Daten- und Methodenbericht August 2022

Florence Baillet | Thorsten Euler | Andreas Franken | Henrike Schmidtchen | Anne Weber

DZHW-Absolventenpanel 2009

Daten- und Methodenbericht zu den Erhebungen der Absolvent*innenkohorte 2009 (1. - 3. Befragungswelle)



Dieses Werk steht unter der Creative Commons Namensnennung – Nicht kommerziell – Weitergabe unter gleichen Bedingungen 3.0 Deutschland Lizenz (CC-BY-NC-SA)

https://creativecommons.org/licenses/by-nc-sa/3.0/de/



Autor*innen

Florence Baillet Thorsten Euler Andreas Franken Henrike Schmidtchen Dr. Anne Weber

Impressum

Herausgeber

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW) Lange Laube 12 | 30159 Hannover | www.dzhw.eu Postfach 2920 | 30029 Hannover Tel.: +49 511 450670-960

Geschäftsführung

Prof. Dr. Monika Jungbauer-Gans Dr. habil. Thorsten Kowalke

Vorsitzender des Aufsichtsrats

Ministerialdirigent Peter Greisler

Registergericht

Amtsgericht Hannover | B 210251 Umsatzsteuer-Identifikationsnummer: DE291239300

August 2022

Inhalt

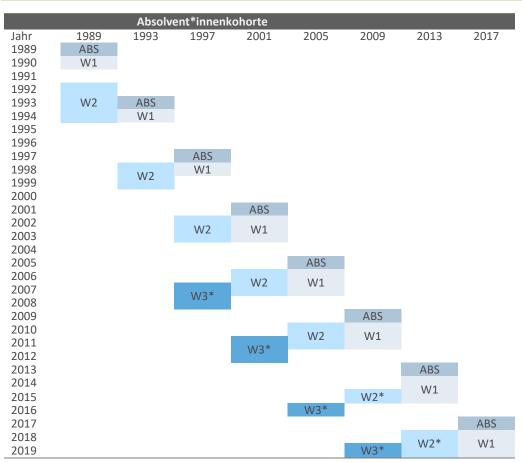
| Tab | ellen-/Ak | obildungsverzeichnis | I |
|---|---|---|----------|
| 1 | Inhalt ur | nd Anlage der Studien | 3 |
| 2 | 2.1 Inhalto | gsinstrumente e der Erhebungsinstrumente | |
| 3 | Grundge | esamtheit und Stichprobenumfang | 12 |
| 4 | Durchfül | hrung der Erhebungen | 14 |
| 5 | Rücklauf | | 17 |
| 789 | 6.1 Daten 6.2 Codien 6.3 Daten 6.4 Gener 6.5 Erstell 6.6 Vergal 6.7 Codien Gewicht Anonym | | |
| Tal | oellen-/A | Abbildungsverzeichnis Kohorten-Panel-Design der DZHW-Absolventenstudienreihe 1989-2017 Anlage der Hauptuntersuchungen der DZHW-Absolventenstudienreihe v | 3 |
| Abbi Tabe | ldung 2: elle 2: (Gesamt) | Kalendarium des DZHW-Absolventenpanels 2009, 2. Welle, Papierversion Brutto-, Nettostichproben und Rücklaufquoten des DZHW-Absolvente 18 Vercodete Merkmale und Codierlisten im DZHW-Absolventenpanel 2009 | enpanels |

| Tabelle 4: | Themengebiete und Kürzel für Variablennamen des DZHW-Absolventenpanels |
|------------------|--|
| 2009 | 24 |
| Tabelle 5: | Systematik des FDZ-DZHW für fehlende Werte (Welle 1 und 2)25 |
| Tabelle 6: | Systematik des Primärforschungsprojekts für fehlende Werte (Welle 3)25 |
| Tabelle 7: | Bereitgestellte Gewichte zum DZHW-Absolventenpanel 2009 |
| Abbildung 3: | Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der |
| Daten des DZHW- | Absolventen panels 200930 |
| Tabelle 8: | Maßnahmen der statistischen Anonymisierung der Daten des DZHW- |
| Ahsolventennanel | s 2009 nach Zugangsweg 31 |

1 Inhalt und Anlage der Studien

[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*innen erfasst werden. Das erste Absolventenpanel wurde 1989¹ durchgeführt, seitdem wird jeder vierte Absolvent*innenjahrgang (Kohorte) untersucht.

Abbildung 1: Kohorten-Panel-Design der DZHW-Absolventenstudienreihe 1989-2017



^{*}Hauptbefragung + Vertiefungsbefragungen

^{**}Abkürzungen: ABS = Abschluss, W1 = Welle 1, W2 = Welle 2, W3 = Welle 3

¹ Schon seit 1974 werden Hochschulabsolvent*innen – neben den Studienabbrecher*innen und Hochschulwechsler*innen – im Rahmen der Exmatrikuliertenbefragung des DZHW befragt. Diese Untersuchungsreihe wird seit Anfang der 2000er Jahre unter dem Namen "Studienabbruch – Umfang und Motive" durchgeführt.

Die Grundgesamtheit einer Kohorte sind Hochschulabsolvent*innen, die im Winter- oder Sommersemester des betreffenden Prüfungsjahrs ihren Studienabschluss an einer Hochschule in der Bundesrepublik Deutschland erworben haben.² Für jede Absolvent*innenkohorte 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 1).

Die Befragungen der Absolvent*innenkohorten 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 2017

| | Absolvent*innenkohorte | | | | | | | |
|-------|--|--|----------------------|-----------------------------------|----------------------|----------------------|----------------------------------|----------------------------------|
| Welle | Befragungs- zeitpunkt | Thematischer Fokus | 1989 u. 1993 | 1997 u. 2001 | 2005 | 2009 | 2013 | 2017 |
| 1 | ca. 1 Jahr nach Studienab- schluss | Studienverlauf und -erfahrung, akademische Weiterqualifi- zierung, Über- gang in den Beruf | Paper & Pencil | Paper & Pencil | Paper & Pencil | Paper & Pencil | Paper & Pencil | Online ^b |
| 2 | ca. 5 Jahre nach Studienab- schluss | Aktuelle Tätig- keit, Erwerbstä- tigkeit, akade- mische und berufliche Wei- terbildung | Paper & Pencil | Paper & Pencil | Paper & Pencil | Online ^b | Online ^b | Online ^b |
| 3 | ca. 10 Jahre nach Studienab- schluss | Aktuelle Tätig- keit, Erwerbstä- tigkeit, akade- mische und berufliche Wei- terbildung, familiäre Situa- tion | | Paper & Pencil ^a | Online ^b | Online ^b | Online ^b (geplant) | Online ^b (geplant) |

a Die Hauptbefragung wurde als Paper & Pencil- und die Vertiefungsbefragungen 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.

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

² Für die Kohorte 1989 wurden ausschließlich Absolvent*innen aus den Bundesländern der damaligen Bundesrepublik Deutschland ausgewählt.

