

Kolja Briedis / Almuth Lietz / Uwe Ruß / Ulrike Schwabe /
Anne Weber / Robert Birkelbach / Ute Hoffstätter

Nacaps 2018

Daten- und Methodenbericht zur National Academics Panel
Study 2018 (1. Befragungswelle - Promovierende)

Daten- und Methodenbericht

Juni 2020

Autor*innen:

Kolja Briedis
Almuth Lietz
Uwe Ruß
Ulrike Schwabe
Anne Weber
Robert Birkelbach
Ute Hoffstätter

Unter Mitarbeit von:

Christian Friedrich
Leonhard Schulz
Susanne de Vogel

Herausgeber:

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

Geschäftsführung:

Prof. Dr. Monika Jungbauer-Gans
Karen Schlüter

Vorsitzender des Aufsichtsrats:

Ministerialdirigent Peter Greisler

Registergericht:

Amtsgericht Hannover | B 210251

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

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



Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Tabellenverzeichnis	V
I Einleitung	1
II Datennutzungshinweise	2
1 Inhalt und Anlage der Studie.....	4
2 Erhebungsinstrument.....	5
2.1 Inhalte des Erhebungsinstruments	5
2.2 Sekundär genutzte Erhebungsinstrumentbestandteile	7
2.3 Pretests	10
3 Grundgesamtheit und Stichprobenverfahren	12
4 Durchführung der Erhebung	13
5 Rücklauf.....	16
6 Datenaufbereitung	18
6.1 Datenübertragung.....	18
6.2 Datenprüfung und -bereinigung	18
6.3 Codierung offener Angaben.....	18
6.4 Datenstruktur und Datenformat.....	19
6.5 Vergabe von Variablenamen, Variablenlabels und Wertelabels	19
6.6 Codierung fehlender Werte	20
7 Repräsentativität der Daten	22
8 Anonymisierung	25
9 Literaturverzeichnis.....	34
Anhang: Nacaps-Partnerhochschulen	36

Abbildungsverzeichnis

Abbildung 1: Datenzugangswege und Analysepotential.....	3
Abbildung 2: Nacaps-Erhebungsdesign.....	4
Abbildung 3: Relativer Anteil beendeter Fragebögen (in %) über den Verlauf der Feldzeit.....	17
Abbildung 4: Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten von Nacaps 2018.....	26

Tabellenverzeichnis

Tabelle 1:	Thematischer Aufbau der Erstbefragung	5
Tabelle 2:	Quellennachweise zu eingesetzten Erhebungsinstrumenten	7
Tabelle 3:	Geschlecht und Fachbereich (DFG) der interviewten Promovierenden	11
Tabelle 4:	Versandtage der Erhebungsunterlagen	14
Tabelle 5:	Beteiligung der Hochschulen	16
Tabelle 6:	Brutto- und Nettorücklaufquote der ersten Nacaps-Befragungswelle (nach AAPOR 2016)	16
Tabelle 7:	Themengebiete und Kürzel für Variablennamen in Nacaps 2018	20
Tabelle 8:	Systematik des FDZ-DZHW für fehlende Werte	21
Tabelle 9:	Verteilung der Promovierenden nach Destatis 2019 und in Nacaps	22
Tabelle 10:	Verteilung von männlichen und weiblichen Promovierenden nach Destatis 2019 und in Nacaps	23
Tabelle 11:	Maßnahmen der statistischen Anonymisierung nach Zugangsweg	28

I Einleitung

Nacaps steht für *National Academics Panel Study* und ist eine Längsschnittstudie über Promovierende und Promovierte in Deutschland. Nacaps wird vom Deutschen Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) durchgeführt, vom Bundesministerium für Bildung und Forschung (BMBF) gefördert und erhebt erstmalig systematisch Daten zu Karriereverläufen von Hochqualifizierten in Deutschland – für Forschung, Hochschulen und Wissenschaftspolitik.¹

Im Forschungsdatenzentrum für Hochschul- und Wissenschaftsforschung am DZHW (FDZ-DZHW) werden die Daten in Zusammenarbeit mit den Projektmitarbeiter*innen von Nacaps 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 werden auch Dokumentationsmaterialien bereitgestellt.

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation zur ersten Befragungswelle der Nacaps-Promovierendenkohorte 2018 (doi: [10.21249/DZHW:nac2018:1.0.0](https://doi.org/10.21249/DZHW:nac2018:1.0.0)). Weitere Dokumentationsmaterialien zur Studie (Datensatzreport, Fragebogen, etc.) können frei im Metadaten-System des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden. Die zentralen Informationen zur Nutzung der Daten dieser Studie folgen in Abschnitt II. Kapitel 1 stellt Inhalt und Anlage von Nacaps im Allgemeinen und der ersten Befragungswelle von Nacaps 2018 im Speziellen vor. Die weitere Gliederung des Berichts orientiert sich im Wesentlichen am Ablauf des Forschungsprozesses. In Kapitel 2 wird das eingesetzte Erhebungsinstrument und in den Kapiteln 3 bis 6 der Erhebungsprozess beschrieben (Stichprobenziehung, Erhebungsablauf, Rücklauf, Datenaufbereitung). In den Kapiteln 7 und 8 folgen Vergleiche mit den Daten des Statistischen Bundesamtes und die Darstellung der vorgenommenen Anonymisierung.

¹ Aktuelle Informationen zu Nacaps können über die Website des Projekts (<https://www.nacaps.de>) abgerufen werden.

II Datennutzungshinweise

[Voraussetzungen der Datennutzung] Die Daten von Nacaps 2018 werden durch das FDZ des DZHW entsprechend der europäischen Datenschutzgrundverordnung (EU-DSGVO) anonymisiert bereitgestellt und ausschließlich zur wissenschaftlichen Nutzung freigegeben.² Das FDZ bietet ein Scientific Use File (SUF) für die wissenschaftliche Sekundärnutzung und ein Campus Use File (CUF) für Lehr- und Übungszwecke an. Beide können über die Website des FDZ (<https://fdz.dzhw.eu>) beantragt werden. Für die Nutzung von SUF wird ein Datennutzungsvertrag abgeschlossen.

[Datenzugang] Das CUF von Nacaps 2018 wird via Download bereitgestellt und kann am lokalen Computer genutzt werden. Das SUF wird über drei Zugangswege angeboten, die hinsichtlich des Speicherortes, der Möglichkeit der eigenständigen Verknüpfung mit externen Daten und der Kontrollmöglichkeiten des FDZ unterschiedlich restriktiv sind.

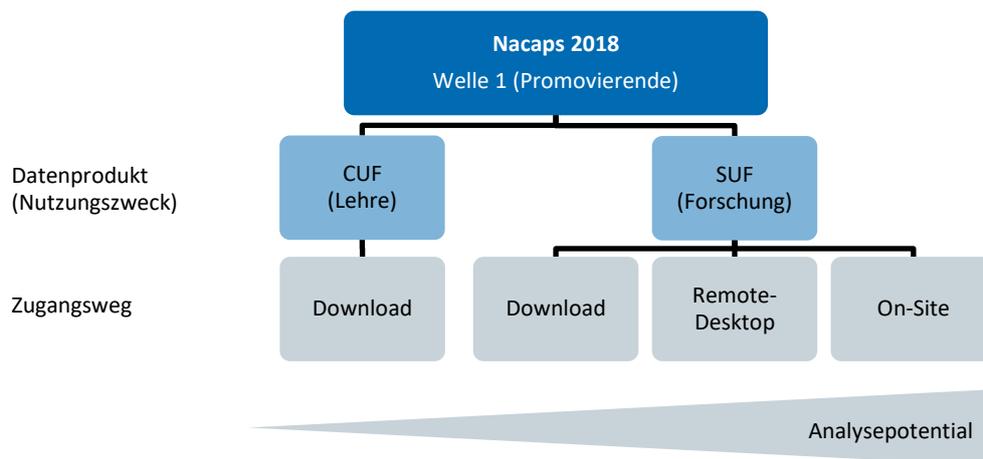
- **Download:** Die Daten werden verschlüsselt auf der Website des FDZ zum Download bereitgestellt. Datennutzer*innen können die Daten auf ihrem lokalen Computer speichern, falls gewünscht selbst mit Daten aus externen Quellen verknüpfen und die Daten mit eigener Software analysieren.
- **Remote-Desktop:**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

