung der Erhebungen | 16 |
| 5 Rücklauf | 18 |
| 6 Datenaufbereitung | 20 |
| 6.1 Datenübertragung | 20 |
| 6.2 Codierung offener Angaben | 20 |
| 6.3 Datenprüfung und Datenbereinigung | 22 |
| 6.4 Generierung von Variablen | 23 |
| 6.5 Erstellung der Datensätze | 24 |
| 6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels | 24 |
| 6.7 Codierung fehlender Werte | 25 |
| 7 Gewichtung | 27 |
| 7.1 Vorgehen und Anwendungshinweise | 27 |
| 7.2 Gewichtung des Datensatzes | 28 |
| 8 Anonymisierung | 30 |
| 9 Literaturverzeichnis | 33 |
| Anhang 1: Probit-Regression zur Erstellung des Panelausfallgewichts in Welle 2 | 36 |
| Anhang 2: Probit-Regression zur Erstellung des Panelausfallgewichts in Welle 3 | 37 |

Abbildungsverzeichnis

| | | |
|--------------|---|----|
| Abbildung 1: | Datenzugangswege und Analysepotential | 6 |
| Abbildung 2: | Kohorten-Panel-Design der DZHW-Studienberechtigtenbefragungsreihe | 9 |
| Abbildung 3: | Entwicklung von Brutto- und Nettostichprobe im Zeitverlauf | 19 |
| Abbildung 4: | Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten des DZHW-Studienberechtigtenpanels 2008 | 31 |

Tabellenverzeichnis

| | | |
|------------|--|----|
| Tabelle 1: | Anlage und konzeptionelle Veränderungen der DZHW-Studienberechtigten- befragungsreihe von 1976 bis 2008 | 10 |
| Tabelle 2: | Brutto-, Nettostichproben und Rücklaufquoten des DZHW- Studienberechtigtenpanels 2008 | 19 |
| Tabelle 3: | Vercodete Merkmale und verwendete Codierlisten im DZHW- Studienberechtigtenpanel 2008 | 21 |
| Tabelle 4: | Themengebiete und Kürzel für Variablennamen des DZHW- Studienberechtigtenpanels 2008 | 25 |
| Tabelle 5: | Systematik des FDZ-DZHW für fehlende Werte | 26 |
| Tabelle 6: | Bereitgestellte Gewichte zum DZHW-Studienberechtigtenpanel 2008 | 28 |
| Tabelle 7: | Maßnahmen der statistischen Anonymisierung der Daten des DZHW- Studienberechtigtenpanels 2008 nach Zugangsweg | 32 |

I Einleitung

Die DZHW-Studienberechtigtenpanels stellen eine Untersuchungsreihe zu den Ausbildungs- und Studienentscheidungen beziehungsweise Ausbildungs- und Studienverläufen von Schulabgänger(inne)n mit Abitur oder Fachhochschulreife dar.¹ Sie werden durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)² durchgeführt, vom Bundesministerium für Bildung und Forschung (BMBF) finanziert und dienen – in Ergänzung zur amtlichen Hochschulstatistik – dem nationalen Bildungsmonitoring sowie der Beantwortung von Fragestellungen der Hochschul- und Wissenschaftsforschung. Seit 1976 werden alle ein bis drei Jahre neue Studienberechtigtenjahrgänge befragt.

Im Rahmen des vom BMBF geförderten Projektes zum Aufbau eines Forschungsdatenzentrums am DZHW (FDZ-DZHW) werden die Daten einiger Studienberechtigtenjahrgänge nachträglich zum Zweck der Nachnutzung aufbereitet und dokumentiert.³ Sie werden über verschiedene Zugangswege als *Scientific Use Files* (SUF) für die wissenschaftliche Sekundärnutzung zur Verfügung gestellt. Neben den Datensätzen der Erhebungen werden auch die Dokumentationsmaterialien zu den Datensätzen und zur Durchführung der Studien bereitgestellt.⁴

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation der ersten bis dritten Befragungswelle des Studienberechtigtenjahrgangs 2008 (doi: 10.21249/DZHW:gsl2008:1.0.0). Weitere Dokumentationsmaterialien zur Studie (Datensatzreports, Fragebögen, Filterführungsdiagramme etc.) können frei im Metadatensuchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden.

Abschnitt II dieses Berichts stellt die Eckdaten der Studie im Überblick dar. Die zentralen Informationen zur Nutzung der Daten dieser Studie folgen in Abschnitt III. Kapitel 1 stellt Inhalt und Anlage der DZHW-Studienberechtigtenbefragungsreihe bis 2008 im Allgemeinen und des Studienberechtigtenpanels 2008 im Speziellen vor. Die weitere Gliederung des Berichts orientiert sich im Wesentlichen am Ablauf des Forschungsprozesses. In Kapitel 2 werden die eingesetzten Erhebungsinstrumente und in den Kapiteln 3 bis 6 der Erhebungsprozess beschrieben (Stichprobenziehung, Erhebungsablauf, Rücklauf, Datenaufbereitung). In den Kapiteln 7 und 8 folgt die Darstellung der vorgenommenen Gewichtung und Anonymisierung.

¹ Aktuelle Informationen zum DZHW-Studienberechtigtenpanel können über die Website des Projektes (http://www.dzhw.eu/projekte/pr_show?pr_id=465) abgerufen werden.

² Das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW, <http://www.dzhw.eu>) entstand im August 2013 durch eine Ausgründung aus der HIS Hochschul-Informationssystem GmbH. Im nachfolgenden Text wird durchgängig der Begriff DZHW verwendet, auch wenn die Studie vor der Ausgründung 2013 durchgeführt wurde.

³ Da zum Erhebungszeitpunkt der Daten keine Datennachnutzung vorgesehen war, sind einige Informationen zur Erhebung nicht mit dem Fokus einer späteren Datennachnutzung dokumentiert worden und teilweise nicht mehr rekonstruierbar. An entsprechenden Stellen ist dies im Text angemerkt.

⁴ Informationen zu verfügbaren Datensätzen und Dokumentationen werden auf der Website des FDZ-DZHW (<https://fdz.dzhw.eu>) zur Verfügung gestellt.

II Übersicht zum DZHW-Studienberechtigtenpanel 2008

| | |
|---|---|
| Studienreihe | DZHW-Studienberechtigtenpanel |
| Kohorte | Studienberechtigtenjahrgang 2008 (17. Kohorte der Studienreihe) |
| Erhebende Institution | Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) |
| Förderung | Bundesministerium für Bildung und Forschung (BMBF) |
| Projektmitarbeiter(innen) Projektleitung | 1. Welle: <u>Heike Spangenberg</u> , Heiko Quast 2. Welle: <u>Heiko Quast</u> , <u>Christoph Heine</u> , Julia Willich 3. Welle: <u>Heiko Quast</u> , Markus Lörz, Percy Scheller |
| Themen | Ausbildungs- und Studienabsichten, Bildungsentscheidungen und –verläufe Wahl ingenieurwissenschaftlicher Studiengänge Studienfinanzierung/ -gebühren |
| Erhebungsdesign | Kohorten-Panel-Design |
| Grundgesamtheit | Schulabgänger(innen) allgemeinbildender und beruflicher Schulen des Schuljahres 2007/2008 mit bundesdeutscher Hochschulreife |
| Stichprobe | Disproportional geschichtete Klumpenstichprobe |
| Erhebungsmethode | Standardisierte postalische Befragung |
| Erhebungszeitraum | 1. Welle: Dezember 2007 bis August 2008 2. Welle: Dezember 2008 bis September 2009 3. Welle: Dezember 2012 bis Dezember 2013 |
| Auswertbare Fälle | 1. Welle: n = 28.182 2. Welle: n = 5.933 3. Welle: n = 3.671 |
| Rücklaufquote | 1. Welle: 48,9 % 2. Welle: 33,8 % 3. Welle: 61,9 % |
| Datenprodukte und Zugangswege | SUF: Remote-Desktop, On-Site |

| | |
|------------------------------|---|
| Datensatzstruktur | Personendatensatz im wide-Format |
| DOI | 10.21249/DZHW:gs 2008:1.0.0 |
| Weitere Informationen | https://fdz.dzhw.eu |

Projektpublikationen

1. Welle: Heine, C. & Quast, H. (2009). *Studierneigung und Berufsausbildungspläne. Studienberechtigte 2008 ein halbes Jahr vor Schulabgang* (HIS Hochschul-Informationssystem GmbH, Hrsg.) (HIS: Forum Hochschule 4/2009). Hannover: HIS.
2. Welle: Heine, C., Quast, H. & Beuße, M. (2010). *Studienberechtigte 2008 ein halbes Jahr nach Schulabschluss. Übergang in Studium, Beruf und Ausbildung* (HIS Hochschul-Informationssystem GmbH, Hrsg.) (HIS: Forum Hochschule 3/2010). Hannover: HIS.
3. Welle: Quast, H., Scheller, P. & Lörz, M. (2014). *Bildungsentscheidungen im nachschulischen Verlauf. Dritte Befragung der Studienberechtigten 2008 viereinhalb Jahre nach Schulabschluss* (HIS Hochschul-Informationssystem GmbH, Hrsg.) (HIS: Forum Hochschule 9/2014). Hannover: HIS.

Alle Projektpublikationen werden auf der Website des Projektes (http://www.dzhw.eu/projekte/pr_show?pr_id=270) zum Download bereitgestellt.

Publikationen zum Datensatz (Auswahl)

Lörz, M., Quast, H. & Roloff, J. (2015). Konsequenzen der Bologna-Reform: Warum bestehen auch am Übergang vom Bachelor- ins Masterstudium soziale Ungleichheiten? *Zeitschrift für Soziologie*, 44 (2), 137–155.

Baier, T. & Helbig, M. (2014). Much ado about € 500: do tuition fees keep German students from entering university? Evidence from a natural experiment using DiD matching methods. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 20 (2), 98–121. doi:10.1080/13803611.2014.881745

Helbig, M., Baier, T. & Kroth, A. J. (2012). Die Auswirkung von Studiengebühren auf die Studierneigung in Deutschland. Evidenz aus einem natürlichen Experiment auf Basis der HIS-Studienberechtigtenbefragung. *Zeitschrift für Soziologie*, 41 (3), 227–246.

Lörz, M. (2012). Mechanismen sozialer Ungleichheit beim Übergang ins Studium: Prozesse der Status- und Kulturreproduktion. In R. Becker & H. Solga (Hrsg.), *Soziologische Bildungsforschung* (Kölner Zeitschrift für Soziologie und Sozialpsychologie: Sonderheft, Bd. 52, S. 302–324). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-658-00120-9_13

III Datennutzungshinweise

[Voraussetzungen der Datennutzung] Die Daten des Studienberechtigtenpanels 2008 werden durch das FDZ-DZHW entsprechend dem Bundesdatenschutzgesetz (vgl. § 40 Abs. 1 und Abs. 2 BDSG) ausschließlich zur wissenschaftlichen Nutzung und anonymisiert bereitgestellt.⁵ Das FDZ bietet ein *Scientific Use File* (SUF) für die wissenschaftliche Sekundärnutzung an.

Voraussetzungen für die Nutzung des SUF sind die Anstellung des Datennutzers/der Datennutzerin an einer wissenschaftlichen Einrichtung und der Abschluss eines Datennutzungsvertrags mit dem FDZ. Studierende oder Promovierende ohne eine Anstellung müssen eine Zusammenarbeit mit einer/einem betreuenden Mitarbeiter(in) einer wissenschaftlichen Einrichtung nachweisen. Im Zuge des Vertragsabschlusses wird durch das FDZ auch das Vorliegen eines wissenschaftlichen Nutzungsinteresses geprüft. Das Formular für den Datennutzungsvertrag kann von der Website des FDZ heruntergeladen werden. Das Formular für den Datennutzungsantrag kann von der Website des FDZ heruntergeladen werden.

[Datenzugang] Das SUF wird über zwei Zugangswege angeboten, die hinsichtlich des Speicherortes, und der Kontrollmöglichkeiten des FDZ unterschiedlich restriktiv sind:

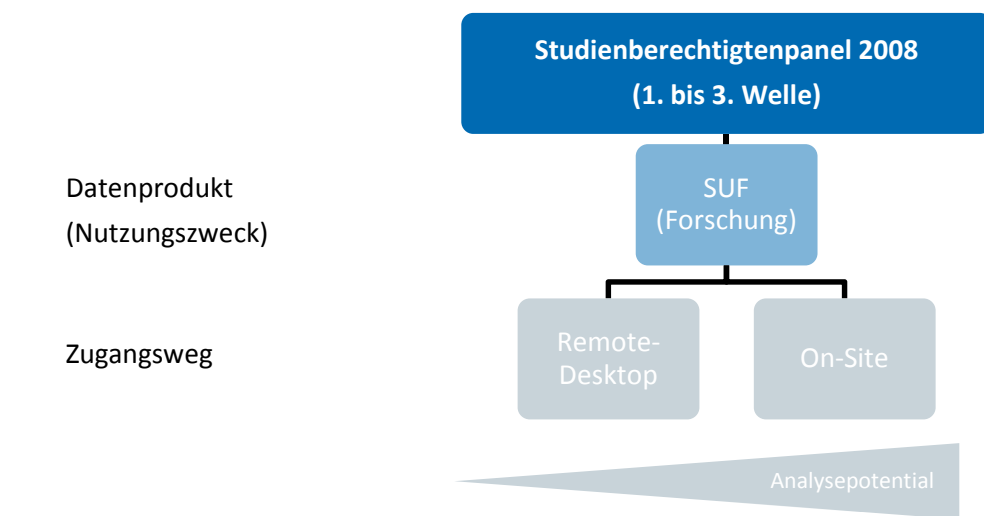
- **Remote-Desktop:** Die Daten werden auf einem Terminal-Server des FDZ bereitgestellt. Über eine besonders gesicherte Verbindung zwischen dem lokalen Computer der nutzenden Person und dem Terminal-Server des FDZ können die Daten mit der auf dem Terminal-Server vorhandenen Software analysiert werden. Das Übertragen der Daten auf den lokalen Computer ist nicht möglich. Analyseergebnisse werden erst nach einer Prüfung auf datenschutzrechtliche Unbedenklichkeit durch das FDZ freigegeben und zur Verfügung gestellt.
- **On-Site:** Die Daten werden in den Räumlichkeiten des FDZ in einer kontrollierten Umgebung an einem speziell gesicherten Computer zur Analyse bereitgestellt. Wie beim Remote-Desktop-Zugang werden Analyseergebnisse erst nach einer Prüfung auf datenschutzrechtliche Unbedenklichkeit durch das FDZ freigegeben und zur Verfügung gestellt.

Die bereitgestellten Daten weisen je nach Zugangsweg einen unterschiedlichen Informationsgehalt auf und unterscheiden sich damit hinsichtlich ihres Analysepotentials (vgl. Abbildung 1). Dabei gilt: Je stärker der Datenzugriff der Nutzer(innen) durch technische und organisatorische Maßnahmen kontrolliert wird, desto mehr Informationen können für die Datennutzer(innen) bereitgestellt werden.⁶ Mit diesem Vorgehen wird ein Höchstmaß an Nutzbarkeit und gleichzeitig ein bestmöglicher Schutz der bereitgestellten Daten sichergestellt.