[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 Verlaufsdaten 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*innen bayerischer Hochschulen begrenzt³ Das Kooperationsprojekt Absolventenstudie (KOAB) des International Centre for Higher Education Research (INCHER) befragt Absolvent*innen seiner Kooperationshochschulen und ermöglicht individuelle Analysen auf Hochschul- und Studiengangsebene, die zur Evaluation und Weiterentwicklung genutzt werden können.⁴

[Spezifika des Absolventenpanels 2009] Neben den allgemeinen Charakteristika der Studienreihe weist die hier betrachtete Absolvent*innenkohorte 2009⁵, für die momentan Daten aus drei Hauptbefragungen und den zwei Vertiefungsbefragungen der zweiten Welle zum Thema "Promotion" und "regionale Mobilität" bereitgestellt werden⁶, folgende Spezifika auf. Wie bei der Absolvent*innenkohorte 2005 ist die Studienphase der Kohorte 2009 durch den Hochschulwandel im Rahmen des Bologna-Prozesses geprägt und die berufliche Einstiegsphase durch die Wirtschaftsund Finanzkrise im Jahr 2008 gekennzeichnet. Im Unterschied zu vorangegangenen Absolvent*innenkohorten weist die Absolvent*innenkohorte 2009 jedoch neben Absolvent*innen mit traditionellen Abschlüssen auch eine große Anzahl an Bachelorabsolvent*innen auf, die im Prüfungsjahr 2009 22,9 % der Gesamtabsolvent*innenzahl umfassten (vgl. Dudek, Glässner & Schröder, 2010, S. 31).⁷ Dies ermöglicht zum einen Vergleiche zwischen Bachelorabsolvent*innen und Absolvent*innen traditioneller Studiengänge. Zum anderen kann betrachtet werden, ob die Bachelorabsolvent*innen nach ihrem Abschluss in den Beruf übergehen, ein Masterstudium aufnehmen oder beide Optionen kombinieren.

³ vgl. dazu http://www.bap.ihf.bayern.de

⁵ Die Grundgesamtheit sind Hochschulabsolvent*innen, 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 Teilnehmer*innen von Bundeswehrhochschulen, Verwaltungsfachhochschulen sowie von dualen, Teilzeit- und Fernstudiengängen). Im Rahmen der ersten Befragungswelle wurde ursprünglich zudem eine kleine Sonderstichprobe mit Masterabsolvent*innen berücksichtigt. Die realisierte Fallzahl war jedoch so gering, dass insbesondere fächerdifferenzierte Analysen damit nicht sinnvoll sind. Daher wurden die Masterabsolvent*innen aus dem Datensatz entfernt und auch in der folgenden Dokumentation nicht weiter berücksichtigt.

⁶ Die Daten der beiden Vertiefungsbefragungen aus der dritten Welle zu den Themen "Mobilität/Tätigkeitsinhalte" und "Vereinbarkeit Beruf und Familie" sind momentan noch nicht als Scientific Use File aufbereitet und werden daher im vorliegenden Daten- und Methodenbericht auch nicht weiter ausgeführt.

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

⁴ vgl. dazu http://koab.uni-kassel.de

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 Vergleich zu den vorangegangen Kohorten außerdem, dass die zweite Befragungswelle erstmals online durchgeführt wurde.

Die Erhebung der dritten Befragungswelle des Absolventenpanels 2009 erfolgte im Rahmen des Verbundprojekts "Bundesweites Absolventenpanel (buwap)"⁸ als eines von drei Modulen. Die beiden anderen Module umfassten die zweite Befragungswelle des Absolventenpanels 2013 und die erste Befragung des Absolventenpanels 2017.

⁸ https://www.dzhw.eu/forschung/projekt?pr_id=606

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 2). Die beiden Vertiefungsbefragungen der zweiten Befragungswelle sowie alle Teilbefragungen der dritten Befragungswelle wurden ausschließlich über einen standardisierten Online-Fragebogen realisiert. Alle eingesetzten Erhebungsinstrumente waren in deutscher Sprache.⁹

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. Daran anknüpfend wird der weitere Werdegang der Absolvent*innen in den Blick genommen.

Zu diesem Zweck wird über die Wellen hinweg der individuelle Tätigkeitsverlauf der Absolvent*innen 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, Frage 1.5 (Papierversion) bzw. Page 5 (Onlineversion) in Welle 2 und Page 103 in Welle 3), in das die Befragten ihre einzelnen Tätigkeiten eintragen sollen (vgl. Abbildung 2). Das Kalendarium wurde vom DZHW konzipiert, um lückenhafte Angaben in der Beschreibung von Tätigkeitsverläufen zu reduzieren.

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, Frage 4.5 (Papierversion) bzw. Page 19 (Onlineversion) in Welle 2 und Page 304-304o), 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. Daneben beinhalten die Erhebungsinstrumente auch Fragen zur subjektiven Einschätzung der eige-

¹¹ Vor der Einführung des Kalendariums wurden die Tätigkeitsverläufe anhand eines Tableaus erfasst.

_

⁹ Die Fragebögen sowie Filterführungsdiagramme können im Metadatensuchsystem des FDZ-DZHW (https://metadata.fdz.dzhw.eu) heruntergeladen werden.

 $^{^{10}}$ vgl. Abschnitt 1 "Studienverlauf und Studienerfahrungen" im Fragebogen der ersten Welle.

¹² Vor der Einführung des Erwerbstätigkeitstableaus wurden – abgesehen von der Art der Institution – generell nur zusätzliche Informationen zur ersten und letzten Erwerbstätigkeit erhoben.

nen beruflichen Situation, etwa hinsichtlich der beruflichen Zufriedenheit, den beruflichen Zukunftsperspektiven und der Beschäftigungsadäquanz.¹³ Ein Großteil der Informationen zu den Erwerbstätigkeiten wird über die Befragungswellen hinweg wiederholt erhoben.¹⁴ 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.

Abbildung 2: Kalendarium des DZHW-Absolventenpanels 2009, 2. Welle, Papierversion¹⁵

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 aufführen. 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 D ΑL w Α 2011 D A 2012

Erwerbstätigkeit, nichtselbstän-Berufsausbildung, Fort-, Weiterbildung (Vollzeit, dig (als Angestellte(r), Beam-Umschulung, Volontariat längerfristig) **AL** Arbeitslosigkeit Praktikum Selbständige Erwerbstätigkeit Referendariat, Anerkennungs-R **EZ** Elternzeit (ohne Werk-/Honorararbeit) praktikum u. Ä. Hausfrau/Hausmann. Trainee ST Studium **Familienarbeit** Jobben D SO Promotion Sonstiges (z. B. Wehr-/Zivil-Werkvertrag, Honorararbeit dienst, längerer Urlaub, Krank-JΡ Juniorprofessur/Habilitation

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

¹³ vgl. dazu auch Kerst & Fehse, 2007.

¹⁴ Teilweise wurde allerdings der Wortlaut abgewandelt. Darauf hinzuweisen ist außerdem, dass sich die Wiederholungsmessung in der zweiten bzw. dritten Welle bei einer beruflichen Veränderung auf eine andere berufliche Situation bzw. Tätigkeit bezieht als in der Vorwelle.

 $^{^{15}}$ Die Onlineversion des Kalendariums der zweiten Welle findet sich in den Screenshots auf Page 5.

