

Florence Baillet / Andreas Franken / Anne Weber

DZHW-Absolventenpanel 2005

Daten- und Methodenbericht zu den Erhebungen der
Absolvent(inn)enkohorte 2005 (1. und 2. Befragungswelle)

Version 1.0.0

Daten- und Methodenbericht

Juni 2017

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

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

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

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

Vorsitzender des Aufsichtsrats:
Ministerialdirigent Peter Greisler

Registergericht:
Amtsgericht Hannover | B 210251

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



Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Tabellenverzeichnis	IV
I Einleitung	1
II Übersicht zum DZHW-Absolventenpanel 2005	2
III Datennutzungshinweise	4
1 Inhalt und Anlage der Studie	8
2 Erhebungsinstrumente	11
2.1 Inhalte der Erhebungsinstrumente	11
2.2 Pretests	13
3 Grundgesamtheit und Stichprobenverfahren	15
4 Durchführung der Erhebungen	17
5 Rücklauf	19
6 Datenaufbereitung	22
6.1 Datenübertragung	22
6.2 Codierung offener Angaben	22
6.3 Datenprüfung und Datenbereinigung	23
6.4 Generierung von Variablen	25
6.5 Erstellung der Datensätze	26
6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels	27
6.7 Codierung fehlender Werte	28
7 Gewichtung	30
7.1 Vorgehen und Anwendungshinweise	30
7.2 Gewichtung der Daten	31
8 Anonymisierung	34
9 Literaturverzeichnis	38
Anhang 1: Probit-Regression zur Erstellung des Panelausfallgewichts in Welle 2	39

Abbildungsverzeichnis

Abbildung 1: Datenzugangswege und Analysepotential.....	5
Abbildung 2: Kohorten-Panel-Design der DZHW-Absolventenstudienreihe.....	8
Abbildung 3: Kalendarium des DZHW-Absolventenpanels 2005, 1. Welle	12
Abbildung 4: Rücklauf des DZHW-Absolventenpanels 2005 im Zeitverlauf, 1. Welle	19
Abbildung 5: Rücklauf des DZHW-Absolventenpanels 2005 im Zeitverlauf, 2. Welle	20
Abbildung 6: Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten des DZHW-Absolventenpanels 2005	35

Tabellenverzeichnis

Tabelle 1: Anlage der Hauptuntersuchungen der DZHW-Absolventenstudienreihe von 1989 bis 2005.....	9
Tabelle 2: Brutto-, Nettostichproben und Rücklaufquoten des DZHW-Absolventenpanels 2005..	20
Tabelle 3: Vercodete Merkmale und Codierlisten im DZHW-Absolventenpanel 2005.....	23
Tabelle 4: Themengebiete und Kürzel für Variablennamen des DZHW-Absolventenpanels 2005.	28
Tabelle 5: Systematik des FDZ-DZHW für fehlende Werte	29
Tabelle 6: Bereitgestellte Gewichte zum DZHW-Absolventenpanel 2005.....	31
Tabelle 7: Maßnahmen der statistischen Anonymisierung der Daten des DZHW- Absolventenpanels 2005 nach Zugangsweg.....	36

I Einleitung

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

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

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation zur ersten und zweiten Befragungswelle des Absolventenpanels 2005 (doi: 10.21249/DZHW:gra2005:1.0.0).⁵ Weitere Dokumentationsmaterialien zur Studie (Datensatzreports, Fragebögen, Filterführungsdiagramme etc.) können frei im Metadatensuchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden.

Abschnitt II dieses Berichts stellt die Eckdaten des Absolventenpanels 2005 im Überblick dar. Die zentralen Informationen zur Nutzung der Daten dieser Studie folgen in Abschnitt III. Kapitel 1 stellt Inhalt und Anlage der Absolventenstudienreihe bis 2005⁶ im Allgemeinen und des Absolventenpanels 2005 im Speziellen vor. Die weitere Gliederung des Berichts orientiert sich im Wesentlichen am Ablauf des Forschungsprozesses. In Kapitel 2 werden die eingesetzten Erhebungsinstrumente und in den Kapiteln 3 bis 6 der Erhebungsprozess beschrieben (Stichprobenziehung, Erhebungsablauf, Rücklauf, Datenaufbereitung). In den Kapiteln 7 und 8 folgt die Darstellung der vorgenommenen Gewichtung und Anonymisierung.

¹ Aktuelle Informationen zur Absolventenstudienreihe können über die Website des Projektes (http://www.dzhw.eu/abteilungen/bildung/projekte/pr_detail?pr_id=467) abgerufen werden.

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

³ Da zum Erhebungszeitpunkt der Daten keine Datennachnutzung vorgesehen war, sind einige Informationen zur Erhebung nicht mit dem Fokus einer späteren Datennachnutzung dokumentiert worden und teilweise nicht mehr rekonstruierbar.

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

⁵ Die Daten der dritten Welle, die im Jahr 2016 durchgeführt wird, können erst zu einem späteren Zeitpunkt veröffentlicht werden.

⁶ Es werden nur die Kohorten bis zur hier dokumentierten Kohorte 2005 berücksichtigt.

II Übersicht zum DZHW-Absolventenpanel 2005

Studienreihe	DZHW-Absolventenstudien
Kohorte	Absolvent(inn)enkohorte 2005 (5. Kohorte der Studienreihe)
Erhebende Institution	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)
Förderung	Bundesministerium für Bildung und Forschung (BMBF)
Projektmitarbeiter(innen) (Projektleitung)	<u>Kolja Briedis</u> , Michael Grotheer, Sören Isleib, <u>Karl-Heinz Minks</u> , Nicolai Netz
Themen	Studienverlauf Übergang in den Beruf Beruflicher Werdegang Weiterqualifizierung
Erhebungsdesign	Kohorten-Panel-Design
Grundgesamtheit	Hochschulabsolvent(inn)en, die im Wintersemester 2004/2005 oder im Sommersemester 2005 ihren ersten berufsqualifizierenden Studienabschluss an einer staatlich anerkannten Hochschule in der Bundesrepublik Deutschland erworben haben (mit Ausnahme der Absolvent(inn)en von Bundeswehrhochschulen, Verwaltungsfachhochschulen, Berufsakademien und Fernhochschulen)
Stichproben	Absolvent(inn)en traditioneller Studiengänge: quotierte geschichtete Klumpenstichprobe Absolvent(inn)en aus Bachelor-Studiengängen: bewusste Auswahl
Erhebungsmethode	Standardisierte postalische Befragung
Erhebungszeitraum	1. Welle: 1. Januar 2006 bis 18. Mai 2007 2. Welle: 6. Dezember 2010 bis 20. September 2011
Auswertbare Fälle	1. Welle: n = 11.788 (davon 1.622 Bachelor-Absolvent(inn)en) 2. Welle: n = 6.459 (davon 797 Bachelor-Absolvent(inn)en)
Rücklaufquote	1. Welle: 24,7 % 2. Welle: 60,3 %
Datenprodukte und Zugangswege	CUF: Download SUF: Download, Remote-Desktop, On-Site

Datensatzstruktur	Personendatensätze im wide-Format Episodendatensätze im long-Format
Besonderheiten der Daten	Getrennte Datensätze für Absolvent(inn)en traditioneller Studiengänge und Bachelorabsolvent(inn)en wegen unterschiedlicher Stichprobenziehung
DOI	10.21249/DZHW:gra2005:1.0.0
Weitere Informationen	https://fdz.dzhw.eu

Projektpublikationen*

Briedis, K. (2007). *Übergänge und Erfahrungen nach dem Hochschulabschluss. Ergebnisse der HIS-Absolventenbefragung 2005* (HIS: Forum Hochschule 13/2007). Hannover: HIS.

Briedis, K. & Minks, K.-H. (2007). *Generation Praktikum. Mythos oder Massenphänomen*. Hannover: HIS.

Grotheer, M., Isleib, S., Netz, N. & Briedis, K. (2012). *Hochqualifiziert und gefragt. Ergebnisse der zweiten HIS-HF Absolventenbefragung des Jahrgangs 2005* (HIS: Forum Hochschule 14/2012). Hannover: HIS.

* Alle Projektpublikationen werden auf der Website des Projektes (http://www.dzhw.eu/projekte/pr_show?pr_id=298) zum Download bereitgestellt.

Publikationen zum Datensatz (Auswahl)

Schaeper, H. (2009). Development of competencies and teaching–learning arrangements in higher education: findings from Germany. *Studies in Higher Education*, 34 (6), 677–697. doi:10.1080/03075070802669207

Jaksztat, S. (2014). Bildungsherkunft und Promotionen: Wie beeinflusst das elterliche Bildungsniveau den Übergang in die Promotionsphase? *Zeitschrift für Soziologie*, 43 (4), 286–301.

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

Kratz, F. & Netz, N. (2016). Which mechanisms explain monetary returns to international student mobility? *Studies in Higher Education*. doi:10.1080/03075079.2016.1172307

III Datennutzungshinweise

[Voraussetzungen der Datennutzung] Die Daten des Absolventenpanels 2005 werden durch das FDZ-DZHW entsprechend dem Bundesdatenschutzgesetz (vgl. § 40 Abs. 1 und Abs. 2 BDSG) ausschließlich zur wissenschaftlichen Nutzung und anonymisiert bereitgestellt.⁷ Das FDZ bietet ein *Scientific Use File* (SUF) für die wissenschaftliche Sekundärnutzung und ein *Campus Use File* (CUF) für Lehr- und Übungszwecke an.

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

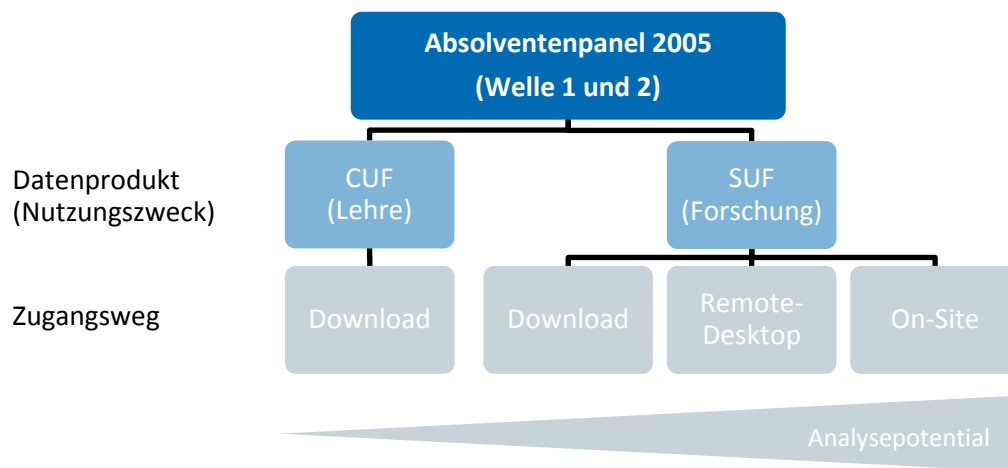
[Datenzugang] Das CUF des Absolventenpanels 2005 kann nach Bereitstellung am lokalen Computer genutzt werden. Das SUF wird über drei Zugangswege angeboten, die hinsichtlich des Speicherortes, der Möglichkeit der eigenständigen Verknüpfung mit externen Daten und der Kontrollmöglichkeiten des FDZ unterschiedlich restriktiv sind.

- **Download:** Die Daten werden verschlüsselt per E-Mail versandt oder auf der Website des FDZ zum Download bereitgestellt. Datennutzer(innen) können die Daten auf ihrem lokalen Computer speichern, falls gewünscht selbst mit Daten aus externen Quellen verknüpfen und die Daten mit eigener Software analysieren.
- **Remote-Desktop:** Die Daten werden auf einem Terminal-Server des FDZ bereitgestellt. Über eine besonders gesicherte Verbindung zwischen dem lokalen Computer der nutzenden Person und dem Terminal-Server des FDZ können die Daten mit der auf dem Terminal-Server vorhandenen Software analysiert werden. Das Übertragen der Daten auf den lokalen Computer ist nicht möglich. Analyseergebnisse werden erst nach einer Prüfung auf datenschutzrechtliche Unbedenklichkeit durch das FDZ freigegeben und zur Verfügung gestellt.
- **On-Site:** Die Daten werden in den Räumlichkeiten des FDZ in einer kontrollierten Umgebung an einem speziell gesicherten Computer zur Analyse bereitgestellt. Wie beim Remote-Desktop-Zugang werden Analyseergebnisse erst nach einer Prüfung auf datenschutzrechtliche Unbedenklichkeit durch das FDZ freigegeben und zur Verfügung gestellt.

⁷ Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, 6ff.), an dem sich bereits das Leibniz-Institut für Bildungswissenschaften (LIfBi) (vgl. Koberg, 2016, 699ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner, Müller, Schmucker & Roß, 2012, 9f.) orientieren. Das FDZ-DZHW hat den Portfolio-Ansatz an die Anforderungen der eigenen Datenbestände angepasst und nutzt vier Kategorien von Maßnahmen zur Sicherstellung des Datenschutzes, die in unterschiedlicher Weise kombiniert werden: Rechtlich-institutionelle Maßnahmen, informationelle Maßnahmen, technische Maßnahmen und statistische Maßnahmen.