⁵ Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, S. 6), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (IfBi) (vgl. Koberg, 2016, S. 693) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner, Müller, Schmucker & Roß, 2012, S. 9) orientieren. Das FDZ-DZHW hat den Portfolio-Ansatz an die Anforderungen der eigenen Datenbestände angepasst und nutzt vier Kategorien von Maßnahmen zur Sicherstellung des Datenschutzes, die in unterschiedlicher Weise kombiniert werden: Rechtlich-institutionelle Maßnahmen, informationelle Maßnahmen, technische Maßnahmen und statistische Maßnahmen. Für detailliertere Informationen siehe Datenschutzkonzept des FDZ auf der Website.

⁶ Zu den verschiedenen Anonymisierungsgraden und Analysepotentialen der verschiedenen SUF-Varianten vgl. Kapitel 8..

Abbildung 1: Datenzugangswege und Analysepotential

**[Datenprodukte]**

Über den *Digital Object Identifier* (DOI) 10.21249/DZHW:gsl2008:1.0.0 erfolgt eine Weiterleitung zu einer Übersichtsseite mit zentralen Informationen über die Studie, den weiteren Dokumentationsmaterialien sowie Hinweisen zu den Versionen der bereitgestellten Daten.

Die Daten des Studienberechtigtenpanels 2008 werden für das On-site oder Remote-Desktop Arbeiten jeweils mit zugangsspezifischem Analysepotential (vgl. Abbildung 1) – zur Verfügung gestellt. Die Daten werden im wide-Format bereitgestellt (vgl. Kapitel 6.5). Der Datensatz wird standardmäßig im Stata-Format herausgegeben.

[Gebühren der Datenbereitstellung] CUF und SUF werden derzeit (Stand: Juni 2017) kostenfrei zur Verfügung gestellt. Änderungen bzw. die aktuelle Gebührenordnung können auf der Website des FDZ (<https://fdz.dzhw.eu>) eingesehen werden.

[Pflichten der Datennutzer(innen)] Die Datennutzer(innen) sind verpflichtet, folgende Regeln⁹ einzuhalten:

- **Wissenschaftliche Nutzung:** Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.
- **De-Anonymisierungsverbot:** Jeder Versuch der Re-Identifikation von Analyseeinheiten (z. B. Personen, Haushalten, Institutionen) ist verboten.
- **Gebot zur Mitteilung von Sicherheitslücken:** Falls Datennutzer(innen) Kenntnis von Sicherheitslücken hinsichtlich Datenschutz bzw. Datensicherheit erlangen, müssen diese dem FDZ-DZHW unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** SUF dürfen nur durch die Person genutzt werden, die den Datennutzungsvertrag abgeschlossen hat.

⁹ Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

- **Bereitstellung/Meldung von Publikationen:** Jede Art von Publikation, die aus der Arbeit mit Daten des FDZ-DZHW hervorgeht, muss dem FDZ im Voraus gemeldet und nach Veröffentlichung unverzüglich in elektronischer Form zur Verfügung gestellt werden. Informationen zu bereits vorhandenen Veröffentlichungen können dem Metadatensuchsystem¹⁰ entnommen werden.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen laut folgenden Vorgaben zitiert werden.

[Zitation]

- **Verwendeter Datensatz:**
Heine, C., Quast, H., Spangenberg, H., Lörz, M., Scheller, P. & Willich, J. (2014). *DZHW-Studienberechtigtenpanel 2008*. Aufbereitet durch Daniel, A., Hoffstätter, U., Huß, B. & Scheller, P., doi: 10.21249/DZHW:gsl2008:1.0.0, DATENSATZNAME¹¹, released 2017. Hannover: FDZ-DZHW.
- **Daten- und Methodenbericht:**
Daniel, A., Hoffstätter, U., Huß, B. & Scheller, P. (2017). *DZHW-Studienberechtigtenpanel 2008. Daten- und Methodenbericht der Erhebungen des Studienberechtigtenjahrgangs 2008 (1. bis 3. Befragungswelle)*. Version 1.0.0. Hannover: FDZ-DZHW.

Zusätzlich ist im Text mit folgender Formulierung auf die verwendeten Daten zu verweisen:

*Diese Arbeit nutzt die Daten des DZHW-Studienberechtigtenpanels 2008. Die Daten sind vom Forschungsdatenzentrum des Deutschen Zentrums für Hochschul- und Wissenschaftsforschung (FDZ-DZHW) unter der doi: 10.21249/DZHW:gsl2008:1.0.0 veröffentlicht worden.*¹²

¹⁰ <https://metadata.fdz.dzhw.eu>

¹¹ An dieser Stelle bitte den genauen Namen der verwendeten Datensatzversion angeben, z.B. gsl2008_r_1_0_0 für das Remote-Desktop-SUF des Studienberechtigtenpanels 2008.

¹² In englischsprachigen Publikationen: This scientific work uses data of the German School Leavers Survey 2008, conducted by the German Center for Higher Education Research and Science Studies (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung; DZHW). The data were published by the research data center of the DZHW, doi: 10.21249/DZHW:gsl2008:1.0.0.

1 Inhalt und Anlage der Studie

[Studienreihe] Das DZHW-Studienberechtigtenpanel 2008 ist Teil der DZHW-Studienberechtigtenbefragungsreihe, in der anhand von standardisierten Mehrfachbefragungen Informationen zu den nachschulischen Werdegängen von Schulabgänger(inne)n mit (schulischer) Hochschulzugangsberechtigung erfasst werden.¹³ Die Untersuchungsreihe beginnt mit der Befragung des Studienberechtigtenjahrgangs 1976 in der damaligen BRD; seitdem wurde jeder zweite oder dritte Studienberechtigtenjahrgang (Kohorte) befragt. Die Schulabgänger(innen) in den neuen Bundesländern wurden erstmalig 1990 einbezogen. Die Grundgesamtheit aller DZHW-Studienberechtigtenbefragungen sind die Schulabgänger(innen) allgemeinbildender und beruflicher Schulen eines Schuljahres mit bundesdeutscher Hochschulreife.

In der Regel¹⁴ werden für jeden Studienberechtigtenjahrgang mehrere Befragungswellen zu unterschiedlichen Zeitpunkten vor und nach Erwerb der Hochschulzugangsberechtigung durchgeführt, das heißt es handelt sich um ein kombiniertes Kohorten-Panel-Design. hebt die konzeptionellen Änderungen hervor. Bis 1986 verfolgte das DZHW das Ziel jeden Studienberechtigtenjahrgang kurz nach sowie 2 ½, 4 ½ und 12 ½ Jahre nach Schulabgang zu befragen. Ab 1990 wurde die Befragung zu verspäteten oder korrigierten Ausbildungs- und Studienentscheidungen zweieinhalb Jahre nach Schulabgang gestrichen.¹⁵ Seit dem Studienberechtigtenjahrgang 2005 werden die Studienberechtigtenjahrgänge zusätzlich bereits ein halbes Jahr vor Schulabgang befragt und somit wieder vier Befragungswellen durchgeführt.

Tabelle 1: Anlage und konzeptionelle Veränderungen der DZHW-Studienberechtigtenbefragungsreihe von 1976 bis 2008

| Befragungszeitpunkt | Thematischer Fokus | Studienberechtigtenjahrgänge ^a | | |
|--------------------------------------|--|---|---------------|-----------|
| | | bis 1986 ^b | 1990 bis 2004 | seit 2005 |
| ½ Jahr vor Schulabgang | Ausbildungs- und Studienabsichten | - | - | 1. Welle |
| ½ Jahr nach Schulabgang | Ausbildungs- und Studienentscheidungen | 1. Welle | 1. Welle | 2. Welle |
| 2 ½ Jahre nach Schulabgang | verspätete oder korrigierte Ausbildungs- und Studienentscheidungen | 2. Welle | - | - |
| 3 ½ bzw. 4 ½ Jahre nach Schulabgang | Ausbildungs- und Studienverläufe | 3. Welle | 2. Welle | 3. Welle |
| 10 ½ bis 20 ½ Jahre nach Schulabgang | Erwerbseintritt und -verläufe | 4. Welle | 3. Welle | 4. Welle |

^a Für einige Studienberechtigtenjahrgänge konnten nicht alle Wellen realisiert werden (siehe Abbildung 2).

^b Die Befragungen der Jahrgänge 1976 bis 1986 richteten sich ausschließlich an Hochschulzugangsberechtigte aus den Bundesländern der damaligen Bundesrepublik Deutschland.

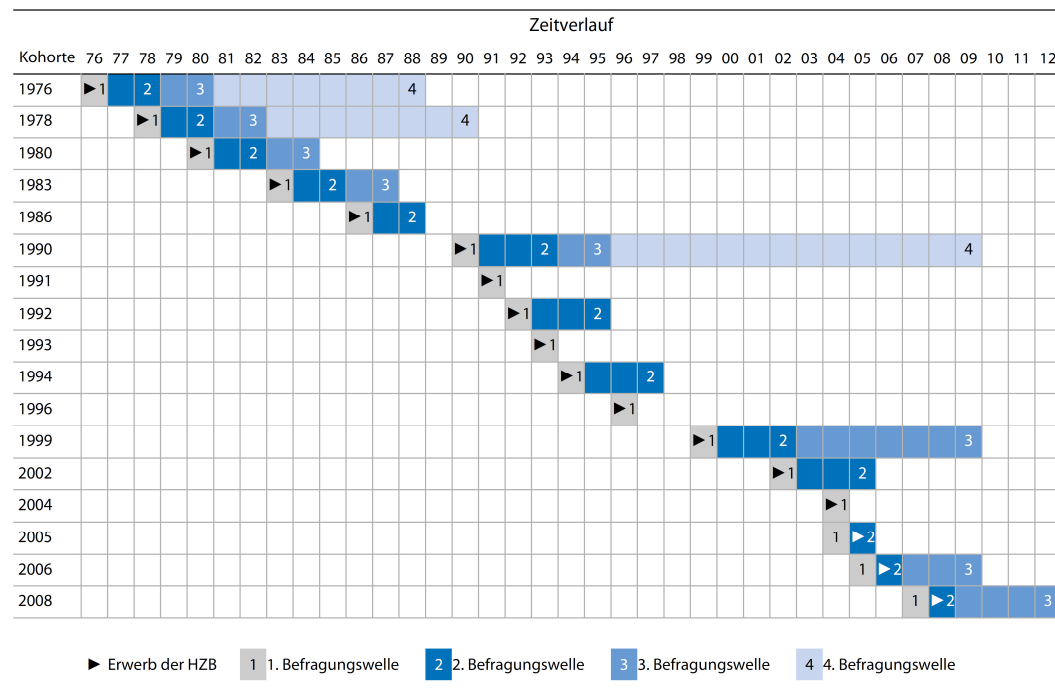
¹³ Aktuelle Informationen zum Studienberechtigtenpanel sowie den bisher verfügbaren Scientific Use Files der Erhebungsreihe finden Sie unter <https://fdz.dzhw.eu>.

¹⁴ Die Ausnahmen bilden die Studienberechtigtenjahrgänge 1991, 1993, 1996 und 2004, die nur einmal befragt wurden.

¹⁵ Eine weitere Ausnahme stellt die, in Kooperation mit dem Bundesinstitut für Berufsbildung (BIBB) durchgeführte, zusätzliche Befragungswelle des Studienberechtigtenjahrgangs 1990 dar, die 5 ½ Jahre nach Schulabgang erfolgte.

Die verschiedenen Erhebungen wurden als schriftlich-postalische Paper-and-Pencil-Befragung (PAPI) durchgeführt. Die Erhebungsinstrumente aller Kohorten enthalten Fragen zu Bildungsin-tentionen, nachschulischem Werdegang, zum Studium, zum Übergang in den Beruf und zu sozio-demographischen und bildungsbiographischen Merkmalen. Der thematische Fokus einer Befragungswelle orientiert sich an der jeweils aktuellen Bildungs-, Berufs- und Lebensphase der Befragten zum Befragungszeitpunkt. Über eine Spanne von fast 40 Jahren bis zur vorliegenden Befragung entstand so eine Zeitreihe von 17 Jahrgängen mit insgesamt 41 Befragungswellen (vgl. Abbildung 2).

Abbildung 2: Kohorten-Panel-Design der DZHW-Studienberechtigtenbefragungsreihe



Anmerkung: Die Befragungen der Jahrgänge 1976 bis 1986 richteten sich ausschließlich an Hochschulzugangsberechtigte aus den Bundesländern der damaligen Bundesrepublik Deutschland.

Quelle: Schneider und Franke (2014); eigene Überarbeitung

[Analysepotential] In allen Kohorten wird je Befragungswelle ein identischer Kern an Informationen erhoben. Auf dieser Grundlage können mit Zeitreihen- bzw. Kohortenvergleichen langfristige Trends im Übergang von der Schule in das weitere Ausbildungssystem und in den Arbeitsmarkt in den Blick genommen werden. Besonders hervorzuheben ist, dass in allen Befragungen von Studienberechtigtenjahrgängen, die mit mehr als einem halben Jahr Abstand nach Erwerb der Hochschulreife erfolgten, halbjahres- oder monatsgenaue kontinuierliche Verlaufsdaten zum individuellen Tätigkeitsverlauf seit dem Schulabgang erfasst werden, die sich für Ereignisdaten- und Sequenzmusteranalysen eignen. Prinzipiell variiert innerhalb eines Befragungsjahrgangs der thematische Fokus von Welle zu Welle. Spezifische Fragen wurden jedoch wiederholt gestellt, sodass die Angaben derselben Befragten aus verschiedenen Wellen direkt miteinander verknüpft und intra-individuelle Veränderungen über den Zeitverlauf betrachtet werden können. Auf Basis des DZHW-Studienberechtigtenpanels können somit Querschnitts-, Trend-, Ereignisdaten-, Sequenzmuster- sowie eingeschränkt kausale Panelanalysen durchgeführt werden. Die Daten der Studienberechtigtenpanels werden in der Regel anhand der Merkmale Geschlecht, Bundesland, Art der Schule und der Hochschulreife gewichtet und an die Grundgesamtheit angeglichen. Zusätzlich werden paneltypische Ausfallprozesse in der Gewichtung der Daten berücksichtigt.

[Einordnung ins Forschungsfeld] Das Stichproben- und Erhebungsdesign sowie die damit verbundenen Analysemöglichkeiten unterscheiden das DZHW-Studienberechtigtenpanel von anderen in Deutschland durchgeführten Befragungen von Studienberechtigten. Keine andere Befragung ermöglicht bundesweite Analysen. Zusätzlich weisen andere Erhebungen in diesem Feld keine oder deutlich kürzere Zeitreihen auf. Andere Studienberechtigtenbefragungen sind beispielsweise die sächsische Abiturientenbefragung¹⁶ (durchgeführt durch das Kompetenzzentrum für Bildungs- und Hochschulplanung an der TU Dresden), die TOSCA-Studie¹⁷ (inzwischen durchgeführt vom Hector-Institut für Empirische Bildungsforschung an der Universität Tübingen)

¹⁶ vgl. dazu: <http://www.kfbh.de/projekte/abiturientenstudie.html>

¹⁷ vgl. dazu: <http://www.wiso.uni-tuebingen.de/faecher/hector-institut-fuer-empirische-bildungsforschung/forschung/laufende-studien/tosca.html>

sowie das Berliner-Studienberechtigtenpanel *Best Up*¹⁸ (durchgeführt vom Deutschen Institut für Wirtschaftsforschung (DIW) und dem Wissenschaftszentrum Berlin (WZB)).

