

Gesche Brandt / Susanne de Vogel / Steffen Jaksztat /
Anne-Marie Lapstich / Carola Teichmann / Sandra Vietgen /
Marten Wallis

DZHW-Promoviertenpanel 2014

Daten- und Methodenbericht zu den Erhebungen der
Promoviertenkohorte 2014 (Befragungswelle 1-5);
Version 4.0.0

Oktober 2020

Autor(inn)en:

Gesche Brandt
Susanne de Vogel
Steffen Jaksztat
Anne-Marie Lapstich
Carola Teichmann
Sandra Vietgen
Marten Wallis

Unter Mitarbeit von:
Kolja Briedis

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/>



Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH
Lange Laube 12 | 30159 Hannover | www.dzhw.eu

Oktober 2020

Inhaltsverzeichnis

I	Einleitung	7
II	Datennutzungshinweise für das SUF	8
1	Inhalt und Anlage der Studie.....	11
2	Erhebungsinstrumente.....	12
2.1	Pretest.....	12
2.2	Kerninstrumente.....	12
2.2.1	Erwerbstätigkeit.....	12
2.2.2	Berufliche und wissenschaftliche Weiterqualifikation.....	13
2.2.3	Wissenschaftliche Aktivitäten.....	13
2.2.4	Internationale Mobilität.....	13
2.2.5	Private Lebenssituation.....	14
2.3	Lernumwelt und Rahmenbedingungen	16
2.4	Wellenspezifische (Schwerpunkt-)Themen	16
2.4.1	Internationale Mobilität.....	16
2.4.2	Individuelles Innovationsverhalten.....	16
2.4.3	Persönlichkeitsmerkmale.....	16
2.4.4	Gesundheit und Beeinträchtigung	17
3	Grundgesamtheit und Kontaktierung	18
3.1	Welle 1	18
3.2	Welle 2	18
3.3	Welle 3	19
3.4	Welle 4.....	19
3.5	Welle 5	19
4	Durchführung.....	20
4.1	Welle 1	20
4.2	Welle 2	21
4.3	Welle 3	22
4.4	Welle 4	23
4.5	Welle 5	25
5	Rücklauf.....	27
5.1	Welle 1	28
5.2	Welle 2	29
5.3	Welle 3	30
5.4	Welle 4	31
5.5	Welle 5	32
6	Datenaufbereitung	35
6.1	Datenübertragung.....	35
6.2	Codierung offener Angaben.....	35
6.3	Datenprüfung und Datenbereinigung.....	38
6.4	Generierung von Variablen	39

6.5	Erstellung der Datensätze	40
6.6	Vergabe von Variablennamen, Variablenlabels und Wertelabels	40
6.7	Codierung fehlender Werte	43
7	Gewichtung	45
7.1	Vorgehen und Anwendungshinweise	45
7.2	Modellierung der Ausfallgewichte	46
7.3	Kalibrierung	47
7.4	Trimmung	48
7.5	Verteilung der Gewichte	48
8	Anonymisierung der SUF und CUF	49
9	Literaturverzeichnis	61

Abbildungsverzeichnis

Abbildung 1: Datenzugangswege und Analysepotential	9
Abbildung 2: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, Welle 1.....	29
Abbildung 3: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, Welle 2.....	30
Abbildung 4: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, Welle 3.....	31
Abbildung 5: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, Welle 4.....	32
Abbildung 6: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, 5. Welle.....	33
Abbildung 7: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, Welle 2 bis 5	34
Abbildung 8: Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten des DZHW- Promoviertenpanels 2014	50

Tabellenverzeichnis

Tabelle 1:	Thematischer Aufbau der Erstbefragung	14
Tabelle 2:	Thematischer Aufbau der Zweitbefragung.....	14
Tabelle 3:	Thematischer Aufbau der Drittbefragung	15
Tabelle 4:	Thematischer Aufbau der Viertbefragung.....	15
Tabelle 5:	Thematischer Aufbau der Fünftbefragung	15
Tabelle 6:	Brutto- und Nettorücklaufquote des DZHW-Promoviertenpanels 2014	27
Tabelle 7:	Beteiligung der Hochschulen.....	28
Tabelle 8:	Vercodete Merkmale und verwendete Codierlisten (* für Welle 1-5).....	37
Tabelle 9:	Systematik der Themenbereiche und Variablenstämme	42
Tabelle 10:	Systematik des FDZ-DZHW für fehlende Werte	44
Tabelle 11:	Bereitgestellte Gewichte zum DZHW-Promoviertenpanel 2014.....	46
Tabelle 12:	Verteilung der Gewichte für Welle 1 bis 5	48
Tabelle 13:	Maßnahmen der statistischen Anonymisierung der Daten des DZHW-Promoviertenpanels 2014 nach Zugangsweg.....	51
Tabelle 14:	Logit-Regression zur Erstellung des Ausfallgewichts für Welle 2	54
Tabelle 15:	Logit-Regression zur Erstellung des Ausfallgewichts für Welle 3	55
Tabelle 16:	Logit-Regression zur Erstellung des Ausfallgewichts für Welle 4	57
Tabelle 17:	Logit-Regression zur Erstellung des Ausfallgewichts für Welle 5	59

I Einleitung

Das DZHW-Promoviertenpanel ist eine Untersuchung der Promotionsbedingungen, des beruflichen Übergangs nach der Promotion sowie des weiteren Berufsverlaufs von Promovierten.¹ Es wird von der Deutschen Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW) 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 Forschungsdatenzentrums für Hochschul- und Wissenschaftsforschung am DZHW (FDZ-DZHW) werden die Daten in Zusammenarbeit mit dem Projektteam 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.²

Der vorliegende Daten- und Methodenbericht dokumentiert die Befragungswellen des Promoviertenpanels 2014 von der ersten bis zur fünften Welle (doi: 10.21249/DZHW:phd2014:4.0.0). Weitere Dokumentationsmaterialien zur Studie (Fragebögen etc.) können frei im Metadatensuchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden. Die Dokumentationsmaterialien des Primärforschungsteams werden vom FDZ zum Download angeboten bzw. sind dem Anhang zu entnehmen. Die zentralen Informationen zur Nutzung der Daten dieser Studie werden in Abschnitt II dargestellt. 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 die Datenerhebungs- und Dateneditierungsprozesse beschrieben (Stichprobenziehung, Erhebungsablauf, Rücklauf, Datenaufbereitung). In den Kapiteln 7 und 8 folgt die Darstellung der vorgenommenen Gewichtung und Anonymisierung.

¹ Aktuelle Informationen zum Promoviertenpanel können über die Website des Projektes (www.promoviertenpanel.de) abgerufen werden.

² Informationen zu verfügbaren Datensätzen und Dokumentationen werden auf der Website des FDZ-DZHW (<https://fdz.dzhw.eu>) zur Verfügung gestellt.

II Datennutzungshinweise für das SUF

[Voraussetzungen der Datennutzung] Für personenbezogene Daten, die in freiwilligen Befragungen durch das DZHW erhoben werden, gelten die EU-Datenschutz-Grundverordnung (EUDSGVO) und das Bundesdatenschutzgesetz in seiner Neufassung vom 30. Juni 2017.³ Das FDZ bietet einen *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 ist neben dem Abschluss eines Datennutzungsvertrags mit dem FDZ eine Anstellung an einer wissenschaftlichen Einrichtung. 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. 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.

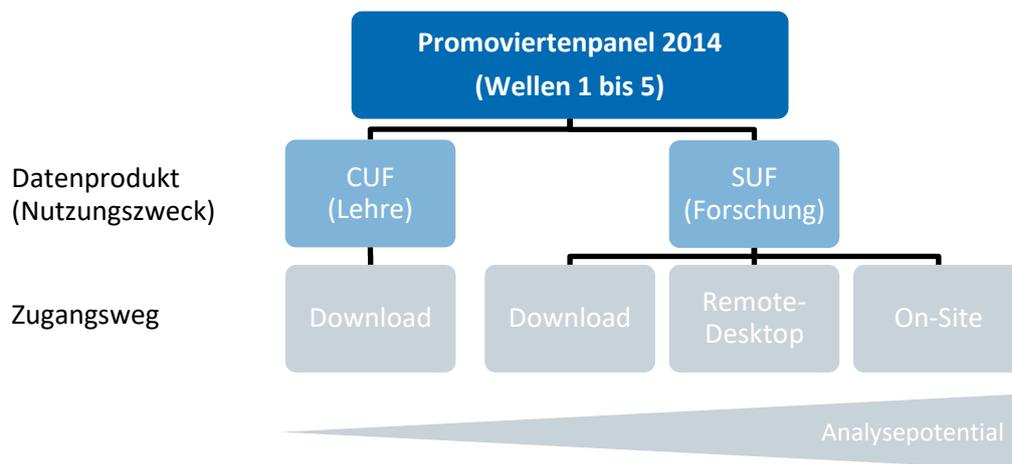
- **Download:** Die Daten werden verschlüsselt per E-Mail versandt oder auf der Website des FDZ zum Download bereitgestellt. Datennutzer*innen können die Daten auf ihrem lokalen Computer speichern, falls gewünscht selbst mit Daten aus externen Quellen verknüpfen und die Daten mit eigener Software analysieren.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

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 organisatori-

³ 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.

sche Maßnahmen kontrolliert wird, desto mehr Informationen können für die Datennutzer*innen bereitgestellt werden.⁴ 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:phd2014:4.0.0 ist eine Website 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 CUF⁵ werden für jeden im FDZ-DZHW angebotenen Zugangsweg beide Datensätze – jeweils mit zugangswegspezifischem Analysepotential (vgl. Abbildung 1) – bereitgestellt.

Die Datensätze sind für jeden Zugangsweg im Stata- und im SPSS-Format verfügbar.

[Gebühren der Datenbereitstellung] CUF und SUF werden derzeit (Stand: Oktober 2020) 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 Regeln⁶ 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.

⁴ Zu den verschiedenen Anonymisierungsgraden und Analysepotentialen des CUF und der verschiedenen SUF-Varianten vgl. Kapitel 8.

⁵ Aus Anonymisierungsgründen werden für den CUF jedoch nur die Daten einer Substichprobe bereitgestellt (vgl. Kapitel 8).

⁶ Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

- **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 Metadatensuchsystem⁷ entnommen werden.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen laut der Vorgaben des FDZ zitiert werden (vgl. Zitationsanleitung unter 10.21249/DZHW:phd2014:4.0.0).

⁷ <https://metadata.fdz.dzhw.eu>

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 Forschung⁸. 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 (für Details zu den ersten beiden Wellen siehe: Brandt et al. 2017b).

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. Die weiteren Folgebefragungswellen erfolgten in jährlichen Abstand jeweils im Frühjahr (Welle 2-5).

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).

⁸ Förderkennzeichen 16FWN014 und 16FWN017

2 Erhebungsinstrumente

In der ersten Befragungswelle des Promoviertenpanels 2014 wurde als Erhebungsinstrument ein standardisierter Papierfragebogen in deutscher Sprache eingesetzt.⁹ Für die Befragungswellen zwei bis fünf wurden zur Datenerhebung standardisierte Online-Fragebögen in deutscher Sprache verwendet.¹⁰ Zu den Kerninstrumenten, die in jeder Welle erneut abgefragt werden, gehören Fragen zu den Themen Erwerbstätigkeit, berufliche und wissenschaftliche Weiterqualifikationen, wissenschaftliche Aktivitäten, Mobilitätserfahrungen und Informationen zur privaten Lebenssituation. Die einzelnen Erhebungswellen enthalten darüber hinaus individuelle Schwerpunktsetzungen. Die Programmierung des Fragebogens und die Durchführung der Befragung erfolgten mit der DZHW-Onlinebefragungssoftware „Zofar“. Der thematische Aufbau der Befragungswellen kann Tabelle 1 entnommen werden.

2.1 Pretest

Vor der Erstbefragung wurden zunächst ein kognitiver und später ein quantitativer Pretest durchgeführt. In diesen Phasen 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 & Jaksztat, 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.

2.2 Kerninstrumente

2.2.1 Erwerbstätigkeit

Die Fragebögen der zweiten, dritten, vierten und fünften Welle enthalten zwei Instrumente zur Erfassung der Tätigkeitsverläufe im Anschluss an die Promotion bzw. seit der letzten Befragung: Zum einen kam ein Berufstableau zur monatsgenauen Erfassung wichtiger beruflicher Episoden und zur Abbildung detaillierter Merkmale der jeweiligen Episode (Beginn und Ende, berufliche Stellung, Vertragsart, Arbeitszeit, Arbeitsort, wissenschaftliche Tätigkeit und Arbeitgeber) zum Einsatz; darüber hinaus wurde ein Kalendarium eingesetzt, um die beruflichen und nicht-beruflichen Tätigkeiten nach der Promotion monatsgenau zu erfassen (z. B. weitere Ausbildungen, Stipendien, Elternzeit, Familienarbeit, Arbeitslosigkeit). Um die Fortschreibung der Daten

⁹ Auf Wunsch konnte von den Befragten ein deutscher oder englischer PDF-Fragebogen per E-Mail angefordert werden.

¹⁰ Die Fragebogen sowie das Filterführungsdiagramm und der Variablenplan zu jeder Welle können von der Website des FDZ heruntergeladen werden.

über die Wellen zu erleichtern, wurden jeweils alle Tätigkeiten seit der letzten Befragung, an der die Promovierten teilnahmen, erfasst. Befragte, die nicht an der zweiten Befragung teilgenommen haben, wurden um Angaben seit Abschluss ihrer Promotion gebeten, um möglichst vollständige Verlaufsdaten zu erhalten. Technisch wurde dies durch die Verwendung von Preloads und die Einblendung alternativer Fragetexte gelöst. Abschließend wurden die Befragten um die Kennzeichnung der Episode gebeten, bei der es sich um ihre aktuelle bzw. letzte Beschäftigung handelt.

Für die aktuelle bzw. letzte Stelle wurde zudem in jeder Welle eine Reihe weiterer Beschäftigungsmerkmale erhoben. Dazu gehören die Berufsbezeichnung, der Tätigkeitsbereich und die beruflichen Aufgaben, anhand derer während der Datenaufbereitung eine Vercodung nach der Klassifikation der Berufe 2010 (KldB2010) vorgenommen wurde. Des Weiteren wurde erfragt, ob es sich um eine Tätigkeit im öffentlichen Dienst handelt, und die Betriebsgröße, die Branche, die berufliche Zufriedenheit, die Adäquanz der Beschäftigung und das Einkommen erhoben. Ab der dritten Welle wurde bei Tätigkeiten in der Wissenschaft erstmalig auch die Personalkategorie erhoben. Schließlich wurden in jeder Welle Fragen zum Thema Selbständigkeit gestellt.

2.2.2 Berufliche und wissenschaftliche Weiterqualifikation

In der zweiten und dritten Welle wurden Informationen zu berufsständisch organisierten Weiterbildungen gesammelt. Da sich derartige Weiterbildungen üblicherweise über mehrere Jahre erstrecken, wurden Details zu berufsständischen Weiterbildungen fortan nur noch in jeder zweiten Befragung erhoben (folglich in W3 und W5). Zugleich wurde der Themenblock ab der dritten Befragung um zwei Fragen zum Abschluss der Weiterqualifikation ergänzt, die jährlich abgefragt werden.

Zum Kerninstrumentarium für die Erhebung ggf. stattfindender wissenschaftlicher Weiterqualifikation gehören Fragen zur Art und zum Fortschritt der Weiterqualifizierung, zur wissenschaftlichen Karriereintention, zur Einschätzung der Karriereaussichten und detailliertere Informationen zur Intention bzw. Art der Professur.

2.2.3 Wissenschaftliche Aktivitäten

Ab der zweiten Befragung wurden Informationen zur Wissenschaftshaltigkeit der Arbeit erfragt. Um die Befragungslast für die Befragten so gering wie möglich zu halten, erfolgte die detaillierte Abfrage wissenschaftlicher Aktivitäten ab der dritten Befragung nach einer einleitenden Filterfrage. Jährlich erfragt wurden Publikationen, Tagungsbesuche, Gutachter*innentätigkeiten, Drittmittelwerbung, Patente, Lehrtätigkeiten und die subjektive Integration in die Scientific Community. In der fünften Welle kamen zwei Instrumente zur Erfassung der Mentoringsituation in der Postdoc-Phase hinzu. Die Mitarbeit in akademischer Selbstverwaltung sowie die Mitgliedschaft in Verbänden und Fachgesellschaften wurden in den Wellen 3 bis 5 nicht mehr erfragt.