Die bereitgestellten Daten weisen je nach Zugangsweg einen unterschiedlich hohen Informationsgehalt auf und unterscheiden sich damit hinsichtlich ihres Analysepotentials (vgl. Abbildung 1). Dabei gilt: Je stärker der Datenzugriff der Nutzer(innen) durch technische und organisatorische Maßnahmen kontrolliert wird, desto mehr Informationen können für die Datennutzer(innen) bereitgestellt werden.⁸ Mit diesem Vorgehen wird ein Höchstmaß an Nutzbarkeit und gleichzeitig ein bestmöglicher Schutz der bereitgestellten Daten sichergestellt.

Abbildung 1: Datenzugangswege und Analysepotential



[Datenprodukte] Über den *Digital Object Identifier* (DOI) 10.21249/DZHW:gra2005:1.0.0 ist eine Webseite mit zentralen Informationen zur Studie, weiteren Dokumentationsmaterialien sowie einer Übersicht der zur Verfügung stehenden Datenprodukte zur Studie erreichbar.

Die bereitgestellten Daten des Absolventenpanels 2005 sind in mehreren Datensätzen abgelegt. Erstens sind die Daten der Bachelorabsolvent(inn)en von den Daten der Absolvent(inn)en traditioneller Studiengänge abgetrennt. Zweitens liegen für beide Absolvent(inn)engruppen jeweils zwei getrennte Datensätze, ein Personendatensatz im wide-Format und ein Episodendatensatz im long-Format, vor (vgl. Kapitel 6.5). Beim SUF werden für jeden im FDZ-DZHW angebotenen Zugangsweg alle vier Datensätze – jeweils mit zugangswegspezifischem Analysepotential (vgl. Abbildung 1) – bereitgestellt. Beim CUF werden lediglich die beiden Datensätze der Absolvent(inn)en traditioneller Studiengänge herausgegeben.⁹

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

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

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

⁹ Aus Anonymisierungsgründen werden dabei jedoch nur die Daten einer Substichprobe bereitgestellt (vgl. Kapitel 8).

[Pflichten der Datennutzer(innen)] Die Datennutzer(innen) sind verpflichtet, folgende Regeln¹⁰ einzuhalten:

- **Wissenschaftliche Nutzung:** Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.
- **De-Anonymisierungsverbot:** Jeder Versuch der Re-Identifikation von Analyseeinheiten (z. B. Personen, Haushalten, Institutionen) ist verboten.
- **Gebot zur Mitteilung von Sicherheitslücken:** Falls Datennutzer(innen) Kenntnis von Sicherheitslücken hinsichtlich Datenschutz bzw. Datensicherheit erlangen, müssen diese dem FDZ-DZHW unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** SUF dürfen nur durch die Person genutzt werden, die den Datennutzungsvertrag abgeschlossen hat. CUF dürfen ausschließlich im Rahmen der angegebenen Lehrveranstaltung weitergegeben werden.
- **Löschungsgebot:** Download-SUF sind nach Ablauf der vereinbarten Nutzungsdauer (in der Regel drei Jahre) von jeglichen Rechnern, Servern und Datenträgern zu löschen. Ebenso müssen alle Sicherungskopien, modifizierte Datensätze (z. B. Arbeits-, Auszugs- oder Hilfsdateien) sowie Ausdrucke vernichtet werden.
- **Bereitstellung/Meldung von Publikationen:** Jede Art von Publikation, die aus der Arbeit mit Daten des FDZ-DZHW hervorgeht, muss dem FDZ im Voraus gemeldet und nach Veröffentlichung unverzüglich in elektronischer Form zur Verfügung gestellt werden. Informationen zu bereits vorhandenen Veröffentlichungen können dem Metadatensuchsystem¹¹ entnommen werden.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen laut folgenden Vorgaben zitiert werden.

[Zitation]

- **Verwendeter Datensatz:**
Minks, K. H., Briedis, K., Grotheer, M., Isleib, S. & Netz, N. (2012). DZHW-Absolventenpanel 2005. Aufbereitet durch Baillet, F., Franken, A. & Weber, A., doi: 10.21249/DZHW:gra2005:1.0.0, DATENSATZNAME¹², released 2017. Hannover: FDZ-DZHW.
- **Daten- und Methodenbericht:**
Baillet, F., Franken, A. & Weber A. (2017). DZHW-Absolventenpanel 2005. Daten- und Methodenbericht zu den Erhebungen der Absolvent(inn)enkohorte 2005 (1. und 2. Befragungswelle). Version 1.0.0. Hannover: FDZ-DZHW.

¹⁰ Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

¹¹ <https://metadata.fdz.dzhw.eu>

¹² An dieser Stelle bitte den genauen Namen der verwendeten Datensatzversion angeben, z. B. gra2005_tS_p_o_1-0-0.dta für das Onsite-SUF des Personendatensatzes der Absolvent(inn)en traditioneller Studiengänge des Prüfungsjahrgangs 2005.

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

*Diese Arbeit nutzt die Daten des DZHW-Absolventenpanels 2005. Die Daten sind vom Forschungsdatenzentrum des Deutschen Zentrums für Hochschul- und Wissenschaftsforschung (FDZ-DZHW) unter der doi: 10.21249/DZHW:gra2005:1.0.0 veröffentlicht worden.*¹³

¹³ In englischsprachigen Publikationen: This scientific work uses data of the graduate panel 2005, conducted by the German Center for Higher Education Research and Science Studies (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung; DZHW). The data were published by the research data center of the DZHW, doi: 10.21249/DZHW:gra2005:1.0.0.

1 Inhalt und Anlage der Studie

[Studienreihe] Das DZHW-Absolventenpanel 2005 ist Teil der DZHW-Absolventenstudienreihe, in der anhand von standardisierten Befragungen Informationen zu Studium, Berufseintritt, Berufsverlauf und Weiterqualifizierung von Hochschulabsolvent(inn)en erfasst werden. Das erste Absolventenpanel wurde 1989¹⁴ durchgeführt, seitdem wird jeder vierte Absolvent(inn)enjahrgang (Kohorte) untersucht. Die Grundgesamtheit einer Kohorte sind Hochschulabsolvent(inn)en, die im Winter- oder Sommersemester des betreffenden Prüfungsjahrs ihren Studienabschluss an einer Hochschule in Deutschland erworben haben.¹⁵

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

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

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

¹⁴ Schon seit 1974 werden Hochschulabsolvent(inn)en – neben den Studienabbrecher(inne)n und Hochschulwechsler(inne)n – im Rahmen der Exmatrikuliertenbefragung des DZHW befragt. Diese Untersuchungsreihe wird seit Anfang der 2000er Jahre unter dem Namen „Studienabbruch – Umfang und Motive“ durchgeführt.

¹⁵ Für die Kohorte 1989 wurden ausschließlich Absolvent(inn)en aus den Bundesländern der damaligen Bundesrepublik Deutschland ausgewählt.

Die Befragungen der Absolvent(inn)enkohorten von 1989 und 1993 umfassten zwei Wellen, seit 1997 wird eine zusätzliche dritte Befragungswelle durchgeführt. Die erste Welle einer Kohorte findet im Mittel ein Jahr nach dem jeweiligen Studienabschluss statt. Die zweite Befragungswelle folgt etwa fünf Jahre nach dem Studienabschluss. Etwa zehn Jahre nach dem Abschluss schließt sich eine dritte Befragungswelle an, die sich aus einer Hauptbefragung und separaten Vertiefungsbefragungen zu spezifischen Themen zusammensetzt. Die verschiedenen Erhebungen wurden als schriftlich-postalische Paper-and-Pencil-Befragung (Paper and Pencil Interview; PAPI) durchgeführt, mittlerweile 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 2005

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

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

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

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

[Analysepotential] In allen Kohorten wird je Befragungswelle ein identischer Kern an Informationen erhoben. Auf dieser Grundlage können mit Zeitreihen- bzw. Kohortenvergleichen langfristige Trends der Hochschulbildung und Arbeitsmarktentwicklung in den Blick genommen werden. Zudem wird innerhalb einer Kohorte ein Teil der Fragen in den verschiedenen Befragungswellen wiederholt gestellt. Dies ermöglicht die Betrachtung intra-individueller Veränderungen zwischen den Wellen (z. B. kausale Panelanalysen). Besonders hervorzuheben ist, dass in allen Kohorten über die Wellen hinweg monatsgenaue kontinuierliche 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 unterscheidet die DZHW-Absolventenstudienreihe von anderen in Deutschland durchgeführten Absolventenstudien. So ist beispielsweise das Bayerische Absolventenpanel (BAP) des Bayerischen Staatsinstituts für Hochschulforschung und Hochschulplanung (IHF) auf Absolvent(inn)en bayrischer Hochschulen begrenzt.¹⁶ Das Kooperationsprojekt Absolventenstudie (KOAB) des International Centre for Higher Education Research (INCHER) befragt Absolvent(inn)en seiner Kooperationshochschulen und ermöglicht individuelle Analysen auf Hochschul- und Studiengangsebene, die zur Evaluation und Weiterentwicklung genutzt werden können.¹⁷

[Spezifika des Absolventenpanels 2005] Neben den allgemeinen Charakteristika der Studienreihe weist die hier betrachtete Absolvent(inn)enkohorte 2005¹⁸, für die momentan Daten aus zwei schriftlich-postalisch durchgeführten Befragungswellen zur Nutzung bereitgestellt werden¹⁹, folgende Spezifika auf. Im Unterschied zu vorangegangenen Absolvent(inn)enkohorten ist die Studienphase der Kohorte 2005 durch den Hochschulwandel im Rahmen des Bologna-Prozesses geprägt. Zur Untersuchung der geforderten Internationalisierung des Studiums und der Verbesserung der Berufsfähigkeit von Absolvent(inn)en wurde das Erhebungsinstrument um neue Fragen zu Auslandsmobilität, Kompetenzentwicklung und Studienorganisation erweitert. Darüber hinaus wurden – im Hinblick auf die Studienstrukturenform – nicht nur Absolvent(inn)en traditioneller Studiengänge, sondern erstmalig auch eine nennenswerte Anzahl an Bachelorabsolvent(inn)en befragt. Zur Untersuchung dieser neuen Absolvent(inn)engruppe, deren Anteil im Prüfungsjahr 2005 4,4 Prozent (vgl. Dudek, Glässner & Schröder, 2010, S. 25) der Gesamtabzolvent(inn)enzahl betrug²⁰, wurde eine gesonderte Stichprobe gezogen (vgl. Kapitel 3). Dabei wurden lediglich Absolvent(inn)en aus Fächern berücksichtigt, für die bereits eine größere Absolvent(inn)enzahl vorlag.²¹

Neben den sich wandelnden Studienbedingungen ergaben sich für die Absolvent(inn)enkohorte 2005 auch andere Arbeitsmarktbedingungen als für die vorangegangenen Kohorten. So ist ihre berufliche Einstiegs- und Konsolidierungsphase durch den Beginn der Wirtschafts- und Finanzkrise im Jahr 2008 gekennzeichnet. Hervorzuheben ist ferner, dass in der ersten Welle des Absolventenpanels 2005 – aufgrund des zum Befragungszeitpunkt aktuellen gesellschaftspolitischen Diskurs zur "Generation Praktikum" – zusätzliche Informationen zu Praktika nach dem Studium erhoben wurden.²²

¹⁶ vgl. dazu <http://www.bap.ihf.bayern.de>

¹⁷ vgl. dazu <http://koab.uni-kassel.de>

¹⁸ Die Grundgesamtheit sind Hochschulabsolvent(inn)en, die im Wintersemester 2004/2005 oder im Sommersemester 2005 ihren ersten berufsqualifizierenden Studienabschluss an einer staatlich anerkannten Hochschule in der Bundesrepublik Deutschland erworben haben (mit Ausnahme der Absolvent(inn)en von Bundeswehrhochschulen, Verwaltungsfachhochschulen, Berufsakademien und Fernhochschulen).

¹⁹ Die dritte Befragung wird 2016 als Online-Befragung durchgeführt. Die Daten dieser Welle können daher erst zu einem späteren Zeitpunkt veröffentlicht werden.

²⁰ Weitere 4 Prozent hatten einen Masterabschluss. Für den Prüfungsjahrgang 2001 lag der Anteil an Bachelor- und Masterabsolvent(inn)en noch bei 0,6 Prozent (nur Bachelor: 0,1 Prozent).

²¹ Aufgrund der unterschiedlichen Ziehungsverfahren für Absolvent(inn)en traditioneller Studiengänge und Bachelorabsolvent(inn)en werden gesonderte Datensätze für die Befragengruppen bereitgestellt.

²² vgl. dazu auch Briedis & Minks, 2007.