[Spezifika des Studienberechtigtenpanels 2008] Neben den allgemeinen Charakteristika der Studienreihe weist die hier betrachtete Befragung des Studienberechtigtenjahrgangs 2008 folgende Spezifika auf. So wurden – wie in der Pilotstudie mit dem Studienberechtigtenjahrgang 2005 beziehungsweise der Befragung des Studienberechtigtenjahrgangs 2006 – in einer ersten Welle bereits ein halbes Jahr vor Schulabgang die Ausbildungs- und Studienabsichten in den Blick genommen. Als Neuerung fand die Befragung des Studienberechtigtenjahrgangs 2008 erstmals im Klassenverbund und, wenn möglich, während des Unterrichts statt.

Des Weiteren ist die Studienphase der Kohorte 2008 durch die 2002 beschlossene Einführung von gestuften Studiengängen (Bachelor und Master) geprägt. Dieser Veränderung wurde mittels einer detaillierten Erfassung der studienrelevanten Bildungsentscheidung Rechnung getragen. Die Verschiebung des Befragungszeitpunktes der dritten Welle von dreieinhalb auf viereinhalb Jahre nach Schulabgang ermöglichte es zudem, einen größeren Anteil von Studienberechtigten nach Aufnahme eines Masterstudiums zu befragen (34 Prozent aller Studienberechtigten 2008 mit aufgenommenen Bachelorstudium hatten zu diesem Zeitpunkt bereits ein Masterstudium aufgenommen (vgl. Quast et al., 2014, S. 115). Und schließlich ermöglichte es die Entscheidung des Bundesverfassungsgerichts im Jahr 2005, sieben Bundesländern zum Zeitpunkt der ersten Befragung des Studienberechtigtenjahrgangs 2008 allgemeine Studiengebühren oder -beträge (von Beginn des Erststudiums) zu erheben. Zudem fand die Befragung vor dem Hintergrund der bildungspolitischen Diskussion eines erwarteten Mangels an Hochschulabsolvent(inn)en in ingenieurwissenschaftlichen Fächern statt; auch dem wurde durch eine detailliertere Erfassung diesbezüglicher Motive Rechnung getragen.

¹⁸ vgl. dazu: <http://www.best-up.eu>

2 Erhebungsinstrumente

Für die Erhebung des Studienberechtigtenpanels 2008 wurde als Erhebungsinstrument jeweils ein standardisierter Papierfragebogen in deutscher Sprache eingesetzt.¹⁹ Kapitel 2.1 stellt die zentralen Inhalte der drei Erhebungsinstrumente vor.²⁰ Kapitel 2.2 beschreibt die zur Prüfung und Verbesserung durchgeführten Pretests.

2.1 Inhalte der Erhebungsinstrumente

[Charakteristika der Studienreihe] Im Fokus des Studienberechtigtenpanels 2008 steht, wie bei den übrigen Kohorten der Studienreihe, die Beschreibung und Erklärung von Bildungsentscheidungen. Der Inhalt der Instrumente zur Erhebung des Bildungshandelns sowie dessen Erklärung orientiert sich – analog zu einer Klassifikation von Mertens (1976) – an dem zu erwartenden chronologischen Auftreten der unterschiedlichen Lebensphasen von der *nachschulischen Übergangsphase* (1) über die *Qualifizierungsphase* (2) bis zu der *Berufseintrittsphase* (3). Neben der Erfassung von Informationen zu Bildungsabsichten und -entscheidungen, werden in allen Erhebungsinstrumenten auch Einstellungen und Überzeugungen erfragt, um die Bildungsentscheidungen theoretisch fundiert erklären zu können.

Die Befragung der Studienberechtigten vor Schulabgang²¹ und Erwerb der Hochschulzugangsberechtigung (1. Welle) fokussiert auf die Ausbildungs- und Studienabsichten. Zu diesem Zweck erfolgt die Erfassung von Persönlichkeitseigenschaften, von Überzeugungen bezüglich Schulzeit, zu Zukunftsplänen für die Zeit nach dem Schulabschluss und der diesbezüglichen Entscheidungsfindung, zur Bedeutung verschiedener Informationswege für die Studien- und Ausbildungsplanung sowie zur Studienfinanzierung.

In der ersten nachschulischen Befragung (2. Welle) rücken die Ausbildungs- und Studienentscheidungen in den Fokus. Dabei werden im Wesentlichen weitere Informationen zur Schulzeit sowie zur Person, die aktuell ausgeübte Tätigkeit, detaillierte Informationen zu den (ggf. beabsichtigten) Bildungsschritten, zu Motiven des nachschulischen Werdegangs, zu Berufsaussichten sowie allgemein die Berufs- und Lebensziele erfasst.

Die weitere nachschulische Befragung etwa drei- bis viereinhalb Jahre nach Schulabgang (3. Welle) fokussiert über alle Kohorten retrospektiv vor allem die Realisierung der Bildungsabsichten. Dazu werden die Ausbildungs- und Studienverläufe in den Blick genommen. Anhand eines Tätigkeitstableaus (Frage 2.1 in der 3. Welle des Studienberechtigtenjahrgangs 2008) werden die Zeiträume²², die Art der Tätigkeiten sowie erläuternde Anmerkungen erhoben. Zum Studium werden die Hauptstudienfächer, die angestrebte Abschlussprüfung sowie der Name und Ort der Hochschule/Berufsakademie erfragt. Zu einer beruflichen Ausbildung oder einer beruflichen Tätigkeit werden ferner die Berufsbezeichnung und gegebenenfalls die Art der Ausbildung er-

¹⁹ Die Fragebögen der drei Wellen können von der Website des FDZ heruntergeladen werden. Ebenso stehen dort Flussdiagramme der Filterführung der drei Fragebögen zur Verfügung.

²⁰ Informationen zu der Herkunft und Entwicklung von verwendeten Skalen liegen nicht vor.

²¹ Eine Befragung vor Schulabgang fand erstmalig in der Befragung des Studienberechtigtenjahrgangs 2005 statt und ist seitdem Standard des Studienberechtigtenpanels. Die vorherigen Kohorten (1976 bis 2004) wurden erst nach Schulabgang kontaktiert und befragt (vgl. Kapitel 1).

²² Die Zeiträume wurden zu Beginn der Studienreihe halbjahres- und seit der Kohorte 1999 monatsgenau erhoben.

fasst. Ergänzt wird die Erhebung und Rekonstruktion der Bildungswege durch die Reflexion der Entscheidungsfindung und die Erhebung weiterer Bildungsabsichten, einer wiederholten Messung allgemeiner Berufs- und Lebensziele sowie durch Bewertungen der Bildungsentscheidung (z. B. des Studienfaches).

[Spezifika des Studienberechtigtenpanels 2008] Neben dem Umstand, dass die Studienberechtigten 2008 erst die dritte Kohorte der Befragungsreihe sind, die bereits vor Schulabgang befragt wurde, ist in Bezug auf die Erhebungsinstrumente der besondere bildungspolitische Kontext hervorzuheben. Im Zusammenhang der Einführung allgemeiner Studiengebühren bzw. -beiträge sowie der Umstellung auf das gestufte Studiensystem von Bachelor- und Masterabschlüssen wurden insbesondere für die zweite und dritte Befragungswelle zusätzliche Fragen aufgenommen. In der ersten und dritten Welle des Studienberechtigtenpanels 2008 wurden (gegenüber den vorangegangenen Kohorten von 2005 und 2006) erstmalig Instrumente zur Unterrichtsbewertung (Frage 4 und 5 in Welle 1 und Frage 1.3 in Welle 3) sowie zu den persönlichen Empfindungen im Hinblick auf die Entscheidung über den nachschulischen Werdegang (Fragen 16 bis 19 in Welle 1 und Frage 1.2 in Welle 3) ergänzt. In der zweiten Befragungswelle wurden spezielle Fragen zur Aufnahme eines ingenieurwissenschaftlichen Studiums gestellt (Fragen 13 und 14 in Welle 2). In der zweiten und dritten Befragungswelle wurden Absichten, ein Masterstudium aufzunehmen (Fragen 27 und 28 in Welle 2) sowie umfangreiche Informationen zum Thema Studiengebühren (Fragen 35 bis 38 in Welle 2) und Studienverzicht (Frage 6.1 in Welle 3) sowie ergänzende Informationen zur Herkunft (Frage 8.4 und 8.5) erhoben. Des Weiteren wurden in der dritten Welle ein rückblickend anders gewählter Ausbildungs- und Berufsweg (Frage 1.5), die Gründe gegen ein Masterstudium (Fragen 5.2) sowie Informationen zu einer Erwerbstätigkeit (Frage 7.2 und 7.4 bis 7.6) präziser als in früheren Kohorten erfragt. Auch wurden erstmals Informationen zur studentischen Arbeitstätigkeit und Studienfinanzierung, wie Jobben und Praktika, erfasst (Frage 4.12 bis 4.14). Zudem wurden in der dritten Befragungswelle dieses Jahrgangs neue Instrumente zur Messung von persönlichen Faktoren eingesetzt, wie etwa Zufriedenheit (Frage 1.1. und 1.8), selbst eingeschätzte Stärken und Schwächen (Frage 1.4) sowie entscheidungsrelevante Persönlichkeitseigenschaften (Frage 1.6 bis 1.7).

2.2 Pretest

[Ziel und Verfahren] Die Erhebungsinstrumente der drei Befragungswellen wurden im Vorfeld der Erhebung durch Pretests geprüft. Dabei sollte erstens für die bereits in vorherigen Kohorten eingesetzten Fragen und Antwortvorgaben geprüft werden, ob sie von dem Studienberechtigtenjahrgang 2008 gleich perzipiert werden würden wie von früheren Jahrgängen.²³ Zweitens sollte für die in allen drei Wellen des Studienberechtigtenpanels 2008 neu eingesetzten Messinstrumente deren Verständlichkeit, Beantwortbarkeit, theoretische Aussagekraft, Reliabilität und Validität getestet werden. Drittens zog die Aufnahme der neuen Fragen grundsätzliche Veränderungen im Aufbau und Layout des Fragebogens sowie an der Befragungsdauer nach sich, die evaluiert werden sollten.

²³ Die Erhebungsinstrumente des Studienberechtigtenpanels 2008 basieren im Wesentlichen auf dem für die Untersuchungsreihe etablierten Design und den in den Befragungen der Studienberechtigten 2005 und 2006 bewährten schriftlichen Fragebögen.

Eine erste Version der Erhebungsinstrumente wurde in allen drei Wellen zunächst im Rahmen von *Expertenbewertungen* (vgl. Häder, 2015, S. 406–407) begutachtet. Neben diesen Bewertungen fanden in den einzelnen Wellen jeweils zielgerichtete Entwicklungs- und Abschluss-Pretests²⁴ statt. Diese wurden als sogenannte *Pretests im Feld* (vgl. Häder, 2015, S. 396), also unter möglichst ähnlichen Bedingungen wie in den tatsächlichen Befragungen, durchgeführt. In den ersten beiden Wellen wurden iterative Pretests (vgl. Prüfer & Rexroth, 2000) durchgeführt, die insgesamt drei Phasen (Welle 1) bzw. zwei Phasen (Welle 2) umfassten.²⁵ Zum Test des Erhebungsinstrumentes der dritten Welle wurde ein Pretest ohne Iterationen durchgeführt.

[Probanden] In allen drei Befragungswellen wurden zum einen fünf bis zehn Mitarbeiter(innen) des DZHW als Expert(inn)en aus dem Bereich der Hochschulforschung einbezogen. Zum anderen wurden 50 bis 100 externe Testpersonen rekrutiert, die im gleichen Alter beziehungsweise in einer ähnlichen Lebenssituation wie die Befragten waren.

[Durchführung] Die Pretests der ersten Befragungswelle fanden zu Beginn des Schuljahres 2007/2008, also ein halbes Jahr vor dem jeweiligen Feldstart, statt. Sie wurden an Schulen verschiedener Schulformen durchgeführt, um die Auswahl der Befragten strukturell möglichst nah an der endgültigen Stichprobe zu halten.

An den Pretests der ersten Welle nahmen insgesamt 58 Schüler(innen) teil. In einem ersten Pretest wurden die Erhebungsinstrumente von Schüler(inne)n des Abschlussjahrgangs eines allgemeinbildenden Gymnasiums ausgefüllt. Das überarbeitete²⁶ Erhebungsinstrument wurde in einer zweiten Phase dem Abschlussjahrgang einer beruflichen Schule vorgelegt. In einer dritten Phase erhielt der Abschlussjahrgang einer Gesamtschule das erneut überarbeitete Erhebungsinstrument. Alle Versionen wurden während des Unterrichts im Beisein von DZHW-Mitarbeiter(inne)n ausgefüllt. Im Anschluss erörterten diese mit den Schüler(inne)n das Verständnis der Fragen und Antwortvorgaben sowie Schwierigkeiten mit dem Erhebungsinstrument.

Für die zweite Welle wurden im Sommer 2008 zwei Pretests durchgeführt. Im ersten Pretest wurde 30 Studienberechtigten des Jahrgangs 2008²⁷ das vorläufige Erhebungsinstrument postalisch zugesandt. Das Ausfüllen der Fragebögen war mit der Bitte verbunden, Ausfüllprobleme, Filterfehler, Unverständlichkeiten sowie unvollständige Antwortvorgaben zu notieren und (porto-frei) an das DZHW zurückzusenden. Ein zweiter Pretest richtete sich an die zahlenmäßig stark vertretenen Studienanfänger(innen). Im Rahmen einer Einführungslehrveranstaltung zur empirischen Sozialforschung an der Universität Hannover füllten 65 Proband(inn)en das Erhebungsinstrument aus. Deren Hinweise zur Verbesserung von Fragestellungen, Antwortvorgaben und Ausfüllhinweisen, insbesondere zum Fokusthema Studiengebühren und Studienverzichtsgründe, wurden in das endgültige Erhebungsinstrument eingearbeitet.

Der dritten Welle ging im Oktober 2012 ein Pretest ebenfalls mit Studierenden der Universität Hannover voraus. Der Pretest mit 50 Teilnehmer(inne)n führte insbesondere im Bereich von neu entwickelten Fragen zum Thema „Jobben und Praktika während des Studiums“ zu einer Verbesserung des Erhebungsinstrumentes.

²⁴ Abschluss-Pretests dienen im Gegensatz zu Entwicklungs-Pretests in der Regel kleineren Restkorrekturen der Frageinstrumente wie Kürzungen oder Umstellungen sowie der Überprüfung neuer Filterführungen oder Veränderungen des Druckbilds, vgl. Schnell, Hill und Esser (2005), S. 348.

²⁵ In jeder Phase wurden die Fragebögen, die jeweils die Überarbeitungen des vorherigen Pretests enthielten, der nächsten Befragtengruppe vorgelegt.

²⁶ Informationen zu den Änderungen am Erhebungsinstrument nach den Pretests liegen nicht vor.

²⁷ Diese Pretestteilnehmer(innen) waren nicht Teil der Stichprobe, hatten aber an der ersten Befragungswelle teilgenommen und damit ihre Adresse angegeben (vgl. Abs. 6.3 Datenprüfung und Datenbereinigung).

3 Grundgesamtheit und Stichprobenverfahren

[Grundgesamtheit] Die Grundgesamtheit der Befragung des Studienberechtigtenjahrgangs 2008 umfasst alle Schülerinnen und Schüler der Bundesrepublik Deutschland, die im Schuljahr 2007/2008 an allgemeinbildenden oder beruflichen Schulen die (fachgebundene) Hochschulreife erwarben.