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 Hauptbefragung¹⁶ 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. 17 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 zuvor in keiner Kohorte vertiefend beleuchtet. Die Erhebungsinhalte der Hauptbefragung der dritten Welle des Absolventenpanels 2009 orientieren sich zum überwiegenden Teil an der zweiten Welle des Absolventenpanels 2009 und der dritten Welle des Absolventenpanels 2005. Damit wird sowohl der Vergleich innerhalb als auch zwischen Kohorten sichergestellt. Die Instrumente umfassten neben der Abfrage der monatsgenauen Lebensverläufe seit der letzten Befragung die aktuelle berufliche Situation, Angaben zur aktuellen Erwerbstätigkeit, Weiterbildungsbedarfe und die private Lebenslage. Im Gegensatz zur dritten Welle 2005 wurden zusätzlich sämtliche akademischen Weiterqualifizierungen seit dem Abschluss im Prüfungsjahr 2008/2009 abgefragt und bei Befragten mit einer Promotionsphase zusätzliche Informationen zu dieser erhoben. Dies erfolgte, da für die dritte Welle keine eigenständige Vertiefungsbefragung zum Thema "Promotion" vorgesehen war.

2.2 Pretests

[Ziel und Verfahren] Die Erhebungsinstrumente der ersten Befragungswelle und der Hauptbefragung der zweiten und dritten Befragungswelle wurden im Vorfeld der Erhebungen jeweils durch

¹⁶ Für die Hauptbefragung gab es zusätzlich auch noch eine Papierversion des Fragebogens (vgl. Kapitel 4). Diese entsprach grundsätzlich dem Fragebogen in der Onlineversion. Jedoch ist auf die folgenden Unterschiede hinzu-weisen: Frage 4.5 Papierversion vs. Page 19-19h Onlineversion; Frage 4.12 Papierversion vs. Page 27 und 27_1 Onlineversion; Frage 5.5 Papierversion vs. Page 51 und 51_1 Onlineversion; Frage 6.5 Papierversion vs. Page 57-77 Onlineversion; Frage 7.1 Papierversion vs. Page 79/80 Onlineversion; Frage 7.2 Papierversion vs. Page 81/82 Onlineversion

¹⁷ 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. doi:10.6102/zis76

Pretests geprüft.¹⁸ Dabei sollte erstens für die bereits in Erhebungsinstrumenten vorheriger Kohorten eingesetzten Fragen und Antwortvorgaben geprüft werden, ob sie von der Absolvent*innenkohorte 2009 gleich perzipiert werden würden wie von vorangegangenen Kohorten. Zweitens sollte überprüft werden, ob die in den vorherigen Kohorten eingesetzten Erhebungsinstrumente auch für die Befragung der neuen Absolvent*innengruppe der Bachelor (vgl. Kapitel 1 und 3) in gleicher Weise 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).¹⁹ Die Erhebungsinstrumente der Hauptbefragung der zweiten und dritten Befragungswelle wurden 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*innen 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. ²⁰ In der zweiten und dritten Befragungswelle wurde das Erhebungsinstrument von etwa fünf Mitarbeiter*innen²¹ des DZHW als Hochschulforschungsexpert*innen geprüft.

[Durchführung] Die 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 und dritten Befragungswelle wurden iterativ in zwei Schritten eingeholt. In einem ersten Schritt wurden die einzelnen Fragen inhaltlich begutachtet. Nachdem die darauf basierenden Änderungsvorschläge eingearbeitet worden waren, wurden die Onlineversion des Fragebogens programmiert sowie – nur in der zweiten Welle – 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.

_

¹⁸ Auch für die Erhebungsinstrumente der beiden Vertiefungsbefragungen gab es Pretests, jedoch nur in geringerem Ausmaß (ca. 5 Testpersonen).

¹⁹ 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.

²⁰ Die genaue Anzahl der Teilnehmer*innen ist nicht mehr rekonstruierbar.

 $^{^{21}}$ Die genaue Anzahl der Teilnehmer st innen ist nicht mehr rekonstruierbar.

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.²² 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. Für die dritte Befragungswelle wurden zusätzlich Erkenntnisse aus den parallel stattgefundenen Pretests und Instrumentenentwicklungen des National Academics Panel Study Projekts (Nacaps) am DZHW eingebunden, die sich vor allem auf die Formulierung von Fragen zur Promotion bezogen.

 $^{^{\}rm 22}$ Die spezifischen Gründe für Änderungen können nicht mehr rekonstruiert werden.

3 Grundgesamtheit und Stichprobenumfang

[Grundgesamtheit] Die Grundgesamtheit des Absolventenpanels 2009 umfasst alle Hochschulabsolvent*innen, die im Wintersemester 2008/2009 oder im Sommersemester 2009²³ ihren ersten berufsqualifizierenden Studienabschluss an einer staatlich anerkannten Hochschule in der Bundesrepublik Deutschland erworben haben. Hochschulabsolvent*innen von Bundeswehrhochschulen, Verwaltungsfachhochschulen sowie von dualen, Teilzeit- und Fernstudiengängen waren dabei von vorherein ausgenommen.

[Stichprobenverfahren] Aufgrund fehlender oder nicht zugänglicher Listen von Hochschulabsolvent*innen 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*innen verzeichnen würden.²⁴ Daraus wurden zwei getrennte geschichtete Klumpenstichproben für zum einen Hochschulabsolvent*innen traditioneller Studiengänge²⁵ und zum anderen Hochschulabsolvent*innen aus Bachelorstudiengängen gezogen.

[Geschichtete Klumpenstichprobe] Die "primary sampling units" (Klumpen) der Klumpenstichproben wurden anhand der Hochschule und dem Studienbereich²⁶ definiert. Die "secondary sampling units" stellten die Hochschulabsolvent*innen des Prüfungsjahrgangs 2009 innerhalb dieser Klumpen dar.²⁷ 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 Hochschulabsolvent*innen 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).

²³ 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.

²⁴ Bei Erwartung annähernd konstanter Absolvent*innenzahlen 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.

²⁵ Hierunter fallen Absolvent*innen aus Studiengängen mit den Abschlüssen Diplom, Magister, Staatsexamen (u.a. Lehramtsstudiengänge) sowie Absolvent*innen mit kirchlichen und künstlerischen Abschlüssen.

²⁶ entsprechend der Aufgliederung nach der amtlichen Statistik (gemäß Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2008/2009 und SoSe 2009).

²⁷ Beispiel für den Personenkreis eines Klumpens: Alle Bachelorabsolvent*innen des Studienbereichs Physik an der Universität A oder alle Diplomabsolvent*innen der Wirtschaftswissenschaften der Fachhochschule B

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*innen 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*innen. Dies sollte bei Analysen berücksichtigt werden.²⁸

-

 $^{^{28}}$ 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).²⁹ Insgesamt wurden mehr als 400 Prüfungsämter kontaktiert. Da die Hochschulen die Kontaktdaten ihrer Absolvent*innen aus Datenschutzgründen nicht herausgeben durften, teilten sie dem DZHW lediglich die jeweilige Absolvent*innenanzahl 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 und dritten 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. 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.

[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

²⁹ 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.

³⁰ Zur Gewährleistung des Datenschutzes wurde der Adressabschnitt vom Fragebogen abgetrennt und die Referenzliste getrennt von den Befragungsdaten auf einem geschützten Server gespeichert.

³¹ Die Befragten wurden dabei sowohl im Anschluss an die jeweils abgeschlossene Welle – im Rahmen eines Informationsschreibens über die Ergebnisse der Befragung – als auch im Vorfeld zur folgenden 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. Innerhalb der Feldphase der dritten Welle wurde erneut auf den Adressaktualisierungsservice der Deutschen Post sowie die Melderegisterauskunft von RISER ID Services GmbH zurückgegriffen, sofern die postalischen Einladungen unzustellbar waren.

