

Florence Baillet / Andreas Franken / Anne Weber

DZHW-Absolventenpanel 2009

Daten- und Methodenbericht zu den Erhebungen der
Absolvent(inn)enkohorte 2009 (1. und 2. Befragungswelle)

Version 1.0.0

Daten- und Methodenbericht

Juni 2017

Autor(inn)en:
Florence Baillet
Andreas Franken
Dr. Anne Weber

Unter Mitarbeit von:
Andreas Daniel (Kapitel 3 und 7)
Dr. Karsten Stephan (Lektorat)

Herausgeber:
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)
Lange Laube 12 | 30159 Hannover | www.dzhw.eu
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960 | info@dzhw.eu

Geschäftsführung:
Prof. Dr. Monika Jungbauer-Gans
Dr. Bernhard Hartung

Vorsitzender des Aufsichtsrats:
Ministerialdirigent Peter Greisler

Registergericht:
Amtsgericht Hannover | B 210251

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/>



Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Tabellenverzeichnis	IV
I Einleitung	1
II Übersicht zum DZHW-Absolventenpanel 2009	2
III Datennutzungshinweise	5
1 Inhalt und Anlage der Studie	8
2 Erhebungsinstrumente	12
2.1 Inhalte der Erhebungsinstrumente	12
2.2 Pretests	15
3 Grundgesamtheit und Stichprobenverfahren	17
4 Durchführung der Erhebungen	18
5 Rücklauf	21
6 Datenaufbereitung	22
6.1 Datenübertragung	22
6.2 Codierung offener Angaben	22
6.3 Datenprüfung und Datenbereinigung	24
6.4 Generierung von Variablen	26
6.5 Erstellung der Datensätze	26
6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels	27
6.7 Codierung fehlender Werte	28
7 Gewichtung	30
8 Anonymisierung	33
9 Literaturverzeichnis	37

Abbildungsverzeichnis

Abbildung 1: Datenzugangswege und Analysepotential.....	6
Abbildung 2: Kohorten-Panel-Design der DZHW-Absolventenstudienreihe.....	8
Abbildung 3: Kalendarium des DZHW-Absolventenpanels 2009, 2. Welle, Papierversion	13
Abbildung 4: Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten des DZHW-Absolventenpanels 2009	34

Tabellenverzeichnis

Tabelle 1:	Anlage der Hauptuntersuchungen der DZHW-Absolventenstudienreihe von 1989 bis 2009.....	9
Tabelle 2:	Vercodete Merkmale und Codierlisten im DZHW-Absolventenpanel 2009	24
Tabelle 3:	Themengebiete und Kürzel für Variablennamen des DZHW-Absolventenpanels 2009	28
Tabelle 4:	Systematik des FDZ-DZHW für fehlende Werte.....	29
Tabelle 5:	Bereitgestellte Gewichte zum DZHW-Absolventenpanel 2009	31
Tabelle 6:	Maßnahmen der statistischen Anonymisierung der Daten des DZHW- Absolventenpanels 2009 nach Zugangsweg	35

I Einleitung

Die DZHW-Absolventenpanels sind eine Untersuchungsreihe zu den beruflichen Werdegängen von Hochschulabsolvent(inn)en.¹ Sie werden durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)² durchgeführt, vom Bundesministerium für Bildung und Forschung (BMBF) gefördert und dienen – in Ergänzung zur amtlichen Hochschulstatistik – dem nationalen Bildungsmonitoring. Seit 1989 wird jeder vierte Absolvent(inn)enjahrgang (Kohorte) mehrfach befragt.

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

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation zur ersten und zweiten Befragungswelle (inkl. zwei Vertiefungsbefragungen) des Absolventenpanels 2009 (doi: 10.21249/DZHW:gra2009:1.0.0).⁵ Weitere Dokumentationsmaterialien zur Studie (Datensatzreport, Fragebögen, Filterführungsdiagramme etc.) können frei im Metadatensuchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden. Die Daten und weitere Dokumentationsmaterialien zur zweiten Welle sowie die Episodendaten aus beiden Wellen können aufgrund von Sperrfristen erst ab Oktober 2017 veröffentlicht werden.

Abschnitt II dieses Berichts stellt die Eckdaten des Absolventenpanels 2009 im Überblick dar. Die zentralen Informationen zur Nutzung der Daten dieser Studie folgen in Abschnitt III. Kapitel 1 stellt Inhalt und Anlage der Absolventenstudienreihe bis 2009⁶ im Allgemeinen und des Absolventenpanels 2009 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 zur Absolventenstudienreihe können über die Website des Projektes (http://www.dzhw.eu/abteilungen/bildung/projekte/pr_detail?pr_id=467) 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-Informationen-System 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.

⁴ Informationen zu verfügbaren Datensätzen und Dokumentationen werden auf der Website des FDZ-DZHW (<https://fdz.dzhw.eu>) zur Verfügung gestellt. Darüber hinaus sind für die Absolvent(inn)enkohorten 1997 und 2001 (jeweils die ersten beiden Befragungswellen) SUF bei GESIS - Leibniz-Institut für Sozialwissenschaften verfügbar (doi:10.4232/1.4272 und doi:10.4232/1.5186).

⁵ Eine dritte Welle ist das Jahr 2019 geplant.

⁶ Es werden nur die Kohorten bis zur hier dokumentierten Kohorte 2009 berücksichtigt.

II Übersicht zum DZHW-Absolventenpanel 2009

Studienreihe	DZHW-Absolventenstudien
Kohorte	Absolvent(inn)enkohorte 2009 (6. Kohorte der Studienreihe)
Erhebende Institution	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)
Förderung	Bundesministerium für Bildung und Forschung (BMBF)
Projektmitarbeiter(innen) (Projektleitung)	Gesche Brandt, <u>Kolja Briedis</u> , Gregor Fabian, Saskia Klüver, Torsten Rehn, Maximilian Trommer
Themen	Studienverlauf Übergang in den Beruf Beruflicher Werdegang Weiterqualifizierung
Erhebungsdesign	Kohorten-Panel-Design
Grundgesamtheit	Hochschulabsolvent(inn)en, die im Wintersemester 2008/2009 oder im Sommersemester 2009 ihren ersten berufsqualifizierenden Studienabschluss an einer staatlich anerkannten Hochschule in der Bundesrepublik Deutschland erworben haben (mit Ausnahme der Absolvent(inn)en von Bundeswehrhochschulen, Verwaltungsfachhochschulen sowie von dualen, Teilzeit- und Fernstudiengängen)
Stichproben	geschichtete Klumpenstichprobe (getrennt für Bachelorabsolvent(inn)en und Absolvent(inn)en traditioneller Studiengänge)
Erhebungsmethode	1. Welle: standardisierte postalische Befragung 2. Welle (Hauptbefragung): standardisierte Online-Befragung bzw. postalische Befragung 2. Welle (Vertiefungsbefragung Promotion): standardisierte Online-Befragung 2. Welle (Vertiefungsbefragung Mobilität): standardisierte Online-Befragung
Erhebungszeitraum	1. Welle: 1. Februar 2010 bis 31. Januar 2011 2. Welle (Hauptbefragung): 16. Februar 2015 bis 2. Oktober 2015 2. Welle (Vertiefungsbefragung Promotion): 4. Juni 2015 bis 15. Juli 2015 2. Welle (Vertiefungsbefragung Mobilität): 16. Juli 2015 bis 31. August 2015

Auswertbare Fälle	1. Welle: n = 10.494 (davon 4.883 Bachelorabsolvent(inn)en) 2. Welle (Hauptbefragung): n = 4.755 (davon 2.110 Bachelorabsolvent(inn)en) 2. Welle (Vertiefungsbefragung Promotion): n = 676 (davon 254 Bachelorabsolvent(inn)en) 2. Welle (Vertiefungsbefragung Mobilität): n = 2.471 (davon 1.065 Bachelorabsolvent(inn)en)
Rücklaufquote	1. Welle: 20 % 2. Welle: <i>werden nachgereicht zur Veröffentlichung im Oktober 2017</i>
Datenprodukte und Zugangswege	CUF: Download SUF: Download, Remote-Desktop, On-Site
Datensatzstruktur	Personendatensätze im wide-Format Episodendatensätze im long-Format
Besonderheiten der Daten	<p>Die Daten der zweiten Befragungswelle (Hauptbefragung und Vertiefungsbefragungen) sowie die Episodendaten aus beiden Wellen können aufgrund von Sperrfristen erst ab Oktober 2017 veröffentlicht werden.</p> <p>Die Bachelorstichprobe enthält auch Bachelorabsolvent(inn)en mit dem Ziel Lehramt. Diese benötigen generell immer einen Master-Abschluss, um in den Beruf übergehen zu können und unterscheiden sich damit von den anderen Bachelorabsolvent(inn)en. Dies sollte bei Analysen berücksichtigt werden. Es handelt sich um Fälle, für die gilt: inlist(astu021f_g1,4,5) & astu03a=1</p>
DOI	10.21249/DZHW:gra2009:1.0.0
Weitere Informationen	https://fdz.dzhw.eu

Projektpublikationen*

Rehn, T., Brandt, G., Fabian, G. & Briedis, K. (2011). *Hochschulabschlüsse im Umbruch. Studium und Übergang von Absolventinnen und Absolventen reformierter und traditioneller Studiengänge des Jahrgangs 2009* (HIS: Forum Hochschule, Hrsg.) (HIS: Forum Hochschule 17/2011). Hannover: HIS.

Briedis, K., Brandt, G., Fabian, G. & Rehn, T. (2011). Bachelorabsolventen im Fokus. In K. Briedis, C. Heine, C. Konegen-Grenier & A.-K. Schröder (Hrsg.), *Mit dem Bachelor in den Beruf. Arbeitsmarktbefähigung und -akzeptanz von Bachelorstudierenden und -absolventen* (S. 53–82). Essen: Edition Stifterverband.

Briedis, K., Klüver, S. & Trommer, M. (2016). *Zwischen Etablierung, Stabilisierung und Aufstieg: Berufliche Entwicklung der Hochschulabsolvent(inn)en 2009* (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Hrsg.) (Forum Hochschule 4/2016). Hannover: DZHW.

* Alle Projektpublikationen werden auf der Website des Projektes (http://www.dzhw.eu/projekte/pr_show?pr_id=302) zum Download bereitgestellt.

Publikationen zum Datensatz (Auswahl)

Schaeper, H., Grotheer, M. & Brandt, G. (2014). Familiengründung von Hochschulabsolventinnen. Eine empirische Untersuchung verschiedener Examenskohorten. In D. Konietzka & M. Kreyenfeld (Hrsg.), *Ein Leben ohne Kinder* (2. Aufl., S. 47–80). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-94149-3_2

Saniter, N. & Siedler, T. (2014). *Door opener or waste of time? The effects of student internships on labor market outcomes* (IZA Discussion Paper No. 8141). Bonn: Forschungsinstitut zur Zukunft der Arbeit (IZA).

III Datennutzungshinweise

[Voraussetzungen der Datennutzung] Die Daten des Absolventenpanels 2009 werden durch das FDZ-DZHW entsprechend dem Bundesdatenschutzgesetz (vgl. § 40 Abs. 1 und Abs. 2 BDSG) ausschließlich zur wissenschaftlichen Nutzung und anonymisiert bereitgestellt.⁷ Das FDZ bietet ein *Scientific Use File* (SUF) für die wissenschaftliche Sekundärnutzung und ein *Campus Use File* (CUF) für Lehr- und Übungszwecke 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 solche 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 Datennutzungsantrag kann von der Website des FDZ heruntergeladen werden. Für die Nutzung des CUF sind lediglich Name und Nutzungszweck anzugeben. Danach wird das CUF durch das FDZ übermittelt.

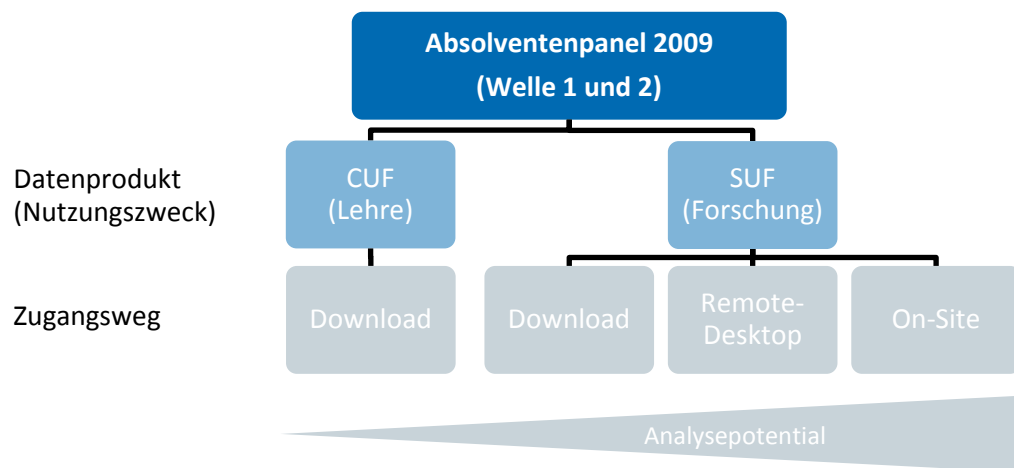
[Datenzugang] Das CUF des Absolventenpanels 2009 kann nach Bereitstellung am lokalen Computer genutzt werden. Das SUF wird über drei Zugangswege angeboten, die hinsichtlich des Speicherortes, der Möglichkeit der eigenständigen Verknüpfung mit externen Daten und der Kontrollmöglichkeiten des FDZ unterschiedlich restriktiv sind.

