

Florence Baillet / Andreas Franken / Anne Weber

DZHW-Absolventenpanel 2005

Daten- und Methodenbericht zu den Erhebungen der
Absolvent(inn)enkohorte 2005 (1., 2. und 3. Befragungswelle)

Daten- und Methodenbericht

Januar 2021

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
Karen Schlüter

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	7
2 Erhebungsinstrumente	10
2.1 Inhalte der Erhebungsinstrumente	10
2.2 Pretests	12
3 Grundgesamtheit und Stichprobenverfahren	14
4 Durchführung der Erhebungen	16
5 Rücklauf	18
6 Datenaufbereitung	21
6.1 Datenübertragung	21
6.2 Datenprüfung und Datenbereinigung	21
6.3 Generierung von Variablen	22
6.4 Erstellung der Datensätze	23
6.5 Vergabe von Variablenamen, Variablenlabels und Wertelabels	24
6.6 Codierung fehlender Werte	26
7 Gewichtung	28
7.1 Vorgehen und Anwendungshinweise	28
7.2 Gewichtung der Daten	29
8 Anonymisierung	32
9 Literaturverzeichnis	36
Anhang 1: Generierte Variablen im DZHW-Absolventenpanel 2005	37

Abbildungsverzeichnis

Abbildung 1: Datenzugangswege und Analysepotential.....	5
Abbildung 2: Kohorten-Panel-Design der DZHW-Absolventenstudienreihe.....	7
Abbildung 3: Kalendarium des DZHW-Absolventenpanels 2005, 1. Welle	11
Abbildung 4: Rücklauf des DZHW-Absolventenpanels 2005 im Zeitverlauf, 1. Welle	18
Abbildung 5: Rücklauf des DZHW-Absolventenpanels 2005 im Zeitverlauf, 2. Welle	19
Abbildung 6: Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten des DZHW-Absolventenpanels 2005	33

Tabellenverzeichnis

Tabelle 1: Anlage der Hauptuntersuchungen der DZHW-Absolventenstudienreihe von 1989 bis 2005.....	8
Tabelle 2: Brutto-, Nettostichproben und Rücklaufquoten des DZHW-Absolventenpanels 2005	20
Tabelle 3: Themengebiete und Kürzel für Variablennamen des DZHW-Absolventenpanels 2005	25
Tabelle 4: Systematik des FDZ-DZHW für fehlende Werte.....	27
Tabelle 5: Bereitgestellte Gewichte zum DZHW-Absolventenpanel 2005	29
Tabelle 6: Maßnahmen der statistischen Anonymisierung der Daten des DZHW-Absolventenpanels 2005 nach Zugangsweg	34

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 Forschungsdatenzentrum 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.

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation zur ersten, zweiten und dritten Befragungswelle des Absolventenpanels 2005 (doi: 10.21249/DZHW:gra2005:2.0.1). Weitere Dokumentationsmaterialien zur Studie (Datensatz-reports, 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 2005⁴ 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 (www.dzhw.eu/absolventen) abgerufen werden.

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

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

⁴ 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)
Gefördert von	Bundesministerium für Bildung und Forschung (BMBF)
Projektmitarbeiter(innen) (Projektleitung)	<u>Kolja Briedis</u> , Thorsten Euler, Michael Grotheer, Sören Isleib, <u>Karl-Heinz Minks</u> , Nicolai Netz, <u>Hildegard Schaeper</u> , Fabian Trennt, Maximilian Trommer
Themen	Studienverlauf Übergang in den Beruf Beruflicher Werdegang Weiterqualifizierung
Erhebungsdesign	Kohorten-Panel-Design
Erhebungsdatentyp	Quantitative Daten
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)
Stichprobe	Absolvent(inn)en traditioneller Studiengänge: quotierte geschichtete Klumpenstichprobe Absolvent(inn)en aus Bachelor-Studiengängen: bewusste Auswahl
Erhebungsmethode	1. Welle: Standardisierte postalische Befragung 2. Welle: Standardisierte postalische Befragung 3. Welle (Hauptbefragung): Standardisierte Onlinebefragung
Feldzeit	1. Welle: 1. Januar 2006 bis 18. Mai 2007 2. Welle: 6. Dezember 2010 bis 21. Februar 2012 3. Welle (Hauptbefragung): 14. April 2016 bis 07. Juni 2016
Nettostichprobe	1. Welle: n = 11.787 (davon 1.622 Bachelor-Absolvent(inn)en) 2. Welle: n = 6.472 (davon 798 Bachelor-Absolvent(inn)en) 3. Welle (Hauptbefragung): n = 4.279 (davon 519 Bachelor-Absolvent(inn)en)

Rücklaufquote	1. Welle: 24,7 % 2. Welle: 60,5 % 3. Welle (Hauptbefragung): 66,3 %
Datenprodukt und Zugangsweg	CUF: Download SUF: Download, Remote-Desktop, On-Site
Datensatzstruktur	Personendatensätze im wide-Format Episodendatensätze im long-Format
DOI	10.21249/DZHW:gra2005:2.0.1
Anmerkungen	Getrennte Datensätze für Absolvent(inn)en traditioneller Studiengänge und Bachelorabsolvent(inn)en wegen unterschiedlicher Stichprobenziehung Der CUF enthält lediglich die Daten der Absolvent(inn)en traditioneller Studiengänge.
Weitere Informationen	https://metadata.fdz.dzhw.eu/#!/de/studies/stu-gra2005 www.dzhw.eu/absolventen2005 https://fdz.dzhw.eu
Projektpublikationen*	
Briedis, K. (2007). <i>Übergänge und Erfahrungen nach dem Hochschulabschluss. Ergebnisse der HIS-Absolventenbefragung 2005</i> (HIS: Forum Hochschule 13/2007). Hannover: HIS.	
Briedis, K. & Minks, K.-H. (2007). <i>Generation Praktikum. Mythos oder Massenphänomen</i> . Hannover: HIS.	
Grotheer, M., Isleib, S., Netz, N. & Briedis, K. (2012). <i>Hochqualifiziert und gefragt. Ergebnisse der zweiten HIS-HF Absolventenbefragung des Jahrgangs 2005</i> (HIS: Forum Hochschule 14/2012). Hannover: HIS.	
Euler, T., Trennt, F., Trommer, M. & Schaeper, H. (2018). <i>Werdegänge der Hochschulabsolventinnen und Hochschulabsolventen 2005. Dritte Befragung des Prüfungsjahrgangs 2005 zehn Jahre nach dem Abschluss</i> . (Forum Hochschule 01/2018). Hannover: DZHW.	
* Alle Projektpublikationen werden auf der Website des Projektes (www.dzhw.eu/absolventen2005) zum Download bereitgestellt.	
Publikationen zum Datensatz (Auswahl)	
Schaeper, H. (2009). Development of competencies and teaching–learning arrangements in higher education: findings from Germany. <i>Studies in Higher Education</i> , 34 (6), 677–697. doi:10.1080/03075070802669207	
Jaksztat, S. (2014). Bildungsherkunft und Promotionen: Wie beeinflusst das elterliche Bildungsniveau den Übergang in die Promotionsphase? <i>Zeitschrift für Soziologie</i> , 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.), <i>Ein Leben ohne Kinder</i> (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? <i>Studies in Higher Education</i> . doi:10.1080/03075079.2016.1172307	

III Datennutzungshinweise

[Voraussetzungen der Datennutzung] Die Daten des Absolventenpanels 2005 werden durch das FDZ des DZHW entsprechend der europäischen Datenschutzgrundverordnung (EU-DSGVO) anonymisiert bereitgestellt und ausschließlich zur wissenschaftlichen Nutzung freigegeben.⁵ 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 Anstellung an einer wissenschaftlichen Einrichtung müssen gemeinsam mit einer/einem betreuenden Mitarbeiter(in) einen Datennutzungsvertrag abschließen. 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 ist eine Registrierung beim FDZ nötig. Danach wird das CUF durch das FDZ übermittelt. Ein Datennutzungsvertrag muss nicht abgeschlossen werden.

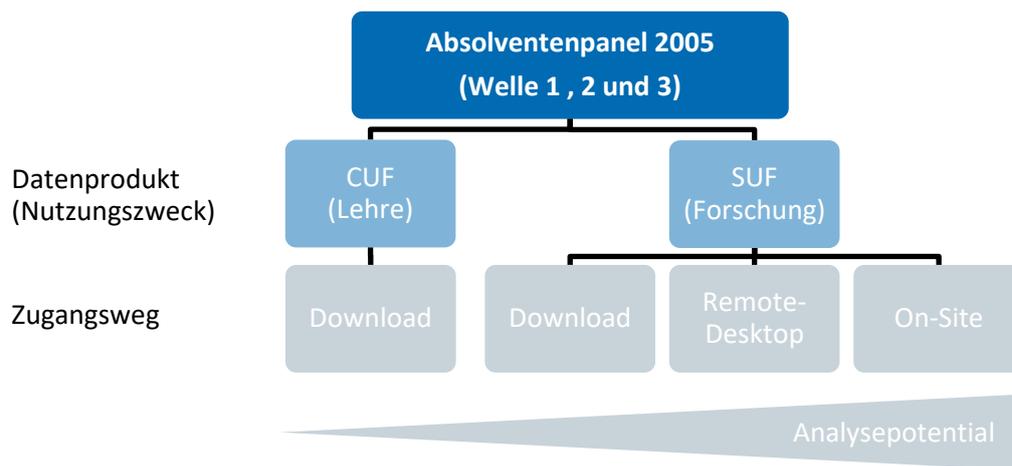
[Datenzugang] Das CUF des Absolventenpanels 2005 kann nach Bereitstellung am lokalen Computer genutzt werden. Das SUF wird über drei Zugangswege angeboten, die hinsichtlich des Speicherortes, der Möglichkeit der eigenständigen Verknüpfung mit externen Daten und der Kontrollmöglichkeiten des FDZ unterschiedlich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

- **Download:** Die Daten werden verschlüsselt 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.