[Stichprobenverfahren] Aufgrund fehlender oder nicht zugänglicher Listen von Studienberechtigten auf Individualebene war es nicht möglich, eine einfache Zufallsstichprobe zu ziehen. So mussten die Individuen über die Schulen rekrutiert werden. Als Stichprobendesign wurde ein disproportional geschichtetes Klumpenstichprobenverfahren als Spezialfall einer mehrstufigen Zufallsauswahl (vgl. Häder, 2015, S. 169) gewählt. Die Schulen bzw. Schulzweige²⁸ stellen dabei die „primary sampling units“ (Klumpen), die Studienberechtigten des Jahrgangs 2008 die „secondary sampling units“ innerhalb dieser Klumpen dar.²⁹ Die Auswahlgesamtheit bildete eine Liste mit allen relevanten Schulen bzw. Schulzweigen. Die Schichtung erfolgte anhand der Merkmale Bundesland und Schultyp. Innerhalb der Schichten wurden Ziehungsquoten für die Klumpen und Individuen festgelegt, die sich zum einen an der Verteilung des Studienberechtigtenjahrgangs 2004/2005 laut amtlicher Statistik orientierten und zum anderen die Rücklaufquoten der letzten Erhebungen berücksichtigten. Im Zuge dessen erhielten Studienberechtigte aus Bremen, Hamburg und dem Saarland sowie Schülerinnen und Schüler von Abendgymnasien und -kollegs überproportionale Auswahlwahrscheinlichkeiten, um in diesen Gruppen ausreichend große Fallzahlen zu generieren.

Im Rahmen der praktischen Durchführung des Ziehungsverfahrens erfolgte zunächst in jeder Schicht die Auswahl der Klumpen (Schulen) über eine einfache Zufallsauswahl. Die ausgewählten Schulen meldeten im Folgenden ihre Schulzweige und die Schülerzahlen zurück, sodass ein zweiter Auswahlsschritt durchgeführt werden konnte, der die Erfüllung der Ziehungsquoten auf Individualebene zum Ziel hatte. In diesem zweiten Schritt wurden in jeder Schicht *sukzessive* so viele Schulen/Schulzweige zufällig gezogen, dass mindestens 97 % der schichtspezifischen Ziehungsquote für die Studienberechtigten erreicht wurde.³⁰ Insgesamt wurden nach diesem Verfahren fünf unterschiedliche Stichproben erzeugt, von denen jene Variante ausgewählt wurde, die am besten geeignet erschien.³¹ Hiernach wurden die gezogenen Schulen (oder Schulzweige an diesen) komplett erhoben.

²⁸ Manche berufsbildenden Schulen bestehen aus verschiedenen Schulzweigen, von denen jedoch nur jene für die Stichprobe relevant waren, die zum Erwerb der (Fach)-Hochschulzugangsberechtigung führten. Daher wurden in manchen Fällen nur bestimmte Schulzweige berücksichtigt.

²⁹ Beispiel für den Personenkreis eines Klumpens: Alle Studienberechtigten des Prüfungsjahrgangs 2008 des allgemeinbildenden Gymnasiums XY in Bundesland YZ.

³⁰ Die Prüfung, ob die Ziehungsquote in einer Schicht erreicht wurde, erfolgte nach jedem Ziehungsschritt. Erfüllte beispielsweise die kumulierte Schülerzahl von fünf Schulen die Ziehungsquote für die Schülerzahl in der spezifischen Schicht, so wurde nach der Auswahl der fünften Schule keine weitere Schule gezogen.

³¹ Da die Ziehung im zweiten Schritt ohne Zurücklegen erfolgte, die Ziehungsreihenfolge jedoch nicht dokumentiert wurde und sich die Auswahlkriterien für die Selektion der letztendlichen Stichprobenvariante nicht rekonstruieren lässt, können die fallspezifischen Auswahlwahrscheinlichkeiten lediglich geschätzt werden (siehe Kapitel 7).

4 Durchführung der Erhebungen

[Kontaktaufnahme und Adresspflege] Die gezogenen Schulen wurden vom DZHW angeschrieben und um ihre Teilnahme gebeten. Zudem teilte das DZHW den Schulen die Kriterien mit, anhand derer sie die Zielpersonen für das Studienberechtigtenpanel 2008 identifizieren sollten (vgl. Kapitel 3).³² Da die Hochschulen die Kontaktdaten ihrer Schüler(innen) aus Datenschutzgründen nicht herausgeben durften, teilten sie dem DZHW lediglich die jeweilige Schülerzahl (ggf. für einzelne Schulzweige) mit. Anschließend sendete das DZHW Pakete mit der passenden Anzahl an Erhebungsunterlagen für die erste Befragungswelle postalisch an die jeweiligen Schulen.³³ Aufgrund des direkten Feldzugangs erfolgte die Befragung über die Schulen in den Klassenverbänden und, wenn möglich, während des Unterrichts.³⁴

Um die weiterhin teilnahmebereiten Personen in den Folgewellen direkt durch das DZHW kontaktieren zu können, wurden im Fragebogen der ersten Welle deren Adressdaten erfasst. Beim Eingang eines Fragebogens im DZHW wurden die Adressangaben auf den eingegangenen Fragebögen neben der eingedruckten Identifikationsnummer mit einer laufenden Nummer versehen. Daraus wurde eine Referenzliste mit Identifikations- und laufender Nummer sowie den zugehörigen Adressen erstellt.³⁵ In der zweiten und dritten Welle wurden alle Personen kontaktiert, die ihre Adresse in der ersten Welle angegeben und an der vorangegangenen Welle teilgenommen hatten. Für die folgenden Befragungswellen wurden die Identifikationsnummern und die Adressen in die Fragebögen gedruckt, sodass die Befragungsunterlagen direkt an die Postadressen der Teilnehmerinnen und Teilnehmer versandt werden konnten.³⁶ In allen drei Erhebungswellen erfolgte vor dem Versand der Erhebungsunterlagen ein Adressabgleich, um auch Personen zu berücksichtigen, die in der Zwischenzeit umgezogen waren. Zudem wurden durch das DZHW Adressen nachrecherchiert, wenn Erhebungsunterlagen unzustellbar waren.³⁷

[Erhebungsunterlagen] Die Erhebungsunterlagen bestanden in allen Befragungswellen pro zu befragender Person aus einem Anschreiben (inkl. Datenschutzinformationen) und dem Fragebogen. Eine portofreie Rücksendung des ausgefüllten Fragebogens wurde über die gesammelte Rücksendung in kleinen Paketen durch die Schulen bzw. über ein an das DZHW adressiertes Antwortkuvert gewährleistet.

[Feldphase] Der Erhebungszeitraum der ersten Befragungswelle erstreckte sich von Dezember 2007 bis Ende August 2008.³⁸ Aufgrund des angewendeten Kontaktverfahrens über die Schulen konnte das DZHW keinen direkten Einfluss auf den genauen Befragungszeitpunkt nehmen.

³² An dieser Stelle können durch die Schulen vereinzelt auch Personen als stichprobenzugehörig identifiziert worden sein, die nicht zur Grundgesamtheit gehörten (Overcoverage). Dies kam beispielsweise vor, wenn Schulabgänger(innen) nur den schulischen Teil der Fachhochschulreife erworben haben ohne anschließend eine volle Fachhochschulreife anzustreben oder die Schulen den Fragebogen an Schüler(innen) eines falschen Jahrgangs verteilt haben.

³³ Dem Schulanschreiben waren eine frankierte Antwortpostkarte, ein Flyer mit ausgewählten Ergebnissen von vorherigen Studienberechtigtenpanels sowie der Hinweis auf Teilnahmeprämien beigelegt. Darüber hinaus gaben in 14 der 16 Bundesländer die Kultusbehörden ein explizites Empfehlungsschreiben an die Schulen ab.

³⁴ Dass Schulen, Lehrer(innen) oder Klassen nicht teilnahmen, stellt eine Ausnahme dar (zu der aber keine näheren Informationen vorliegen).

³⁵ Zur Gewährleistung des Datenschutzes wurde der Adressabschnitt vom Fragebogen abgetrennt und die Referenzliste getrennt von den Befragungsdaten auf einem geschützten Server gespeichert.

³⁶ In den Fragebögen konnten Teilnehmerinnen und Teilnehmer in einem Adressabschnitt ihre Adressdaten korrigieren.

³⁷ Neben dem Adressaktualisierungsservice der Deutschen Post sowie der Europäische Melderegisterauskunft wurden auch Einwohnermeldeamtsabfragen vorgenommen.

³⁸ Die Feldzeit wurde so lange wie möglich – parallel zu den weiteren Projektschritten – ausgedehnt und jeder noch eintreffende Fragebogen bis August 2008 berücksichtigt.

Die Erhebungszeiträume der zweiten und dritten Erhebungswelle erstreckten sich von Dezember 2008 bis Anfang September 2009 bzw. Dezember 2012 bis Dezember 2013.³⁹ Durch den nun im DZHW vorliegenden Adressbestand konnten konkrete Versandzeitpunkte für die Erhebungsunterlagen festgelegt werden.⁴⁰ Zudem konnten die Erinnerungsschreiben etwa zwei Monate nach Beginn der Erhebung gezielt nur an diejenigen Zielpersonen verschickt werden, die sich noch nicht an der Erhebung beteiligt hatten.

[Rücklaufsteigernde Maßnahmen] Die rücklaufsteigernden Maßnahmen zielten zum einen auf Anreize für die Schulen, die Befragung organisatorisch zu unterstützen und zum anderen auf die individuelle Motivation der Befragten. Für den Erstkontakt bestand die organisatorische Unterstützung der Schulen (und verantwortlicher Ansprechpartner(innen)) darin, Unterrichtszeit zum Ausfüllen bereitzustellen, die Fragebögen am Ende der Unterrichtsstunden einzusammeln sowie gegebenenfalls bei einzelnen Klassen die Rückgabe der Fragebögen zu unterstützen. Als Gegenleistung wurden den Schulen schulspezifische Auswertungen angeboten. Auch wurden für eine hohe Teilnahmebereitschaft Preise für die Schulen (drei Mal 500 Euro) ausgelobt. Als nützlich erwiesen sich zudem ein Hinweis auf die gesamtgesellschaftliche Bedeutung der Erhebung in den Anschreiben, ministerielle Aufforderungen die Befragung zu unterstützen sowie Anschreiben und die telefonische Kontaktierung bei Schulen, von denen noch keine oder nur wenige Fragebögen eingegangen waren. Die auf die individuelle Motivation der Befragten zielenden rücklaufsteigernden Maßnahmen bestanden in einem Hinweis auf den gesamtgesellschaftlichen und altersgruppenspezifischen Nutzen im Anschreiben an die Befragten, einem zusammen mit dem Anschreiben an diese verschickten Flyer, einer Projektwebsite⁴¹ (mit Informationen zum Projekt und daraus entstandenen Publikationen) und der Versand eines Erinnerungsschreibens (in Welle 2 und 3). Als materieller individueller Anreiz bestand in der ersten Welle für die Befragten die Möglichkeit zum Gewinn von 15 Büchergutscheinen im Wert von jeweils 100 Euro. In der zweiten Welle wurde auf den Einsatz von rücklaufsteigernden materiellen Anreizen verzichtet. In der dritten Welle nahmen alle Teilnehmer(innen) der Befragung an einer Verlosung mit folgenden Gewinnen teil: ein Apple iPad4, zwei Apple iPod nano sowie zwanzig Amazon-Gutscheine im Wert von jeweils 50 Euro.

³⁹ Nach Juni 2013 wurden nur noch vereinzelt Fragebögen an das DZHW zurückgeschickt. Der letzte im Datensatz enthaltene Fragebogen ging im Dezember 2013 ein.

⁴⁰ Die Versandzeitpunkte können nicht mehr exakt rekonstruiert werden.

⁴¹ Die Projektwebsite (<http://www.panel2008.de/>) wurde im Laufe der Erhebung für jede Welle aktualisiert.

5 Rücklauf

[Rücklauf] Auf der Grundlage des Stichproben- und Erhebungsdesigns (siehe Kapitel 3) wurden für die Befragung der ersten Welle 1.363 Schulen angeschrieben. Von diesen Schulen meldeten 959 (70 Prozent) die Anzahl ihrer zu erwartenden Studienberechtigten zurück. Dies entsprach einer Zahl von 91.311 Schülerinnen und Schülern. Die daraus gezogene bereinigte Bruttostichprobe der ersten Befragungswelle umfasste 57.622 Studienberechtigte, die kontaktiert wurden (13% der Grundgesamtheit, vgl. Heine et al., 2010, S. 10).⁴² Die Rücklaufquote betrug 48,9 Prozent, sodass eine Nettostichprobe von 28.182 auswertbaren Fragebögen zu verzeichnen war. 17.558 Personen, also rund 62 Prozent der 28.182 Teilnehmer(innen) der ersten Welle, erklärten sich mit einer Kontaktierung für weitere Befragungen einverstanden.

In der zweiten Erhebung, welche auf postalischem Wege durchgeführt wurde, konnten folglich 17.558 Personen in die Bruttostichprobe einbezogen werden. Die Rücklaufquote der zweiten Befragung war im Vergleich zur ersten Welle mit 33,8 Prozent deutlich geringer, was zu einer Nettostichprobe von 5.933 Personen führte.

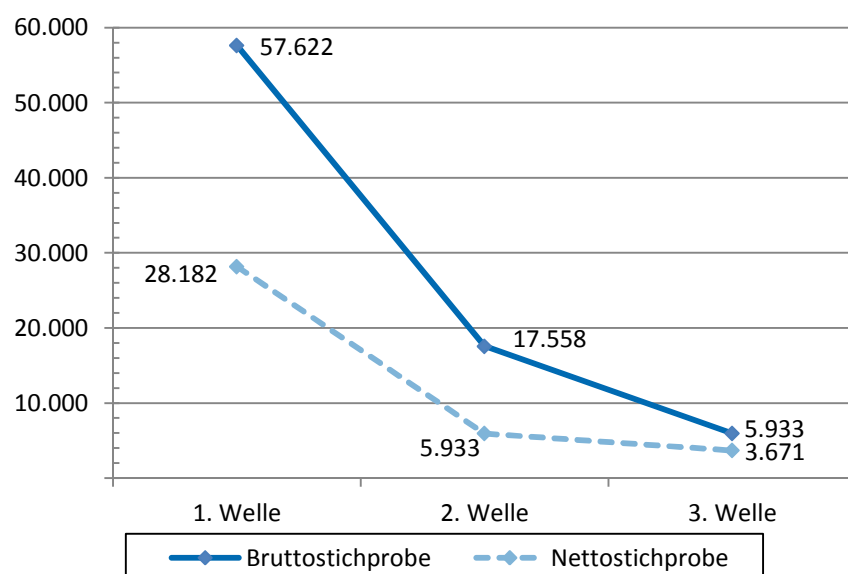
Diese 5.933 Personen, welche sowohl an der ersten als auch an der zweiten Befragung teilgenommen hatten, bildeten die Bruttostichprobe zur dritten Erhebungswelle. Die Rücklaufquote dieser Erhebungswelle betrug 61,9 Prozent (3.671 Personen).

[Panelausfälle] Das Studienberechtigtenpanel 2008 ist von paneltypischen Ausfallprozessen betroffen.⁴³ Diese äußern sich in der grundsätzlichen Verweigerung zur Teilnahme an Folgebefragungen (keine Angabe von Adressangaben für die Kontaktierung in der zweiten Welle) oder in Form von Nichtteilnahmen nach einer (versuchten) Kontaktaufnahme in der zweiten Erhebungswelle. In Abbildung 3 und Tabelle 2 ist die beschriebene Entwicklung der Brutto- und Nettostichprobe im Zeitverlauf dargestellt.

