

Gesche Brandt / Susanne de Vogel / Steffen Jaksztat /  
Carola Teichmann / Kerstin Lange / Percy Scheller /  
Sandra Vietgen

# DZHW-Promoviertenpanel 2014

Daten- und Methodenbericht zu den Erhebungen der  
Promoviertenkohorte 2014 (1. und 2. Befragungswelle)

## Daten- und Methodenbericht

Mai 2018

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



Autor(inn)en:

Gesche Brandt

Susanne de Vogel

Steffen Jaksztat

Carola Teichmann

Kerstin Lange

Percy Scheller

Sandra Vietgen

Unter Mitarbeit von:

Kolja Briedis

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH

Forschungsdatenzentrum

Lange Laube 12 | 30159 Hannover | <https://fdz.dzhw.eu/>

Mai 2018

## Inhaltsverzeichnis

Abbildungsverzeichnis .....	IV
Tabellenverzeichnis .....	IV
I Einleitung .....	1
II Übersicht zum DZHW-Promoviertenpanel 2014.....	2
III Datennutzungshinweise .....	4
1 Inhalt und Anlage der Studie.....	8
2 Erhebungsinstrumente.....	9
2.1 Inhalte der Erhebungsinstrumente .....	9
2.2 Pretest.....	11
3 Grundgesamtheit und Kontaktierung .....	12
4 Durchführung .....	13
5 Rücklauf.....	15
6 Datenaufbereitung.....	18
6.1 Datenübertragung.....	18
6.2 Codierung offener Angaben.....	18
6.3 Datenprüfung und Datenbereinigung.....	19
6.4 Generierung von Variablen .....	20
6.5 Erstellung der Datensätze .....	21
6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels .....	21
6.7 Codierung fehlender Werte .....	23
7 Gewichtung .....	25
7.1 Vorgehen und Anwendungshinweise .....	25
7.2 Gewichtung des Datensatzes .....	25
8 Anonymisierung .....	26
9 Literaturverzeichnis.....	32
Anhang 1: Logit-Regression zur Erstellung des Panelausfallgewichts in Welle 2 .....	34

## Abbildungsverzeichnis

Abbildung 1: Datenzugangswege und Analysepotential .....	5
Abbildung 2: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, 1. Welle .....	16
Abbildung 3: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, 2. Welle .....	17
Abbildung 4: Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten des DZHW- Promoviertenpanels 2014 .....	28

## Tabellenverzeichnis

Tabelle 1: Thematischer Aufbau der Erstbefragung .....	9
Tabelle 2: Thematischer Aufbau der Zweitbefragung.....	10
Tabelle 3: Beteiligung der Hochschulen.....	15
Tabelle 4: Brutto- und Nettorücklaufquote des DZHW-Promoviertenpanels 2014 .....	15
Tabelle 5: Vercodete Merkmale und verwendete Codierlisten im DZHW-Promoviertenpanel 2014	19
Tabelle 6: Themengebiete und Kürzel für Variablennamen des DZHW-Promoviertenpanels 2014.....	22
Tabelle 7: Systematik des FDZ-DZHW für fehlende Werte .....	24
Tabelle 8: Gewichtungsvariablen im Scientific-Use-File .....	25
Tabelle 9: Maßnahmen der statistischen Anonymisierung der Daten des DZHW-Promoviertenpanels 2014 nach Zugangsweg.....	29

## I Einleitung

Das DZHW-Promoviertenpanel mit dem Projekttitel „Karrieren Promovierter“ ist eine Untersuchung der Promotionsbedingungen, des beruflichen Übergangs nach der Promotion sowie des weiteren Berufsverlaufs von Promovierten.<sup>1</sup> Es wird vom Deutschen Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)<sup>2</sup> durchgeführt, vom Bundesministerium für Bildung und Forschung (BMBF) gefördert und dient – in Ergänzung zur amtlichen Hochschulstatistik – dem nationalen Bildungsmonitoring.

Im Rahmen des vom BMBF geförderten Projektes zum Aufbau eines Forschungsdatenzentrums für Hochschul- und Wissenschaftsforschung am DZHW (FDZ-DZHW) werden die Daten in Zusammenarbeit mit den Projektmitarbeiter(inne)n des Promoviertenpanels 2014 nachträglich zum Zweck der Datennachnutzung aufbereitet und dokumentiert. 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.<sup>3</sup>

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation zur ersten und zweiten Befragungswelle des Promoviertenpanels 2014 (doi: 10.21249/DZHW:phd2014:2.0.0).<sup>4</sup> Weitere Dokumentationsmaterialien zur Studie (Datensatzreports, Fragebögen, Filterführungsdiagramme etc.) können frei im Metadaten- und Dokumentensystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden. Abschnitt II dieses Berichts stellt die Eckdaten des Promoviertenpanels 2014 im Überblick dar. Die zentralen Informationen zur Nutzung der Daten dieser Studie folgen in Abschnitt III. Kapitel 1 stellt Inhalt und Anlage des Promoviertenpanels im Allgemeinen 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.

---

<sup>1</sup> Aktuelle Informationen zum Promoviertenpanel können über die Website des Projektes ([www.promoviertenpanel.de](http://www.promoviertenpanel.de)) abgerufen werden.

<sup>2</sup> Das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW, <http://www.dzhw.eu>) entstand im August 2013 durch eine Ausgründung aus der HIS Hochschul-Informationen-System GmbH.

<sup>3</sup> Informationen zu verfügbaren Datensätzen und Dokumentationen werden auf der Website des FDZ-DZHW (<https://fdz.dzhw.eu>) zur Verfügung gestellt.

<sup>4</sup> Die Daten der dritten Welle, die im Jahr 2017 durchgeführt wird, sowie die Daten der zwei weiteren geplanten Wellen können erst zu einem späteren Zeitpunkt veröffentlicht werden.

## II Übersicht zum DZHW-Promoviertenpanel 2014

<b>Studie</b>	DZHW-Promoviertenpanel
<b>Kohorte</b>	Promovierte des Prüfungsjahres 2013/2014
<b>Erhebende Institution</b>	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)
<b>Förderung</b>	Bundesministerium für Bildung und Forschung (BMBF)
<b>Projektmitarbeiter(innen) (Projektleitung)</b>	Gesche Brandt, <a href="#">Kolja Briedis</a> , Susanne de Vogel, Steffen Jaksztat, Carola Teichmann
<b>Themen</b>	Bildungs- und Beschäftigungsbiografie vor Beginn der Promotion Beruflicher Werdegang seit Beendigung der Promotion Rahmenbedingungen der Promotion Promotionsergebnis Wissenschaftliche und berufliche Weiterqualifikation Wissenschaftliche Aktivitäten Mobilitätserfahrungen Ziele, Motive und Persönlichkeit Soziale Herkunft
<b>Erhebungsdesign</b>	Panel-Design
<b>Grundgesamtheit</b>	Personen, die im Wintersemester 2013/2014 oder im Sommersemester 2014 eine Promotion an einer Hochschule mit Promotionsrecht in der Bundesrepublik Deutschland abgeschlossen haben
<b>Stichprobe</b>	Vollerhebung
<b>Erhebungsmethode</b>	1. Welle: standardisierte postalische Befragung 2. Welle: standardisierte Online-Befragung
<b>Erhebungszeitraum</b>	1. Welle: 15.12.2014 bis 17.02.2016 2. Welle: 15.03.2016 bis 18.04.2016
<b>Auswertbare Fälle</b>	1. Welle: n = 5.410 2. Welle: n = 3.184
<b>Rücklaufquote</b>	1. Welle: 27,2 % 2. Welle: 66,1 %
<b>Datenprodukte und Zugangswege</b>	CUF: Download SUF: Download, Remote-Desktop, On-Site
<b>Datensatzstruktur</b>	Personendatensatz im wide-Format Episodendatensatz im long-Format

---

**DOI** 10.21249/DZHW:phd2014:2.0.0

---

**Weitere Informationen** <https://fdz.dzhw.eu>

---

**Projektpublikationen\***

Brandt, G., de Vogel, S., & Jaksztat, S. (2016). Entwicklung und Testung eines Instruments zur Erfassung der Lernumwelt in der Promotionsphase. Ergebnisse der Entwicklungsstudie. Werkstattbericht. DZHW: Hannover.

\* Alle Projektpublikationen werden auf der Website des Projektes ([www.dzhw.eu/promovierte](http://www.dzhw.eu/promovierte)) zum Download bereitgestellt.

---

### III Datennutzungshinweise

**[Voraussetzungen der Datennutzung]** Die Daten des Promoviertenpanels 2014 werden durch das FDZ-DZHW entsprechend dem Bundesdatenschutzgesetz (vgl. § 40 Abs. 1 und Abs. 2 BDSG) ausschließlich zur wissenschaftlichen Nutzung und anonymisiert bereitgestellt.<sup>5</sup> Das FDZ bietet ein *Scientific Use File* (SUF) für die wissenschaftliche Sekundärnutzung und ein *Campus Use File* (CUF) für Lehr- und Übungszwecke an.

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

**[Datenzugang]** Das CUF des Promoviertenpanels 2014 kann nach Bereitstellung am lokalen Computer genutzt werden. Das SUF wird über drei Zugangswege angeboten, die hinsichtlich des Speicherortes, der Möglichkeit der eigenständigen Verknüpfung mit externen Daten und der Kontrollmöglichkeiten des FDZ unterschiedlich restriktiv sind.

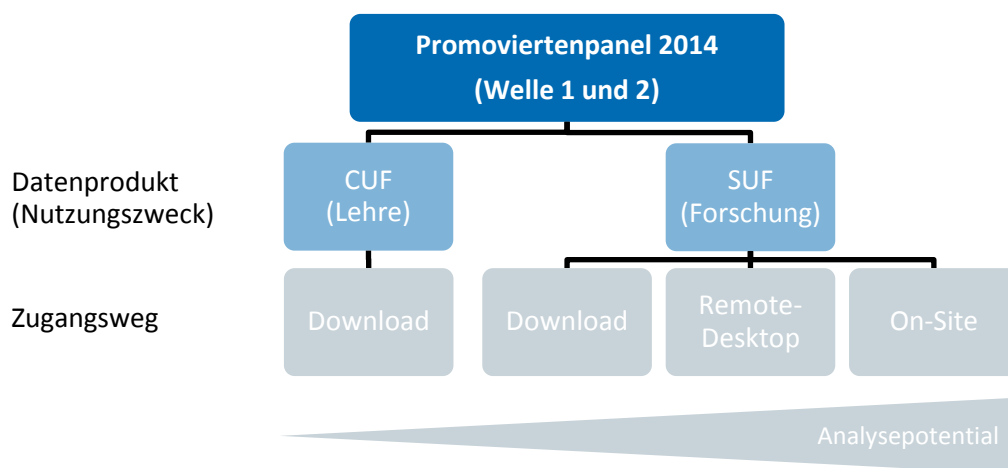
- **Download:** Die Daten werden verschlüsselt per E-Mail versandt oder auf der Website des FDZ zum Download bereitgestellt. Datennutzer(innen) können die Daten auf ihrem lokalen Computer speichern, falls gewünscht selbst mit Daten aus externen Quellen verknüpfen und die Daten mit eigener Software analysieren.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

<sup>5</sup> Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, 6ff.), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (LifBi) (vgl. Koberg, 2016, 699ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner, Müller, Schmucker & Roß, 2012, 9f.) orientieren. Das FDZ-DZHW hat den Portfolio-Ansatz an die Anforderungen der eigenen Datenbestände angepasst und nutzt vier Kategorien von Maßnahmen zur Sicherstellung des Datenschutzes, die in unterschiedlicher Weise kombiniert werden: Rechtlich-institutionelle Maßnahmen, informationelle Maßnahmen, technische Maßnahmen und statistische Maßnahmen.