- **Download:** Die Daten werden verschlüsselt per E-Mail versandt oder auf der Website des FDZ zum Download bereitgestellt. Datennutzer(innen) können die Daten auf ihrem lokalen Computer speichern, falls gewünscht selbst mit Daten aus externen Quellen verknüpfen und die Daten mit eigener Software analysieren.
- **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.

⁷ Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, 6ff.), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (LifBi) (vgl. Koberg, 2016, 699ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner, Müller, Schmucker & Roß, 2012, 9f.) 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.

Die bereitgestellten Daten weisen je nach Zugangsweg einen unterschiedlich hohen 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.⁸ Mit diesem Vorgehen wird ein Höchstmaß an Nutzbarkeit und gleichzeitig ein bestmöglicher Schutz der bereitgestellten Daten sichergestellt.

Abbildung 1: Datenzugangswege und Analysepotential



[Datenprodukte] Über den *Digital Object Identifier* (DOI) 10.21249/DZHW:gra2009:1.0.0 ist eine Webseite mit zentralen Informationen zur Studie, weiteren Dokumentationsmaterialien sowie einer Übersicht der zur Verfügung stehenden Datenprodukte zur Studie erreichbar.

Die bereitgestellten Daten des Absolventenpanels 2009 sind in zwei Datensätzen, einem Personendatensatz im wide-Format und einem Episodendatensatz im long-Format (vgl. Kapitel 6.5) abgelegt. Beide Datensätze werden für jeden im FDZ-DZHW angebotenen Zugangsweg – jeweils mit zugangswegspezifischem Analysepotential (vgl. Abbildung 1) – bereitgestellt.

Das Download-CUF und das Download-SUF sind jeweils im Stata- und im SPSS-Format verfügbar. Für die Zugangswege Remote-Desktop und On-Site werden die Datensätze standardmäßig im Stata-Format bereitgestellt.

[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.

⁸ Zu den verschiedenen Anonymisierungsgraden und Analysepotentialen des CUF und der verschiedenen SUF-Varianten (vgl. Kapitel 8).

⁹ Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

- **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. CUF dürfen ausschließlich im Rahmen der angegebenen Lehrveranstaltung weitergegeben werden.
- **Lösungsgebot:** Download-SUF sind nach Ablauf der vereinbarten Nutzungsdauer (in der Regel drei Jahre) von jeglichen Rechnern, Servern und Datenträgern zu löschen. Ebenso müssen alle Sicherungskopien, modifizierten Datensätze (z. B. Arbeits-, Auszugs- oder Hilfsdateien) sowie Ausdrucke vernichtet werden.
- **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:**
Brandt, G., Briedis, K., Fabian, G., Klüver, S., Rehn, T. & Trommer, M. (2016). DZHW-Absolventenpanel 2009. Aufbereitet durch Baillet, F., Franken, A. & Weber, A., doi: 10.21249/DZHW:gra2009:1.0.0, DATENSATZNAME¹¹, released 2017. Hannover: FDZ-DZHW.
- **Daten- und Methodenbericht:**
Baillet, F., Franken, A. & Weber A. (2017). DZHW-Absolventenpanel 2009. Daten- und Methodenbericht zu den Erhebungen der Absolvent(inn)enkohorte 2009 (1. und 2. 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-Absolventenpanels 2009. Die Daten sind vom Forschungsdatenzentrum des Deutschen Zentrums für Hochschul- und Wissenschaftsforschung (FDZ-DZHW) unter der doi: 10.21249/DZHW:gra2009:1.0.0 veröffentlicht worden.*¹²

¹⁰ <https://metadata.fdz.dzhw.eu>

¹¹ An dieser Stelle bitte den genauen Namen der verwendeten Datensatzversion angeben, z. B. gra2009_p_o_1-0-0.dta für das Onsite-SUF des Personendatensatzes des Prüfungsjahrgangs 2009.

¹² In englischsprachigen Publikationen: This scientific work uses data of the graduate panel 2009, 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:gra2009:1.0.0.

1 Inhalt und Anlage der Studie

[Studienreihe] Das DZHW-Absolventenpanel 2009 ist Teil der DZHW-Absolventenstudienreihe, in der anhand von standardisierten Befragungen Informationen zu Studium, Berufseintritt, Berufsverlauf und Weiterqualifizierung von Hochschulabsolvent(inn)en erfasst werden. Das erste Absolventenpanel wurde 1989¹³ durchgeführt, seitdem wird jeder vierte Absolvent(inn)enjahrgang (Kohorte) untersucht. Die Grundgesamtheit einer Kohorte sind Hochschulabsolvent(inn)en, die im Winter- oder Sommersemester des betreffenden Prüfungsjahrs ihren Studienabschluss an einer Hochschule in der Bundesrepublik Deutschland erworben haben.¹⁴

Für jede Absolvent(inn)enkohorte werden mehrere Befragungswellen durchgeführt, wobei jede Welle in unterschiedlichem zeitlichem Abstand zum Studienabschluss stattfindet. Es handelt sich somit um ein kombiniertes Kohorten-Panel-Design (vgl. Abbildung 2).

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

Jahr	Absolvent(inn)enkohorte					
	1989	1993	1997	2001	2005	2009
1989	Abschluss					
1990	1. Welle					
1991						
1992						
1993	2. Welle	Abschluss				
1994		1. Welle				
1995						
1996						
1997			Abschluss			
1998		2. Welle	1. Welle			
1999						
2000						
2001				Abschluss		
2002			2. Welle	1. Welle		
2003						
2004						
2005					Abschluss	
2006				2. Welle	1. Welle	
2007			3. Welle*			
2008						
2009						Abschluss
2010					2. Welle	1. Welle
2011				3. Welle*		
2012						
2013						
2014						
2015						2. Welle*
2016					3. Welle*	

*Hauptbefragung + Vertiefungsbefragungen

¹³ Schon seit 1974 werden Hochschulabsolvent(inn)en – neben den Studienabbrecher(inne)n und Hochschulwechsler(inne)n – im Rahmen der Exmatrikuliertenbefragung des DZHW befragt. Diese Untersuchungsreihe wird seit Anfang der 2000er Jahre unter dem Namen „Studienabbruch – Umfang und Motive“ durchgeführt.

¹⁴ Für die Kohorte 1989 wurden ausschließlich Absolvent(inn)en aus den Bundesländern der damaligen Bundesrepublik Deutschland ausgewählt.

Die Befragungen der Absolvent(inn)enkohorten von 1989 und 1993 umfassten zwei Wellen; seit 1997 wird eine zusätzliche dritte Befragungswelle durchgeführt. Die erste Welle einer Kohorte findet im Mittel ein Jahr nach dem jeweiligen Studienabschluss statt. Die zweite Befragungswelle folgt etwa fünf Jahre nach dem Studienabschluss. Etwa zehn Jahre nach dem Abschluss schließt sich die dritte Befragungswelle an. Teilweise setzen sich die zweiten bzw. dritten Wellen aus einer Hauptbefragung und gesonderten Vertiefungsbefragungen zu spezifischen Themen zusammen.

Die verschiedenen Erhebungen wurden vorwiegend als schriftlich-postalische Paper-and-Pencil-Befragung (Paper and Pencil Interview; PAPI) durchgeführt, in den letzten Jahren aber zunehmend auch als Online-Befragung (Computer Assisted Web Interview; CAWI) (vgl. Tabelle 1).

Tabelle 1: Anlage der Hauptuntersuchungen der DZHW-Absolventenstudienreihe von 1989 bis 2009

Welle	Befragungszeitpunkt	Thematischer Fokus	Absolvent(inn)enkohorte			
			1989 u. 1993	1997 u. 2001	2005	2009
1	ca. 1 Jahr nach Studienabschluss	Studienverlauf und -erfahrung, akademische Weiterqualifizierung, Übergang in den Beruf	Paper & Pencil	Paper & Pencil	Paper & Pencil	Paper & Pencil
2	ca. 5 Jahre nach Studienabschluss	Aktuelle Tätigkeit, Erwerbstätigkeit, akademische und berufliche Weiterbildung	Paper & Pencil	Paper & Pencil	Paper & Pencil	Online ^b
3	ca. 10 Jahre nach Studienabschluss	Aktuelle Tätigkeit, Erwerbstätigkeit, akademische und berufliche Weiterbildung, familiäre Situation	---	Paper & Pencil ^a	Online ^b	Online (geplant)

^a Die Hauptbefragung wurde als Paper & Pencil- und die Vertiefungsbefragungen als Online-Befragung durchgeführt.

^b Sowohl die Hauptbefragung als auch die Vertiefungsbefragungen wurden als Online-Befragung durchgeführt.

Die Erhebungsinstrumente aller Kohorten enthalten Fragen zum Studium, zum Übergang in den Beruf, zur akademischen und beruflichen Weiterbildung, zur Erwerbstätigkeit und zu soziodemographischen und bildungsbiographischen Merkmalen. Der thematische Fokus einer Befragungswelle orientiert sich an der jeweils typischen Bildungs-, Berufs- und Lebensphase der Befragten zum Befragungszeitpunkt.

[Analysepotential] In allen Kohorten wird je Befragungswelle ein identischer Kern an Informationen erhoben. Auf dieser Grundlage können mit Zeitreihen- bzw. Kohortenvergleichen langfristige Trends der Hochschulbildung und Arbeitsmarktentwicklung in den Blick genommen werden. Zudem wird innerhalb einer Kohorte ein Teil der Fragen in den verschiedenen Befragungswellen wiederholt gestellt. Dies ermöglicht die Betrachtung intra-individueller Veränderungen zwischen den Wellen (z. B. kausale Panelanalysen). Besonders hervorzuheben ist, dass in allen Kohorten über die Wellen hinweg monatsgenaue kontinuierliche Verlaufsda-

ten zum individuellen Tätigkeitsverlauf seit dem Studienabschluss erfasst werden, die sich für Ereignisdaten- und Sequenzmusteranalysen eignen. Darüber hinaus werden in Abhängigkeit von aktuellen Entwicklungen und Forschungsinteressen in einzelnen Kohorten bestimmte Aspekte vertiefend oder ergänzend abgefragt.

[Einordnung ins Forschungsfeld] Das Stichproben- und Erhebungsdesign sowie die damit verbundenen Analysemöglichkeiten unterscheiden die DZHW-Absolventenstudienreihe von anderen in Deutschland durchgeführten Absolventenstudien. So ist beispielsweise das Bayerische Absolventenpanel (BAP) des Bayerischen Staatsinstituts für Hochschulforschung und Hochschulplanung (IHF) auf Absolvent(inn)en bayerischer Hochschulen begrenzt.¹⁵ Das Kooperationsprojekt Absolventenstudie (KOAB) des International Centre for Higher Education Research (INCHER) befragt Absolvent(inn)en seiner Kooperationshochschulen und ermöglicht individuelle Analysen auf Hochschul- und Studiengangsebene, die zur Evaluation und Weiterentwicklung genutzt werden können.¹⁶

[Spezifika des Absolventenpanels 2009] Neben den allgemeinen Charakteristika der Studienreihe weist die hier betrachtete Absolvent(inn)enkohorte 2009¹⁷, für die momentan Daten aus zwei Hauptbefragungen und zwei Vertiefungsbefragungen zum Thema „Promotion“ und „regionale Mobilität“ vorliegen¹⁸, folgende Spezifika auf. Wie bei der Absolvent(inn)enkohorte 2005 ist die Studienphase der Kohorte 2009 durch den Hochschulwandel im Rahmen des Bologna-Prozesses geprägt und die berufliche Einstiegsphase durch die Wirtschafts- und Finanzkrise im Jahr 2008 gekennzeichnet. Im Unterschied zu vorangegangenen Absolvent(inn)enkohorten weist die Absolvent(inn)enkohorte 2009 jedoch neben Absolvent(inn)en mit traditionellen Abschlüssen auch eine große Anzahl an Bachelorabsolvent(inn)en auf, die im Prüfungsjahr 2009 22,9 % der Gesamtabzolvent(inn)enzahl umfassten (vgl. Dudek, Glässner & Schröder, 2010, S. 31).¹⁹ Dies ermöglicht zum einen Vergleiche zwischen Bachelorabsolvent(inn)en und Absolvent(inn)en traditioneller Studiengängen. Zum anderen kann betrachtet werden, ob die Bachelorabsolvent(inn)en nach ihrem Abschluss in den Beruf übergehen, ein Masterstudium aufnehmen oder beide Optionen kombinieren.