Da in der zweiten und dritten Welle jeweils nur jene Personen kontaktiert wurden, die zu den vorherigen Zeitpunkten teilgenommen hatten, reduziert sich die Bruttostichprobe von Zeitpunkt zu Zeitpunkt deutlich. Im Drei-Wellen-Panel verbleiben somit, gemessen an der Ausgangsstichprobe, lediglich 6,4 Prozent.

⁴² Im Laufe der Rückmeldung von den Schulen wurden auch Personen im Nachhinein als nicht stichprobenzugehörig eingestuft, weil sie im relevanten Prüfungsjahr 2008 nur den schulischen Teil der Fachhochschulreife absolvierten. Diese Art von *Overcoverage* trat auf, da manche Schulen mehrere Schulzweige aufweisen, die nicht alle Teil der Stichprobe waren, jedoch von den Schulen zurückgemeldet wurden. Sowohl die Brutto- als auch die Nettostichprobe sind bereits um diese Zahlen bereinigt.

⁴³ Zu paneltypischen Ausfallprozessen vgl. Schnell et al. (2005), S. 421.

Abbildung 3: Entwicklung von Brutto- und Nettostichprobe im Zeitverlauf

Anmerkung zur Brutto- und Nettostichprobe: Nur Personen mit Stichprobenberechtigung

Die deutliche Diskrepanz zwischen der Rücklaufquote der zweiten Welle der Befragung von 2008 und der zweiten Wellen von vorherigen Studienberechtigtenjahrgängen kann teilweise durch die Form der Erhebung in der ersten Welle erklärt werden: Durch die Befragung im Klassenverbund wurden vermutlich auch Personen zur Teilnahme motiviert, die andernfalls nicht teilgenommen hätten. Diese fehlende grundsätzliche Teilnahmebereitschaft trug möglicherweise zu einer höheren Verweigerungsquote in der – auf postalischem Weg zugestellten – zweiten Welle bei und äußerte sich ebenfalls in der recht hohen Anzahl an Bögen ohne Adressrückmeldungen. Zugleich wurde in der zweiten Welle keine materielle Incentivierung der Befragten vorgenommen. Auch dies könnte der geringeren Teilnahmebereitschaft zuträglich gewesen sein.

Tabelle 2: Brutto-, Nettostichproben und Rücklaufquoten des DZHW-Studienberechtigtenpanels 2008

| | Welle 1 | Welle 2 | Welle 3 |
|---|---------|---------|---------|
| Bruttostichprobe | 57.622 | 17.558 | 5.933 |
| Nettostichprobe | 28.182 | 5.933 | 3.671 |
| Rücklaufquote | 48,9 % | 33,8 % | 61,9 % |
| Anteil der Nettostichprobe an der vorangegangenen Nettostichprobe | -- | 21,1 % | 61,9 % |
| Anteil Nettostichprobe an der Bruttostichprobe t1 | 48,9 % | 10,3 % | 6,4 % |
| Anteil der Bruttostichprobe an der vorangegangenen Bruttostichprobe | -- | 30,5 % | 33,8 % |
| Anteil der Bruttostichprobe an der Bruttostichprobe t1 | -- | 30,5% | 10,3 % |

6 Datenaufbereitung

Im Folgenden werden die verschiedenen Arbeitsschritte der Datenaufbereitung beschrieben. Diese erfolgten in den drei Befragungswellen analog. Die in den Kapiteln 6.1 bis 6.3 beschriebenen Tätigkeiten wurden bereits durch das Primärforschungsprojekt durchgeführt. Die Generierung von Variablen (Kapitel 6.4) wurde sowohl durch das Primärforschungsprojekt als auch im Rahmen der Datenedition durch das FDZ-DZHW vorgenommen. Die in den Kapiteln 6.5 bis 6.7 dargestellten Tätigkeiten wurden durch das FDZ durchgeführt, teilweise aufbauend auf Vorarbeiten des Primärforschungsprojektes. Die im Rahmen der Datenedition vorgenommenen Aufbereitungsprozesse der Gewichtung und Anonymisierung werden in den beiden folgenden Kapiteln 7 und 8 gesondert erläutert.

6.1 Datenübertragung

Zur weiteren Verarbeitung wurden die Angaben der Befragten aus den Papierfragebögen auf Basis eines Codeplans in ein computerlesbares Format übertragen. Zuvor wurden noch auf den Papierfragebögen numerische Codierungen für die offenen Angaben vermerkt (vgl. Kapitel 6.2) sowie manuelle Vorkorrekturen zur Erleichterung der Datenübertragung vorgenommen (vgl. Kapitel 6.3).

[Erstellung eines Codeplans] Auf Basis des Fragebogens der Befragung wurde ein Codeplan erstellt. Dabei wurde vermerkt, welcher Frage bzw. Teilfrage eine Variable zugeordnet ist, welchen Namen diese Variable trägt und welche numerischen Codierungen für die standardisierten Antworten der Befragten verwendet werden. Um die Erfassungsreihenfolge festzulegen, wurden die Variablen zusätzlich nummeriert.⁴⁴

[Datenerfassung] Für die Datenübertragung wurden der Codeplan, weitere Anweisungen zur Datenerfassung sowie die vorbereiteten Papierfragebögen an einen externen Dienstleister übergeben. Die Erfassung der Angaben erfolgte dort manuell durch Schreibkräfte.

6.2 Codierung offener Angaben

Vor der Datenübertragung erfolgte eine Codierung der (halb-)offenen Angaben. Dabei wurden diesen anhand einer Codierliste numerische Codierungen zugeordnet. Je nach Variable wurden unterschiedliche Codierlisten verwendet. Es handelt sich um Klassifikationsschlüssel der amtlichen Statistik (z. B. Klassifikation der Berufe, Schlüsselverzeichnis der Studenten- und Prüfungsstatistik etc.) oder um bereits in vorherigen Studien eingesetzte Schlüssel. Für einige Variablen wurden neue Codierlisten auf Basis der in den Daten der Befragung des Studienberechtigtenjahrgangs 2008 vorkommenden Nennungen entwickelt. Für einige halboffene Fragen wurden keine neuen Variablen mit numerischen Codierungen erstellt, sondern die Nennungen nur – sofern möglich – den vorhandenen (geschlossenen) Antwortkategorien zugeordnet. Einzelne offene

⁴⁴ Die Daten wurden in einem einfachen spaltenorientierten Textformat ohne eine die Variablennamen enthaltene Kopfzeile erfasst. Der Codeplan musste daher festlegen, in welcher Reihenfolge die Daten zu erfassen waren, damit die zu einer Variable zugehörigen Daten in der richtigen Spalte eingetragen werden konnten.

Fragen wurden nicht vercodet, weil sie vorwiegend als Kontextinformationen zur Codierung anderer offener Angaben erfasst wurden.⁴⁵

In Tabelle 3 sind die codierten Merkmale sowie die jeweils verwendete Codierliste im Überblick dargestellt. Der Datensatz beinhaltet ausschließlich die codierten numerischen Variablen, die offenen Nennungen selbst sind nicht im Datensatz enthalten. Die Ausprägungen der einzelnen Variablen sind im Datensatzreport sowie im Metadaten suchsystem⁴⁶ dokumentiert.

Tabelle 3: Vercodete Merkmale und verwendete Codierlisten im DZHW-Studienberechtigten-panel 2008

| Merkm al | Codierliste | Codierlisten-ID ^a |
|---------------------------------|--|---|
| Schulart (ermittelt) | Projekteigene Codierung | cl-dzhw-6 |
| Schulart | Projekteigene Codierung | cl-dzhw-20 |
| Schulfächer, -gruppen | Projekteigene Codierung | cl-dzhw-7 |
| Bezeichnung (Ausbildungs-)Beruf | Destatis Klassifikation der Berufe 1992 | cl-destatis-kldb-1992 |
| Studienfächer | Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2007/08) | cl-destatis-studienfach-2008 ^b |
| Studienabschluss (2. Welle) | Projekteigene Codierung | cl-dzhw-12 |
| Studienabschluss (3. Welle) | Projekteigene Codierung | cl-dzhw-13 |
| Hochschule | Deutsche Hochschulen: Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2007/08) | cl-destatis-hochschule-2008 ^c |
| | Ausländische Hochschulen: Projekteigene Codierung | cl-dzhw-9 |
| Hochschulart | Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2007/08) | cl-destatis-hochschule-2008 ^c |
| Bundesland | Destatis Bundeslandschlüssel (entsprechend der ersten beiden Ziffern des Amtlichen Gemeindecchlüssels (AGS)) | cl-destatis-bundesland-1990 ^d |
| Staatsbürgerschaft (andere) | Projekteigene Codierung | cl-dzhw-17 |
| Staat (Ausland) | Projekteigene Codierung | cl-dzhw-19 |
| Sonstige offene Abfragen | Projekteigene Codierung oder Zuordnung zu vorgegebenen Kategorien | --- |

^a eine Codierlisten-ID wurde nur dann vergeben, wenn die Kategorien nicht aus den tatsächlichen Nennungen im Datensatz hergeleitet wurden, sondern sich aus einer Systematik ergeben

^b ergänzt um projekteigene Codierung für fehlende Fächer

^c ergänzt um projekteigene Codierung für fehlende Hochschulen

^d ergänzt um projekteigene Codierung für ausländische Staaten

⁴⁵ Dies betrifft die offene Frage nach den auf den Schulabschluss folgenden nächsten Schritten in der ersten Welle (Frage 11) und die offene Abschlussfrage in der zweiten und dritten Welle. Diese Angaben wurden erhoben, um zusätzliche Informationen bei Unklarheiten im Tätigkeitstableau zu erhalten.

⁴⁶ <https://metadata.fdz.dzhw.eu/https://metadata.fdz.dzhw.eu/>

6.3 Datenprüfung und Datenbereinigung

[Manuelle Vorkorrektur] Bereits vor der Übertragung der Daten wurde auf den Papierfragebögen eine manuelle Prüfung und gegebenenfalls eine Anpassung von Angaben der Befragten durchgeführt.⁴⁷ Dies sollte zum einen die Erfassbarkeit der Daten erleichtern. Dafür wurde die Form, aber nicht der Inhalt der bestehenden Angaben verändert. Beispielsweise wurden schwer lesbare Angaben oder Streichungen der Befragten verdeutlicht, Zahlenangaben rechtsbündig in die dafür vorgesehenen Kästchen eingetragen, verbale Angaben von Noten in Ziffern übersetzt (z. B. „gut“ = 2,0) oder Einträge im Tätigkeitstableau auf mehrere Zeilen aufgeteilt, wenn die Befragten verschiedene Phasen in eine Zeile geschrieben hatten (z. B. bei zwei parallelen Studiengängen).

Zum anderen zielte die manuelle Prüfung darauf ab, schon vor der softwaregestützten Korrektur (siehe unten) erste Fehler oder Inkonsistenzen in den Angaben der Befragten zu bereinigen. Dabei wurden Konsistenzprüfungen innerhalb einer Welle durchgeführt, bei denen inhaltliche und zeitbezogene Angaben zu gleichen Themengebieten miteinander abgeglichen wurden. Beispielsweise erfolgte im Fragebogen der dritten Welle ein Abgleich zwischen den Angaben im Tätigkeitstableau (Frage 2.1) und den Angaben zur Berufsausbildung (Fragen 3.1 bis 3.4 und 3.8), den Angaben zu Studienfach, Abschluss und Hochschule (Frage 4.1 bis Frage 5.9) oder den Angaben zur Erwerbstätigkeit (Frage 7.1 und 7.2). Festgestellte Inkonsistenzen wurden, falls möglich, durch den Abgleich mit anderen Nennungen im Fragebogen aufgelöst oder andernfalls ein entsprechender Missingcode (vgl. Kapitel 6.7) vergeben.

[Softwaregestützte Korrektur] Im Anschluss an die Datenübertragung erfolgte eine umfassende Prüfung und Korrektur der Daten mit Hilfe einer DZHW-eigenen Software. Dabei sollten zum einen Fehler bei der vorherigen manuellen Vorkorrektur und Datenübertragung, zum anderen weitere inkonsistente Angaben der Befragten, die bei der Vorkorrektur nicht geprüft werden konnten, identifiziert werden.

Zu diesem Zweck wurden die erfassten Fragebogen-Daten (sowie in der zweiten (und dritten) Befragungswelle auch Daten der ersten (und zweiten) Befragungswelle) in eine Datenbank eingelesen. Anschließend wurden anhand formaler Regeln gültige Wertebereiche und Antwortkombinationen definiert und geprüft. Folgende Typen von Prüfungen wurden vorgenommen:

- *Prüfung von Wertebereichen:* Es wurde geprüft, ob die erfasste Ausprägung einer Variable in dem für diese Variable definierten Wertebereich lag.
- *Prüfung der Einhaltung der Filterführung:* Es wurde auf der einen Seite geprüft, ob aufgrund der im Fragebogen vorgesehenen Filterführung für die jeweilige befragte Person Angaben zu erwarten gewesen wären (Vollständigkeitsprüfung), und auf der anderen Seite, ob für die jeweilige Person keine Angaben hätten erfolgen dürfen (Filterverstöße).⁴⁸

⁴⁷ Die Zahl der vorgenommenen Korrekturen wurde nicht zentral, sondern nur auf den Papierfragebögen dokumentiert und ist daher nicht mehr systematisch rekonstruierbar.

⁴⁸ Die Eingangsfilter der einzelnen Fragen sind im Datensatzreport sowie im Metadatensuchsystem (<https://metadata.fdz.dzhw.eu/>) dokumentiert. Sie definieren für die jeweilige Variable, welche Befragtengruppe die zugehörige Frage beantworten sollte, sofern sie nicht an alle Personen gestellt wurde.

- **Prüfung von Merkmalskombinationen:** Es wurde die Konsistenz der Angaben geprüft. Über die Merkmalskombinationen hinaus, die bereits bei der manuellen Vorkorrektur überprüft wurden, konnten hier auch komplexere Merkmalskombinationen innerhalb der einzelnen Befragungswellen überprüft werden. Zudem wurde hier auch die Konsistenz von Angaben zwischen den Befragungswellen geprüft. Besonders hervorzuheben ist hier der Abgleich zwischen den Angaben zu Ausbildungsschritten in der zweiten Welle (Frage 17 bis 20 und Frage 22) und dem Tätigkeitstableau in der dritten Welle (Frage 2.1).⁴⁹

Insgesamt wurden mehrere hundert Konsistenzregeln definiert und getestet. Bei fehlenden, fehlerhaften oder unplausiblen Werten wurde zunächst mit Hilfe des Papierfragebogens geprüft, ob der entsprechende Wert falsch (bzw. nicht) übertragen worden war. Ansonsten wurde versucht, den korrekten Wert anhand anderer Fragebogen-Angaben zu erschließen.⁵⁰ Im Zweifelsfall wurde ein spezifischer Missingcode vergeben (vgl. Kapitel 6.7). Fehlerkorrekturen wurden dokumentiert⁵¹ und von mindestens einer weiteren Person geprüft.

[Löschung von Fällen] In allen drei Wellen wurden Fälle aus dem Datensatz entfernt. Ein Fall wurde gelöscht, wenn gar keine Angaben zum Bildungsweg gemacht wurden oder wenn zu viele Inkonsistenzen vorhanden waren. Diese Fälle wurden als nicht auswertbar eingestuft und entfernt. Darüber hinaus wurden in der ersten Welle Fälle identifiziert, die nicht zur Grundgesamtheit gehörten.⁵² Diese wurden ebenfalls aus dem Datensatz entfernt. Insgesamt wurden in der ersten Befragungswelle 301 Fälle, in der zweiten Befragungswelle 84 Fälle und in der dritten Befragungswelle acht Fälle gelöscht.