2 Erhebungsinstrumente

In beiden Befragungswellen des Absolventenpanels 2005 wurde als Erhebungsinstrument ein standardisierter Papierfragebogen in deutscher Sprache eingesetzt.²³ Kapitel 2.1 stellt die zentralen Inhalte der beiden 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 2005 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.²⁴ Daran anknüpfend wird der weitere Werdegang der Absolvent(inn)en in den Blick genommen.

Zu diesem Zweck wird über die Wellen hinweg der individuelle Tätigkeitsverlauf der Absolvent(inn)en seit ihrem Studienabschluss erfasst. Pro Tätigkeit wird jeweils deren Art (z. B. Erwerbstätigkeit, Promotion, Elternzeit) sowie ihr Anfangs- und Endzeitpunkt monatsgenau erhoben. Dies geschieht seit der ersten Welle der Kohorte 2001 bzw. der zweiten Welle der Kohorte 1997 in Form eines Kalendariums (Frage 4.7 in Welle 1 und Frage 1.7 in Welle 2), in das die Befragten ihre einzelnen Tätigkeiten eintragen sollen (vgl. Abbildung 3).²⁵ Das Kalendarium wurde vom DZHW konzipiert, um lückenhafte Angaben in der Beschreibung von Tätigkeitsverläufen zu reduzieren.

²³ Die Fragebögen sowie Filterführungsdiagramme beider Wellen können im Metadatensuchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden.

²⁴ vgl. Abschnitt 1 „Studienverlauf und Studienerfahrungen“ im Fragebogen der ersten Welle.

²⁵ Vor der Einführung des Kalendariums wurden die Tätigkeitsverläufe anhand eines Tableaus erfasst.

Abbildung 3: Kalendarium des DZHW-Absolventenpanels 2005, 1. Welle

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

Bitte kennzeichnen Sie den Monat, in dem Sie die letzte Prüfungsleistung Ihres abgeschlossenen Studiums erbracht haben, mit einem **X** und tragen Sie für die Zeit vom Studienabschluss bis heute Ihre Tätigkeiten anhand der aufgeführten Kennbuchstaben in den Kalender ein. Haben Sie mehrere Tätigkeiten gleichzeitig ausgeübt, können Sie diese untereinander aufführen. Wichtig ist, dass es keine zeitlichen Lücken gibt.

Beispiel:

Im Oktober 2004 erbrachten Sie Ihre letzte Prüfungsleistung (**X**). Zwischen November 2004 und Februar 2005 machten Sie zunächst ein Praktikum (**P**), danach waren sie arbeitslos (**AL**). Ab März 2005 bis zum Zeitpunkt der Befragung übten Sie (evtl. auf verschiedenen Stellen) eine nichtselbständige Erwerbstätigkeit aus (**A**). Parallel zu Ihrer Erwerbstätigkeit (z. B. auf einer Promotionsstelle) arbeiteten Sie seit Juni 2005 an Ihrer Dissertation (**D**).

	Jan.	Feb.	März	April	Mai	Juni	Juli	Aug.	Sept.	Okt.	Nov.	Dez.
2004										X	P	AL
2005	AL		A			D						
2006	A D											

Erwerbstätigkeiten

R	Referendariat, Anerkennungspraktikum u. Ä.
W	Werkvertrag, Honorararbeit
J	Jobben
SE	Selbständige Erwerbsarbeit (ohne Werk-/Honorararbeit)
A	Nichtselbständige Erwerbsarbeit (z. B. als Angestellte/r oder Beamte/r)

Andere Tätigkeiten

D	Promotion	EZ	Elternzeit
ST	Studium	H	Hausfrau, Hausmann, Familienarbeit
P	Praktikum	AL	Arbeitslosigkeit
B	Berufsausbildung, Um- schulung, Volontariat	SO	Sonstiges (z. B. Wehr-/ Zivildienst, Urlaub)
F	Fort-/Weiterbildung (Vollzeit, längerfristig)		

Ergänzend zum Kalendarium werden zu einzelnen Tätigkeitsarten zusätzliche Informationen erhoben. Ein Großteil der Fragen bezieht sich dabei auf die Erwerbstätigkeiten. Anhand des Erwerbstätigkeitstableaus (Frage 5.2 in Welle 1 und Frage 4.3 in Welle 2), das zeitgleich mit dem Kalendarium eingeführt wurde, werden für alle angegebenen Erwerbstätigkeiten der Zeitraum, die Art des Arbeitsverhältnisses, der Arbeitszeitumfang, die berufliche Stellung und der Arbeitsort erhoben. Darüber hinaus werden für die erste und die letzte Erwerbstätigkeit noch weitere Merkmale erfasst, etwa das Einkommen und die Branche.²⁶ Daneben beinhalten die Erhebungsinstrumente auch Fragen zur subjektiven Einschätzung der eigenen beruflichen Situation, etwa hinsichtlich der beruflichen Zufriedenheit, den beruflichen Zukunftsperspektiven und der Beschäftigungsadäquanz²⁷. Ein Großteil der Informationen zu den Erwerbstätigkeiten wird über die Befragungswellen hinweg wiederholt erhoben.²⁸ 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.

²⁶ Vor der Einführung des Erwerbstätigkeitstableaus wurden – abgesehen von der Art der Institution – generell nur zusätzliche Informationen zur ersten und letzten Erwerbstätigkeit erhoben.

²⁷ vgl. dazu auch Kerst & Fehse, 2007.

²⁸ Teilweise wurde allerdings der Wortlaut abgewandelt. Darauf hinzuweisen ist außerdem, dass sich die Wiederholungsmessung in der zweiten Welle bei einer beruflichen Veränderung auf eine andere berufliche Situation bzw. Tätigkeit bezieht als in der ersten Welle.

Neben den Erwerbstätigkeiten werden auch andere Tätigkeitsarten vertiefend behandelt. So werden etwa Fragen zu akademischen Weiterqualifizierungen, vor allem Promotionen, 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 2005] Innerhalb dieser allgemeinen Themenblöcke, die in den Erhebungsinstrumenten aller Kohorten der Studienreihe enthalten sind, wurden in die Fragebögen des Absolventenpanels 2005 spezifische Fragen aufgenommen, die in den vorangegangenen Kohorten nicht gestellt wurden. Mit der Internationalisierung des Studiums und der Berufsfähigkeit von Hochschulabsolvent(inn)en wurden in der ersten Welle zwei zentrale Ziele des Bologna-Prozesses in den Blick genommen. Dabei wurde zum einen eine Frage zu Auslandserfahrungen im Studium neu in den Fragebogen eingefügt (Frage 1.6). Zum anderen wurden die Informationen zur Studienorganisation, zu Lehr- und Lernformen sowie zur Qualifikations- und Kompetenzentwicklung detaillierter als zuvor erhoben (Fragen 1.14 bis 1.18).²⁹ Darüber hinaus wurden in der ersten Welle zur Prüfung der These der „Generation Praktikum“ verschiedene Informationen zu Praktika nach dem Studium erfasst (Fragen 4.10 bis 4.16).³⁰ Das Erhebungsinstrument der zweiten Welle umfasst im Vergleich zu vorangegangenen Kohorten nur einzelne spezifische Fragen (z. B. Frage 8.11 und Frage 8.12 zur räumlichen Mobilität). Insbesondere bestehen viele Vergleichsmöglichkeiten mit dem entsprechenden Erhebungsinstrument der Kohorte 2001.

2.2 Pretests

[Ziel und Verfahren] Die Erhebungsinstrumente beider Wellen wurden im Vorfeld der Erhebungen durch Pretests geprüft. Dabei sollte erstens für die bereits in Erhebungsinstrumenten vorheriger Kohorten eingesetzten Fragen und Antwortvorgaben geprüft werden, ob sie von der Absolvent(inn)enkohorte 2005 gleich perzipiert werden würden wie von vorangegangenen Kohorten. Zweitens sollte überprüft werden, ob die Erhebungsinstrumente auch für die Befragung der neuen Absolvent(inn)engruppe der Bachelor (vgl. Kapitel 1 und 3) geeignet waren. Drittens sollte für die neu eingesetzten Fragen die Verständlichkeit und Beantwortbarkeit getestet werden. Die Aufnahme der neuen Fragen zog außerdem Veränderungen am Aufbau und Layout des Fragebogens sowie der Befragungsdauer nach sich, die evaluiert werden sollten. Zur Prüfung dieser verschiedenen Aspekte kam 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).³¹

²⁹ Frage 1.17 war in dieser Form bereits im Jahrgang 2001 Bestandteil des Erhebungsinstrumentes. Im Vergleich zu den Vorgängerbefragungen wird nicht nur erhoben, inwieweit die aufgeführten Kompetenzen im Beruf gefordert werden, sondern auch, in welchem Maße die Befragten nach eigener Einschätzung über diese Kenntnisse und Fähigkeiten verfügen. Zudem wurde die dreistufige Antwortskala durch eine fünfstufige ersetzt (vgl. Schaeper & Briedis, 2004, S. 8).

³⁰ vgl. dazu auch Briedis & Minks, 2007.

³¹ 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.

[Probanden] In beiden Befragungswellen wurden als Testpersonen zum einen Mitarbeiter(innen) des DZHW ausgewählt, die ihren Hochschulabschluss im Prüfungsjahrgang 2005 oder einem angrenzenden Prüfungsjahrgang erworben hatten. Durch die Befragung dieser Testpersonen wurden so gleichzeitig Expert(inn)en aus dem Bereich der Hochschulforschung einbezogen. Zum anderen wurden über persönliche oder projektbezogene Kontakte verschiedene externe Personen rekrutiert. In beiden Wellen nahmen zwischen 10 und 15 Personen am Pretest teil.³²

[Durchführung] Beide Pretests fanden etwa zwei bis drei Monate vor dem jeweiligen Feldstart statt. Dabei wurden die Testpersonen gebeten, den für die Hauptbefragung vorgesehenen standardisierten Fragebogen zu bearbeiten und Verständnisprobleme, Kritik oder Anmerkungen zu notieren. Im Anschluss an die Bearbeitung des Fragebogens wurde den Probanden ein Fragenkatalog zu verschiedenen Aspekten der Befragung vorgelegt. 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. In der ersten Welle wurden außerdem konkrete Nachfragen zu neuen oder veränderten Fragen gestellt.³³ Auf Grundlage der Pretestergebnisse wurden die Formulierungen verschiedener Fragetexte präzisiert sowie die Reihenfolgen von Items und Antwortkategorien in einzelnen Itembatterien und Mehrfachnennungen überarbeitet.³⁴ Der grundsätzliche Aufbau, der Umfang und das Layout der Fragebögen wurden unverändert beibehalten.

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

³³ Außerdem wurde in der ersten Befragungswelle im Rahmen des Pretests das Anschreiben an die Befragten im Hinblick auf Inhalte, Verständlichkeit und Länge evaluiert.

³⁴ Die spezifischen Gründe für Änderungen können nicht mehr rekonstruiert werden.

3 Grundgesamtheit und Stichprobenverfahren

[Grundgesamtheit] Die Grundgesamtheit des Absolventenpanels 2005 umfasst alle Hochschulabsolvent(inn)en, die im Wintersemester 2004/2005 oder im Sommersemester 2005³⁵ ihren ersten berufsqualifizierenden Studienabschluss an einer staatlich anerkannten Hochschule in der Bundesrepublik Deutschland erworben haben. Hochschulabsolvent(inn)en von Bundeswehrhochschulen, Verwaltungsfachhochschulen, Berufsakademien und Fernhochschulen waren dabei von vornherein ausgenommen.

[Stichprobenverfahren] Aufgrund fehlender oder nicht zugänglicher Listen von Hochschulabsolvent(inn)en mussten die Individuen über die Hochschulen rekrutiert werden. Dabei erfolgten getrennte Ziehungsprozesse für Hochschulabsolvent(inn)en traditioneller Studiengänge³⁶ und Hochschulabsolvent(inn)en aus Bachelorstudiengängen. Für Absolvent(inn)en traditioneller Studiengänge wurde eine quotierte geschichtete Klumpenstichprobe gezogen. Die Stichprobe der Bachelorabsolvent(inn)en basierte auf einer bewussten Auswahl.³⁷

[Quotierte geschichtete Klumpenstichprobe] Die „primary sampling units“ (Klumpen) der Klumpenstichprobe wurden anhand der Hochschule, dem Studienbereich³⁸ sowie der Abschlussart³⁹ definiert. Die „secondary sampling units“ stellten die Hochschulabsolvent(inn)en des Prüfungsjahrgangs 2005 innerhalb dieser Klumpen dar.⁴⁰ Die Umsetzung dieses Designs erfolgte in mehreren Schritten. Im ersten Schritt wurde eine mehrstufige Zufallsstichprobe gezogen.⁴¹ Die Klumpen wurden dabei zunächst nach alten und neuen Bundesländern disproportional geschichtet, um die Stichprobengröße der neuen Bundesländer zu erhöhen. Der Auswahlatz für die Klumpen betrug 30 Prozent in den neuen (150 Klumpen) und 18 Prozent in den alten Bundesländern (400 Klumpen). Innerhalb dieser beiden Schichten wurde eine proportional geschichtete Klumpenstichprobe gezogen. Dieses Vorgehen erlaubt eine genauere Schätzung der Schichtungsmerkmale in der Population als eine einfache Zufallsstichprobe. Die Schichtungsmerkmale waren die Bundesländer⁴², Größe und Art der Hochschule⁴³. Für diese Schichtung wurde die Zufallsallokation/-rundung nach Cox (1987) verwendet.