Die Erhebungsinstrumente der Hauptbefragungen des Absolventenpanels 2009 sind inhaltlich stark angelehnt an die Fragebögen des Absolventenpanels 2005 und bieten entsprechende Vergleichsmöglichkeiten. Durch die beiden Vertiefungsbefragungen wurden jedoch in der zweiten Welle insgesamt umfangreichere Informationen als in vorherigen zweiten Befragungswellen erfasst. Eine methodische Neuerung der Absolventenkohorte 2009 ist im Ver-

¹⁵ vgl. dazu <http://www.bap.ihf.bayern.de>

¹⁶ vgl. dazu <http://koab.uni-kassel.de>

¹⁷ Die Grundgesamtheit sind Hochschulabsolvent(inn)en, die im Wintersemester 2008/2009 oder im Sommersemester 2009 ihren ersten berufsqualifizierenden Studienabschluss an einer staatlich anerkannten Hochschule in der Bundesrepublik Deutschland erworben haben (mit Ausnahme der Absolvent(inn)en von Bundeswehrhochschulen, Verwaltungsfachhochschulen sowie von dualen, Teilzeit- und Fernstudiengängen).

Im Rahmen der ersten Befragungswelle wurde ursprünglich zudem eine kleine Sonderstichprobe mit Masterabsolvent(inn)en berücksichtigt. Die realisierte Fallzahl war jedoch so gering, dass insbesondere fächerdifferenzierte Analysen damit nicht sinnvoll sind. Daher wurden die Masterabsolvent(inn)en aus dem Datensatz entfernt und auch in der folgenden Dokumentation nicht weiter berücksichtigt.

¹⁸ Die Daten der zweiten Befragungswelle (Hauptbefragung und Vertiefungsbefragungen) sowie die Episodendaten aus beiden Wellen können aufgrund von Sperrfristen erst ab Oktober 2017 veröffentlicht werden. Eine dritte Befragungswelle ist für das Jahr 2019 geplant.

¹⁹ In der Absolvent(inn)enkohorte 2005 lag der Anteil noch bei 4,4 % (vgl. Dudek, Glässner & Krause, 2007, S. 25). Daher wurden hier lediglich Absolvent(inn)en aus Fächern berücksichtigt, für die bereits eine größere Absolvent(inn)enzahl vorlag. Aus diesem Grund eignete sich die Stichprobe der Bachelorabsolvent(inn)en in der Kohorte 2005 nicht, um Aussagen über diese Gruppe zu treffen.

gleich zu den vorangegangenen Kohorten außerdem, dass die zweite Befragungswelle erstmals online durchgeführt wurde.

2 Erhebungsinstrumente

In der ersten Befragungswelle des Absolventenpanels 2009 wurde als Erhebungsinstrument ein standardisierter Papierfragebogen eingesetzt. In der Hauptbefragung der zweiten Befragungswelle kamen ein standardisierter Fragebogen in einer Online- und einer Papierversion zum Einsatz (vgl. Kapitel 4). Die beiden Vertiefungsbefragungen der zweiten Befragungswelle wurden ausschließlich über einen standardisierten Online-Fragebogen realisiert. Alle eingesetzten Erhebungsinstrumente waren in deutscher Sprache.²⁰

Kapitel 2.1 stellt die zentralen Inhalte der Erhebungsinstrumente vor. Kapitel 2.2 beschreibt die zur Prüfung und Verbesserung der Fragebögen durchgeführten Pretests.

2.1 Inhalte der Erhebungsinstrumente

[Charakteristika der Studienreihe] Im Fokus des Absolventenpanels 2009 stehen, wie bei den übrigen Kohorten der Studienreihe, der Übergang von der Hochschule in die Beschäftigung und der Zusammenhang zwischen Studium und Berufserfolg. Ansatzpunkt im Erhebungsinstrument ist ein Rückblick auf das Studium. Dabei werden Informationen zum Studienverlauf und Studienerfolg, zur Bewertung der Studienbedingungen sowie zu an der Hochschule erworbenen Qualifikationen erhoben.²¹ Daran anknüpfend wird der weitere Werdegang der Absolvent(inn)en in den Blick genommen.

Zu diesem Zweck wird über die Wellen hinweg der individuelle Tätigkeitsverlauf der Absolvent(inn)en seit ihrem Studienabschluss erfasst. Pro Tätigkeit wird jeweils deren Art (z. B. Erwerbstätigkeit, Promotion, Elternzeit) sowie ihr Anfangs- und Endzeitpunkt monatsgenau erhoben. Dies geschieht seit der ersten Welle der Kohorte 2001 bzw. der zweiten Welle der Kohorte 1997 in Form eines Kalendariums (vgl. Frage 4.1 in Welle 1 und Frage 1.5 (Papierversion) bzw. Page 5 (Onlineversion) in Welle 2), in das die Befragten ihre einzelnen Tätigkeiten eintragen sollen (vgl. Abbildung 3).²² Das Kalendarium wurde vom DZHW konzipiert, um lückenhafte Angaben in der Beschreibung von Tätigkeitsverläufen zu reduzieren.

²⁰ Die Fragebögen sowie Filterführungsdiagramme können im Metadatensuchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden (für die zweite Welle erst ab Oktober 2017).

²¹ vgl. Abschnitt 1 „Studienverlauf und Studienerfahrungen“ im Fragebogen der ersten Welle.

²² Vor der Einführung des Kalendariums wurden die Tätigkeitsverläufe anhand eines Tableaus erfasst.

Abbildung 3: Kalendarium des DZHW-Absolventenpanels 2009, 2. Welle, Papierversion²³

1.5 Um die Wege beim Übergang aus dem Studium in das Berufsleben und in andere Lebensbereiche besser verstehen zu können, bitten wir Sie, Ihre seit Anfang 2010 ausgeübten Tätigkeiten in den folgenden Kalender einzutragen.

Bitte tragen Sie für die Zeit von Januar 2010 bis heute Ihre Tätigkeiten anhand der aufgeführten Kennbuchstaben in Ihren persönlichen Kalender ein. Haben Sie mehrere Tätigkeiten gleichzeitig ausgeübt, können Sie diese untereinander auflisten. Wichtig ist, dass es keine zeitlichen Lücken gibt.

Beispiel:

Von Anfang 2010 an waren Sie als Angestellte(r) erwerbstätig (A). Parallel dazu arbeiteten Sie an Ihrer Promotion (D), die Sie im Juni 2011 beendeten. Nach Auslaufen Ihrer Stelle waren Sie im Oktober 2011 arbeitslos (AL), hatten danach bis Dezember einen Werkvertrag (W) und begannen im Januar 2012 auf einer neuen Stelle (A).

	Januar	Februar	März	April	Mai	Juni	Juli	August	Sept.	Oktober	Nov.	Dez.
2010	A											
	D											
2011	A									AL	W	
	D											
2012	A											
usw.												

A Erwerbstätigkeit, nichtselbstständig (als Angestellte(r), Beamte(r))

SE Selbstständige Erwerbstätigkeit (ohne Werk-/Honorararbeit)

T Trainee

J Jobben

W Werkvertrag, Honorararbeit

B Berufsausbildung, Umschulung, Volontariat

P Praktikum

R Referendariat, Anerkennungspraktikum u. Ä.

ST Studium

D Promotion

JP Juniorprofessur/Habilitation

F Fort-, Weiterbildung (Vollzeit, längerfristig)

AL Arbeitslosigkeit

EZ Elternzeit

H Hausfrau/Hausmann, Familienarbeit

SO Sonstiges (z. B. Wehr-/Zivildienst, längerer Urlaub, Krankheit)

Ergänzend zum Kalendarium werden zu einzelnen Tätigkeitsarten zusätzliche Informationen erhoben. Ein Großteil der Fragen bezieht sich dabei auf die Erwerbstätigkeiten. Anhand des Erwerbstätigkeitstableaus (vgl. Frage 5.4 in Welle 1 und Frage 4.5 (Papierversion) bzw. Page 19 (Onlineversion) in Welle 2), das zeitgleich mit dem Kalendarium eingeführt wurde, werden für alle angegebenen Erwerbstätigkeiten der Zeitraum, die Art des Arbeitsverhältnisses, der Arbeitszeitumfang, die berufliche Stellung und der Arbeitsort erhoben. Darüber hinaus werden für die erste und die letzte Erwerbstätigkeit noch weitere Merkmale erfasst, etwa das Einkommen und die Branche.²⁴ Daneben beinhalten die Erhebungsinstrumente auch Fragen zur subjektiven Einschätzung der eigenen beruflichen Situation, etwa hinsichtlich der beruflichen Zufriedenheit, den beruflichen Zukunftsperspektiven und der Beschäftigungsadäquanz²⁵. Ein Großteil der Informationen zu den Erwerbstätigkeiten wird über die Befragungswellen hinweg wiederholt erhoben.²⁶ Mit Blick auf den Übergang in den Beruf werden zudem Fragen zu speziellen Erwerbstätigkeitsformen in zweiten Ausbildungsphasen, z. B. Referendariaten oder Vikariaten, sowie zu den Wegen der Berufsfindung, z. B. zur Stellensuche, gestellt.

Neben den Erwerbstätigkeiten werden auch andere Tätigkeitsarten vertiefend behandelt. So werden etwa Fragen zu akademischen Weiterqualifizierungen sowie zu beruflichen Fort- und Weiterbildungen gestellt.

²³ Die Onlineversion des Kalendariums der zweiten Welle findet sich in den Screenshots auf Page 5.

²⁴ Vor der Einführung des Erwerbstätigkeitstableaus wurden – abgesehen von der Art der Institution – generell nur zusätzliche Informationen zur ersten und letzten Erwerbstätigkeit erhoben.

²⁵ vgl. dazu auch Kerst & Fehse, 2007.

²⁶ Teilweise wurde allerdings der Wortlaut abgewandelt. Darauf hinzuweisen ist außerdem, dass sich die Wiederholungsmessung in der zweiten Welle bei einer beruflichen Veränderung auf eine andere berufliche Situation bzw. Tätigkeit bezieht als in der ersten Welle.

Darüber hinaus werden mit den Erhebungsinstrumenten verschiedene soziodemographische bzw. bildungsbiographische Merkmale erfasst, etwa zum elterlichen Bildungshintergrund oder zu Partnerschaft und Kindern.

[Spezifika des Absolventenpanels 2009] Wie bereits bei der Absolvent(inn)enkohorte 2005 ist auch die Studienphase der Kohorte 2009 durch die Studienstrukturreform im Rahmen des Bologna-Prozesses geprägt und die berufliche Einstiegsphase durch die Wirtschafts- und Finanzkrise im Jahr 2008 gekennzeichnet. Aus diesem Grund wurde das Erhebungsinstrument der ersten Befragungswelle inhaltlich stark angelehnt an den Fragebogen des Absolventenpanels 2005. So wurden lediglich einzelne Fragen innerhalb der bestehenden Themenblöcke ausgetauscht bzw. neu aufgenommen. Beispielsweise wurden ergänzende Fragen zur akademischen Weiterqualifizierung (vgl. Frage 2.2, 2.3 und 2.5), zum Praktikum (vgl. Frage 4.12, 4.13 und 4.17), zu Karrieremaßnahmen (vgl. Frage 5.8) sowie zu Arbeits- und Lebenszielen (vgl. Frage 5.19) gestellt.

In der zweiten Befragungswelle gab es im Vergleich zu vorangegangenen Kohorten die methodische Neuerung, dass sie erstmals online durchgeführt wurde. In diesem Zuge wurde sie in eine Hauptbefragung²⁷ und zwei ergänzende Vertiefungsbefragungen zu den Themen „Promotion“ und „regionale Mobilität“ unterteilt. Das Erhebungsinstrument der Hauptbefragung orientiert sich wiederum am Fragebogen der zweiten Welle des Absolventenpanels 2005. Zusätzliche Fragen wurden vor allem zu den beruflichen Fort- und Weiterbildungen (vgl. Fragenblock 6) sowie zur aktuellen beruflichen Tätigkeit (Frage 4.7, 4.11, 4.13, 4.16 und 4.17) aufgenommen. In der Vertiefungsbefragung zur Promotion wurden Promovierende und Promovierte zu ihren Erfahrungen in der Promotionsphase befragt, beispielsweise zu Motiven für die Promotion, zum institutionellen Rahmen oder zur Betreuungssituation der Promotion. In der Vertiefungsbefragung zur regionalen Mobilität wurden unter anderem der Wohnverlauf seit dem Studienabschluss, Mobilitätseinstellungen und -absichten sowie Auslandsaufenthalte erfasst.²⁸ Für die Vertiefungsbefragungen wurden einige Fragen aus dem Frageinstrumentarium der vorherigen zweiten Befragungswellen ausgelagert, größtenteils aber zusätzliche Fragen gestellt. Insgesamt ist die zweite Befragungswelle des Absolventenpanels 2009 damit deutlich umfangreicher als die zweiten Befragungswellen vorheriger Kohorten. Während es zur Promotion ebenfalls in den dritten Wellen der Kohorten 1997 und 2001 jeweils eine Vertiefungsbefragung gab, wurde die Mobilität bislang in keiner Kohorte vertiefend beleuchtet.