Die bereitgestellten Daten weisen je nach Zugangsweg einen unterschiedlich hohen Informationsgehalt auf und unterscheiden sich damit hinsichtlich ihres Analysepotentials (vgl. Abbildung 1). Dabei gilt: Je stärker der Datenzugriff der Nutzer*innen durch technische und organisatorische Maßnahmen kontrolliert wird, desto mehr Informationen können für die Datennutzer*innen bereitgestellt werden.³ Mit diesem Vorgehen wird ein Höchstmaß an Nutzbarkeit und gleichzeitig ein bestmöglicher Schutz der bereitgestellten Daten sichergestellt.

² Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, 6ff.), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (LifBi) (vgl. Koberg, 2016a, 699ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner, Müller, Schmucker & Roß, 2012, 9f.) orientieren. Das FDZ des DZHW hat diesen Ansatz an die Anforderungen der eigenen Datenbestände angepasst und nutzt vier Kategorien von Maßnahmen zur Sicherstellung des Datenschutzes, die in unterschiedlicher Weise kombiniert werden: Rechtlich-institutionelle Maßnahmen, informationelle Maßnahmen, technische Maßnahmen und statistische Maßnahmen.

³ Zu den verschiedenen Anonymisierungsgraden und Analysepotentialen des CUF und der verschiedenen SUF-Varianten vgl. Kapitel 8.

Abbildung 1: Datenzugangswege und Analysepotential



[Kosten der Datenbereitstellung] CUF und SUF werden derzeit (Stand: Juni 2020) kostenfrei zur Verfügung gestellt. Änderungen bzw. das aktuelle Kostenmodell können auf der Website des FDZ (<https://fdz.dzhw.eu>) eingesehen werden.

[Pflichten der Datennutzer*innen] Die Datennutzer*innen sind verpflichtet, folgende Regeln⁴ einzuhalten:

- **Wissenschaftliche Nutzung:** Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.
- **De-Anonymisierungsverbot:** Jeder Versuch der Re-Identifikation von Analyseeinheiten (z. B. Personen, Haushalten, Institutionen) ist verboten.
- **Gebot zur Mitteilung von Sicherheitslücken:** Falls Datennutzer*innen Kenntnis von Sicherheitslücken hinsichtlich Datenschutz bzw. Datensicherheit erlangen, müssen diese dem FDZ-DZHW unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** SUF dürfen nur durch die Personen genutzt werden, die einen Datennutzungsvertrag abgeschlossen haben. CUF dürfen ausschließlich im Rahmen von dem FDZ gemeldeten Lehrveranstaltungen weitergegeben werden.
- **Löschungsgebot:** Download-SUF sind nach Ablauf der vereinbarten Nutzungsdauer (in der Regel 1,5 Jahre) von jeglichen Rechnern, Servern und Datenträgern zu löschen. Ebenso müssen alle Sicherungskopien, modifizierten Datensätze (z. B. Arbeits-, Auszugs- oder Hilfsdateien) sowie Ausdrücke vernichtet werden.
- **Bereitstellung/Meldung von Publikationen:** Jede Art von wissenschaftlichen Arbeiten (z. B. Publikationen, Abschlussarbeiten, Dissertationen), die aus der Arbeit mit Daten des FDZ hervorgehen, ist dem FDZ anzuzeigen. Dabei ist dem FDZ eine elektronische Version der Druckfassung zur Verfügung zu stellen.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen laut der Vorgaben des FDZ zitiert werden (vgl. Zitationsanleitung unter [10.21249/DZHW:nac2018:1.0.0](https://fdz.dzhw.eu/10.21249/DZHW:nac2018:1.0.0)).

⁴ Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

1 Inhalt und Anlage der Studie

[Studienreihe] Nacaps 2018 ist Teil der Studienreihe *National Academics Panel Study*, die anhand standardisierter Online-Befragungen von Promovierenden und Promovierten Informationen zu Karriere- und Lebensverläufen von Hochqualifizierten in Deutschland liefert (Wegner/Briedis 2020).

Die Nacaps-Studienreihe weist ein Multi-Kohorten-Panel-Design auf (vgl. Abbildung 2), bei dem im Abstand von zwei Jahren Promovierende kontaktiert und in der Folge jährlich befragt werden. Zur Auffrischung und partiellen Ergänzung der Promovierendenerhebungen sollen ab 2023 zusätzlich Promoviertenjahrgänge befragt und in das bestehende Panel der Promovierenden integriert werden.

Nacaps 2018 bildet als erste Promovierendenkohorte den Auftakt der Studienreihe. Die Grundgesamtheit beinhaltet alle Personen, die zum Stichtag 1. Dezember 2018 an einer deutschen Hochschule zur Promotion registriert waren. Die erste Befragungswelle fand von Februar bis Mai 2019 statt.

Abbildung 2: Nacaps-Erhebungsdesign



2 Erhebungsinstrument

In Nacaps 2018 wurde als Erhebungsinstrument ein standardisierter Online-Fragebogen in deutscher und englischer Sprache eingesetzt. Kapitel 2.1 stellt die zentralen Themen des Instruments vor. Kapitel 2.2 führt die sekundär genutzten Erhebungsinstrumentbestandteile auf. Kapitel 2.3 beschreibt die zur Prüfung und Optimierung der Fragebögen durchgeführten kognitiven und quantitativen Pretests.

2.1 Inhalte des Erhebungsinstruments

Der Fragebogen der ersten Befragungswelle war in sechs thematische Blöcke gegliedert. Es wurden *Angaben zur Promotion (A)*, *das Promotionsthema und die Promotionsbedingungen (B)*, *der Arbeitsalltag in der Promotionsphase (C)*, *die persönliche Lebenssituation und die Karriereziele (D)*, *die Soziodemografie und die vorangegangenen Bildungsetappen (E)* und zuletzt Aspekte wie *Panelbereitschaft, Newsletter, Verlosung und Abfrage der Kontaktdaten (F)* erfasst.

Zu Beginn der Befragung wurden die Befragten zudem nach ihrem derzeitigen Promotionsstatus⁵ gefragt. Je nach Promotionsstatus wurden im weiteren Verlauf der Befragung verschiedene Fragen angezeigt.

Tabelle 1: Thematischer Aufbau der Erstbefragung

Thema	Fragennummern
A Angaben zur Promotion	A1 bis A20
B Promotionsthema und Promotionsbedingungen	B1 bis B47
C Arbeitsalltag in der Promotionsphase	C1 bis C48
D Persönliche Lebenssituation und Karriereziele	D1 bis D41
E Soziodemografie und vorangegangene Bildungsetappen	E1 bis E22
F Panelbereitschaft, Newsletter, Verlosung, Abfrage der Kontaktdaten	F1 bis F4

⁵ Mit den vier Antwortmöglichkeiten: „ich promoviere“, „ich habe das Promotionsverfahren abgeschlossen“, „ich habe mein Promotionsvorhaben unterbrochen“ und „ich habe mein Promotionsvorhaben abgebrochen“.