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

**Abbildung 1: Datenzugangswege und Analysepotential**



**[Datenprodukte]** Über den *Digital Object Identifier* (DOI) 10.21249/DZHW:phd2014:2.0.0 ist eine Webseite mit zentralen Informationen zur Studie, weiteren Dokumentationsmaterialien sowie einer Übersicht der zur Verfügung stehenden Datenprodukte zur Studie erreichbar.

Die bereitgestellten Daten des Promoviertenpanels 2014 sind in zwei Datensätzen abgelegt. Es liegen ein Personendatensatz im wide-Format und ein Episodendatensatz im long-Format vor (vgl. Kapitel 6.5). Für SUF und CUF<sup>7</sup> werden für jeden im FDZ-DZHW angebotenen Zugangsweg beide Datensätze – jeweils mit zugangswegspezifischem Analysepotential (vgl. Abbildung 1) – bereitgestellt.

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

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

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

<sup>7</sup> Aus Anonymisierungsgründen werden dabei jedoch nur die Daten einer Substichprobe bereitgestellt (vgl. Kapitel 8).

**[Pflichten der Datennutzer(innen)]** Die Datennutzer(innen) sind verpflichtet, folgende Regeln<sup>8</sup> einzuhalten:

- **Wissenschaftliche Nutzung:** Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.
- **De-Anonymisierungsverbot:** Jeder Versuch der Re-Identifikation von Analyseeinheiten (z. B. Personen, Haushalten, Institutionen) ist verboten.
- **Gebot zur Mitteilung von Sicherheitslücken:** Falls Datennutzer(innen) Kenntnis von Sicherheitslücken hinsichtlich Datenschutz bzw. Datensicherheit erlangen, müssen diese dem FDZ-DZHW unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** SUF dürfen nur durch die Person genutzt werden, die den Datennutzungsvertrag abgeschlossen hat. CUF dürfen ausschließlich im Rahmen der angegebenen Lehrveranstaltung weitergegeben werden.
- **Löschungsgebot:** Download-SUF sind nach Ablauf der vereinbarten Nutzungsdauer (in der Regel drei Jahre) von alljeglichen 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-DZHW hervorgeht, muss dem FDZ im Voraus gemeldet und nach Veröffentlichung unverzüglich in elektronischer Form zur Verfügung gestellt werden. Informationen zu bereits vorhandenen Veröffentlichungen können dem Metadaten-suchsystem<sup>9</sup> entnommen werden.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen laut folgenden Vorgaben zitiert werden.

#### [Zitation]

- **Verwendeter Datensatz:**  
Brandt, G., Briedis, K., de Vogel, S., Jaksztat, S. & Teichmann, C. (2016). *DZHW-Promoviertenpanel 2014*. Aufbereitet durch Lange, K., Scheller, P. & Vietgen, S., doi: 10.21249/DZHW:phd2014:2.0.0, DATENSATZNAME<sup>10</sup>, released 2018. Hannover: FDZ-DZHW.
- **Daten- und Methodenbericht:**  
Brandt, G., de Vogel, S., Jaksztat, S., Teichmann, C., Lange, K., Scheller, P. & Vietgen, S. (2018). *DZHW-Promoviertenpanel 2014. Daten- und Methodenbericht zu den Erhebungen der Promoviertenkohorte 2014 (1. und 2. Befragungswelle)*. Hannover: FDZ-DZHW.

<sup>8</sup> Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

<sup>9</sup> <https://metadata.fdz.dzhw.eu>

<sup>10</sup> An dieser Stelle bitte den genauen Namen der verwendeten Datensatzversion angeben, z. B. phd2014\_p\_d\_2-0-0 für das Download-SUF des Personendatensatzes der Promoviertenkohorte 2014.

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

*Diese Arbeit nutzt die Daten des DZHW-Promoviertenpanels 2014. Die Daten sind vom Forschungsdatenzentrum des Deutschen Zentrums für Hochschul- und Wissenschaftsforschung (FDZ-DZHW) unter der doi: 10.21249/DZHW:phd2014:2.0.0 veröffentlicht worden.<sup>11</sup>*

---

<sup>11</sup> In englischsprachigen Publikationen: This scientific work uses data of the DZHW PhD Panel 2014, conducted by the German Center for Higher Education Research and Science Studies (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung; DZHW). The data were published by the research data center of the DZHW, doi: 10.21249/DZHW:phd2014:2.0.0.

## 1 Inhalt und Anlage der Studie

Das DZHW-Promoviertenpanel startete im Jahr 2013 im Rahmen der Förderlinie „Forschung zum wissenschaftlichen Nachwuchs“ (FoWiN) des Bundesministeriums für Bildung und Forschung. Im Rahmen der Studie wird untersucht, welche Einflüsse die formalen Promotionskontexte und die konkreten Lern- und Entwicklungsbedingungen, die Promovierte während ihrer Promotionsphase vorgefunden haben, auf den beruflichen Übergang nach der Promotion und auf den weiteren Berufsverlauf innerhalb und außerhalb der Wissenschaft ausüben.

Die Grundgesamtheit der Befragung umfasst alle 28.147 Personen (Statistisches Bundesamt, 2015), die im Prüfungsjahr 2014 in Deutschland an einer Hochschule mit Promotionsrecht eine Promotion abgeschlossen haben. Die Befragung ist als Vollerhebung angelegt, um eine ausreichend große Fallzahl zu erhalten. D. h. es erfolgte keine Stichprobenziehung, sondern alle Promovierten des Jahrgangs wurden zur Teilnahme an der Erstbefragung eingeladen.

Die Auftaktbefragung fand im Jahr 2015 statt; hierzu wurde ein Paper-Pencil-Fragebogen circa ein halbes bis anderthalb Jahre nach Abschluss der Promotion postalisch versandt. Die Zweitbefragung des Prüfungsjahrgangs erfolgte etwa ein Jahr später als Onlinebefragung. Drei weitere Wellen sind geplant.

Neben Grunddaten für die Bildungsberichterstattung enthält der Datensatz detaillierte Informationen über die Lern- und Entwicklungsbedingungen während der Promotionsphase sowie über die Lebensläufe Promovierter im Anschluss an die Promotion. Darüber hinaus beinhaltet der Datensatz eine Reihe von Persönlichkeitsmerkmalen (*Big Five*, Selbstwirksamkeit, interne/externe Kontrollüberzeugungen) sowie sozio- und bildungsbiografische Hintergrundinformationen. Dadurch bietet sich ein großes, bisher nicht vorhandenes Analysepotential für die Hochschul- und Wissenschaftsforschung. Das Paneldesign und die Erhebung von monatsgenauen Verlaufsdaten ermöglichen Kausalanalysen auf der Individualebene im Zeitverlauf (z. B. in Form von Ereignisdaten- und Sequenzmusteranalysen).

## 2 Erhebungsinstrumente

In der ersten Befragungswelle des Promoviertenpanels 2014 wurde als Erhebungsinstrument ein standardisierter Papierfragebogen in deutscher Sprache eingesetzt.<sup>12</sup> Die zweite Befragung wurde über einen standardisierten Online-Fragebogen in deutscher Sprache realisiert.<sup>13</sup> Die Erhebungsinstrumente umfassen einerseits Kernelemente, die in jeder Welle erneut abgefragt werden, hierzu gehört ein Tableau zur monatsgenauen Erfassung wichtiger beruflicher Episoden, Merkmale beruflicher und wissenschaftlicher Tätigkeiten und Informationen zur familiären Situation. Andererseits enthalten die einzelnen Erhebungswellen individuelle Schwerpunktsetzungen, wie etwa die retrospektive Erfassung der Promotionsphase in der ersten Welle.

### 2.1 Inhalte der Erhebungsinstrumente

Der Fokus der Erstbefragung lag auf der Lernumwelt und den Rahmenbedingungen der Promotion sowie den wissenschaftlichen Aktivitäten und praktischen Berufserfahrungen während der Promotionsphase (vgl. Tabelle 1).

**Tabelle 1: Thematischer Aufbau der Erstbefragung**

Thema	Fragennummern
Rahmenbedingungen /Angaben zur Promotionsphase	1.1 bis 1.21
Finanzierung	2.1 bis 2.3
Betreuung und Unterstützung	3.1 bis 3.10
Wissenschaftliche Aktivitäten	4.1 bis 4.10
Praktische Erfahrungen	5.1 bis 5.6
Eigenschaften, Ziele und Motive /Persönliche Merkmale/Einstellungen	6.1 bis 6.6
Berufliche Entwicklung	7.1 bis 7.5
Beschäftigung und Erwerbstätigkeit	2.3; 7; 5; 8.1 bis 8.13
Soziodemographische Merkmale	9.1 bis 10.4
Vorbildung/Hochschulzugang	9.9 bis 9.11
Soziale Herkunft	10.1 bis 10.4

Zur Erfassung der Lernumwelt in der Promotionsphase wurde ein theoretisch fundiertes und empirisch getestetes Modell entwickelt. Die theoretische Konzeption der Lernumwelten basiert auf dem SSCO-Modell (*Structure - Support - Challenge - Orientation*) von Bäumer, Preis, Roßbach, Stecher & Klieme, 2011. Ein weiteres Kernstück der Erstbefragung ist das Tätigkeitstableau (Frage 2.3), mit dem die Tätigkeiten seit Beginn der Promotion chronologisch erfasst wurden. Erfragt wurden Informationen zum Beginn und Ende der Tätigkeit, zur beruflichen Stellung, zur Arbeitszeit, zur Befristung, ggf. zu Arbeitgeberwechseln sowie zum Wissenschaftsbezug der Tätigkeit. Zusätzlich wurden in den Fragen 8.1 bis 8.13 weitere Angaben zur ersten Stelle nach

<sup>12</sup> Auf Wunsch konnte von den Befragten ein deutscher oder englischer PDF-Fragebogen per E-Mail angefordert werden.

<sup>13</sup> Die Fragebögen beider Wellen können von der Website des FDZ heruntergeladen werden. Ebenso stehen dort Filterführungsdiagramme beider Fragebögen zur Verfügung.

Abschluss der Promotion und zur aktuellen Beschäftigung erfragt (z. B. Beruf, Branche, Adäquanz). Des Weiteren wurden psychometrische Kurzskalen zur Messung allgemeiner Selbstwirksamkeit (ASKU) (Beierlein, Kovaleva, Kemper & Rammstedt, 2014), der „Big Five-Persönlichkeitsfaktoren“ (BFI-10) (Rammstedt, Kemper, Klein, Beierlein & Kovaleva, 2014) sowie internaler und externaler Kontrollüberzeugungen (IE-4) (Kovaleva, Beierlein, Kemper & Rammstedt, 2014) nach Vorlagen und Richtlinien der GESIS in den Fragebogen integriert. Mit einer leicht abgewandelten Version der Skala „Allgemeine geografische Mobilitätsbereitschaft (modifizierte Fassung)“ (Otto, Glaser & Dalbert, 2004) wurde die Bereitschaft, berufsbedingt innerhalb Deutschlands umzuziehen oder für eine begrenzte Zeit ins Ausland zu gehen, erfasst. Einzelne Items wurden (teilweise in leicht abgewandelter Form) aus unterschiedlichen Befragungen zum Thema Promotion und Hochqualifizierte entnommen.<sup>14</sup>

Ein Teil der Fragen der Erstbefragung wurde in der zweiten Befragung modifiziert verwendet. Dies war aus mehreren Gründen erforderlich: Erstens fand ein Moduswechsel statt und nicht alle Fragen waren gleichermaßen für Paper-Pencil- und Onlinebefragungen geeignet. Zweitens war bei einigen Items eine Präzisierung der Fragestellung erforderlich. Drittens wurde in vielen Fragen der zeitliche Bezug geändert, um die Fortschreibung der Daten in späteren Panelwellen zu erleichtern.