⁵ Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, 6ff.), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (LfBi) (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 des DZHW hat diesen 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.⁶ Mit diesem Vorgehen wird ein Höchstmaß an Nutzbarkeit und gleichzeitig ein bestmöglicher Schutz der bereitgestellten Daten sichergestellt.

Abbildung 1: Datenzugangswege und Analysepotential



[Gebühren der Datenbereitstellung] CUF und SUF werden derzeit (Stand: März 2019) kostenfrei zur Verfügung gestellt. Änderungen bzw. die aktuelle Gebührenordnung können auf der Website des FDZ (<https://fdz.dzhw.eu>) eingesehen werden.

[Pflichten der Datennutzer(innen)] Die Datennutzer(innen) sind verpflichtet, folgende Regeln⁷ 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.

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

⁷ Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

- **Löschungsgebot:** Download-SUF sind nach Ablauf der vereinbarten Nutzungsdauer (in der Regel 1,5 Jahre) von jeglichen Rechnern, Servern und Datenträgern zu löschen. Ebenso müssen alle Sicherungskopien, modifizierten Datensätze (z. B. Arbeits-, Auszugs- oder Hilfsdateien) sowie Ausdrücke vernichtet werden.
- **Bereitstellung/Meldung von Publikationen:** Jede Art von Publikation, die aus der Arbeit mit Daten des FDZ hervorgeht, ist dem FDZ unmittelbar nach Veröffentlichung anzuzeigen. Dabei ist dem FDZ eine elektronische Version der Druckfassung zur Verfügung zu stellen.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen laut der Vorgaben des FDZ zitiert werden.

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 1989⁸ 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.⁹

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	2. Welle				
1993		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

⁸ 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.

⁹ 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. Teilweise setzen sich die zweiten bzw. dritten Wellen aus einer Hauptbefragung und gesonderten Vertiefungsbefragungen zu spezifischen Themen zusammen.

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 unterscheiden die DZHW-Absolventenstudienreihe von anderen in Deutschland durchgeführten Absolventenstudien. So ist beispielsweise das Bayerische Absolventenpanel (BAP) des Bayerischen Staatsinstituts für Hochschulforschung und Hochschulplanung (IHF) auf Absolvent(inn)en bayrischer Hochschulen begrenzt.¹⁰ 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.¹¹

[Spezifika des Absolventenpanels 2005] Neben den allgemeinen Charakteristika der Studienreihe weist die hier betrachtete Absolvent(inn)enkohorte 2005¹² 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 Studienstrukturreform – 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 Gesamtabsolvent(inn)enzahl betrug¹³, 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.¹⁴

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.¹⁵

¹⁰ vgl. dazu <http://www.bap.ihf.bayern.de>

¹¹ vgl. dazu <http://koab.uni-kassel.de>

¹² 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).

¹³ 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).

¹⁴ Aufgrund der unterschiedlichen Ziehungsverfahren für Absolvent(inn)en traditioneller Studiengänge und Bachelorabsolvent(inn)en werden gesonderte Datensätze für die Befragengruppen bereitgestellt.

¹⁵ vgl. dazu auch Briedis & Minks, 2007.

2 Erhebungsinstrumente

In den ersten beiden Befragungswellen des Absolventenpanels 2005 wurde als Erhebungsinstrument ein standardisierter Papierfragebogen in deutscher Sprache eingesetzt. Die dritte Befragungswelle wurde als Online-Erhebung durchgeführt.¹⁶ Kapitel 2.1 stellt die zentralen Inhalte der Erhebungsinstrumente vor. Kapitel 2.2 beschreibt die zur Prüfung und Verbesserung der Fragebögen durchgeführten Pretests.

2.1 Inhalte der Erhebungsinstrumente

[Charakteristika der Studienreihe] Im Fokus des Absolventenpanels 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.¹⁷ 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; Frage 1.7 in Welle 2; Page 9 in Welle 3), in das die Befragten ihre einzelnen Tätigkeiten eintragen sollen (vgl. Abbildung 3).¹⁸ Das Kalendarium wurde vom DZHW konzipiert, um lückenhafte Angaben in der Beschreibung von Tätigkeitsverläufen zu reduzieren.

¹⁶ Die Fragebögen sowie Filterführungsdiagramme aller Befragungswellen können im Metadaten-suchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden.

¹⁷ vgl. Abschnitt 1 „Studienverlauf und Studienerfahrungen“ im Fragebogen der ersten Welle.

¹⁸ 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 auflisten. 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
ST Studium
P Praktikum
B Berufsausbildung, Umschulung, Volontariat
F Fort-/Weiterbildung (Vollzeit, längerfristig)
EZ Elternzeit
H Hausfrau, Hausmann, Familienarbeit
AL Arbeitslosigkeit
SO Sonstiges (z. B. Wehr-/Zivildienst, Urlaub)

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; Frage 4.3 in Welle 2; Page 24-24h in Welle 3), 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.¹⁹ Daneben beinhalten die Erhebungsinstrumente auch Fragen zur subjektiven Einschätzung der eigenen beruflichen Situation, etwa hinsichtlich der beruflichen Zufriedenheit, den beruflichen Zukunftsperspektiven und der Beschäftigungsadäquanz²⁰. Ein Großteil der Informationen zu den Erwerbstätigkeiten wird über die Befragungswellen hinweg wiederholt erhoben.²¹ 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.

¹⁹ Vor der Einführung des Erwerbstätigkeitstableaus wurden – abgesehen von der Art der Institution – generell nur zusätzliche Informationen zur ersten und letzten Erwerbstätigkeit erhoben.

²⁰ vgl. dazu auch Kerst & Fehse, 2007.

²¹ Teilweise wurde allerdings der Wortlaut abgewandelt. Darauf hinzuweisen ist außerdem, dass sich die Wiederholungsmessungen in der zweiten bzw. dritten Welle bei einer beruflichen Veränderung auf eine andere berufliche Situation bzw. Tätigkeit beziehen als in der ersten bzw. zweiten 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).²² 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).²³ 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. Ebenso ist das Erhebungsinstrument der Hauptbefragung der dritten Welle inhaltlich komplett an das Instrument der Kohorte 2001 angelehnt, jedoch in Form eines Online-Fragebogens.

2.2 Pretests

[Ziel und Verfahren] Die Erhebungsinstrumente 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. Viertens zog die Aufnahme neuer Fragen bzw. die veränderte Fragenauswahl auch Veränderungen am Aufbau und Layout des Fragebogens sowie der Befragungsdauer nach sich, die evaluiert werden sollten. In der dritten Befragungswelle musste zudem die Umsetzung der Onlineversion des Fragebogens geprüft werden.

Zur Prüfung dieser verschiedenen Aspekte kam in den ersten beiden Befragungswellen ein sogenanntes *Pretestverfahren im Feld* zum Einsatz. Dieses Verfahren hat zum Ziel, dass die am Pretest teilnehmenden Personen „unter möglichst ähnlichen Bedingungen untersucht [wer-

²² 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).

²³ vgl. dazu auch Briedis & Minks, 2007.

den], wie sie später für die eigentliche Erhebung vorgesehen sind“ (Häder, 2015, S. 396).²⁴ Das Erhebungsinstrument der Hauptbefragung der dritten Befragungswelle wurde im Rahmen von *Expertenbewertungen* (vgl. Häder, 2015, S. 406–407) begutachtet.

[Probanden] In den ersten 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.²⁵ In der dritten Befragungswelle wurde das Erhebungsinstrument von etwa fünf Mitarbeiter(innen)²⁶ des DZHW als Hochschulforschungsexpert(inn)en geprüft.

[Durchführung] Die Pretests der ersten beiden Befragungswellen fanden beide etwa zwei bis drei Monate vor dem jeweiligen Feldstart statt. Dabei wurden die Testpersonen gebeten, den für die Befragung vorgesehenen standardisierten Fragebogen zu bearbeiten und Verständnisprobleme, Kritik oder Anmerkungen zu notieren. Im Anschluss an die Bearbeitung des Fragebogens wurde 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.²⁷ Auf Grundlage der Pretestergebnisse wurden die Formulierungen verschiedener Frage- bzw. Itemtexte präzisiert, die Reihenfolgen von einzelnen Fragen bzw. Items und Antwortkategorien überarbeitet, einzelne Fragen bzw. Items gestrichen oder neu aufgenommen sowie das Layout entsprechend angepasst.²⁸ Der grundsätzliche Aufbau und der Umfang der Fragebögen wurden unverändert beibehalten.

In den Expertenbewertungen der dritten Befragungswelle wurde insbesondere begutachtet, ob die Fragen aus dem etablierten Welle-3-Fragebogen der Kohorte 2001 auch für die Kohorte 2005 inhaltlich passend waren und übernommen werden konnten.

²⁴ 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.

²⁵ Die genaue Anzahl der Teilnehmer(innen) ist nicht mehr rekonstruierbar.

²⁶ Die genaue Anzahl der Teilnehmer(innen) ist nicht mehr rekonstruierbar.

²⁷ Außerdem wurde in der ersten Befragungswelle im Rahmen des Pretests das Anschreiben an die Befragten im Hinblick auf Inhalte, Verständlichkeit und Länge evaluiert.