[Angaben zur Promotion] Der Themenblock *Angaben zur Promotion* umfasste Fragen zur Motivation bei Promotionsbeginn, zum inhaltlichen Beginn und zum Zeitpunkt der Registrierung an der Hochschule, zum Namen und Standort der Hochschule, zu Gründen für die Hochschulwahl, zu Kooperationen mit Hochschulen im Ausland oder Fachhochschulen sowie zum Promotionsfach. Sofern die Promotion bereits abgeschlossen wurde, wurde ferner nach dem Datum des Abschlusses gefragt; sofern die Promotion abgebrochen wurde, wurden das Datum und die Gründe des Abbruchs erhoben.

[Promotionsthema und Promotionsbedingungen] Der Themenblock *Promotionsthema und Promotionsbedingungen* enthielt Fragen zu kooperativen Promotionsformen in Forschungsprojekten, zum Promotionsthema, zum subjektiv wahrgenommenen Promotionsfortschritt, zu strukturierten Promotionsprogrammen, zu Stipendien, zu Kursen und Lehrveranstaltungen (inkl. der Zufriedenheit damit), zu Kooperationen mit externen Organisationen, zur Form der Promotion (Monographie vs. Kumulative Promotion), zu Abbruchintentionen, zum Vorhandensein einer Promotions- bzw. Betreuungsvereinbarung sowie zur Anzahl an Betreuer*innen. Ferner umfasste er Fragen zum/zur Betreuer*in der Promotion (Affiliation, Geschlecht, formaler Betreuungsstatus, Position, Austauschhäufigkeit, Betreuungstil, Zufriedenheit sowie Sympathieeinschätzung), zur guten wissenschaftlichen Praxis, zu der Finanzierungsquelle sowie spezifische Fragen für Promovierende, die an einer Hochschule oder Forschungseinrichtung arbeiten.

[Arbeitsalltag in der Promotionsphase] Der Themenblock *Arbeitsalltag in der Promotionsphase* setzte sich zusammen aus Fragen zur Beschäftigung (wie Befristung, Stellenumfang), zu den Qualifizierungszielen sowie zum thematischen Bezug zum Promotionsprojekt. Weiterhin wurde nach konkreten wissenschaftlichen Tätigkeiten, der täglichen Motivation zur Arbeit an der Promotion, den jüngsten kooperativen Projekten mit anderen Forscher*innen oder Partner*innen außerhalb der Wissenschaft, nach der Anzahl unterrichteter Semesterwochenstunden und nach Auslandsaufenthalten im Rahmen des Promotionsvorhabens gefragt.

[Persönliche Lebenssituation und Karriereziele] Der Themenblock *Persönliche Lebenssituation und Karriereziele* erfasste die allgemeine Lebenszufriedenheit, enthielt Fragen zur Partner- und Elternschaft, zur Zufriedenheit mit der Vereinbarkeit von Arbeits- und Privatleben, zum allgemeinen Gesundheitszustand, zum körperlichen und seelischen Befinden sowie zum Vorhandensein einer Behinderung. Weiterhin wurden als psychologische Konstrukte die Big Five (Fünf-Faktoren-Modell), die allgemeine Risikobereitschaft, die allgemeinen Kontrollüberzeugungen und die allgemeinen Selbstwirksamkeitserwartungen erfasst. Ferner wurden Instrumente entwickelt und eingesetzt, die die Wichtigkeit von bestimmten Stelleneigenschaften, die Attraktivität von verschiedenen Stellen, die Intention zum Verbleib im Wissenschaftsbereich, die Sektorenpräferenz, das Anstreben einer Professur, den Wunsch nach dauerhafter Beschäftigung in der Wissenschaft und die spezifischen Selbstwirksamkeitserwartungen mit Blick auf eine Tätigkeit innerhalb oder außerhalb der Wissenschaft erfassen.

[Soziodemografie und vorangegangene Bildungsetappen] Der Themenblock *Soziodemografie und vorangegangene Bildungsetappen* enthielt zum einen Fragen nach Geschlecht, Geburtsjahr und -ort sowie Staatsangehörigkeit(en) der Befragten, zum anderen nach Geburtsort und Staatsangehörigkeiten, Schulabschluss, Ausbildungsabschluss und beruflicher Stellung und Bildungsaspirationen der Eltern. Außerdem wurden der Ort und die Note der Studienberechtigung und Angaben zum promotionsrelevanten Studium erhoben.

[Panelbereitschaft, Verlosung, Abfrage der Kontaktdaten] Die Befragung schloss mit Fragen zur Panelbereitschaft und Teilnahme an der Verlosung inkl. Abfrage der Kontaktdaten und bot die Möglichkeit zur Angabe von Kommentaren.

2.2 Sekundär genutzte Erhebungsinstrumentbestandteile

[Quellenangaben zu Erhebungsinstrumenten] Teile des Erhebungsinstrumentes wurden sekundär genutzt, die entsprechenden Quellennachweise sind in Tabelle 2 zusammengestellt.

Tabelle 2: Quellennachweise zu eingesetzten Erhebungsinstrumenten

Fragebogenblock	Fragennummer und Konstrukt	Quelle und Zitation
A – Angaben zur Promotion	---	---
B – Promotionsthema und Promotionsbedingungen	B32: Betreuungsstil	<i>Modifiziertes Instrument nach:</i> Mainhard, T., van der Rijst, R., & van Tartwijk, J. (2009). A model for the supervisor–doctoral student relationship. <i>Higher Education</i> 58, 359–373. https://doi.org/10.1007/s10734-009-9199-8 .
C – Arbeitsalltag in der Promotionsphase	C37: Motivation zur Arbeit an der Promotion	<i>Modifiziertes Instrument nach:</i> Tremblay, M. A., Blanchard, C. M., Taylor, S., Pelletier, L. G., & Villeneuve, M. (2009). Work Extrinsic and Intrinsic Motivation Scale: Its value for organizational psychology research. <i>Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Com-</i>

Fragebogenblock	Fragennummer und Konstrukt	Quelle und Zitation
		<p>portement, 41(4), 213–226. doi:10.1037/a0015167.</p> <p><i>und</i></p> <p>Gagné, M., Forest, J., Vansteenkiste, M., Crevier-Braud, L., Van den Broeck, A., Aspeli, A. K., Bellerose, J., Benabou, C., Chemolli, E., Güntert, S. T., Halvari, H., Indiyastuti, D. L., Johnson, P. A., Hauan Molstad, M., Naudin, M., Ndao, A., Hagen Olafsen, A., Roussel, P., Wang, Z., & Westbye, C. (2015). The Multidimensional Work Motivation Scale: Validation evidence in seven languages and nine countries. <i>European Journal of Work and Organizational Psychology</i> 24 (2), 178-196. doi: 10.1080/1359432X.2013.877892.</p>
D – Persönliche Lebenssituation und Karriereziele	D01: Allgemeine Lebenszufriedenheit	<p><i>Originalinstrument:</i></p> <p>Beierlein, C., Kovaleva, A., László, Z., Kemper, C. J., & Rammstedt, B. (2015). Kurzskala zur Erfassung der Allgemeinen Lebenszufriedenheit (L-1). Zusammenstellung sozialwissenschaftlicher Items und Skalen. doi:10.6102/zis229.</p>
	D11: Gründe gegen Realisierung Kinderwunsch	<p><i>Modifiziertes Instrument nach:</i></p> <p>Lind, I., & Samjeske, K. (2010). Schlussbericht zum Projekt "Balancierung von Wissenschaft und Elternschaft" (BAWIE). Bonn.</p>
	D13: Allgemeiner Gesundheitszustand ("Self-rated Health")	<p><i>Originalinstrument:</i></p> <p>GESIS - Leibniz-Institut für Sozialwissenschaften (2015). Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2014. GESIS Datenarchiv, Köln. ZA5240 Fragebogen.</p> <p><i>und</i></p> <p>Robert-Koch-Institut (2015). Gesundheit in Deutschland aktuell: GEDA 2014/2015. Berlin.</p>
	D24: Big Five Inventory-SOEP (BFI-S)	<p><i>Originalinstrument:</i></p> <p>Schupp, J., & Gerlitz, J.-Y. (2014). Big Five Inventory-SOEP (BFI-S). Zusammenstellung sozialwissenschaftlicher Items und Skalen.</p>