In der zweiten Erhebung wurde des Weiteren eine Reihe von Informationen zum Bildungsverlauf retrospektiv erhoben, die in der ersten Befragung aus Platzgründen nicht erhoben werden konnten (Tabelle 2).

**Tabelle 2: Thematischer Aufbau der Zweitbefragung**

Thema	Fragennummern
Promotionsergebnis	3.1 bis 3.4; 6.3 bis 6.4, 6.16 bis 6.17
Wissenschaftliche und berufliche Weiterqualifikation	5.10 bis 5.12; 6.10
Wissenschaftliche Aktivitäten	4.2 bis 4.7; 5.1 bis 5.10; 5.13 bis 5.19; 6.1 bis 6.3; 6.10
Bildungs- und Beschäftigungsbiografie vor Beginn der Promotion	1.1 bis 1.2; 2.1 bis 2.4
Beruflicher Werdegang seit Beendigung der Promotion	4.1; 4.8 bis 4.9; 6.7 bis 6.20
Mobilität	1.3; 1.6; 2.5 bis 2.6; 6.21 bis 6.24
Personenmerkmale/private Lebenssituation	6.25 bis 6.34

Der Fragebogen zur Zweitbefragung beinhaltete zur Erfassung der Tätigkeitsverläufe im Anschluss an die Promotion zwei Instrumente: Wie in der ersten Welle kam ein Berufstableau zur monatsgenauen Erfassung wichtiger beruflicher Episoden und zur Abbildung detaillierter Merkmale der jeweiligen Episode zum Einsatz. Das Tableau wurde noch um die Frage nach dem Ar-

<sup>14</sup> Etwa Bundesamt für Statistik [BFS] (2011) (Welle 1, Frage 2.1), Jungbauer-Gans und Gross (2013) (Welle 1, Frage 3.3), Grünh, Hecht, Rubelt und Schmidt (2009) (Welle 1, Frage 4.1), Auriol, Felix und Schaaper (2012) (Welle 1, Frage 5.4), Blickle, Kuhnert und Rieck (2003) (Welle 1, Welle 7.1), Egel, Gottschalk, Rammer und Spielkamp (2003) (Welle 2, Frage 6.3), Statistisches Bundesamt (2013) (Welle 2, Frage 6.4). Das Institut für Gründung und Innovation der Universität Potsdam [BIEM-CEIP] (2010) (Welle 1, Frage 4.10) sowie Wissenschaftliches Zentrum für Berufs- und Hochschulforschung [Universität Kassel] (2009) (Welle 1, Frage 7.2).

beitsort ergänzt. Zusätzlich wurde ein Kalendarium in den Fragebogen der zweiten Welle integriert, mit dem berufliche und nicht-berufliche Tätigkeiten nach der Promotion monatsgenau erfasst wurden (z. B. weitere Ausbildungen, Stipendien, Elternzeit, Familienarbeit, Arbeitslosigkeit).

### 2.2 Pretest

Vor der Erstbefragung wurden zunächst ein kognitiver und später ein quantitativer Pretest durchgeführt. Hier sollten vor allem die neu entwickelten Items zur Erhebung der Lernumwelt in der Promotionsphase getestet und die Anwendbarkeit für die Befragung von Promovierenden und Promovierten verschiedener Fächer und Promotionsformen geprüft werden. Das Ziel der Pretests war es, das Frageverständnis und das Antwortverhalten der Promovierenden und Promovierten zu eruieren, die Beantwortungsdauer zu ermitteln und etwaige Reihenfolgeeffekte aufzudecken. Ausführlich ist dies in einem Werkstattbericht dokumentiert (Brandt, Vogel & Jakszat, 2016).

Auch der Fragebogen der Zweitbefragung wurde vorab im Rahmen eines kognitiven Pretests mit Promovierten unterschiedlicher formaler Promotionsformen und Fachbereiche getestet. Die Programmierung des Fragebogens und die Durchführung der Befragung wurden unter Verwendung der DZHW-Onlinebefragungssoftware „Zofar“ durchgeführt.

### 3 Grundgesamtheit und Kontaktierung

Um ein möglichst repräsentatives Bild und aussagekräftige Schlussfolgerungen über die Promotionsbedingungen, die berufliche und private Entwicklung von Promovierten ziehen zu können und fach- sowie kontextspezifische Analysen zu ermöglichen, wurde die Erhebung als Vollerhebung konzipiert. Die Grundgesamtheit umfasst alle Promovierten, die im Prüfungsjahr 2014 (Wintersemester 2013/14 und Sommersemester 2014) an einer deutschen Hochschule mit Promotionsrecht eine Promotion abgeschlossen haben. Die amtliche Statistik weist für das betreffende Prüfungsjahr 2014 bundesweit 28.147 Promovierte aus (Statistisches Bundesamt, 2015).

Aus Datenschutzgründen mussten die Erstsprache der Promovierten und der Versand der Befragungsunterlagen über die Hochschulen stattfinden (Adressmittlungsverfahren). Im Vorfeld wurden die Hochschulleitungen aller 146 Hochschulen mit Promotionsrecht über die geplante Erhebung informiert und um Unterstützung des Forschungsvorhabens gebeten. Sofern die Hochschulleitung keine Absage erteilte, erfolgten Anfragen an die Stellen, die innerhalb der Hochschulen für die Verwaltung der Promotionsakten zuständig waren.



## 4 Durchführung

**[Kontaktaufnahme und Adresspflege]** Der Versand der Fragebögen für die Erstbefragung fand aus Datenschutzgründen über die Dekanate bzw. die für Promotionen zuständigen Prüfungsämter statt. In den meisten Fällen lag die Verantwortlichkeit innerhalb der einzelnen Dekanate. An wenigen Hochschulen wurden zentrale bzw. dezentrale Prüfungsämter oder Graduiertenakademien als zuständige Stellen identifiziert.

Um die weiterhin teilnahmebereiten Personen in der zweiten Befragungswelle direkt durch das DZHW kontaktieren zu können, wurden im Fragebogen der ersten Welle deren Kontaktdaten erfasst. Beim Eingang eines Fragebogens im DZHW wurde sowohl auf den Fragebogen als auch auf den Adressabschnitt des Fragebogens per Paginierstempel eine eindeutige Identifikationsnummer gestempelt und aus allen Adressabschnitten eine Referenzliste von der Identifikationsnummer zur jeweils zugehörigen Adresse erstellt.<sup>15</sup> Um dabei auch Personen zu berücksichtigen, die in der Zwischenzeit umgezogen waren, wurden die Adressbestände zwischen den Wellen geprüft und ggf. aktualisiert.

Zwischen der Erstbefragung und der Zweitbefragungswelle wurde eine Kontaktierung durchgeführt. Alle Befragten, die eine E-Mailadresse angegeben haben, wurden per E-Mail über den Stand des Projektes informiert und ihnen wurde für die bisherige Teilnahme gedankt. Darüber hinaus wurde in dem Anschreiben die anstehende Zweitbefragung angekündigt. Durch die Zwischenkontaktierung konnte der Bestand an E-Mailadressen für die Zweitbefragungswelle auf Aktualität und mögliche Erfassungsfehler überprüft werden. Sofern keine (gültige) E-Mailadresse vorlag, erfolgte der Zwischenkontakt postalisch. Mit einer beigefügten Postkarte wurden die 194 Panelteilnehmer(innen) ohne (gültige) E-Mailadresse gebeten, kostenfrei ihre aktuelle E-Mail- und Postadresse mitzuteilen. Auf diese Weise konnten die E-Mailadressen von 68 Befragten gewonnen werden. Eine weitere E-Mailadresse konnte durch eine Namenskorrektur nach einer Anschriftenberichtigung der Deutschen Post dazugewonnen werden. So konnten insgesamt 4.822 Adressdaten in den Panelbestand eingepflegt werden.<sup>16</sup>

Die Zweitbefragung wurde als Onlineerhebung durchgeführt. Hierfür konnte auf den aktualisierten Adressbestand von 4.816 Befragten aus der Erstbefragung, die ihre Kontaktdaten hinterlassen und sich zur Teilnahme an Folgebefragungen bereit erklärt haben, zurückgegriffen werden.

**[Erhebungsunterlagen]** Die Erhebungsunterlagen bestanden in der ersten Befragungswelle pro zu befragender Person aus einem Anschreiben (inkl. Datenschutzinformationen), dem Fragebogen, einem Begleitschreiben des Bundesministeriums für Bildung und Forschung (BMBF) und einem an das DZHW adressierten portofreien Umschlag zur Rücksendung des ausgefüllten Fragebogens. Zudem wurden drei Erinnerungsschreiben verschickt. Die Befragung der zweiten Welle wurde mittels der DZHW-Onlinebefragungssoftware „Zofar“ programmiert. Die drei Erinnerungsschreiben zur Zweitbefragung wurden gezielt nur an diejenigen verschickt, die sich noch nicht an der zweiten Erhebung beteiligt hatten.

**[Feldphase]** Der Erhebungszeitraum der ersten Befragungswelle erstreckte sich vom 15. Dezember 2014 bis zum 17. Februar 2016. Die beiden Erinnerungsschreiben wurden etwa vier bzw.

<sup>15</sup> Zur Gewährleistung des Datenschutzes wurde der Adressabschnitt vom Fragebogen abgetrennt und die Referenzliste getrennt von den Befragungsdaten auf einem geschützten Server gespeichert.

<sup>16</sup> Davon haben vier Befragte in der Erstbefragung keine inhaltlichen Angaben gemacht und weitere sechs Befragte haben die Teilnahme an der Zweitbefragung vorab verweigert.

acht Wochen nach Feldstart verschickt.<sup>17</sup> Dieser vergleichsweise lange Zeitraum ist dadurch begründet, dass die Verwaltungen der Hochschulen unterschiedlich schnell auf die Anfrage des Projektes reagiert haben. Nicht alle Verwaltungsstellen konnten für den Versand der Erinnerungen gewonnen werden.<sup>18</sup> Bei dem Großteil der Verwaltungsstellen konnte der Versand der Befragungsunterlagen und der zwei Erinnerungsschreiben jedoch planmäßig durchgeführt werden. Aufgrund des angewendeten Kontaktverfahrens über die Prüfungsämter konnte das DZHW keinen direkten Einfluss auf den genauen Versandzeitpunkt der Erhebungsunterlagen nehmen.

Für die zweite Welle konnte durch den aus der ersten Welle dem DZHW vorliegende Adressbestand ein konkreter Termin zur Einladungen zur Befragung festgelegt werden. Die Befragung der zweiten Welle war vom 15. März 2016 bis zum 18. April 2016 für die Befragten erreichbar.

**[Rücklaufsteigernde Maßnahmen]** Die rücklaufsteigernden Maßnahmen zielten zum einen auf Anreize für die Hochschulen, die Befragung organisatorisch zu unterstützen und zum anderen auf die individuelle Motivation der Befragten. Als Dank für die Unterstützung erhielten die Hochschulen exklusive Vorabergebnisse aus der Befragung. Die Fakultäten konnten auf Wunsch Sonderauswertungen für die eigene Hochschule erhalten, sofern dafür ausreichende Fallzahlen vorlagen. Die auf die individuelle Motivation der Befragten zielenden rücklaufsteigernden Maßnahmen bestanden darin, dass zunächst in dem Anschreiben an die Befragten die subjektive und gesellschaftliche Relevanz des Themas hervorgehoben wurde. Zusätzlich stellte das Begleitschreiben des Bundesministeriums für Bildung und Forschung (BMBF) in der ersten Welle die hochschulpolitische Relevanz der Studie dar und warb um Unterstützung. Des Weiteren erfolgte der Versand von Erinnerungsschreiben.<sup>19</sup> Mittels einer Rückmeldung ausgewählter Studienergebnisse<sup>20</sup> nach der ersten Welle sollte zum einen die Bindung der Studienteilnehmer(innen) an das Panel erhöht werden, um diese zur Teilnahme an Folgewellen zu motivieren. Zum anderen diente der Ergebnisversand zur erneuten Aktualisierung des Adressbestandes. Darüber hinaus wurden unter allen Teilnehmer(inne)n der ersten Welle ein Laptop, drei iPads sowie mehrere Büchergutscheine verlost. Unter allen Teilnehmenden der zweiten Welle wurden ein Smartphone sowie mehrere Reisegutscheine verlost.