2.2.4 Internationale Mobilität

Seit der dritten Welle gehören Fragen zu internationalen Mobilitätserfahrungen und -plänen zum jährlichen Frageprogramm. Erfragt werden die absolvierten Auslandsaufenthalte seit der letzten Befragung und die Pläne, zukünftig ins Ausland zu gehen.

2.2.5 Private Lebenssituation

In diesem Themenbereich wurden in jeder Welle Informationen zur Partnerschaftssituation und Elternschaft erhoben. Dazu gehören beispielsweise der Familienstand, ggf. Angaben zum Bildungsstand des Partners/der Partnerin, zur Erwerbstätigkeit und zum Tätigkeitsbereich der Partner*innen sowie zur Haushaltssituation. Die Fragen nach vorhandenen Kindern unterscheiden sich in der dritten und vierten Welle im Vergleich zu den Vorwellen dahingehend, dass nur noch nach Kindern gefragt wurde, die seit der letzten Teilnahme dazu gekommen sind. In der fünften Welle wurden wieder die Gesamtzahl der Kinder und alle Geburtsdaten erhoben.

Tabelle 1: Thematischer Aufbau der Erstbefragung

Thema	Fragenummern aus Welle 1
Rahmenbedingungen /Angaben zur Promotionsphase	1.1 - 1.21
Finanzierung	2.1 - 2.3
Betreuung und Unterstützung	3.1 - 3.10
Wissenschaftliche Aktivitäten	4.1 - 4.10
Praktische Erfahrungen	5.1 - 5.6
Eigenschaften, Ziele und Motive /Persönliche Merkmale/Einstellungen	6.1 - 6.6
Berufliche Entwicklung	7.1 - 7.5
Beschäftigung und Erwerbstätigkeit	2.3; 7; 5; 8.1 - 8.13
Soziodemographische Merkmale	9.1 - 10.4
Vorbildung/Hochschulzugang	9.9 - 9.11
Soziale Herkunft	10.1 - 10.4

Tabelle 2: Thematischer Aufbau der Zweitbefragung

Thema	Fragenummern aus Welle 2
Promotionsergebnis	3.1 - 3.4; 6.3 - 6.4, 6.16 - 6.17
Wissenschaftliche und berufliche Weiterqualifikation	5.10 - 5.12; 6.10
Wissenschaftliche Aktivitäten	4.2 - 4.7; 5.1 - 5.10; 5.13 - 5.19; 6.1 - 6.3; 6.10
Bildungs- und Beschäftigungsbiografie vor Beginn der Promotion	1.1 - 1.2; 2.1 - 2.4
Beruflicher Werdegang seit Beendigung der Promotion	4.1; 4.8 - 4.9; 6.7 - 6.20
Mobilität	1.3; 1.6; 2.5 - 2.6; 6.21 - 6.24
Personenmerkmale/private Lebenssituation	6.25 - 6.34

Tabelle 3: Thematischer Aufbau der Drittbefragung

Thema	Fragennummern aus Welle 3
Erwerbstätigkeit	1.1 – 1.20
Weiterbildung	2.1 – 2.11
Wissenschaftliche Aktivitäten	3.1 – 3.18
Promotion	4.1 – 4.3
Mobilität	5.1 – 5.7
Partnerschaft und Kinder	6.1 – 6.9

Tabelle 4: Thematischer Aufbau der Viertbefragung

Thema	Fragennummern aus Welle 4
Erwerbstätigkeit	1.1 – 1.18
Dynamik des beruflichen Umfelds, Problemlösung, Entscheidungsautonomie	1.19 – 1.21
Selbständigkeit	1.22 – 1.23.2
Weiterbildung	2.1 – 2.2
Wissenschaftliche Aktivitäten	2.3 – 2.5, 3.1 – 3.17
Professur	2.6 – 2.9
Mobilität	4.1 – 4.3
Partnerschaft und Kinder	5.1 – 5.8

Tabelle 5: Thematischer Aufbau der Fünftbefragung

Thema	Fragennummern aus Welle 5
Erwerbstätigkeit	1.1 – 1.19
Selbständigkeit	1.20 – 1.21
Weiterbildung	2.1 – 2.4
Wissenschaftliche Aktivitäten	2.5 – 2.9, 3.1 – 3.13
Professur	2.10 – 2.13
Mobilität	3.16 – 4.3
Partnerschaft und Kinder	5.1 – 5.10
Gesundheit- und Beeinträchtigung	6.1 – 6.6

2.3 Lernumwelt und Rahmenbedingungen

Zur Erfassung der Lernumwelt in der Promotionsphase wurde für die erste Welle ein theoretisch fundiertes und empirisch getestetes Modell entwickelt. Die theoretische Konzeption der Lernumwelten basiert auf dem SSCO-Modell (*Structure - Support - Challenge - Orientation*) von Bäumler, Preis, Roßbach, Stecher & Klieme (2011).

2.4 Wellenspezifische (Schwerpunkt-)Themen

In der dritten Welle wurden bei Personen, die nicht an der zweiten Welle teilgenommen haben, zentrale Fragen aus der zweiten Welle nacherhoben. Dazu gehören z. B. Informationen zur Promotion (Veröffentlichung, Datum der Urkundenverleihung, erhaltene Preise), zur internationalen Mobilität vor und während des Studiums und zur Staatsangehörigkeit. Darüber hinaus wurden in jeder Welle spezifische (Schwerpunkt-)Themen wie internationale Mobilität (Welle 3), individuelles Innovationsverhalten (Welle 4) sowie Gesundheit und Behinderung (Welle 5) gesetzt.

2.4.1 Internationale Mobilität

In der dritten Welle wurde ein Schwerpunkt auf internationale Mobilitätserfahrungen gelegt. Neben der Nacherhebung der Auslandsaufenthalte vor dem Promotionsabschluss und neben den Kerninstrumenten zur internationalen Mobilität (vgl. 2.2.4) wurden auch die zukünftigen Mobilitätspläne sowie die Nutzung von Informations- und Kommunikationstechnologien (IKT) für den internationalen Austausch erhoben.

2.4.2 Individuelles Innovationsverhalten

Der Fragebogen der vierten Welle enthielt mehrere etablierte Instrumente zum Thema individuelles Innovationsverhalten: Erhoben wurden die Dynamik des beruflichen Umfeldes (deutschsprachige Adaption von Schilke (2014)), Problemlösung und Entscheidungsautonomie im Beruf (Morgeson und Humphrey (2006), deutschsprachige Adaption von Stegmann et al. (2010)) sowie die individuelle Ambidexterität (Mom et al. (2009), deutsche Adaption von Kobarg et al. (2017)).

2.4.3 Persönlichkeitsmerkmale

Die Instrumente zur Erfassung zentraler Persönlichkeitsmerkmale, die bereits in die Erstbefragung integriert wurden, wurden in der vierten Welle erneut eingesetzt. Dazu zählen die psychometrischen Kurzskalen zur Messung der allgemeinen Selbstwirksamkeit (ASKU) (Beierlein et al. 2014b), zur Messung der „Big Five-Persönlichkeitsfaktoren“ (BFI-10) (Rammstedt et al. 2014) sowie internaler und externaler Kontrollüberzeugungen (IE-4) (Kovaleva et al. 2014).

2.4.4 Gesundheit und Beeinträchtigung

Thematischer Schwerpunkt der fünften Welle war der Bereich Gesundheit und Behinderung. Hierzu kam ein Kurzinstrument zur Erfassung des subjektiven allgemeinen Gesundheitszustands zum Einsatz. Gestellt wurden außerdem drei Fragen zur Erhebung der körperlichen und psychischen Gesundheit sowie gesundheitsbedingten Beeinträchtigungen im Alltag. Bei diesen Fragen handelt es sich um eine deutschsprachige Adaption des Healthy Days Core Module (CDC HRQOL- 4) (Centers for Disease Control and Prevention (CDC) 2000), die beispielsweise auch in der Startkohorte 5 des NEPS (Nationales Bildungspanel (NEPS) 2018) eingesetzt wurden. Schließlich wurden Informationen zu Behinderungen und ggf. die Art, das Jahr der Feststellung und der Grad der Behinderung abgefragt. Diese Fragen sind Instrumente, die bereits in der Startkohorte 6 des NEPS (Nationales Bildungspanel (NEPS) 2013) genutzt wurden. Alle gesundheitsbezogenen Instrumente sind auch Teil des Nacaps-Frageprogramms. Da es sich bei Daten zum Gesundheitszustand nach DSGVO um besonders schützenswerte Daten handelt, wurde für diese Fragen noch einmal das explizite Einverständnis der Befragten eingeholt.

3 Grundgesamtheit und Kontaktierung

3.1 Welle 1

Um ein möglichst repräsentatives Bild und aussagekräftige Schlussfolgerungen über die Promotionsbedingungen sowie 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 (damals) 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.

3.2 Welle 2

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.¹¹ Um Personen zu berücksichtigen, die in der Zwischenzeit umgezogen waren, wurden die Adressbestände zwischen den Wellen geprüft und ggf. aktualisiert.

Zwischen der ersten und der zweiten Befragung 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 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. Abzüglich der Personen,

¹¹ Zur Gewährleistung des Datenschutzes wurde der Adressabschnitt vom Fragebogen abgetrennt und die Referenzliste getrennt von den Befragungsdaten auf einem geschützten Server gespeichert.

die keine Zustimmung für eine Teilnahme an einer Folgebefragung gegeben haben oder bei denen die Verknüpfung zu einem Befragungsdatensatz fehlte (anonyme Beteiligung an der Befragung), wurden letztlich zur Zweitbefragung 4.816 Befragte per E-Mail eingeladen.

3.3 Welle 3

Zur Ausgangsstichprobe der dritten Welle des Promoviertenpanels gehörten alle Befragten, die an der ersten Welle teilgenommen haben, die der Teilnahme an Folgebefragungen zugestimmt haben und von denen aktuelle Kontaktdaten vorlagen. Somit wurden zur dritten Welle 4.809 Promovierte zur Beantwortung des Onlinefragebogens der dritten Welle eingeladen.

3.4 Welle 4

Auch in der vierten Welle wurde die vollständige Ausgangsstichprobe herangezogen. Allerdings hat sich diese um diejenigen Befragten verringert, welche aufgrund von veralteten Kontaktdaten nicht mehr kontaktiert werden konnten oder aufgrund einer Teilnahmeverweigerung nicht mehr kontaktiert werden wollten. Es wurden insgesamt 4.795 Befragte eingeladen, sich an der vierten Befragungswelle zu beteiligen.

3.5 Welle 5

Wie bereits in den vorherigen Folgebefragungen wurden auch für die fünfte Welle alle Befragten kontaktiert, die an der ersten Welle teilgenommen sowie der Teilnahme an den weiteren Befragungen zugestimmt haben und von denen aktuelle Kontaktdaten vorlagen. Abzüglich derjenigen, die ihre Teilnahme zwischenzeitlich verweigert hatten, wurden zur Beantwortung der fünften Welle insgesamt 4.778 Befragte per E-Mail eingeladen.

4 Durchführung

4.1 Welle 1

[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.

[Erhebungsunterlagen] Die Erhebungsunterlagen bestanden in der ersten Befragungswelle pro zu befragender Person aus einem Anschreiben (inkl. Datenschutzinformationen), dem Fragebogen (Paper-Pencil), 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 ebenfalls postalisch verschickt.

[Feldphase] Der Erhebungszeitraum der ersten Befragungswelle erstreckte sich vom 15. Dezember 2014 bis zum 17. Februar 2016. Die beiden Erinnerungsschreiben wurden etwa vier bzw. acht Wochen nach Feldstart verschickt.¹² 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.¹³ 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.

[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

¹² 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.

¹³ Gründe hierfür waren z. B. ein zu großer zeitlicher oder personeller Aufwand, längere Abwesenheiten des zuständigen Personals oder personelle Veränderungen der zuständigen Verwaltungsstelle.

der Versand von Erinnerungsschreiben.¹⁴ Mittels einer Rückmeldung ausgewählter Studienergebnisse¹⁵ nach der ersten Welle sollte zum einen die Bindung der Studienteilnehmenden 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 Teilnehmenden der ersten Welle ein Laptop, drei iPads sowie mehrere Bücher Gutscheine verlost.

4.2 Welle 2

[Kontaktaufnahme und Adresspflege] Aus den Adressangaben der Befragten in der ersten Befragung wurde ein Kontaktdatenatz erstellt. Um die Daten den Befragungsdatensatz zuordnen zu können und zugleich die Datenschutzrichtlinien des DZHW einzuhalten, wurden im Fragebogen der ersten Welle und die dazugehörigen Kontaktdaten bei Fragebogeneingang mit einer Identifikationsnummer versehen. Aus Datenschutzgründen werden die Adressdaten auf einem separaten und geschützten Server ohne Internetzugang, getrennt vom Befragungsdatensatz als Adressdatenbank verwaltet. Zur Aktualisierung des Kontaktdatenatzes werden die Adressbestände zwischen den Wellen geprüft und ggf. aktualisiert.

Der Kontaktdatenatz umfasst die folgenden Kontaktangaben, (soweit vorhanden): Name, Postanschrift, E-Mail-Adresse, Status der Postanschrift, Status der E-Mail-Adresse, offenes Feld zum Aktualisierungsverlauf, Versandart, Versanddatum je Aktion, Teilnahmestatus je Welle, Identifikationsnummer und Verschlüsselungscode je Welle, Versand von Zwischenkontaktaktionen, Incentive-Gewinn.

Wie in Abschnitt 3.2 bereits erwähnt, fanden zwischen der Erstbefragung und der Zweitbefragungswelle zwei Kontaktaktionen statt. Zunächst wurde eine E-Mail an alle Befragten mit einer Danksagung für die Teilnahme und Informationen zum Stand des Projektes verschickt. Als zweite Aktion wurde die Zweitbefragung unmittelbar vor der Feldphase (und der Einladung per E-Mail) auf postalischem Weg angekündigt.

[Erhebungsunterlagen] Die Befragung der zweiten Welle wurde mittels der DZHW-Onlinebefragungssoftware „Zofar“ programmiert. Per E-Mail wurden die Panelteilnehmer*innen zur Befragung eingeladen. Diese enthielt neben kurzen Anschreiben, den Link zur Befragungsteilnahme, Hinweise zum Datenschutz und die Ankündigung der Verlosung von explizit genannten Preisen. Die drei Erinnerungsschreiben zur Zweitbefragung wurden gezielt nur an diejenigen verschickt, die sich noch nicht an der zweiten Erhebung beteiligt hatten.

[Feldphase] Für die zweite Welle konnte aufgrund des vorliegenden Adressbestands aus der ersten Welle ein konkreter Termin zur Einladungen zur Befragung festgelegt werden. Die Befragung der zweiten Welle war vom 15. März 2016 bis zum 19. April 2016 für die Befragten erreichbar. Es wurden insgesamt drei Erinnerungsschreiben versandt am neunten, am 21. Und 30. Tag nach Befragungsstart.

¹⁴ 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.

¹⁵ Ein Kurzüberblick über die wichtigsten Befunde konnte in einem online bereitgestellten Flyer abgerufen werden. Ebenso werden detailliertere Informationen nach Fachgruppen zur Verfügung gestellt.

[Rücklaufsteigernde Maßnahmen] Unter allen Teilnehmenden der zweiten Welle wurden als Incentives ein hochwertiges und zum Befragungszeitpunkt neu erscheinendes Smartphone sowie mehrere Reisegutscheine verlost.

4.3 Welle 3

[Kontaktaufnahme und Adresspflege] Gelegenheit zu einer Aktualisierung der Kontaktdaten ergab sich aus zwei Kontakten zwischen der zweiten und dritten Befragung: Einmal durch den Versand von kompakten Ergebnissen aus den bisherigen Befragungen und einmal durch eine Ankündigungsmail zur dritten Befragung.