Im zweiten Schritt erfolgte eine quotierte Ziehung⁴⁴ von weiteren 87 Klumpen⁴⁵, die teilweise die im ersten Schritt gezogenen Klumpen ersetzten und teilweise zusätzlich berücksichtigt wurden. Ziel dabei war, eine möglichst gute Angleichung der Verteilungen von

³⁵ Das Prüfungsjahr 2005 begann an Fachhochschulen im September 2004 und endete im August 2005, an Universitäten begann es im Oktober 2004 und endete im September 2005.

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

³⁷ Zum Verfahren der bewussten Auswahl vgl. Schnell, Hill & Esser, 2005, 298ff.

³⁸ entsprechend der Aufgliederung nach der amtlichen Statistik (gemäß Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2004/2005 und SoSe 2005).

³⁹ Hierbei wurde zwischen einem Diplom und vergleichbaren Abschlüssen, etwa einem Magister oder unterschiedlichen Lehramtsabschlüssen, differenziert.

⁴⁰ Beispiel für den Personenkreis eines Klumpens: Alle Hochschulabsolvent(inn)en der Leibniz-Universität Hannover, die im Studienbereich „Mathematik“ ihr Diplom abgelegt haben.

⁴¹ Diese Auswahl wurde durch das „Zentrum für Umfragen, Methoden und Analysen (ZUMA)“ in Mannheim durchgeführt.

⁴² Teilweise wurden Bundesländer zu einer Schicht zusammengefügt. Folgende Einheiten gingen in die Schichtung ein: Bayern, Baden-Württemberg, Niedersachsen, Nordrhein-Westfalen, die verbleibenden westlichen Bundesländer (Berlin, Bremen, Hamburg, Hessen, Rheinland-Pfalz, Saarland, Schleswig-Holstein), Sachsen und die verbleibenden östlichen Bundesländer (Brandenburg, Mecklenburg-Vorpommern, Sachsen-Anhalt, Thüringen).

⁴³ Es wurde zwischen Universität und Fachhochschule differenziert.

⁴⁴ Quotierungsmerkmale: Studienbereiche, Abschlussarten und Hochschulen.

⁴⁵ Diese Ziehung wurde durch das DZHW durchgeführt.

Studienbereich, Abschlussart und Hochschule der Stichprobe an die Verteilungen in der Grundgesamtheit zu erreichen.

Im dritten Schritt wurde beim Ausfall eines Klumpens (z. B. bei Teilnahmeverweigerung auf Hochschul- oder Fakultätsebene) ein – hinsichtlich der Schichtungsmerkmale – möglichst ähnlicher Klumpen als Ersatz gesucht.

[Bewusste Auswahl Bachelorabsolvent(inn)en] Die Auswahl der Bachelorabsolvent(inn)en erfolgte über eine bewusste Auswahl. Dieses Auswahlverfahren wurde gewählt, da im Prüfungsjahrgang 2005 erst wenige Hochschulabsolvent(inn)en mit einem Bachelor abgeschlossen hatten, jedoch möglichst viele Personen dieser neuen Absolvent(inn)engruppe befragt werden sollten, um erste explorative Erkenntnisse zu gewinnen. Zu diesem Zweck wurde auf der Grundlage vorangegangener Studienanfänger(innen)zahlen, durchschnittlicher Studienabbruchquoten sowie durchschnittlicher Studiendauern geschätzt, in welchen Fachrichtungen größere Anzahlen von Bachelorabsolvent(inn)en zu erwarten waren. Auf Basis dieser Schätzung wurden alle Bachelorabsolvent(inn)en der identifizierten Fachrichtungen ausgewählt.

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, Studienbereich, Abschlussart) mit, anhand derer sie die Zielpersonen für das Absolventenpanel 2005 identifizieren sollten (vgl. Kapitel 3).⁴⁶ Da die Hochschulen die Kontaktdaten ihrer Absolvent(inn)en aus Datenschutzgründen nicht herausgeben durften, teilten sie dem DZHW lediglich die jeweilige Absolvent(inn)enanzahl mit. Anschließend sendete das DZHW die passende Anzahl an Erhebungsunterlagen für die erste Befragungswelle postalisch an die jeweiligen Prüfungsämter, die diese an die Zielpersonen weiterverschickten.

Um die weiterhin teilnahmebereiten Personen in der zweiten Befragungswelle direkt durch das DZHW kontaktieren zu können, wurden im Fragebogen der ersten Welle deren Kontaktdaten 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.⁴⁷ Für die zweite Befragungswelle wurde diese Identifikationsnummer direkt in den Fragebogen gedruckt und die Erhebungsunterlagen gezielt an die zugeordnete Postadresse versandt.⁴⁸ Um dabei auch Personen zu berücksichtigen, die in der Zwischenzeit umgezogen waren, wurden die Adressbestände zwischen den Wellen geprüft und ggf. aktualisiert.⁴⁹

[Erhebungsunterlagen] Die Erhebungsunterlagen bestanden in beiden Befragungswellen pro zu befragender Person aus einem Anschreiben (inkl. Datenschutzinformationen), dem Fragebogen, 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.

[Feldphase] Der Erhebungszeitraum der ersten Befragungswelle erstreckte sich vom 1. Januar 2006 bis 18. Mai 2007.⁵⁰ Die beiden Erinnerungsschreiben wurden etwa vier bzw. acht Wochen nach Feldstart verschickt. Aufgrund des angewendeten Kontaktverfahrens über die Prüfungsämter konnte das DZHW keinen direkten Einfluss auf den genauen Versandzeitpunkt

⁴⁶ 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 2005 angehörten (dies tritt z.B. bei Verzögerung der Zeugniserstellung auf) oder es sich nicht um den Erstabschluss handelte.

⁴⁷ Zur Gewährleistung des Datenschutzes wurde der Adressabschnitt vom Fragebogen abgetrennt und die Referenzliste getrennt von den Befragungsdaten auf einem geschützten Server gespeichert.

⁴⁸ Auf Nachfrage bestand außerdem die Möglichkeit, den Fragebogen in elektronischer Form per E-Mail zur erhalten. Dieses Angebot wurde von 140 Personen in Anspruch genommen (vgl. Grotheer, Isleib, Netz & Briedis, 2012, S. 424).

⁴⁹ Die Befragten wurden dabei sowohl im Anschluss an die erste Welle als auch im Vorfeld zur zweiten Welle postalisch und per E-Mail angeschrieben und um die Aktualisierung ihrer Postadresse 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. Für die dritte Welle wurden nach Abschluss der zweiten Welle weitere Adressaktualisierungen vorgenommen.

⁵⁰ Die Feldzeit wurde so lange wie möglich – parallel zu den weiteren Projektschritten – ausgedehnt und jeder noch eintreffende Fragebogen bis Mai 2007 berücksichtigt.

der Erhebungsunterlagen nehmen.⁵¹ 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 zweiten Befragungswelle erstreckte sich vom 6. Dezember 2010 bis 20. September 2011. Durch den nun im DZHW vorliegenden Adressbestand konnten konkrete Versandzeitpunkte für die Erhebungsunterlagen festgelegt werden.⁵² Zudem konnten die Erinnerungsschreiben gezielt nur an diejenigen Zielpersonen verschickt werden, die sich noch nicht an der Erhebung beteiligt hatten.

[Rücklaufsteigernde Maßnahmen] Neben dem für die Rücklaufsteigerung besonders wirksamen Instrument der Erinnerungsschreiben und dem zusammen mit dem Anschreiben verschickten Flyer wurde im Anschreiben angekündigt, eine Zusammenfassung der zentralen Ergebnisse der Studie im Anschluss an die Erhebung zu versenden. Des Weiteren wurden auf einer Projekthomepage Informationen zum Projekt und daraus entstandenen Publikationen bereitgestellt. Darüber hinaus wurde unter allen Befragungsteilnehmer(inne)n eine Verlosung verschiedener Sachpreise durchgeführt. In der ersten Befragungswelle wurden dabei 40 Büchergutscheine à 50 Euro verlost, in der zweiten Welle ein Notebook à etwa 1.200 Euro, fünf Fluggutscheine à 200 Euro und zehn Büchergutscheine à 30 Euro.

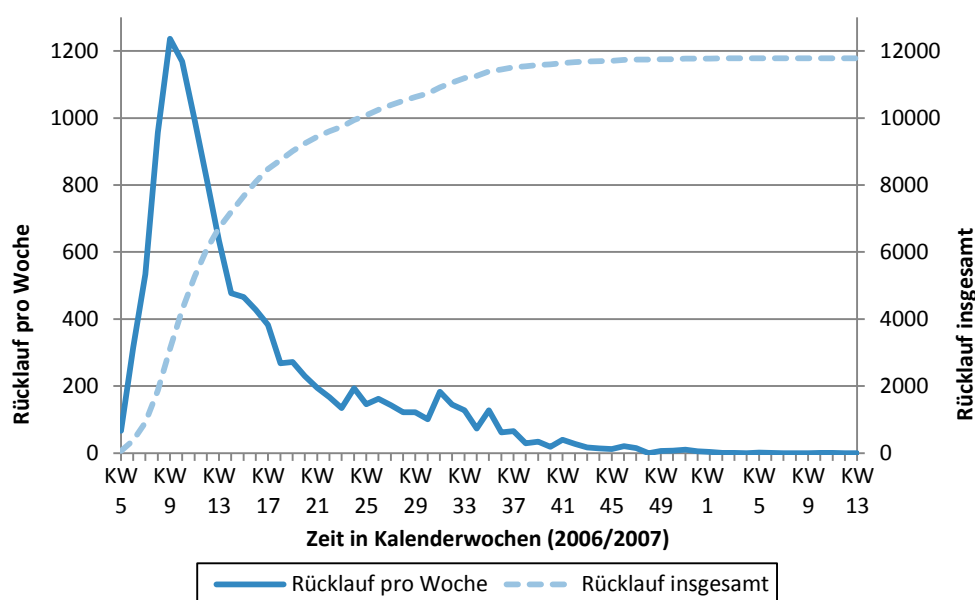
⁵¹ Nach dem Versand der Erstanschreiben meldeten die Prüfungsämter jeweils den genauen Versandtermin. Das DZHW schickte dann im Abstand von vier Wochen zu diesem Versandtermin das erste Erinnerungsschreiben an das Prüfungsamt, welches dieses wiederum an die Zielpersonen weiterschickte. Anschließend meldete das Prüfungsamt wiederum den Versandtermin dieses Erinnerungsschreibens. Das Verfahren für die zweite Erinnerung erfolgte analog.

⁵² Einladung: 6.-10.12.2010; Erste Erinnerung: 14.01.2011; Zweite Erinnerung: 11.02.2011.

5 Rücklauf

[Rücklauf] Die Bruttostichprobe der ersten Befragungswelle umfasste ca. 47.800 Absolvent(inn)en, die bei den Prüfungsämtern der im Stichprobenverfahren gezogenen Hochschulen (vgl. Kapitel 3) gemeldet waren und kontaktiert wurden.⁵³ Insgesamt wurden 12.114 Fragebögen an das DZHW zurückgeschickt. Von diesen wurden 326 entfernt, da sie nicht zur Grundgesamtheit gehörten (stichprobenneutrale Ausfälle) oder nicht auswertbar waren (stichprobenrelevante Ausfälle) (vgl. Kapitel 6.3). Somit verblieben in der Nettostichprobe der ersten Befragungswelle 11.788 Fälle, davon 1.622 Bachelorabsolvent(inn)en. Im Hinblick auf die Bruttostichprobe liegt die Rücklaufquote damit bei rund 25 Prozent (vgl. Tabelle 2). Abbildung 4 stellt den Rücklauf der Fragebögen in der Feldphase der ersten Befragungswelle im Zeitverlauf dar. Wie sich zeigt, hat ein Großteil der ausgefüllten Fragebögen das DZHW während der ersten Hälfte der Feldphase erreicht, in der auch die Erinnerungen verschickt wurden. Gleichzeitig ist festzustellen, dass auch zu späteren Zeitpunkten – hier lag auch die zweite Erinnerung bereits länger zurück – noch Fragebögen zurückgesandt wurden.

Abbildung 4: Rücklauf des DZHW-Absolventenpanels 2005 im Zeitverlauf, 1. Welle



Dargestellt sind nur die auswertbaren Fälle.