6.4 Generierung von Variablen

Neben den Variablen, die die codierten Antworten der Befragten enthalten, beinhaltet der Datensatz zum Studienberechtigtenjahrgang 2008 auch generierte Variablen. Dabei handelt es sich zum einen um Variablen mit numerischen Codierungen von ursprünglich offenen Nennungen (vgl. Kapitel 6.2). Zum anderen wurden Variablen des Datenschutzes wegen verändert (vgl. Kapitel 8) und im Forschungsfeld häufiger benötigte Variablen aus den Werten einer Quellvariablen generiert (z. B. Aggregation der Studienfächer zu Studienbereichen und Fächergruppen oder Ableitung von Hochschultyp und Hochschulort aus den Hochschulvariablen). Die überwiegende Anzahl der generierten Variablen wurde bereits durch das Primärforschungsprojekt erstellt. Der Variablenname einer generierten Variable ist durch das Suffix „_g#“ gekennzeichnet. Eine Übersicht aller für das Studienberechtigtenpanel 2008 generierten Variablen sowie eine detaillierte Dokumentation der einzelnen Variablen mit Angabe ihrer jeweiligen Ausprägungen und Berechnungsvorschriften findet sich im Datensatzreport sowie im MetadatenSuchsystem⁵³.

⁴⁹ Für das Tätigkeitstableau und die Einzelfragen zu Ausbildungsschritten waren Überschneidungen in den Tätigkeitsangaben zwischen den beiden Wellen bewusst konzipiert worden.

⁵⁰ Inkonsistenzen zwischen den Befragungswellen wurden nur in den Daten der dritten Welle korrigiert, während sie in den vorausgegangenen Erhebungen unkorrigiert blieben.

⁵¹ Die Dokumentation der Fehlerkorrekturen erfolgte handschriftlich auf den Papierfragebögen und ist daher nicht systematisch rekonstruierbar.

⁵² Dies kam beispielsweise vor, wenn Schulabgänger(innen) nur den schulischen Teil der Fachhochschulreife erworben haben ohne anschließend eine volle Fachhochschulreife anzustreben oder die Schulen den Fragebogen an Schüler(innen) eines falschen Jahrgangs verteilt haben.

⁵³ <https://metadata.fdz.dzhw.eu/>

Generierte Variablen wurden im Datensatz – sofern möglich – nach der jeweiligen Ausgangsvariable positioniert. Sofern die Ausgangsvariable aufgrund von Anonymisierung nicht mehr im Datensatz vorhanden ist (vgl. Kapitel 8), nimmt die generierte Variable ihren Platz im Datensatz ein. Wurde eine Variable aus verschiedenen Quellvariablen generiert, wurde sie hinter jene Variable eingefügt, die ihr thematisch am nächsten ist. Falls eine eindeutige Zuordnung (wie bei Gewichtungsvariablen) nicht möglich war, wurde die generierte Variable am Ende des Datensatzes eingefügt. Generierte Variablen im Zuge von Anonymisierungsmaßnahmen wurden bis auf wenige Ausnahmen durch das FDZ erstellt.

6.5 Erstellung der Datensätze

[Zusammenführung der Wellen] Die Daten der drei Befragungswellen wurden zusammengeführt. Die Zuordnung der Daten der drei Befragungswellen erfolgte über die im Rahmen der Feldphase vergebene Identifikationsnummer der Befragten (vgl. Kapitel 4).

[Erstellung des Datensatzes] Die so zusammengeführten Daten wurden in einem Datensatz abgelegt und um die zusätzlich generierten Variablen ergänzt. Pro befragte Person existiert eine Datenzeile (wide-Format). Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehörigen Fragen im Fragebogen.

[Datenformat] Der Datensatz wird standardmäßig im Stata-Format bereitgestellt. Möglichkeiten zur Nutzung weiterer Formate bzw. Analysesoftware sind der Website zu entnehmen (vgl. auch Abschnitt III).

6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels

[Variablen- und Wertelabelvergabe] Für Variablen- und Wertelabels wurden Formulierungen des Fragebogens übernommen oder prägnante Kurzformen dieser Formulierungen gewählt. Dabei basieren die Variablenlabels in der Regel auf dem entsprechenden Fragetext. Grundlage für die Wertelabels sind je nach Fragetyp die Texte der Antwortoptionen bzw. eine Kombination der Texte von Frage und Antwortoption. Bei generierten Variablen, denen bestimmte Klassifikationen zugrunde liegen, wurden für die Wertelabels die Bezeichnungen der Schlüssel der Klassifikation wortgetreu übernommen. Die Variablen- und Wertelabels sind als zweisprachige Labels (Deutsch und Englisch) im gleichen Datensatz hinterlegt.

[Variablenbenennung] Für die Variablenbenennung wurde im FDZ-DZHW eine einheitliche Benennungssystematik erstellt. Mit Ausnahme der Identifikatorvariablen *pid* sowie der Wellenvariablen *wave*⁵⁴ werden die Variablennamen nach einem Präfix-Stamm-Suffix-Schema gebildet. Damit liefern die Variablennamen auch Metainformationen zur entsprechenden Variable. Das Präfix enthält die Wellenkennung anhand eines Buchstabens. Im Stamm geht der Themenbereich, dem die Variable zugeordnet ist, aus einem dreistelligen englischen Buchstabenkürzel hervor. Tabelle 4 stellt die verschiedenen Themenbereiche der Befragung des Studienberechtigtenjahrgangs 2008 sowie die zugehörigen Kürzel für den Stamm des Variablennamens im Überblick dar. Das anhand eines Unterstriches vom Stamm abgetrennte Suffix enthält verschiedene Zusatzinformationen, wie die Kenntlichmachung von generierten Variablen sowie

⁵⁴ Diese enthält die Information, welche Fälle nur an der ersten Welle bzw. an den weiteren Wellen teilgenommen haben.

verschiedenen Datenzugangswegen. Für Indikatoren, die in mehreren Befragungswellen verwendet werden, wurden die Namen der zugehörigen Variablen durch die Vergabe eines identischen Stammes harmonisiert.

Detaillierte Informationen zur Variablenbenennung im Studienberechtigtenpanel 2008 befinden sich im Datensatzreport.

Tabelle 4: Themengebiete und Kürzel für Variablennamen des DZHW-Studienberechtigtenpanels 2008

| Themengebiete-Kürzel | Themengebiet (englisch) | Themengebiet (deutsch) |
|----------------------|-------------------------------|--|
| act | activity | Tätigkeiten |
| dec | decisions | Entscheidungen |
| dem | demographic informations | demografische Informationen |
| eng | engineering | Wahl ingenieurwissenschaftlicher Studiengänge |
| fea | fears | (Studienbezogene) Sorgen |
| fin | financing | Finanzierung |
| fut | future prospects | Zukunftsaussichten |
| goa | goals (occupatioal, life) | (Berufs- und) Lebensziele |
| inc | income | Einkommen |
| inf | information | Information(-sverhalten) |
| int | interests | Interessen |
| job | job | Jobs |
| mot | motives | Tätigkeitsmotive (Studium/Ausbildung/Erwerbstätigkeit) |
| per | personality | Persönlichkeit |
| sat | satisfaction | (Berufs-)Zufriedenheit |
| sch | school | Schulzeit |
| ski | skills | Fähigkeiten |
| stu | studies | Studium |
| sys | system variables | Systemvariablen |
| tra | transition | Übergang (Schule-Beruf) |
| voc | vocational training/education | (Berufs-)Ausbildung |
| wgh | weights | Gewichtungsvariablen |

6.7 Codierung fehlender Werte

Zur Codierung fehlender Werte wurde im FDZ-DZHW eine übergreifende Systematik erstellt, um über verschiedene Datensätze des DZHW hinweg eine einheitliche Missingcodierung gewährleisten zu können. Fehlende Angaben werden dabei durch dreistellige negative Werte codiert. Tabelle 5 stellt die verwendete Missingsystematik dar. Die im Studienberechtigtenpanel 2008 verwendeten Missingcodierungen sind hervorgehoben.

Sie lassen sich vier verschiedenen Gruppen zuordnen. Die erste Gruppe beinhaltet fehlende Werte aufgrund von Nicht-Beantwortung von Fragen seitens der Befragten (Nonresponse), die zweite Gruppe fehlende Werte aufgrund der Filterführung bzw. für Befragte nicht relevante Fra-

gen nicht zutreffend). Die dritte Gruppe beinhaltet Missingcodierungen, die durch das Primärforschungsprojekt oder das FDZ im Zuge der Datenaufbereitung vergeben wurden (editierter fehlender Wert). Zu dieser Gruppe gehören auch die Codierungen, die aufgrund von Anonymisierungsmaßnahmen für bestimmte Variablen gesetzt wurden. Die vierte Gruppe umfasst spezielle Missingcodierungen, die im Rahmen der Datenaufbereitung eines konkreten Datensatzes nur für einzelne Items vergeben wurden (z. B. „Abbruch im weiten Sinn“ bei dem Item `cstu30_g1`, Frage 4.9, 3. Welle).

Tabelle 5: Systematik des FDZ-DZHW für fehlende Werte

| Wertebereich | Code | Wertelabel |
|---|-------------|--|
| -999 bis -990: Nonresponse | -999 | weiß nicht |
| | -998 | keine Angabe |
| | -997 | keine Angabe (Antwortkategorie) |
| | -996 | Interviewabbruch |
| | -995 | keine Teilnahme (Panel) |
| | -994 | verweigert |
| -989 bis -970: Nicht zutreffend | -989 | filterbedingt fehlend |
| | -988 | trifft nicht zu |
| | -987 | designbedingt fehlend (Fragebogensplit) |
| | -986 | designbedingt fehlend (Welle) ^a |
| | -985 | designbedingt fehlend (Kohorte) ^b |
| -969 bis -950: Editierter fehlender Wert | -969 | unbekannter fehlender Wert^c |
| | -968 | unplausibler Wert^d |
| | -967 | anonymisiert |
| | -966 | nicht bestimmbar^e |
| | -965 | ungültige Mehrfachnennung |
| -949 bis -930: Item-spezifische fehlende Werte^f | -949 | Zweitstudium bereits aufgenommen |
| | -948 | Doppelstudium |
| | -947 | Abbruch eines Studiums in weiter Definition (BA/VerFH-Studium) |
| | -946 | Studium ohne nähere Angabe |
| | -945 | Gewählt ohne Reihenfolge |
| | -944 | kein Studium an Uni/PH/KunstHS/Theol.HS/FH/VerwFH, aber Studium an BA |
| | -943 | kein Uni/FH-Studium, aber Studium im weiteren Sinne (VerwFH/BA) |
| -929 bis -920: Andere fehlende Werte | -929 | Datenverlust |

^a dieser Wert wird nur für Datensätze im Long-Format vergeben

^b dieser Wert wird nur in gepoolten Datensätzen vergeben

^c dieser Wert wird vergeben, wenn keinerlei Ursache rekonstruiert werden kann

^d Angaben, die aufgrund unterschiedlicher Faktoren in der Codierphase als nicht plausibel eingestuft werden, erhalten diesen Wert. Eine exakte Rekonstruktion ist ggf. nicht mehr möglich

^e diese Kategorie wird vergeben, wenn eine eindeutige Codierung nicht möglich ist, z. B. offene Angabe, die nicht vercodet werden konnte, da sie nicht lesbar ist

^f die Werte dieser Kategorie sind definitionsgemäß für jeden Datensatz spezifisch

7 Gewichtung

Die Gewichtung der Daten dient dem Ausgleich von Verzerrungen der Stichprobe im Vergleich zur definierten Grundgesamtheit. Es folgt zunächst eine allgemeinere Einführung in die Vorgehensweise und die Darstellung der erstellten Gewichte. Im Anschluss wird die Gewichtungsprozedur im Detail beschrieben.

7.1 Vorgehen und Anwendungshinweise

[Ursachen für die Verzerrungen der Stichprobe] Maßgeblich für die Verzerrungen der Stichprobe sind zwei Prozesse:

- **Designbedingte Verzerrung:** Disproportionalitäten werden bewusst erzeugt, um in bestimmten relevanten Subgruppen die Fallzahlen zu erhöhen (vgl. Kapitel 3).
- **Verzerrung durch Nonresponse:** Ausfallprozesse (z. B. Nichtteilnahmen, fehlende Erreichbarkeit, Verlust auf dem Postweg) führen zu einem verringerten Rücklauf und somit zu einer Differenz zwischen Brutto- und Nettostichprobe (vgl. Kapitel 5). Wenn diese Ausfallprozesse unsystematisch sind (Missing Completely at Random), können sie ignoriert werden.⁵⁵ Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (Missing at Random, Not Missing at Random), der einer Modellierung bedarf.⁵⁶

[Konzeptuelles Vorgehen] Im Zuge einer Gewichtungsprozedur sollten idealerweise zunächst designbedingte Disproportionalitäten ausgeglichen werden. Die hierfür benötigten *Designgewichte* ergeben sich bei zufallsgesteuerten Auswahlverfahren direkt aus dem Stichprobendesign. Im Anschluss sollte eine Adjustierung der Designgewichte mit Hilfe von *Nonresponsegewichten* im Quer- und Längsschnitt erfolgen, die auf der Grundlage von Informationen über Teilnehmer(innen) und Nichtteilnehmer(innen) auf Individualebene erzeugt werden. In einem letzten Schritt können die nonresponse-adjustierten Designgewichte anhand von Merkmalsverteilungen aus der Grundgesamtheit kalibriert werden (Kalibrierung).

Vor dem Hintergrund der Erläuterungen zur Grundgesamtheit und zum Stichprobenverfahren in Kapitel 3 wird ersichtlich, dass ein derartiges idealtypisches Vorgehen für die Daten des Studienberechtigtenpanels 2008 nicht durchführbar ist. Das Stichprobendesign ermöglicht keine Ableitung von exakten Auswahlwahrscheinlichkeiten. Da in der ersten Befragungswelle zudem keine Informationen über Nichtteilnehmer(innen) vorliegen, können hier auch keine individuellen Ausfallgewichte gebildet werden. Als Querschnittsgewicht für die erste Welle wird daher ein geschätztes Designgewicht berechnet, das anhand von Informationen aus der Grundgesamtheit kalibriert wird. Aus der Merkmalsverteilung der Grundgesamtheit werden zugleich Informationen zu Nichtteilnehmer(inne)n gewonnen, sodass hiermit ebenfalls eine Form der Nonresponse-Adjustierung erfolgt. Für die zweite und dritte Welle wird zusätzlich ein Nonresponsegewicht berechnet, das mit Hilfe von Informationen aus den jeweiligen vorangegangenen Wellen die Nicht-

⁵⁵ Insofern die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

⁵⁶ Siehe grundlegend zu den unterschiedlichen Formen von Ausfallprozessen Rubin (1976).

teilnahme in der zweiten bzw. dritten Welle modelliert. In Tabelle 6 sind die im Datensatz bereitgestellten Gewichte dargestellt.