Die Befragten wurden bei jedem Kontakt gebeten, etwaige Änderungen ihrer Post- oder E-Mail-Adressen mitzuteilen. Darüber hinaus erfolgt eine Aktualisierung der nicht-zustellbaren Adressen über Versandmitteilungen von Anschriftenberichtigungen der Deutschen Post AG und durch Anfragen bei Einwohnermeldeämtern. Nach der letzten Aktualisierung der Adressbestände konnten 4.809 Panelteilnehmende für die dritte Befragungsrunde eingeladen werden. Davon wurden 151 postalisch und 4.658 per E-Mail angeschrieben. Sieben Teilnehmende konnten aufgrund von veralteten Adressdaten (drei Fälle), doppelter Teilnahme (ein Fall) oder einer Teilnahmeverweigerung/Absage (drei Fälle) nicht mehr befragt werden.

[Erhebungsunterlagen] Die Erhebungsunterlagen der dritten Befragung bestanden aus einem Anschreiben per E-Mail mit dem Link zur Onlinebefragung und der Ankündigung der Gewinnverlosung sowie Hinweise zum Datenschutz. In der dritten Welle wurden drei Erinnerungsschreiben und eine postalische Zwischeneinladung verschickt. Erinnerungsschreiben und Zwischeneinladung wurden gezielt an jene Befragte verschickt, die noch nicht an der laufenden Erhebung teilgenommen hatten¹⁶. Die postalische Zwischeneinladung erfolgte an sämtliche Personen, bei denen eine Postadresse vorlag. In dieser wurde mit einem nicht-personalisierten Anschreiben auf den Link und Identifikations-Code in der E-Mail hingewiesen. Der Onlinefragebogen der Folgewellen wurde, wie auch in der zweiten Welle, mit der DZHW-Onlinebefragungssoftware „Zofar“ programmiert.

[Feldphase] Der Versand der Einladungen zur dritten Befragung erfolgte am 30. März 2017. Acht Tage später erfolgten die erste Erinnerung, die zweite 20. Tage und die dritte Erinnerung 36. Tag nach Erhalt der Einladung. Die postalische Zwischeneinladung wurde zwischen dem 27. und 29. Tag versendet. Die Feldphase endete am 18. Mai 2017.

[Rücklaufsteigernde Maßnahmen] Die rücklaufsteigernden Maßnahmen zielten auf die individuelle Motivation der Befragten ab. Im März 2017 wurde die neue Projektwebsite (www.promoviertenpanel.de) online gestellt, welche Informationen über Studienergebnisse, aktuelle Beiträge im Projekt sowie des Projektteams in der Wissenschaft bietet. Die Rückmeldungen ausgewählter Studienergebnisse auf der Website soll die Bindung der Befragten an das (Promovierten-)Panel erhöhen, um diese zur Teilnahme an Folgewellen zu motivieren.

Circa zwei Wochen vor Durchführung der dritten Befragung wurde in einer E-Mail die anstehende Befragung angekündigt sowie auf die neue Website aufmerksam gemacht. Inhaltlich wurde in der Ankündigung sowie in der Einladung zur dritten Befragung die subjektive und gesell-

¹⁶ Als Teilnahmekriterium galt das Erreichen der Seite 70 des Fragebogens.

schaftliche Relevanz des Themas hervorgehoben. Der Versand ausgewählter Studienergebnisse sowie die Ankündigung der dritten Welle dienten darüber hinaus der Aktualisierung des Adressbestandes.

Die Erinnerungsschreiben haben wiederum messbar zur Rücklaufsteigerung beigetragen (vgl. Abbildung 4). Die postalische Zwischenkontaktaktion in der dritten Befragungswelle sollte diesen Effekt zusätzlich stärken. Für die letzte Erinnerung wurden zwei Erinnerungsschreiben erstellt, eines für jene, die nur in der ersten Welle teilnahmen, und ein zweites für Befragte, die an der ersten und zweiten Befragung teilgenommen haben. Dies sollte insbesondere die Motivation der kontinuierlich Teilnehmenden durch Hervorhebung der Bedeutsamkeit ihrer weiteren Teilnahme anregen.

Unter allen Teilnehmenden der dritten Welle wurden zudem Incentives verlost. Hauptgewinn war ein hochwertiger aktueller Tablet-PC von einem renommierten Hersteller; Nebengewinne waren 10 Reisegutscheine.

4.4 Welle 4

[Kontaktaufnahme und Adresspflege] Im Unterschied zur dritten Befragung wurde die vierte Befragung circa acht Tage vor Feldstart durch ein persönliches Anschreiben und Unterstützungsschreiben des BMBF postalisch angekündigt. Die persönlichen Anschreiben richteten sich zum einen gezielt an die paneltreuen Befragten (mit einer Teilnahme an zwei oder mehr Befragungen) mit dem Verweis auf die Bedeutung von Längsschnittdaten. In einer zweiten Ausführung richteten sich die Anschreiben gezielt an Personen, die bisher nur an der ersten Befragung des Promoviertenpanels teilgenommen hatten (Non-Responder). Der Versand dieser Vorankündigung erfolgte an 4.351 Panelteilnehmenden, die eine gültige Postadresse angegebenen hatten (die Einladung mit Zugangslink zur Befragung erfolgte jedoch wieder hauptsächlich per E-Mail). Neben der Ankündigung der nächsten Befragung konnte so auch eine Adressaktualisierung insbesondere der veralteten oder fehlenden Adressen durchgeführt werden. Eine Rückmeldung der aktuellen Post- oder E-Mail-Adresse erfolgte per E-Mail durch die Befragten.

Erneut wurden die nicht-zustellbaren Adressen über Versandmitteilungen (Unzustellbarkeitsnachricht von E-Mails), durch Anschriftenberichtigung der Deutschen Post AG und durch Anfragen bei Einwohnermeldeämtern aktualisiert. Nach der letzten Aktualisierung der Adressbestände kurz vor Start der vierten Befragung konnten so 4.795 Promovierte eingeladen werden. Davon wurden 157 postalisch und 4.643 per E-Mail eingeladen. Im Laufe der Feldphase verweigerten neun Personen ihre Teilnahme durch Mitteilung per E-Mail und sechs Personen konnten aufgrund von veralteten Adressen während der Feldphase nicht erreicht werden.

[Erhebungsunterlagen] Circa acht Tage nach Zustellung des postalischen Vorankündigungsschreibens wurde den 4.795 Panelteilnehmer(inne)n per E-Mail oder Brief das Einladungsschreiben mit persönlichem Zugangs- bzw. Verschlüsselungscode und angekündigter Gewinnverlosung zugesandt. Im Unterschied zur dritten Befragung enthielt das Einladungsschreiben nur einen verkürzten datenschutzrechtlichen Hinweis, da dieser ausführlich auf der Startseite der Befragung mit zusätzlichem Link zu den weiteren „Hinweise[n] zum Datenschutz“ im PDF-Format beschrieben ist. Diesen Datenschutzbestimmungen musste nach neuer Datenschutz-Grundverordnung (DSGVO) jetzt aktiv zugestimmt werden durch die „Opt-In“-Funktion (aktives Anklicken).

Wie auch in den Vorwellen wurden drei Erinnerungsschreiben an alle, die noch nicht teilgenommen¹⁶ hatten, versandt. Der Versand von Einladungen und Erinnerungsschreiben erfolgte hauptsächlich per E-Mail. Nur Panelteilnehmende ohne (gültige) E-Mail wurden postalisch kontaktiert. In der dritten Erinnerungsaktion erfolgte mit unterschiedlichen Anschreiben, die sich zum einen gezielt an eher paneltreue Nicht-Teilnehmer*innen (Teilnahme an mehreren Wellen) und zum anderen an stete Nicht-Teilnehmer*innen (seit Erstbefragung) richteten, eine spezifische Ansprache der Gruppen.

Wie auch in der zweiten und dritten Welle wurde der Online-Fragebogen mit der DZHW-Onlinebefragungssoftware „Zofar“ programmiert.

[Feldphase] Am 20. März 2018 ist die vierte Befragungswelle mit dem Versand der Einladungsschreiben gestartet. Die drei Erinnerungsschreiben wurden am zehnten, 22. und 31. Tag versendet. Die Feldphase endete am 2. Mai 2018.

[Rücklaufsteigernde Maßnahmen] Mit den postalischen Ankündigungsschreiben sollte auf die Wichtigkeit, Bedeutung und Vertrauenswürdigkeit der Studie bzw. der durchführenden Institution hingewiesen werden. Zu diesem Zweck wurde dem Schreiben neben dem Anschreiben auch ein Unterstützungsschreiben des BMBF beigelegt. Zum anderen sollte durch den postalischen Versand Rückbezug auf die Erstbefragung genommen werden, damit diese wiedererkannt wird und in Erinnerung bleibt. In den personalisierten Ankündigungsschreiben selbst wurde die angegebene E-Mailadresse zur Prüfung auf Aktualität abgedruckt. Auf diese Weise konnte neben der fortlaufenden Adressaktualisierung über Anschriftenbenachrichtigungen (Premiumadressservice der Deutschen Post AG) und Einwohnermeldeamtsrecherchen auch eine Prüfung durch die Befragten selbst stattfinden.

Die im Vorfeld der dritten Welle angekündigte Projektwebsite wurde weiter als Plattform genutzt, um aktuelle Meldungen zu Publikationen, Vorträgen oder anderen Beiträgen und Informationen über das Forschungsprojekt den Befragten und der Öffentlichkeit zugänglich zu machen. In den Kontaktschreiben der Vorkontaktaktion und der weiteren Anschreiben wurde für weiterführende Informationen und Projektergebnisse auf die Projekthomepage verwiesen.

Vor Beginn der Kontaktaktionen der vierten Welle wurden neue Ergebnisse auf der Projekthomepage präsentiert, das Anonymisierungskonzept und Vorgehen ausführlich dargestellt und als aktuelle Meldung zeitnah vor Feldstart online gesetzt.

Neben den motivationssteigernden Maßnahmen, wie der Incentivierung der Vorkontaktaktion mit dem begleitenden BMBF-Anschreiben, sollten durch die Ergebnisdarstellung der Nutzen der Studie für die Befragten selbst und auch für die Wissenschaft deutlich gemacht werden. Bedenken und Vorbehalten gegenüber der Studie sollte durch die Aufklärung zum Anonymisierungs- und Datenschutzkonzept entgegengewirkt werden. Zudem wurde versucht, durch die postalische Kontaktaktion und Projektwebsite eine Bindung zum Projekt aufzubauen.

Neben diesen neuen Maßnahmen wurden die bewährten Verfahren der Vorwellen wie Adressaktualisierungen (Premiumadressservice der Deutschen Post AG und Einwohnermeldeamtsrecherchen), der Versand von drei zum Teil gruppenspezifischen Erinnerungsschreiben und die Ankündigung der Verlosung von Incentives fortgeführt. Mit Unterschied zu den Vorwellen wurden zwar ebenfalls ein Hauptpreis und mehrere Nebenpreise verlost, als Hauptpreis konnte dieses Mal jedoch zwischen zwei verschiedenen hochwertigen Technikprodukten unterschiedlicher bekannter Anbieter ausgewählt werden. Verlost wurde als Hauptpreis wahlweise ein Mac

Book oder ein Convertible Notebook im Wert von 1.000 Euro und als Nebenpreise fünf E-Book-Reader.

4.5 Welle 5

[Kontaktaufnahme und Adresspflege] Als weitere Zwischenkontaktaktion wurden zuvor im Dezember (erstmalig) Weihnachtspostkarten versandt, um den Kontakt zu den Promovierten zu halten. Außerdem sollte so gewährleistet werden, dass die Adressbestände stets aktuell sind. Die Weihnachtspostkarten wurden an 4.246 Personen versandt.

Wie bereits bei der vierten Befragungswelle wurde auch die fünfte Welle eine Woche vor Start mit einem postalischen Anschreiben angekündigt. In diesem Fall handelte es sich um ein einfaches Anschreiben. Diese Vorkontaktaktion wurde ausschließlich postalisch verschickt an 3.591 Befragte, von denen eine gültige Post-Anschrift vorlag. Bei dieser Zwischenkontaktaktion wurden die Post-Adressen der Teilnehmenden auf Aktualität überprüft. Im Anschreiben der Vorkontaktaktion wurde die hinterlegte E-Mail-Adresse aufgeführt, um im Fall, dass diese nicht aktuell ist, die Befragten zur Rückmeldung zu animieren. Auch in den anderen Anschreiben wurde auf Ansprechpersonen zur Mitteilung veränderter Angaben hingewiesen, sodass eine Prüfung der Kontaktdaten zusätzlich zu dem Verfahren der Adressaktualisierung der Deutschen Post AG und des Einwohnermeldeamts durch die Befragten selbst stattfand.

Die Einladung zur Befragung erfolgte erneut hauptsächlich per E-Mail. So wurden insgesamt 4.778 Promovierte zur Teilnahme an der fünften Befragung eingeladen (davon 4.616 per E-Mail und 162 per Post). Innerhalb der Kontaktierung der fünften Welle verweigerten 17 Personen ihre Teilnahme und sieben weitere konnten wegen veralteter Kontaktdaten nicht mehr erreicht werden.

[Erhebungsunterlagen] Die Erhebungsunterlagen der fünften Welle wurden eine Woche nach dem postalischen Vorankündigungsschreiben an 4.778 Promovierte per E-Mail oder Brief verschickt. Diese enthielten das Einladungsschreiben inklusive Link zur Onlinebefragung und den genannten Preise für die Verlosung nach Befragungsende sowie den persönlichen Zugangscode. Wie bereits bei der vierten Welle wurde auch in diesem Anschreiben für die datenschutzrechtlichen Aspekte lediglich auf die Website verwiesen. Die Teilnehmenden mussten diesen Bestimmungen vor Beginn der Umfrage erneut per „Opt-In“-Verfahren aktiv zustimmen.

In der fünften Welle wurden – wie in den Vorwellen auch – drei Erinnerungsschreiben verschickt. Diese gingen an alle, die bis zu dem Zeitpunkt noch nicht (vollständig) an der aktuellen Befragung teilgenommen hatten. Im Gegensatz zu den Vorwellen wurde hier nicht mit unterschiedlichen Anschreiben gearbeitet, sondern lediglich zwischen postalischen und Anschreiben per E-Mail unterschieden. Erneut wurde der Onlinefragebogen mit der DZHW-Onlinebefragungssoftware „Zofar“ programmiert.

[Feldphase] Die fünfte Befragung startete am 20. März 2019 mit dem Versand der Einladungsschreiben. Die drei Erinnerungsschreiben wurden am 10., 17. und 36. Tag der Befragung versendet. Am 07. Mai 2019 endete die Feldphase.

[Rücklaufsteigernde Maßnahmen] Das erneute postalische Ankündigungsschreiben sollte wie im Jahr zuvor die Wichtigkeit, Bedeutung und Vertrauenswürdigkeit der Studie hervorheben. Außerdem wurde betont, dass jeder einzelne Werdegang und dadurch jede Teilnahme eine

hohe Relevanz für das Projekt hat. Die zur Kontaktpflege versandten Weihnachtskarten dienten zugleich der Panelbindung.

Die seit 2017 bestehende Projektwebsite wurde fortlaufend durch neue Inhalte ergänzt, sodass in den Anschreiben der fünften Welle auf diese verwiesen werden konnte.

Abbildung 4 zufolge sind die Erinnerungsschreiben hilfreich, um den Rücklauf zu steigern, sodass diese Maßnahme für die fünfte Welle erneut genutzt wurde.

Es wurde ebenfalls eine Verlosung von Incentives für die Teilnahme angekündigt. Es wurden zwei hochwertige Tablet-PCs von zwei verschiedenen Anbietern dem bzw. der Gewinner*in zur Auswahl angeboten. Darüber hinaus wurden zehn Wunschgutscheine als Nebenpreis verlost.