10.706 Personen, also über 90 Prozent der 11.788 Teilnehmer(innen) der ersten Welle, erklärten sich mit einer Kontaktierung für weitere Befragungen einverstanden. Diese ergeben die Bruttostichprobe der zweiten Welle. Aufgrund von verschiedenen stichprobenrelevanten Ausfällen (Nichtteilnahmen, ungültige Adressen⁵⁴ und nicht auswertbare Fragebögen⁵⁵) um-

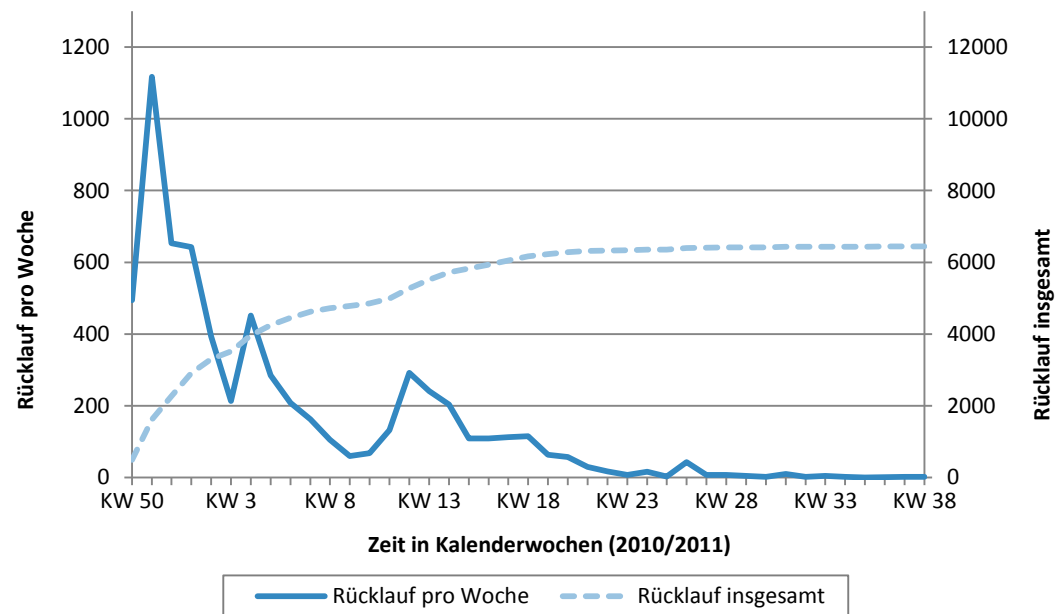
⁵³ Da eine Differenzierung zwischen einer (unbereinigten) Brutto-Ausgangsstichprobe und einer bereinigten Bruttostichprobe nicht möglich ist, wird im Folgenden nur die Bezeichnung „Bruttostichprobe“ verwendet. Ebenfalls ist nicht rekonstruierbar, wie viele dieser Absolvent(inn)en Bachelorabsolvent(inn)en waren.

⁵⁴ Dies betraf etwa 300 Fälle.

⁵⁵ Dies betraf 4 Fälle.

fasst die Nettostichprobe der zweiten Welle 6.459 Fälle, darunter 797 Bachelorabsolvent(inn)en. Bezogen auf die 10.706 Fälle der Bruttostichprobe liegt die Rücklaufquote damit bei 60 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).

Abbildung 5: Rücklauf des DZHW-Absolventenpanels 2005 im Zeitverlauf, 2. Welle



Dargestellt sind nur die auswertbaren Fälle.

Insgesamt sind über 70 Prozent der eingegangenen Fragebögen zwischen Dezember 2010 (KW 50) und Februar 2011 (KW 8) zurückgesendet worden (vgl. Abbildung 5). Innerhalb dieses Zeitraums erfolgte auch der Versand der Erinnerungsschreiben. Im März und April 2011 (KW 9 bis KW 17) gingen weitere 20 Prozent der Fragebögen ein.

Tabelle 2: Brutto-, Nettostichproben und Rücklaufquoten des DZHW-Absolventenpanels 2005

	Welle 1	Welle 2
Bruttostichprobe	47.800	10.706
Nettostichprobe	11.788 ^a	6.459 ^b
Rücklaufquote	24,7 %	60,3 %
Anteil Bruttostichprobe Welle 2 an Bruttostichprobe Welle 1	--	22,4 %
Anteil Nettostichprobe Welle 2 an Nettostichprobe Welle 1	--	54,8 %
Anteil Nettostichprobe Welle 2 an Bruttostichprobe Welle 1	--	13,5 %

^a darunter 1.622 Bachelorabsolvent(inn)en.

^b darunter 797 Bachelorabsolvent(inn)en.

[Panelausfälle] Das Absolventenpanel 2005 ist von paneltypischen Ausfallprozessen⁵⁶ 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(inn)en zu verzeichnen ist (vgl. Fabian & Briedis, 2009, 71ff.).

Die Betrachtung der Bruttostichprobe im Zeitverlauf zeigt, dass diese in der zweiten Welle nur noch etwa 22 Prozent der Bruttostichprobe der ersten Welle umfasst. Von den 11.788 Fällen aus der Nettostichprobe der ersten Welle konnten etwa 55 Prozent der Befragten in der zweiten 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 knapp 14 Prozent der ersten Bruttostichprobe an beiden Befragungswellen teilgenommen haben.

⁵⁶ Zu paneltypischen Ausfallprozessen vgl. Schnell, Hill & Esser, 2005, S. 241.

6 Datenaufbereitung

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

6.1 Datenübertragung

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

[Erstellung eines Codeplans] Auf Basis des Fragebogens der Befragung wurde ein Codeplan erstellt. Dabei wurde vermerkt, welcher Frage bzw. Teilfrage eine Variable zugeordnet ist, welchen Namen diese Variable trägt und welche numerischen Codierungen für die standardisierten Antworten der Befragten verwendet werden. Um die Erfassungsreihenfolge festzulegen, wurden die Variablen zusätzlich nummeriert.⁵⁷

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

6.2 Codierung offener Angaben

Vor der Datenübertragung erfolgte eine Codierung der (halb-)offenen Angaben. Dabei wurden diesen anhand einer Codierliste numerische Codierungen zugeordnet. Je nach Variable wurden unterschiedliche Codierlisten verwendet. Es handelt sich um Klassifikationsschlüssel der amtlichen Statistik (z. B. Klassifikation der Berufe, Schlüsselverzeichnis der Studenten- und Prüfungsstatistik etc.) oder um bereits in vorherigen Absolventenpanels eingesetzte Schlüssel. Für einige Variablen wurden neue Codierlisten auf Basis der in den Daten des Absolventenpanels 2005 vorkommenden Nennungen entwickelt. Für einige halboffene Fragen wurden keine neuen Variablen mit numerischen Codierungen erstellt, sondern die Nennungen nur – sofern möglich – den vorhandenen (geschlossenen) Antwortkategorien zugeordnet. Einzelne offene Fragen wurden nicht vercodet, weil sie vorwiegend als Kontextinformationen zur Codierung

⁵⁷ Die Daten wurden in einem einfachen spaltenorientierten Textformat ohne eine die Variablennamen enthaltene Kopfzeile erfasst. Der Codeplan musste daher festlegen, in welcher Reihenfolge die Daten zu erfassen waren, damit die zu einer Variable zugehörigen Daten in der richtigen Spalte eingetragen werden konnten.

anderer offener Angaben erfasst wurden.⁵⁸ Die durch das Primärforschungsprojekt vorgenommenen Codierungsentscheidungen wurden unverändert beibehalten.

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 im Datensatzreport sowie im MetadatenSuchsystem⁵⁹ dokumentiert.

Tabelle 3: Vercodete Merkmale und Codierlisten im DZHW-Absolventenpanel 2005

Merkmal	Codierliste	Codierlisten-ID ^b
Studienfach	Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2004/2005 und SoSe 2005), Schlüssel 3.1	cl-destatis-studienfach-2005 ^c
Studienabschluss	projekteigene Codierung	cl-dzhw-2
Fachlicher Studienschwerpunkt (für ausgewählte wirtschaftswissenschaftliche Fächer)	projekteigene Codierung	cl-dzhw-3
Hochschule	Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2004/2005 und SoSe 2005), Schlüssel 2.2	cl-destatis-hochschule-2005 ^d
Bundesland	Destatis Bundeslandschlüssel (entsprechend der ersten beiden Ziffern des Amtlichen Gemeindecchlüssels (AGS))	cl-destatis-bundesland-1990 ^e
Ausland / Staatsangehörigkeit	projekteigene Codierung	cl-dzhw-1
Berufsbezeichnung	Welle 1: Destatis Klassifikation der Berufe 1992 Welle 2: Destatis Klassifikation der Berufe 2010	cl-destatis-kldb-1992 cl-destatis-kldb-2010Vorversion
Beruflicher Aufgabenbereich ^a	projekteigene Codierung	cl-dzhw-4
sonstige offene Abfragen	Zuordnung zu vorgegebenen Kategorien oder projekteigene Codierung	---

^a vgl. Frage 5.1 (Welle 1) und Frage 4.12 (Welle 2)

^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 Codes aus älteren Schlüsselverzeichnissen, wenn Studienfächer nicht mehr im aktuellen Verzeichnis enthalten waren

^d ergänzt um projekteigene Codes, wenn nicht zuordenbar (z. B. bei Hochschulen im Ausland)

^e ergänzt um projekteigene Codes, wenn nicht zuordenbar

6.3 Datenprüfung und Datenbereinigung

[Manuelle Vorkorrektur] Bereits vor der Übertragung der Daten wurde auf den Papierfragebögen eine manuelle Prüfung und gegebenenfalls eine Anpassung von Angaben der Befragten

⁵⁸ Dies betrifft in beiden Wellen die typischen Arbeitsschwerpunkte, die neben der Berufsbezeichnung und dem Aufgabenbereich in Frage 5.1 in der ersten und Frage 4.12 in der zweiten Welle erhoben wurden. Die Angaben zu den typischen Arbeitsschwerpunkten dienten nur dazu, zusätzliche Informationen für die Codierung der ebenfalls offen abgefragten Berufsbezeichnung sowie des beruflichen Aufgabenbereichs zu erhalten.

⁵⁹ <https://metadata.fdz.dzhw.eu>

durchgeführt.⁶⁰ Dies sollte zum einen die Erfassbarkeit der Daten erleichtern. Dafür wurde die Form der bestehenden Angaben verändert. Beispielsweise wurden schwer lesbare Angaben oder Streichungen der Befragten verdeutlicht, Zahlenangaben rechtsbündig in die dafür vorgesehenen Kästchen eingetragen, verbale Angaben von Noten in Ziffern übersetzt (z. B. „gut“ = 2,0) oder Einträge im Studienverlaufstableau auf mehrere Zeilen aufgeteilt, wenn die Befragten verschiedene Phasen in eine Zeile geschrieben hatten (z. B. bei zwei parallelen Studiengängen).

Zum anderen zielte die manuelle Prüfung darauf ab, schon vor der softwaregestützten Korrektur (siehe unten) erste Fehler oder Inkonsistenzen in den Angaben der Befragten zu bereinigen. Dabei wurden zunächst Konsistenzprüfungen innerhalb einer Welle durchgeführt, bei denen inhaltliche und zeitbezogene Angaben zu gleichen Themengebieten miteinander abgeglichen wurden. Beispielsweise erfolgte im Fragebogen der ersten Welle ein Abgleich der Angaben zu Studienfach, Abschluss und Hochschule zwischen dem Studienverlaufstableau (Frage 1.1) und dem Studienabschlusstableau (Frage 1.2). In beiden Wellen wurde jeweils geprüft, ob die Angaben über Berufstätigkeiten im Erwerbstätigkeitstableau mit den entsprechenden Angaben im Kalendarium übereinstimmten (Fragen 5.2 und 4.7 in Welle 1 sowie Fragen 4.3 und 1.7 in Welle 2). Festgestellte Inkonsistenzen wurden – falls möglich – durch den Abgleich mit anderen Nennungen im Fragebogen aufgelöst oder andernfalls ein entsprechender Missingcode (vgl. Kapitel 6.7) vergeben.

Neben diesen Konsistenzprüfungen von Angaben der gleichen Welle wurde auch die Konsistenz von Angaben über die Wellen hinweg geprüft. Besonders hervorzuheben ist hier der Abgleich von Angaben zwischen den Erwerbstätigkeitstableaus bzw. Kalendarien beider Wellen. Sowohl für die Kalendarien als auch für die Erwerbstätigkeitstableaus waren Überschneidungen in den Tätigkeitsangaben zwischen den beiden Wellen bewusst konzipiert worden.⁶¹ Sofern es keine Widersprüche zwischen den Angaben gab, wurden die sich überschneidenden Monate im Fragebogen der zweiten Welle gestrichen. Bei widersprüchlichen Angaben wurden auf Einzelfallebene Anpassungen vorgenommen bzw. Missingcodes (vgl. Kapitel 6.7) vergeben.

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

Zu diesem Zweck wurden die erfassten Fragebogen-Daten in eine Datenbank eingelesen. Anschließend wurden anhand formaler Regeln gültige Wertebereiche und Antwortkombinationen definiert und geprüft. Folgende Typen von Prüfungen wurden vorgenommen:

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

⁶⁰ Die Zahl der vorgenommenen Korrekturen wurde nicht zentral, sondern nur auf den Papierfragebögen dokumentiert und ist daher nicht mehr systematisch rekonstruierbar.