²⁸ 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 2005²⁹ 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 Studiengänge³⁰ 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.³¹

[Quotierte geschichtete Klumpenstichprobe] Die „primary sampling units“ (Klumpen) der Klumpenstichprobe wurden anhand der Hochschule, dem Studienbereich³² sowie der Abschlussart³³ definiert. Die „secondary sampling units“ stellten die Hochschulabsolvent(inn)en des Prüfungsjahrgangs 2005 innerhalb dieser Klumpen dar.³⁴ Die Umsetzung dieses Designs erfolgte in mehreren Schritten. Im ersten Schritt wurde eine mehrstufige Zufallsstichprobe gezogen.³⁵ 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 Bundesländer³⁶, Größe und Art der Hochschule³⁷. Für diese Schichtung wurde die Zufallsallokation/-rundung nach Cox (1987) verwendet.

Im zweiten Schritt erfolgte eine quotierte Ziehung³⁸ von weiteren 87 Klumpen³⁹, die teilweise die im ersten Schritt gezogenen Klumpen ersetzten und teilweise zusätzlich berücksichtigt wurden. Ziel dabei war, eine möglichst gute Angleichung der Verteilungen von

²⁹ 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.

³⁰ 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.

³¹ Zum Verfahren der bewussten Auswahl vgl. Schnell, Hill & Esser, 2005, 298ff..

³² entsprechend der Aufgliederung nach der amtlichen Statistik (gemäß Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2004/2005 und SoSe 2005).

³³ Hierbei wurde zwischen einem Diplom und vergleichbaren Abschlüssen, etwa einem Magister oder unterschiedlichen Lehramtsabschlüssen, differenziert.

³⁴ Beispiel für den Personenkreis eines Klumpens: Alle Hochschulabsolvent(inn)en der Leibniz-Universität Hannover, die im Studienbereich „Mathematik“ ihr Diplom abgelegt haben.

³⁵ Diese Auswahl wurde durch das „Zentrum für Umfragen, Methoden und Analysen (ZUMA)“ in Mannheim durchgeführt.

³⁶ 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).

³⁷ Es wurde zwischen Universität und Fachhochschule differenziert.

³⁸ Quotierungsmerkmale: Studienbereiche, Abschlussarten und Hochschulen.

³⁹ 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).⁴⁰ Da die Hochschulen die Kontaktdaten ihrer Absolvent(inn)en aus Datenschutzgründen nicht herausgeben durften, teilten sie dem DZHW lediglich die jeweilige Absolvent(inn)enanzahl mit. Anschließend sendete das DZHW die passende Anzahl an Erhebungsunterlagen für die erste Befragungswelle postalisch an die jeweiligen Prüfungsämter, die diese an die Zielpersonen weiterverschickten.

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

[Erhebungsunterlagen] Die Erhebungsunterlagen bestanden in den ersten 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.⁴³ Zudem wurden zwei Erinnerungsschreiben verschickt.

In der Hauptbefragung der dritten Befragungswelle wurde per E-Mail ein Anschreiben mit einem Link zur Online-Befragung, einem individuellen Passwort (Token) und einem Link zum Datenschutzblatt verschickt. Sofern keine valide E-Mail-Adresse vorlag, wurden die Erhebungsunterlagen zusätzlich auch postalisch verschickt. Ergänzend wurden drei Erinnerungsschreiben per E-Mail verschickt.

⁴⁰ An dieser Stelle können durch die Prüfungsämter vereinzelt auch Personen als stichprobenzugehörig identifiziert worden sein, die nicht zur Grundgesamtheit gehörten (Overcoverage), wenn sie beispielsweise einem anderen Prüfungsjahrgang als dem Jahrgang 2005 angehörten (dies tritt z.B. bei Verzögerung der Zeugniserstellung auf) oder es sich nicht um den Erstabschluss handelte.

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

⁴² Die Befragten wurden dabei sowohl im Anschluss an die erste Welle als auch im Vorfeld zur zweiten Welle angeschrieben und um die Aktualisierung ihrer Kontaktinformationen 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, vor Feldstart sowie innerhalb der Erhebungszeit weitere Adressaktualisierungen vorgenommen.

⁴³ In der zweiten Welle bestand außerdem auf Nachfrage 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).

[Feldphase] Der Erhebungszeitraum der ersten Befragungswelle erstreckte sich vom 1. Januar 2006 bis 18. Mai 2007.⁴⁴ 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 der Erhebungsunterlagen nehmen.⁴⁵ 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 dauerte vom 6. Dezember 2010 bis 21. Februar 2012.⁴⁶ Durch den nun im DZHW vorliegenden Adressbestand konnten konkrete Versandzeitpunkte für die Erhebungsunterlagen festgelegt werden.⁴⁷ Zudem konnten die Erinnerungsschreiben gezielt nur an diejenigen Zielpersonen verschickt werden, die sich noch nicht an der Erhebung beteiligt hatten.

Der Erhebungszeitraum der Hauptbefragung der dritten Befragungswelle erstreckte sich vom 14. April 2016 bis 7. Juni 2016. Wie in der zweiten Befragungswelle konnten auch in der dritten Welle konkrete Versandzeitpunkte⁴⁸ bestimmt sowie gezielt Erinnerungen an Nicht-Teilnehmer(inne)n versandt werden.

[Rücklaufsteigernde Maßnahmen] Neben dem für die Rücklaufsteigerung besonders wirksamen Instrument der Erinnerungsschreiben und dem zusammen mit dem Anschreiben verschickten Flyer wurde im Anschreiben der ersten und zweiten Welle 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. In der dritten Welle umfasste die Verlosung ein Tablet PC à etwa 1.000 Euro, zwei Smartphones à etwa 550 Euro sowie fünf Bahn-Reisegutscheine à 100 Euro.

⁴⁴ Die Feldzeit wurde so lange wie möglich – parallel zu den weiteren Projektschritten – ausgedehnt und jeder noch eintreffende Fragebogen bis Mai 2007 berücksichtigt.

⁴⁵ 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.

⁴⁶ Die Feldzeit weicht von der SUF-Version 1.0.0 ab, weil nachträglich noch Nachzügler(innen)-Fälle in den Datensatz aufgenommen wurden.

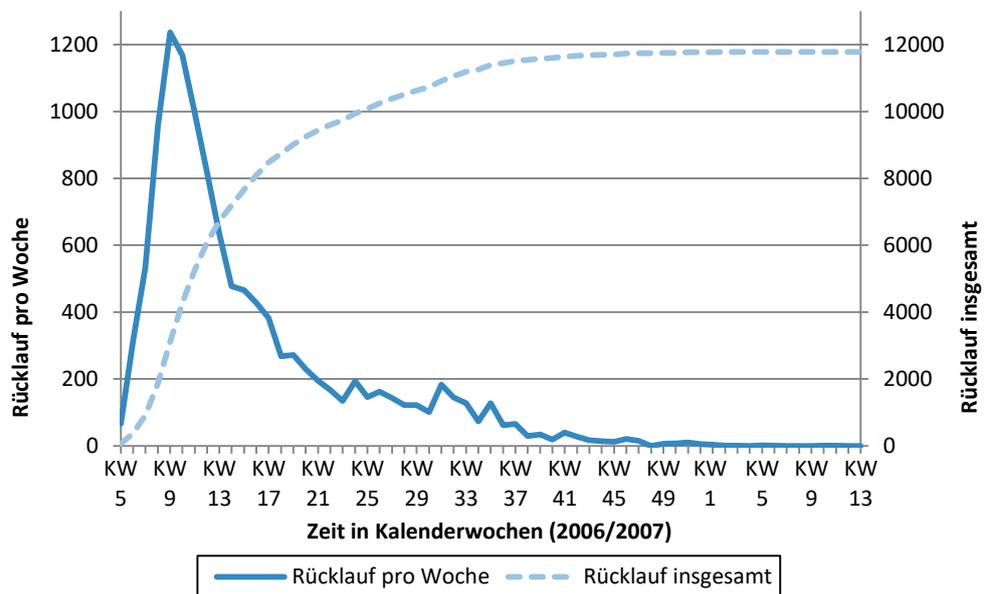
⁴⁷ Einladung: 6.-10.12.2010; Erste Erinnerung: 14.01.2011; Zweite Erinnerung: 11.02.2011.

⁴⁸ Hauptbefragung: Einladung: 14.04.2016 (E-Mail), 21./22.04.2016 (postalisch); Erste Erinnerung: 03.05.2016 (E-Mail); Zweite Erinnerung: 19.05.2016 (E-Mail); Dritte Erinnerung: 31.05.2016 (E-Mail)

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.⁴⁹ Insgesamt wurden 12.114 Fragebögen an das DZHW zurückgeschickt. Von diesen wurden 327⁵⁰ entfernt, da sie nicht zur Grundgesamtheit gehörten (stichprobenneutrale Ausfälle) oder nicht auswertbar waren (stichprobenrelevante Ausfälle) (vgl. Kapitel 6.2). Somit verblieben in der Nettostichprobe der ersten Befragungswelle 11.787⁵⁰ 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.