5 Rücklauf

Tabelle 6 gibt einen Überblick über die Anzahl der versandten Fragebögen sowie die Rücklaufquoten der einzelnen Befragungswellen. Von 19.000 versandten Befragungsunterlagen der Erstbefragung wurden 5.423 beantwortet. Letztlich verwertbar¹⁷ für den Befragungsdatensatz (und SUF) waren 5.408. Dies ist zum Teil durch Panelausfälle (s. u.) erklärbar. Neben klassischen Widerrufsfällen, sind in der Erstbefragung mehrere leere Fragebögen und einige außerhalb der Feldzeit eingegangen.¹⁸

[Panelausfälle] Während der Folgebefragungen wurden mehrere (schriftliche) Verweigerungen der Teilnahme verzeichnet – teilweise auch mit Widerruf der Befragungsdaten und Kontaktdaten. Bei vollständigem Widerruf der Teilnahme müssen die bestehenden Datensätze (auch der Vorwellen) gelöscht werden. Weiter wurden während der folgenden Befragungswellen zwei doppelte Teilnahmen aufgedeckt, um die auch der Befragungsdatensatz bereinigt werden musste. Die Anzahl der verwertbaren Datensätze der Erstbefragung hat sich daher auf 5.408 Fälle verringert.¹⁹

Tabelle 6: Brutto- und Nettorücklaufquote des DZHW-Promoviertenpanels 2014

	Welle 1	Welle 2	Welle 3	Welle 4	Welle 5
Befragungsunterlagen versendet	19.916	4.816	4.809	4.795	4.778
davon Versand bestätigt	19.900 ²⁰	4.815	4.806	4.789	4.771
Befragungsunterlagen Rücklauf	5.423	3.334	2.976	3.037	3.088
Befragungsunterlagen verwertbar ¹⁷	5.408	3.183	2.924	2.981	3.037
Rücklaufquote (verwertbar/versendet)	27,2 %	66,1 %	60,8 %	62,2 %	63,6 %
Anteil an Welle 1 (verwertbare Fälle/verwertbare Fälle Welle 1)	-	58,9 %	54,1 %	55,1 %	56,2 %
Teilnahme an allen bisherigen Wellen	-	100 %	84,4 %	71,4 %	64,1 %

¹⁷ Die verwertbaren Befragungsunterlagen umfassen alle Promovierenden, die in die Teilnahme eingewilligt haben und mindestens eine Antwort im Onlinefragebogen gegeben haben.

¹⁸ Drei Fragebögen sind nach Beendigung der Datenerfassung der Erstbefragung eingegangen und konnten daher nicht mehr aufgenommen werden. Vier Fragebögen sind unausgefüllt eingegangen, vier weitere mussten wegen stark unplausibler Angaben aussortiert werden. Zudem wurde ein zweifach ausgefüllter Fragebogen einer Person entfernt. Eine solche doppelte Teilnahme wurde während späterer Wellen erneut festgestellt, sodass zwei Fälle aufgrund einer doppelten Teilnahme gelöscht werden mussten. Zwei Personen verlangten die Löschung ihrer Befragungsdaten.

¹⁹ Abweichungen der gesamten Fallzahl im SUF der ersten und zweiten Welle (5.410) ergeben sich aus diesen unbereinigten Fällen (ein Widerrufsfall und ein Fall der doppelten Teilnahme).

²⁰ 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.

5.1 Welle 1

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 7). An 19 Hochschulen mit Promotionsrecht fand keine Befragung statt, da dort im untersuchten Zeitraum keine Promotion abgeschlossen wurde.

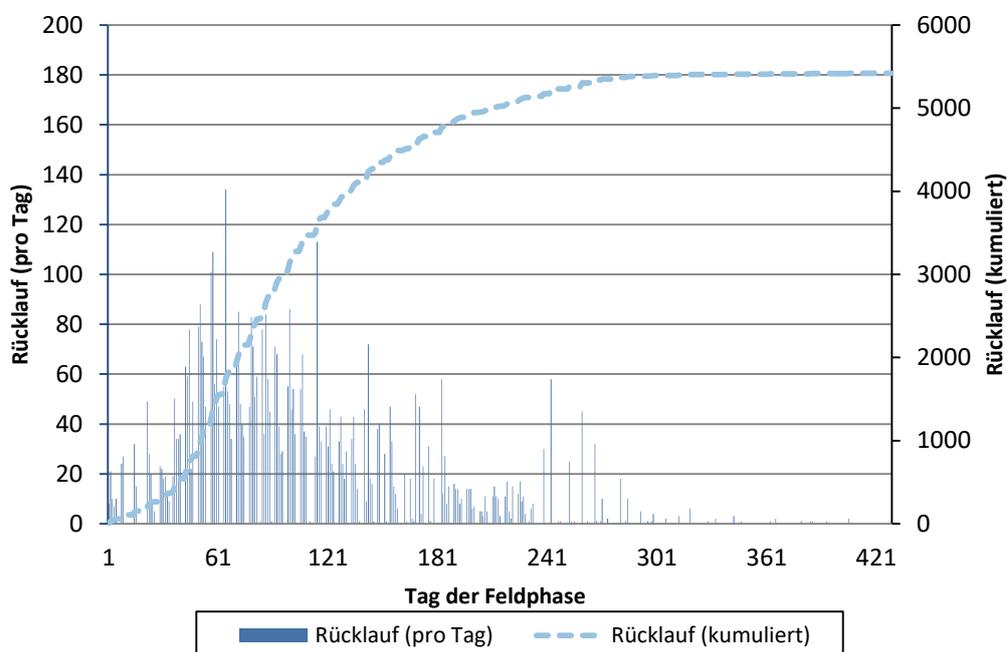
Tabelle 7: Beteiligung der Hochschulen

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

Von insgesamt 19.916 versandten Fragebögen wurden 5.423 Fragebögen von den Befragten zurückgeschickt (vgl. Tabelle 6). Abzüglich leerer, doppelter und unlesbarer Fragebögen wurden 5.410 Fragebögen elektronisch erfasst (vgl. Kapitel 6.3). Zwei weitere Fälle mussten im weiteren Panelverlauf gelöscht werden (ein Widerrufsfall und ein Fall der doppelten Teilnahme), wodurch sich die aktuelle Zahl von 5.408 Fällen ergibt.

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, Welle 1



5.2 Welle 2

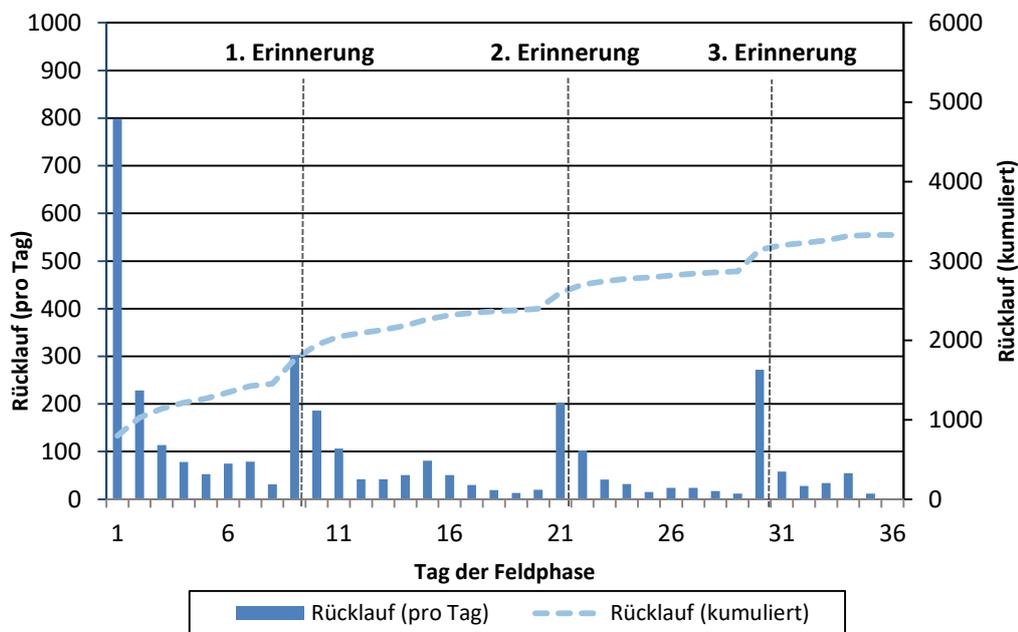
4.816 Personen, also 89 Prozent der 5.410 Teilnehmer(innen) der ersten Welle, erklärten sich mit einer Kontaktierung für weitere Befragungen einverstanden. An der zweiten Befragung beteiligten sich insgesamt 3.188 Promovierte, wovon 3.184 Fälle auswertbar waren. Ein Fall war hiervon aufgrund fehlender Zustimmung zum Datenschutz nicht verwertbar. 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 ersten Folgebefragung (zweite Welle) eingeladen. Am ersten Tag der Befragung nahmen bereits 24 Prozent und in den ersten drei Tagen 35 Prozent der an der Welle 2 teilnehmenden Befragten teil. Außerdem gab es deutliche Effekte der Erinnerungen, die jeweils einen erkennbaren Anstieg des Rücklaufs zur Folge hatten.

Der Anteil der verwertbaren Fälle der zweiten Welle an der Erstbefragung beträgt circa 60 % und die Rücklaufquote rund 66 % (Anteil an verwertbaren Fällen zu versendeten Einladungen).

[Panelausfälle] Nachdem die Erstbefragung durch Erstkontaktierung seitens der eigenen Hochschule durch postalisch versandte Befragungsunterlagen (Paper-Pencil-Befragung) stattgefunden hat, ist für die Zweitbefragung das DZHW durch ein per E-Mail versandtes Anschreiben mit Einladung zur Online-Befragung herantreten. Eine aktive Verweigerung der Panelteilnahme in schriftlicher Form (E-Mail) ist nur in drei Fällen verzeichnet gewesen. In 150 Fällen wurde jedoch der Befragungslink zwar aufgerufen, aber der Fragebogen nicht beantwortet (leerer Fragebogen). Diese weiche Form der Verweigerung war in der zweiten Befragungswelle im Vergleich zu den anderen Wellen am höchsten. Da diese Personen jedoch auch in den darauffolgenden Wellen zur Befragung eingeladen worden sind, sind diese keine Ausfälle für das gesamte Panel.

Durch die zuvor erfolgten Zwischenkontaktaktionen war der Ausfall aufgrund nicht zustellbarer Einladungen zudem sehr gering (ein Fall).

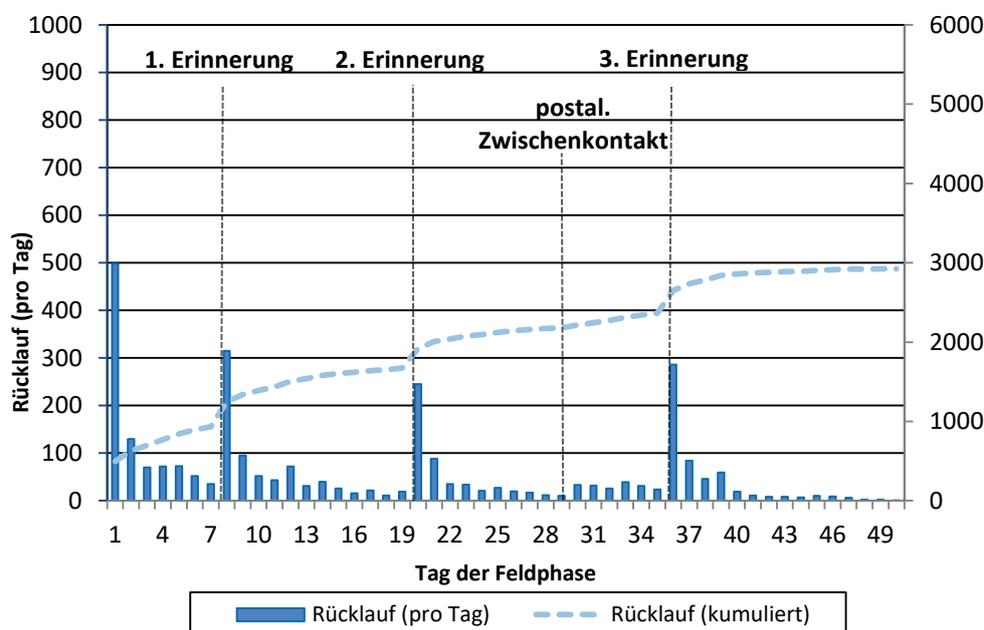
Abbildung 3: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, Welle 2



5.3 Welle 3

An der dritten Befragung beteiligten sich 2.977 Promovierte, davon sind die Angaben von 2.924 Personen verwertbar. Als nicht verwertbar gelten alle Fragebögen, die lediglich aufgerufen wurden, ohne dass mindestens eine Frage beantwortet wurde (49 Fälle), oder bei denen keine Zustimmung zu den Datenschutzbestimmungen gegeben wurde (3 Fälle). Die Rücklaufquote von 60,8 Prozent bezieht sich auf das Verhältnis von verwertbaren Fällen zu den versendeten Einladungen. Die Quote ist geringfügig niedriger als in der Vorgängerwelle. In Abbildung 4 sind der Rücklauf seit Beginn der Befragung und die Zeitpunkte der drei Erinnerungskaktionen und der Zwischenkontaktaktion dargestellt. Bereits am ersten Tag der Befragung nahm ein Großteil der Befragten teil. Die Erinnerungsschreiben hatten jeweils einen erkennbaren Anstieg des Rücklaufs zur Folge (Abbildung 4).

Abbildung 4: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, Welle 3



[Panelausfälle] Das Promoviertenpanel 2014 ist auch in der dritten Welle von paneltypischen Ausfallprozessen²¹ betroffen. Dazu gehören vor allem die grundsätzliche Verweigerung der Teilnahme an Folgebefragungen (14 Fälle) und die Nichtteilnahme (nur) an der aktuellen Befragung nach Aufrufen des Befragungslinks („weiche“ Verweigerung durch Senden eines leeren Fragebogens: 49 Fälle). Veraltete Adressdaten (drei Fälle) und die Ablehnung der Datenschutzbestimmungen (drei Fälle) sind hingegen nur Randphänomene in der dritten Befragungswelle. Der Datensatz zu der dritten Befragungswelle umfasst 54,1 Prozent der Befragten aus der ersten Befragungswelle. 84,4 Prozent der Teilnehmenden der dritten Welle haben sich bisher an allen drei Wellen beteiligt. Die Ausgangsstichprobe in Welle 3 umfasste im Vergleich zur zweiten Welle sieben Personen weniger. Davon hatten drei Personen in der Zwischenzeit die Bereitschaft zur Panelteilnahme zurückgezogen, von drei Personen lagen weder eine gültige E-Mailadresse noch eine gültige Postanschrift vor und eine Person wurde doppelt im Panel registriert.

5.4 Welle 4

Die Befragung der vierten Welle wurde insgesamt von 3.037 Panelteilnehmenden aufgerufen. Davon waren letztlich 2.981 Fälle verwertbar. In vier Fällen erfolgte keine Zustimmung zu den Datenschutzbestimmungen und in 52 Fällen wurde die Befragung ohne Abgabe einer Angabe angebrochen.

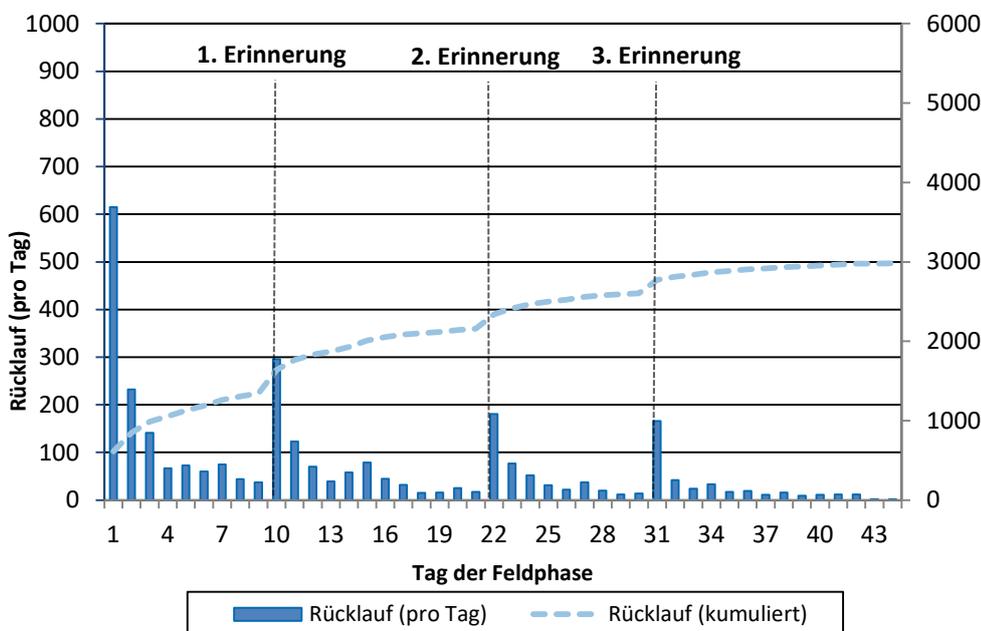
Die Rücklaufquote von 62,2 Prozent gibt das Verhältnis von verwertbaren Fällen zu den versendeten Einladungen an. Der Rücklauf ist somit im Vergleich zur Vorwelle leicht höher. In Abbildung 5 ist der Rücklauf aller verwertbaren Fälle (2.981) über die Feldphase hinweg dargestellt. Die Wirkung der Erinnerungsaktionen wird durch den deutlichen Tagesanstieg ersichtlich.