²⁷ Für die Hauptbefragung gab es zusätzlich auch noch eine Papierversion des Fragebogens (vgl. Kapitel 4).

²⁸ Darauf hinzuweisen ist, dass in der Mobilitätsbefragung die Big-Five-Kurzskala BFI-10 zum Einsatz kam, vgl. dazu:

- Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less. A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41 (1), 203–212. doi:10.1016/j.jrp.2006.02.001
- Rammstedt, B. (2007). The 10-Item Big Five Inventory. Norm values and investigation of sociodemographic effects based on a German population representative sample. *European Journal of Psychological Assessment*, 23 (3), 193–201. doi:10.1027/1015-5759.23.3.193
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2013). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit. 10 Item Big Five Inventory (BFI-10). *methoden, daten, analysen*, 7(2), 233–249. doi:10.12758/mda.2013.013
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2014). *Big Five Inventory (BFI-10). Zusammenstellung sozialwissenschaftlicher Items und Skalen* (GESIS – Leibniz-Institut für Sozialwissenschaften, Hrsg.). doi:10.6102/zis76

2.2 Pretests

[Ziel und Verfahren] Die Erhebungsinstrumente der ersten Befragungswelle und der Hauptbefragung der zweiten Befragungswelle wurden im Vorfeld der Erhebungen jeweils durch Pretests geprüft.²⁹ Dabei sollte erstens für die bereits in Erhebungsinstrumenten vorheriger Kohorten eingesetzten Fragen und Antwortvorgaben geprüft werden, ob sie von der Absolvent(inn)enkohorte 2009 gleich perzipiert werden würden wie von vorangegangenen Kohorten. Zweitens sollte überprüft werden, ob die Erhebungsinstrumente auch für die Befragung der neuen Absolvent(inn)engruppe der Bachelor (vgl. Kapitel 1 und 3) geeignet waren. Drittens sollte für neu eingesetzte Fragen die Verständlichkeit und Beantwortbarkeit getestet werden. Zudem zog die Aufnahme neuer Fragen bzw. die veränderte Fragenauswahl auch Veränderungen am Aufbau und Layout des Fragebogens sowie der Befragungsdauer nach sich, die evaluiert werden sollten. In der zweiten Befragungswelle musste viertens die Umsetzung der Onlineversion des Fragebogens geprüft werden.

Zur Prüfung dieser verschiedenen Aspekte kam in der ersten Befragungswelle ein sogenanntes *Pretestverfahren im Feld* zum Einsatz. Dieses Verfahren hat zum Ziel, dass die am Pretest teilnehmenden Personen „unter möglichst ähnlichen Bedingungen untersucht [werden], wie sie später für die eigentliche Erhebung vorgesehen sind“ (Häder, 2015, S. 396).³⁰ Das Erhebungsinstrument der Hauptbefragung der zweiten Befragungswelle wurde im Rahmen von *Expertenbewertungen* (vgl. Häder, 2015, S. 406–407) begutachtet.

[Probanden] In der ersten Befragungswelle wurden als Testpersonen zum einen Mitarbeiter(innen) des DZHW ausgewählt, die ihren Hochschulabschluss im Prüfungsjahrgang 2009 oder einem angrenzenden Prüfungsjahrgang erworben hatten. Durch die Befragung dieser Testpersonen wurden so gleichzeitig Expert(inn)en aus dem Bereich der Hochschulforschung einbezogen. Zum anderen wurden über persönliche oder projektbezogene Kontakte verschiedene externe Personen rekrutiert. Insgesamt nahmen in der ersten Befragungswelle zwischen 10 und 15 Personen am Pretest teil.³¹ In der zweiten Befragungswelle wurde das Erhebungsinstrument von etwa fünf Mitarbeiter(innen)³² des DZHW als Hochschulforschungsexpert(inn)en geprüft.

[Durchführung] Beide Pretests fanden etwa zwei bis drei Monate vor dem jeweiligen Feldstart statt. In der ersten Befragungswelle wurden die Testpersonen gebeten, den für die Befragung vorgesehenen standardisierten Fragebogen zu bearbeiten und Verständnisprobleme, Kritik oder Anmerkungen zu notieren. Im Anschluss an die Bearbeitung des Fragebogens wurde mit den Probanden ein Fragenkatalog zu verschiedenen Aspekten der Befragung durchgegangen. Dabei wurden Informationen zur Ausfülldauer, zu Inhalt und Länge des Fragebogens, zu Aufbau und Layout, zur Verständlichkeit der Fragen und Ausfüllanweisungen sowie zur Vollständigkeit der Antwortmöglichkeiten erhoben. Außerdem wurden konkrete Nachfragen zu neuen oder veränderten Fragen gestellt.

Die Expertenbewertungen in der zweiten Befragungswelle wurden iterativ in zwei Schritten eingeholt. In einem ersten Schritt wurden die einzelnen Fragen inhaltlich begutachtet.

²⁹ Für die Erhebungsinstrumente der beiden Vertiefungsbefragungen gab es keine Pretests.

³⁰ Darauf hinzuweisen ist jedoch, dass die Testpersonen der durchgeführten Pretests nicht alle – wie bei diesen Pretestverfahren üblich – der Grundgesamtheit entstammten, sondern auch andere Prüfungsjahrgänge berücksichtigt wurden.

³¹ Die genaue Anzahl der Teilnehmer(innen) ist nicht mehr rekonstruierbar.

³² Die genaue Anzahl der Teilnehmer(innen) ist nicht mehr rekonstruierbar.

Nachdem die darauf basierenden Änderungsvorschläge eingearbeitet worden waren, wurden die Onlineversion des Fragebogens programmiert sowie die Papierversion des Fragebogens gesetzt. Diese wurden dann wiederum anhand eines Fragenkatalogs zu Dauer, Aufbau, Ausfüllanweisungen, Inhalt, Antwortmöglichkeiten sowie zum konkreten Themenblock der Weiterbildung begutachtet.

Auf Grundlage der Pretestergebnisse wurden die Formulierungen verschiedener Frage- bzw. Itemtexte präzisiert, die Reihenfolgen von einzelnen Fragen bzw. Items und Antwortkategorien überarbeitet, einzelne Fragen bzw. Items gestrichen oder neu aufgenommen sowie das Layout entsprechend angepasst.³³ Darauf hinzuweisen ist, dass in der zweiten Befragungswelle in diesem Zuge die Filterführung innerhalb des Themenblocks der Weiterbildung in der Onlineversion anders gestaltet wurde als in der Papierversion. Der grundsätzliche Aufbau und Umfang der Fragebögen wurden unverändert beibehalten.

³³ Die spezifischen Gründe für Änderungen können nicht mehr rekonstruiert werden.

3 Grundgesamtheit und Stichprobenverfahren

[Grundgesamtheit] Die Grundgesamtheit des Absolventenpanels 2009 umfasst alle Hochschulabsolvent(inn)en, die im Wintersemester 2008/2009 oder im Sommersemester 2009³⁴ ihren ersten berufsqualifizierenden Studienabschluss an einer staatlich anerkannten Hochschule in der Bundesrepublik Deutschland erworben haben. Hochschulabsolvent(inn)en von Bundeswehrhochschulen, Verwaltungsfachhochschulen sowie von dualen, Teilzeit- und Fernstudiengängen waren dabei von vornherein ausgenommen.

[Stichprobenverfahren] Aufgrund fehlender oder nicht zugänglicher Listen von Hochschulabsolvent(inn)en mussten die Individuen über die Hochschulen rekrutiert werden. Zu diesem Zweck wurden zunächst auf Basis der vorliegenden Studienanfängerzahlen und der Studiendauern aller Studierenden an Universitäten und Fachhochschulen für jedes Bundesland einzeln diejenigen Studiengänge prognostiziert, die für den Prüfungsjahrgang 2009 voraussichtlich Absolvent(inn)en verzeichnen würden.³⁵ Daraus wurden zwei getrennte geschichtete Klumpenstichproben für zum einen Hochschulabsolvent(inn)en traditioneller Studiengänge³⁶ und zum anderen Hochschulabsolvent(inn)en aus Bachelorstudiengängen gezogen.

[Geschichtete Klumpenstichprobe] Die „primary sampling units“ (Klumpen) der Klumpenstichproben wurden anhand der Hochschule und dem Studienbereich³⁷ definiert. Die „secondary sampling units“ stellten die Hochschulabsolvent(inn)en des Prüfungsjahrgangs 2009 innerhalb dieser Klumpen dar.³⁸ Die Umsetzung dieses Designs erfolgte über die Ziehung von einfachen Zufallsstichproben. Die Klumpen wurden dabei nach Region (alte Länder, neue Länder einschließlich Berlin) geschichtet. In den jeweiligen Schichten der Stichproben bestand das Ziel, eine im Vergleich zur Grundgesamtheit proportionale Verteilung zu erreichen. Erhebungsbedingt ergaben sich jedoch leichte Disproportionalitäten bei den Hochschulabsolventen traditioneller Studiengänge, sodass ein relativ höherer Anteil dieser Gruppe in Ostdeutschland in die Stichprobe gelangte. Diese designbedingte Verzerrung kann durch die Verwendung der erstellten kalibrierten Designgewichte ausgeglichen werden (siehe Kapitel 7).

Beim Ausfall eines Klumpens (z. B. bei Teilnahmeverweigerung auf Hochschul- oder Fakultätsebene) wurde ein – hinsichtlich der Merkmalskombination Studienbereich, Hochschulart, Abschlussart, Region – möglichst ähnlicher Klumpen als Ersatz gesucht. Bei mehreren Klumpen mit ähnlichen Merkmalskombinationen wurde der größte Klumpen ausgewählt.

Darauf hinzuweisen ist, dass die Bachelorstichprobe auch Bachelorabsolvent(inn)en mit dem Ziel Lehramt enthält. Diese benötigen generell immer einen Master-Abschluss, um in den Beruf übergehen zu können und unterscheiden sich damit von den anderen Bachelorabsolvent(inn)en. Dies sollte bei Analysen berücksichtigt werden.³⁹

³⁴ Das Prüfungsjahr 2009 begann an Fachhochschulen im September 2008 und endete im August 2009, an Universitäten begann es im Oktober 2008 und endete im September 2009.

³⁵ Bei Erwartung annähernd konstanter Absolvent(inn)enzahlen hätten die Zahlen des vorherigen Prüfungsjahrgangs herangezogen werden können. Durch die Umstellung auf die neuen Studiengänge unterschieden sich die Statistiken zwischen den Prüfungsjahrgängen jedoch zu stark.

³⁶ Hierunter fallen Absolvent(inn)en aus Studiengängen mit den Abschlüssen Diplom, Magister, Staatsexamen (u.a. Lehramtsstudiengänge) sowie Absolvent(inn)en mit kirchlichen und künstlerischen Abschlüssen.

³⁷ entsprechend der Aufgliederung nach der amtlichen Statistik (gemäß Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2008/2009 und SoSe 2009).

³⁸ Beispiel für den Personenkreis eines Klumpens: Alle Bachelorabsolvent(inn)en des Studienbereichs Physik an der Universität A oder alle Diplomabsolvent(inn)en der Wirtschaftswissenschaften der Fachhochschule B

³⁹ Es handelt sich um Fälle, für die gilt: inlist(astu021f_g1,4,5) & astu03a=1

4 Durchführung der Erhebungen

[Kontaktaufnahme und Adresspflege] Die gezogenen Hochschulen wurden vom DZHW angeschrieben und um ihre Teilnahme gebeten. Zudem teilte das DZHW den Hochschulen bzw. deren Prüfungsämtern die Kriterien (Prüfungsjahr, Erstabschluss, Studienfach, Abschlussart) mit, anhand derer sie die Zielpersonen für das Absolventenpanel 2009 identifizieren sollten (vgl. Kapitel 3).⁴⁰ Insgesamt wurden mehr als 400 Prüfungsämter kontaktiert. Da die Hochschulen die Kontaktdaten ihrer Absolvent(inn)en aus Datenschutzgründen nicht herausgeben durften, teilten sie dem DZHW lediglich die jeweilige Absolvent(inn)enanzahl mit. Anschließend sendete das DZHW die passende Anzahl an Erhebungsunterlagen für die erste Befragungswelle postalisch an die jeweiligen Prüfungsämter, die diese an die Zielpersonen weiterverschickten.