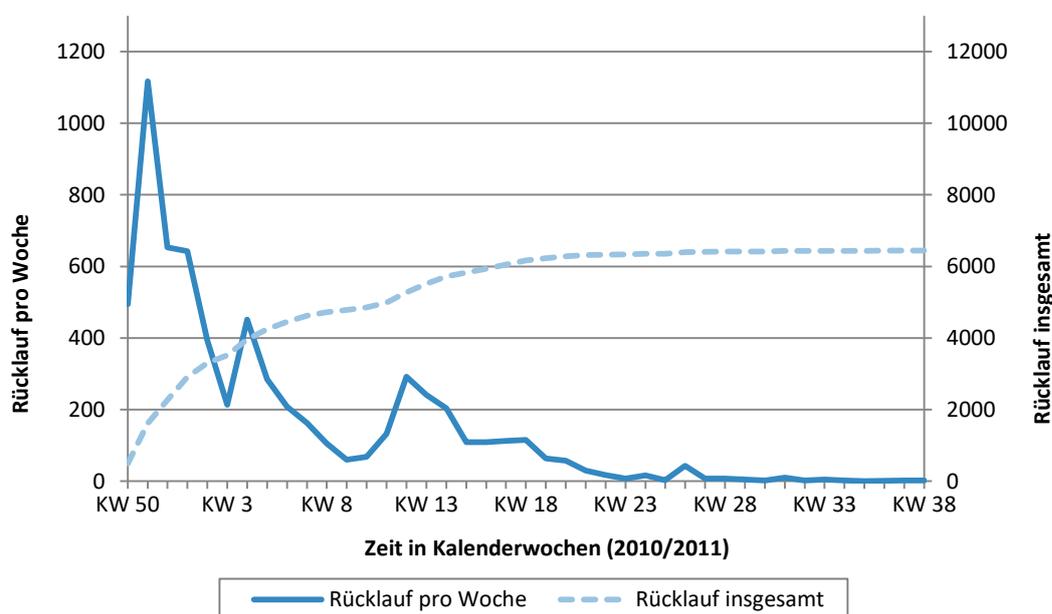
10.706 Personen, also über 90 Prozent der 11.787 Teilnehmer(innen) der ersten Welle, erklärten sich mit einer Kontaktierung für weitere Befragungen einverstanden. Diese ergeben die Bruttostichprobe der zweiten Welle. Aufgrund von verschiedenen stichprobenrelevanten

⁴⁹ 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.

⁵⁰ Aus der SUF-Version 1.0.0 musste nachträglich noch ein Fall wegen einer Dopplung in der (Original-)ID gelöscht werden.

Ausfällen (Nichtteilnahmen, ungültige Adressen⁵¹ und nicht auswertbare Fragebögen⁵²) umfasst die Nettostichprobe der zweiten Welle 6.472 Fälle, darunter 798 Bachelorabsolvent(inn)en. Bezogen auf die 10.706 Fälle der Bruttostichprobe liegt die Rücklaufquote damit bei 60,5 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.

Von den 6.472 Fällen aus der zweiten Welle wurden für die dritte Welle 6.453 Fälle kontaktiert.⁵³ Abzüglich der stichprobenrelevanten Ausfälle (Nichtteilnahmen, nicht auswertbare Fragebögen) verblieben in der Nettostichprobe der dritten Welle 4.279 Fälle, darunter 519 Bachelorabsolvent(inn)en und 3.760 Absolvent(inn)en traditioneller Studiengänge. Die Rücklaufquote liegt damit bei 66,3 % und ist höher als in der zweiten Welle, jedoch deutlich geringer als der Rücklauf bei den dritten Befragungswellen der vorherigen Absolventenkohorten 1997 (89 %) und 2001 (88 %) (Euler et al., 2018, S. 10).

⁵¹ Dies betraf etwa 300 Fälle.

⁵² Dies betraf 4 Fälle.

⁵³ Die Ausfälle ergeben sich zum einen aus Teilnahmeverweigerungen für Folgewellen, zum anderen aus ungültigen Adressen.

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

	Welle 1	Welle 2	Welle 3
Bruttostichprobe	47.800	10.706	6.453
Nettostichprobe	11.787 ^a	6.472 ^b	4.279 ^c
Rücklaufquote	24,7 %	60,5 %	66,3 %
Anteil Bruttostichprobe W2/3 an Bruttostichprobe W1	--	22,4 %	13,5 %
Anteil Nettostichprobe W2/3 an Nettostichprobe W1	--	54,9 %	36,3 %
Anteil Nettostichprobe W2/3 an Bruttostichprobe W1	--	13,5 %	9,0 %

^a darunter 1.622 Bachelorabsolvent(inn)en

^b darunter 798 Bachelorabsolvent(inn)en

^c darunter 519 Bachelorabsolvent(inn)en

[Panelausfälle] Das Absolventenpanel 2005 ist von paneltypischen Ausfallprozessen⁵⁴ 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 bzw. dritten 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 11.787 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. Diese Entwicklung setzt sich in der dritten Welle fort.

⁵⁴ Zu paneltypischen Ausfallprozessen vgl. Schnell, Hill & Esser, 2005, S. 241.

6 Datenaufbereitung

Im Folgenden werden die verschiedenen Schritte der Datenaufbereitung beschrieben. Die in den Kapiteln 6.1 und 6.2 beschriebenen Tätigkeiten wurden bereits durch das Primärforschungsprojekt durchgeführt. Die Generierung von Variablen (Kapitel 6.3) wurde sowohl durch das Primärforschungsprojekt als auch im Rahmen der Datenedition durch das FDZ-DZHW vorgenommen. Die in den Kapiteln 6.4 bis 6.6 dargestellten Tätigkeiten wurden durch das FDZ durchgeführt, teilweise aufbauend auf Vorarbeiten des Primärforschungsprojektes. Die im Rahmen der Datenedition vorgenommenen Aufbereitungsprozesse der Gewichtung und Anonymisierung werden in den beiden folgenden Kapiteln 7 und 8 gesondert erläutert.

6.1 Datenübertragung

[PAPI-Befragungen] Bei den PAPI-Befragungen mussten die Angaben der Befragten aus den Papierfragebögen zunächst auf Basis eines Codeplans zur weiteren Verarbeitung in ein computerlesbares Format übertragen werden. Zu diesem Zweck wurde auf dem Fragebogen vermerkt, welcher Frage bzw. Teilfrage eine Variable zugeordnet ist, welchen Namen diese Variable trägt und welche numerischen Codierungen für die Antworten der Befragten verwendet werden sollten. Dabei wurden auch bereits numerische Codierungen für die offenen Angaben vermerkt (vgl. Kapitel 6.3) sowie manuelle Vorkorrekturen zur Erleichterung der Datenübertragung vorgenommen (vgl. Kapitel 6.2). Um die Erfassungsreihenfolge festzulegen, wurden die Variablen zusätzlich nummeriert. 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.

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

6.2 Datenprüfung und Datenbereinigung

[Konsistenzprüfungen] Für alle Befragungswellen wurden verschiedene Konsistenzprüfungen durchgeführt. Folgende Typen von Prüfungen wurden vorgenommen:

- *Prüfung von Wertebereichen:* Es wurde geprüft, ob die erfasste Ausprägung einer Variablen in dem für diese Variable definierten Wertebereich lag.
- *Prüfung der Einhaltung der Filterführung:* Auf Grundlage der definierten Filterführung des Fragebogens wurde auf der einen Seite geprüft, ob für die jeweilige befragte Person Angaben zu erwarten gewesen wären, die aber nicht vorhanden waren (Vollständigkeitsprüfung), und auf der anderen Seite, ob für die jeweilige Person, Angaben vorhanden waren, die nicht hätten erfolgen dürfen (Filterverstöße).

- *Prüfung von Merkmalskombinationen:* Es wurde die Konsistenz der Angaben innerhalb eines Fragebogens sowie zwischen den Befragungen geprüft. Insbesondere wurden dabei inhaltliche und zeitbezogene Angaben zu gleichen Themengebieten miteinander abgeglichen. Beispielsweise wurde innerhalb beider Wellen jeweils geprüft, ob die Angaben über Berufstätigkeiten im Erwerbstätigkeitstableau mit den entsprechenden Angaben im Kalendarium übereinstimmten.

Festgestellte Inkonsistenzen wurden – falls möglich – durch den Abgleich mit anderen Nennungen im Fragebogen (ggf. mit Hilfe der Papierfragebögen) aufgelöst oder andernfalls ein entsprechender Missingcode (vgl. Kapitel 6.6) vergeben.

In den ersten beiden Befragungswellen wurden erste Konsistenzprüfungen bereits vor der Übertragung der Daten manuell auf den Papierfragebögen vorgenommen.⁵⁵ Darauf aufbauend erfolgte im Anschluss an die Datenübertragung eine umfassende Prüfung und Korrektur der Daten mit Hilfe einer DZHW-eigenen Software.⁵⁶ In der dritten Befragungswelle erfolgte die Konsistenzprüfung zum einen ebenfalls softwaregestützt, zum anderen – für das Kalendarium – über Stata-Do-Files.

[Löschung von Fällen] In allen Wellen wurden Fälle aus dem Datensatz entfernt. Ein Fall wurde gelöscht, wenn rund die Hälfte der Fragen bzw. zentrale Fragen (z. B. zum Studienfach) nicht beantwortet worden oder zu viele Inkonsistenzen vorhanden waren. Diese Fälle wurden als nicht auswertbar eingestuft. Darüber hinaus wurden in der ersten Welle Fälle identifiziert, die nicht zur Grundgesamtheit gehörten.⁵⁷ Diese wurden ebenfalls aus dem Datensatz entfernt.⁵⁸

6.3 Generierung von Variablen

Neben den Variablen, die die unmittelbaren 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 (halb-)offenen Nennungen.⁵⁹ Für die Papierfragebögen erfolgte die Codierung der (halb-)offenen Angaben bereits im Rahmen der Datenübertragung (vgl. Kapitel 6.1). Für die Online-Daten wurde die Codierung

⁵⁵ Die Zahl der vorgenommenen Korrekturen wurde nicht zentral, sondern nur auf den Papierfragebögen dokumentiert und ist daher nicht mehr systematisch rekonstruierbar.

⁵⁶ Zu diesem Zweck wurden die erfassten Fragebogen-Daten in eine Datenbank eingelesen. Anschließend wurden anhand formaler Regeln gültige Wertebereiche und Antwortkombinationen definiert und geprüft.

⁵⁷ 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.

⁵⁸ 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.