²¹ Zu paneltypischen Ausfallprozessen vgl. Schnell et al. 2005, S. 241.

Insgesamt umfasst der Befragungsdatensatz 55,1 Prozent der Befragten aus der ersten Befragungswelle. 71,4 Prozent der Teilnehmenden der vierten Welle haben sich bisher an allen Wellen beteiligt.

[Panelausfälle] Sechs Teilnehmende konnten im Vergleich zum Vorjahr aufgrund veralteter Adressdaten (drei Fälle) oder einer zwischenzeitlichen Verweigerung bzw. Widerruf (drei Fälle) nicht angeschrieben werden. Während der vierten Befragung gab es neun weitere Verweigerungen bzw. (aktive) Widerrufe und in den bereits erwähnten 52 Fällen eine „weiche“ Verweigerung durch den Interviewabbruch auf der ersten Seite. In vier Fällen erfolgte keine Zustimmung zu den Datenschutzbestimmungen, sodass diese Fälle nicht verwertbar waren. In sechs weiteren Fällen konnte die Einladung nicht zugestellt werden aufgrund veralteter Adressdaten.

Abbildung 5: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, Welle 4



5.5 Welle 5

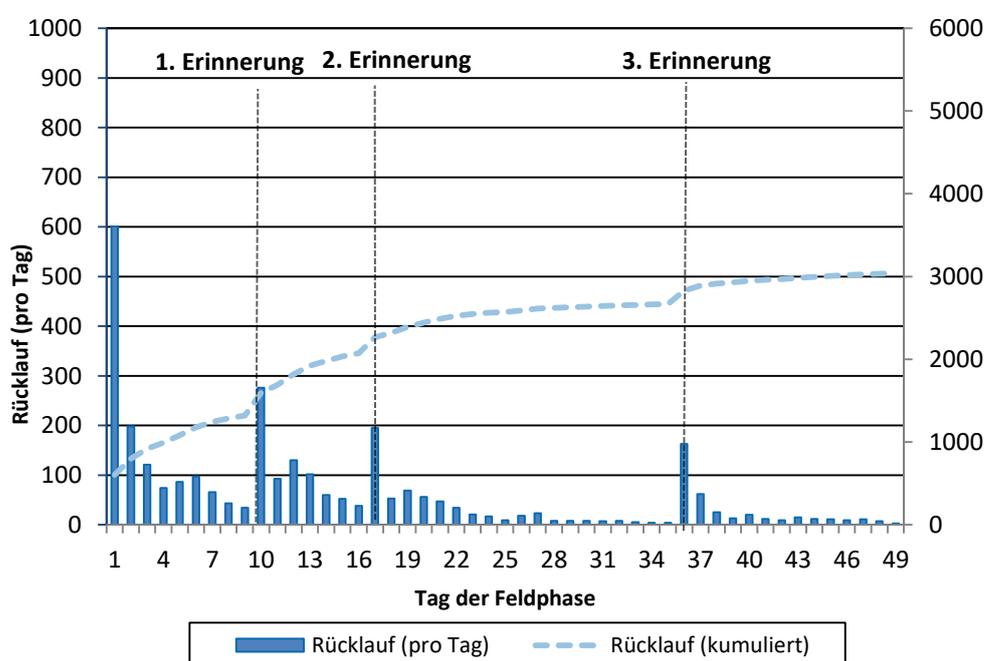
In der fünften Befragungswelle konnte der Rücklauf stabilisiert werden. Da alle Befragungsteilnehmer*innen mit gültiger Adresse angeschrieben worden waren und somit auch eine Teilnahme an nur einzelnen Wellen möglich ist, konnte eine leichte Steigerung des Rücklaufs von 56 Fällen im Vergleich zur Vorwelle erreicht werden. Insgesamt folgten 3.088 Personen der Einladung zur Befragungsstartseite. Davon waren 3.037 Befragungsdatensätze verwertbar. In 48 Fällen wurde nur die Startseite aufgerufen und die Befragung danach abgebrochen und in drei Fällen erfolgte eine weitere Beantwortung des Fragebogens, aber ohne eine Zustimmung zu den Datenschutzbestimmungen.

Mit der Rücklaufquote von 63,6 Prozent ist das Verhältnis der verwertbaren Fälle zu den versandten Einladungen zur fünften Befragung im Vergleich zu den Vorwellen leicht angestiegen. Der Befragungsdatensatz umfasst 56,2 Prozent der Befragten aus der ersten Befragungswelle.

64,1 Prozent der Teilnehmenden der fünften Welle haben sich bisher an allen Befragungen des Promoviertenpanels 2014 beteiligt.

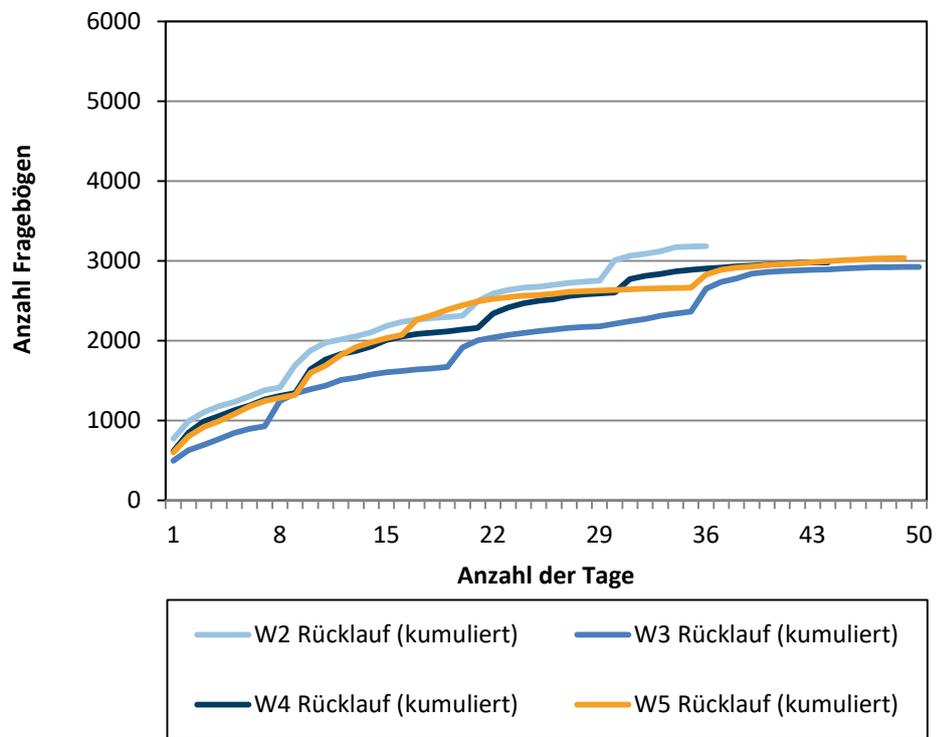
[Panelausfälle] Die Anzahl der Personen, die zur fünften Welle eingeladen werden konnten, hat sich zur Vorwelle um 17 Fälle verringert. Davon entfallen sieben aufgrund ungültiger Adressdaten und der Rest durch (aktive) Verweigerung bzw. Widerruf. Während der Feldphase waren weitere 17 Panelausfälle aufgrund von Teilnahmeverweigerung (meist auch für weitere Befragungen) und sieben Ausfälle aufgrund von veralteten Adressdaten zu verzeichnen.

Abbildung 6: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, 5. Welle



In Abbildung 7 ist der Rücklauf der zweiten bis fünften Befragung des Promoviertenpanels dargestellt. Im Vergleich zur dritten Welle weisen die Verläufe der zweiten, vierten und fünften Welle einen steileren Verlauf sowohl kurz nach Befragungsstart als auch nach den einzelnen Erinnerungsaktionen auf. In den beiden letzten Wellen wurde vor Beginn der Befragung das postalische Vorankündigungsschreiben verschickt. Bei der dritten Welle wurde dieses während der Befragung versendet. Die Rückläufe nähern sich zum Befragungsende an, jedoch liegen die Rücklaufzahlen- und verläufe der vierten und fünften Welle leicht über der dritten Welle.

Abbildung 7: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf, Welle 2 bis 5



6 Datenaufbereitung

Die im Folgenden beschriebenen Schritte der Datenaufbereitung erfolgten im Anschluss an die Datenerhebung. Die in den Kapiteln 6.1 bis 6.3 beschriebenen Arbeitsschritte – Datenübertragung, Codierung offener Angaben, Datenprüfung und Datenbereinigung – wurden von dem Primärforschungsprojekt durchgeführt. Die Generierung der Variablen (Kapitel 6.4) wurde sowohl von dem Primärforschungsprojekt als auch im Rahmen der Datenedition des FDZ-DZHW vorgenommen. Die Erstellung der Datensätze, Vergabe von Variablennamen, Variablenlabels und Wertelabels sowie die Codierung fehlender Werte (Kapitel 6.5 bis 6.7) wurden vom Primärforschungsteam in Absprache mit dem FDZ erstellt (Kapitel 6.2 bis 6.7). Die Schritte der Gewichtung und Anonymisierung werden in den Kapiteln 7 und 8 gesondert ausgeführt.

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 Erstbefragung 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.²²

[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.

Ab der zweiten Befragungswelle wurden die Daten mit dem Online-Befragungstool „Zofar“ erhoben und folglich kein Codeplan oder manuelle Datenerfassung verwendet. Der Rohdatensatz wurde auf einem geschützten Server gespeichert.

6.2 Codierung offener Angaben

[Erstbefragung] 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 halb-

²² 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.

offene Fragen wurden keine neuen Variablen mit numerischen Codierungen erstellt, sondern die Nennungen nur – sofern möglich – den vorhandenen (geschlossenen) Antwortkategorien zugeordnet. Einzelne offene Fragen wurden nicht vercodet, weil sie vorwiegend als Kontextinformationen zur Codierung anderer offener Angaben erfasst wurden.²³ Die durch das Primärforschungsprojekt vorgenommenen Codierungsentscheidungen wurden unverändert beibehalten.

[Folgebefragungen] Offene Angaben wurden wenn möglich anhand bereits vorliegender Codierschemata aus den Vorwelen und anhand von Klassifikationsschlüsseln der amtlichen Statistik (z. B. Klassifikation der Berufe) codiert. Bei einigen der offenen Fragen wurden keine neuen Variablen erstellt, sondern die Nennungen – sofern möglich – den vorhandenen (geschlossenen) Antwortkategorien zugeordnet. Einzelne offene Fragen wurden nicht codiert, da sie vorwiegend als Kontextinformationen zur Codierung anderer offener Angaben erfasst wurden.²³

Zur Codierung der beruflichen Tätigkeiten in Form von Codes nach KldB2010 wurde die Firma STR Coding beauftragt (die Dokumentation zur Codierung ist im Metadatenystem abrufbar). Es wurden neben den offen erhobenen Berufsangaben zur Berufsbezeichnung, dem hauptsächlichen Tätigkeitsbereich und den beruflichen Aufgaben auch die angegebene berufliche Stellung und in Einzelfällen weitere Kontextinformationen, wie der Wirtschaftsbereich (Branche) und das Promotionsfachgenutzt. Der vergebene Code bildet das Tätigkeitsfeld und die berufliche Stellung ab. So werden bei gleichbleibender Stelle Veränderungen der Tätigkeiten (wissenschaftlich, administrativ oder beratend etc.) und auch Veränderungen wie der Aufstieg in Führungspositionen abgebildet. Eine Änderung des Berufscodes ist demnach aber nicht immer gleichbedeutend mit einem Jobwechsel.

In Tabelle 8 sind die codierten Merkmale sowie die verwendeten Codierlisten der jeweiligen Welle dargestellt. Der Datensatz des Primärforschungsteams beinhaltet die codierten numerischen Variablen sowie die offenen Nennungen. Im SUF und CUF hingegen werden zum Datenschutz nur die numerischen Variablen ausgegeben (vgl. Kapitel 8). Die Ausprägungen der einzelnen Variablen sind für den SUF und CUF im Metadatenuchsystem²⁴ dokumentiert (vgl. auch Kapitel 6.6).

²³ Dies betrifft die beruflichen Tätigkeitsbereiche und Aufgaben. Die offenen Angaben dienen als zusätzliche Informationen für die Codierung der ebenfalls offen abgefragten Berufsbezeichnung.

²⁴ <https://metadata.fdz.dzhw.eu/#!/de/studies/stu-phd2014>

Tabelle 8: Vercodete Merkmale und verwendete Codierlisten (* für Welle 1-5)

Variablenamen ²⁵	Merkmal	Codierliste
d1prohs	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
d1profach1-3, b1studber1-4, b1abs1-4	Promotionsfach, Studienfach, Studienabschluss	Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2015/2016 und SoSe 2016)
c1eberuf	Beruf	Destatis Klassifikation der Berufe 2010
c*jberuf [c1-5jberuf]	Beruf	Destatis Klassifikation der Berufe 2010
c*taet1plz – c*taet10plz	Ort der beruflichen Tätigkeit	Postleitzahlenverzeichnis
c*taet1land – c*taet6land	Land der beruflichen Tätigkeit	Destatis Staats- und Gebietssystematik 2014
m*studkomplla, m*studsempla, m*studprakla, m*studkursla, m*studsonla	Land des Studienaufenthaltes	Destatis Staats- und Gebietssystematik 2014
m*grad1land, m*grad2land, m*grad3land, m*grad4land	Land des Auslandsaufenthaltes	Destatis Staats- und Gebietssystematik 2014
m*plandaofla, m*plandamfla, m*planzeitla, m*planforla, m*planwbila, m*plansonla	Land des geplanten Auslandsaufenthaltes	Destatis Staats- und Gebietssystematik 2014
k*natausl	Nationalität	Destatis Staats- und Gebietssystematik 2014
c*weibfarzt, c*weibzarzt, c*weibpsycho, c*weibtarzt, c*weibfanw	Sonstige berufsständische Weiterbildung	Einordnung in vorgegebene Kategorien der berufsständigen Weiterbildungen (Frage 2.2), wenn möglich
c*jbranche	Sonstige Wirtschaftsbereiche	Einordnung in vorgegebene Kategorien der Wirtschaftsbereiche (Frage 1.13), wenn möglich

²⁵ Die angegebenen Variablenamen wurden von der Primärforschungsgruppe vergeben.

x*anmerkungen	Offene Anmerkungen	Soweit möglich wurden verwertbare offene Angaben in die Angaben des jeweiligen Falls eingearbeitet
---------------	--------------------	--

Für die folgenden Variablen wurde auf eine Codierung offener Angaben verzichtet: Angabe des Promotionspreises (nur Welle 3, Frage 4.3), Tätigkeitsbereiche und berufliche Aufgaben (Welle 2 bis Welle 5, als Berufscodierung in c1eberuf bzw. c*jberuf codiert), die Kategorie für sonstige Angaben bei der Abfrage der Institution an der eine Professur angesiedelt ist (nur Welle 3, Frage 2.9) sowie ebenfalls die unter „Sonstige“ angegebene Art der Professur (Welle 3 bis Welle 5), bei den unter „Sonstiges“ offen genannten Personalkategorien (Welle 2 bis Welle 5) und der unter „sonstiger“ offen abgefragter beruflicher Abschluss des Partners oder der Partnerin (Welle 2 bis Welle 5).