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 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.

Für die Hauptbefragung der dritten Befragungswelle, welche ebenfalls ausschließlich online durchgeführt wurde, wurden sowohl die postalischen als auch die E-Mail-Adressen der Befragten zur Kontaktaufnahme genutzt. Zunächst wurde den Absolvent*innen, von denen eine Postadresse vorlag, eine postalische Einladung mit einem Link zur Befragung und einem individuellen Passwort (Token) geschickt. Eine Woche später erfolgte eine weitere Einladung per E-Mail (inkl. direktem Link und Passwort zur Befragung) an all jene Absolvent*innen, von denen eine valide E-Mail-Adresse vorlag. Darüber hinaus wurden eine postalische Erinnerung und drei Erinnerungen per E-Mail verschickt. Die Anzahl der Kontaktierungen pro Person hing also von den jeweils vorliegenden Kontaktinformationen ab und variierte entsprechend zwischen zwei (nur postalisch) bis zu sechs (postalisch & E-Mail) Kontaktierungen.

[Feldphase] Der Erhebungszeitraum der ersten Befragungswelle erstreckte sich vom 1. Februar 2010 bis 31. Januar 2011.³² 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.³³ 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 2015.³⁴ Die Feldphasen der Vertiefungsbefragungen, zu der nur Teilnehmer*innen der Hauptbefragung der zweiten Befragungswelle eingeladen wurden³⁵, fanden vom 4. Juni 2015 bis 15. Juli 2015 (Vertiefungsbefragung Promotion) bzw. 16. Juli 2015 bis 31. August 2015 (Vertiefungsbefragung Mobilität) statt. Durch den nun im DZHW vorliegenden Adressbestand konnten konkrete Versandzeitpunkte für die Erhebungsunterlagen festgelegt werden.^{36,37} Zudem

³² Die Feldzeit wurde so lange wie möglich – parallel zu den weiteren Projektschritten – ausgedehnt und jeder noch eintreffende Fragebogen bis Januar 2011 berücksichtigt.

³³ Nach dem Versand der Erstanschreiben 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.

³⁴ Auch hier wurde die Feldzeit so lange wie möglich offen gehalten, sodass Nachzügler bis Oktober 2015 berücksichtigt werden konnten.

³⁵ Zur Vertiefungsbefragung Promotion wurden zudem nur die Teilnehmer*innen eingeladen, die in der Hauptbefragung angegeben hatten, nach ihrem Studienabschluss eine Promotion aufgenommen zu haben.

³⁶ Hauptbefragung: 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

konnten die Erinnerungsschreiben gezielt nur an diejenigen Zielpersonen verschickt werden, die sich noch nicht an der Erhebung beteiligt hatten.

Die Daten zur Hauptbefragung der dritten Befragungswelle wurden vom 12. April 2019 bis 03. Juli 2019 erhoben. Auch hier konnten konkrete Versandzeitpunkte³⁸ bestimmt sowie gezielte Erinnerungen von Nicht-Teilnehmer*innen 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 rücklaufsteigernde Maßnahmen vor allem in den 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 (ohne Flyer). Des Weiteren wurden auf einer Projekthomepage Informationen zum Projekt und daraus entstandenen Publikationen bereitgestellt. Darüber hinaus wurde unter allen Befragungsteilnehmer*innen 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 und in der dritten Welle ein aktuelles Tablet, ein hochwertiges Smartphone und Reisegutscheine in einem Gesamtwert von 2.000 Euro

³⁷ Vertiefungsbefragung Promotion: Einladung: 04.06.2015; Erinnerung: 18.06.2015; Vertiefungsbefragung Mobilität: Einladung: 16.07.2015, Erinnerung: 30.07.2015

³⁸ Hauptbefragung: Einladung (postalisch): 11.04.2019; Einladung (E-Mail): 18.04.2019; Erste Erinnerung: 02.05.2019 (E-Mail), 16.05.2019 (postalisch); Zweite Erinnerung: 23.05.2019 (E-Mail); Dritte Erinnerung: 06.06.2019 (E-Mail)

5 Rücklauf

[Rücklauf] Die Bruttostichprobe der ersten Befragungswelle umfasste 52.550 Absolvent*innen, die bei den Prüfungsämtern der im Stichprobenverfahren gezogenen Hochschulen (vgl. Kapitel 3) gemeldet waren und kontaktiert wurden, darunter 27.702 Bachelorabsolvent*innen und 24.848 Absolvent*innen 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 10.492 Fälle³⁹, davon 4.882 Bachelorabsolvent*innen und 5.610 Absolvent*innen traditioneller Studiengänge. Im Hinblick auf die Bruttostichprobe liegt die Rücklaufquote damit insgesamt bei 20 Prozent. Für die Absolvent*innen traditioneller Studiengänge (22,6 %) ist sie etwas höher als für die Bachelorabsolvent*innen (17,6 %).

9.086 Personen, also über 86 Prozent der 10.492 Teilnehmer*innen der ersten Welle, wurden im Rahmen einer Kontaktaktion, die im Vorfeld der zweiten Befragungswelle zur Adressaktualisierung durchgeführt wurde, angeschrieben und über die geplante Befragung informiert. Diese ergeben die Bruttostichprobe der zweiten Welle. Von diesen 9.086 Fällen konnten letztendlich 9.004 Fälle zur Befragung eingeladen werden. Aufgrund von weiteren stichprobenrelevanten Ausfällen (Nichtteilnahmen und nicht auswertbare Fragebögen) umfasst die Nettostichprobe der zweiten Welle 4.755 Fälle, darunter 2.110 Bachelorabsolvent*innen und 2.645 Absolvent*innen traditioneller Studiengänge. Bezogen auf die 9.086 Fälle der Bruttostichprobe liegt die Gesamt-Rücklaufquote damit bei 52,3 Prozent. Da zur zweiten Welle lediglich diejenigen Personen eingeladen wurden, die im Rahmen der ersten Welle einer erneuten Kontaktierung zugestimmt hatten, ist die Rücklaufquote der zweiten Welle deutlich höher als die der ersten Welle (vgl. Tabelle 2).

Zur Vertiefungsbefragung zum Thema "Mobilität" wurden alle Teilnehmer*innen der zweiten Hauptbefragung erneut kontaktiert. Die Bruttostichprobe umfasste entsprechend 4.755 Fälle. Es konnte eine Nettostichprobe von 2.465 Fällen realisiert werden, dies entspricht einer Rücklaufquote von 51,8 Prozent.

In die Vertiefungsbefragung zum Thema "Promotion" wurden nur diejenigen 1.136 Personen einbezogen, die in der zweiten Hauptbefragung angegeben hatten, dass sie nach ihrem Studienabschluss eine Promotion aufgenommen haben. ⁴¹ Hier ergab sich eine Nettostichprobe von 676 Fällen und damit eine Rücklaufquote von 59,5 Prozent.

³⁹ Im Rahmen der Erstellung der SUF-Version 2.0.0 wurden nachträglich zwei Fälle wegen einer Dopplung in der ID gelöscht.

 $^{^{40}}$ Die Ausfälle ergeben sich zum einen aus Teilnahmeverweigerungen für Folgewellen, zum anderen aus ungültigen Adressen.