⁵⁹ Einzelne offene Fragen wurden nicht vercodet, weil sie vorwiegend als Kontextinformationen zur Codierung anderer offener Angaben erfasst wurden oder aufgrund von unzureichenden zeitlichen Ressourcen. Dies betrifft beispielsweise die typischen Arbeitsschwerpunkte, die neben der Berufsbezeichnung und dem Aufgabenbereich erhoben wurden (Frage 5.1 in der ersten Welle; Frage 4.12 in der zweiten Welle; Page 14 in der dritten Welle). 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.

in der Regel mit Excel durchgeführt. Die durch das Primärforschungsprojekt vorgenommenen Codierungsentscheidungen wurden dabei unverändert beibehalten. Zum anderen wurden Variablen des Datenschutzes wegen verändert (vgl. Kapitel 8) (z. B. Aggregation der Studienfächer zu Studienbereichen und Fächergruppen oder Ableitung von Hochschultyp und Hochschulort aus den Hochschulvariablen). Generierte Variablen im Zuge von Anonymisierungsmaßnahmen wurden bis auf wenige Ausnahmen durch das FDZ erstellt. Darüber hinaus wurden vom Primärforschungsprojekt Hilfsvariablen zur Erleichterung der Arbeit mit dem Erwerbstätigkeitstableau generiert.⁶⁰

Der Variablenname einer generierten Variablen ist im Datensatz durch das Suffix „_g#“ gekennzeichnet (vgl. Kapitel 6.5). 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.

Eine Übersicht aller für das Absolventenpanel 2005 generierten Variablen sowie Informationen zu ihrer Generierung findet sich in Anhang 1. Eine detaillierte Dokumentation der einzelnen Variablen mit Angabe ihrer jeweiligen Ausprägungen und Berechnungsvorschriften ist im Metadatensuchsystem auf der jeweiligen Variablendetailseite zu finden.⁶¹

6.4 Erstellung der Datensätze

[Zusammenführung der Wellen] Die Daten der einzelnen Befragungswellen 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, Page 9 der dritten Welle). Für jede befragte Person werden ein oder mehrere Episoden gespeichert. Dabei ist eine Episode definiert als ein Zeitraum, in dem eine bestimmte Tätigkeitsart (z. B. Erwerbstätigkeit, Ausbildung) ausgeübt wird bzw. ein konkreter Status (z. B. Elternzeit, Arbeitslosigkeit) besteht. Für jede Episode einer Person existiert jeweils eine Datenzeile (long format). Die Struktur entspricht der gängigen Struktur für Episodendaten (vgl. Scherer & Brüderl, 2010, S. 1042). Die Episoden wurden fallweise sortiert, das heißt alle Episoden einer Person folgen direkt aufeinander. Verschiedene Tätigkeitsarten im selben Zeitraum wurden jeweils als eigenständige Episode codiert. Wenn Tätigkeiten derselben Art unmittelbar aufeinander folgten oder parallel ausgeübt wurden, wurden sie zu einer Episode zusammengefasst. Daher geht aus den Episodendaten nicht hervor, ob eine Episode eine oder mehrere Tätigkeiten derselben Art umfasst. Für Episoden der Tätigkeitsarten Erwerbstätigkeit und akademische Weiterqualifizie-

⁶⁰ vgl. dazu auf [https://metadata.fdz.dzhw.eu/#/de/instruments/ins-gra2005-ins3\\$](https://metadata.fdz.dzhw.eu/#/de/instruments/ins-gra2005-ins3$) die Abgleichsziffer-Dokumentation sowie die Elternzeit-Flag-Variablen-Dokumentation

⁶¹ z.B. [https://metadata.fdz.dzhw.eu/#/de/variables/var-gra2005-ds1-astu011e_g1o\\$](https://metadata.fdz.dzhw.eu/#/de/variables/var-gra2005-ds1-astu011e_g1o$)

rung 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 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.

6.5 Vergabe von Variablennamen, Variablenlabels und Wertelabels

[Variablen- und Wertelabelvergabe] Für Variablen- und Wertelabel wurden Formulierungen des Fragebogens übernommen oder prägnante Kurzformen dieser Formulierungen gewählt. Dabei basieren die Variablenlabel in der Regel auf dem entsprechenden Fragetext. Grundlage für die Wertelabel 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 Wertelabel die Bezeichnungen der Schlüssel der Klassifikation wortgetreu übernommen. Die Variablen- und Wertelabel 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.⁶²

[Variablenbenennung im Personendatensatz] Für die Variablenbenennung wurde im FDZ-DZHW eine einheitliche Benennungssystematik erstellt. Mit Ausnahme der Identifikatorvariablen *pid* sowie der Wellenvariablen *wave*⁶³ 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.⁶⁴ Der Stamm setzt sich aus drei Komponenten zusammen. Mit einem dreistelligen englischen Buchstabenkürzel wird die Variable erstens einem inhaltlichen Themengebiet zugeordnet. Tabelle 3 stellt die verschiedenen Themenbereiche des Absolventenpanels 2005 sowie das zugehörige Kürzel für den Stamm des Variablennamens im Überblick dar. Zweitens werden die Variablen innerhalb der definierten Themenbereiche auf zwei bis vier Stellen durchnummeriert. Mit Hilfe eines Buchstabens am Ende des Stamms können drittens verschiedene Variablen, die zur gleichen Frage gehören und dadurch die gleiche Themendifferenzierung und Nummerierung aufweisen (z. B. bei Itembatterien, Mehrfachnennungen oder Fragen, in denen geschlossene und offene Fragen kombiniert werden), gekennzeichnet werden (z. B. *stu01a*, *stu01b*, *stu01c*, ...). Für Indikatoren, die in mehreren Befragungswellen erfasst wurden, sind die Namen der zuge-

⁶² Mit dem Befehl „label language en“ erfolgt ein Wechsel zu den englischen Labels. Über „label language de“ kann zu den deutschen Labels zurückgewechselt werden.

⁶³ Diese enthält die Information, welche Fälle an welchen Wellen teilgenommen haben.

⁶⁴ Welle 1: a; Welle 2: b; Welle 3 (Hauptbefragung): c

hörigen Variablen („Panelvariablen“) durch die Vergabe eines identischen Stammes harmonisiert.

Tabelle 3: Themengebiete und Kürzel für Variablennamen des DZHW-Absolventenpanels 2005

Themengebiete-Kürzel	Themengebiet (englisch)	Themengebiet (deutsch)
stu	studies	Studium
occ	occupation	Beschäftigung
ski	skills	Fähigkeiten
fmt	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

Das anhand eines Unterstriches vom Stamm abgetrennte Suffix enthält verschiedene Zusatzinformationen:

- Generierte Variablen (vgl. Kapitel 6.3) wurden mit dem Kürzel **g#** (g1 bzw. bei weiteren Derivaten g2, g3, ...) markiert. Unter den Typus der generierten Variablen fallen alle Variablen, die aus einer oder mehreren Variablen des Ursprungsdatensatzes erzeugt wurden (z.B. vercodete Variablen, Indizes, Aggregationen).
- Bei Informationen, die in mehreren Wellen wiederholt, jedoch mit abgewandelten Fragen-, Item- oder Antwortkategorienformulierungen abgefragt wurden⁶⁵, haben die neuen Variablenversionen zwar denselben Variablenstamm wie die Ursprungsvariable, jedoch ergänzt um ein Versionskürzel **v#** im Suffix (_v1 für die erste Änderung, _v2 für die zweite Änderung usw.). Die ursprüngliche Version der Variable erhält keine Kennzeichnung.
- Bestimmte Variable sind aus Anonymisierungsgründen nicht über alle potentiellen Zugangswege (Download-CUF, Download-SUF, Remote-Desktop-SUF, On-Site-SUF; vgl. Kapitel III) einsehbar. In diesen Fällen wird im Suffix des Variablennamens derjenige Zugangsweg angegeben, ab dem die Variable nutzbar⁶⁶ ist:
 - d:** Variable ist nicht im CUF, aber im Download-SUF, im Remote-Desktop-SUF und im On-Site-SUF nutzbar.
 - r:** Variable ist nicht im CUF und im Download-SUF, aber im Remote-Desktop-SUF und im On-Site-SUF nutzbar.
 - o:** Variable ist nicht im CUF, im Download-SUF und im Remote-Desktop-SUF, aber im On-Site-SUF nutzbar.
 - a:** Variable ist über keinen Zugangsweg nutzbar. Sie wird aber dokumentiert, da es zugehörige Fragen im Fragebogen gibt.

⁶⁵ Die Darstellung und die Anordnung der jeweiligen Frage im Erhebungsinstrument spielen keine Rolle. Ebenso ist nicht notwendig, dass die zugehörige Frage auch den gleichen Personengruppen wiederholt gestellt wurde.

⁶⁶ „Nutzbar“ heißt: die Variable enthält nicht das Missing „anonymisiert“.

- Gewichtungsvariablen enthalten (zusätzlich zu dem festen Stamm „wgt“) das Suffix **t#** als Indikator für die Wellen, auf die sie sich beziehen (vgl. Kapitel 7.1).⁶⁷

[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.

6.6 Codierung fehlender Werte

Zur Codierung fehlender Werte wurde im FDZ-DZHW eine übergreifende Systematik erstellt, um über verschiedene Datensätze des DZHW hinweg eine einheitliche Missingcodierung gewährleisten zu können. Fehlende Angaben wurden dabei durch dreistellige negative Werte codiert. Tabelle 4 stellt die verwendete Missingsystematik dar. Die im Absolventenpanel 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).

⁶⁷ Sie enthalten daher kein Wellenpräfix.