Fragebogenblock	Fragennummer und Konstrukt	Quelle und Zitation
		doi:10.6102/zis54.
	D25: Allgemeine Risikobereitschaft	<i>Originalinstrument:</i> Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2014). Eine Single-Item-Skala zur Erfassung von Risikobereitschaft: Die Kurzsкала Risikobereitschaft-1 (R-1). (GESIS-Working Papers, 2014/34). Mannheim: GESIS Leibniz-Institut für Sozialwissenschaften.
	D26: Kontrollüberzeugung	<i>Originalinstrument:</i> Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2012). Eine Kurzsкала zur Messung von Kontrollüberzeugung: Die Skala Internale-Externale-Kontrollüberzeugung-4 (IE-4). Hg. v. GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS-Working Papers, 19).
	D26a: Allgemeine Selbstwirksamkeitserwartung	<i>Originalinstrument:</i> Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2012). Ein Messinstrument zur Erfassung subjektiver Kompetenzerwartungen: Allgemeine Selbstwirksamkeit Kurzsкала (ASKU) Hg. v. GESIS - Leibniz-Institut für Sozialwissenschaften. (GESIS Working Papers, 17).
	D27: Wichtigkeit von Stellenmerkmalen	<i>Modifiziertes Instrument nach:</i> Roach, M., & Sauermann, H. (2010). A taste for science? PhD scientists' academic orientation and selfselection into research careers in industry. <i>Research Policy</i> 39 (3), 422–434.
	D28: Attraktivität von Stellen	<i>Modifiziertes Instrument nach:</i> Krempkow, R., Sembritzki, T., Schürmann, R., & Winde, M. (2016). Personalentwicklung für den wissenschaftlichen Nachwuchs. Bedarf, Angebote und Perspektiven – eine empirische Bestandsaufnahme im Zeitvergleich. Essen.

Fragebogenblock	Fragennummer und Konstrukt	Quelle und Zitation
	D29: Attraktivität von Stellen II	<i>Modifiziertes Instrument nach:</i> Roach, M., & Sauermann, H. (2010). A taste for science? PhD scientists' academic orientation and selfselection into research careers in industry. <i>Research Policy</i> 39 (3), 422–434.
E – Soziodemografie und vorangegangene Bildungs- etappen	---	---
F - Panelbereitschaft, Newsletter, Verlosung, Ab- frage der Kontaktdaten	---	---

2.3 Pretests

[Ziel und Verfahren] Das Erhebungsinstrument wurde vor der Erstbefragung zunächst durch einen kognitiven und anschließend durch einen quantitativen Pretest geprüft. Ziel des kognitiven Pretests war es zu validieren, ob das Erhebungsinstrument für die Zielgruppe der Promovierenden geeignet war, ob sich neu entwickelte Items für unterschiedliche Promovierendengruppen gleichermaßen eigneten und ob diese verständlich waren. Ziel des quantitativen Pretests war es, das Erhebungsinstrument im Feld unter ähnlichen Bedingungen wie in der Haupterhebung zu testen. Einerseits sollten die Validität und Reliabilität neu entwickelter Instrumente getestet werden. Andererseits sollte die Dauer der Befragung, der Aufbau des Fragebogens sowie die generelle Durchführbarkeit der Online-Befragung und die Wirksamkeit der Incentivierung geprüft werden.

[Kognitiver Pretest] Vom 6. bis 15. Dezember 2017 wurden 17 kognitive Interviews mit Promovierenden an den Standorten Hannover und Berlin durch vier Nacaps-Projektmitarbeiter*innen *face-to-face* durchgeführt. Es wurden 90 Minuten pro Interview einkalkuliert, wobei die tatsächliche Interviewlänge zwischen 72 Minuten und 110 Minuten lag. Als Aufwandsentschädigung erhielten die Interviewten 50 Euro. Abhängig von den konkret zu testenden Fragen und dem jeweiligen Erkenntnisinteresse kamen Befragungstechniken, wie *Probing* (z. B. Information Retrieval Probing, Comprehension Probing), *Confidence Rating* und *General Probing* zum Einsatz (Lenzner/Neuert/Otto 2015, Prüfer/Rexroth 2000, Weichbold 2019). Dabei wurden Personen aus unterschiedlichen Fächern und sowohl Frauen als auch Männer interviewt (Tabelle 3). Es wurden verschiedene Fragen und Itembatterien zu den Themen *Weg zum Promotionsthema*, *Promotionstypen*, *Soziale Netzwerke und Unterstützung*, *wissenschaftliche Tätigkeit* und *Betreuerverhältnis* getestet. In zwei verschiedenen Versionen wurde jeweils die Reihenfolge der Themen variiert.

Tabelle 3: Geschlecht und Fachbereich (DFG) der interviewten Promovierenden

Fachbereich	Geschlecht		Gesamt
	weiblich	männlich	
Geistes- und Sozialwissenschaften	4	1	5
Lebenswissenschaften	3	0	3
Naturwissenschaften	3	2	5
Ingenieurwissenschaften	0	4	4
Gesamt	10	7	17

[Quantitativer Pretest] Für den quantitativen Pretest wurden 10.089 Namen und E-Mail-Adressen von Promovierenden und wissenschaftlichen Mitarbeiter*innen ohne Dokortitel auf frei zugänglichen Webseiten von 29 promotionsberechtigten deutschen Hochschulen von Januar bis März 2018 recherchiert. Die Programmierung des Fragebogens und die Durchführung der Befragung wurden unter Verwendung der DZHW-Onlinebefragungssoftware *Zofar* durchgeführt. Die Feldphase begann am 9. Mai und endete am 31. Mai 2018. In diesem Zeitraum wurden zwei Reminder (am 17. und 23. Mai 2018) versandt. Von den 10.089 angeschriebenen Personen haben 1.692 vollständig und 3.109 teilweise an der Befragung teilgenommen. Somit betrug die Rücklaufquote 16,7 bzw. 30,8 Prozent.

3 Grundgesamtheit und Stichprobenverfahren

[Grundgesamtheit] Als wissenschaftlicher Nachwuchs werden in Anlehnung an die Definition in der Berichterstattung des Bundes Hochschulabsolvent*innen verstanden, die nach dem Studium innerhalb oder außerhalb von Hochschulen oder außeruniversitären Forschungseinrichtungen wissenschaftlich tätig sind oder promovieren, sowie Promovierte, die innerhalb und außerhalb von Hochschulen und außeruniversitären Forschungseinrichtungen einer wissenschaftlichen Tätigkeit nachgehen oder nachgehen wollen. Die genaue Definition, wer als Promovierende*r gilt, ist lange Zeit nicht einheitlich geregelt gewesen. Nacaps orientiert sich daher an der Definition des Hochschulstatistikgesetzes⁶ sowie an der 2017 eingeführten Promovierendenstatistik des Statistischen Bundesamts. Demnach bestand die Grundgesamtheit der Promovierenden bei Nacaps 2018 aus denjenigen Personen, die zum Stichtag 1. Dezember 2018 an einer deutschen Hochschule als Doktorand*in registriert waren. Nach Angaben des Statistischen Bundesamts belief sich die Anzahl der so definierten Promovierenden in Deutschland im Jahr 2018 auf 173.339 Personen (Statistisches Bundesamt 2019c).