Um die weiterhin teilnahmebereiten Personen in der zweiten Befragungswelle direkt durch das DZHW kontaktieren zu können, wurden im Fragebogen der ersten Welle deren Kontaktdaten (Post- und E-Mail-Adresse) erfasst. Beim Eingang eines Fragebogens im DZHW wurde sowohl auf den Fragebogen als auch auf den Adressabschnitt des Fragebogens per Paginierstempel eine eindeutige Identifikationsnummer gestempelt und aus allen Adressabschnitten eine Referenzliste von der Identifikationsnummer zur jeweils zugehörigen Adresse erstellt.⁴¹ Die Adressbestände wurden zwischen den Wellen geprüft und ggf. aktualisiert, um auch Personen berücksichtigen zu können, deren E-Mail-Adressen nach der ersten Befragung noch nicht vorlagen oder sich in der Zwischenzeit verändert hatten bzw. die umgezogen waren.⁴²

[Erhebungsunterlagen] Die Erhebungsunterlagen der ersten Befragungswelle bestanden pro zu befragender Person aus einem postalischen Anschreiben (inkl. Datenschutzinformationen), dem Papierfragebogen, einem Flyer mit zentralen Informationen zur Studie und einem an das DZHW adressierten portofreien Umschlag zur Rücksendung des ausgefüllten Fragebogens. Zudem wurden zwei Erinnerungsschreiben verschickt.

Für die Hauptbefragung der zweiten Befragungswelle unterschieden sich die Erhebungsunterlagen nach deren Versandart. Sofern eine valide E-Mail-Adresse vorlag, wurde ein Anschreiben mit einem Link zur Online-Befragung, einem individuellen Passwort (Token) und einem Link zum Datenschutzblatt verschickt. Andernfalls erfolgte der Versand des Anschreibens (inkl. Link und Passwort zur Befragung), des Datenschutzblatts sowie eines Flyers postalisch. Dabei wurde eine Papierversion des Fragebogens mit einem frankierten Rückumschlag mitverschickt, sodass die Befragten entscheiden konnten, über welchen Erhebungsmodus sie

⁴⁰ An dieser Stelle können durch die Prüfungsämter vereinzelt auch Personen als stichprobenzugehörig identifiziert worden sein, die nicht zur Grundgesamtheit gehörten (Overcoverage), wenn sie beispielsweise einem anderen Prüfungsjahrgang als dem Jahrgang 2009 angehörten (dies tritt z.B. bei Verzögerung der Zeugniserstellung auf) oder es sich nicht um den Erstabschluss handelte.

⁴¹ Zur Gewährleistung des Datenschutzes wurde der Adressabschnitt vom Fragebogen abgetrennt und die Referenzliste getrennt von den Befragungsdaten auf einem geschützten Server gespeichert.

⁴² Die Befragten wurden dabei sowohl im Anschluss an die erste Welle – im Rahmen eines Informationsschreibens über die Ergebnisse der Befragung – als auch im Vorfeld zur zweiten Welle angeschrieben und um die Aktualisierung ihrer Adresse gebeten. Die Adressen von unzustellbaren Postsendungen wurden über den Adressaktualisierungsservice der Deutschen Post sowie die Melderegisterauskunft von RISER ID Services GmbH aktualisiert. Innerhalb der Feldphase der zweiten Welle wurden außerdem durch das DZHW Adressen nachrecherchiert, wenn Erhebungsunterlagen unzustellbar waren. In Vorbereitung auf die dritte Welle wurden nach Abschluss der zweiten Welle weitere Adressaktualisierungen im Zuge eines erneuten Ergebnisversands vorgenommen.

an der Befragung teilnehmen wollten. Darüber hinaus wurden per E-Mail drei und per Post zwei Erinnerungsschreiben versandt.

Die Erhebungsunterlagen der beiden Vertiefungsbefragungen, die ausschließlich als Online-Befragung durchgeführt wurden, umfassten – sowohl beim E-Mail- als auch beim Post-Versand – jeweils ein Anschreiben mit Link und individuellem Passwort zur Befragung. Pro Vertiefungsbefragung wurde außerdem ein Erinnerungsschreiben verschickt.

[Feldphase] Der Erhebungszeitraum der ersten Befragungswelle erstreckte sich vom 1. Februar 2010 bis 31. Januar 2011.⁴³ Die beiden Erinnerungsschreiben wurden etwa zwei bzw. vier Wochen nach Feldstart verschickt. Aufgrund des angewendeten Kontaktverfahrens über die Prüfungsämter konnte das DZHW keinen direkten Einfluss auf den genauen Versandzeitpunkt der Erhebungsunterlagen nehmen.⁴⁴ Ebenfalls wurden die Erinnerungsschreiben an alle Personen der Stichprobe – also auch diejenigen, die sich bereits an der Befragung beteiligt hatten – verschickt, da den Prüfungsämtern unbekannt war, welche Personen bereits einen Fragebogen an das DZHW zurückgeschickt hatten.

Der Erhebungszeitraum der Hauptbefragung der zweiten Befragungswelle erstreckte sich vom 16. Februar 2015 bis 2. Oktober 2015.⁴⁵ Durch den nun im DZHW vorliegenden Adressbestand konnten konkrete Versandzeitpunkte für die Erhebungsunterlagen festgelegt werden.⁴⁶ Zudem konnten die Erinnerungsschreiben gezielt nur an diejenigen Zielpersonen verschickt werden, die sich noch nicht an der Erhebung beteiligt hatten.

Die Feldphasen der Vertiefungsbefragungen, zu der nur Teilnehmer(innen) der Hauptbefragung der zweiten Befragungswelle eingeladen wurden⁴⁷, fanden vom 4. Juni 2015 bis 15. Juli 2015 (Vertiefungsbefragung Promotion) bzw. 16. Juli 2015 bis 31. August 2015 (Vertiefungsbefragung Mobilität) statt. Auch hier konnten konkrete Versandzeitpunkte⁴⁸ bestimmt sowie gezielte Erinnerungen von Nicht-Teilnehmer(inne)n vorgenommen werden.

[Rücklaufsteigernde Maßnahmen] Die rücklaufsteigernden Maßnahmen zielten zum einen auf Anreize für die Hochschulen, die Befragung beim Erstkontakt organisatorisch zu unterstützen, und zum anderen auf die individuelle Motivation der Befragten, ab. Die Hochschulen erhielten zusammen mit dem Anschreiben vom DZHW ein Begleitschreiben des BMBF, in dem auf die gesamtgesellschaftliche Bedeutung der Erhebung hingewiesen wurde. Zudem wurde – mit Blick auf folgende Erhebungen – im Anschluss an die erste Welle eine Zusammenfassung der Ergebnisse an die Hochschulen versandt. Dieses wurde sowohl an Hochschulen, die an der Befragung teilgenommen hatten, als auch an Hochschulen, die ihre Teilnahme verweigert hatten, verschickt. Auf der Ebene der Befragten wurden, neben dem in der Haupt- und den Vertiefungsbefragungen eingesetzten Instrument der Erinnerungsschreiben, weitere rücklauf-

⁴³ Die Feldzeit wurde so lange wie möglich – parallel zu den weiteren Projektschritten – ausgedehnt und jeder noch eintreffende Fragebogen bis Januar 2011 berücksichtigt.

⁴⁴ Nach dem Versand der Erstanschriften meldeten die Prüfungsämter jeweils den genauen Versandtermin. Das DZHW schickte dann im Abstand von zwei Wochen zu diesem Versandtermin das erste Erinnerungsschreiben an das Prüfungsamt, welches dieses wiederum an die Zielpersonen weiterschickte. Anschließend meldete das Prüfungsamt wiederum den Versandtermin dieses Erinnerungsschreibens. Das Verfahren für die zweite Erinnerung erfolgte analog.

⁴⁵ Auch hier wurde die Feldzeit so lange wie möglich offen gehalten, sodass Nachzügler bis Oktober 2015 berücksichtigt werden konnten.

⁴⁶ Einladung: 16.02.2015; Erste Erinnerung: 26.02.2015 (E-Mail), 05.03.2015 (postalisch); Zweite Erinnerung: 12.03.2015 (E-Mail), 19.03.2015 (postalisch); Dritte Erinnerung: 20.03.2015 (E-Mail), postalisch entfällt

⁴⁷ Zur Vertiefungsbefragung Promotion wurden zudem nur die Teilnehmer(innen) eingeladen, die in der Hauptbefragung angegeben hatten, nach ihrem Studienabschluss eine Promotion aufgenommen zu haben.

⁴⁸ Vertiefungsbefragung Promotion: Einladung: 04.06.2015; Erinnerung: 18.06.2015; Vertiefungsbefragung Mobilität: Einladung: 16.07.2015, Erinnerung: 30.07.2015

steigernde Maßnahmen vor allem in den beiden Haupterhebungen eingesetzt. So wurde bereits im Anschreiben angekündigt, eine Zusammenfassung der zentralen Ergebnisse der Studie im Anschluss an die Erhebung zu versenden. In der ersten Befragungswelle wurde dem Anschreiben zudem ein Flyer beigelegt, in der zweiten Befragungswelle erfolgte dies nur beim postalischen Versand. Des Weiteren wurden auf einer Projekthomepage Informationen zum Projekt und daraus entstandenen Publikationen bereitgestellt. Darüber hinaus wurde unter allen Befragungsteilnehmer(inne)n eine Verlosung verschiedener Sachpreise durchgeführt. In der ersten Befragungswelle wurden dabei ein Notebook, fünf iPod nano und fünf USB-Sticks verlost, in der zweiten Welle ein Notebook, zehn iPod nano, 15 Bahngutscheine à 50 Euro und 20 USB-Sticks.

5 Rücklauf

[Rücklauf] Die Bruttostichprobe der ersten Befragungswelle umfasste 52.550 Absolvent(inn)en, die bei den Prüfungsämtern der im Stichprobenverfahren gezogenen Hochschulen (vgl. Kapitel 3) gemeldet waren und kontaktiert wurden, darunter 27.702 Bachelorabsolvent(inn)en und 24.848 Absolvent(inn)en traditioneller Studiengänge. Diese stammen von 237 Hochschulen aus allen Bundesländern.

Nachdem von den an das DZHW zurückgeschickten Fragebögen diejenigen entfernt worden waren, die nicht zur Grundgesamtheit gehörten (stichprobenneutrale Ausfälle) oder nicht auswertbar waren (stichprobenrelevante Ausfälle) (vgl. Kapitel 6.3), verblieben in der Nettostichprobe der ersten Befragungswelle 10.494 Fälle, davon 4.883 Bachelorabsolvent(inn)en und 5.611 Absolvent(inn)en traditioneller Studiengänge. Im Hinblick auf die Bruttostichprobe liegt die Rücklaufquote damit insgesamt bei 20 Prozent. Für die Absolvent(inn)en traditioneller Studiengänge (22,6 %) ist sie etwas höher als für die Bachelorabsolvent(inn)en (17,4 %).

Die Rücklaufzahlen zur zweiten Befragungswelle (inkl. Vertiefungsbefragungen), die aufgrund von Sperrfristen erst im Oktober 2017 veröffentlicht wird, werden nachgereicht.⁴⁹

⁴⁹ Für die Hauptbefragung der zweiten Welle werden jedoch voraussichtlich 4.755 Fälle, darunter 2.110 Bachelorabsolvent(inn)en und 2.645 Absolvent(inn)en traditioneller Studiengänge, zur Verfügung stehen. Die Vertiefungsbefragung zum Thema Mobilität umfasst voraussichtlich 2.471 Fälle und die Vertiefungsbefragung zum Thema Promotion 676 Fälle.

6 Datenaufbereitung

Im Folgenden werden die verschiedenen Schritte der Datenaufbereitung beschrieben. 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

[PAPI-Befragungen] Bei den PAPI-Befragungen mussten die Angaben der Befragten aus den Papierfragebögen zunächst auf Basis eines Codeplans zur weiteren Verarbeitung in ein computerlesbares Format übertragen werden. Zu diesem Zweck wurde auf dem Fragebogen vermerkt, welcher Frage bzw. Teilfrage eine Variable zugeordnet ist, welchen Namen diese Variable trägt und welche numerischen Codierungen für die Antworten der Befragten verwendet werden sollten. Dabei wurden auch bereits numerische Codierungen für die offenen Angaben vermerkt (vgl. Kapitel 6.2).⁵⁰ Um die Erfassungsreihenfolge festzulegen, wurden die Variablen zusätzlich nummeriert.

In der ersten Befragungswelle 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. In der zweiten Befragungswelle wurden die Daten der Papierfragebögen mit der Erfassungssoftware EpiData zweifach⁵¹ von jeweils unterschiedlichen DZHW-Mitarbeiter(inne)n erfasst. Die doppelt erfassten Daten wurden in Stata abgeglichen, bei Abweichungen erneut geprüft und bereinigt.

[Online-Befragungen] Die Angaben aus den Online-Befragungen konnten direkt aus der Befragungssoftware als .csv-Datei exportiert und weiterverarbeitet werden.