⁴¹ if inlist(bfec12,1,2,3,4); Für einen Fall, der an der Promotionsbefragung teilgenommen hat, gilt jedoch bfec12=5. Es ist nicht mehr rekonstruierbar, wie es zu dieser Ausnahme gekommen ist. Der Fall wurde jedoch in den Daten belassen.

Grundlage der Hauptbefragung der dritten Befragungswelle waren die Teilnehmer*innen der zweiten Befragungswelle, die im Rahmen vorheriger Befragungen ihre persönlichen Kontaktadressen zur Verfügung gestellt und der Wiederkontaktierung im Rahmen des Panels zugestimmt hatten. Insgesamt wurden 4733 Personen kontaktiert und zur Befragung eingeladen. 3.642 auswertbare Fragebögen gingen bis zum Ende der Feldphase ein, wodurch sich ein Rücklauf von 76,9% ergibt.

[Panelausfälle] Die Betrachtung der Bruttostichprobe im Zeitverlauf zeigt, dass diese in der Hauptbefragung der zweiten Welle nur noch etwa 17 Prozent der Bruttostichprobe der ersten Welle umfasst; in der Hauptbefragung der dritten Befragungswelle sind es nur noch etwa 9 Prozent. Von den 10.492 Fällen aus der Nettostichprobe der ersten Welle konnten etwa 45 Prozent der Befragten in der zweiten Befragungswelle und 34,7 Prozent in der dritten Befragungswelle erneut befragt werden (vgl. Tabelle 2). Darüber hinaus erweist sich im Vergleich der Nettostichprobe der zweiten Welle mit der Bruttostichprobe der ersten Welle, dass nur 9 Prozent der ersten Bruttostichprobe an beiden Befragungswellen teilgenommen haben. Bei den Vertiefungsbefragungen fallen die Rücklaufquoten entsprechend geringer aus. Zusätzlich auch an der Hauptbefragung der dritten Befragungswelle haben sich nur knapp 7 Prozent der ersten Bruttostichprobe beteiligt.

Das Absolventenpanel 2009 ist damit von paneltypischen Ausfallprozessen⁴² betroffen. Hier sind die grundsätzliche Verweigerung der Teilnahme an Folgebefragungen (keine Angabe von Adressangaben für die Kontaktierung in der zweiten Welle) und die Nichtteilnahme nach einer (versuchten) Kontaktaufnahme in der zweiten Befragungswelle zu nennen. Weiterhin treten Ausfälle aufgrund einer erschwerten Kontaktaufnahme (Adressänderungen) auf, da besonders in der ersten Zeit nach dem Studienabschluss eine große Mobilität von Absolvent*innen zu verzeichnen ist (vgl. Fabian & Briedis, 2009, 71ff.).

Tabelle 2: Brutto-, Nettostichproben und Rücklaufquoten des DZHW-Absolventenpanels 2009 (Gesamt)

| | Welle 1 | Welle 2 (Hauptbefragung) | Welle 2 (Promotion) | Welle 2 (Mobilität) | Welle 3 (Hauptbefragung) |
|--|---------|-----------------------------|------------------------|------------------------|-----------------------------|
| Bruttostichprobe | 52.550 | 9.086 | 1.136 | 4.755 | 4733 |
| Nettostichprobe | 10.492 | 4.755 | 676 | 2.465 | 3.642 |
| Rücklaufquote | 20,0 % | 52,3 % | 59,5 % | 51,8 % | 76,9% |
| Anteil Brutto- stichprobe an Bruttostichprobe Welle 1 | | 17,3 % | 2,2 % | 9,0 % | 9,0% |
| Anteil Netto- stichprobe an Nettostichprobe Welle 1 | | 45,3 % | 6,4 % | 23,5 % | 34,7 % |
| Anteil Netto- stichprobe an Bruttostichprobe Welle 1 | | 9,0 % | 1,3 % | 4,7 % | 6,9 % |

_

⁴² Zu paneltypischen Ausfallprozessen vgl. Schnell, Hill & Esser, 2005, S. 241.

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). 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 zweifach⁴⁴ von jeweils unterschiedlichen DZHW-Mitarbeiter*innen 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 wur-

_

⁴³ In der ersten Befragungswelle wurden zudem bereits manuelle Vorkorrekturen zur Erleichterung der Datenübertragung vorgenommen (vgl. Kapitel 6.3).

⁴⁴ mit Ausnahme der Kalenderdaten, die nur einfach erfasst wurden

den keine neuen Variablen mit numerischen Codierungen erstellt, sondern die Nennungen nur sofern möglich – den vorhandenen (geschlossenen) Antwortkategorien zugeordnet. Einzelne offene Fragen wurden nicht vercodet, weil sie nur als Kontextinformationen zur Codierung anderer offener Angaben erfasst wurden oder aufgrund von unzureichenden zeitlichen Ressourcen.

In Tabelle 3 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 - sofern datenschutzrechtlich unbedenklich – im Datensatzreport sowie im Metadatensuchsystem⁴⁵ dokumentiert.

Tabelle 3: Vercodete Merkmale und Codierlisten im DZHW-Absolventenpanel 2009

| Merkmal | Codierliste | Codierlisten-ID ^b |
|---|--|---|
| Studienfach | Welle 1 und 2: Destatis- Fächersystematik 2008/2009 [gemäß Schlüsselverzeichnis für die Studenten- und Prüfungssta- tistik (WiSe 2008/2009 und SoSe 2009), Schlüssel 3.1] | cl-destatis-studienfach-2009 ^c |
| | Welle 3: Destatis- Fächersystematik 2019/2020 [gemäß Schlüsselverzeichnis für die Studenten- und Prüfungssta- tistik (WiSe 2019/2020 und SoSe 2020), Schlüssel 4.1] | cl-destatis-studienfach-2020 ^c |
| Studienabschluss | projekteigene Codierung | cl-dzhw-22 |
| Hochschule | Welle 1 und 2: Destatis- Schlüsselverzeichnis für die Stu- denten- und Prüfungsstatistik (WiSe 2014/2015 und SoSe 2015), Schlüssel 2.2 | cl-destatis-hochschule-2015 ^c |
| | Welle 3: Destatis- Schlüsselverzeichnis für die Stu- denten- und Prüfungsstatistik, Promovierendenstatistik und Gasthörerstatistik (WiSe 2019/2020 und SoSe 2020), Schlüssel 2.2 | cl-destatis-hochschule-2020 ^c |
| Bundesland | Destatis-Bundeslandschlüssel (entsprechend der ersten beiden Ziffern des Amtlichen Gemeinde- schlüssels (AGS)) | cl-destatis-bundesland-1990 ^c |
| Ausland/Staatsangehörigkeit | Welle 1: projekteigene Codierung | cl-dzhw-23 |
| | Welle 2 und 3: Destatis- Staatsangehörigkeits- und Ge- bietsschlüssel 2015 | cl-destatis-ausland-2015 |
| Berufsbezeichnung | Beruf der Befragten: Destatis- Klassifikation der Berufe 2010 | cl-destatis-kldb-2010 ^c |
| | Ausbildungs-/Elternberuf: Desta- tis-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 Codie- | |

⁴⁵ https://metadata.fdz.dzhw.eu

rung oder Löschung

- a vgl. Frage 5.2 in Welle 1 (in Welle 2 und 3 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 einige Zusatzcodes

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.