<sup>17</sup> Vom Dezember 2014 bis zum Januar 2015 lagen zwischen dem Versand der Befragungsunterlagen und dem Versand der ersten Erinnerung aufgrund von Lieferungsverzögerungen zum Teil mehr als vier Wochen. Ein wesentlicher Grund dafür war eine kurzfristige Umstellung der Portokosten, die eine Nachfrankierung mit Sonderbriefmarken erforderlich machte.

<sup>18</sup> Gründe hierfür waren z. B. ein zu großer zeitlicher oder personeller Aufwand, längere Abwesenheiten des zuständigen Personals oder personelle Veränderungen der zuständigen Verwaltungsstelle.

<sup>19</sup> Aufgrund des angewendeten Kontaktverfahrens über die Prüfungsämter wurden die Erinnerungsschreiben in der ersten Welle 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. Die Nicht-Teilnehmenden der zweiten Welle wurden dreimal erinnert.

<sup>20</sup> Ein Kurzüberblick über die wichtigsten Befunde konnte in einem online bereitgestellten Flyer abgerufen werden. Ebenso werden detailliertere Informationen nach Fachgruppen zur Verfügung gestellt.

## 5 Rücklauf

**[Rücklauf]** Insgesamt fand die Befragung der Promovierten des Prüfungsjahrgangs 2014 an 112 Hochschulen mit Promotionsrecht statt. Der überwiegende Teil der Hochschulen unterstützte die Befragung durch eine zentrale Teilnahme aller Fakultäten (80 Hochschulen), 32 Hochschulen beteiligten sich mit einzelnen Fakultäten, 15 Hochschulen nahmen nicht teil (siehe Tabelle 3). An 19 Hochschulen mit Promotionsrecht fand keine Befragung statt, da dort im untersuchten Zeitraum keine Promotion abgeschlossen wurde.

**Tabelle 3: Beteiligung der Hochschulen**

Beteiligte	Anzahl
Hochschulen mit Promotionsrecht	146
Keine abgeschlossenen Promotionen im Prüfungsjahr 2014	19
Teilgenommen	112
vollständig teilgenommen	80
teilweise teilgenommen	32
nicht teilgenommen	15

Tabelle 4 zeigt die Anzahl der versandten Fragebögen sowie die Rücklaufquoten. Während in der ersten Welle rund 27 Prozent der eingeladenen Promovierten teilnahmen, konnten für die zweite Welle rund 66 Prozent der teilnahmebereiten Personen der ersten Welle gewonnen werden.

**Tabelle 4: Brutto- und Nettorücklaufquote des DZHW-Promoviertenpanels 2014**

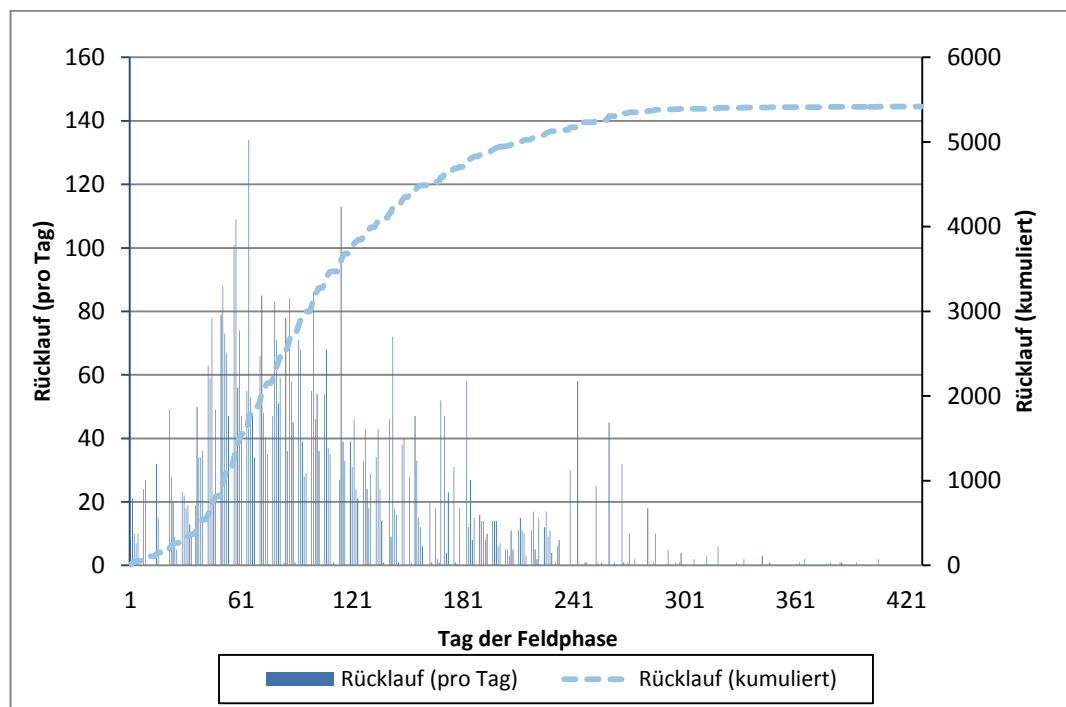
	Welle 1	Welle 2
Befragungsunterlagen versendet	19.916	4.816
Befragungsunterlagen Versand bestätigt <sup>21</sup>	19.900	4.816
Befragungsunterlagen Rücklauf	5.423	3.188
Befragungsunterlagen verwertbar	5.410	3.184
Rücklaufquote (verwertbar/Versand bestätigt)	27,2 %	66,1 %
Rücklaufquote (verwertbar/versendet)	27,2 %	66,1 %
Anteil Welle 2 (brutto) an Welle 1 (brutto)		24,2 %
Anteil Welle 2 (netto) an Welle 1 (netto)		58,8 %
Anteil Welle 2 (netto) an Welle 1 (brutto)	-	16,0 %

In der ersten Welle wurden 19.916 Fragebögen an 127 Hochschulen versandt. Davon wurden 5.423 Fragebögen von den Befragten zurückgesandt (vgl. Tabelle 4).<sup>22</sup> Abzüglich leerer, doppel-

<sup>21</sup> Da ein direkter Versand aus Datenschutzgründen nicht möglich war, muss der tatsächliche Versand von Befragungsunterlagen geschätzt werden. Als Grundlage für die Schätzung dient die Rückmeldung der Prüfungsämter über ein Formular, auf dem vermerkt wurde, wie viele Fragebögen tatsächlich versendet wurden.

ter und unlesbarer Fragebögen konnten 5.410 Fragebögen elektronisch erfasst werden (vgl. Kapitel 6.3). Abbildung 2 stellt den Rücklauf der Fragebögen in der Feldphase der ersten Befragungswelle im Zeitverlauf dar. Ein Großteil der ausgefüllten Fragebögen hat 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 2: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, 1. Welle**

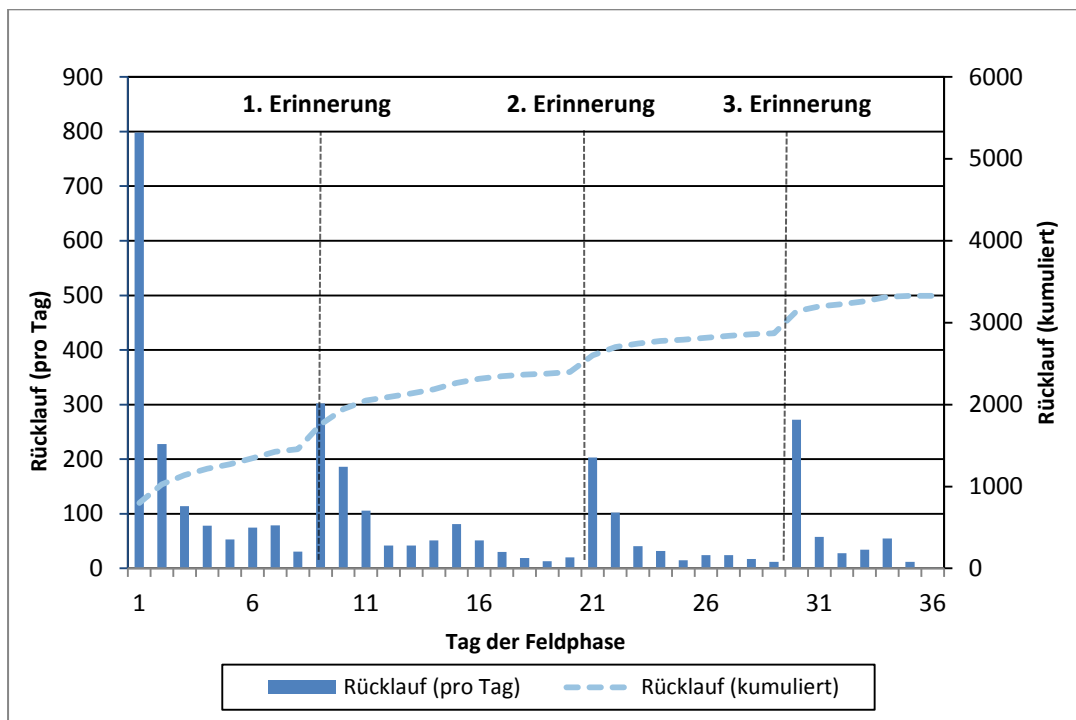


4.816 Personen, also 89 Prozent der 5.410 Teilnehmer(innen) der ersten Welle, erklärten sich mit einer Kontaktierung für weitere Befragungen einverstanden. An der zweiten Befragung beteiligten sich insgesamt 3.188 Promovierte, wovon 3.184 Befragungen auswertbar waren.<sup>23</sup> Die Rücklaufkurve in Abbildung 3 stellt den Rücklauf seit Beginn der Befragung und die Zeitpunkte der drei Erinnerungsaktionen dar. Am 15.03.2016 wurden 4.816 Befragte zur Teilnahme an der Folgebefragung (zweite Welle) eingeladen (abzüglich elf nicht zustellbaren Einladungen). Am ersten Tag der Befragung nahmen bereits 24 Prozent und in den ersten drei Tagen 35 Prozent der Befragten teil. Es zeigten sich deutliche Effekte der Erinnerungen, die jeweils einen erkennbaren Anstieg des Rücklaufs zur Folge hatten.

<sup>22</sup> Drei Fragebögen sind nach Beendigung der Datenerfassung eingegangen und konnten daher nicht mehr aufgenommen werden. Fünf Fragebögen sind unausgefüllt eingegangen, drei weitere mussten wegen stark unplausibler Angaben aussortiert werden. Zudem wurde ein zweifach ausgefüllter Fragebogen einer Person entfernt. Eine Person verlangte die Löschung seiner Befragungsdaten.

<sup>23</sup> Eine Person wurde als teilnehmend gewertet, wenn sie mindestens eine Frage beantwortet hatte.