⁶¹ Im Rahmen der ersten Welle wurden sowohl im Erwerbstätigkeitstableau als auch im Kalendarium die jeweiligen Tätigkeiten seit dem Studienabschluss bis zum Befragungszeitpunkt erfasst. Dieser lag – je nachdem, wann die Befragten den Fragebogen ausfüllten – zwischen Januar 2006 und Mai 2007 (vgl. Kapitel 4). In der zweiten Welle setzte das Kalendarium bzw. das Erwerbstätigkeitstableau Anfang Januar 2006 an, sodass Tätigkeiten unter Umständen mehrfach erfasst wurden.

- *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).⁶²
- *Prüfung von Merkmalskombinationen:* Es wurde die Konsistenz der Angaben innerhalb eines Fragebogens sowie zwischen den Befragungswellen geprüft. Über die Merkmalskombinationen hinaus, die bereits bei der manuellen Vorkorrektur überprüft wurden, konnten hier auch komplexere Merkmalskombinationen überprüft werden.

Insgesamt wurden mehrere hundert Konsistenzregeln definiert und getestet. Bei fehlenden, fehlerhaften oder unplausiblen Werten wurde zunächst mit Hilfe des Papierfragebogens geprüft, ob der entsprechende Wert falsch (bzw. nicht) übertragen worden war. Ansonsten wurde versucht, den korrekten Wert anhand anderer Angaben im Fragebogen zu erschließen. Im Zweifelsfall wurde ein spezifischer Missingcode vergeben (vgl. Kapitel 6.7). Fehlerkorrekturen wurden dokumentiert⁶³ und von mindestens einer weiteren Person geprüft.

[Löschung von Fällen] In beiden Wellen wurden Fälle aus dem Datensatz entfernt. Ein Fall wurde gelöscht, wenn rund die Hälfte der Fragen oder zentrale Fragen (z. B. zum Studienfach) nicht beantwortet worden waren oder wenn zu viele Inkonsistenzen vorhanden waren. Diese Fälle wurden als nicht auswertbar eingestuft und entfernt. Darüber hinaus wurden in der ersten Welle Fälle identifiziert, die nicht zur Grundgesamtheit gehörten.⁶⁴ Diese wurden ebenfalls aus dem Datensatz entfernt. Insgesamt wurden in der ersten Befragungswelle 326 Fälle und in der zweiten Befragungswelle 4 Fälle gelöscht.⁶⁵

6.4 Generierung von Variablen

Neben den Variablen, die die codierten Antworten der Befragten enthalten, beinhaltet der Datensatz des Absolventenpanels 2005 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

⁶² Die Eingangsfilter der den einzelnen Fragen zugeordneten Variablen sind im Datensatzreport sowie im Metadaten- suchsystem (<https://metadata.fdz.dzhw.eu>) dokumentiert. Sie definieren für die jeweilige Variable, welche Befragten- gruppe die zugehörige Frage beantworten sollte.

⁶³ Die Dokumentation der Fehlerkorrekturen erfolgte handschriftlich auf den Papierfragebögen und ist daher nicht systematisch rekonstruierbar.

⁶⁴ Dies kam beispielsweise vor, wenn die Prüfungsämter versehentlich Personen angeschrieben hatten, die einem anderen Prüfungsjahrgang angehörten oder wenn es sich um Absolvent(inn)en handelte, die bereits einen weiteren Abschluss hatten. 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 2005 liegt. Diese wurden beibehalten, weil es an einigen Hochschulen üblich ist, dass Fälle noch dem vorherigen Prüfungsjahr zugerechnet werden, wenn sich eine Prüfung ohne Zutun der Absolvent(inn)en verschoben hat (z. B. durch Erkrankung des Prüfers). Umgekehrt gibt es auch Fälle, bei denen das Abschlussdatum einige Monate vor dem eigentlichen Prüfungszeitraum des Jahrgangs 2005 liegt, die Fälle aber von den Hochschulen trotzdem zum Prüfungsjahrgang 2005 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.

⁶⁵ Dies entspricht in der ersten Welle 2,7 Prozent bzw. in der zweiten Welle 0,1 Prozent der Fälle. Es ist nicht rekonstruierbar, wie viele dieser Ausfälle stichprobenneutral und wie viele Ausfälle stichprobenrelevant sind.

Quellvariablen generiert (z. B. Aggregation der Studienfächer zu Studienbereichen und Fächergruppen oder Ableitung von Hochschultyp und Hochschulort aus den Hochschulvariablen). Der Variablenname einer generierten Variablen ist im Datensatz durch das Suffix „_g#“ gekennzeichnet. Eine Übersicht aller für das Absolventenpanel 2005 generierten Variablen sowie eine detaillierte Dokumentation der einzelnen Variablen mit Angabe ihrer jeweiligen Ausprägungen und Berechnungsvorschriften findet sich im Datensatzreport sowie im Metadaten-System⁶⁶.

Generierte Variablen wurden im Datensatz – sofern möglich – nach der jeweiligen Ausgangsvariable positioniert. Wurde eine Variable aus verschiedenen Quellvariablen generiert, wurde sie hinter jene Variable eingefügt, die ihr thematisch am nächsten ist. Falls eine eindeutige Zuordnung nicht möglich war, wurde die generierte Variable am Ende des Datensatzes eingefügt. Generierte Variablen im Zuge von Anonymisierungsmaßnahmen wurden bis auf wenige Ausnahmen durch das FDZ erstellt.

6.5 Erstellung der Datensätze

[Zusammenführung der Wellen] Die Daten der ersten und zweiten Befragungswelle wurden zusammengeführt. Die Zuordnung der Fälle erfolgte über die im Rahmen der Feldphase vergebenen Identifikationsnummern der Befragten (vgl. Kapitel 4).

[Erstellung von Personen- und Episodendatensatz] Die so zusammengeführten Daten wurden in zwei getrennten Datensätzen abgelegt. Der *Personendatensatz* enthält den Großteil der Befragungsdaten sowie die zusätzlich generierten Variablen. Pro befragter Person existiert eine Datenzeile (wide-Format). Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehörigen Fragen im Fragebogen. Der *Episodendatensatz* enthält nur die Antworten aus den Kalendarien (Frage 4.7 der ersten Welle, Frage 1.7 der zweiten Welle). Für jede befragte Person werden ein oder mehrere Episoden gespeichert. Dabei ist eine Episode definiert als ein Zeitraum, in dem eine bestimmte Tätigkeitsart (z. B. Erwerbstätigkeit, Ausbildung) ausgeübt wird bzw. ein konkreter Status (z. B. Elternzeit, Arbeitslosigkeit) besteht. Für jede Episode einer Person existiert jeweils eine Datenzeile (long format). Die Struktur entspricht der gängigen Struktur für Episodendaten (vgl. Scherer & Brüderl, 2010, S. 1042). Die 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.

[Abtrennung der Bachelor-Daten] Aufgrund des angewendeten Stichprobenverfahrens (vgl. Kapitel 3) eignet sich die Stichprobe der Bachelorabsolvent(inn)en nicht, um Aussagen zu treffen, die sich auf die Grundgesamtheit dieser Gruppe beziehen. Aus diesem Grund wurden

⁶⁶ <https://metadata.fdz.dzhw.eu>

die Daten der Bachelorabsolvent(inn)en von denen der Absolvent(inn)en traditioneller Studiengänge sowohl im Personen- als auch im Episodendatensatz abgetrennt und in zwei gesonderten Datensätzen gespeichert.

[Dateiformat] Alle Datensätze werden sowohl im Stata- als auch im SPSS-Format bereitgestellt (vgl. Abschnitt III).

6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels

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

[Variablenbenennung im Personendatensatz] Für die Variablenbenennung wurde im FDZ-DZHW eine einheitliche Benennungssystematik erstellt. Mit Ausnahme der Identifikatorvariablen pid sowie der Wellenvariablen wave⁶⁷ 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 Wellenkennung anhand eines Buchstabens. Im Stamm geht der Themenbereich, dem die Variable zugeordnet ist, aus einem dreistelligen englischen Buchstabenkürzel hervor. Tabelle 4 stellt die verschiedenen Themenbereiche des Absolventenpanels 2005 sowie das zugehörige Kürzel für den Stamm des Variablennamens im Überblick dar. Das anhand eines Unterstriches vom Stamm abgetrennte Suffix enthält verschiedene Zusatzinformationen, wie die Kenntlichmachung von generierten Variablen sowie verschiedenen Datenzugangswegen.

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

Detaillierte Informationen zur Variablenbenennung im Absolventenpanel 2005 befinden sich im Datensatzreport.

⁶⁷ Diese enthält die Information, welche Fälle nur an der ersten Welle bzw. an beiden Wellen teilgenommen haben.

Tabelle 4: Themengebiete und Kürzel für Variablennamen des DZHW-Absolventenpanels 2005

Themengebiete-Kürzel	Themengebiet (englisch)	Themengebiet (deutsch)
stu	studies	Studium
occ	occupation	Beschäftigung
ski	skills	Fähigkeiten
fvt	further vocational training	Berufliche Fort- und Weiterbildung
fec	further education	Aus- und Weiterbildung
dem	demographic information	demographische Informationen
wgt	weights	Gewichtungsvariablen
sys	system variables	Systemvariablen

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

6.7 Codierung fehlender Werte

Zur Codierung fehlender Werte wurde im FDZ-DZHW eine übergreifende Systematik erstellt, um über verschiedene Datensätze des DZHW hinweg eine einheitliche Missingcodierung gewährleisten zu können. Fehlende Angaben wurden dabei durch dreistellige negative Werte codiert. Tabelle 5 stellt die verwendete Missingsystematik dar. Die im Absolventenpanel 2005 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. „nicht gegeben“ bei den Items aocc17a, aocc17b und aocc17c, Frage 4.16, 1. Welle).

Tabelle 5: Systematik des FDZ-DZHW für fehlende Werte

Wertebereich	Code	Wertelabel
-999 bis -990: Nonresponse	-999	weiß nicht
	-998	keine Angabe
	-997	keine Angabe (Antwortkategorie)
	-996	Interviewabbruch
	-995	keine Teilnahme (Panel)
	-994	verweigert
-989 bis -970: Nicht zutreffend	-989	filterbedingt fehlend
	-988	trifft nicht zu
	-987	designbedingt fehlend (Fragebogensplit)
	-986	designbedingt fehlend (Welle) ^a
	-985	designbedingt fehlend (Kohorte) ^b
-969 bis -950: Editierter fehlender Wert	-969	unbekannter fehlender Wert ^c
	-968	unplausibler Wert^d
	-967	anonymisiert
	-966	nicht bestimmbar^e
	-965	ungültige Mehrfachnennung
-949 bis -930: Item-spezifische fehlende Werte^f	-949	nicht gegeben
	-948	läuft noch
-929 bis -920: Andere fehlende Werte	-929	Datenverlust

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

^b Dieser Wert wird nur in gepoolten Datensätzen vergeben.

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

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

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

^f Die Ausprägungen dieser Missingkategorie sind definitionsgemäß für jeden Datensatz spezifisch

.

7 Gewichtung

Die Gewichtung der Daten dient dem Ausgleich von Verzerrungen der Stichprobe im Vergleich zur definierten Grundgesamtheit. Eine Gewichtung wurde nur für die Daten der Absolvent(inn)en traditioneller Studiengänge vorgenommen, für den Datensatz der Bachelorabsolvent(inn)en erfolgte aufgrund seines explorativen Charakters und des entsprechend gewählten Stichprobendesigns (vgl. Kapitel 3) keine Gewichtung. Es folgt zunächst eine allgemeinere Einführung in die Vorgehensweise und die Darstellung der erstellten Gewichte. Im Anschluss wird die Gewichtungsprozedur im Detail beschrieben.

7.1 Vorgehen und Anwendungshinweise

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

- **Designbedingte Verzerrung:** Disproportionalitäten werden bewusst erzeugt, um in bestimmten relevanten Subgruppen die Fallzahlen zu erhöhen (vgl. Kapitel 3).
- **Verzerrung durch Nonresponse:** Ausfallprozesse (z. B. Nichtteilnahmen, fehlende Erreichbarkeit, Verlust auf dem Postweg) führen zu einem verringerten Rücklauf und somit zu einer Differenz zwischen Brutto- und Nettostichprobe (vgl. Kapitel 5). Wenn diese Ausfallsprozesse unsystematisch sind (Missing Completely at Random), können sie ignoriert werden.⁶⁸ Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (Missing at Random, Not Missing at Random), der einer Modellierung bedarf.⁶⁹

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

Vor dem Hintergrund der Erläuterungen zur Grundgesamtheit und zum Stichprobenverfahren in Kapitel 3 wird ersichtlich, dass ein derartiges idealtypisches Vorgehen für die Daten des Absolventenpanels 2005 nicht durchführbar ist. Das Stichprobendesign ermöglicht keine Ableitung von exakten Inklusionswahrscheinlichkeiten. Da in der ersten Befragungswelle zudem keine Informationen über Nichtteilnehmer(innen) vorliegen, können hier auch keine individuellen Nonresponsegewichte gebildet werden. Als Querschnittsgewicht für die erste Welle wird daher ein geschätztes Designgewicht berechnet, das anhand von Informationen aus der Grundgesamtheit kalibriert wird. Aus der Merkmalsverteilung der Grundgesamtheit werden zugleich Informationen zu Nichtteilnehmer(inne)n gewonnen, sodass hiermit ebenfalls eine

⁶⁸ Insofern die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

⁶⁹ Zu den unterschiedlichen Formen von Ausfallprozessen siehe grundlegend Rubin, 1976.