Tabelle 6: Bereitgestellte Gewichte zum DZHW-Studienberechtigtenpanel 2008

| Variablenname | Beschreibung |
|---------------|---|
| wgt_t1 | Querschnittsgewicht Welle 1 |
| wgt_t1t2 | Längsschnittgewicht 2-Wellen-Panel t1t2 |
| wgt_t1t2t3 | Längsschnittgewicht 3-Wellen-Panel t1t2t3 |

[Hinweise zur Anwendung der Gewichte] Bei den erstellten Gewichten handelt es sich um *probability weights*, die in Stata mit Hilfe ado-spezifischer Optionen berücksichtigt werden können.⁵⁷ Das Gewicht wgt_t1 ist für Auswertungen der ersten Welle vorgesehen, die Gewichte wgt_t1t2 und wgt_t1t2t3 für Auswertungen des Zwei- bzw. Dreiwellenpanels. Grundlegend ist zu beachten, dass Gewichte nur für jene Analysen sinnvolle Korrekturgrößen darstellen, in denen die zur Gewichtung verwendeten Variablen in dem Analysemodell enthalten sind oder mit diesen in einem Zusammenhang stehen. Aus diesem Grund sollten Gewichte immer mit Bedacht und Fokus auf die analysierte Fragestellung verwendet werden. Im Folgenden wird die Vorgehensweise bei der Erstellung der Gewichte für den Datensatz näher dargestellt.

7.2 Gewichtung des Datensatzes

[Querschnittsgewichtung] Aufgrund des Stichprobendesigns konnten keine exakten Designgewichte abgeleitet werden. Diese mussten vielmehr geschätzt werden. Die Schätzung des Designgewichts erfolgte in jeder Schicht s wie folgt⁵⁸:

$$\widehat{dwgt}_{sci} = \frac{n_{cs}^{-1}}{N_{cs}}$$

Aufgrund der fehlenden Informationen über die Nichtteilnehmer(innen) in der ersten Welle, war auch für den Ausfallprozess durch Nichtteilnahmen (Nonresponse) keine umfängliche Adjustierung der geschätzten Designgewichte auf Individualebene möglich. Jedoch wurde eine Kalibrierung vorgenommen, welche die Anpassung der geschätzten Designgewichte an Merkmalsträger der Grundgesamtheit zum Ziel hatte. Als Merkmale wurden das Geschlecht, der Schulzweig und das Bundesland herangezogen.⁵⁹ Da die Merkmalsträger in der Grundgesamtheit auch Informationen über Nichtteilnehmer(innen) enthielten, erfolgt durch die Kalibrierung zusätzlich eine Nonresponse-Adjustierung im Hinblick auf die verwendeten Merkmale. Die Kalibrierung des geschätzten Designgewichts \widehat{dwgt}_{sci} wurde mittels des Raking-Algorithmus⁶⁰ durchgeführt. Daraus ergibt sich das Querschnittsgewicht der ersten Welle des Studienberechtigtenpanels 2008 wgt_{t1i} . Für die zweite und dritte Erhebungswelle wurde kein Querschnittsgewicht erstellt, da

⁵⁷ Siehe hierzu die Stata-Hilfe (Befehl: *help weights*).

⁵⁸ Wobei n_s der Anzahl der Individuen in einer Schicht und N_s der Anzahl der Individuen in der jeweiligen Schicht der Grundgesamtheit entspricht.

⁵⁹ Die Informationen aus der Grundgesamtheit wurden aus Daten des Statistischen Bundesamtes (2009a); (2009b) abgeleitet.

⁶⁰ Raking ist auch unter dem Begriff des iterative proportional fitting (ipf) bekannt (vgl. Kolenikov, 2014).

weder neu hinzugekommene Individuen befragt wurden (Auffrischungsstichprobe), noch Personen die Möglichkeit hatten, an Welle zwei oder drei teilzunehmen, wenn sie nicht Teil der vorangegangenen Erhebungswelle waren.

[Längsschnittgewichtung] Für die Gewichtung des Zwei- bzw. Drei-Wellen-Panels mussten die Ausfallprozesse im Zeitverlauf berücksichtigt werden (Panelmortalität; siehe auch Kapitel 5). Zu diesem Zweck wurde jeweils ein Ausfallgewicht errechnet, das die Wahrscheinlichkeit abbildete, an der nächsten Welle teilzunehmen. Im Unterschied zu der Nonresponse-Adjustierung in der ersten Welle lagen für die Nichtteilnehmer(innen) der Wellen zwei und drei eine Vielzahl an Informationen aus den jeweils vorangegangenen Wellen vor (σ_t). Diese dienten als Kovariaten in einem Probit-Regressionsmodell, das die Vorhersage der Teilnahmewahrscheinlichkeit im folgenden Zeitpunkt $P(R_{t+1})$ zum Ziel hatte. Im Falle des Vorliegens fehlender Werte wurden diese jeweils als zusätzliche Kategorien der Variablen in das Modell aufgenommen, sodass auch Fälle mit Item-Nonresponse in das Modell inkludiert werden konnten. Hierdurch konnte weiterhin die Annahme geprüft werden, dass Item-Nonresponse einen bedeutsamen Prädiktor für Unit-Nonresponse in künftigen Wellen darstellt. Eine Reihe von Prädiktoren erwies sich als bedeutsam für die Vorhersage der Teilnahmewahrscheinlichkeit in der zweiten bzw. dritten Welle. Aus dem Modell konnte die bedingte Teilnahmewahrscheinlichkeit abgeleitet werden, deren Kehrwert das Ausfallgewicht für die zweite bzw. dritte Welle darstellt⁶¹:

$$NR_{gew_{t+1_i}} = P(Res_{t+1_i} | \sigma_{t_i})^{-1}$$

Das Gesamtgewicht (nonresponse-adjustierte Designgewicht) für das jeweilige Teilpanel ergibt sich aus dem Produkt des geschätzten Designgewichts und der jeweiligen Nonresponsegewichte:

$$wgt_{i_{t1t2}} = \widehat{dwgt}_{sci} \times NR_{i_{gew_{t2}}}$$

$$wgt_{i_{t1t2t3}} = \widehat{dwgt}_{sci} \times NR_{i_{gew_{t2}}} \times NR_{i_{gew_{t3}}}$$

Im Anschluss wurde das jeweilige Gesamtgewicht mittels des Raking-Algorithmus auf Merkmale der Grundgesamtheit kalibriert. Hierfür wurden dieselben Merkmale verwendet, wie auch für die Kalibrierung des Designgewichts in der ersten Erhebungswelle.

[Normierung auf die Fallzahl der Stichprobe] Die errechneten Gewichte wurden, wie in der sozialwissenschaftlichen Forschungspraxis üblich, auf die Fallzahl der Stichprobe normiert.

[Trimming der Gewichte] Die initial berechneten Gewichte wiesen einen kleinen Teil an Gewichtungsfaktoren auf, die Ausreißer darstellten. Um diese zu beseitigen, wurden alle Gewichte einer Trimmung nach Potter (1990) (vgl. auch Valliant, Dever & Kreuter, 2013, S. 388f.) unterzogen. Dem Verfahren liegt die Annahme zugrunde, dass die Gewichte einer Wahrscheinlichkeitsverteilung (Betaverteilung) folgen. All jene Gewichte, die über dem 99-Prozent-Quantil lagen, wurden auf diese Grenze trunziert. Der Überschuss jenseits der Trunkierung wurde im Folgenden unter den verbleibenden Gewichten verteilt.

⁶¹ Das Verfahren entspricht seiner Logik nach dem *Propensity Score Matching*, das auf Rosenbaum und Rubin (1983) zurückgeht, siehe auch Blumenstiel und Gummer (2015).

8 Anonymisierung

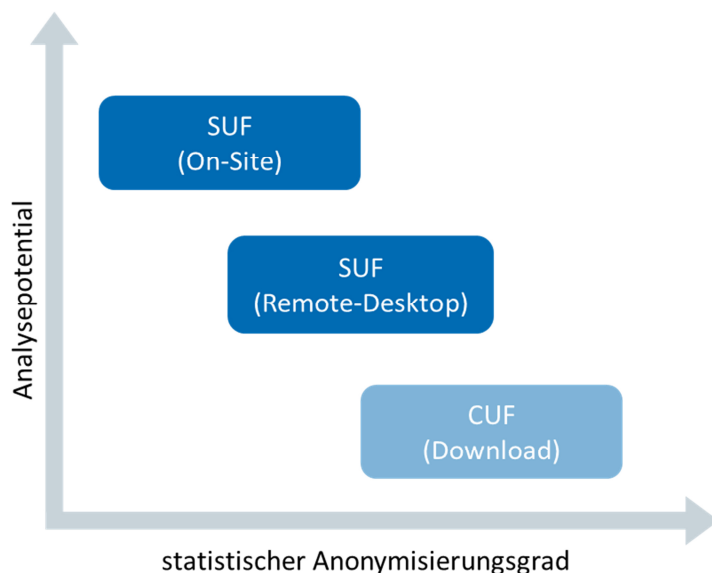
[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten, die in freiwilligen Befragungen durch das DZHW erhoben werden, gilt das Bundesdatenschutzgesetz (BDSG).⁶² Danach dürfen personenbezogene Daten, die im Rahmen wissenschaftlicher Forschung erhoben worden sind, ausschließlich zum Zweck wissenschaftlicher Forschung verarbeitet oder genutzt werden (vgl. §40 Abs. 1 BDSG). Darüber hinaus sind personenbezogene Daten im Forschungskontext zum Schutz der Befragten zu anonymisieren (vgl. §40 Abs. 2 BDSG). Im BDSG wird der Vorgang der Anonymisierung definiert als „das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können.“ (§3 Abs. 6 BDSG). Das heißt, für die Weitergabe von Daten aus wissenschaftlichen Forschungsprojekten an Dritte sind die Daten entweder absolut zu anonymisieren, sodass kein Bezug zur Person mehr hergestellt werden kann, oder mindestens faktisch zu anonymisieren, sodass die Herstellung eines Personenbezugs mit einem unverhältnismäßig hohen Aufwand verbunden und daher die Wahrscheinlichkeit der Re-Identifikation einer Person minimal ist.

[Datenzugang, Anonymisierungsgrad und Analysepotential] Das FDZ-DZHW stellt für das Studienberechtigtenpanel 2008 ein faktisch anonymisiertes SUF für die wissenschaftliche Sekundärnutzung zur Verfügung. Die Anonymität der Befragten wird dabei über eine Kombination aus statistischen Maßnahmen und technischen Zugriffsbeschränkungen sichergestellt. Je stärker der Datenzugang technisch kontrolliert wird, desto geringer ist das Risiko einer De-Anonymisierung der Daten, desto weniger müssen die Daten mittels statistischer Maßnahmen um Informationen reduziert werden und desto größer bleibt ihr Analysepotential.

Die SUF werden über zwei verschiedene Zugangswege angeboten: Remote-Desktop und On-Site (für weiterführende Informationen vgl. Abschnitt III). Für jeden Zugangsweg wird eine andere SUF-Variante bereitgestellt, die unterschiedlich stark anonymisiert worden ist und entsprechend weniger oder mehr Informationen umfasst. Abbildung 4 gibt einen Überblick über den jeweiligen Grad der statistischen Anonymisierung und dem damit verbundenen Analysepotential. Im Folgenden werden die durchgeführten statistischen Anonymisierungsmaßnahmen in Abhängigkeit von Datenprodukt (SUF) und Zugangsweg erläutert.

⁶² Das BDSG kommt zur Anwendung, da die DZHW GmbH juristisch als öffentliche Stelle des Bundes betrachtet wird (vgl. § 2 Abs. 3 BDSG). Der Bund hält die absolute Mehrheit der Anteile der DZHW GmbH und das Institut erfüllt Aufgaben der öffentlichen Verwaltung des Bundes im weitesten Sinn. Zur Auslegung einzelner rechtlicher Aspekte wird ergänzend die Europäische Datenschutzrichtlinie hinzugezogen.

Abbildung 4: Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten des DZHW-Studienberechtigtenpanels 2008



[Statistische Anonymisierungsmaßnahmen] Im Rahmen der Anonymisierung sind zunächst alle Informationen, mit denen sich Personen oder Institutionen direkt identifizieren lassen, zu löschen. Diese sogenannten *direkten Identifikatoren*, wie Namen, Adressen und E-Mail Adressen, wurden im Studienberechtigtenpanel 2008 bereits während der Feldphase in einem separaten Datensatz erfasst (vgl. Kapitel 4) und nicht in den verschiedenen SUF-Varianten enthalten. Um einen Rückbezug auf diesen Datensatz zu unterbinden, wurden zudem die Original-Identifikationsnummern der Befragten sowie der Institutionen (Schule, Hochschule) entfernt und durch neue zufällig vergebene Identifikationsnummern ersetzt.

Zur Gewährleistung der faktischen Anonymisierung der Daten des SUF wurden die *Quasi-Identifikatoren* bestimmt, also Informationen, die in Kombination oder durch die Anspielung externer Informationen geeignet sind, eine Person indirekt zu identifizieren.⁶⁴ Für das Studienberechtigtenpanel 2008 wurden die folgenden Quasi-Identifikatoren identifiziert, die sowohl in externen Datenquellen⁶⁵ als auch in den Studienberechtigtenpanelen vorliegen: regionale Informationen (Geburtsland, Ort des Erwerbs der Studienberechtigung, Hochschule und Arbeitsort), Staatsangehörigkeit, Sprache im Elternhaus, Schulart, schulische Prüfungsfächer, Hochschule, Studienfach, Abschlussart, Berufsangaben. Um eine eindeutige Zuordnung der Studienberechtigtenpanelen zu unterbinden, wurden diese Schlüsselmerkmale – je nach Datenform bzw. Zugangsweg – aggregiert oder gelöscht (vgl. Tabelle 7). Beispielsweise ist die Sprache im Elternhaus On-Site ohne Einschränkungen verfügbar, während im Remote-Desktop-SUF 13 Sprachen einzeln ausgewiesen werden und der Rest zu „sonstige Sprachen“ zusammengefasst wird. Offene Anga-

⁶⁴ Dabei ist darauf hinzuweisen, dass die Identifikation einer Person bereits durch die Stichprobenauswahl erschwert wird, da eine Ungewissheit darüber besteht, ob eine befragte Person eine einzigartige Merkmalskombination in der Population aufweist.

⁶⁵ Z. B. Studenten- und Prüfungsstatistik des statistischen Bundesamtes, Alumninetzwerke der Hochschulen oder auch Berufsnetzwerke.

ben sind ebenfalls Quasi-Identifikatoren (vgl. Ebel, 2015, S. 3). Sie wurden bereits vor der Datenerfassung vercodet oder nicht wortgetreu übertragen (siehe Kapitel 6.2).

Zuletzt wurde geprüft, ob in den Daten *sensible Informationen*, z. B. zur Gesundheit, sexuellen Orientierung und zu politischen Einstellungen, enthalten waren. Diese eignen sich zwar nicht zur Re-Identifikation von Individuen oder Institutionen, jedoch können die Informationen im Falle einer De-Anonymisierung nutzbringend sein (vgl. Koberg, 2016, S. 694) und sind daher besonders schützenswert (vgl. §3 Abs. 9 BDSG, Art. 8 Abs. 1 und 2a EG-DSRL). Im Studienberechtigtenpanel 2008 wurden Gesundheitsinformationen erhoben, für die bei den Befragten kein zusätzliches Einverständnis für die Sekundärnutzung eingeholt wurde. Daher wurden diese Antworten in allen SUF-Varianten gelöscht.