Abbildung 3: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, 2. Welle



**[Panelausfälle]** Das Promoviertenpanel 2014 ist von paneltypischen Ausfallprozessen<sup>24</sup> betroffen. Hier sind die grundsätzliche Verweigerung der Teilnahme an Folgebefragungen (keine Angabe von Adressangaben für die Kontaktierung in der zweiten Welle) und die Nichtteilnahme nach einer (versuchten) Kontaktaufnahme in der zweiten Befragungswelle zu nennen.

Die Betrachtung der Rücklaufquote im Zeitverlauf zeigt, dass die Bruttostichprobe in der zweiten Welle nur noch 24 Prozent der versendeten Befragungsunterlagen der ersten Welle umfasst. Von den 5.410 verwertbaren Fällen aus der ersten Welle konnten 59 Prozent der Befragten in der zweiten Befragungswelle erneut befragt werden (vgl. Tabelle 4). Darüber hinaus haben nur 16 Prozent der in der ersten Welle angeschriebenen Promovierten an beiden Befragungswellen teilgenommen.

<sup>24</sup> Zu paneltypischen Ausfallprozessen vgl. Schnell, Hill & Esser, 2005, S. 241.

## 6 Datenaufbereitung

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

### 6.1 Datenübertragung

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

**[Erstellung eines Codeplans]** Auf Basis des Fragebogens der Befragung wurde ein Codeplan erstellt. Dabei wurde vermerkt, welcher Frage bzw. Teilfrage eine Variable zugeordnet ist und welchen Namen diese Variable trägt. Um die Erfassungsreihenfolge festzulegen, wurden die Variablen zusätzlich nummeriert.<sup>25</sup>

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

### 6.2 Codierung offener Angaben

Vor der Datenübertragung erfolgte eine Codierung der meisten (halb-)offenen Angaben. Dabei wurden diesen anhand einer Codierliste numerische Codierungen zugeordnet. Je nach Variable wurden unterschiedliche Codierlisten verwendet. Es handelt sich um Klassifikationsschlüssel der amtlichen Statistik (z. B. Klassifikation der Berufe, Schlüsselverzeichnis der Studenten- und Prüfungsstatistik etc.). Für einige Variablen wurden Codierlisten auf Basis der in den Daten des Promoviertenpanel 2014 vorkommenden Nennungen entwickelt. Für einige halboffene Fragen wurden keine neuen Variablen mit numerischen Codierungen erstellt, sondern die Nennungen nur – sofern möglich – den vorhandenen (geschlossenen) Antwortkategorien zugeordnet. Einzelne offene Fragen wurden nicht vercodet, weil sie vorwiegend als Kontextinformationen zur Codie-

<sup>25</sup> Die Daten wurden in einem einfachen spaltenorientierten Textformat ohne eine die Variablennamen enthaltene Kopfzeile erfasst. Der Codeplan musste daher festlegen, in welcher Reihenfolge die Daten zu erfassen waren, damit die zu einer Variable zugehörigen Daten in der richtigen Spalte eingetragen werden konnten.

nung anderer offener Angaben erfasst wurden.<sup>26</sup> Die durch das Primärforschungsprojekt vorgenommenen Codierungsentscheidungen wurden unverändert beibehalten.

In Tabelle 5 sind die codierten Merkmale sowie die jeweils verwendete Codierliste dargestellt. Der Datensatz beinhaltet ausschließlich die codierten numerischen Variablen, die offenen Nennungen selbst sind nicht im Datensatz enthalten. Die Ausprägungen der einzelnen Variablen sind im Datensatzreport sowie im Metadatensuchsystem<sup>27</sup> dokumentiert.

**Tabelle 5: Vercodete Merkmale und verwendete Codierlisten im DZHW-Promoviertenpanel 2014**

Merkmal	Codierliste
Promotionshochschule, Hochschule des Studienabschlusses	Deutsche Hochschule: Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2013/2014 und SoSe 2014) Ausländische Hochschule: DZHW-Codierung auf Basis der Ländercodes
Promotionsfach, Studienfach	Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2015/2016 und SoSe 2016)
Beruf, Berufe der Eltern	Destatis Klassifikation der Berufe 2010
Ort der beruflichen Tätigkeit	Zuordnung der Postleitzahlen
Land der beruflichen Tätigkeit	Staats- und Gebietssystematik 2014
Geburtsort, Ort des Erwerbs der Studienberechtigung	Zuordnung der Postleitzahlen
Geburtsland, Land des Erwerbs der Studienberechtigung	Staats- und Gebietssystematik 2014
Studienabschluss	DZHW-Codierung auf Basis des Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2013/2014 und SoSe 2014)
sonstige offene Abfragen	Zuordnung zu vorgegebenen Kategorien oder projekteigene Codierung

### 6.3 Datenprüfung und Datenbereinigung

**[Manuelle Vorkorrektur]** In der ersten Welle wurde bereits vor der Übertragung der Daten eine manuelle Prüfung und gegebenenfalls eine Anpassung von Angaben der Befragten auf den Papierfragebögen durchgeführt.<sup>28</sup> Dies sollte zum einen die Erfassbarkeit der Daten erleichtern. Dafür wurde die Form der bestehenden Angaben verändert. Beispielsweise wurden schwer lesbare Angaben oder Streichungen der Befragten verdeutlicht, Zahlenangaben rechtsbündig in die dafür vorgesehenen Kästchen eingetragen, verbale Angaben von Noten in Ziffern übersetzt (z. B. „gut“ = 2,0).

<sup>26</sup> Dies betrifft in beiden Wellen die beruflichen Tätigkeitsbereiche und Aufgaben, die neben der Berufsbezeichnung in Frage 8.4 und 8.5 in der ersten Welle sowie in Frage 6.12 in der zweiten Welle erhoben wurden. Die Angaben dienen nur dazu, zusätzliche Informationen für die Codierung der ebenfalls offen abgefragten Berufsbezeichnung zu erhalten.

<sup>27</sup> <https://metadata.fdz.dzhw.eu/>

<sup>28</sup> Die Zahl der vorgenommenen Korrekturen wurde nicht zentral, sondern nur auf den Papierfragebögen dokumentiert und ist daher nicht mehr systematisch rekonstruierbar.

Zum anderen zielte die manuelle Prüfung darauf ab, schon vor der softwaregestützten Korrektur (siehe unten) erste Fehler oder Inkonsistenzen in den Angaben der Befragten zu bereinigen. Beispielsweise musste die Angabe „keine“ gestrichen werden, wenn ein oder mehr Aufsätze in einer Zeile bei „Anzahl insgesamt“ genannt wurden. Ferner wurde geprüft, ob die Angaben über Berufstätigkeiten im Tätigkeitstableau mit den entsprechenden Angaben zur Erwerbstätigkeit übereinstimmten (Fragen 2.1 und 8.1-8.13) Festgestellte Inkonsistenzen wurden – falls möglich – durch den Abgleich mit anderen Nennungen im Fragebogen aufgelöst oder andernfalls ein entsprechender Missingcode (vgl. Kapitel 6.7) vergeben.

**[Softwaregestützte Korrektur]** Im Anschluss an die Datenübertragung in der ersten Welle erfolgte eine umfassende Prüfung und Korrektur der Daten mit Hilfe einer DZHW-eigenen Software und in der zweiten Welle mit Hilfe des Statistikprogrammes Stata. Dabei sollten zum einen Fehler bei der vorherigen manuellen Vorkorrektur und Datenübertragung, zum anderen weitere inkonsistente Angaben der Befragten, die bei der Vorkorrektur nicht geprüft werden konnten, identifiziert werden. Anhand formaler Regeln wurden gültige Wertebereiche und Antwortkombinationen definiert und geprüft. Folgende Typen von Prüfungen wurden vorgenommen:

- *Prüfung von Wertebereichen:* Es wurde geprüft, ob die erfasste Ausprägung einer Variablen in dem für diese Variable definierten Wertebereich lag.
- *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 Befragungswellen geprüft. Über die Merkmalskombinationen hinaus, die bereits bei der manuellen Vorkorrektur überprüft wurden, konnten hier auch komplexere Merkmalskombinationen überprüft werden.

Bei fehlenden, fehlerhaften oder unplausiblen Werten wurde zunächst mit Hilfe des Papierfragebogens geprüft, ob der entsprechende Wert falsch (bzw. nicht) übertragen worden war. Ansonsten wurde versucht, den korrekten Wert anhand anderer Angaben im Fragebogen zu erschließen. Im Zweifelsfall wurde ein spezifischer Missingcode vergeben (vgl. Kapitel 6.7). Fehlerkorrekturen wurden dokumentiert<sup>29</sup> und von mindestens einer weiteren Person geprüft.

**[Löschung von Fällen]** In beiden Wellen wurden Fälle aus dem Datensatz entfernt. Ein Fall wurde gelöscht, wenn er doppelt eingegangen war (ein Fall), wenn weniger als eine Frage beantwortet worden war (fünf Fälle) oder wenn zu viele Inkonsistenzen vorhanden waren (drei Fälle). Insgesamt wurden in der ersten Befragungswelle acht Fälle und in der zweiten Befragungswelle drei Fälle gelöscht.

## 6.4 Generierung von Variablen

Neben den Variablen, die die codierten Antworten der Befragten enthalten, beinhaltet der Datensatz des Promoviertenpanel 2014 auch generierte Variablen. Dabei handelt es sich zum einen

<sup>29</sup> Die Dokumentation der Fehlerkorrekturen erfolgte handschriftlich auf den Papierfragebögen und ist daher nicht systematisch rekonstruierbar.



um Variablen mit numerischen Codierungen von ursprünglich offenen Nennungen (vgl. Kapitel 6.2). Zum anderen wurden Variablen des Datenschutzes wegen verändert (vgl. Kapitel 8) und im Forschungsfeld häufiger benötigte Variablen aus den Werten einer oder mehrerer Quellvariablen generiert (z. B. Aggregation der Studienfächer zu Studienbereichen und Fächergruppen oder Ableitung von Hochschultyp und Hochschulort aus den Hochschulvariablen). Der Variablenname einer generierten Variablen ist im Datensatz durch das Suffix „\_g#“ gekennzeichnet. Eine Übersicht aller für das Promoviertenpanel 2014 generierten Variablen sowie eine detaillierte Dokumentation der einzelnen Variablen mit Angabe ihrer jeweiligen Ausprägungen und Berechnungsvorschriften findet sich im Datensatzreport sowie im Metadaten suchsystem.<sup>30</sup>

## 6.5 Erstellung der Datensätze

**[Zusammenführung der Wellen]** Die Daten der ersten und Zweitbefragungswelle 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 befragte 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 dem Kalendarium (Frage 4.8 der zweiten Welle). Für jede befragte Person werden eine oder mehrere Episoden gespeichert. Dabei ist eine Episode definiert als ein Zeitraum, in dem eine bestimmte Tätigkeitsart (z. B. Erwerbstätigkeit, Weiterbildung) 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 sind detaillierte Informationen jedoch in den entsprechenden Variablen des Personendatensatzes enthalten (Welle 1 Frage 2.3, Welle 2 Frage 4.9). Die Daten dieser Variablen können mit den Episodendaten verbunden werden. Das Zusammenführen von Personendatensatz und Episodendatensatz wird über die Identifikationsnummer der Person (Variable: *pid*) ermöglicht.