Form der Nonresponse-Adjustierung erfolgt. Für die zweite Welle wird zusätzlich ein Non-responsegewicht berechnet, das mit Hilfe von Informationen aus der ersten Welle die Nichtteilnahme in der zweiten Welle modelliert. Zusammengenommen mit dem Querschnittsgewicht der ersten Welle ergibt sich hieraus das Längsschnittgewicht für das Zwei-Wellen-Panel. In Tabelle 6 sind die erstellten Gewichte dargestellt.

Tabelle 6: Bereitgestellte Gewichte zum DZHW-Absolventenpanel 2005

Variablenname	Beschreibung
wgt_t1d	Querschnittsgewicht Welle 1
wgt_t1t2d	Längsschnittgewicht 2-Wellen-Panel

[Hinweise zur Anwendung der Gewichte] Beide Gewichte wurden nur für den Datensatz der traditionellen Studierenden erstellt. Bei den erstellten Gewichten handelt es sich um probability weights, die in Stata mit Hilfe .ado-spezifischer Optionen berücksichtigt werden können.⁷⁰ Das Gewicht wgt_t1d ist für Auswertungen der ersten Welle vorgesehen, das Gewicht wgt_t1t2d für Auswertungen des Zwei-Wellen-Panels.⁷¹ 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. Im Folgenden wird die Vorgehensweise zur Erstellung der Gewichte näher dargestellt.

7.2 Gewichtung der Daten

[Querschnittsgewichtung] Aufgrund des Stichprobendesigns konnten keine exakten Designgewichte abgeleitet werden. Diese mussten vielmehr geschätzt werden. Die Schätzung des Designgewichts erfolgte in jeder Schicht s wie folgt⁷²:

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

⁷⁰ Siehe hierzu die Stata-Hilfe (Befehl: *help weights*).

⁷¹ Hierbei ist zu beachten, dass auch bei Auswertungen, die sich lediglich auf Variablen der zweiten Welle beziehen, das Längsschnittgewicht anzuwenden ist.

⁷² Wobei n_{sc} der Anzahl der Klumpen in einer Schicht und N_{sc} der Anzahl der Klumpen in der jeweiligen Schicht der Grundgesamtheit entspricht. Da die Klumpen vollständig erhoben wurden, entspricht die Auswahlwahrscheinlichkeit eines Individuums der Auswahlwahrscheinlichkeit des zugehörigen Klumpens.

Aufgrund der fehlenden Informationen über die Nichtteilnehmer(innen) in der ersten Welle, war auch für den Ausfallprozess durch Nichtteilnahmen (Nonresponse) keine umfängliche Adjustierung der geschätzten Designgewichte auf Individualebene möglich. Jedoch wurde eine Kalibrierung vorgenommen, welche die Anpassung der geschätzten Designgewichte an Merkmalsträger der Grundgesamtheit zum Ziel hatte. Als Merkmale wurden das Bundesland, das Geschlecht, die Abschlussart und die Fächergruppen herangezogen.⁷³ 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 geschätzten Designgewichts \widehat{dwgt}_{sci} wurde mittels des Raking-Algorithmus⁷⁴ durchgeführt. Daraus ergibt sich das Querschnittsgewicht der ersten Welle des Absolventenpanels 2005 $wgt_{i_{t1}}$. Für die zweite Befragungswelle wurde kein Querschnittsgewicht erstellt, da weder neu hinzugekommene Individuen befragt wurden (Auffrischungsstichprobe) noch Personen die Möglichkeit hatten, an Welle 2 teilzunehmen, wenn sie nicht Teil der ersten Befragungswelle waren.

[Längsschnittgewichtung] Für die Gewichtung des Zwei-Wellen-Panels mussten die Ausfallprozesse im Zeitverlauf (vgl. Kapitel 5) berücksichtigt werden. Zu diesem Zweck wurde ein Ausfallgewicht errechnet, das die Wahrscheinlichkeit abbildete, an der zweiten Welle teilzunehmen. Im Unterschied zu der Nonresponse-Adjustierung in der ersten Welle lag für die Nichtteilnehmer(innen) der zweiten Welle eine Vielzahl an Informationen aus der vorangegangenen ersten Welle vor. Diese 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.⁷⁵ Aus dem Modell konnte die bedingte Teilnahmewahrscheinlichkeit abgeleitet werden, deren Kehrwert das Ausfallgewicht für die zweite Welle darstellt:⁷⁶

$$NR_{i_{gewt2}} = P(Res_{t2} | \sigma_{t1})^{-1}$$

Das Gesamtgewicht für das Zwei-Wellen-Panels des Datensatzes ergibt sich aus dem Produkt des geschätzten Designgewichts (\widehat{dwgt}_{sci}) mit dem Nonresponsegewicht ($NR_{i_{gewt2}}$):

$$wgt_{i_{t1t2}} = \widehat{dwgt}_{sci} \times NR_{i_{gewt2}}$$

⁷³ Die Informationen aus der Grundgesamtheit wurden aus Daten des Statistischen Bundesamtes (Prüfungstatistik 2005) abgeleitet.

⁷⁴ Raking ist auch unter dem Begriff des iterative proportional fitting (ipf) bekannt (vgl. Kolenikov, 2014).

⁷⁵ Das komplette Modell ist im Anhang 1 dargestellt, statistisch nicht bedeutsame Prädiktoren wurden aus dem Modell entfernt.

⁷⁶ Das Verfahren entspricht seiner Logik nach dem *Propensity Score Matching*, das auf Rosenbaum und Rubin, 1983 zurückgeht (vgl. Blumenstiel & Gummer, 2015).

Im Anschluss wurde das nonresponse-adjustierte Längsschnittgewicht mittels des Raking-Algorithmus auf Merkmale der Grundgesamtheit kalibriert.⁷⁷

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

[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 trunziert. Der Überschuss jenseits der Trunkierung wird im Folgenden unter den verbleibenden Gewichten verteilt.

⁷⁷ Hierfür wurden dieselben Merkmale verwendet, welche auch für die Kalibrierung des geschätzten Designgewichtes in der ersten Erhebungswelle eingesetzt wurden.

8 Anonymisierung

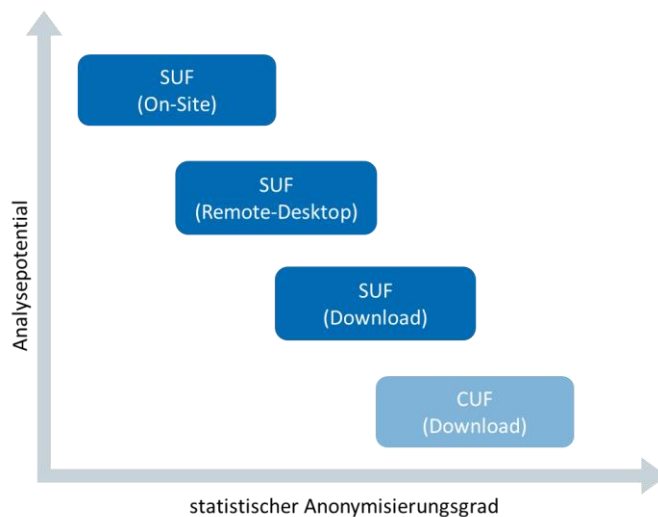
[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten, die in freiwilligen Befragungen durch das DZHW erhoben werden, gilt das Bundesdatenschutzgesetz (BDSG).⁷⁸ Danach dürfen personenbezogene Daten, die im Rahmen wissenschaftlicher Forschung erhoben worden sind, ausschließlich zum Zweck wissenschaftlicher Forschung verarbeitet oder genutzt werden (vgl. §40 Abs. 1 BDSG). Darüber hinaus sind personenbezogene Daten im Forschungskontext zum Schutz der Befragten zu anonymisieren (vgl. §40 Abs. 2 BDSG). Im BDSG wird der Vorgang der Anonymisierung definiert als „das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbar natürlichen Person zugeordnet werden können.“ (§3 Abs. 6 BDSG). Das heißt, für die Weitergabe von Daten aus wissenschaftlichen Forschungsprojekten an Dritte sind die Daten entweder absolut zu anonymisieren, sodass kein Bezug zur Person mehr hergestellt werden kann, oder mindestens faktisch zu anonymisieren, sodass die Herstellung eines Personenbezugs mit einem unverhältnismäßig hohen Aufwand verbunden und daher die Wahrscheinlichkeit der Re-Identifikation einer Person minimal ist.

[Datenzugang, Anonymisierungsgrad und Analysepotential] Das FDZ-DZHW stellt für das Absolventenpanel 2005 ein faktisch anonymisiertes SUF für die wissenschaftliche Sekundärnutzung und ein absolut anonymisiertes CUF für Lehr- und Übungszwecke zur Verfügung. Die Anonymität der Befragten wird dabei über eine Kombination aus statistischen Maßnahmen und technischen Zugriffsbeschränkungen sichergestellt. Je stärker der Datenzugang technisch kontrolliert wird, desto geringer ist das Risiko einer De-Anonymisierung der Daten, desto weniger müssen die Daten mittels statistischer Maßnahmen um Informationen reduziert werden und desto größer bleibt ihr Analysepotential.

Während das CUF nach einer Registrierung direkt durch das FDZ-DZHW übermittelt wird, wird das SUF über drei verschiedene Zugangswege angeboten: Download, Remote-Desktop und On-Site (für weiterführende Informationen vgl. Abschnitt III). Für jeden Zugangsweg wird eine andere SUF-Variante bereitgestellt, die unterschiedlich stark anonymisiert worden ist und entsprechend weniger oder mehr Informationen umfasst. Abbildung 6 gibt einen Überblick über den jeweiligen Grad der statistischen Anonymisierung und dem damit verbundenen Analysepotential. Im Folgenden werden die durchgeführten statistischen Anonymisierungsmaßnahmen in Abhängigkeit von Datenprodukt (SUF/CUF) und Zugangsweg erläutert.

⁷⁸ Das BDSG kommt zur Anwendung, da die DZHW GmbH juristisch als öffentliche Stelle des Bundes betrachtet wird (vgl. § 2 Abs. 3 BDSG). Der Bund hält die absolute Mehrheit der Anteile der DZHW GmbH und das Institut erfüllt Aufgaben der öffentlichen Verwaltung des Bundes im weitesten Sinn. Zur Auslegung einzelner rechtlicher Aspekte wird ergänzend die Europäische Datenschutzrichtlinie hinzugezogen.

Abbildung 6: Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten des DZHW-Absolventenpanels 2005



[Statistische Anonymisierungsmaßnahmen] Im Rahmen der Anonymisierung sind zunächst alle Informationen, mit denen sich Personen oder Institutionen direkt identifizieren lassen, zu löschen. Diese sogenannten *direkten Identifikatoren*, wie Namen, Adressen und E-Mail Adressen, wurden im Absolventenpanel 2005 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.⁷⁹ Für das Absolventenpanel 2005 wurden die folgenden Quasi-Identifikatoren identifiziert, die sowohl in externen Datenquellen⁸⁰ als auch in den Absolvent(inn)endaten vorliegen: Hochschule, Studienfach, Abschlussart, Berufsangaben, regionale Informationen (zur Hochschule, zum Ort des Erwerbs der Studienberechtigung und Arbeitsort), Staatsangehörigkeit und Geburtsland. Um eine eindeutige Zuordnung der Absolvent(inn)endaten zu unterbinden, wurden diese Schlüsselmerkmale – je nach Datenprodukt bzw. Zugangsweg – aggregiert oder gelöscht (vgl. Tabelle 7). 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.

⁷⁹ 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.

⁸⁰ z.B. Studenten- und Prüfungsstatistik des statistischen Bundesamtes, Alumninetzwerke der Hochschulen oder auch Berufsnetzwerke.

Zuletzt wurde geprüft, ob in den Daten *sensible Informationen*, z. B. zur Gesundheit, sexuellen Orientierung und zu politischen Einstellungen, enthalten waren. Diese eignen sich zwar nicht zur Re-Identifikation von Individuen oder Institutionen, jedoch können die Informationen im Falle einer De-Anonymisierung nutzbringend sein (vgl. Koberg, 2016, S. 694) und sind daher besonders schützenswert (vgl. §3 Abs. 9 BDSG, Art. 8 Abs. 1 und 2a EG-DSRL). Im Absolventenpanel 2005 wurden Gesundheitsinformationen erhoben, für die bei den Befragten kein zusätzliches Einverständnis für die Sekundärnutzung eingeholt wurde. Daher wurden diese Antworten im CUF- und allen SUF-Varianten gelöscht.