6.2 Codierung offener Angaben

Die (halb-)offenen Angaben wurden nur in codierter Form in das SUF/CUF aufgenommen. Die durch das Primärforschungsprojekt vorgenommenen Codierungsentscheidungen wurden dabei unverändert beibehalten. 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 Absolventenpanels eingesetzte Schlüssel. Für einige Variablen wurden neue Codierlisten auf Basis der in den Daten des Absolventenpanels 2009 vorkommenden Nennungen entwickelt. Für einige halboffene Fragen wurden keine neuen Variablen mit numerischen Codierungen erstellt, sondern die Nennungen nur – sofern möglich – den vorhandenen (ge-

⁵⁰ In der ersten Befragungswelle wurden zudem bereits manuelle Vorkorrekturen zur Erleichterung der Datenübertragung vorgenommen (vgl. Kapitel 6.3).

⁵¹ mit Ausnahme der Kalenderdaten, die nur einfach erfasst wurden

schlossenen) Antwortkategorien zugeordnet. Einzelne offene Fragen wurden nicht vercodet, weil sie als Kontextinformationen zur Codierung anderer offener Angaben erfasst wurden oder aufgrund von unzureichenden zeitlichen Ressourcen.⁵²

In Tabelle 2 sind die codierten Merkmale sowie die jeweils verwendete Codierliste 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 Metadatensuchsystem⁵³ dokumentiert.

Für die Papierfragebögen erfolgte die Codierung der (halb-)offenen Angaben bereits im Rahmen der Datenübertragung (vgl. Kapitel 6.1). Für die Online-Daten wurde die Codierung in der Regel mit Excel durchgeführt. Damit fand die Codierung der PAPI-Daten in der zweiten Befragungswelle zwar technisch getrennt von der Codierung der Online-Daten statt, für die Erstellung neuer Codierlisten wurden jedoch immer die Daten beider Erhebungsmodi berücksichtigt. Eine Ausnahme von diesem getrennten Vorgehen wurde lediglich bei der Codierung der offenen Berufsangaben in der zweiten Befragungswelle vollzogen. Hier wurden die offenen Angaben aus den Papierfragebögen und Online-Daten zusammengespielt und mit Hilfe einer Codierungssoftware, in der die Klassifikation der Berufe von 2010 hinterlegt war, gemeinsam codiert. Die Software weist dabei soweit möglich automatisiert Berufscodes zu, für die verbleibenden Nennungen können zusätzlich manuell Codes vergeben werden.

⁵² Dies betrifft insbesondere die offenen Angaben in der Vertiefungsbefragung zur Mobilität, teilweise aber auch offene Nennungen in den anderen Befragungen. So wurden in der ersten Befragungswelle auf die Codierung der Studienschwerpunkte (Frage 1.8) verzichtet. Des Weiteren erfolgte sowohl in der ersten Befragungswelle als auch in der Hauptbefragung der zweiten Befragungswelle keine Codierung der typischen Arbeitsschwerpunkte, die neben der Berufsbezeichnung und dem Aufgabenbereich erhoben wurden (Frage 5.2 in Welle 1 bzw. Frage 4.9 in Welle 2 (Papierversion)). Die beruflichen Aufgabenbereiche in der gleichen Frage wurden nur für die erste Welle vercodet.

⁵³ <https://metadata.fdz.dzhw.eu>

Tabelle 2: Vercodete Merkmale und Codierlisten im DZHW-Absolventenpanel 2009

Merkmals	Codierliste	Codierlisten-ID ^b
Studienfach	Destatis Fächersystematik 2008/2009 [gemäß Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2008/2009 und SoSe 2009), Schlüssel 3.1]	cl-destatis-studienfach-2009 ^c
Studienabschluss	projekteigene Codierung	cl-dzhw-22
Hochschule	Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2014/2015 und SoSe 2015), Schlüssel 2.2	cl-destatis-hochschule-2015 ^d
Bundesland	Destatis Bundeslandschlüssel (entsprechend der ersten beiden Ziffern des Amtlichen Gemeindegchlüssels (AGS))	cl-destatis-bundesland-1990 ^d
Ausland / Staatsangehörigkeit	Welle 1: projekteigene Codierung	cl-dzhw-23
	Welle 2: Destatis-Staatsangehörigkeits- und Gebietsschlüssel 2015	cl-destatis-ausland-2015
Berufsbezeichnung	Beruf der Befragten: Destatis Klassifikation der Berufe 2010	cl-destatis-kldb-2010 ^d
	Ausbildungs-/Elternberuf: Destatis Klassifikation der Berufe 1992	cl-destatis-kldb-1992
Beruflicher Aufgabenbereich ^a	projekteigene Codierung	cl-dzhw-4
Schulform Lehramtsausrichtung	projekteigene Codierung	cl-dzhw-21
sonstige offene Abfragen	Zuordnung zu vorgegebenen Kategorien, projekteigene Codierung oder Löschung	---

^a vgl. Frage 5.2 in Welle 1 (in Welle 2 wurden die Angaben zum beruflichen Aufgabenbereich nicht vercodet)

^b 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.

^c ergänzt um einen Code aus dem Schlüsselverzeichnis 2012/2013

^d ergänzt um projekteigene Codes, wenn nicht zuordenbar

6.3 Datenprüfung und Datenbereinigung

[Konsistenzprüfungen] Für beide Befragungswellen wurden – insbesondere für die beiden Hauptbefragungen – verschiedene Konsistenzprüfungen durchgeführt. Folgende Typen von Prüfungen wurden vorgenommen:

- *Prüfung von Wertebereichen:* Es wurde geprüft, ob die erfasste Ausprägung einer Variablen in dem für diese Variable definierten Wertebereich lag.

- *Prüfung der Einhaltung der Filterführung:* Auf Grundlage der definierten Filterführung des Fragebogens wurde auf der einen Seite geprüft, ob für die jeweilige befragte Person Angaben zu erwarten gewesen wären, die aber nicht vorhanden waren (Vollständigkeitsprüfung), und auf der anderen Seite, ob für die jeweilige Person, Angaben vorhanden waren, die nicht hätten erfolgen dürfen (Filterverstöße).⁵⁴
- *Prüfung von Merkmalskombinationen:* Es wurde die Konsistenz der Angaben innerhalb eines Fragebogens sowie zwischen den Befragungen geprüft. Insbesondere wurden dabei inhaltliche und zeitbezogene Angaben zu gleichen Themengebieten miteinander abgeglichen.⁵⁵ Beispielsweise wurde innerhalb beider Wellen jeweils geprüft, ob die Angaben über Berufstätigkeiten im Erwerbstätigkeitstableau mit den entsprechenden Angaben im Kalendarium übereinstimmten.

Festgestellte Inkonsistenzen wurden – falls möglich – durch den Abgleich mit anderen Nennungen im Fragebogen aufgelöst oder andernfalls ein entsprechender Missingcode (vgl. Kapitel 6.7) vergeben.

In der ersten Befragungswelle wurden erste Konsistenzprüfungen bereits vor der Übertragung der Daten manuell auf den Papierfragebögen vorgenommen.⁵⁶ Darauf aufbauend erfolgte im Anschluss an die Datenübertragung eine umfassende Prüfung und Korrektur der Daten mit Hilfe einer DZHW-eigenen Software.⁵⁷ In der zweiten Befragungswelle erfolgte die Konsistenzprüfung zum einen ebenfalls softwaregestützt, zum anderen über Stata-Do-Files.

[Löschung von Fällen] In beiden Wellen wurden Fälle aus dem Datensatz entfernt. Ein Fall wurde gelöscht, wenn rund die Hälfte der Fragen bzw. zentrale Fragen (z. B. zum Studienfach) nicht beantwortet worden oder 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.⁵⁸ Diese wurden ebenfalls aus dem Datensatz entfernt.⁵⁹

⁵⁴ Die Eingangsfilter der den einzelnen Fragen zugeordneten Variablen sind im Datensatzreport sowie im Metadaten suchsystem (<https://metadata.fdz.dzhw.eu>) dokumentiert. Sie definieren für die jeweilige Variable, welche Befragtengruppe die zugehörige Frage beantworten sollte.

⁵⁵ Da die Vertiefungsbefragung zur Mobilität aufgrund der thematisch klar abgegrenzten Inhalte kaum Überschneidungen mit der Hauptbefragung der zweiten Welle hatte, war eine intensive Plausibilisierung im Abgleich der Daten nicht notwendig. Die Vertiefungsbefragung zur Promotion hingegen überschneidet sich in einigen Punkten (wie z. B. den Angaben zu Promotionsdaten und zum akademischen Werdegang) durchaus mit der Hauptbefragung.

⁵⁶ Die Zahl der vorgenommenen Korrekturen wurde nicht zentral, sondern nur auf den Papierfragebögen dokumentiert und ist daher nicht mehr systematisch rekonstruierbar.

⁵⁷ Zu diesem Zweck wurden die erfassten Fragebogen-Daten in eine Datenbank eingelesen. Anschließend wurden anhand formaler Regeln gültige Wertebereiche und Antwortkombinationen definiert und geprüft.

⁵⁸ Dies kam beispielsweise vor, wenn die Prüfungsämter versehentlich Personen angeschrieben hatten, die einem anderen Prüfungsjahrgang angehörten oder wenn es sich um Absolvent(inn)en handelte, die bereits einen weiteren Abschluss hatten.

⁵⁹ Darauf hinzuweisen ist jedoch, dass auch noch einige Fälle im Datensatz enthalten sind, bei denen das Abschlussdatum einige Monate hinter dem eigentlichen Prüfungszeitraum des Jahrgangs 2009 liegt. Diese wurden beibehalten, weil es an einigen Hochschulen üblich ist, dass Fälle noch dem vorherigen Prüfungsjahr zugerechnet werden, wenn sich eine Prüfung ohne Zutun der Absolvent(inn)en verschoben hat (z. B. durch Erkrankung des Prüfers). Umgekehrt gibt es auch Fälle, bei denen das Abschlussdatum einige Monate vor dem eigentlichen Prüfungszeitraum des Jahrgangs 2009 liegt, die Fälle aber von den Hochschulen trotzdem zum Prüfungsjahrgang 2009 gerechnet wurden, weil z. B. einige Formalitäten fehlten (z. B. ein Praktikumsnachweis) oder Nachprüfungen notwendig waren. Aufgrund dieser Praxis an den Hochschulen wurden Fälle, die etwas vom eigentlichen Prüfungsjahr abwichen, nur dann gelöscht, wenn sie offensichtlich nicht zur Grundgesamtheit gehörten.

6.4 Generierung von Variablen

Neben den Variablen, die die codierten Antworten der Befragten enthalten, beinhaltet der Datensatz des Absolventenpanels 2009 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 oder mehrerer Quellvariablen generiert (z. B. Aggregation der Studienfächer zu Studienbereichen und Fächergruppen oder Ableitung von Hochschultyp und Hochschulort aus den Hochschulvariablen). Der Variablenname einer generierten Variablen ist im Datensatz durch das Suffix „_g#“ gekennzeichnet. Eine Übersicht aller für das Absolventenpanel 2009 generierten Variablen sowie eine detaillierte Dokumentation der einzelnen Variablen mit Angabe ihrer jeweiligen Ausprägungen und Berechnungsvorschriften findet sich im Datensatzreport sowie im Metadaten-System⁶⁰.

Generierte Variablen wurden im Datensatz – sofern möglich – nach der jeweiligen Ausgangsvariable positioniert. Wurde eine Variable aus verschiedenen Quellvariablen generiert, wurde sie hinter jene Variable eingefügt, die ihr thematisch am nächsten ist. Falls eine eindeutige Zuordnung 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 ersten und zweiten Befragungswelle (inkl. Vertiefungsbefragungen) wurden zusammengeführt.⁶¹ Die Zuordnung der Fälle erfolgte über die im Rahmen der Feldphase vergebenen Identifikationsnummern der Befragten (vgl. Kapitel 4).

[Erstellung von Personen- und Episodendatensatz] Die so zusammengeführten Daten wurden in zwei getrennten Datensätzen abgelegt. Der *Personendatensatz* enthält den Großteil der Befragungsdaten sowie die zusätzlich generierten Variablen. Pro befragter Person existiert eine Datenzeile (wide-Format). Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehörigen Fragen im Fragebogen. Der *Episodendatensatz* enthält nur die Antworten aus den Kalendarien (Frage 4.1 der ersten Welle, Frage 1.5 bzw. Page 5 der zweiten Welle). Für jede befragte Person werden ein oder mehrere Episoden gespeichert. Dabei ist eine Episode definiert als ein Zeitraum, in dem eine bestimmte Tätigkeitsart (z. B. Erwerbstätigkeit, Ausbildung) ausgeübt wird bzw. ein konkreter Status (z. B. Elternzeit, Arbeitslosigkeit) besteht. Für jede Episode einer Person existiert jeweils eine Datenzeile (long format). Die Struktur entspricht der gängigen Struktur für Episodendaten (vgl. Scherer & Brüderl, 2010, S. 1042). Die

⁶⁰ <https://metadata.fdz.dzhw.eu>

⁶¹ Darauf hinzuweisen ist, dass in den Kalendarien und Erwerbstätigkeitstableaus der beiden Befragungswellen bewusst Überschneidungen konzipiert worden waren. So wurden im Rahmen der ersten Welle die jeweiligen Tätigkeiten seit dem Studienabschluss bis zum Befragungszeitpunkt erfasst. Dieser lag – je nachdem, wann die Befragten den Fragebogen ausfüllten – zwischen Februar 2010 und Januar 2011 (vgl. Kapitel 4). In der zweiten Welle setzte das Kalendarium bzw. das Erwerbstätigkeitstableau Anfang 2010 an, sodass Tätigkeiten unter Umständen mehrfach erfasst wurden. Die sich überschneidenden Monate wurden grundsätzlich aus den Daten der zweiten Welle gestrichen und entsprechend die Angaben aus der ersten Welle beibehalten.