Tabelle 4: Systematik des FDZ-DZHW für fehlende Werte

Wertebereich	Code	Wertelabel
-999 bis -990: Nonresponse	-999	weiß nicht
	-998	keine Angabe
	-997	keine Angabe (Antwortkategorie)
	-996	Interviewabbruch
	-995	keine Teilnahme (Panel)
	-994	verweigert
-989 bis -970: Nicht zutreffend	-989	filterbedingt fehlend
	-988	trifft nicht zu
	-987	designbedingt fehlend (Fragebogensplit)
	-986	designbedingt fehlend (Welle) ^a
	-985	designbedingt fehlend (Kohorte) ^b
-969 bis -950: Editierter fehlender Wert	-969	unbekannter fehlender Wert ^c
	-968	unplausibler Wert^d
	-967	anonymisiert
	-966	nicht bestimmbar^e
	-965	ungültige Mehrfachnennung
-949 bis -930: Item-spezifische fehlende Werte^f	-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 und dritte Welle wird zusätzlich ein Nonresponsegewicht berechnet, das mit Hilfe von Informationen aus den jeweiligen vorangegangenen Wellen die Nichtteilnahme in der zweiten bzw. dritten Welle modelliert. Daraus wurden wiederum die Längsschnittgewichte berechnet. In Tabelle 5 sind die erstellten Gewichte dargestellt.

Tabelle 5: Bereitgestellte Gewichte zum DZHW-Absolventenpanel 2005

Variablenname	Beschreibung
wgt_t1d	Querschnittsgewicht Welle 1
wgt_t1t2d	Längsschnittgewicht 2-Wellen-Panel
wgt_t1t2t3d	Längsschnittgewicht 3-Wellen-Panel

[Hinweise zur Anwendung der Gewichte] Alle Gewichte wurden nur für den Datensatz der Absolvent(innen) traditioneller Studiengänge 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, die Gewichte wgt_t1t2d und wgt_t1t2t3d für Auswertungen des Zwei- bzw. Drei-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. Gleiches gilt für Auswertungen in der dritten Welle.

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 umfangreiche 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 $wgt_{i_{t_1}}$. Für die zweite und dritte Befragungswelle wurde kein Querschnittsgewicht erstellt, da weder neu hinzugekommene Individuen befragt wurden (Auffrischungsstichprobe) noch Personen die Möglichkeit hatten, an Welle 2 bzw. 3 teilzunehmen, wenn sie nicht Teil der ersten Befragungswelle waren.

[Längsschnittgewichtung] Für die Gewichtung des Zwei- bzw. Drei-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 nächsten Welle teilzunehmen. Im Unterschied zu der Nonresponse-Adjustierung in der ersten Welle lag für die Nichtteilnehmer(innen) der Wellen 2 und 3 eine Vielzahl an Informationen aus den jeweils vorangegangenen Wellen vor. Diese dienten als Kovariaten (σ_t) in einem Probit-Regressionsmodell, das die Vorhersage der Teilnahmewahrscheinlichkeit im folgenden 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 erwies sich als bedeutsam für die Vorhersage der Teilnahmewahrscheinlichkeit in der zweiten bzw. dritten Welle. Aus dem Modell konnte die bedingte Teilnahmewahrscheinlichkeit abgeleitet werden, deren Kehrwert das Ausfallgewicht für die zweite bzw. dritte Welle darstellt.⁷⁵

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

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

$$wgt_{i_{t_1t_2}} = \widehat{dwgt}_{sci} \times NR_{i_{gew_{t_2}}}$$

$$wgt_{i_{t_1t_2t_3}} = \widehat{dwgt}_{sci} \times NR_{i_{gew_{t_2}}} \times NR_{i_{gew_{t_3}}}$$

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

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

⁷⁵ 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.⁷⁶

[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.

⁷⁶ 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 Daten⁷⁷, die in freiwilligen Befragungen durch das DZHW erhoben werden, gelten die EU-Datenschutz-Grundverordnung (EU-DSGVO) und das Bundesdatenschutzgesetz in seiner Neufassung vom 30. Juni 2017.⁷⁸ Danach sind personenbezogene Daten für die Weitergabe zur wissenschaftlichen Sekundärnutzung (ohne Vorliegen einer Einverständniserklärung zur Sekundärnutzung der personenbezogenen Daten) in der Regel derart aufzubereiten, dass „die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden können“ (Art. 4 Abs. 5 DSGVO; s. auch Art. 89 DSGVO sowie Erwägungsgrund 26 DSGVO). Das heißt, für die Weitergabe von Daten aus wissenschaftlichen Forschungsprojekten an Dritte sind die Daten derart zu anonymisieren, dass kein Bezug zur Person mehr hergestellt werden kann.

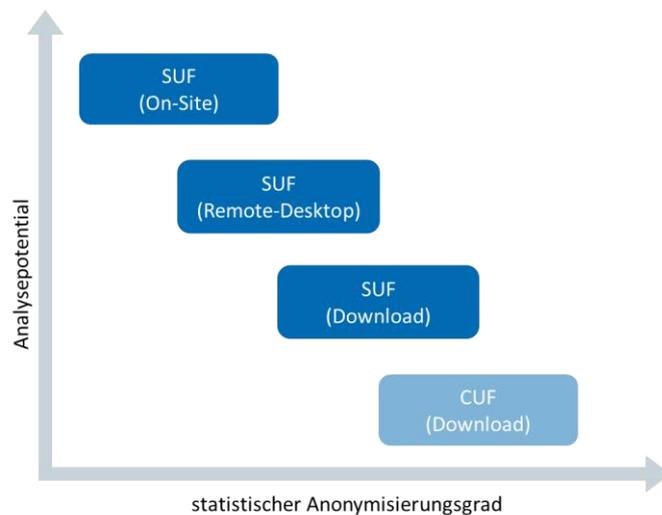
[Datenzugang, Anonymisierungsgrad und Analysepotential] Das FDZ des DZHW stellt für das Absolventenpanel 2005 ein SUF für die wissenschaftliche Sekundärnutzung und ein CUF für Lehr- und Übungszwecke zur Verfügung. Die Anonymität der Befragten wird dabei über eine Kombination aus statistischen Maßnahmen 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 des 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.

⁷⁷ „Personenbezogene Daten (sind) alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person“) beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden kann, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind“ (Art. 4 DSGVO, S. 1).

⁷⁸ Die DSGVO gilt grundsätzlich innerhalb der EU und somit ebenfalls für das DZHW. Das BDSG in seiner Neufassung vom 30. Juni 2017 (Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680 (Datenschutz-Anpassungs- und Umsetzungsgesetz EU DSAnpUG-EU)) kommt teils zusätzlich zur Anwendung, da die DZHW GmbH juristisch als öffentliche Stelle des Bundes betrachtet wird (§ 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.

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: Hochschule, Studienfach, Abschlussart, Berufsangaben, regionale Informationen (zur Hochschule, zum Ort des Erwerbs der Studienberechtigung und Arbeitsort), Staatsangehörigkeit und Geburtsland. Um eine eindeutige Zuordnung der Absolvent(inn)endaten zu unterbinden, wurden diese Schlüsselmerkmale – je nach Datenprodukt bzw. Zugangsweg – aggregiert oder gelöscht (vgl. Tabelle 6). Beispielsweise wird das Merkmal „Hochschule“ in dem SUF für die On-Site Nutzung zu NUTS-2-Regionen, im Remote-Desktop-SUF zu Bundesländern, im Download-SUF und im Download-CUF zu den zwei Kategorien alte vs. neue Bundesländern aggregiert.

Darüber hinaus empfehlen Ebel und Meyermann, offene Angaben zu löschen „selbst wenn die jeweiligen Fragestellungen an sich unproblematisch sind. Denn es besteht die Gefahr, dass Studienteilnehmer/-innen bei eigentlich unbedenklichen Fragen mit offener Antwortmöglichkeit kritische Informationen preisgegeben haben, die zu einer Identifikation

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.

führen könnten“ (Ebel & Meyermann, 2015, S. 5). Die offenen Angaben wurden größtenteils bereits im Rahmen der Datenaufbereitung durch das Primärforschungsprojekt vercodet und werden in dieser Form (teilweise aggregiert) zur Verfügung gestellt. Nicht codierte offene Angaben wurden im CUF und in allen SUF-Varianten gelöscht.

Zuletzt wurde geprüft, ob in den Daten *sensible Informationen*, z. B. zur Gesundheit, sexuellen Orientierung und zu politischen Einstellungen, enthalten waren. Diese eignen sich zwar nicht notwendig 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 (Art. 9 DSGVO, Erwägungsgrund 51 DSGVO). 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 Realisierung der Anonymität der Befragten im CUF wurden zum einen im Vergleich zu den SUF-Varianten restriktivere statistische Anonymisierungsmaßnahmen auf Variablenebene vorgenommen. Zum anderen wurde eine per Zufallsauswahl gewonnene Substichprobe der Daten (2/3 der befragten Absolvent(inn)en traditioneller Studiengänge) gezogen.

Die nachfolgende Tabelle 6 stellt in Kurzform die durchgeführten statistischen Anonymisierungsmaßnahmen je nach Datenform bzw. Zugangsweg dar.