Zur Gewährleistung der absoluten Anonymisierung der Daten des CUF wurden zum einen im Vergleich zu den SUF-Varianten restriktivere statistische Anonymisierungsmaßnahmen auf Variablenebene vorgenommen (vgl. Tabelle 7). Zum anderen wurde eine per Zufallsauswahl gewonnene Substichprobe der Daten (10 Prozent der befragten Absolvent(inn)en traditioneller Studiengänge) gezogen.

Tabelle 7: Maßnahmen der statistischen Anonymisierung der Daten des DZHW-Absolventenpanels 2005 nach Zugangsweg⁸¹

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

⁸¹ Detaillierte Informationen zu den anonymisierten Variablen sind dem Datensatzreport sowie dem Metadaten-suchsystem (<https://metadata.fdz.dzhw.eu>) zu entnehmen.

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

^a Nach dem Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2004/2005 und SoSe 2005 von Destatis.

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

^c Nach Klassifikation der Berufe von 1992 bzw. 2010 von Destatis.

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

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

9 Literaturverzeichnis

- Blumenstiel, J. E. & Gummer, T. (2015). Prävention, Korrektur oder beides? Drei Wege zur Reduzierung von Non-response Bias mit Propensity Scores. In J. Schupp & C. Wolf (Hrsg.), *Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen* (S. 13–44). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-10459-7
- Briedis, K. (2007). *Übergänge und Erfahrungen nach dem Hochschulabschluss. Ergebnisse der HIS-Absolventenbefragung 2005* (HIS: Forum Hochschule 13/2007). Hannover: HIS.
- Briedis, K. & Minks, K.-H. (2007). *Generation Praktikum. Mythos oder Massenphänomen*. Hannover: HIS.
- Cox, L. H. (1987). A constructive procedure for unbiased controlled rounding. *Journal of the American Statistical Association*, 82 (398), 520–524.
- 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.
- Grotheer, M., Isleib, S., Netz, N. & Briedis, K. (2012). *Hochqualifiziert und gefragt. Ergebnisse der zweiten HIS-HF Absolventenbefragung des Jahrgangs 2005* (HIS: Forum Hochschule 14/2012). Hannover: HIS.
- Häder, M. (2015). *Empirische Sozialforschung. Eine Einführung* (3. Aufl.). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-531-19675-6
- Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. (2012). *FDZ-Methodenreport. Datenschutz am Forschungsdatenzentrum* (Nr. 06). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Jaksztat, S. (2014). Bildungsherkunft und Promotionen: Wie beeinflusst das elterliche Bildungsniveau den Übergang in die Promotionsphase? *Zeitschrift für Soziologie*, 43 (4), 286–301.
- 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.
- Kratz, F. & Netz, N. (2016). Which mechanisms explain monetary returns to international student mobility? *Studies in Higher Education*. doi:10.1080/03075079.2016.1172307
- Lane, J., Heus, P. & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. *Transactions on Data Privacy*, 1 (1), 2–16.
- Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. *Proceedings of the Survey Research Methods Section*, 225–230.
- 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.
- Schaeper, H. (2009). Development of competencies and teaching–learning arrangements in higher education: findings from Germany. *Studies in Higher Education*, 34 (6), 677–697. doi:10.1080/03075070802669207
- Schaeper, H. & Briedis, K. (2004). *Kompetenzen von Hochschulabsolventinnen und Hochschulabsolventen, berufliche Anforderungen und Folgerungen für die Hochschulreform*. Hannover: HIS.
- Schaeper, H., Grotheer, M. & Brandt, G. (2014). Familiengründung von Hochschulabsolventinnen. Eine empirische Untersuchung verschiedener Examenkohorten. In D. Konietzka & M. Kreyenfeld (Hrsg.), *Ein Leben ohne Kinder* (2. Aufl., S. 47–80). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-94149-3_2
- Scherer, S. & Brüderl, J. (2010). Sequenzdatenanalyse. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 1031–1051). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schnell, R., Hill, P. B. & Esser, E. (2005). *Methoden der empirischen Sozialforschung* (7. Aufl.). München: Oldenbourg.
- Valliant, R., Dever, J. A. & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York (NY): Springer New York. doi:10.1007/978-1-4614-6449-5

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

Variable (AV: Teilnahme Welle 2)	b-Koeffizient	Z-Wert
Fragebogeneingang nach erstem Fragebogeneingang (Tage)	-0,003	0,0007
Fragebogeneingang nach erstem Fragebogeneingang ^2	0,000	0,0000
Letzte Berufliche Stellung (Referenz: leitende Angestellte)		
wissenschaftlich Angestellter	0,145	0,0908
Angestellter	0,152	0,0940
Selbstständig	0,003	0,2193
Beamter	0,543	0,2360
Sonstige	0,171	0,1396
fehlender Wert	0,521	0,1887
Geschlecht (Referenz: männlich)		
weiblich	0,063	0,0299
Alter (Referenz: 20 bis 22 Jahre)		
23 bis 27 Jahre	-0,084	0,2472
28 bis 29 Jahre	-0,105	0,2487
30 bis 32 Jahre	-0,214	0,2506
33 bis 35 Jahre	-0,154	0,2578
36 bis 39 Jahre	0,136	0,2674
40 bis 44 Jahre	-0,120	0,2742
45 bis 67 Jahre	0,091	0,3074
fehlender Wert	-0,346	0,3664
CASMIN (Eltern/Dominanzmodell) (Referenz: niedrig)		
mittel	0,189	0,0840
hoch	0,196	0,0840
fehlender Wert	-0,138	0,1623
Abschlussnote (gerundet) (Referenz: 1,0 bis 1,9)		
Note 2,0 bis 2,9	-0,085	0,0328
Note 3,0 bis 3,9	-0,101	0,0466
Note 4,0 bis 4,9	-0,226	0,1631
Juristen	-0,441	0,3608
fehlender Wert	-0,197	0,0803
Zufriedenheit mit Beschäftigung (Referenz: in hohem Maße)		
2	-0,020	0,0406
3	-0,063	0,0445
4	-0,100	0,0567
gar nicht	-0,053	0,0677
fehlender Wert	-0,050	0,0960
Zeitarbeit/Leiharbeit (Referenz: ja)		
nein	0,133	0,0566
fehlender Wert	0,216	0,1084
Anzahl der Bewerbungen fürs Praktikum (Referenz: keine Bewerbungen)		
1 bis 3	-0,063	0,0934
4 bis 7	-0,043	0,0978
8 bis 15	-0,016	0,0965
16 bis 30	-0,081	0,0970
31 bis 99	-0,094	0,0980
mehr als 100	-0,107	0,1234
fehlender Wert	-0,194	0,1008
Beginn Stellensuche (Referenz: noch nicht begonnen)		
nach Examen	-0,038	0,0725
während Examen	0,022	0,0724
vor Examen	0,029	0,0755
fehlender Wert	-0,247	0,1000

Variable (AV: Teilnahme Welle 2)	b-Koeffizient	Z-Wert
Abschlussart (letzter Abschluss) (Referenz: Diplom)		
Magister	0,028	0,0779
Staatsexamen	0,660	0,3643
Lehramt	0,003	0,0844
sonstige	0,160	0,2309
Art der letzten Tätigkeit (Referenz: unbefristet)		
befristet (Zeitvertrag)	0,015	0,0381
befristet (über ABM)	-0,513	0,2960
Ausbildungsverhältnis	0,075	0,0733
Honorar-/Werkvertrag	0,098	0,2163
selbständig/freiberuflich	0,112	0,2106
Sonstiges	-0,156	0,1173
fehlender Wert	-0,364	0,1670
Bundesland Hochschule (Referenz: Schleswig-Holstein)		
Hamburg	0,182	0,1223
Niedersachsen	0,042	0,0974
Bremen	0,124	0,1481
Nordrhein-Westfalen	-0,024	0,0964
Hessen	0,098	0,1116
Rheinland-Pfalz	-0,058	0,1139
Baden-Württemberg	-0,009	0,0963
Bayern	0,031	0,0958
Saarland	0,454	0,2846
Berlin	0,073	0,1129
Brandenburg	0,086	0,1198
Mecklenburg-Vorpommern	-0,085	0,1175
Sachsen	0,021	0,0987
Sachsen-Anhalt	-0,035	0,1124
Thüringen	0,195	0,1099
Einkommen der aktuellen Tätigkeit (Referenz: 1 bis 749 Euro)		
750 bis 999 Euro	-0,077	0,0737
1000 bis 1249 Euro	-0,018	0,0685
1250 bis 1499 Euro	-0,054	0,0821
1500 bis 1749 Euro	0,066	0,0750
1750 bis 1999 Euro	0,057	0,0875
2000 bis 2249 Euro	0,058	0,0740
2250 bis 2499 Euro	0,072	0,0834
2500 bis 2749 Euro	-0,020	0,0758
2750 bis 2999 Euro	-0,009	0,0832
3000 bis 3249 Euro	0,041	0,0746
3250 bis 3499 Euro	0,073	0,0937
3500 bis 3749 Euro	0,084	0,0987
3750 bis 3999 Euro	0,219	0,1414
4000 bis 4499 Euro	0,058	0,1194
4500 bis 5000 Euro	0,147	0,1506
über 5000 Euro	-0,005	0,2980
keine Angabe/noch keine Stelle gehabt	-0,195	0,0678
Studienbereich Hochschulabschluss (Referenz: Sprach- und Kulturwissenschaften allgemein)		
Evang. Theologie, -Religionslehre	-0,010	0,3023
Kath. Theologie, -Religionslehre	-0,238	0,2660
Philosophie	-0,701	0,2884
Geschichte	-0,113	0,2204
Bibliothekswissenschaft, Dokumentation, Publizistik	-0,414	0,2043
Allgemeine und vergleichende Literatur- und Sprachwissenschaft	-0,510	0,2235

Variable (AV: Teilnahme Welle 2)	b-Koeffizient	Z-Wert
Altphilologie (klass. Philologie), Neugriechisch	-0,498	0,4824
Germanistik (Deutsch, germanische Sprachen ohne Anglistik)	-0,296	0,1969
Anglistik, Amerikanistik	-0,444	0,2074
Romanistik	-0,315	0,2496
Slawistik, Baltistik, Finno-Ugristik	-1,295	0,6864
Außereuropäische Sprach- und Kulturwissenschaften	-1,308	0,5011
Kulturwissenschaften i.e.S.	-0,236	0,3757
Psychologie	-0,244	0,2021
Erziehungswissenschaften	-0,271	0,1979
Sonderpädagogik	-0,104	0,2166
Sport, Sportwissenschaft	-0,674	0,2269
Wirtschafts- und Gesellschaftslehre allgemein	-0,702	0,6525
Politikwissenschaften	-0,148	0,2198
Sozialwissenschaften	-0,010	0,2171
Sozialwesen	-0,372	0,1922
Rechtswissenschaft	-0,771	0,2554
Verwaltungswissenschaft	-0,237	0,2341
Wirtschaftswissenschaften	-0,466	0,1874
Wirtschaftsingenieurwesen	-0,377	0,1994
Mathematik, Naturwissenschaften allgemein	-0,365	0,5384
Mathematik	-0,242	0,2047
Informatik	-0,473	0,1947
Physik, Astronomie	-0,081	0,2350
Chemie	-0,173	0,2113
Pharmazie	-1,030	0,4297
Biologie	-0,374	0,2014
Geowissenschaften (ohne Geographie)	-0,560	0,2706
Geographie	-0,292	0,2092
Gesundheitswissenschaften allgemein	-0,295	0,2849
Humanmedizin	-0,967	0,4148
Zahnmedizin	-1,139	0,4307
Veterinärmedizin	-0,972	0,4177
Landespflege, Umweltgestaltung	-0,233	0,2086
Agrarwissenschaften, Lebensmittel- und Getränketechnologie	-0,415	0,2072
Forstwissenschaft, Holzwirtschaft	-0,412	0,2780
Ernährungs- und Haushaltswissenschaften	-0,150	0,2349
Ingenieurwesen allgemein	-0,433	0,3131
Bergbau, Hüttenwesen	-0,821	0,4443
Maschinenbau, Verfahrenstechnik	-0,431	0,1919
Elektrotechnik	-0,365	0,2001
Verkehrstechnik, Nautik	-0,277	0,2304
Architektur, Innenarchitektur	-0,573	0,1946
Raumplanung	-0,093	0,2738
Bauingenieurwesen	-0,382	0,2002
Vermessungswesen	-0,238	0,2229
Kunst, Kunstwissenschaft allgemein	-0,251	0,2350
Bildende Kunst	-0,549	0,3008
Gestaltung	-0,430	0,2119
Darstellende Kunst, Film und Fernsehen, Theaterwissenschaften	-0,194	0,3574
Musik, Musikwissenschaft	-0,615	0,2237
Konstante	0,447	0,3751
Anzahl der Beobachtungen	10.159	