6.3 Datenprüfung und Datenbereinigung

[Konsistenzprüfungen] Für alle Befragungswellen – insbesondere für die Hauptbefragungen – wurden 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).
- 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. 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.⁴⁷ Darauf aufbauend erfolgte im Anschluss an die Datenübertragung eine umfassende Prüfung und Korrektur der Daten mit Hilfe einer DZHW-

⁴⁶ 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 überschnitt sich in einigen Punkten (wie z. B. den Angaben zu Promotionsdaten und zum akademischen Werdegang) durchaus mit der Hauptbefragung.

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

eigenen Software.⁴⁸ In der zweiten und dritten 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. ⁴⁹ Diese wurden ebenfalls aus dem Datensatz entfernt. ⁵⁰

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. Der Variablenname einer generierten Variablen ist im Datensatz durch das Suffix "_g#" gekennzeichnet.

Die Dokumentation der einzelnen generierten Variablen mit Angabe ihrer jeweiligen Ausprägungen und Generierungsvorschriften findet sich im Datensatzreport sowie im Metadatensuchsystem⁵¹. Darüber hinaus existiert eine Sonderdokumentation für bestimmte Hilfsvariablen, die vom Primärforschungsprojekt bzgl. der erfassten Informationen zu Erwerbstätigkeiten generiert wurden, um Auswertungen zu vereinfachen und standardisieren zu können. 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.

6.5 Erstellung der Datensätze

[Zusammenführung der Wellen] Die Daten der drei Befragungswellen wurden zusammengeführt.⁵² Die Zuordnung der Fälle erfolgte über die im Rahmen der Feldphase vergebenen Identifikationsnummern der Befragten (vgl. Kapitel 4).

-

⁴⁸ Zu diesem Zweck wurden die erfassten Fragebogen-Daten in eine Datenbank eingelesen. Anschließend wurden anhand formaler Regeln gültige Wertebereiche und Antwortkombinationen definiert und geprüft.

⁴⁹ Dies kam beispielsweise vor, wenn die Prüfungsämter versehentlich Personen angeschrieben hatten, die einem anderen Prüfungsjahrgang angehörten oder wenn es sich um Absolvent*innen handelte, die bereits einen weiteren Abschluss hatten.

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*innen 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.

⁵¹ https://metadata.fdz.dzhw.eu

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 neuen Welle gestrichen und entsprechend die Angaben aus der Vorwelle beibehalten.

[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, Page 103 der dritten 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 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.

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 SPSSFormat 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 wave⁵³ 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.⁵⁴ Im Stamm geht der Themenbereich, dem die Variable zugeordnet ist, aus einem dreistelligen englischen Buchstabenkürzel hervor. Tabelle 4 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 gene-

⁵⁴ Welle 1: a, Welle 2 - Hauptbefragung: b, Welle 2 - Vertiefungsbefragung Mobilität: m, Welle 2- Vertiefungsbefragung Promotion: p, Welle 3 - Hauptbefragung: c

⁵³ Diese enthält die Information, welche Fälle an welchen Wellen teilgenommen haben.

rierten 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.

Tabelle 4: Themengebiete und Kürzel für Variablennamen des DZHW-Absolventenpanels 2009

| Themengebiets-Kürzel | Themengebiet (englisch) | Themengebiet (deutsch) |
|----------------------|-----------------------------|---|
| stu | studies | Studium |
| осс | occupation | Beschäftigung |
| ski | skills | Fähigkeiten |
| fvt | further vocational training | Berufliche Fort- und Weiterbil- dung |
| 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 Episodenzeit-raums, der über vier Variablen (Monat: begin_m und end_m; Jahr: begin_y; end_y) codiert wird. Für die Episoden, die dem Kalendarium der dritten Welle entstammen, wurde vom Erhebungsprojekt zusätzlich eine Quellvariable (source) ergänzt. Diese gibt an, ob die Episode von den Befragten im Kalendarium angegeben oder im Zuge der Datenaufbereitung durch das Erhebungsprojekt ergänzt wurde. Genauere Informationen finden sich in der Sonderdokumentation.

6.7 Codierung fehlender Werte

Die Codierung fehlender Werte variiert zwischen der ersten und zweiten Welle im Vergleich zur dritten Welle.

In der ersten und zweiten Welle wurde die Missingsystematik des FDZ-DZHW verwendet, in der fehlende Angaben durch dreistellige negative Werte codiert werden. Tabelle 5 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 5: Systematik des FDZ-DZHW für fehlende Werte (Welle 1 und 2)

| Wertebereich | Code | Wertelabel |
|---|--|---|
| -999 bis -990: Nonresponse | -999 - 998 -997 -996 -995 -994 | weiß nicht keine Angabe keine Angabe (Antwortkategorie) Interviewabbruch keine Teilnahme (Panel) verweigert |
| -989 bis -970: Nicht zutreffend | -989 -988 -987 -986 -985 | filterbedingt fehlend trifft nicht zu designbedingt fehlend (Fragebogensplit) designbedingt fehlend (Welle) ^a designbedingt fehlend (Kohorte) ^b |
| -969 bis -950: Editierter fehlender Wert | -969 - 968 - 967 - 966 -965 | unbekannter fehlender Wert ^c unplausibler Wert ^d anonymisiert nicht bestimmbar ^e 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

In der dritten Befragungswelle wurden die vom Primärforschungsprojekt vergebenen, einstelligen negativen Missingcodes übernommen (vgl. Tabelle 6).

Tabelle 6: Systematik des Primärforschungsprojekts für fehlende Werte (Welle 3)

| Code | Wertelabel |
|------|---|
| -9 | keine Teilnahme (Panel) |
| -8 | unbekannt aufgrund nachgetragener Episode |
| -7 | n. teilgenommen an Vertiefungsbefragung |
| -6 | keine Zuordnung möglich |
| -4 | keine Episoden vorhanden |
| -3 | noch laufend |
| -2 | Filter |
| -1 | keine Angabe |

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.⁵⁵ Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (Missing at Random, Not Missing at Random), der einer Modellierung bedarf.⁵⁶

[Konzeptuelles Vorgehen] Im Zuge einer Gewichtungsprozedur sollten idealerweise zunächst designbedingte Disproportionalitäten ausgeglichen werden. Die hierfür benötigten *Designgewichte* ergeben sich bei zufallsgesteuerten Auswahlverfahren direkt aus dem 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 Nicht-Teilnehmer*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 für die erste Befragungswelle⁵⁷ gebildet: ein Gesamtgewicht für die Auswertung aller Absolvent*innen und zwei gesonderte Gewichte für die Auswertung von Absolvent*innen traditioneller Studiengänge bzw. von Bachelorabsolvent*innen.

Das Designgewicht ergibt sich dabei für jede Schicht s wie folgt⁵⁸:

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

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

⁵⁶ Zu den unterschiedlichen Formen von Ausfallprozessen siehe grundlegend Rubin, 1976.

⁵⁷ Für die zweite und dritte Befragungswelle wurden keine Querschnittsgewichte erstellt, da weder neu hinzugekommene Individuen befragt wurden (Auffrischungsstichprobe) noch Personen die Möglichkeit hatten, an Welle 2 und 3 teilzunehmen, wenn sie nicht Teil der ersten Befragungswelle waren.