Tabelle 6: Maßnahmen der statistischen Anonymisierung der Daten des DZHW-Absolventenpanels 2005 nach Zugangsweg⁸⁰

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
Fragebogeneingangsdatum	Freigabe	Freigabe von Monat/Jahr	Freigabe von Monat/Jahr	Löschung
Studienfach ^a	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

⁸⁰ 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)
Arbeitsort (Bundesland/Land)	Freigabe	Freigabe	Aggregation zu Bundesländern und Ausland	Aggregation zu alte, neue Bundesländer und Ausland
Arbeitsort (PLZ-3-Steller)	Freigabe	Aggregation zu NUTS 2: Basisregionen für regionalpolitische Maßnahmen ^b	Aggregation zu NUTS 2: Basisregionen für regionalpolitische Maßnahmen ^b	Löschung
Ort der Studienberechtigung (Bundesland/Ausland)	Freigabe	Freigabe	Aggregation zu Bundesländern und Ausland	Aggregation zu alte, neue Bundesländer und Ausland
Ort der Studienberechtigung (PLZ-3-Steller)	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-3-Steller)	Freigabe	Aggregation zu NUTS 2: Basisregionen für regionalpolitische Maßnahmen ^b	Aggregation zu NUTS 2: Basisregionen für regionalpolitische Maßnahmen ^b	Aggregation zu Missing/ Deutschland
Beruf (KIdB-Berufsklassen (4-Steller) bzw. Berufsgattungen (5-Steller)) ^c	Freigabe	Aggregation zu Berufsordnungen /-gruppen (3-Steller) ^c	Aggregation zu Berufsordnungen /-gruppen (3-Steller) ^c	Aggregation zu Berufsgruppen /Berufshauptgruppen (2-Steller) ^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/ Löschung	Vercodung/ Löschung	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 (KIdB) von 1992 (Welle 1) bzw. 2010 (Welle 2 und 3)

^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. Zugriff am 17.03.2016 UTC.
- 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.
- Euler, T., Trennt, F., Trommer, M. & Schaeper, H. (2018). Werdegänge der Hochschulabsolventinnen und Hochschulabsolventen 2005. Dritte Befragung des Prüfungsjahrgangs 2005 zehn Jahre nach dem Abschluss. (Forum Hochschule 01/2018). Hannover: DZHW.
- 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). *Datenschutz am Forschungsdatenzentrum* (FDZ-Methodenreport Nr. 6). 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. & Briedis, K. (2004). Kompetenzen von Hochschulabsolventinnen und Hochschulabsolventen, berufliche Anforderungen und Folgerungen für die Hochschulreform. Hannover: HIS.
- 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., Grotheer, M. & Brandt, G. (2014). Familiengründung von Hochschulabsolventinnen. Eine empirische Untersuchung verschiedener Examenskohorten. In D. Konietzka & M. Kreyenfeld (Hrsg.), *Ein Leben ohne Kinder* (2. Aufl., S. 47–80). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-94149-3_2
- Scherer, S. & Brüderl, J. (2010). Sequenzdatenanalyse. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 1031–1051). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schnell, R., Hill, P. B. & Esser, E. (2005). *Methoden der empirischen Sozialforschung* (7. Aufl.). München: Oldenbourg.
- Valliant, R., Dever, J. A. & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York (NY): Springer New York. doi:10.1007/978-1-4614-6449-5

Anhang 1: Generierte Variablen im DZHW-Absolventenpanel 2005

Variablen	Generierungsinformationen ^a
astu011e_g1o	Vercodung offener Angaben zum Studienfach Codierliste: cl-destatis-studienfach-2005 ^b (Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2004/2005 und SoSe 2005), Schlüssel 3.1)
astu011g_g1o	
astu011i_g1o	
astu012e_g1o	
astu012g_g1o	
astu012i_g1o	
astu013e_g1o	
astu013g_g1o	
astu013i_g1o	
astu014e_g1o	
astu014g_g1o	
astu014i_g1o	
astu015e_g1o	
astu015g_g1o	
astu015i_g1o	
astu021c_g1o	
astu021d_g1o	
astu021e_g1o	
astu022c_g1o	
astu022d_g1o	
astu022e_g1o	
astu023c_g1o	
astu023d_g1o	
astu023e_g1o	
afec021h_g1o	
afec022h_g1o	
bfec161g_g1o	
bfec162g_g1o	
bfec163g_g1o	
astu011e_g2d	
astu011g_g2d	
astu011i_g2d	
astu012e_g2d	
astu012g_g2d	
astu012i_g2d	
astu013e_g2d	
astu013g_g2d	
astu013i_g2d	
astu014e_g2d	
astu014g_g2d	
astu014i_g2d	
astu015e_g2d	
astu015g_g2d	
astu015i_g2d	
astu021c_g2d	
astu021d_g2d	
astu021e_g2d	
astu022c_g2d	
astu022d_g2d	
astu022e_g2d	
astu023c_g2d	
astu023d_g2d	
astu023e_g2d	

Variablen	Generierungsinformationen ^a
afec021h_g2d afec022h_g2d bfec161g_g2d bfec162g_g2d bfec163g_g2d	
astu011e_g3 astu011g_g3 astu011i_g3 astu012e_g3 astu012g_g3 astu012i_g3 astu013e_g3 astu013g_g3 astu013i_g3 astu014e_g3 astu014g_g3 astu014i_g3 astu015e_g3 astu015g_g3 astu015i_g3 astu021c_g3 astu021d_g3 astu021e_g3 astu022c_g3 astu022d_g3 astu022e_g3 astu023c_g3 astu023d_g3 astu023e_g3 afec021h_g3 afec022h_g3 bfec161g_g3 bfec162g_g3 bfec163g_g3	Aggregation der Studienfächer zu Fächergruppen Codierliste: cl-destatis-studienfach-2005 (Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2004/2005 und SoSe 2005), Schlüssel 3.1)
astu011f_g1 astu011h_g1 astu011j_g1 astu012f_g1 astu012h_g1 astu012j_g1 astu013f_g1 astu013h_g1 astu013j_g1 astu014f_g1 astu014h_g1 astu014j_g1 astu015f_g1 astu015h_g1 astu015j_g1 astu021f_g1 astu022f_g1 astu023f_g1	Vercodung offener Angaben zum Studienabschluss Codierliste: cl-dzhw-2 (projekteigene Codierliste)
astu09a_g1r astu09b_g1r	Vercodung offener Angaben zum fachlichen Studienschwerpunkt (für ausgewählte wirtschaftswissenschaftliche Fächer) Codierliste: cl-dzhw-3 (projekteigene Codierliste)
astu011k_g1a astu012k_g1a astu013k_g1a	Vercodung offener Angaben zur Hochschule Codierliste: cl-destatis-hochschule-2005 ^b (Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2004/2005 und SoSe 2005),

Variablen	Generierungsinformationen ^a
aocc224j_g1r aocc225j_g1r aocc226j_g1r bocc221j_g1v1r bocc222j_g1v1r bocc223j_g1v1r bocc224j_g1v1r bocc225j_g1v1r bocc226j_g1v1r bocc227j_g1r bocc228j_g1r bocc229j_g1r cocc221j_g1v1r cocc222j_g1v1r cocc223j_g1v1r cocc224j_g1v1r cocc225j_g1v1r cocc226j_g1v1r cocc227j_g1r cocc228j_g1r cocc229j_g1r cocc2210j_g1r cocc2211j_g1r cocc2212j_g1r	sels (AGS)); cl-dzhw-1 (projekteigene Codierliste, Welle 1 und 2);cl-destatis-ausland-2016 ^b (Destatis-Staats- und Gebietssystematik 2016, Welle 3)
aocc221j_g2d aocc222j_g2d aocc223j_g2d aocc224j_g2d aocc225j_g2d aocc226j_g2d bocc221j_g2v1d bocc222j_g2v1d bocc223j_g2v1d bocc224j_g2v1d bocc225j_g2v1d bocc226j_g2v1d bocc227j_g2d bocc228j_g2d bocc229j_g2d cocc221j_g2v1d cocc222j_g2v1d cocc223j_g2v1d cocc224j_g2v1d cocc225j_g2v1d cocc226j_g2v1d cocc227j_g2d cocc228j_g2d cocc229j_g2d cocc2210j_g2d cocc2211j_g2d cocc2212j_g2d	Aggregation der Länder des Arbeitsortes (bei Ausland) zur Kategorie „Ausland“ (Bundesländer wurden beibehalten)
aocc221j_g3 aocc222j_g3 aocc223j_g3 aocc224j_g3 aocc225j_g3 aocc226j_g3 bocc221j_g3v1	Aggregation des Arbeitsortes (Bundesländer) zu neue/alte Bundesländer bzw. Länder des Arbeitsortes (bei Ausland) zur Kategorie „Ausland“

Variablen	Generierungsinformationen ^a
bocc222j_g3v1	
bocc223j_g3v1	
bocc224j_g3v1	
bocc225j_g3v1	
bocc226j_g3v1	
bocc227j_g3	
bocc228j_g3	
bocc229j_g3	
cocc221j_g3v1	
cocc222j_g3v1	
cocc223j_g3v1	
cocc224j_g3v1	
cocc225j_g3v1	
cocc226j_g3v1	
cocc227j_g3	
cocc228j_g3	
cocc229j_g3	
cocc2210j_g3	
cocc2211j_g3	
cocc2212j_g3	
cocc221k_g2v2o	Aggregation des Arbeitsortes (5-stellige Postleitzahl) zur 3-stelligen Postleitzahl
cocc222k_g2v2o	
cocc223k_g2v2o	
cocc224k_g2v2o	
cocc225k_g2v2o	
cocc226k_g2v2o	
cocc227k_g2v1o	
cocc228k_g2v1o	
cocc229k_g2v1o	
cocc2210k_g2o	
cocc2211k_g2o	
cocc2212k_g2o	
aocc221k_g1d	Aggregation des Arbeitsortes (Postleitzahl) zu NUTS-2-Regionen Codierliste: cl-eurostat-nuts-2010 (NUTS-Klassifikation des Statistischen Amtes der Europäischen Union (Eurostat) von 2010, Welle 1 und 2); cl-dzhw-5 (projekteigene Codierliste, Welle 1 und 2); cl-eurostat-deplznuts-2010 (Tercet NUTS-postal codes matching tables des Statistischen Amtes der Europäischen Union (Eurostat) von 2010, Welle 3)
aocc222k_g1d	
aocc223k_g1d	
aocc224k_g1d	
aocc225k_g1d	
aocc226k_g1d	
bocc221k_g1v1d	
bocc222k_g1v1d	
bocc223k_g1v1d	
bocc224k_g1v1d	
bocc225k_g1v1d	
bocc226k_g1v1d	
bocc227k_g1d	
bocc228k_g1d	
bocc229k_g1d	
cocc221k_g1v2d	
cocc222k_g1v2d	
cocc223k_g1v2d	
cocc224k_g1v2d	
cocc225k_g1v2d	
cocc226k_g1v2d	
cocc227k_g1v1d	
cocc228k_g1v1d	
cocc229k_g1v1d	
cocc2210k_g1d	
cocc2211k_g1d	