[Vollerhebung] Im ersten Schritt wurden zunächst alle 155 promotionsberechtigten Hochschulen (Stand Oktober 2017) zu einer Teilnahme aufgefordert, anschließend wurden in einem zweiten Schritt alle (registrierten) Promovierenden der 53 teilnehmenden Hochschulen (vgl. Anhang: Nacaps-Partnerhochschulen⁷) zur Befragung eingeladen. Da weder eine Auswahl bzw. Ziehung von promotionsberechtigten Hochschulen noch Promovierenden erfolgte, handelt es sich bei dem gewählten Verfahren um eine Vollerhebung. Nicht-teilnehmende Hochschulen und Promovierende werden dabei als Non-Response behandelt.

⁶ „Als Promovierende gelten Personen, die von einer zur Promotion berechtigten Einrichtung eine schriftliche Bestätigung über die Annahme als Doktorandin oder Doktorand in dieser Einrichtung erhalten haben.“ (§5 (1) HStatG)

⁷ Es ist möglich, dass Personen zum Zeitpunkt der Kontaktierung noch an einer der Partnerhochschulen registriert, faktisch aber bereits an eine andere Hochschule gewechselt waren und die Angaben in der Befragung sich schon auf diese neue Hochschule beziehen.

4 Durchführung der Erhebung

[Kontaktaufnahme und Adresspflege] Von den insgesamt 53 teilnehmenden Hochschulen erfolgte an 48 teilnehmenden Hochschulen der Einladungs- und Remindersversand durch die Hochschulen selbst (Adressmittlung) und drei Hochschulen übermittelten Kontaktdaten an das DZHW, das in diesen Fällen den Einladungs- und Remindersversand übernahm (Adressübermittlung). Zwei teilnehmende Hochschulen konnten weder Kontaktdaten an das DZHW übermitteln noch direkt eine Einladung zur Befragung selbst verschicken. Deswegen wurden die Promovierenden an diesen beiden Hochschulen gebeten, sich für die Befragung auf einer separaten Website zu registrieren. Anschließend wurden die so registrierten Promovierenden durch das DZHW zur Befragung eingeladen.

An den teilnehmenden Hochschulen war der Einladungs- und Remindersversand größtenteils an zentralen Stellen wie Graduiertenzentren und Evaluationsbüros angesiedelt, die den Promovierenden der Hochschule in vielen Fällen bereits bekannt waren. Um einen möglichst einheitlichen Einladungs- und Remindersversand zu gewährleisten, wurden denjenigen Hochschulen, die ihre Promovierenden selbst zur Befragung einluden, verschiedene Materialien durch das DZHW zur Verfügung gestellt. Dabei handelte es sich um eine Handreichung und ein Video-Tutorial zum Einladungs- und Remindersversand sowie Serienbriefvorlagen für die Vorankündigung, das Einladungs- und die Reminderanschreiben. Zusätzlich wurden die teilnehmenden Hochschulen zu einem Workshop zur Befragungsdurchführung an das DZHW in Hannover eingeladen.

[Erhebungsunterlagen] Die Erhebungsunterlagen umfassten ein Vorankündigungsschreiben, ein Einladungsschreiben und drei Reminder. Mit der Einladungs-E-Mail sowie mit allen nachfolgenden Remindern erhielten die Befragten einen personalisierten Link zum Online-Fragebogen. Die Befragung wurde mit der DZHW-Onlinebefragungssoftware Zofar programmiert und durchgeführt. Reminder wurden nur an diejenigen Personen verschickt, die noch nicht vollständig an der Befragung teilgenommen hatten. Informationen zum Datenschutz wurden zu Beginn des Online-Fragebogens bei der Abfrage der Zustimmung zu den Datenschutzbestimmungen bereitgestellt. Alle Anschreiben wurden den beteiligten Hochschulen durch das DZHW zur Verfügung gestellt und durch diese nur unwesentlich verändert, indem die jeweils übliche Anrede und die jeweilige Grußformel der versendenden Stelle der Hochschule eingefügt wurden.

[Feldphase] Die gesamte Feldphase lief vom 18. Februar bis 1. Mai 2019. Vorankündigung, Einladungen und Reminder zur Befragungsteilnahme wurden an folgenden Tagen verschickt (vgl. Tabelle 4).

Tabelle 4: Versandtage der Erhebungsunterlagen

Versandtage	Erhebungsunterlage
11.02.2019	Vorankündigung
18.02.2019	Einladung / Beginn der Feldphase
05.03.2019	1. Reminder
20.03.2019	2. Reminder
04.04.2019	3. Reminder

An einigen Hochschulen wurde die Einladung wenige Tage später als geplant verschickt. An fünf Hochschulen konnte die Einladungs-E-Mail aus organisatorischen Gründen erst am 5. März 2019 verschickt werden; in einem weiteren Fall wurde die Einladungs-E-Mail erst am 22. März 2019 versendet. Auf Grund dieser verspäteten Einladungen wurde die Feldphase erst am 1. Mai 2019 beendet. Dadurch konnte auch für die Promovierenden dieser Hochschulen eine ausreichend lange Feldphase gewährleistet und ebenfalls bis zu drei Reminder verschickt werden. Einige wenige Hochschulen hatten sich im Vorfeld gegen das Versenden von drei Remindern ausgesprochen und stattdessen nur einen oder zwei Reminder versendet.

[Rücklaufsteigernde Maßnahmen] Zur Steigerung des Rücklaufs wurden mehrere Maßnahmen ergriffen. Zum einen wurde mit der Befragungseinladung und den Remindern ein nicht-monetäres promised Incentive in Form einer Verlosung angeboten. Unter allen Teilnehmer*innen wurden dabei folgende Sachpreise verlost:

- 1 x Apple iPad Pro 11“ mit 256 GB im Wert von 1.049 Euro
- 2 x Microsoft Surface Go mit 64 GB im Wert von je 429 Euro
- 10 x Wunschgutscheine im Wert von je 50 Euro
- 25 x Netflix-Gutscheine im Wert von je 25 Euro

In einem Fall wurde durch eine teilnehmende Hochschule eine zusätzliche eigene Verlosung angeboten.

Zum anderen wurden die teilnehmenden Hochschulen gebeten, bereits vor der Feldphase ihre Promovierenden auf die bevorstehende Nacaps-Befragung aufmerksam zu machen. Zu diesem Zweck wurden den Hochschulen Plakate, Flyer, Beispieltexthe für Newseinträge auf Webseiten und Bildmaterialien zur Verfügung gestellt. Zusätzlich wurden die Hochschulen gebeten, eine Woche vor Beginn der Feldphase die Vorankündigungs-E-Mail an ihre Promovierenden zu verschicken. Einerseits sollten damit die Promovierenden auf die anstehende Befragung und das Projekt aufmerksam gemacht werden. Andererseits sollte damit der E-Mail-Versand durch die Hochschulen getestet werden. Neben der direkten Ansprache der Promovierenden wurden schließlich verschiedene begleitende Maßnahmen der Öffentlichkeitsarbeit durchgeführt, die das Interesse der Promovierenden am Projekt steigern sollten. So wurden bereits einige Monate vor Beginn der Feldphase eine Projekt-Homepage (<https://www.nacaps.de/>) und ein Twitter-Account (@Nacaps_Panel) angelegt, über die insbesondere vor und während der Feldphase zielgruppenrelevante Beiträge geteilt wurden.