Episoden wurden fallweise sortiert, das heißt alle Episoden einer Person folgen direkt aufeinander. Verschiedene Tätigkeitsarten im selben Zeitraum wurden jeweils als eigenständige Episode codiert. Wenn Tätigkeiten derselben Art unmittelbar aufeinander folgten oder parallel ausgeübt wurden, wurden sie zu einer Episode zusammengefasst. Daher geht aus den Episodendaten nicht hervor, ob eine Episode eine oder mehrere Tätigkeiten derselben Art umfasst. Für Episoden der Tätigkeitsarten Erwerbstätigkeit und akademische Weiterqualifizierung sind detailliertere Informationen jedoch in den entsprechenden Variablen des Personendatensatzes enthalten. Die Daten dieser Variablen können mit den Episodendaten verbunden werden. Das Zusammenführen von Personendatensatz und Episodendatensatz wird über die Identifikationsnummer der Person (Variable: *pid*) ermöglicht.

[Dateiformat] Alle Datensätze werden sowohl im Stata- als auch im SPSS-Format bereitgestellt (vgl. 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 liegen auf Deutsch und auf Englisch vor. Im SPSS-Format existiert für jede Sprache ein eigener Datensatz. Im Stata-Format wurden zweisprachige Labels im gleichen Datensatz hinterlegt.

[Variablenbenennung im Personendatensatz] Für die Variablenbenennung wurde im FDZ-DZHW eine einheitliche Benennungssystematik erstellt. Mit Ausnahme der Identifikatorvariablen *pid* sowie der Wellenvariablen *wave*⁶² wurden die Variablennamen im Personendatensatz nach einem Präfix-Stamm-Suffix-Schema, das eine automatisierte Verarbeitung erleichtert, gebildet. Zudem liefern die Variablennamen Metainformationen zur entsprechenden Variable. Das Präfix der Variable enthält die Erhebungskennung anhand eines Buchstabens.⁶³ Im Stamm geht der Themenbereich, dem die Variable zugeordnet ist, aus einem dreistelligen englischen Buchstabenkürzel hervor. Tabelle 3 stellt die verschiedenen Themenbereiche des Absolventenpanels 2009 sowie das zugehörige 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 verschiedenen Datenzugangswegen. Für Indikatoren, die in beiden Befragungswellen verwendet werden, wurden die Namen der zugehörigen Variablen durch die Vergabe eines identischen Stammes harmonisiert. Detaillierte Informationen zur Variablenbenennung im Absolventenpanel 2009 befinden sich im Datensatzreport.

⁶² Diese enthält die Information, welche Fälle nur an der ersten Welle bzw. an beiden Wellen teilgenommen haben.

⁶³ Welle 1: **a**
 Welle 2 - Hauptbefragung: **b**
 Welle 2 - Vertiefungsbefragung Mobilität: **m**
 Welle 2 - Vertiefungsbefragung Promotion: **p**

Tabelle 3: Themengebiete und Kürzel für Variablennamen des DZHW-Absolventenpanels 2009

Themengebiete-Kürzel	Themengebiet (englisch)	Themengebiet (deutsch)
stu	studies	Studium
occ	occupation	Beschäftigung
ski	skills	Fähigkeiten
fvt	further vocational training	Berufliche Fort- und Weiterbildung
fec	further education	Aus- und Weiterbildung
dem	demographic information	demographische Informationen
abr	(experiences) abroad	Auslandserfahrung
mov	move	Umzug
per	personality	Persönlichkeit
res	residence	Wohnsitz
rsa	research activity	Forschungsaktivität
wgt	weights	Gewichtungsvariablen
sys	system variables	Systemvariablen

[Variablenbenennung im Episodendatensatz] Die Variablen im Episodendatensatz sind die Identifikationsnummer der befragten Person (pid), die Identifikationsnummer der jeweiligen Episode (eid), die ausgeübte Tätigkeitsart (status) sowie Beginn und Ende des Episodenzeitraums, der über vier Variablen (Monat: begin_m und end_m; Jahr: begin_y; end_y) codiert wird.

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 wurden dabei durch dreistellige negative Werte codiert. Tabelle 4 stellt die verwendete Missingsystematik dar. Die im Absolventenpanel 2009 verwendeten Missingcodierungen sind hervorgehoben.

Sie lassen sich vier verschiedenen Gruppen zuordnen. In den ersten beiden Gruppen wird zwischen fehlenden Werten aufgrund von Nicht-Beantwortung von Fragen seitens der Befragten (Nonresponse) und fehlenden Werten aufgrund der Filterführung bzw. für Befragte nicht relevanten Fragen unterschieden (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 gehört auch die Codierung, die aufgrund von Anonymisierungsmaßnahmen (vgl. Kapitel 8) für bestimmte Variablen vergeben wurde. Die vierte Gruppe umfasst spezielle Missingcodierungen, die im Rahmen der Datenaufbereitung eines konkreten Datensatzes nur für einzelne Items vergeben wurden (z. B. „läuft noch“ bei den Items astu012c und astu012d, Frage 1.1, 1. Welle).

Tabelle 4: 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	-948	läuft noch
-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 Ausprägungen dieser Missingkategorie 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. Im Anschluss wird die Gewichtungsprozedur im Detail beschrieben und es werden Anwendungshinweise gegeben.

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

- **Designbedingte Verzerrung:** Disproportionalitäten werden bewusst erzeugt (zur Erhöhung der Fallzahl in Subgruppen) oder ergeben sich aus erhebungsbedingten Ursachen (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 Ausfallsprozesse unsystematisch sind (Missing Completely at Random), können sie ignoriert werden.⁶⁴ Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (Missing at Random, Not Missing at Random), der einer Modellierung bedarf.⁶⁵

[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 Stichprobenplan. 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) erzeugt werden. In einem letzten Schritt können die nonresponse-adjustierten Designgewichte anhand von Merkmalsverteilungen aus der Grundgesamtheit kalibriert werden.

[Querschnittsgewichtung] Für das Absolventenpanel 2009 wurden drei Querschnittsgewichte gebildet: ein Gesamtgewicht für die Auswertung aller Absolvent(inn)en und zwei gesonderte Gewichte für die Auswertung von Absolvent(inn)en traditioneller Studiengänge bzw. von Bachelorabsolvent(inn)en.

Das Designgewicht ergibt sich dabei für jede Schicht s wie folgt⁶⁶:

$$dwgt_{sci} = \frac{n_{sc}^{-1}}{N_{sc}}$$

⁶⁴ Insofern die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

⁶⁵ Zu den unterschiedlichen Formen von Ausfallprozessen siehe grundlegend Rubin, 1976.

⁶⁶ Wobei n_{sc} der Anzahl der Klumpen in einer Schicht und N_{sc} der Anzahl der Klumpen in der jeweiligen Schicht der Grundgesamtheit entspricht. Da die Klumpen vollständig erhoben wurden, entspricht die Auswahlwahrscheinlichkeit eines Individuums der Auswahlwahrscheinlichkeit des zugehörigen Klumpens.

Aufgrund der fehlenden Informationen über die Nichtteilnehmer(innen) in der ersten Welle, war auch für den Ausfallprozess durch Nichtteilnahmen (Nonresponse) keine umfängliche Adjustierung der Designgewichte auf Individualebene möglich. Jedoch wurde eine Kalibrierung vorgenommen, welche die Anpassung der Designgewichte an Merkmalsträger der Grundgesamtheit zum Ziel hatte. Die Kalibrierung erfolgte sowohl für die vollständige Stichprobe als auch anhand von Randverteilungen aller Bachelorabsolventen sowie aller Absolventen traditioneller Studiengänge. Als Merkmale wurden die Region (Ost/West), das Geschlecht, die Fächergruppen sowie für das Gesamtgewicht zusätzlich auch die Abschlussart herangezogen.⁶⁷ Da die Merkmalsträger in der Grundgesamtheit ebenfalls Informationen über Nichtteilnehmer(innen) enthielten, erfolgt durch die Kalibrierung zusätzlich eine Nonresponse-Adjustierung im Hinblick auf die verwendeten Merkmale. Die Kalibrierung des Designgewichts $dwgt_{sci}$ wurde mittels des Raking-Algorithmus⁶⁸ durchgeführt.

[Längsschnittgewichtung] Für die Hauptbefragung der zweiten Welle wurde zusätzlich ein Nonresponsegewicht berechnet, das mit Hilfe von Informationen aus der ersten Welle die Nichtteilnahme in der zweiten Welle modelliert. Zusammengenommen mit dem Querschnittsgewicht der ersten Welle ergibt sich hieraus das Längsschnittgewicht für das Zwei-Wellen-Panel. Im Unterschied zu vorangegangenen Absolventenbefragungen wurden im Rahmen der zweiten Befragungswelle des Absolventenpanels 2009 im Anschluss an die Hauptbefragung auch zwei Vertiefungsbefragungen durchgeführt, die bezüglich der Gewichtung wie eigenständige Befragungswellen behandelt werden müssen. Dementsprechend wurden auch für diese beiden Erhebungen Nonresponse-Adjustierungen vorgenommen. Aufgrund von Sperrfristen wird die Veröffentlichung der Daten der zweiten Befragungswelle (inkl. Vertiefungsbefragungen) sowie der zugehörigen detaillierten Informationen zur Gewichtung erst im Oktober 2017 erfolgen.

[Hinweise zur Anwendung der Gewichte] In Tabelle 5 sind die erstellten Gewichte für die erste Erhebungswelle dargestellt.

Tabelle 5: Bereitgestellte Gewichte zum DZHW-Absolventenpanel 2009

Variablenname	Beschreibung
wgt01_t1d	Querschnittsgewicht Welle 1: Absolvent(inn)en insgesamt
wgt02_t1d	Querschnittsgewicht Welle 1: Absolvent(inn)en traditioneller Studiengänge
wgt03_t1d	Querschnittsgewicht Welle 1: Bachelorabsolvent(inn)en

Bei den erstellten Gewichten handelt es sich um probability weights, die in Stata mit Hilfe .ado-spezifischer Optionen berücksichtigt werden können.⁶⁹ Das Gewicht wgt01_t1d ist für Auswertungen aller Absolvent(inn)en vorgesehen. Für die getrennte Auswertung von Absolvent(inn)en traditioneller Studiengänge und Bachelorabsolvent(inn)en sind die Gewichte wgt02_t1d bzw. wgt03_t1d zu verwenden. Grundlegend ist zu beachten, dass Gewichte nur dann sinnvolle Korrekturgrößen darstellen, wenn das verwendete Analysemodell die zur Gewichtung herangezogenen Variablen enthält oder mit diesen in einem Zusammenhang steht.

⁶⁷ Die Informationen aus der Grundgesamtheit wurden aus Daten des Statistischen Bundesamtes (Prüfungstatistik 2008/2009) abgeleitet.

⁶⁸ Raking ist auch unter dem Begriff des iterative proportional fitting (ipf) bekannt (vgl. Kolenikov, 2014).

⁶⁹ Siehe hierzu die Stata-Hilfe (Befehl: *help weights*).

Aus diesem Grund müssen Gewichte immer mit Fokus auf die analysierte Fragestellung verwendet werden. Die hier erstellten Gewichte beziehen sich auf die Bundesebene, Auswertungen auf der Ebene einzelner Bundesländer oder Regionen sind mit diesen Gewichten somit nicht sinnvoll durchführbar.