Variablen	Generierungsinformationen ^a
cocc2212k_g1d	
adem05a_g1r	Vercodung offener Angaben zum Ort der Studienberechtigung (Bundesland/Land) Codierliste: cl-destatis-bundesland-1990 ^b (Destatis Bundeslandschlüssel (entsprechend der ersten beiden Ziffern des Amtlichen Gemeindecodes (AGS))); cl-dzhw-1 (projekteigene Codierliste)
adem05a_g2d	Aggregation der Länder des Ortes der Studienberechtigung (bei Ausland) zur Kategorie „Ausland“ (Bundesländer wurden beibehalten)
adem05a_g3	Aggregation der Bundesländer des Ortes der Studienberechtigung zu neue/alte Bundesländer bzw. Länder des Arbeitsortes (bei Ausland) zur Kategorie „Ausland“
adem05b_g1d	Aggregation des Ortes der Studienberechtigung (Postleitzahl) zu NUTS-2-Regionen Codierliste: cl-eurostat-nuts-2010 (NUTS-Klassifikation des Statistischen Amtes der Europäischen Union (Eurostat) von 2010); cl-dzhw-5 (projekteigene Codierliste)
astu061c_g1	Vercodung offener Angaben zum Auslandsaufenthalt
astu062c_g1	Codierliste: cl-dzhw-1 (projekteigene Codierliste)
astu063c_g1	
astu064c_g1	
afec021k_g1r	Vercodung offener Angaben zum Land der weiteren akademischen Qualifikation
afec022k_g1r	Codierliste: cl-dzhw-1 (projekteigene Codierliste)
afec021k_g2	Aggregation des Landes der weiteren akademischen Qualifikation zu den Kategorien „Deutschland“ und „Ausland“
afec022k_g2	
bocc46b_g1	Vercodung offener Angaben zum Land der Unternehmenszentrale Codierliste: cl-dzhw-1 (projekteigene Codierliste)
bdem18b_g1o	Vercodung offener Angaben zum Geburtsland bzw. der Staatsangehörigkeit
bdem19b_g1o	Codierliste: cl-dzhw-1 (projekteigene Codierliste)
bdem18b_g2r	Aggregation des Geburtslandes bzw. der Staatsangehörigkeit zu NEPS-Regionen (angepasst)
bdem19b_g2r	Codierliste: cl-dzhw-1 (NEPS-Regionen)
bdem18b_g3d	Aggregation des Geburtslandes bzw. der Staatsangehörigkeit zu Weltregionen
bdem19b_g3d	Codierliste: cl-dzhw-1 (projekteigene Codierliste)
aocc19_g1o	Vercodung offener Angaben zum Beruf (KldB-1992-4-Steller)
aocc19_g2d	Codierliste: cl-destatis-kldb-1992 (Destatis-Klassifikation der Berufe 1992)
aocc19_g2d	Aggregation des Berufs (KldB-1992-4-Steller) zum KldB-1992-3-Steller
aocc19_g3	Codierliste: cl-destatis-kldb-1992 (Destatis-Klassifikation der Berufe 1992)
aocc19_g3	Aggregation des Berufs (KldB-1992-4-Steller) zum KldB-1992-2-Steller
aocc35b_g1d	Codierliste: cl-destatis-kldb-1992 (Destatis-Klassifikation der Berufe 1992)
aocc35b_g1d	Vercodung offener Angaben zum Beruf (KldB-1992-3-Steller)
aocc35b_g2	Codierliste: cl-destatis-kldb-1992 (Destatis-Klassifikation der Berufe 1992)
aocc35b_g2	Aggregation des Berufs (KldB-1992-3-Steller) zum KldB-1992-2-Steller
bocc19_g1v1o	Codierliste: cl-destatis-kldb-1992 (Destatis-Klassifikation der Berufe 1992)
cocc19_g1v1o	Vercodung offener Angaben zum Beruf (KldB-2010-5-Steller)
bocc19_g2v1d	Codierliste: cl-destatis-kldb-2010 ^b (Destatis-Klassifikation der Berufe 2010, Welle 2); cl-destatis-kldb-2010 ^b (Destatis-Klassifikation der Berufe 2010, Welle 3)
cocc19_g2v1d	Aggregation des Berufs (KldB-2010-5-Steller) zum KldB-2010-3-Steller
bocc19_g3v1	Codierliste: cl-destatis-kldb-2010 ^b (Destatis-Klassifikation der Berufe 2010)
cocc19_g3v1	Aggregation des Berufs (KldB-2010-5-Steller) zum KldB-2010-2-Steller
aocc20a_g1r	Codierliste: cl-destatis-kldb-2010 ^b (Destatis-Klassifikation der Berufe 2010)
aocc20b_g1r	Vercodung offener Angaben zum beruflichen Aufgabenbereich
bocc20a_g1v1r	Codierliste: cl-dzhw-4 (projekteigene Codierliste)

Variablen	Generierungsinformationen ^a	
bocc20b_g1v1r		
bdem231c_g1d	Aggregation des Wohnsitzes (Postleitzahl) zu NUTS-2-Regionen Codierliste: cl-eurostat-nuts-2010 (NUTS-Klassifikation des Statistischen Amtes der Europäischen Union (Eurostat) von 2010); cl-dzhw-5 (projektei- gene Codierliste)	
bdem232c_g1d		
bdem233c_g1d		
bdem234c_g1d		
bdem235c_g1d		
bdem231c_g2	Aggregation des Wohnsitzes (Postleitzahl) zur Kategorie „Deutschland“	
bdem232c_g2		
bdem233c_g2		
bdem234c_g2		
bdem235c_g2		
bdem231d_g1r	Verkodung offener Angaben zum Wohnsitz im Ausland Codierliste: cl-dzhw-1 (projekteigene Codierliste)	
bdem232d_g1r		
bdem233d_g1r		
bdem234d_g1r		
bdem235d_g1r		
bdem231d_g2	Aggregation des Landes des Wohnsitzes im Ausland zur Kategorie „Aus- land“	
bdem232d_g2		
bdem233d_g2		
bdem234d_g2		
bdem235d_g2		
bocc221m_g1	vgl. Abgleichsziffer-Dokumentation https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-gra2005-ins3\$	
bocc222m_g1		
bocc223m_g1		
bocc224m_g1		
bocc225m_g1		
bocc226m_g1		
bocc227m_g1		
bocc228m_g1		
bocc229m_g1		
cocc221m_g1		
cocc222m_g1		
cocc223m_g1		
cocc224m_g1		
cocc225m_g1		
cocc226m_g1		
cocc227m_g1		
cocc228m_g1		
cocc229m_g1		
cocc2210m_g1		
cocc2211m_g1		
cocc2212m_g1		
cocc221n_g1		vgl. Elternzeit-Flag-Variablen-Dokumentation https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-gra2005-ins3\$
cocc222n_g1		
cocc223n_g1		
cocc224n_g1		
cocc225n_g1		
cocc226n_g1		
cocc227n_g1		
cocc228n_g1		
cocc229n_g1		
cocc2210n_g1		
cocc2211n_g1		
cocc2212n_g1		
adem11_g1	Diese Variable wurde in der Regel anhand der angegebenen Geburtsdaten der Kinder gebildet. Die Anzahl der Kinder entspricht jedoch in Sonderfäl- len nicht immer der Summe der angegebenen Kinder (Geburtsdaten).	
bdem11_g1		
aocc14a_g1	Verkodung offener Angaben zu Branchen des Praktikums	

Variablen	Generierungsinformationen ^a
aocc14b_g1	Codierliste: analog der Antwortkategorien bei Frage 5.7 in Welle 1
adem07_g1	Aggregation des Alters des Befragten (Top-Codierung für Geburtsjahre bis 1959)
astu064d_g1r	Vercodung sonstiger offener Angaben
afec03o_g1r	Zuordnung zu vorgegebenen Kategorien oder projekteigene Codierliste
aocc05l_g1r	
aocc06s_g1r	
aocc15h_g1r	
aocc231v_g1r	
aocc232v_g1r	
aocc261b_g1r	
aocc262b_g1r	
aocc271b_g1r	
aocc272b_g1r	
adem01b_g1r	
adem02b_g1r	
afvt01l_g1r	
bocc37p_g1r	
bocc06s_g1v1r	
bfec03o_g1v1r	
bfec11h_g1r	
bocc232v_g1v1r	
bocc272b_g1v1r	
bocc262b_g1v1r	
bocc292j_g1r	
bdem16o_g1r	
bfec161j_g1r	
bfec162j_g1r	
bfec163j_g1r	
cocc272b_g1v1r	

^a Eine Codierlisten-ID wurde nur dann vergeben, wenn die Kategorien nicht aus den tatsächlichen Nennungen im Datensatz hergeleitet wurden, sondern sich aus einer Systematik ergeben. Alle Codierlisten sind im Metadatensuchsystem bei den Erhebungsinstrumenten zu finden.

^b ggf. ergänzt um Sondercodes, wenn nicht zuordenbar