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

## 6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels

**[Variablen- und Wertelabelvergabe]** Für Variablen- und Wertelabels wurden Formulierungen des Fragebogens übernommen oder prägnante Kurzformen dieser Formulierungen gewählt. Dabei basieren die Variablenlabels in der Regel auf dem entsprechenden Fragetext. Grundlage

<sup>30</sup> <https://metadata.fdz.dzhw.eu/>

für die Wertelabels sind je nach Fragetyp die Texte der Antwortoptionen bzw. eine Kombination der Texte von Frage und Antwortoption. Bei generierten Variablen, denen bestimmte Klassifikationen zugrunde liegen, wurden für die Wertelabels die Bezeichnungen der Schlüssel der Klassifikation wortgetreu übernommen. Die Variablen- und Wertelabels liegen auf Deutsch und auf Englisch vor. Im SPSS-Format existiert für jede Sprache ein eigener Datensatz. Im Stata-Format wurden zweisprachige Labels im gleichen Datensatz hinterlegt.

**[Variablenbenennung im Personendatensatz]** Für die Variablenbenennung wurde im FDZ-DZHW eine einheitliche Benennungssystematik erstellt. Mit Ausnahme der Identifikatorvariablen *pid* sowie der Wellenvariablen *wave*<sup>31</sup> wurden die Variablennamen im Personendatensatz nach einem Präfix-Stamm-Suffix-Schema, das eine automatisierte Verarbeitung erleichtert, gebildet. Zudem liefern die Variablennamen Metainformationen zur entsprechenden Variable. Das Präfix der Variable enthält die Wellenkennung anhand eines Buchstabens. Im Stamm geht der Themenbereich, dem die Variable zugeordnet ist, aus einem dreistelligen englischen Buchstabenkürzel hervor. Tabelle 6 stellt die verschiedenen Themenbereiche des Promoviertenpanels 2014 sowie das zugehörige Kürzel für den Stamm des Variablennamens im Überblick dar. Das anhand eines Unterstriches vom Stamm abgetrennte Suffix enthält verschiedene Zusatzinformationen, wie die Kenntlichmachung von generierten Variablen sowie die verschiedenen Datenzugangswege.

Detaillierte Informationen zur Variablenbenennung im Promoviertenpanel 2014 befinden sich im Datensatzreport.

**Tabelle 6: Themengebiete und Kürzel für Variablennamen des DZHW-Promoviertenpanels 2014**

Themengebiete-Kürzel	Themengebiet (englisch)	Themengebiet (deutsch)
<b>sys</b>	system variables	Systemvariablen
<b>stu</b>	studies	Studium
<b>occ</b>	occupation	Beschäftigung
<b>ski</b>	skills	Fähigkeiten
<b>fvt</b>	further vocational training	Berufliche Fort- und Weiterbildung
<b>dem</b>	demographic information	demographische Informationen
<b>abr</b>	(experiences) abroad	Auslandserfahrung
<b>fin</b>	financing	Finanzierung
<b>fut</b>	future prospects	Zukunftsaussichten
<b>goa</b>	goals (occupational, life)	(Berufs- und) Lebensziele
<b>inc</b>	income	Einkommen
<b>job</b>	job	Jobs
<b>mot</b>	motives	(Tätigkeits-)Motive (für Studium / Ausbildung / Promotion / Erwerbstätigkeit)
<b>mov</b>	move	Regionale Mobilität/Umzüge
<b>net</b>	network	Netzwerk
<b>par</b>	Information about partner	Informationen über Partner
<b>sat</b>	satisfaction	(Berufs-)Zufriedenheit

<sup>31</sup> Diese enthält die Information, welche Fälle nur an der ersten Welle bzw. an beiden Wellen teilgenommen haben.

<b>sch</b>	school	Schulzeit
<b>sci</b>	scientific experiences	wissenschaftliche Aktivität
<b>voc</b>	Vocational training/education	(Berufs-)Ausbildung
<b>con</b>	conditions of doing a phd	Rahmenbedingungen der Promotion
<b>phd</b>	to do a phd	Promotion
<b>scc</b>	Structure-Support-Challenge-Orientation-Scale (SSCO)	Dimension zur Erfassung von Lernumwelten
<b>sel</b>	self-efficacy scale	allgemeine Selbstwirksamkeit Kurzskala
<b>bfi</b>	Big-Five-Inventory (short)	Kurzskala zur Erfassung der fünf wichtigsten Persönlichkeitsdimensionen
<b>wgt</b>	weights	Gewichtungsvariablen

**[Variablenbenennung im Episodendatensatz]** Die Variablen im Episodendatensatz sind die Identifikationsnummer der befragten Person (*pid*), die Identifikationsnummer der jeweiligen Episode (*eid*), die ausgeübte Tätigkeitsart (*status*) sowie Beginn und Ende des Episodenzeitraums, der über vier Variablen (Monat: *begin\_m* und *end\_m*; Jahr: *begin\_y*; *end\_y*) codiert wird.

## 6.7 Codierung fehlender Werte

Zur Codierung fehlender Werte wurde im FDZ-DZHW eine übergreifende Systematik erstellt, um über verschiedene Datensätze des DZHW hinweg eine einheitliche Missingcodierung gewährleisten zu können. Fehlende Angaben wurden dabei durch dreistellige negative Werte codiert. Tabelle 7 stellt die verwendete Missingsystematik dar. Die im Promoviertenpanel 2014 verwendeten Missingcodierungen sind hervorgehoben.

Sie lassen sich drei bis 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.<sup>32</sup>

<sup>32</sup> Eine vierte Gruppe würde spezielle Missingcodierungen umfassen, die im Rahmen der Datenaufbereitung eines konkreten Datensatzes nur für einzelne Items vergeben wurden.

Tabelle 7: Systematik des FDZ-DZHW für fehlende Werte

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

<sup>a</sup> Dieser Wert ist nur für Datensätze im Long-Format vergeben.

<sup>b</sup> Dieser Wert wird nur in gepoolten Datensätzen vergeben.

<sup>c</sup> Dieser Wert wird vergeben, wenn keinerlei Ursache rekonstruiert werden kann.

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

<sup>e</sup> 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.

## 7 Gewichtung

### 7.1 Vorgehen und Anwendungshinweise

Im Rahmen des Projektes wurden zwei getrennte Gewichtungsschritte durchgeführt. Im ersten Schritt wurde die Ausgangsstichprobe der ersten Welle möglichst optimal an die Grundgesamtheit angepasst. Das realisierte Sample der ersten Welle wurde mit der Grundgesamtheit der Promovierten des Prüfungsjahres 2014 verglichen. Der Vergleich zwischen dem realisierten Sample und der Grundgesamtheit zeigte, dass die Ausgangsstichprobe offenbar ein gutes Abbild der Zielpopulation darstellt. Abweichungen zwischen Sample und Grundgesamtheit konnten durch die Berechnung von Querschnittsgewichten (Redressment) ausgeglichen werden (siehe Abschnitt 7.2). Im Rahmen von Ausfallanalysen auf Hochschulebene wurde zusätzlich geprüft, ob die Teilnahmewahrscheinlichkeit einer Hochschule an der Studie mit bestimmten Hochschulmerkmalen in Zusammenhang steht. Die berücksichtigten Hochschulmerkmale waren die Hochschulart, die Zahl der Promotionen im Prüfungsjahr 2014 sowie das Bundesland. Es konnten keine systematischen Ausfälle festgestellt werden. Eine Überprüfung der Ausfälle auf Ebene der Verwaltungsstellen war aufgrund fehlender Informationen zu möglicherweise relevanten Merkmalen nicht möglich.

Für die zweite Welle wurden zusätzlich Panelgewichte berechnet, um mögliche Selektionsmechanismen bei der Teilnahme an der Folgebefragung auszugleichen. Im Gegensatz zu den Nicht-Teilnehmer(inne)n der ersten Welle lagen für die Nicht-Teilnehmer(inne)n der zweiten Welle aus der Erstbefragung umfangreiche Informationen vor, die zur Modellierung der Teilnahmewahrscheinlichkeit genutzt werden konnten.

**Tabelle 8: Gewichtungsvariablen im Scientific-Use-File**

Variablenname	Beschreibung
wgt_t1	Querschnittsgewicht Welle 1 (Redressmentgewicht)
wgt_t1t2	Längsschnittgewicht 2-Wellen-Panel (getrimmt)

Die bereitgestellten Gewichte können in Stata mit Hilfe .ado-spezifischer Optionen berücksichtigt werden.<sup>33</sup> Das Gewicht wgt\_t1 ist für Auswertungen der ersten Welle vorgesehen, das Gewicht wgt\_t1t2 für Auswertungen des Zwei-Wellen-Panels.

### 7.2 Gewichtung des Datensatzes

Für die Berechnung der Querschnittsgewichte in der ersten Welle wurden die Verteilungen der Promovierten nach Studienbereich x Region (Ost/West) x Geschlecht in der Stichprobe und in der Grundgesamtheit gegenübergestellt. Die Referenzdaten zur Grundgesamtheit wurden vom Statistischen Bundesamt erhoben.<sup>34</sup> Die Kombination der drei Gewichtungsmerkmale ergab eine Kreuztabelle mit  $59 \times 2 \times 2 = 236$  Zellen, für die jeweils ein Soll-Ist-Abgleich vorgenommen wur-

<sup>33</sup> Siehe hierzu die Stata-Hilfe (Befehl: help weights).

<sup>34</sup> Statistisches Bundesamt, Hauptberichte. Auswertung aus der ICE-Datenbank der Länderministerien (ICE = Information, Controlling, Entscheidung). Bestand: 50001

de. Jeder Merkmalskombination wurde ein Gewicht zugewiesen, das sich aus dem Quotienten des Anteils der jeweiligen Merkmalskombination in der Grundgesamtheit und deren Anteil in der Sample ergibt. Einzelne Zellen mussten für die Berechnung der Gewichte zusammengefasst werden.<sup>35</sup>

Für die zweite Welle mussten zusätzlich wellenspezifische Ausfallprozesse modelliert werden. Für die Nonresponseanalysen standen die Befragungsdaten aus der ersten Welle zur Verfügung. Grundlage der Panelgewichtung ist ein logistisches Schätzmodell, das die Teilnahmewahrscheinlichkeit einer Person an der zweiten Welle vorhersagt. In einem explorativem Vorgehen wurden individuelle Merkmale identifiziert, die einen Beitrag zur Erklärung der Teilnahmewahrscheinlichkeit leisten. Das endgültige Schätzmodell beinhaltet die Variablen Geschlecht, Alter, Promotionsrahmen, Promotionsnote, Geburtsort, Fächergruppe, aktuelles/letztes Einkommen, die Branche in der die Person in Welle 1 tätig war sowie ein Item zur Selbsteinschätzung der Persönlichkeit aus dem Big-Five-Inventory („Ich bin bequem, neige zu Faulheit“). Um für jede Person die Teilnahmewahrscheinlichkeit schätzen zu können, wurden fehlende Werte bei allen Variablen als zusätzliche Kategorie in das Modell integriert.<sup>36</sup> Aus dem Modell konnte die bedingte Teilnahmewahrscheinlichkeit abgeleitet werden, deren Kehrwert das Ausfallgewicht für die zweite Welle darstellt:<sup>37</sup>

$$NR_{gewt_{2i}} = P(Res_{t_{2i}} | \sigma_{t_i})^{-1}$$

Das Gesamtgewicht für das Zwei-Wellen-Panel des Datensatzes ergibt sich aus dem Produkt des Querschnittsgewichts ( $wgt_{i_{t1}}$ ) mit dem Längsschnittgewicht ( $NR_{gewt_{2i}}$ ):

$$wgt_{i_{t1t2}} = wgt_{i_{t1}} \times NR_{gewt_{2i}}$$

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

## 8 Anonymisierung

**[Datenschutzrechtlicher Rahmen]** Für personenbezogene Daten, die in freiwilligen Befragungen durch das DZHW erhoben werden, gilt das Bundesdatenschutzgesetz (BDSG).<sup>38</sup> Danach dürfen personenbezogene Daten, die im Rahmen wissenschaftlicher Forschung erhoben worden sind,

<sup>35</sup> Das betraf in der Regel Zellen, deren Merkmalskombination in der Grundgesamtheit sehr selten vorkam und die daher durch die Stichprobe nicht abgebildet wurden.