6.3 Datenprüfung und Datenbereinigung

[Manuelle Vorkorrektur der Erstbefragung] 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.²⁶ 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).

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 (Welle 1, 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 ab 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 Variable in dem für diese Variable definierten Wertebereich lag.

²⁶ Die Zahl der vorgenommenen Korrekturen wurde nicht zentral, sondern nur auf den Papierfragebögen dokumentiert und ist daher nicht mehr systematisch rekonstruierbar.

- *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. Detaillierte Informationen zur Datenedition des Primärforschungsteams finden sich in den Dateneditionskonzepten und im Datenversionierungskonzept.

Bei fehlenden, fehlerhaften oder unplausiblen Werten wurde zunächst mit Hilfe des Papierfragebogens bzw. des Befragungsdatensatzes 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 und von mindestens einer weiteren Person geprüft.

[Löschung von Fällen] Im Panelverlauf wurden Fälle aus dem Datensatz entfernt. Zwei Fälle wurden gelöscht, weil eine doppelte Teilnahme festgestellt wurde (zwei Fälle), wenn weniger als eine Frage beantwortet worden war (fünf Fälle) oder wenn zu viele Inkonsistenzen vorhanden waren (drei Fälle). In drei Fällen musste der komplette Fall aus dem Datensatz genommen werden, da ein Widerruf vorlag.

In der ersten Befragungswelle wurden zehn Fälle und in bzw. nach der zweiten Befragungswelle fünf Fälle gelöscht. In Welle 3 wurde ein Fall gelöscht, da der/die Befragte um Löschung aller Daten gebeten hatte. Drei weitere Fälle konnten wegen fehlender Teilnahmeerlaubnis nicht verwertet werden. In Welle 4 musste ein kompletter Fall gelöscht werden, da sich eine doppelte Teilnahme bestätigt hatte. Weiter mussten in der vierten Welle vier Fälle wegen fehlender Zustimmung zu den Datenschutzbestimmungen gelöscht werden. In Welle 5 musste ein weiterer Fall komplett gelöscht werden aufgrund eines Widerrufs.

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 um Variablen mit numerischen Codierungen von ursprünglich offenen Nennungen (vgl. Kapitel 6.2). Zum anderen wurden Variablen für den SUF und CUF aus Datenschutzgründen verändert (vgl. Kapitel 8). Außerdem wurden im Forschungsfeld häufiger benötigte Variablen aus den Werten einer oder mehrerer Quellvariablen generiert (z. B. Ableitung der Kinderzahl aus den Angaben im Tableau zu Angaben über Kinder). Generierte Variablen können über einen Zusatz im Variablenlabel identifiziert werden (generiert aus VarX). Generierte Variablen sind in den Variablenplänen gesondert aufgeführt und durch einen grünen Rahmen als „generierte Variable“ gekennzeichnet.

6.5 Erstellung der Datensätze

[Zusammenführung der Wellen] Die Daten der ersten bis fünften Befragungswelle wurden zwar separat erhoben, jedoch in einem Datensatz zusammengeführt. Die Zuordnung der Fälle erfolgte über die im Rahmen der ersten Feldphase vergebenen Identifikationsnummern der Befragten (vgl. Kapitel 4).

[Erstellung von Personen- und Episodendatensatz] Die so zusammengeführten Daten wurden vom FDZ für den SUF und CUF in zwei getrennten Datensätzen abgelegt. Der Personendatensatz enthält den Großteil der Befragungsdaten sowie die zusätzlich generierten Variablen. Für jede befragte Person existiert eine Datenzeile (wide-Format). Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehörigen Fragen im Fragebogen.

Der Episodendatensatz enthält die Antworten aus den Kalendarien der zweiten bis zur fünften Welle (vgl. Kapitel 6.4). 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), welche monatsgenau den Anfang- und Endzeitpunkt der Tätigkeit und die Episodennummer angibt. Die Struktur entspricht der gängigen Struktur für Episodendaten (vgl. Scherer und 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 Tätigkeitsart Erwerbstätigkeit sind detailliertere Informationen jedoch in den entsprechenden Variablen des Personendatensatzes enthalten. Die Daten dieser Variablen können mit den Episodendaten verbunden werden. Das Zusammenführen von Personendatensatz und Episodendatensatz wird über die Identifikationsnummer der Person (Variable: pid) ermöglicht.

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

6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels

[Variablen- und Wertelabelvergabe] Für Variablen- und Wertelabels wurden Formulierungen des Fragebogens übernommen oder prägnante Kurzformen dieser Formulierungen gewählt. Dabei basieren die Variablenlabels in der Regel auf dem entsprechenden Fragetext. Grundlage für die Wertelabels sind je nach Fragetyp die Texte der Antwortoptionen bzw. eine Kombination der Texte von Frage und Antwortoption. Bei generierten Variablen, denen bestimmte Klassifikationen zugrunde liegen, wurden für die Wertelabels die Bezeichnungen der Schlüssel der Klassifikation wortgetreu übernommen.

[Variablenbenennung] Vom Primärforschungsteam wurden die Variablennamen nach einem festen Schema vergeben. Der erste Buchstabe des Variablennamens kennzeichnet den Themenbereich, dem eine Variable zugeordnet werden kann (Beispiel: „c“ bedeutet Themenbereich

„Beschäftigungsverlauf, Tätigkeitsmerkmale“). Die zweite Stelle kennzeichnet die Welle, in der die Variable erhoben wurde.²⁷ Sofern eine Variable einem bestimmten Variablenblock zugeordnet werden kann, wurde ein Variablenstamm vergeben.²⁸ Ein sprechender Namenszusatz gibt schließlich den konkreten Inhalt der jeweiligen Variable wieder.²⁹ Textvariablen zu offenen Fragen wurden mit der Endung „txt“ gekennzeichnet. Eine Auflistung der Themenbereiche und Variablenstämme, die für die Erstellung der Variablennamen verwendet wurden, befindet sich in Tabelle 9.

²⁷ Beispiel: c3 bedeutet, dass die Variable aus der dritten Welle stammt

²⁸ Beispiel: c3j bedeutet, dass die Variable zum Block „Aktuelle/letzte Stelle“ gehört

²⁹ Beispiel: c3jbranche kennzeichnet die Variable „Welchem Wirtschaftsbereich gehören/gehörten der Betrieb bzw. die Einrichtung, in dem/der Sie arbeiten/arbeiteten schwerpunktmäßig an?“

Tabelle 9: Systematik der Themenbereiche und Variablenstämme

Themenbereiche	Variablenstämme
A	Vorhandene Kompetenzen
B	Studienverlauf mop = Abschlussmonat jp = Abschlussjahr fachsem = Fachsemester examnote = Examensnote punkte = Punktzahl abs = Abschlussart studber = Studienbereich hs = Hochschule
C	Beschäftigungsverlauf; Tätigkeitsmerkmale job = Angaben im Tätigkeitstableau zu = Empfehlungen für berufliche Zukunft ak = Wissenschaftliche Aktivitäten ko = Kontakte neustell = Stelle nach Abschluss der Prom. e = Erste Stelle j = Aktuelle/letzte Stelle zuf = Bereichsspezifische Zufriedenheit stud = Tätigkeiten zu Studienzeiten weib = berufsständische Weiterbildung ww = wissenschaftliche Weiterqualifizierungen beaus = antizipierte Beschäftigungsaussichten
D	Promotionsverlauf; Rahmen- bedingungen der Promotion pv = Promotionsvereinbarungen pub = wissenschaftliche Publikationen kon = Konferenzen/Tagungen mob = Mobilitätserfahrungen pat = Patente nw = Netzwerken bpraxis = Berufspraxis außerhalb der Wiss. fin = Finanzierung
K	Soziales Umfeld (Partnerschaft, Kinder, Demografie, Schulab- schluss und soziale Herkunft) geb = Geburt (Datum/Ort) part = Partnerschaft kind = Kinder studb = Studienberechtigung nat = Staatsangehörigkeit
P	Persönlichkeitseigenschaften, Lebensziele und Motive m = Promotionsmotive asku = Allgemeine Selbstwirksamkeit (Kurzskala) kont = Kontrollüberzeugungen lz = Wichtigkeit Lebensziele bfi = Big-Five-Inventary mob = Mobilitätsbereitschaft

S	SSCO-Dimensionen	<p>strs = Structure_Sicherheit und Stabilität strr = Structure_Regelungen und Kontrolle strg = Structure_Grundlgn./Ausstattung /Beding./Gelegenheitsstruk.</p> <p>chak = Challenge_Kooperatives Forschen chal = Challenge_Leistungsdruck allgemein chav = Challenge_Veröffentlichungsdruck</p> <p>orid = Orientation_Interdisziplinarität orii = Orientation_Internationalität</p> <p>supf = Support_Fachliche Unterstützung supe = Support_Emotionale Unterstützung supk = Support_Karriereplanung supw = Support_Wissenschaftl. Aktivitäten sups = Support_Schlüsselkompetenzen supa = Support_Ausbau wissenschaftlicher Netzwerke</p> <p>orif = Orientation_Forschungsorient. orip = Orientation_Praxisorient.</p>
W	Wissenschaftliche Aktivitäten nach der Promotion	<p>pub = wissenschaftliche Publikationen kon = Konferenzen pat = Patente wisskar = wissenschaftliche Karriere prof = Professur gut = Gutachtertätigkeiten foan = Anträge zur Forschungsförderung mitgl = Mitgliedschaften in Berufsverbänden/Fachgesellschaften asv = akademische Selbstverwaltung</p>
G	Nicht-Erwerbstätigkeit	
M	Mobilität	plan = Mobilitätspläne
X	Informationen zum Fragebogen	

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 10 stellt die verwendete Missingsystematik dar.

Die fehlenden Angaben lassen sich drei 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. Irrelevanz der Frage für die/den Befragte/n 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.³⁰

Tabelle 10: 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	<i>(nicht vergeben)</i>	
-929 bis -920: Andere fehlende Werte	-929	Datenverlust

^a Dieser Wert ist 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. bei einer offenen Angabe, die nicht vercodet werden konnte, da sie nicht lesbar ist.

³⁰ Eine mögliche vierte Gruppe würde spezielle Missingcodierungen umfassen, die im Rahmen der Datenaufbereitung eines konkreten Datensatzes nur für einzelne Items vergeben wurden.

7 Gewichtung

Die Gewichtung der Daten dient dem Ausgleich von Verzerrungen im Vergleich zur definierten Grundgesamtheit. Es folgt zunächst eine allgemeine Einführung in die Vorgehensweise und eine 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 in der Regel zwei Prozesse:

- **Designbedingte Verzerrung:** Disproportionalitäten werden bewusst erzeugt, um in bestimmten relevanten Subgruppen die Fallzahlen zu erhöhen.
- **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.³¹ Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (*Missing at Random, Missing Not at Random*), der einer Modellierung bedarf.³²

[Konzeptuelles Vorgehen] Da es sich bei dem DZHW-Promoviertenpanel um eine Vollerhebung handelt, wurden keine designbedingten Disproportionalitäten erzeugt, die ausgeglichen werden müssten. *Designgewichte* mussten folglich nicht erstellt werden, da alle Individuen die gleiche Auswahlwahrscheinlichkeit aufweisen. Für diesen Datensatz sollte jedoch eine Adjustierung mit Hilfe von *Ausfallgewichten* im Querschnitt erfolgen, die auf der Grundlage von Informationen über Teilnehmer*innen und Nichtteilnehmer*innen erzeugt werden.

In Zusammenarbeit mit Gewichtungsexpert*innen der Gesis und des DZHW wurde das Gewichtungsverfahren im Vergleich zum SUF der ersten beiden Erhebungswellen (DOI: 10.21249/DZHW:phd2014:2.0.0) optimiert. Weil das Studiendesign des Promoviertenpanels Ausfälle von Individuen in einzelnen Erhebungswellen zulässt, werden im vorliegenden SUF ausschließlich Querschnittsgewichte bereitgestellt. Dafür werden im ersten Schritt die wellenspezifischen Ausfallgewichte anhand eines logistischen Regressionsmodells mit Prädiktoren aus der Erstbefragung erstellt (propensity scores). Diese Ausfallgewichte werden im zweiten Schritt mithilfe eines Raking-Algorithmus auf die Merkmalsverteilung der Grundgesamtheit kalibriert. Dadurch ergeben sich im Vergleich insgesamt kleinere Gewichte mit geringeren Varianzen. Ausreißer i. S. v. sehr kleinen oder sehr großen Gewichten werden schließlich getrimmt.

Darüber hinaus wurde ein Fehler bei der Zuordnung von Promovierten von Berliner Hochschulen bei der Kalibrierung auf die Merkmalsverteilung in der Grundgesamtheit berichtigt. Im Zuge des aktuellen SUF Release werden also auch die Gewichte der ersten beiden Erhebungswellen korrigiert. In Tabelle 11 sind die bereitgestellten Gewichte dargestellt.

³¹ Insofern die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

³² Siehe grundlegend zu den unterschiedlichen Formen von Ausfallprozessen Rubin 1976.

Tabelle 11: Bereitgestellte Gewichte zum DZHW-Promoviertenpanel 2014

Variablenname	Beschreibung
x1gewi	Kalibriertes Gewicht für Querschnittsanalysen der 1. Erhebungswelle (getrimmt)
x2geaus	Ausfallgewicht für Querschnittsanalysen der 2. Erhebungswelle (getrimmt)
x2gewi	kalibriertes Ausfallgewicht (x2geaus) für Querschnittsanalysen der 3. Erhebungswelle (getrimmt)
x3geaus	Ausfallgewicht für Querschnittsanalysen der 3. Erhebungswelle (getrimmt)
x3gewi	kalibriertes Ausfallgewicht (x3geaus) für Querschnittsanalysen der 3. Erhebungswelle (getrimmt)
x4geaus	Ausfallgewicht für Querschnittsanalysen der 4. Erhebungswelle (getrimmt)
x4gewi	kalibriertes Ausfallgewicht (x4geaus) für Querschnittsanalysen der 4. Erhebungswelle (getrimmt)
x5geaus	Ausfallgewicht für Querschnittsanalysen der 5. Erhebungswelle (getrimmt)
x5gewi	kalibriertes Ausfallgewicht (x5geaus) für Querschnittsanalysen der 5. Erhebungswelle (getrimmt)

[Hinweise zur Anwendung der Gewichte] Bei den erstellten Gewichten handelt es sich um probability weights, die in Stata mit Hilfe .ado-spezifischer Optionen berücksichtigt werden können.³³ Das Gewicht x1gewi dient der Gewichtung der ersten Erhebungswelle, die Gewichte x2geaus und x2gewi dienen der Gewichtung bei Analysen der Zweiten-Welle-Daten usw. Die bereitgestellten Gewichte sind nur für Querschnittsanalysen geeignet. Da Panelaus- und Wiedereintritte zwischen Wellen möglich sind, wurde auf die Bereitstellung von Längsschnittgewichten im Rahmen des SUF verzichtet, können nutzerseitig jedoch selbst erstellt werden. Grundsätzlich 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 bei der Erstellung der Gewichte für die Wellen 1 bis 5 näher dargestellt.

7.2 Modellierung der Ausfallgewichte

[Propensity Score Matching] Für die Wellen 2 bis 5 wurden die Ausfallgewichte mithilfe von Propensity Scores berechnet.³⁴ Hierzu wurde anhand eines Logit-Schätzmodells die individuelle Teilnahmewahrscheinlichkeit einer Person für die entsprechende Welle vorhersagt (RP). In einem explorativem Vorgehen wurden individuelle Merkmale identifiziert, die einen Beitrag zur Erklärung der Teilnahmewahrscheinlichkeit leisten.³⁵ Um für jede Person die Teilnahmewahrscheinlichkeit schätzen zu können, wurden fehlende Werte bei allen Variablen als zusätzliche

³³ Siehe hierzu die Stata-Hilfe (Befehl: *help weights*).

³⁴ Das Verfahren entspricht seiner Logik nach dem *Propensity Score Matching*, das auf Rosenbaum und Rubin 1983 zurückgeht (siehe auch Blumenstiel und Gummer 2015).