⁵⁸ 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 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 Bachelorabsolvent*innen sowie aller Absolvent*innen traditioneller Studiengänge. Als Merkmale wurden die Region (Ost/West), das Geschlecht, die Studienbereiche sowie für das Gesamtgewicht zusätzlich auch die Abschlussart herangezogen. 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-Algorithmus durchgeführt.

[Längsschnittgewichtung] Für die Hauptbefragung der zweiten und dritten Welle wurde zusätzlich ein Nonresponsegewicht berechnet, das mit Hilfe von Informationen aus der ersten Welle die Nichtteilnahme in der zweiten Welle modelliert. Diese Informationen dienten als Kovariaten (σ_t) in einem Probit-Regressionsmodell, das die Vorhersage der Teilnahmewahrscheinlichkeit zum Zeitpunkt $P(Res_{t+1})$ zum Ziel hatte. Bei Variablen mit fehlenden Werten wurden diese jeweils als zusätzliche Kategorien der Variablen in das Modell aufgenommen, sodass auch Fälle mit Item-Nonresponse in das Modell inkludiert werden konnten. Hierdurch konnte weiterhin die Annahme geprüft werden, dass Item-Nonresponse einen bedeutsamen Prädiktor für Unit-Nonresponse in künftigen Wellen darstellt. Eine Reihe von Prädiktoren aus der ersten Welle erwies sich als bedeutsam für die Vorhersage der Teilnahmewahrscheinlichkeit in der zweiten Welle. Aus dem Modell konnte die bedingte Teilnahmewahrscheinlichkeit abgeleitet werden, deren Kehrwert das Ausfallgewicht für die zweite Welle darstellt:

$$NR_{i_{gew_{t2}}} = P(Res_{t2_i} | \sigma_{t1_i})^{-1}$$

Das Längsschnittgewicht für das Zwei- bzw. Drei-Wellen-Panel des Datensatzes (Hauptbefragung) ergibt sich aus dem Produkt des Designgewichts ($dwgt_{sci}$) mit dem Nonresponsegewicht ($NR_{i_{gew_{r}}}$):

$$wgt_{i_{t1t2}} = dwgt_{sci} \times NR_{i_{gew_{t2}}}$$

Das Längsschnittgewicht für das Drei-Wellen-Panel des Datensatzes (Hauptbefragung) ergibt sich aus dem Produkt des Designgewichts $(dwgt_{sci})$ mit dem Nonresponsegewicht der zweiten Welle $(NR_{i_{gewt_2}})$ und dem Nonresponsegewicht der dritten Welle $(NR_{i_{gewt_3}})$:

$$wgt_{i_{t1t2t3}} = dwgt_{sci} \times NR_{i_{gew_{t2}}} \times NR_{i_{gew_{t3}}}$$

Im Anschluss wurde das nonresponse-adjustierte Längsschnittgewicht mittels des Raking-Algorithmus auf Merkmale der Grundgesamtheit kalibriert.⁶²

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

⁵⁹ Die Informationen aus der Grundgesamtheit wurden aus Daten des Statistischen Bundesamtes (Prüfungsstatistik 2008/2009) abgeleitet.

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

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

⁶² Hierfür wurden dieselben Merkmale verwendet, welche auch für die Kalibrierung des Designgewichtes in der ersten Befragungswelle eingesetzt wurden.

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

[Hinweise zur Anwendung der Gewichte] In Tabelle 7 sind die erstellten Gewichte dargestellt.

Tabelle 7: Bereitgestellte Gewichte zum DZHW-Absolventenpanel 2009

| Variablennamen | Beschreibung |
|----------------|---|
| wgt01_t1d | Querschnittsgewicht Welle 1: Absolvent*innen insgesamt |
| wgt02_t1d | Querschnittsgewicht Welle 1: Absolvent*innen traditioneller Studiengänge |
| wgt03_t1d | Querschnittsgewicht Welle 1: Bachelorabsolvent*innen |
| wgt04_t1t2d | Längsschnittgewicht 2-Wellen-Panel (Hauptbefragung): Absolvent*innen insgesamt |
| wgt05_t1t2d | Längsschnittgewicht 2-Wellen-Panel (Hauptbefragung): Absolvent*innen traditioneller Studiengänge |
| wgt06_t1t2d | Längsschnittgewicht 2-Wellen-Panel (Hauptbefragung): Bachelorabsolvent*innen |
| wgt07_t1t2t3d | Längsschnittgewicht 3-Wellen-Panel (Hauptbefragung): Absolvent*innen insgesamt |
| wgt08_t1t2t3d | Längsschnittgewicht 3-Wellen-Panel (Hauptbefragung): Absolvent*innen traditioneller Studiengänge |
| wgt09_t1t2t3d | Längsschnittgewicht 3-Wellen-Panel (Hauptbefragung): Bachelorabsolvent*innen |

Bei den erstellten Gewichten handelt es sich um probability weights, die in Stata mit Hilfe .ado-spezifischer Optionen berücksichtigt werden können.⁶³ Es gibt für jede Haupterhebung jeweils ein Gesamtgewicht zur Auswertung aller Absolvent*innen sowie zwei Gewichte für die getrennte Auswertung von Absolvent*innen traditioneller Studiengänge und Bachelorabsolvent*innen. Werden nur Fälle eines Subsamples betrachtet, wird die Verwendung des entsprechenden Gewichts empfohlen, in allen anderen Fällen die Verwendung des Gesamtgewichts. Für die Vertiefungsbefragungen wurden keine Gewichte erstellt.

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. 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.

⁶³ Siehe hierzu die Stata-Hilfe (Befehl: *help weights*).

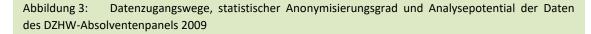
8 Anonymisierung

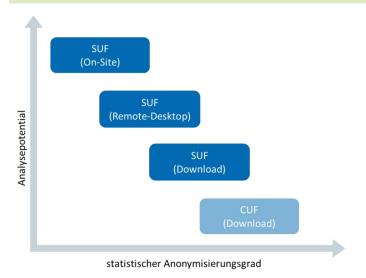
[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten⁶⁴, die in freiwilligen Befragungen durch das DZHW erhoben werden, gelten die EU-Datenschutz-Grundverordnung (EU-DSGVO) und das Bundesdatenschutzgesetz. Danach sind personenbezogene Daten für die Weitergabe zur wissenschaftlichen Sekundärnutzung (ohne Vorliegen einer Einverständniserklärung zur Sekundärnutzung der personenbezogenen Daten) in der Regel derart aufzubereiten, dass "die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden können" (Art. 4 Abs. 5 DSGVO; s. auch Art. 89 DSGVO sowie Erwägungsgrund 26 DSGVO). Das heißt, für die Weitergabe von Daten aus wissenschaftlichen Forschungsprojekten an Dritte sind die Daten derart zu anonymisieren, dass kein Bezug zur Person mehr hergestellt werden kann.

[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 jeden Zugangsweg wird eine andere SUF-Variante bereitgestellt, die unterschiedlich stark anonymisiert worden ist und entsprechend weniger oder mehr Informationen umfasst. Abbildung 3 gibt einen Überblick über den jeweiligen Grad der statistischen Anonymisierung und das damit verbundene Analysepotential. Im Folgenden werden die durchgeführten statistischen Anonymisierungsmaßnahmen in Abhängigkeit von Datenprodukt (SUF/CUF) und Zugangsweg erläutert.