Des Weiteren wurden die Einladung und die drei Reminder jeweils an verschiedenen Wochentagen verschickt. So wurde die Einladung am Montag verschickt, der erste Reminder zwei Wochen später folgte dann an einem Dienstag. Der zweite Reminder weitere zwei Wochen

später wurde wiederum an einem Mittwoch und der dritte Reminder ebenfalls zwei Wochen später am Donnerstag versendet.

Da die Durchführung des Projekts einerseits und die Erreichbarkeit der Promovierenden andererseits in starkem Maße von der Kooperationsbereitschaft promotionsberechtigter Hochschulen abhängen, wurden auch die promotionsberechtigten Hochschulen mit einer Reihe von Maßnahmen incentiviert. So erhielten teilnehmende Hochschulen Zugang zu einem Datenportal, das hochschulspezifische Auswertungen der eigenen Befragungsdaten ermöglicht. Des Weiteren wurden die teilnehmenden Hochschulen mit einem Willkommenspaket (Nacaps-Kaffebecher, Kühlschrankmagnet, Post-It, Flyer, Poster usw.) begrüßt.

5 Rücklauf

Die erste Befragungswelle von Nacaps 2018 erfolgte an 53 deutschen Hochschulen, wobei alle 155 promotionsberechtigten Hochschulen in Deutschland angefragt worden waren, an der Befragung teilzunehmen. Auf Hochschulebene fand eine Vollerhebung statt, d. h. alle zum Stichtag 1. Dezember 2018 registrierten Promovierenden wurden von den kooperierenden Hochschulen eingeladen.

Tabelle 5: Beteiligung der Hochschulen

Beteiligte	Anzahl
Hochschulen mit Promotionsrecht	155
Teilnehmende Hochschulen	53

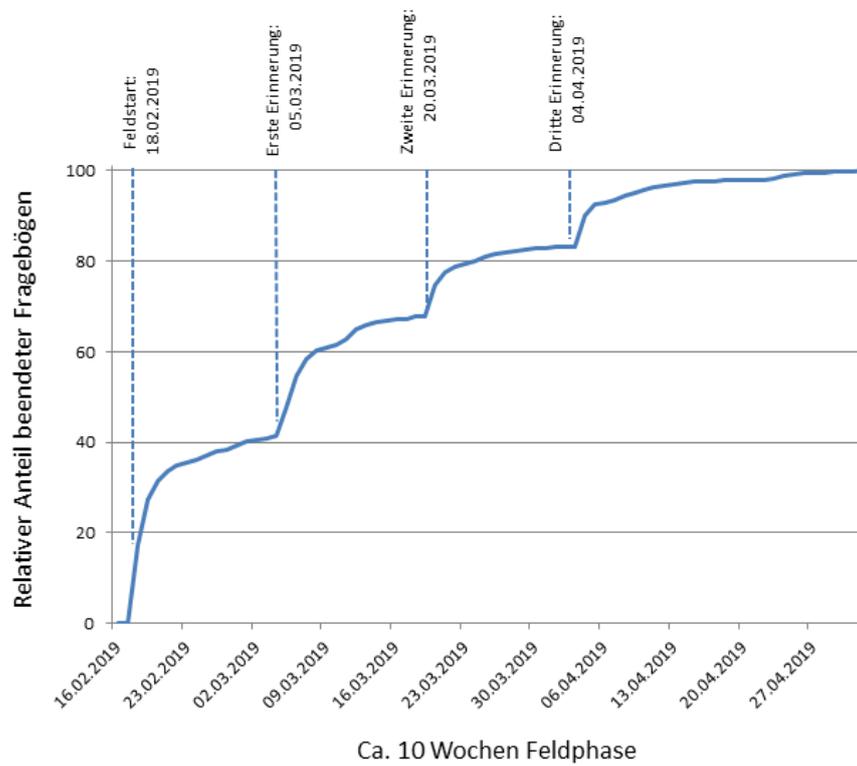
Insgesamt wurden 77.509 Einladungen zur Befragung erfolgreich verschickt. Mit 28.368 Fragebögen, bei denen die Befragten den Datenschutzbestimmungen zugestimmt haben, ergibt sich somit eine (bereinigte) Rücklaufquote von 36,6 Prozent (Tabelle 6).

Tabelle 6: Brutto- und Nettorücklaufquote der ersten Nacaps-Befragungswelle (nach AAPOR 2016)

Welle 1	
Versendete Einladungen [1]	80.086
Unzustellbar	2.577
Bereinigte versendete Einladungen [2]	77.509
Fragebogenlink geöffnet [3]	29.842
Datenschutz zugestimmt [4]	28.368
Fragebogen inhaltlich abgeschlossen [5]	23.612
Rücklaufquote 1 ([3]/[1])	37,3 %
Rücklaufquote 2 ([3]/[2])	38,5 %
Rücklaufquote 3 ([4]/[1])	35,4 %
Rücklaufquote 4 ([4]/[2])	36,6 %
Rücklaufquote 5 ([5]/[1])	29,5 %
Rücklaufquote 6 ([5]/[2])	30,5 %

In Abbildung 3 ist die Entwicklung des Rücklaufs über den gesamten Verlauf der Feldzeit als relativer Anteil beendeter Fragebögen dargestellt. Zusätzlich sind die Zeitpunkte des Versands der drei Erinnerungsschreiben gekennzeichnet, welche insgesamt eine Rücklauf-steigernde Wirkung zeigen.

Abbildung 3: Relativer Anteil beendeter Fragebögen (in %) über den Verlauf der Feldzeit



6 Datenaufbereitung

Im Folgenden werden die verschiedenen Arbeitsschritte der Datenaufbereitung beschrieben, die im Rahmen des Projekts mit der Statistiksoftware Stata durchgeführt wurden. Die nachträgliche Anonymisierung wurde durch das FDZ-DZHW vorgenommen (Kapitel 8).

6.1 Datenübertragung

Die Befragung wurde mit dem DZHW-eigenen Onlinebefragungssystem *Zofar* durchgeführt. Die Angaben aus der Online-Befragung wurden von *Zofar* als .csv-Datei an das Nacaps-Team geliefert und daraufhin als Stata-Datei (.dta) weiterverarbeitet.

6.2 Datenprüfung und -bereinigung

[Definierte Wertebereiche] Um von vorneherein das Ausmaß an unplausiblen Werten einzuschränken, wurden innerhalb der Fragebogenprogrammierung Wertebereiche definiert, die nicht unter- bzw. überschritten werden konnten. Im Falle einer Unter- bzw. Überschreitung der Wertebereiche wurden die Befragten in roter Schrift darauf hingewiesen, dass ihre Eingabe unplausibel erscheint und sie wurden darum gebeten, den Wert zu korrigieren. Erfolgte keine Korrektur nach dem Hinweis, konnte die Befragung dennoch fortgesetzt werden.

[Nachträgliche Prüfung] Nach der Befragung wurde für jede Variable überprüft, ob die erfasste Ausprägung auch tatsächlich in dem für diese Variable definierten Wertebereich lag (Prüfung von Wertebereichen). Haben die Befragten beispielsweise mehr als die in der Ausfüllanweisung geforderten Dezimalstellen angegeben, wurden die Werte nachträglich gerundet. Weiterhin wurden einerseits Vollständigkeitsprüfungen hinsichtlich der Filterführung vorgenommen, andererseits auf Filterverstöße getestet (Prüfung der Einhaltung der Filterführung). Da es in der Befragung möglich war, durch Vor- und Zurückspringen Angaben nachträglich zu ändern, ist in den Daten nicht zu erkennen, welche Angabe tatsächlich korrekt ist. Daher wurden keine Änderungen in den Angaben vorgenommen, jedoch Filterverstöße mit so genannten „Flag-Variablen“ markiert, die den Wert 1 haben, wenn ein Filterverstoß vorliegt, und den Wert 0, wenn dies nicht der Fall ist. Die Entscheidung, welche Angabe für die Analysen verwendet werden soll, kann von den Datennutzer*innen selbst getroffen werden.