Tabelle 7: Maßnahmen der statistischen Anonymisierung der Daten des DZHW-Studienberechtigtenpanels 2008 nach Zugangsweg⁶⁶

| Merkmal | On-Site-SUF | Remote-Desktop-SUF |
|---|---|---|
| Direkte Identifikatoren | Löschung und Vergabe einer zufälligen ID für Schulen und Schüler(innen) | Löschung und Vergabe einer zufälligen ID für Schulen und Schüler(innen) |
| Schulort | Aggregation zum Bundesland | Aggregation zum Bundesland |
| Schulart | Aggregation von geringbesetzten Schularten | Aggregation zu allgemeinbildenden und beruflichen Schulen |
| Prüfungsfächer/Schwerpunktfach (Schule, Berufsschule) | Freigabe | Aggregation zu (Berufs-)Schul-Fächergruppen |
| Studienfach | Freigabe | Aggregation zu Studienbereichen ^a |
| Hochschule | Information zur Hochschulart | Information zur Hochschulart ^b |
| Hochschulort, Aufenthaltsort im Dezember 2008 | Aggregation zum Bundesland und Staat einer ausl. Hochschule | Aggregation zum Bundesland und Staat einer ausl. Hochschule ^c |
| Arbeitsort (Postleitzahl) | Freigabe | Reduktion auf zwei Stellen |
| Beruf | Freigabe | Aggregation zu Berufsordnungen ^d |
| Staatsangehörigkeit | Freigabe | 5 Staaten einzeln ausgewiesen, ansonsten Aggregation (Staatengruppen, Weltregionen) |
| Sprache im Elternhaus | Freigabe | 13 Sprachen einzeln ausgewiesen, andere Sprachen zu „sonstige Sprachen“ aggregiert |
| Merkmale zur Gesundheit | Löschung | Löschung |

^a nach Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2007/08 von Destatis

^b nur Unterscheidung von Fachhochschule und Universität (inklusive Pädagogische Hochschulen, Theologische Hochschulen, Kunst- und Musikhochschulen) und nicht für das Saarland (nur eine Universität)

^c nur Staaten im Ausland mit mehr als einer Hochschule

^d nach Klassifikation der Berufe von 1992 von Destatis

⁶⁶ Detaillierte Informationen zu den anonymisierten Variablen sind dem Datensatzreport sowie dem MetadatenSuchsystem (<https://metadata.fdz.dzhw.eu/>) zu entnehmen.

9 Literaturverzeichnis

- Baier, T. & Helbig, M. (2014). Much ado about € 500: do tuition fees keep German students from entering university? Evidence from a natural experiment using DiD matching methods. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 20 (2), 98–121. doi:10.1080/13803611.2014.881745
- Blumenstiel, J. E. & Gummer, T. (2015). Prävention, Korrektur oder beides? Drei Wege zur Reduzierung von Nonresponse Bias mit Propensity Scores. In J. Schupp & C. Wolf (Hrsg.), *Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen* (S. 13–44). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-10459-7
- Ebel, T. (2015). *Empfehlungen zur Anonymisierung quantitativer Daten*. Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften.
- Häder, M. (2015). *Empirische Sozialforschung. Eine Einführung* (3. Aufl.). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-531-19675-6
- Heine, C. & Quast, H. (2009). *Studierneigung und Berufsausbildungspläne. Studienberechtigte 2008 ein halbes Jahr vor Schulabgang* (HIS Hochschul-Informationen-System GmbH, Hrsg.) (HIS: Forum Hochschule 4/2009). Hannover: HIS.
- Heine, C., Quast, H. & Beuße, M. (2010). *Studienberechtigte 2008 ein halbes Jahr nach Schulabschluss. Übergang in Studium, Beruf und Ausbildung* (HIS Hochschul-Informationen-System GmbH, Hrsg.) (HIS: Forum Hochschule 3/2010). Hannover: HIS.
- Helbig, M., Baier, T. & Kroth, A. J. (2012). Die Auswirkung von Studiengebühren auf die Studierneigung in Deutschland. Evidenz aus einem natürlichen Experiment auf Basis der HIS-Studienberechtigtenbefragung. *Zeitschrift für Soziologie*, 41 (3), 227–246.
- Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. (2012). *FDZ-Methodenreport. Datenschutz am Forschungsdatenzentrum* (Nr. 06). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer & J. Skopek (Hrsg.), *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study* (S. 691–708). Wiesbaden: Springer VS. doi:10.1007/978-3-658-11994-2
- Kolenikov, S. (2014). Calibrating survey data using iterative proportional fitting (raking). *The Stata Journal*, 14 (1), 22–59.
- Lane, J., Heus, P. & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. *Transactions on Data Privacy*, 1 (1), 2–16.
- Lörz, M. (2012). Mechanismen sozialer Ungleichheit beim Übergang ins Studium: Prozesse der Status- und Kulturreproduktion. In R. Becker & H. Solga (Hrsg.), *Soziologische*

- Bildungsforschung* (Kölner Zeitschrift für Soziologie und Sozialpsychologie: Sonderheft, Bd. 52, S. 302–324). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-658-00120-9_13
- Lörz, M., Quast, H. & Roloff, J. (2015). Konsequenzen der Bologna-Reform: Warum bestehen auch am Übergang vom Bachelor- ins Masterstudium soziale Ungleichheiten? *Zeitschrift für Soziologie*, 44 (2), 137–155.
- Mertens, D. (1976). Beziehungen zwischen Qualifikation und Arbeitsmarkt. In W. Schläffke (Hrsg.), *Jugendarbeitslosigkeit. Unlösbare Aufgabe für das Bildungs- und Beschäftigungssystem?* (S. 68–117). Köln: Dt. Inst.-Verl.
- Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. *Proceedings of the Survey Research Methods Section*, 225–230.
- Prüfer, P. & Rexroth, M. (2000). Zwei-Phasen-Pretesting. In P. P. Mohler & P. Lüttinger (Hrsg.), *Querschnitt. Festschrift für Max Kaase* (S. 203–219). Mannheim: ZUMA.
- Quast, H., Scheller, P. & Lörz, M. (2014). *Bildungsentscheidungen im nachschulischen Verlauf. Dritte Befragung der Studienberechtigten 2008 viereinhalb Jahre nach Schulabschluss* (HIS Hochschul-Informationen-System GmbH, Hrsg.) (HIS: Forum Hochschule 9/2014). Hannover: HIS.
- Quast, H. & Spangenberg, H. (2008). *Mehrmalige Befragung der Studienberechtigten 2007/08. Erste Befragung. Sachstand März 2008 (unveröffentlicht)*. Forschungsprojekt im Auftrag des Bundesministeriums für Bildung und Forschung (HIS Hochschul-Informationen-System GmbH, Hrsg.) (HIS: Sachstandsbericht). Hannover: HIS Hochschul-Informationen-System GmbH.
- Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 41–55. doi:10.2307/2335942
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63 (2), 581–592.
- Schneider, H. & Franke, B. (2014). *Bildungsentscheidungen von Studienberechtigten. Studienberechtigte 2012 ein halbes Jahr vor und ein halbes Jahr nach Schulabschluss* (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Hrsg.) (Forum Hochschule Nr. 6). Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).
- Schnell, R., Hill, P. B. & Esser, E. (2005). *Methoden der empirischen Sozialforschung* (7. Aufl.). München: Oldenbourg.
- Statistisches Bundesamt. (2009a). *Bildung und Kultur. Allgemeinbildende Schulen. Schuljahr 2008/09* (Statistisches Bundesamt, W., Hrsg.) (DESTATIS. wissen. nutzen. Nr. 1), Wiesbaden
- Statistisches Bundesamt. (2009b). *Bildung und Kultur. Berufliche Schulen. Schuljahr 2008/09* (Statistisches Bundesamt, W., Hrsg.) (DESTATIS. wissen. nutzen. Nr. 2), Wiesbaden

Valliant, R., Dever, J. A. & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York (NY): Springer New York. doi:10.1007/978-1-4614-6449-5

Anhang 1: Probit-Regression zur Erstellung des Panelausfallgewichts in Welle 2

| Variable (AV: Teilnahme Welle 2) | b-Koeffizient | Z-Wert |
|--|----------------------|---------------|
| Geschlecht (Referenz: weiblich) | -0,286 | -15,7900 |
| Mutter: Hochschulabschluss (Referenz: ja) | | |
| nein | 0,065 | 3,1700 |
| keine Angabe | -0,319 | -4,5700 |
| Vater: Hochschulabschluss (Referenz: ja) | | |
| nein | 0,058 | 2,8900 |
| keine Angabe | -0,115 | -1,8900 |
| Studienabsicht (Referenz: ja) | | |
| nein | -0,174 | -8,4800 |
| keine Angabe | -0,644 | -2,5000 |
| Aufenthaltort: Bundesland (Referenz: Nordrhein-Westfalen) | | |
| Schleswig-Holstein | 0,118 | 1,9300 |
| Hamburg | -0,138 | -2,0700 |
| Niedersachsen | 0,116 | 2,9600 |
| Bremen | -0,109 | -1,5500 |
| Hessen | 0,065 | 1,6600 |
| Rheinland-Pfalz | 0,134 | 3,1200 |
| Baden-Württemberg | 0,060 | 1,8300 |
| Bayern | 0,255 | 7,2100 |
| Saarland | -0,241 | -2,8400 |
| Berlin | -0,172 | -3,2400 |
| Brandenburg | -0,071 | -1,4600 |
| Mecklenburg-Vorpommern | -0,054 | -1,2600 |
| Sachsen | -0,096 | -2,0100 |
| Sachsen-Anhalt | -0,055 | -1,2800 |
| Thüringen | 0,050 | 1,0600 |
| Schulform (Referenz: Gymnasium) | | |
| Gesamtschule | -0,009 | -0,2200 |
| Freie Waldorfschule | 0,330 | 1,9400 |
| Abendgymnasium | 0,025 | 0,2800 |
| Kolleg (nicht Berufskolleg) | 0,041 | 0,6400 |
| Fachgymnasium, Berufliches Gymnasium | -0,159 | -4,8800 |
| Gymnasiale Oberstufe einer Berufsfachschule | -0,216 | -2,2900 |
| Gymnasiale Oberstufe eines Oberstufenzentrums | -0,062 | -0,3200 |
| Berufsoberschule | -0,281 | -5,4000 |
| Fachoberschule | -0,428 | -14,6400 |
| (Höhere) Berufsfachschule | -0,423 | -9,2600 |
| Fachschule | -0,679 | -12,8500 |
| Fachakademie | -0,563 | -2,8600 |
| Abschlussnote (gerundet) (Referenz: 3,0 bis 3,9) | | |
| Note 1,0 bis 1,9 | 0,686 | 17,6000 |
| Note 2,0 bis 2,9 | 0,315 | 16,0400 |
| Note 4,0 bis 4,9 | -0,313 | -4,2200 |
| Note 5,0 bis 6,0 | -0,767 | -2,5000 |
| keine Angabe | -0,159 | -3,0200 |
| Konstante | -0,752 | -25,3500 |
| Anzahl der Beobachtungen | 27.997 ⁶⁷ | |

⁶⁷ Aufgrund von fehlenden Werten auf der Variable Geschlecht können für 185 Fälle aus der Bruttostichprobe keine Teilnahmewahrscheinlichkeiten erzeugt werden. Dies betrifft jedoch nur Fälle, die nicht an der zweiten Welle teilnehmen, daher erhalten alle 5.933 Fälle aus dem Zwei-Wellen-Panel ein Gewicht.

Anhang 2: Probit-Regression zur Erstellung des Panelausfallgewichts in Welle 3

| Variable (AV: Teilnahme Welle 3) | b-Koeffizient | Z-Wert |
|--|----------------------|---------------|
| Geschlecht (Referenz: weiblich) | -0,184 | -5,0200 |
| Vater: Hochschulabschluss (Referenz: ja) | | |
| nein | 0,001 | 0,0300 |
| keine Angabe | -0,348 | -2,9200 |
| Geburtsland: Deutschland (Referenz: ja) | | |
| nein | 0,154 | 1,7300 |
| keine Angabe | 0,312 | 0,7800 |
| Aufenthaltort: Bundesland (Referenz: Nordrhein-Westfalen) | | |
| Schleswig-Holstein | -0,022 | -0,1900 |
| Hamburg | 0,034 | 0,3100 |
| Niedersachsen | -0,030 | -0,4000 |
| Bremen | -0,098 | -0,6900 |
| Hessen | -0,090 | -1,1400 |
| Rheinland-Pfalz | 0,027 | 0,3100 |
| Baden-Württemberg | -0,033 | -0,5400 |
| Bayern | 0,022 | 0,3400 |
| Saarland | -0,171 | -0,9400 |
| Berlin | -0,198 | -2,1200 |
| Brandenburg | 0,237 | 1,9900 |
| Mecklenburg-Vorpommern | -0,215 | -2,1900 |
| Sachsen | 0,042 | 0,4600 |
| Sachsen-Anhalt | -0,230 | -2,4900 |
| Thüringen | -0,043 | -0,4300 |
| Ausland | 0,032 | 0,3500 |
| keine Angabe | -0,140 | -0,4900 |
| Schulform (Referenz: Gymnasium) | | |
| Gesamtschule | -0,210 | -2,6800 |
| Freie Waldorfschule | -0,311 | -1,1100 |
| Abendgymnasium | -0,401 | -2,5400 |
| Kolleg (nicht Berufskolleg) | -0,454 | -3,9600 |
| Fachgymnasium, Berufliches Gymnasium | -0,160 | -2,5400 |
| Gymnasiale Oberstufe einer Berufsfachschule | -0,073 | -0,3800 |
| Gymnasiale Oberstufe eines Oberstufenzentrums | -0,522 | -1,3300 |
| Berufsoberschule | -0,131 | -1,3600 |
| Fachoberschule | -0,281 | -4,6200 |
| (Höhere) Berufsfachschule | -0,431 | -4,3500 |
| Fachschule | -0,255 | -2,1500 |
| Fachakademie | -0,657 | -1,7000 |
| Studiengebühren (Referenz: ja) | | |
| nein | 0,096 | 2,2000 |
| keine Angabe | -0,115 | -0,6700 |
| Elternhaus: Anzahl Bücher (Referenz: 201 bis 300) | | |
| 0 bis 50 | -0,148 | -2,2000 |
| 51 bis 100 | -0,003 | -0,0400 |
| 101 bis 200 | 0,001 | 0,0200 |
| 301 bis 500 | 0,005 | 0,0900 |
| 501 bis 1000 | 0,097 | 1,5500 |
| 1000 und mehr | -0,039 | -0,5600 |
| keine Angabe | 0,126 | 0,6900 |
| Geburtsland Vater: Deutschland (Referenz: ja) | | |
| nein | -0,247 | -3,6700 |
| keine Angabe | -0,547 | -2,5100 |
| Abschlussnote Mathematik (gerundet) (Referenz: 3,0 bis 3,9) | | |

| | | |
|---|--------|---------|
| Note 1,0 bis 1,9 | 0,261 | 4,8900 |
| Note 2,0 bis 2,9 | 0,170 | 3,6300 |
| Note 4,0 bis 4,9 | -0,065 | -1,1300 |
| Note 5,0 bis 6,0 | -0,147 | -2,0200 |
| keine Angabe | -0,077 | -0,3600 |
| Abschlussnote Deutsch (gerundet) (Referenz: 3,0 bis 3,9) | | |
| Note 1,0 bis 1,9 | -0,062 | -1,1100 |
| Note 2,0 bis 2,9 | -0,044 | -1,0400 |
| Note 4,0 bis 4,9 | -0,094 | -1,4100 |
| Note 5,0 bis 6,0 | -0,323 | -2,0500 |
| keine Angabe | -0,066 | -0,3100 |
| Konstante | 0,428 | 5,4500 |
| Anzahl der Beobachtungen | 5.933 | |