<sup>36</sup> Das komplette Modell ist im Anhang 1 dargestellt, statistisch nicht bedeutsame Prädiktoren wurden aus dem Modell entfernt.

<sup>37</sup> Das Verfahren entspricht seiner Logik nach dem *Propensity Score Matching*, das auf Rosenbaum und Rubin (1983) zurückgeht (siehe auch Blumenstiel und Gummer (2015)).

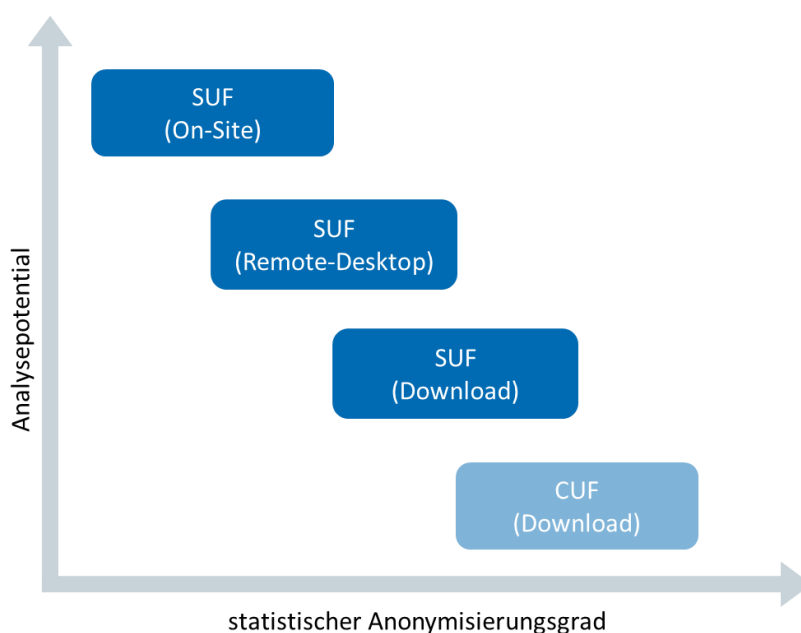
<sup>38</sup> Das BDSG kommt zur Anwendung, da die DZHW GmbH juristisch als öffentliche Stelle des Bundes betrachtet wird (vgl. §2 Abs. 3 BDSG). Der Bund hält die absolute Mehrheit der Anteile der DZHW GmbH und das Institut erfüllt Aufgaben der öffentlichen Verwaltung des Bundes im weitesten Sinn. Zur Auslegung einzelner rechtlicher Aspekte wird ergänzend die Europäische Datenschutzrichtlinie hinzugezogen.

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

**[Datenzugang, Anonymisierungsgrad und Analysepotential]** Das FDZ-DZHW stellt für das Promoviertenpanel 2014 ein faktisch anonymisiertes SUF für die wissenschaftliche Sekundärnutzung und ein absolut anonymisiertes CUF für Lehr- und Übungszwecke zur Verfügung. Die Anonymität der Befragten wird dabei über eine Kombination aus statistischen Maßnahmen und technischen Zugriffsbeschränkungen sichergestellt. Je stärker der Datenzugang technisch kontrolliert wird, desto geringer ist das Risiko einer De-Anonymisierung der Daten, desto weniger müssen die Daten mittels statistischer Maßnahmen um Informationen reduziert werden und desto größer bleibt ihr Analysepotential.

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

**Abbildung 4: Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten des DZHW- Promoviertenpanels 2014**



**[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 Promoviertenpanel 2014 bereits während der Feldphase in einem separaten Datensatz erfasst (vgl. Kapitel 4, Fußnote 15) 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.<sup>39</sup> Für das Promoviertenpanel 2014 wurden die folgenden Quasi-Identifikatoren identifiziert, die sowohl in externen Datenquellen<sup>40</sup> als auch in den Promoviertendaten vorliegen: Hochschule, Promotionsfach, Studienfach, Abschlussart, Preis für Promotion, Berufsangaben, regionale Informationen (zur Hochschule, zum Ort des Erwerbs der Studienberechtigung, zum Arbeitsort oder zu Auslandsaufenthalten) und persönliche Daten (z.B. Geburtsjahr, Angaben zu eigenen Kindern, Staatsangehörigkeit und Geburtsland). Um eine eindeutige Zuordnung der Promoviertendaten zu unterbinden, wurden diese Schlüsselmerkmale – je nach Datenprodukt bzw. Zugangsweg – aggregiert oder gelöscht (vgl. Tabelle 9). Beispielsweise werden bei dem Merkmal „Geburtsort“ in dem SUF für die On-Site Nutzung die ersten drei Stellen der Postleitzahl

<sup>39</sup> Dabei ist darauf hinzuweisen, dass die Identifikation einer Person bereits durch die Nichtteilnahme anderer Personen erschwert wird, da eine Ungewissheit darüber besteht, ob eine befragte Person eine einzigartige Merkmalskombination in der Population aufweist.

<sup>40</sup> Z.B. Studenten- und Prüfungsstatistik des statistischen Bundesamtes, Alumninetzwerke der Hochschulen oder auch Berufszetzwerke.



zahl, im Remote-Desktop-SUF die ersten zwei Stellen der Postleitzahl und im Download-SUF und im Download-CUF eine Zuordnung zu aggregierten Bundesländern herausgegeben. Offene Angaben sind ebenfalls Quasi-Identifikatoren (vgl. Ebel, 2015, S. 3) und wurden im Rahmen der Anonymisierung vercodet oder gelöscht.

Zuletzt wurde geprüft, ob in den Daten *sensible Informationen*, z. B. zur Gesundheit, sexuellen Orientierung oder zu politischen Einstellungen, enthalten waren. Diese eignen sich zwar nicht zur Re-Identifikation von Individuen oder Institutionen, jedoch können die Informationen im Falle einer De-Anonymisierung nutzbringend sein (vgl. Koberg, 2016, S. 694) und sind daher besonders schützenswert (vgl. §3 Abs. 9 BDSG, Art. 8 Abs. 1 und 2a EG-DSRL). Im Promoviertenpanel 2014 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 mit der Kategorie „Sonstiges“ zusammengefasst. Zur Gewährleistung der absoluten Anonymisierung der Daten des CUF wurde eine per Zufallsauswahl gewonnene Stichprobe der Daten (20 Prozent der befragten Promovierten) gezogen.

**Tabelle 9: Maßnahmen der statistischen Anonymisierung der Daten des DZHW-Promoviertenpanels 2014 nach Zugangsweg<sup>41</sup>**

Merkmal	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (Stichprobe)
Direkte Identifikatoren	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID
Fragebogeneingang	Freigabe	Löschung	Löschung	Löschung
Promotions-/ Studienfach	Aggregation zu Studienbereichen <sup>a</sup>	Aggregation zu Studienbereichen <sup>a</sup>	Aggregation zu Fächergruppen <sup>a</sup>	Aggregation zu Fächergruppen <sup>a</sup>
Hochschule	Aggregation zu Hochschulart <sup>b</sup>	Aggregation zu Hochschulart <sup>b</sup>	Löschung	Löschung
Hochschulort	Aggregation zu Bundesländern	Aggregation zu Bundesländer-Gruppen	Aggregation zu Bundesländer-Gruppen	Aggregation zu Bundesländer-Gruppen
Preis für Promotion	Freigabe	Freigabe	Löschung	Löschung
Weiterer akademischer Abschluss (Abschlussart)	Freigabe	Aggregation zu Magister, Staatsexamen, Diplom, Bachelor, Master, Sonstiges	Aggregation zu Magister, Staatsexamen, Diplom, Bachelor, Master, Sonstiges	Aggregation zu Magister, Staatsexamen, Diplom, Bachelor, Master, Sonstiges
Arbeitsort und Ort der Studienberechtigung (PLZ)	Deutschland: Postleitzahl (Stellen 1 bis 3) Ausland: Freigabe	Deutschland: Postleitzahl (Stellen 1 und 2) Ausland: Aggregation zu Weltregionen <sup>c</sup>	Deutschland: Vier Bundesländer einzeln ausgewiesen; ansonsten Aggregation zu fünf Bundesländer-Gruppen Ausland: Aggregation zu Weltregionen <sup>c</sup>	Deutschland: Vier Bundesländer einzeln ausgewiesen; ansonsten Aggregation zu fünf Bundesländer-Gruppen Ausland: Aggregation zu Weltregionen <sup>c</sup>

<sup>41</sup> Detaillierte Informationen zu den anonymisierten Variablen sind dem Datensatzreport sowie dem MetadatenSuchsystem (<https://metadata.fdz.dzhw.eu/>) zu entnehmen.

Merkmal	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (Stichprobe)
Auslandsaufenthalte (Land)	Freigabe	Freigabe	Aggregation zu Weltregionen <sup>c</sup>	Aggregation zu Weltregionen <sup>c</sup>
Beruf	Aggregation zu Berufsgruppen <sup>d</sup>	Aggregation zu Berufshauptgruppen <sup>d</sup>	Aggregation zu Berufshauptgruppen <sup>d</sup>	Aggregation zu Berufshauptgruppen <sup>d</sup>
Personalkategorien	Freigabe	Freigabe	Aggregation: WHK zu sonstige; PD zu Dozent; akad. Oberat a. Z. zu akad. Oberrat	Aggregation: WHK zu sonstige; PD zu Dozent; akad. Oberat a. Z. zu akad. Oberrat
Betriebsgröße	Freigabe	Freigabe	Aggregation: Welle 1: 1-20; 21-249; 250-1000; 1001 und mehr Welle 2: 1-19; 20-249; 250-999; 1000 und mehr	Aggregation: Welle 1: 1-20; 21-249; 250-1000; 1001 und mehr Welle 2: 1-19; 20-249; 250-999; 1000 und mehr
Staatsangehörigkeit (Ausland)	Freigabe	Freigabe	Aggregation zu Weltregionen <sup>c</sup>	Aggregation zu Weltregionen <sup>c</sup>
Deutsche Staatsangehörigkeit seit	Freigabe	Freigabe	Aggregation: bis 1989; 1990-1999; 2000-2009; ab 2010	Aggregation: bis 1989; 1990-1999; 2000-2009; ab 2010
Jahr der Zuwanderung	Freigabe	Freigabe	Aggregation: bis 1989; 1990-1999; ab 2000	Aggregation: bis 1989; 1990-1999; ab 2000
Geburtsjahr	1961 bis 1988 einzeln ausgewiesen, ansonsten Aggregation: bis 1949; 1950-1954; 1955-1960; 1989 und jünger	1961 bis 1988 einzeln ausgewiesen, ansonsten Aggregation: bis 1949; 1950-1954; 1955-1960; 1989 und jünger	Aggregation: bis 1959; 1960-1969; 1970-1979; 1980-1981; 1982-1983; 1984-1985; 1986-1987; ab 1988	Aggregation: bis 1959; 1960-1969; 1970-1979; 1980-1981; 1982-1983; 1984-1985; 1986-1987; ab 1988
Anzahl der Kinder	Freigabe	Freigabe	Top-Codierung <sup>e</sup>	Top-Codierung <sup>e</sup>
Geburtsjahr und -monat der Kinder	Freigabe	Geburtsjahr: Aggregation: bis 1997; 1998-2003; 2004-2009; 2010-2012; ab 2013 Geburtsmonat: Löschung	Geburtsjahr: (nur für die vier jüngsten Kinder) Aggregation: bis 1997; 1998-2009; ab 2010 Geburtsmonat: Löschung	Geburtsjahr: (nur für die vier jüngsten Kinder) Aggregation: bis 1997; 1998-2009; ab 2010 Geburtsmonat: Löschung
Angaben zu den Kindern (eigenes Kind, wohnhaft im Haushalt)	Freigabe	Freigabe	nur für die vier jüngsten Kinder	nur für die vier jüngsten Kinder
Beruf der Eltern	Aggregation zu Berufsuntergruppen <sup>d</sup>	Aggregation zu Berufsgruppen <sup>d</sup>	Aggregation zu Berufshauptgruppen <sup>d</sup>	Aggregation zu Berufshauptgruppen <sup>d</sup>
Antwortkategorien zur Gesundheit	Zusammenfassen mit der Kategorie „an-	Zusammenfassen mit der Kategorie „andere Gründe“	Zusammenfassen mit der Kategorie „andere Gründe“	Zusammenfassen mit der Kategorie „andere Gründe“

## 8 Anonymisierung

---

Merkmal	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (Stichprobe)
dere Gründe“				
Sonstige offene Angaben	Löschung	Löschung	Löschung	Löschung

<sup>a</sup> Nach dem Schlüsselerzeichnis der Studenten- und Prüfungsstatistik WiSe 2015/2016 und SoSe 2016 von Destatis.