⁶⁴ "Personenbezogene Daten (sind) alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden "betroffene Person") beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden kann, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind" (Art. 4 DSGVO, S. 1).





[Statistische Anonymisierungsmaßnahmen] Im Rahmen der Anonymisierung wurden zunächst alle Informationen, mit denen sich Personen oder Institutionen direkt identifizieren lassen, gelöscht. 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. Eür das Absolventenpanel 2009 wurden die folgenden Quasi-Identifikatoren identifiziert, die sowohl in externen Datenquellen als auch in den Absolvent*innendaten vorliegen: Hochschule, Studienfach, Abschlussart, Promotions- und Berufsangaben, regionale Informationen (zur Hochschule, zum Ort des Erwerbs der Studienberechtigung und Arbeitsort), Staatsangehörigkeit und Geburtsland. Um eine eindeutige Zuordnung der Absolvent*innendaten zu unterbinden, wurden diese Schlüsselmerkmale – je nach Datenprodukt bzw. Zugangsweg – aggregiert oder gelöscht (vgl. Tabelle 8). 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.

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 Reldentifikation 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 Ge-

_

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.

⁶⁶ z.B. Studenten- und Prüfungsstatistik des statistischen Bundesamtes, Alumninetzwerke der Hochschulen oder auch Berufsnetzwerke.

sundheitsinformationen 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 8). Zum anderen wurde eine per Zufallsauswahl gewonnene 2/3-Substichprobe der Daten gezogen

Tabelle 8: Maßnahmen der statistischen Anonymisierung der Daten des DZHW-Absolventenpanels 2009 nach Zugangsweg

| Merkmal | On-Site-SUF | Remote-Desktop- SUF | Download-SUF | Download-CUF (Substichprobe) |
|---|---|--|--|--|
| Direkte Identifikatoren | Löschung und Vergabe einer zufäl- ligen ID | Löschung und Vergabe einer zufälligen ID | Löschung und Vergabe einer zufälligen ID | Löschung und Vergabe einer zufäl- ligen ID |
| Fragebogen- eingang | Freigabe | Freigabe von Monat/Jahr | Freigabe von Monat/Jahr | Löschung |
| Studienfach | Vercodung zu Studienfächern ^a | Aggregation zu Studienbereichen ^a | Aggregation zu Studienbereichen ^a | Aggregation zu Fächergruppen ^a |
| Hochschule | Aggregation zu Hochschulart (6 Kategorien) und Hochschulort- (NUTS-2-Regionen) ^b | Aggregation zu Hochschulart (6 Kategorien) und Hochschulort- (Bundesländern) | Aggregation zu Hochschulart (Uni/FH) und Hochschulort (neue bzw. alte Bundesländer) | Aggregation zu Hochschulart (Uni/FH) und Hoch- schulort (neue bzw. alte Bundesländer) |
| Arbeitsort (Bundesland/ Land) | Freigabe | Freigabe | Aggregation zu Bundesländern und Ausland | Aggregation zu alte, neue Bundesländer und Ausland |
| Arbeitsort (PLZ-3-Steller) | Freigabe ^h | Aggregation zu NUTS-2-Regionen ^b | Aggregation zu NUTS-2-Regionen ^b | Löschung |
| Ort der Studienbe- rechtigung (Bundesland/ Land) | Freigabe | Freigabe | Aggregation zu Bundesländern und Ausland | Aggregation zu alte, neue Bundesländer und Ausland |
| Ort der Studienbe- rechtigung (PLZ) | Freigabe | Aggregation zu NUTS-2-Regionen ^b | Aggregation zu NUTS-2-Regionen ^b | Löschung |
| Wohnsitz (PLZ) | Freigabe | Aggregation zu NUTS-2-Regionen ^b | Aggregation zu NUTS-2-Regionen ^b | Löschung |
| Wohnsitz (Bundesland/ Land) | Freigabe | Freigabe | Aggregation zu Bundesländern und Ausland | Aggregation zu alte, neue Bundesländer und Ausland |
| Beruf | Vercodung zu KldB- 5-Steller ^c | Aggregation zu KldB-3-Steller ^c | Aggregation zu KldB-3-Steller ^c | Aggregation zu KldB-2-Steller ^c |
| Leitende Position | Freigabe | Aggregation zu ja/nein | Aggregation zu ja/nein | Aggregation zu ja/nein |
| Ausbildungsberuf | Vercodung zu KldB- 3-Steller ^d | Vercodung zu KldB-3-Steller ^d | Vercodung zu KldB-3-Steller ^d | Aggregation zu KldB-2-Steller ^d |
| Beruf Eltern | Vercodung zu KldB- 3-Steller ^d | Vercodung zu KldB-3-Steller ^d | Vercodung zu KldB-3-Steller ^d | Aggregation zu KldB-2-Steller ^d |
| Zeitpunkte Promotion: - Zulassung - inhaltlicher Beginn ⁱ - letzte mündliche Abschlussprüfung ⁱ | Freigabe | Freigabe | Aggregation der Monate zu Quar- talen Top-/Bottom- Codierung der Jahre ^f | Aggregation der Monate zu Quarta- len Top-/Bottom- Codierung der Jahre ^f |
| Staatsangehörig- keit | Freigabe | Aggregation nach NEPS Klassifikati- on ^e | Aggregation zu Weltregionen | Löschung |

| Geburtsland | Freigabe | Aggregation nach NEPS Klassifikati- on ^e | Aggregation zu Weltregionen | Löschung |
|---|------------------------|---|--------------------------------|----------------------------|
| Alter | Freigabe | on Freigabe | Freigabe | Top-Codierung ^f |
| Merkmale zur Gesundheit | Löschung | Löschung | Löschung | Löschung |
| sonstige offene Angaben ^g | Vercodung/ Löschung | Vercodung/ Löschung | Löschung | Löschung |

- Nach dem Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2008/2009 und SoSe 2009 (Welle 1 und 2) und dem dem Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2019/2020 und SoSe 2020 (Welle 3) von Destatis.
- b Statistisches Amt der Europäischen Union (Eurostat): Nomenclature of territorial units for statistics (NUTS) http://ec.europa.eu/eurostat/web/nuts/overview.
- 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 Angaben, die oberhalb (Top-Codierung) bzw. unterhalb (Bottom-Codierung) 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).
- h In Welle 3 wurde der PLZ-5-Steller erhoben, er wird jedoch nicht freigegeben, sondern zum PLZ-3-Steller umgewandelt.
- i nur in Welle 3

9 Literaturverzeichnis

- Blumenstiel, J. E. & Gummer, T. (2015). Prävention, Korrektur oder beides? Drei Wege zur Reduzierung von Nonresponse Bias mit Propensity Scores. In J. Schupp & C. Wolf (Hrsg.), *Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen* (S. 13–44). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-10459-7
- Dudek, K., Glässner, B. & Krause, I. (2007). *Statistische Daten zur Einführung von Bachelor- und Masterstudiengängen. Sommersemester 2007* (Statistiken zur Hochschulpolitik 1/2007). Bonn: Hochschulrektorenkonferenz.
- Dudek, K., Glässner, B. & Schröder, M. (2010). *Statistische Daten zur Einführung von Bachelor- und Masterstudiengängen. Wintersemester 2010/2011* (Statistiken zur Hochschulpolitik 2/2010). Bonn: Hochschulrektorenkonferenz.
- 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
- 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.
- 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
- 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.
- 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