6.3 Codierung offener Angaben

Im Rahmen der Datenaufbereitung erfolgte eine Codierung zentraler (halb-)offener Angaben. Je nach Variable wurden verschiedene Codierlisten herangezogen, mit deren Hilfe den (halb-)offenen Angaben numerische Codes zugeordnet wurden. Unklare Fälle, die nicht eindeutig einem numerischen Wert der entsprechenden Codierliste zugeordnet werden konnten, wurden zunächst als Missing codiert und erneut separiert voneinander betrachtet und recherchiert, um sie in die Systematik einzugliedern. Führt dies zu keiner eindeutigen Einteilung oder handelte es sich um offensichtliche Falschangaben, verblieben diese weiterhin als Missing. Codierte Variablen sind im Datensatz durch das Suffix „_g#“ im Variablennamen als gene-

rierte Variable gekennzeichnet (vgl. Kapitel 6.5). Genauere Informationen zur Codierung der offenen Angaben sowie zu den verwendeten Codierlisten sind im Datensatzreport zu finden.

6.4 Datenstruktur und Datenformat

Der Datensatz enthält die Befragungsdaten sowie die zusätzlich generierten Variablen. Pro befragter Person existiert eine Datenzeile (wide-Format). Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehörigen Fragen im Fragebogen. Der Datensatz wird sowohl im Stata- als auch im SPSS-Format bereitgestellt.

6.5 Vergabe von Variablenamen, Variablenlabels und Wertelabels

[Variablen- und Wertelabelvergabe] Für Variablen- und Wertelabels wurden Kurzformen der Fragebogenformulierungen gewählt. In den Variablenlabels ist außerdem ein Kürzel für den Zugangsweg enthalten, ab dem die jeweilige Variable nutzbar⁸ ist (vgl. Kapitel II und 8):

- **c:** Variable ist im CUF und in allen SUF-Varianten (Download-SUF, Remote-Desktop-SUF, On-Site-SUF) nutzbar.
- **d:** Variable ist nicht im CUF, aber im Download-SUF, im Remote-Desktop-SUF und im On-Site-SUF nutzbar.
- **r:** Variable ist nicht im CUF und im Download-SUF, aber im Remote-Desktop-SUF und im On-Site-SUF nutzbar.
- **o:** Variable ist nicht im CUF, im Download-SUF und im Remote-Desktop-SUF, aber im On-Site-SUF nutzbar.
- **a:** Variable ist über keinen Zugangsweg nutzbar. Sie wird aber dokumentiert, da es zugehörige Fragen im Fragebogen gibt.

Die Variablen- und Wertelabel liegen auf Deutsch und auf Englisch vor. Im Stata-Format wurden zweisprachige Labels im gleichen Datensatz hinterlegt.⁹ Im SPSS-Format existiert für jede Sprache ein eigener Datensatz.

[Variablenbenennung] Für die Variablenbenennung wurde ein Präfix-Stamm-Suffix-Schema verwendet. Das Präfix enthält die Wellenkennung anhand des Buchstabens a.

Aus dem Stamm geht erstens der Themenbereich, dem die Variable zugeordnet ist, hervor. Tabelle 7 stellt die verschiedenen Themenbereiche sowie das zugehörige Kürzel für den Stamm des Variablennamens im Überblick dar.

⁸ „Nutzbar“ heißt: die Variable enthält nicht das Missing „anonymisiert“.

⁹ Mit dem Befehl „label language en“ erfolgt ein Wechsel zu den englischen Labels. Über „label language de“ kann zu den deutschen Labels zurückgewechselt werden.

Tabelle 7: Themengebiete und Kürzel für Variablenamen in Nacaps 2018

Themengebiete- kürzel	Themengebiet (englisch)	Themengebiet (deutsch)
ict	informed consent	Informierte Einwilligung
dbi	basic information on doctoral phase	Eckdaten zur Promotion
did	interruption/drop-out	Promotionsunterbrechung und -abbruch
dtc	thesis topic and motives for PhD	Promotionsthema und -motive
dcd	conditions during doctoral course	Promotionsbedingungen
dsv	supervisor	Betreuer(in)
fin	finances	Finanzierung
emp	employment	Beschäftigungssituation
dwr	daily scientific working routine	Wissenschaftlicher Arbeitsalltag
abr	international mobility	Auslandsaufenthalte/Mobilität
psy	personality/psychic traits	Persönlichkeit/ psychische Eigenschaften
lcd	living conditions	Persönliche Lebenssituation
hea	health	Gesundheit
crg	career goals	Karriereziele
dem	demography	Demographie
par	social background	Soziale Herkunft
edb	educational biography	Bildungsbiographie
inf	information	Information
cmt	comment	Kommentar

Zweitens sind die Variablen innerhalb der definierten Themenbereiche nummeriert. Mit Hilfe eines Buchstabens am Ende des Stamms sind drittens Variablen differenziert, die zur gleichen Frage gehören und dadurch die gleiche Themendifferenzierung und Nummerierung aufweisen (z. B. bei Itembatterien, Mehrfachnennungen oder Fragen, in denen geschlossene und offene Fragen kombiniert werden).

Darüber hinaus sind generierte Variablen¹⁰ und Flagvariablen mit einem anhand eines Unterstrichs vom Stamm abgetrennten Suffix gekennzeichnet (*_g#* bzw. *_flag*).

6.6 Codierung fehlender Werte

Zur Codierung fehlender Werte wurde im FDZ-DZHW eine übergreifende Systematik erstellt, um über verschiedene Datensätze des DZHW hinweg eine einheitliche Missingcodierung gewährleisten zu können. Fehlende Angaben werden dabei durch dreistellige negative Werte codiert. Die Missingcodierung der Nacaps-Daten basiert auf dieser Systematik. Die verwendeten Missingcodierungen sind in Tabelle 8 fett hervorgehoben.

¹⁰ Unter den Typus der generierten Variablen fallen alle Variablen, die aus einer oder mehreren Variablen des Ursprungsdatensatzes erzeugt wurden (z.B. vercodete Variablen, Indizes, Aggregationen).

Tabelle 8: 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 wird nur für Datensätze im Long-Format vergeben.

^b Dieser Wert wird nur in gepoolten Datensätzen vergeben.

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

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

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

7 Repräsentativität der Daten

Die Anpassung der Daten der realisierten Stichprobe an ausgewählte Merkmale der Grundgesamtheit mittels Gewichtung ist ein übliches Vorgehen, um systematischen Selektionsprozessen entgegenzuwirken und damit die Repräsentativität der Daten zu optimieren. Aufgrund der aktuellen Datenlage ist eine Anpassungsgewichtung an die Grundgesamtheit jedoch noch nicht möglich, da es derzeit keine unverzerrten offiziellen amtlichen Daten zu den Promovierenden in Deutschland gibt. Zwar wurden zum 1. Dezember 2017 erstmalig die registrierten Promovierenden vom Statistischen Bundesamt erfasst, allerdings waren die Promovierenden in dieser Statistik untererfasst, da die erforderlichen organisatorischen und technischen Strukturen, die eine unverzerrte Erfassung ermöglicht hätten, zu diesem Zeitpunkt noch nicht vorhanden waren (vgl. Vollmar 2019). Auch die zweite Erhebung zum Stichtag 1. Dezember 2018 weist weiterhin eine Untererfassung auf (Statistisches Bundesamt 2019c). Um dennoch einen Überblick über eine mögliche ungleiche Verteilung zwischen Stichprobe und der Grundgesamtheit zu erhalten, wurde die Verteilung der Personen nach Fächern zwischen der realisierten Stichprobe mit den Daten des Statistischen Bundesamtes verglichen (Tabelle 9).