<sup>b</sup> In den Bundesländern, wo die Hochschulart nur selten vorkommt (< 3 Mal), wird die Hochschulart anonymisiert.

<sup>c</sup> Nach der Staats- und Gebietssystematik 2014 von Destatis.

<sup>d</sup> Nach Klassifikation der Berufe von 2010 von Destatis.

<sup>e</sup> Angaben von vier und mehr Kindern wurden zu einer Kategorie zusammengefasst.

## 9 Literaturverzeichnis

- Auriol, L., Felix, B. & Schaaper, M. (2012). *Mapping Careers and Mobility of Doctorate Holders* (OECD Science, Technology and Industry Working Papers, 2012/07). doi:10.1787/5k4dnq2h4n5c-en
- Bäumer, T., Preis, N., Roßbach, H.-G., Stecher, L. & Klieme, E. (2011). *6 Education processes in life-course-specific learning environments* (2. Aufl.) (Nr. 14). doi:10.1007/s11618-011-0183-6
- Beierlein, C., Kovaleva, A., Kemper, C. J. & Rammstedt, B. (2014). *Allgemeine Selbstwirksamkeit Kurzsкала (ASKU). Zusammenstellung sozialwissenschaftlicher Items und Skalen* (GESIS – Leibniz-Institut für Sozialwissenschaften, Hrsg.). doi:10.6102/zis35
- Blickle, G., Kuhnert, B. & Rieck, S. (2003). Laufbahnförderung durch ein Unterstützungsnetzwerk. *Zeitschrift für Personalpsychologie*, 2 (3), 118–128. doi:10.1026//1617-6391.2.3.118
- Blumenstiel, J. E. & Gummer, T. (2015). Prävention, Korrektur oder beides? Drei Wege zur Reduzierung von Nonresponse Bias mit Propensity Scores. In J. Schupp & C. Wolf (Hrsg.), *Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen* (S. 13–44). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-10459-7
- Brandt, G., Vogel, S. de & Jakszat, S. (2016). *Entwicklung und Testung eines Instruments zur Erfassung der Lernumwelt in der Promotionsphase. Ergebnisse der Entwicklungsstudie.* : Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).
- Bundesamt für Statistik (Hrsg.). (2011). *Von der Hochschule ins Berufsleben. Erste Ergebnisse der Hochschulabsolventenbefragung 2009* (Statistik der Schweiz: Bildung und Wissenschaft), Neuchâtel
- Das Institut für Gründung und Innovation der Universität Potsdam. (2010). *Onlinebefragung des BIEM-CEIP zur Karriereentwicklung von Wissenschaftlerinnen und Wissenschaftlern in Forschungsteams.*
- Ebel, T. (2015). *Empfehlungen zur Anonymisierung quantitativer Daten.* Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften.
- Egeln, J., Gottschalk, S., Rammer, C. & Spielkamp, A. (2003). *Spinoff-Gründungen aus der öffentlichen Forschung in Deutschland* (Dokumentation Nr. 03-02). Mannheim: Zentrum für Europäische Wirtschaftsforschung GmbH (ZEW).
- Grühn, D., Hecht, H., Rubelt, J. & Schmidt, B. (2009). *Der wissenschaftliche „Mittelbau“ an deutschen Hochschulen. Zwischen Karriereaussichten und Abbruchtendenzen.* Berlin: ver.di - Vereinte Dienstleistungsgewerkschaft.
- Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. (2012). *FDZ-Methodenreport. Datenschutz am Forschungsdatenzentrum* (Nr. 06). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Jungbauer-Gans, M. & Gross, C. (2013). Determinants of Success in University Careers. Findings from the German Academic Labor Market / Erfolgsfaktoren in der Wissenschaft – Ergebnisse aus einer Habilitiertenbefragung an deutschen Universitäten. *Zeitschrift für Soziologie*, 42 (1). doi:10.1515/zfsoz-2013-0106
- 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.
- Kovaleva, A., Beierlein, C., Kemper, C. J. & Rammstedt, B. (2014). *Internale-Externale-Kontrollüberzeugung-4 (IE-4). Zusammenstellung sozialwissenschaftlicher Items und Skalen* (GESIS - Leibniz-Institut für Sozialwissenschaften, Hrsg.). doi:10.6102/zis184
- 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.
- Otto, K., Glaser, D. & Dalbert, C. (2004). *Hallesche Berichte zur Pädagogischen Psychologie. Skalendokumentation "Geografische und berufliche Mobilitätsbereitschaft"* (Dalbert, C., Hrsg.) (Nr. 8).
- Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. *Proceedings of the Survey Research Methods Section*, 225–230.

- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2014). *Big Five Inventory (BFI-10). Zusammenstellung sozialwissenschaftlicher Items und Skalen* (GESIS – Leibniz-Institut für Sozialwissenschaften, Hrsg.). doi:10.6102/zis76
- Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 41–55. doi:10.2307/2335942
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63 (2), 581–592.
- Scherer, S. & Brüderl, J. (2010). Sequenzdatenanalyse. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 1031–1051). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schnell, R., Hill, P. B. & Esser, E. (2005). *Methoden der empirischen Sozialforschung* (7. Aufl.). München: Oldenbourg.
- Statistisches Bundesamt. (2013). *Hochqualifizierte in Deutschland. Erhebung zu Karriereverläufen und internationaler Mobilität von Hochqualifizierten*, Wiesbaden
- Statistisches Bundesamt. (2015). *Bildung und Kultur. Prüfungen an Hochschulen 2014* (Statistisches Bundesamt, Hrsg.). Wiesbaden: Statistisches Bundesamt.
- 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
- Wissenschaftliches Zentrum für Berufs- und Hochschulforschung. (2009). *Neue Ausbildungsformen – andere Werdegänge? Eine Untersuchung zu Ausbildungs- und Berufsverläufen ehemaliger Doktoranden*. Fragebogen.

## Anhang 1: Logit-Regression zur Erstellung des Panellausfallgewichts in Welle 2

	Odds Ratio	Std. Err	z	P> z
<b>Geschlecht</b>				
Männlich	Ref.			
Weiblich	1,06	0,07	0,91	0,365
keine Angabe	0,80	0,36	-0,49	0,625
<b>Alter</b>				
23-30	1,13	0,08	1,58	0,115
31-34	Ref.			
35-39	1,03	0,08	0,36	0,719
40-44	1,21	0,17	1,33	0,185
45-49	2,18	0,44	3,81	0,000
50-60	1,80	0,40	2,67	0,008
61-79	1,14	0,44	0,33	0,744
keine Angabe	0,14	0,08	-3,55	0,000
<b>Promotionsrahmen</b>				
wiss. Mitarbeiter Haushaltstelle	0,98	0,08	-0,29	0,775
wiss. Mitarbeiter Drittmittelstelle	Ref.			
strukturiertes Promotionsprogramm	0,77	0,77	-2,18	0,029
Stipendienprogramm	0,83	0,09	-1,68	0,092
frei promovierend	1,06	0,10	0,62	0,534
keine Angabe	0,46	0,19	-1,84	0,066
<b>Promotionsnote</b>				
summa cum laude	1,19	0,12	1,78	0,074
magna cum laude	1,14	0,09	1,67	0,094
cum laude	Ref.			
satis bene	0,97	0,37	-0,07	0,946
rite	1,00	0,21	0,00	0,997
keine Angabe	1,11	0,35	0,33	0,742
<b>Geburtsort</b>				
Deutschland	Ref.			
In einem anderen Land	0,78	0,08	-2,56	0,010
keine Angabe	0,69	0,16	-1,58	0,113
<b>"Ich bin bequem, neige zu Faulheit"</b>				
trifft überhaupt nicht zu	0,79	0,79	-2,94	0,003
2	0,87	0,07	-1,77	0,077
3	Ref.			
4	1,27	0,16	1,97	0,048
trifft voll und ganz zu	0,98	0,23	-0,07	0,946
keine Angabe	0,28	0,15	-2,32	0,020
<b>Fächergruppe</b>				
Geisteswissenschaften	1,06	0,14	0,42	0,677
Sport	0,82	0,29	-0,56	0,577

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

Rechts-, Wirtschafts- und Sozialwiss.	0,76	0,07	-2,87	0,004
Mathematik, Naturwiss.	Ref.			
Humanmed./Gesundheitswiss.	0,63	0,08	-3,7	0,000
Agrar-, Forst-, Ernährungswiss, Veterinärm.	1,05	0,16	0,31	0,756
Ingenieurwissenschaften	0,68	0,07	-3,59	0,000
Kunst, Kunstwissenschaft	1,28	0,36	0,86	0,392
keine Angabe	0,69	0,08	-3,13	0,002
<b>aktuelles/letztes Bruttoeinkommen</b>				
bis 3200€	1,05	0,08	0,66	0,507
3201€ bis 5000€	Ref.			
über 5000€	1,08	0,09	0,94	0,348
kein Einkommen	1,40	0,26	1,84	0,065
keine Angabe	0,61	0,07	-4,37	0,000
<b>Branche</b>				
Landwirt,Fischerei,Energie,Wasserwirt.	0,91	0,21	-0,42	0,675
Chemische Industrie	0,62	0,09	-3,14	0,002
Maschinen-,Fahrzeugbau	0,60	0,10	-3,03	0,002
E-Technik,Elektronik,EDV-Geräte	0,69	0,14	-1,85	0,065
Sonstiges verarbeitendes Gewerbe	0,69	0,14	-1,77	0,077
Handel,Banken,Versicherungen	0,57	0,12	-2,67	0,007
Softwareentwicklung	1,04	0,22	0,19	0,848
Rechts-,Wirtschafts-,Personalberatung	0,57	0,09	-3,38	0,001
Medien,Verlagswesen,Werbung,Öff	0,47	0,12	-2,89	0,004
Gesundheitswesen	0,79	0,09	-2,06	0,040
Sonstige Dienstleistungen	0,89	0,14	-0,75	0,452
Schulen	0,85	0,20	-0,72	0,474
Hochschulen	Ref			
Forschungseinrichtungen	1,03	0,13	0,19	0,849
Sonstige Bildung,Forschung,Kultur	1,25	0,35	0,79	0,429
Öffentliche Verwaltung	0,77	0,77	-1,57	0,116
Sonstige nicht gewinnorient. Organisationen	0,75	0,16	-1,35	0,177
keine Angabe	0,55	0,09	-3,74	0,000
Konstante	2,16	0,29	5,77	0,000

$N=5412$ ;  $Pseudo R^2=0,0344$