8 Anonymisierung

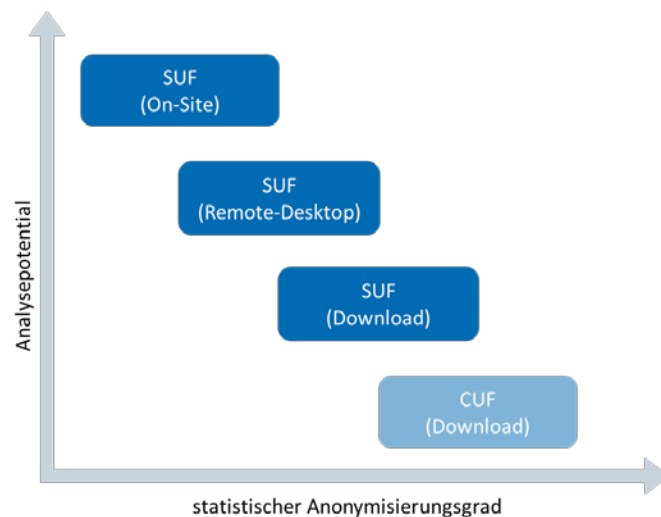
[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten, die in freiwilligen Befragungen durch das DZHW erhoben werden, gilt das Bundesdatenschutzgesetz (BDSG).⁷⁰ Danach dürfen personenbezogene Daten, die im Rahmen wissenschaftlicher Forschung erhoben worden sind, ausschließlich zum Zweck wissenschaftlicher 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 bestimmbarer 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 Absolventenpanel 2009 ein faktisch anonymisiertes SUF für die wissenschaftliche Sekundärnutzung und ein absolut anonymisiertes CUF für Lehr- und Übungszwecke 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.

Während das CUF nach einer Registrierung direkt durch das FDZ-DZHW übermittelt wird, wird das SUF über drei verschiedene Zugangswege angeboten: Download, 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/CUF) und Zugangsweg erläutert.

⁷⁰ 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-Absolventenpanels 2009



[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 Absolventenpanel 2009 bereits während der Feldphase in einem separaten Datensatz erfasst (vgl. Kapitel 4) und sind somit weder im CUF noch in den verschiedenen SUF-Varianten enthalten. Um einen Rückbezug auf diesen Datensatz zu unterbinden, wurde zudem die Original-Identifikationsnummer entfernt und durch eine neue zufällig vergebene Identifikationsnummer ersetzt.

Anschließend wurden die *Quasi-Identifikatoren* bestimmt, also Informationen, die in Kombination oder durch die Anspielung externer Informationen geeignet sind, eine Person indirekt zu identifizieren.⁷¹ Für das Absolventenpanel 2009 wurden die folgenden Quasi-Identifikatoren identifiziert, die sowohl in externen Datenquellen⁷² als auch in den Absolvent(inn)endaten vorliegen: Hochschule, Studienfach, Abschlussart, Berufsangaben, regionale Informationen (zur Hochschule, zum Ort des Erwerbs der Studienberechtigung und Arbeitsort), Staatsangehörigkeit und Geburtsland. Um eine eindeutige Zuordnung der Absolvent(inn)endaten zu unterbinden, wurden diese Schlüsselmerkmale – je nach Datenprodukt bzw. Zugangsweg – aggregiert oder gelöscht (vgl. Tabelle 6). Beispielsweise wird das Merkmal „Hochschule“ in dem SUF für die On-Site Nutzung zu NUTS-2-Regionen, im Remote-Desktop-SUF zu Bundesländern, im Download-SUF und im Download-CUF zu den zwei Kategorien alte vs. neue Bundesländern aggregiert. Offene Angaben sind ebenfalls Quasi-Identifikatoren (vgl. Ebel, 2015, S. 3) und wurden im Rahmen der Anonymisierung vercodet oder gelöscht.

⁷¹ 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.

⁷² z.B. Studenten- und Prüfungsstatistik des statistischen Bundesamtes, Alumninetzwerke der Hochschulen oder auch Berufsnetzwerke.

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 Absolventenpanel 2009 wurden Gesundheitsinformationen erhoben, für die bei den Befragten kein zusätzliches Einverständnis für die Sekundärnutzung eingeholt wurde. Daher wurden diese Antworten im CUF- und allen SUF-Varianten gelöscht.

Zur Gewährleistung der absoluten Anonymisierung der Daten des CUF wurden zum einen im Vergleich zu den SUF-Varianten restriktivere statistische Anonymisierungsmaßnahmen auf Variablenebene vorgenommen (vgl. Tabelle 6). Zum anderen wurde eine per Zufallsauswahl gewonnene Substichprobe der Daten (10 Prozent der befragten Absolvent(inn)en traditioneller Studiengänge) gezogen.

Tabelle 6: Maßnahmen der statistischen Anonymisierung der Daten des DZHW-Absolventenpanels 2009 nach Zugangsweg⁷³

Merkmal	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (Substichprobe)
Direkte Identifikatoren	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID
Fragebogeneingang	Freigabe	Löschung	Löschung	Löschung
Studienfach	Freigabe	Aggregation zu Studienbereichen ^a	Aggregation zu Studienbereichen ^a	Aggregation zu Fächergruppen ^a
Hochschule	Aggregation zu Hochschulart und Hochschulort zu NUTS 2: Basisregionen für regionalpolitische Maßnahmen ^b	Aggregation zu Hochschulart und Hochschulort zu Bundesländern	Aggregation zu Hochschulart und Hochschulort zu neuen bzw. alten Bundesländern	Aggregation zu Hochschulart und Hochschulort zu neuen bzw. alten Bundesländern
Arbeitsort (Bundesland/Ausland)	Freigabe	Freigabe	Aggregation zu Bundesländern und Ausland	Aggregation zu alte, neue Bundesländer und Ausland
Arbeitsort (PLZ)	Freigabe	Aggregation zu NUTS 2: Basisregionen für regionalpolitische Maßnahmen ^b	Aggregation zu NUTS 2: Basisregionen für regionalpolitische Maßnahmen ^b	Löschung

⁷³ Detaillierte Informationen zu den anonymisierten Variablen sind dem Datensatzreport sowie dem Metadaten-suchsystem (<https://metadata.fdz.dzhw.eu>) zu entnehmen.

Merkmal	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (Substichprobe)
Ort der Studienberechtigung (Bundesland/ Ausland)	Freigabe	Freigabe	Aggregation zu Bundesländern und Ausland	Aggregation zu alte, neue Bundesländer und Ausland
Ort der Studienberechtigung (PLZ)	Freigabe	Aggregation zu NUTS 2: Basisregionen für regionalpolitische Maßnahmen ^b	Aggregation zu NUTS 2: Basisregionen für regionalpolitische Maßnahmen ^b	Löschung
Wohnsitz (PLZ)	Freigabe	Aggregation zu NUTS 2: Basisregionen für regionalpolitische Maßnahmen ^b	Aggregation zu NUTS 2: Basisregionen für regionalpolitische Maßnahmen ^b	Löschung
Wohnsitz (Bundesland/ Ausland)	Freigabe	Freigabe	Aggregation zu Bundesländern und Ausland	Aggregation zu alte, neue Bundesländer und Ausland
Beruf	Vercodung zu Berufsgattungen (5-Steller) ^c	Aggregation zu Berufsgruppen (3-Steller) ^c	Aggregation zu Berufsgruppen (3-Steller) ^c	Aggregation zu Berufshauptgruppen (2-Steller) ^c
Ausbildungsberuf	Vercodung zu Berufsordnungen (3-Steller) ^d	Vercodung zu Berufsordnungen (3-Steller) ^d	Vercodung zu Berufsordnungen (3-Steller) ^d	Aggregation zu Berufsgruppen (2-Steller) ^d
Beruf Eltern	Vercodung zu Berufsordnungen (3-Steller) ^d	Vercodung zu Berufsordnungen (3-Steller) ^d	Vercodung zu Berufsordnungen (3-Steller) ^d	Aggregation zu Berufsgruppen (2-Steller) ^d
Staatsangehörigkeit (Ausland)	Freigabe	Aggregation nach NEPS Klassifikation ^e	Aggregation zu Weltregionen	Löschung
Geburtsland (Ausland)	Freigabe	Aggregation nach NEPS Klassifikation ^e	Aggregation zu Weltregionen	Löschung
Alter	Freigabe	Freigabe	Freigabe	Top-Codierung ^f
Merkmale zur Gesundheit	Löschung	Löschung	Löschung	Löschung
sonstige offene Angaben ^g	Vercodung/ Löschung	Vercodung/ Löschung	Löschung	Löschung

^a Nach dem Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2008/2009 und SoSe 2009 von Destatis.

^b Statistisches Amt der Europäischen Union (Eurostat): Nomenclature of territorial units for statistics (NUTS) <http://ec.europa.eu/eurostat/web/nuts/overview>.

^c Nach der Klassifikation der Berufe von 2010 von Destatis.

^d Nach der Klassifikation der Berufe von 1992 von Destatis.

^e Die Aggregation der Staaten zu Weltregionen ist angelehnt an die Klassifikation des NEPS https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/6-0-0/SC5_6-0-0_Anonymisation.pdf (in der ersten Welle mit Anpassungen bei europäischen Ländern).

^f Dabei wurden Altersangaben, die oberhalb einer bestimmten Grenze lagen, zu einer Kategorie aggregiert.

^g Einzelne Angaben wurden auch im On-Site- und Remote-SUF nicht vercodet, sondern komplett gelöscht (vgl. Kapitel 6.2).

9 Literaturverzeichnis

- Blumenstiel, J. E. & Gummer, T. (2015). Prävention, Korrektur oder beides? Drei Wege zur Reduzierung von Non-response 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
- Briedis, K., Brandt, G., Fabian, G. & Rehn, T. (2011). Bachelorabsolventen im Fokus. In K. Briedis, C. Heine, C. Konegen-Grenier & A.-K. Schröder (Hrsg.), *Mit dem Bachelor in den Beruf. Arbeitsmarktbefähigung und -akzeptanz von Bachelorstudierenden und -absolventen* (S. 53–82). Essen: Edition Stifterverband.
- Briedis, K., Klüver, S. & Trommer, M. (2016). *Zwischen Etablierung, Stabilisierung und Aufstieg: Berufliche Entwicklung der Hochschulabsolvent(inn)en 2009* (Forum Hochschule 4/2016). Hannover: DZHW.
- Dudek, K., Glässner, B. & Krause, I. (2007). *Statistische Daten zur Einführung von Bachelor- und Masterstudiengängen. Sommersemester 2007* (Statistiken zur Hochschulpolitik 1/2007), Bonn
- Dudek, K., Glässner, B. & Schröder, M. (2010). *Statistische Daten zur Einführung von Bachelor- und Masterstudiengängen. Wintersemester 2010/2011* (Statistiken zur Hochschulpolitik 2/2010), Bonn
- Ebel, T. (2015). *Empfehlungen zur Anonymisierung quantitativer Daten*. Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften.
- Fabian, G. & Briedis, K. (2009). *Aufgestiegen und erfolgreich. Ergebnisse der dritten HIS-Absolventenbefragung des Jahrgangs 1997 zehn Jahre nach dem Examen* (HIS: Forum Hochschule 2/2009). Hannover: HIS.
- Häder, M. (2015). *Empirische Sozialforschung. Eine Einführung* (3. Aufl.). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-531-19675-6
- Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. (2012). *FDZ-Methodenreport. Datenschutz am Forschungsdatenzentrum* (Nr. 06). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Kerst, C. & Fehse, S. (2007). Arbeiten unter Wert? Vertikal und horizontal inadäquate Beschäftigung von Hochschulabsolventen der Abschlussjahrgänge 1997 und 2001. *Beiträge zur Hochschulforschung*, 29 (1), 72–98.
- 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.
- Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. *Proceedings of the Survey Research Methods Section*, 225–230.
- Rammstedt, B. (2007). The 10-Item Big Five Inventory. Norm values and investigation of sociodemographic effects based on a German population representative sample. *European Journal of Psychological Assessment*, 23 (3), 193–201. doi:10.1027/1015-5759.23.3.193
- Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less. A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41 (1), 203–212. doi:10.1016/j.jrp.2006.02.001
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2013). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit. 10 Item Big Five Inventory (BFI-10). *methoden, daten, analysen*, 7(2), 233–249. doi:10.12758/mda.2013.013
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2014). *Big Five Inventory (BFI-10). Zusammenstellung sozialwissenschaftlicher Items und Skalen*. doi:10.6102/zis76
- Rehn, T., Brandt, G., Fabian, G. & Briedis, K. (2011). *Hochschulabschlüsse im Umbruch. Studium und Übergang von Absolventinnen und Absolventen reformierter und traditioneller Studiengänge des Jahrgangs 2009* (HIS: Forum Hochschule 17/2011). Hannover: HIS.
- 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.
- Saniter, N. & Siedler, T. (2014). *Door opener or waste of time? The effects of student internships on labor market outcomes* (IZA Discussion Paper No. 8141). Bonn: Forschungsinstitut zur Zukunft der Arbeit (IZA).
- Schaeper, H., Grotheer, M. & Brandt, G. (2014). Familiengründung von Hochschulabsolventinnen. Eine empirische Untersuchung verschiedener Examenskohorten. In D. Konietzka & M. Kreyenfeld (Hrsg.), *Ein Leben ohne Kinder* (2. Aufl., S. 47–80). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-94149-3_2

- Scherer, S. & Brüderl, J. (2010). Sequenzdatenanalyse. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 1031–1051). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schnell, R., Hill, P. B. & Esser, E. (2005). *Methoden der empirischen Sozialforschung* (7. Aufl.). München: Oldenbourg.
- 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