³⁵ Siehe hierzu im Anhang das Schätzmodell.

Kategorie in das Modell integriert. Maßgebliches Kriterium für die Modellierung der Ausfallprozesse war, Modelle mit größtmöglicher Erklärungskraft bei gleichzeitiger Sparsamkeit zu finden, die die Varianzerhöhung so gering wie möglich halten. Herangezogen wurden dabei ausschließlich Prädiktoren aus der Erstbefragung.³⁶ Aus dem Modell konnte dann in einem zweiten Schritt die bedingte Teilnahmewahrscheinlichkeit abgeleitet werden, deren Kehrwert das Ausfallgewicht ($x_{2\text{geaus}} - x_{5\text{geaus}}$) für die jeweilige Welle darstellt: $w = 1/RP$.

[Schätzmodelle] Die Schätzmodelle, die zur Erstellung der Ausfallgewichte genutzt werden, beinhalten als Prädiktoren die Variablen *Geschlecht*, *Alter*, *Geburtsort im Ausland*, *Promotionsform*, *Fach*, *Promotionsnote*, *Faulheit (Big-5)* und das *Brutto-Einkommen*. Die Logit-Modelle der Wellen zwei bis fünf können dem Anhang in Tabelle 14, Tabelle 15, Tabelle 16 und Tabelle 17 entnommen werden.

7.3 Kalibrierung

[IPF-Raking-Verfahren] Die Ausfallgewichte der Wellen 2 bis 5 ($x_{2\text{geaus}} - x_{5\text{geaus}}$) wurden mithilfe des IPF-Raking-Verfahrens (Iterative Proportional Fitting) an die Randverteilung bekannter Merkmale in der Grundgesamtheit angepasst.³⁷ Da für das Raking-Verfahren Informationen zur Randverteilung der Merkmale ausreichen und keine Kreuztabellen notwendig sind, müssen kaum Zellen zusammengefasst und weniger Einzelfallentscheidungen getroffen werden. Die Anpassung der Gewichte folgt dabei einem iterativen Prozess, bis keine Verbesserung mehr möglich ist. Da für die Erstbefragung keine Ausfallgewichte erstellt wurden, erfolgt für die erste Befragung ausschließlich die Kalibrierung.³⁸

[Merkmalsverteilung der Grundgesamtheit] Der Abgleich der Stichproben mit der Grundgesamtheit erfolgte anhand der Merkmale Studienbereich, Geschlecht und Region der Promotionshochschule (Ost/West). Die Informationen zur Grundgesamtheit wurden den Daten des Statistischen Bundesamtes entnommen (Datenportal ICE-Land; Bestand: 50001). Personengruppen, deren Anteil in der jeweiligen Welle nicht dem Anteil in der Grundgesamt entspricht, werden unter Verwendung des kalibrierten Ausfallgewichts ($x_{1\text{gewi}} - x_{5\text{gewi}}$) so entsprechend hoch- bzw. runtergewichtet.

[Fehlende Werte] Bei der Verwendung der kalibrierten Ausfallgewichte gilt es zu beachten, dass Personen, die keine Angaben bei dem für das Raking relevanten Variablen (Geschlecht, Studienbereich, Bundesland der Promotionshochschule) gemacht haben, keine Kalibrierung durchgeführt werden konnte. Dementsprechend erhalten diese Personen keine kalibrierten Ausfallgewichte. Die Fallzahlen (n) können bei Analysen unter Verwendung kalibrierter Gewichte demnach geringer ausfallen.

³⁶ Versuche, den Panelcharakter der vorliegenden Datenbasis zu berücksichtigen, z. B. indem auch die Teilnahme (ja/nein), der Teilnahmezeitpunkt oder Prädiktoren aus weiteren vorherigen Wellen in den Schätzmodellen berücksichtigt werden, erhöhten zwar die Erklärungskraft der Ausfallmodelle, mussten aufgrund zu großer Gewichte und Varianzen jedoch verworfen werden.

³⁷ Das IPF-Raking Verfahren geht auf Deming und Stephan (1940) zurück (siehe auch Lomax und Norman (2016)) und wurde von Kolenikov (2014) im stata-.ado *ipfraking* implementiert.

³⁸ Technisch erfolgte dies, indem für alle Teilnehmenden der Erstbefragung im ersten Schritt ein Ausfallgewicht mit dem Wert 1 erstellt wurde, das im zweiten Schritt dann anhand des Raking-Verfahrens kalibriert wurde.

7.4 Trimmung

Um Extremwerte zu beseitigen, wurden alle Gewichte einer Trimmung nach Potter (1990) (vgl. auch Valliant et al. 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.

7.5 Verteilung der Gewichte

Die Verteilungsmaße der bereitgestellten Gewichte für die Wellen eins bis fünf lässt sich Tabelle 12 entnehmen:

Tabelle 12: Verteilung der Gewichte für Welle 1 bis 5

Gewicht	n	Min	Max	SD	Varianz
x1gewi	5.318	0,226	1,832	0,264	0,069
x2geaus	3.183	1,163	2,326	0,230	0,053
x2gewi	3.146	0,178	2,162	0,349	0,122
x3geaus	2.924	1,264	2,603	0,272	0,074
x3gewi	2.890	0,166	2,162	0,349	0,122
x4geaus	2.981	1,241	2,559	0,267	0,072
x4gewi	2.949	0,209	2,147	0,342	0,117
x5geaus	3.037	1,246	2,432	0,239	0,057
x5gewi	3.003	0,214	2,055	0,323	0,104

8 Anonymisierung der SUF und CUF

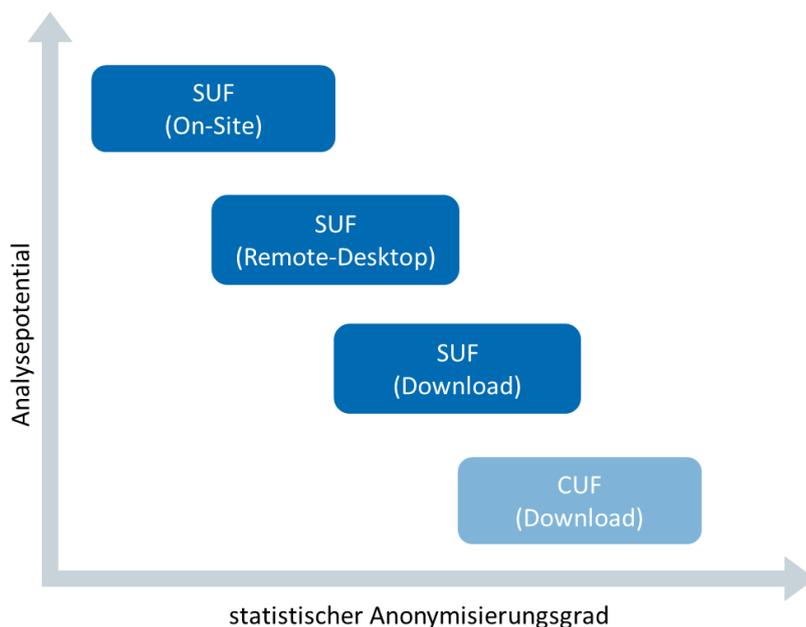
[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten³⁹, die in freiwilligen Befragungen durch das DZHW erhoben werden, gelten die EU-Datenschutz-Grundverordnung (EU-DSGVO) und das Bundesdatenschutzgesetz in seiner Neufassung vom 30. Juni 2017. Danach sind personenbezogene Daten für die Weitergabe zur wissenschaftlichen Sekundärnutzung (ohne Vorliegen einer Einverständniserklärung zur Sekundärnutzung der personenbezogenen Daten) in der Regel derart aufzubereiten, dass „die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden können“ (Art. 4 Abs. 5 DSGVO; s. auch Art. 89 DSGVO sowie Erwägungsgrund 26 DSGVO). Das heißt, für die Weitergabe von Daten aus wissenschaftlichen Forschungsprojekten an Dritte sind die Daten derart zu anonymisieren, dass kein Bezug zur Person mehr hergestellt werden kann.

[Datenzugang, Anonymisierungsgrad und Analysepotential] Das FDZ-DZHW stellt für das 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 (Abbildung 8).

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 II). Für jeden Zugangsweg wird eine andere SUF-Variante bereitgestellt, die unterschiedlich stark anonymisiert worden ist und entsprechend weniger oder mehr Informationen umfasst. Abbildung 8 gibt einen Überblick über den jeweiligen Grad der statistischen Anonymisierung und das damit verbundene Analysepotential. Im Folgenden werden die durchgeführten statistischen Anonymisierungsmaßnahmen in Abhängigkeit von Datenprodukt (SUF/CUF) und Zugangsweg erläutert.

³⁹ „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).

Abbildung 8: 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 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.⁴⁰ Für das Promoviertenpanel 2014 wurden die folgenden Quasi-Identifikatoren identifiziert, die sowohl in externen Datenquellen⁴¹ 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 13). Beispielsweise werden bei dem Merkmal „Geburtsort“ in dem SUF für die On-Site Nutzung die ersten drei Stellen der Postleitzahl, 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 An-

⁴⁰ 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.

⁴¹ Z.B. Studenten- und Prüfungsstatistik des statistischen Bundesamtes, Alumninetzwerke der Hochschulen oder auch Berufsnetzwerke.

gaben 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 13: Maßnahmen der statistischen Anonymisierung der Daten des DZHW-Promoviertenpanels 2014 nach Zugangsweg⁴²

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 ^a	Aggregation zu Studienbereichen ^a	Aggregation zu Fächergruppen ^a	Aggregation zu Fächergruppen ^a
Hochschule	Aggregation zu Hochschulart ^b	Aggregation zu Hochschulart ^b	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 ^c	Deutschland: Vier Bundesländer einzeln ausgewiesen; ansonsten Aggregation zu fünf Bundesländer-Gruppen Ausland: Aggregation zu Weltregionen ^c	Deutschland: Vier Bundesländer einzeln ausgewiesen; ansonsten Aggregation zu fünf Bundesländer-Gruppen Ausland: Aggregation zu Weltregionen ^c
Auslandsaufenthalte (Land)	Freigabe	Freigabe	Aggregation zu Weltregionen ^c	Aggregation zu Weltregionen ^c
Beruf		Aggregation zu	Aggregation zu	Aggregation zu

⁴² 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)
	Aggregation zu Berufsgruppen ^d	Berufshauptgruppen ^d	Berufshauptgruppen ^d	Berufshauptgruppen ^d
Personalkategorien	Freigabe	Freigabe	Aggregation: WHK zu sonstige; PD zu Dozent; akad. Oberrat a. Z. zu akad. Oberrat	Aggregation: WHK zu sonstige; PD zu Dozent; akad. Oberrat 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 ^c	Aggregation zu Weltregionen ^c
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 ^e	Top-Codierung ^e
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 ^d	Aggregation zu Berufsgruppen ^d	Aggregation zu Berufshauptgruppen ^d	Aggregation zu Berufshauptgruppen ^d
Antwortkategorien zur Gesundheit	Zusammenfassen mit der Kategorie „andere Gründe“	Zusammenfassen mit der Kategorie „andere Gründe“	Zusammenfassen mit der Kategorie „andere Gründe“	Zusammenfassen mit der Kategorie „andere Gründe“
Sonstige offene Angaben	Löschung	Löschung	Löschung	Löschung

^a Nach dem Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2015/2016 und SoSe 2016 von Destatis.

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

^c Nach der Staats- und Gebietsystematik 2014 von Destatis.

- ^d Nach Klassifikation der Berufe von 2010 von Destatis.
- ^e Angaben von vier und mehr Kindern wurden zu einer Kategorie zusammengefasst.

Anhang

Logit-Regressionen zur Erstellung der Panelfallgewichte:

Tabelle 14: Logit-Regression zur Erstellung des Ausfallgewichts für Welle 2

	Odds Ratio	Std. Err	z	P> z
Geschlecht				
Männlich	Ref.			
Weiblich	1,08	0,07	1,22	0,223
keine Angabe	0,86	0,39	-0,33	0,742
Alter				
23-30	1,10	0,08	1,29	0,196
31-34	Ref.			
35-39	1,05	0,08	0,66	0,509
40-44	1,25	0,18	1,61	0,108
45-49	2,37	0,49	4,21	0,000
50-60	1,96	0,43	3,05	0,002
61-79	1,23	0,47	0,53	0,594
keine Angabe	0,13	0,07	-3,65	0,000
Promotionsrahmen				
wiss. Mitarbeiter Haushaltstelle	0,97	0,08	-0,41	0,683
wiss. Mitarbeiter Drittmittelstelle	Ref.			
strukturiertes Promotionsprogramm	0,77	0,09	-2,18	0,029
Stipendienprogramm	0,82	0,09	-1,84	0,066
frei promovierend	0,99	0,09	-0,1	0,916
keine Angabe	0,45	0,19	-1,93	0,054
Promotionsnote				
summa cum laude	1,28	0,13	2,49	0,013
magna cum laude	1,16	0,09	1,96	0,050
cum laude	Ref.			
satis bene	0,87	0,33	-0,37	0,713
rite	0,99	0,21	-0,04	0,968
keine Angabe	1,17	0,36	0,5	0,619
Geburtsort				
Deutschland	Ref.			
In einem anderen Land	0,78	0,07	-2,56	0,010

keine Angabe	0,69	0,16	-1,59	0,112
"Ich bin bequem, neige zu Faulheit"				
trifft überhaupt nicht zu	0,78	0,06	-3,12	0,002
2	0,87	0,07	-1,82	0,069
3	Ref.			
4	1,29	0,16	2,06	0,04
trifft voll und ganz zu	1,01	0,24	0,04	0,97
keine Angabe	0,27	0,15	-2,35	0,02
Fächergruppe				
Geisteswissenschaften	1,09	0,13	0,69	0,49
Sport	0,90	0,32	-0,30	0,76
Rechts-, Wirtschafts- und Sozialwiss.	0,74	0,07	-3,39	0,00
Mathematik, Naturwiss.	Ref.			
Humanmed./Gesundheitswiss.	0,67	0,07	-3,98	0,00
Agrar-, Forst-, Ernährungswiss, Veterinärm.	1,05	0,16	0,34	0,73
Ingenieurwissenschaften	0,67	0,07	-3,97	0,00
Kunst, Kunstwissenschaft	1,46	0,40	1,39	0,16
keine Angabe	1,02	0,42	0,05	0,96
aktuelles/letztes Bruttoeinkommen				
bis 3200€	1,08	0,08	1,04	0,30
3201€ bis 5000€	Ref.			
über 5000€	0,96	0,08	-0,54	0,59
kein Einkommen	0,95	0,12	-0,44	0,66
keine Angabe	0,61	0,07	-4,34	0,00
Konstante	1,76	0,21	4,72	0,00
<i>N=5408; Pseudo R²=0,028</i>				

Tabelle 15: Logit-Regression zur Erstellung des Ausfallgewichts für Welle 3

	Odds Ratio	Std. Err	z	P> z
Geschlecht				
Männlich	Ref.			
Weiblich	1,10	0,07	1,49	0,14
keine Angabe	1,13	0,51	0,28	0,78
Alter				
23-30	1,04	0,08	0,57	0,57
31-34	Ref.			

35-39	1,01	0,08	0,16	0,87
40-44	1,25	0,17	1,64	0,10
45-49	1,90	0,36	3,38	0,00
50-60	1,75	0,37	2,66	0,01
61-79	2,17	0,86	1,97	0,05
keine Angabe	0,12	0,07	-3,71	0,00
Promotionsrahmen				
wiss. Mitarbeiter Haushaltstelle	0,95	0,08	-0,56	0,58
wiss. Mitarbeiter Drittmittelstelle	Ref.			
strukturiertes Promotionsprogramm	1,03	0,12	0,28	0,78
Stipendienprogramm	0,79	0,09	-2,14	0,03
frei promovierend	0,92	0,09	-0,91	0,36
keine Angabe	0,62	0,25	-1,17	0,24
Promotionsnote				
summa cum laude	1,32	0,13	2,89	0,00
magna cum laude	1,17	0,09	2,08	0,04
cum laude	Ref.			
satis bene	1,24	0,47	0,56	0,57
rite	0,96	0,20	-0,22	0,83
keine Angabe	0,52	0,17	-2,01	0,04
Geburtsort				
Deutschland	Ref.			
In einem anderen Land	0,68	0,07	-3,98	0,00
keine Angabe	0,90	0,21	-0,44	0,66
"Ich bin bequem, neige zu Faulheit"				
trifft überhaupt nicht zu	0,78	0,06	-3,21	0,00
2	0,90	0,07	-1,28	0,20
3	Ref.			
4	1,12	0,13	0,93	0,35
trifft voll und ganz zu	0,87	0,20	-0,61	0,54
keine Angabe	0,33	0,18	-2,01	0,05
Fächergruppe				
Geisteswissenschaften	1,19	0,14	1,43	0,15
Sport	0,68	0,23	-1,11	0,27
Rechts-, Wirtschafts- und Sozialwiss.	0,82	0,07	-2,23	0,03
Mathematik, Naturwiss.	Ref.			
Humanmed./Gesundheitswiss.	0,73	0,07	-3,21	0,00
Agrar-, Forst-, Ernährungswiss, Veterinärm.	0,91	0,13	-0,66	0,51

Ingenieurwissenschaften	0,82	0,08	-2,04	0,04
Kunst, Kunstwissenschaft	1,02	0,25	0,08	0,94
keine Angabe	1,06	0,43	0,15	0,88
aktuelles/letztes Bruttoeinkommen				
bis 3200€	1,01	0,08	0,11	0,91
3201€ bis 5000€	Ref.			
über 5000€	0,88	0,07	-1,63	0,10
kein Einkommen	0,83	0,10	-1,51	0,13
keine Angabe	0,56	0,06	-5,08	0,00
Konstante	1,45	0,17	3,16	0,00
<i>N=5408; Pseudo R²=0,025</i>				

Tabelle 16: Logit-Regression zur Erstellung des Ausfallgewichts für Welle 4

	Odds Ratio	Std. Err	z	P> z
Geschlecht				
Männlich	Ref.			
Weiblich	1,07	0,07	1,17	0,24
keine Angabe	0,83	0,37	-0,43	0,67
Alter				
23-30	1,05	0,08	0,70	0,48
31-34	Ref.			
35-39	0,89	0,07	-1,57	0,12
40-44	0,95	0,13	-0,36	0,72
45-49	1,81	0,34	3,13	0,00
50-60	1,52	0,32	2,03	0,04
61-79	1,25	0,47	0,60	0,55
keine Angabe	0,11	0,06	-3,87	0,00
Promotionsrahmen				
wiss. Mitarbeiter Haushaltstelle	0,93	0,08	-0,94	0,35
wiss. Mitarbeiter Drittmittelstelle	Ref.			
strukturiertes Promotionsprogramm	0,95	0,11	-0,45	0,66
Stipendienprogramm	0,87	0,09	-1,26	0,21
frei promovierend	0,85	0,08	-1,78	0,08
keine Angabe	0,60	0,25	-1,25	0,21
Promotionsnote				

summa cum laude	1,27	0,12	2,44	0,02
magna cum laude	1,12	0,09	1,51	0,13
cum laude	Ref.			
satis bene	1,07	0,41	0,18	0,86
rite	0,75	0,16	-1,37	0,17
keine Angabe	0,59	0,19	-1,66	0,10
Geburtsort				
Deutschland	Ref.			
In einem anderen Land	0,66	0,06	-4,43	0,00
keine Angabe	1,06	0,25	0,23	0,82
"Ich bin bequem, neige zu Faulheit"				
trifft überhaupt nicht zu	0,84	0,07	-2,26	0,02
2	0,95	0,07	-0,72	0,47
3	Ref.			
4	1,28	0,15	2,10	0,04
trifft voll und ganz zu	1,12	0,26	0,50	0,62
keine Angabe	0,68	0,35	-0,75	0,45
Fächergruppe				
Geisteswissenschaften	0,99	0,12	-0,09	0,93
Sport	0,71	0,24	-1,00	0,32
Rechts-, Wirtschafts- und Sozialwiss.	0,87	0,08	-1,59	0,11
Mathematik, Naturwiss.	Ref.			
Humanmed./Gesundheitswiss.	0,69	0,07	-3,66	0,00
Agrar-, Forst-, Ernährungswiss, Veterinärm.	0,90	0,13	-0,69	0,49
Ingenieurwissenschaften	0,84	0,08	-1,76	0,08
Kunst, Kunstwissenschaft	1,10	0,27	0,37	0,71
keine Angabe	1,07	0,44	0,17	0,87
aktuelles/letztes Bruttoeinkommen				
bis 3200€	1,02	0,08	0,31	0,76
3201€ bis 5000€	Ref.			
über 5000€	0,91	0,07	-1,18	0,24
kein Einkommen	0,95	0,12	-0,39	0,69
keine Angabe	0,55	0,06	-5,22	0,00
Konstante	1,61	0,19	4,04	0,00
<i>N=5408; Pseudo R²=0,026</i>				

Tabelle 17: Logit-Regression zur Erstellung des Ausfallgewichts für Welle 5

	Odds Ratio	Std. Err	z	P> z
Geschlecht				
Männlich	Ref.			
Weiblich	1,05	0,07	0,85	0,40
keine Angabe	1,11	0,51	0,23	0,82
Alter				
23-30	1,05	0,08	0,67	0,50
31-34	Ref.			
35-39	0,99	0,08	-0,11	0,92
40-44	1,34	0,19	2,10	0,04
45-49	2,10	0,41	3,78	0,00
50-60	1,59	0,33	2,20	0,03
61-79	1,00	0,37	0,00	1,00
keine Angabe	0,08	0,05	-3,93	0,00
Promotionsrahmen				
wiss. Mitarbeiter Haushaltstelle	0,83	0,07	-2,32	0,02
wiss. Mitarbeiter Drittmittelstelle	Ref.			
strukturiertes Promotionsprogramm	0,98	0,12	-0,20	0,84
Stipendienprogramm	0,85	0,09	-1,53	0,13
frei promovierend	0,89	0,08	-1,28	0,20
keine Angabe	0,55	0,23	-1,44	0,15
Promotionsnote				
summa cum laude	1,27	0,12	2,44	0,02
magna cum laude	1,18	0,09	2,15	0,03
cum laude	Ref.			
satis bene	0,89	0,34	-0,30	0,76
rite	1,17	0,24	0,75	0,45
keine Angabe	0,83	0,26	-0,61	0,54
Geburtsort				
Deutschland	Ref.			
In einem anderen Land	0,72	0,07	-3,47	0,00
keine Angabe	0,93	0,22	-0,29	0,78
"Ich bin bequem, neige zu Faulheit"				
trifft überhaupt nicht zu	0,76	0,06	-3,47	0,00
2	0,83	0,06	-2,45	0,01

3	Ref.			
4	1,07	0,13	0,53	0,59
trifft voll und ganz zu	0,83	0,19	-0,79	0,43
keine Angabe	0,57	0,29	-1,12	0,26
Fächergruppe				
Geisteswissenschaften	1,07	0,13	0,60	0,55
Sport	1,11	0,39	0,29	0,77
Rechts-, Wirtschafts- und Sozialwiss.	0,84	0,07	-2,03	0,04
Mathematik, Naturwiss.	Ref.			
Humanmed./Gesundheitswiss.	0,73	0,07	-3,10	0,00
Agrar-, Forst-, Ernährungswiss, Veterinärw.	0,76	0,11	-1,94	0,05
Ingenieurwissenschaften	0,82	0,08	-2,05	0,04
Kunst, Kunstwissenschaft	0,89	0,22	-0,50	0,62
keine Angabe	1,22	0,51	0,48	0,63
aktuelles/letztes Bruttoeinkommen				
bis 3200€	0,97	0,07	-0,37	0,71
3201€ bis 5000€	Ref.			
über 5000€	0,91	0,07	-1,26	0,21
kein Einkommen	0,89	0,11	-0,94	0,35
keine Angabe	0,50	0,06	-6,16	0,00
Konstante	1,77	0,21	4,86	0,00
<i>N=5408; Pseudo R²=0,025</i>				

9 Literaturverzeichnis

- Beierlein, Constanze; Kovaleva, Anastassiya; László, Zsuzsa; Kempler, Christoph J.; Rammstedt, Beatrice (2014a): Eine Single-Item-Skala zur Erfassung der Allgemeinen Lebenszufriedenheit: Die Kurzsкала Lebenszufriedenheit-1 (L-1). Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS-Working Papers, 2014|33).
- Beierlein, Constanze; Kovaleva, Anastassiya; Kemper, Christoph J.; Rammstedt, Beatrice (2014b): Allgemeine Selbstwirksamkeit Kurzsкала (ASKU). Zusammenstellung sozialwissenschaftlicher Items und Skalen. Hg. v. GESIS – Leibniz-Institut für Sozialwissenschaften.
- Blumenstiel, Jan Eric; Gummer, Tobias (2015): Prävention, Korrektur oder beides? Drei Wege zur Reduzierung von Nonresponse Bias mit Propensity Scores. In: Jürgen Schupp und Christof Wolf (Hg.): Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen. Wiesbaden: Springer Fachmedien Wiesbaden, S. 13–44.
- Brandt, Gesche; de Vogel, Susanne; Jaksztat, Steffen (2016): Entwicklung und Testung eines Instruments zur Erfassung der Lernumwelt in der Promotionsphase. Ergebnisse der Entwicklungsstudie. Werkstattbericht. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung. Hannover.
- Brandt, Gesche; de Vogel, Susanne; Jaksztat, Steffen (2017a): Ein Instrument zur Erfassung der Lernumwelt Promotionsphase. In: *Zeitschrift für empirische Hochschulforschung* (1), S. 24–44.
- Brandt, Gesche; de Vogel, Susanne; Jaksztat, Steffen; Teichmann, Carola; Lange, K.; Scheller, Percy; Vietgen, Sandra (2017b): DZHW-Promoviertenpanel 2014. Daten- und Methodenbericht zu den Erhebungen der Promoviertenkohorte 2014 (1. und 2. Befragungswelle). Version 1.0.0. FDZ-DZHW. Hannover.
- Centers for Disease Control and Prevention (CDC) (2000): Measuring Healthy Days: Population Assessment of Health-Related Quality of Life. Atlanta, Georgia: CDC. Online verfügbar unter <https://www.cdc.gov/hrqol/pdfs/mhd.pdf>, zuletzt geprüft am 09.12.2019.
- Deming, W. Edwards; Stephan, Frederick F. (1940): On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. In: *Ann. Math. Statist.* 11 (4), S. 427–444. DOI: 10.1214/aoms/1177731829.
- Ebel, Thomas (2015): Empfehlungen zur Anonymisierung quantitativer Daten. GESIS - Leibniz-Institut für Sozialwissenschaften. Mannheim.
- Hochfellner, Daniela; Müller, Dana; Schmucker, Alexandra; Roß, Elisabeth (2012): FDZ-Methodenreport. Datenschutz am Forschungsdatenzentrum. Institut für Arbeitsmarkt- und Berufsforschung (IAB). Nürnberg (06).
- Jaksztat, Steffen (2017): Geschlecht und wissenschaftliche Produktivität. In: *Zeitschrift für Soziologie* 46 (5), S. 347–361.

- Jaksztat, Steffen; Brandt, Gesche; de Vogel, Susanne; Briedis, Kolja (2017): Gekommen, um zu bleiben? Die Promotion als Wegbereiter wissenschaftlicher Karrieren. In: *WSI-Mitteilungen* 70 (5), S. 321–329.
- Kobarg, Sebastian; Wollersheim, Jutta; Welpel, Isabell M.; Spörrle, Matthias (2017): Individual Ambidexterity and Performance in the Public Sector: A Multilevel Analysis. In: *International Public Management Journal* 20 (2), S. 226–260. DOI: 10.1080/10967494.2015.1129379.
- Koberg, Tobias (2016): Disclosing the National Educational Panel Study. In: Hans-Peter Blossfeld, Jutta von Maurice, Michael Bayer und Jan Skopek (Hg.): *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study*. Wiesbaden: Springer VS, S. 691–708.
- Kolenikov, Stanislav (2014): Calibrating survey data using iterative proportional fitting (raking). In: *The Stata Journal* 14 (1), S. 22–59.
- Kovaleva, Anastassiya; Beierlein, Constanze; Kemper, Christoph J.; Rammstedt, Beatrice (2014): Internale-Externale-Kontrollüberzeugung-4 (IE-4). Zusammenstellung sozialwissenschaftlicher Items und Skalen. Hg. v. GESIS - Leibniz-Institut für Sozialwissenschaften.
- Lane, Julia; Heus, Pascal; Mulcahy, Tim (2008): Data access in a cyber world: Making use of cyberinfrastructure. In: *Transactions on Data Privacy* 1 (1), S. 2–16.
- Lomax, Nik; Norman, Paul (2016): Estimating Population Attribute Values in a Table: “Get Me Started in” Iterative Proportional Fitting. In: *The Professional Geographer* 68 (3), S. 451–461. DOI: 10.1080/00330124.2015.1099449.
- Mom, Tom J. M.; van den Bosch, Frans A. J.; Volberda, Henk W. (2009): Understanding Variation in Managers’ Ambidexterity: Investigating Direct and Interaction Effects of Formal Structural and Personal Coordination Mechanisms. In: *Organization Science* 20 (4), S. 812–828. DOI: 10.1287/orsc.1090.0427.
- Morgeson, Frederick P.; Humphrey, Stephen E. (2006): The Work Design Questionnaire (WDQ): developing and validating a comprehensive measure for assessing job design and the nature of work. In: *The Journal of applied psychology* 91 (6), S. 1321–1339. DOI: 10.1037/0021-9010.91.6.1321.
- Nationales Bildungspanel (NEPS) (2013): Startkohorte 6: Erwachsene (SC6). Wellen 2 und 3. Erhebungsinstrumente (Feldversion). Bamberg: LfBi.
- Nationales Bildungspanel (NEPS) (2018): Erhebungsinstrumente (Feldversion). NEPS Startkohorte 5 — Studierende. Hochschulstudium und Übergang in den Beruf, Welle 10. Bamberg: LfBi.
- Potter, Frank J. (1990): A study of procedures to identify and trim extreme sampling weights. In: *Proceedings of the Survey Research Methods Section*, S. 225–230.
- Rammstedt, Beatrice; Kemper, Christoph J.; Klein, Mira Céline; Beierlein, Constanze; Kovaleva, Anastassiya (2014): Big Five Inventory (BFI-10). Zusammenstellung sozialwissenschaftlicher Items und Skalen. Hg. v. GESIS – Leibniz-Institut für Sozialwissenschaften.
- Rosenbaum, Paul R.; Rubin, Donald B. (1983): The central role of the propensity score in observational studies for causal effects. In: *Biometrika* 70 (1), S. 41–55.
- Rubin, Donald B. (1976): Inference and missing data. In: *Biometrika* 63 (2), S. 581–592.

- Scherer, Stefani; Brüderl, Josef (2010): Sequenzdatenanalyse. In: Christof Wolf und Henning Best (Hg.): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 1031–1051.
- Schilke, Oliver (2014): On the contingent value of dynamic capabilities for competitive advantage: The nonlinear moderating effect of environmental dynamism. In: *Strat. Mgmt. J.* 35 (2), S. 179–203. DOI: 10.1002/smj.2099.
- Schnell, Rainer; Hill, Paul B.; Esser, Elke (2005): Methoden der empirischen Sozialforschung. 7. Aufl. München: Oldenbourg.
- Statistisches Bundesamt (2015): Bildung und Kultur. Prüfungen an Hochschulen 2014. Hg. v. Statistisches Bundesamt. Statistisches Bundesamt. Wiesbaden.
- Stegmann, Sebastian; van Dick, Rolf; Ullrich, Johannes; Charalambous, Julie; Menzel, Birgit; E-gold, Nikolai; Wu, Tina Tai-Chi (2010): Der Work Design Questionnaire. In: *Zeitschrift für Arbeits- und Organisationspsychologie A&O* 54 (1), S. 1–28. DOI: 10.1026/0932-4089/a000002.
- Valliant, Richard; Dever, Jill A.; Kreuter, Frauke (2013): Practical tools for designing and weighting survey samples. New York (NY): Springer New York.