Tabelle 9: Verteilung der Promovierenden nach Destatis 2019 und in Nacaps

Fächergruppen	Destatis 2019		Nacaps	
	Absolut	in %	in %	Diff.
Geisteswissenschaften	21.688	12,5	12,4	-0,3
Sport	1.099	0,6	0,7	0,1
Rechts-, Wirtschafts- und Sozialwissenschaften	30.049	17,3	20,3	2,8
Mathematik, Naturwissenschaften	41.810	24,1	29,0	5,0
Humanmedizin/Gesundheitswissenschaften	36.925	21,2	15,0	-6,0
Agrar-, Forst- und Ernährungswissenschaften, Veterinärmedizin	5.158	3,0	3,0	0,0
Ingenieurwissenschaften	30.944	17,8	17,7	0,0
Kunst, Kunstwissenschaft	3.787	2,2	2,0	-0,2
Außerhalb der Studienbereichsgliederung	2.319	1,2	0	-1,3
	173.779	100,0	100,0	

Demnach sind Personen, die in den Rechts-, Wirtschafts- und Sozialwissenschaften sowie in der Fächergruppe Mathematik, Naturwissenschaften promovieren, in der Erhebung leicht überrepräsentiert. Dagegen sind Promovierende aus der Fächergruppe Humanmedizin/Gesundheitswissenschaften etwas unterrepräsentiert. In den anderen Fächergruppen sind die Verteilungen in beiden Datenquellen nahezu gleich.

Ein Vergleich der Verteilungen nach den Merkmalen Geschlecht und Fächergruppe lässt erkennen, dass in Nacaps Frauen aus der Fächergruppe Rechts-, Wirtschafts- und Sozialwissenschaften leicht überrepräsentiert sind. In der Fächergruppe Mathematik/Naturwissenschaften sind dagegen sowohl Männer als auch Frauen etwas überrepräsentiert. In der Fächergruppe Humanmedizin/Gesundheitswissenschaften sind so-

wohl Männer als auch Frauen unterrepräsentiert; der Unterschied ist bei Frauen jedoch deutlich größer als bei Männern. In den übrigen Fächergruppen gibt es keine nennenswerten Unterschiede. Ebenso sind die Unterschiede in der Verteilung nach dem Geschlecht insgesamt zwischen der Nacaps-Stichprobe (51% Männer, 48% Frauen, 1% divers) und den Informationen aus der amtlichen Statistik (53% Männer und 47% Frauen) ebenfalls eher gering.

Tabelle 10: Verteilung von männlichen und weiblichen Promovierenden nach Destatis 2019 und in Nacaps

Fächergruppen	Destatis 2019				Nacaps	
	absolut, männlich	in % männlich	absolut, weiblich	in % weiblich	in % männlich	in % weiblich
Geisteswissenschaften	9.541	10,3	12.147	15,0	9,1	15,3
Sport	605	0,7	494	0,6	0,6	0,8
Rechts-, Wirtschafts- und Sozialwissenschaften	15.600	16,8	14.449	17,9	17,8	21,8
Mathematik, Naturwissenschaften	24.230	26,3	17.380	21,5	30,6	25,9
Humanmedizin/Gesundheitswissenschaften	13.876	14,9	23.049	28,5	11,0	18,0
Agrar-, Forst- und Ernährungswissenschaften, Veterinärmedizin	1.890	2,0	3.268	4,0	1,7	4,5
Ingenieurwissenschaften	24.418	26,3	6.526	8,1	25,8	7,8
Kunst, Kunstwissenschaft	1.297	1,4	2.490	3,1	1,0	3,1
Außerhalb der Studienbereichsgliederung	1.221	1,3	1.098	1,4	0	0
	92.878	100,0	80.902	100,0	100,0	100,0

Darüber hinaus können die Verteilungen zudem mit weiteren Datenquellen abgeglichen werden, die jedoch ihrerseits jeweils Unsicherheiten bezüglich ihrer Repräsentativität aufweisen. Dazu zählt u. a. die Statistik zu abgeschlossenen Promotionen des Jahres 2018 (Statistisches Bundesamt 2019d). Ein Problem bezüglich des Abgleichs mit dieser Datenquelle ist jedoch, dass der Abbruch von Promotionen sich möglicherweise zwischen Fächern, Geschlechtern oder anderen Merkmalen systematisch unterscheidet. Insofern kann ein solcher Abgleich zwar vorgenommen werden, ist jedoch vermutlich mit größeren Unsicherheiten behaftet als ein Vergleich mit der oben vorgenommenen Statistik.¹¹ Ebenso ist ein Abgleich mit einer älteren Erhebung des Statistischen Bundesamtes zu Promovierenden in Deutschland möglich. Dabei handelt es sich um eine sog. §7-Erhebung, die außerhalb der regelmäßigen statistischen Erhebungen erfolgt (Statistisches Bundesamt 2016). Nach §7 des Bundesstatistikgesetzes können Sondererhebungen (max. zweimalig) durchgeführt werden, um Daten für besondere Zwecke zu generieren. Da eine Promovierendenstatistik bis zur Novelle des Hochschulstatistikgesetzes im Jahr 2016 nicht existierte und die Studierendenstatistik eine systematische Untererfassung der Promovierenden aufweist, wurde die §7-Erhebung erstmalig für das Wintersemester 2010/11 und dann erneut für das Wintersemester 2014/15 durchgeführt. Damit ist auch die

¹¹ Nimmt man diesen Vergleich dennoch vor, sind Promovierende der Geisteswissenschaften (+5%) und der Rechts-, Wirtschafts- und Sozialwissenschaften (+5%) in Nacaps jeweils überrepräsentiert, Promovierende der Humanmedizin/Gesundheitswissenschaften dagegen unterrepräsentiert (-10%).

Erhebung für das Wintersemester 2014/15 einerseits etwas älteren Datums und andererseits basieren die Informationen auf Befragungen von Professor*innen und Promovierenden. Insofern ist ein Vergleich zwischen den Informationen über die Grundgesamtheit aus dieser Erhebung und Nacaps ebenso mit Unsicherheiten behaftet.¹² Eine weitere Datenquelle für einen Abgleich der Verteilungen ist die amtliche Statistik über immatrikulierte Studierende, in der ebenfalls eingeschriebene Promovierende erfasst sind (Statistisches Bundesamt 2019e). In dieser Datenquelle gibt es jedoch eine deutliche Untererfassung der Promovierenden gegenüber den anderen beiden Datenquellen über Promovierende, denn es sind hier nur rund 110.000 Promovierende verzeichnet. Somit ist auch hier zu erwarten, dass ein Abgleich zur Verteilung in Nacaps mit Unsicherheiten behaftet ist.¹³

Da derzeit keine vollkommen zuverlässigen Datenquellen zum Abgleich von Fächer- und Geschlechterverteilungen für Nacaps vorliegen, kann keine sichere Bewertung der Repräsentativität der Daten vorgenommen werden. Aus dem gleichen Grund können keine GewichtungsvARIABLEN erstellt werden. Die dargestellten statistischen Abgleiche sollen die Datennutzenden jedoch ermächtigen, eine erste Bewertung der Repräsentativität der Daten vornehmen zu können.

¹² Bei einem Vergleich dieser Erhebungen gibt es wiederum Unterschiede im Bereich der Geisteswissenschaften. Dieses Mal sind Promovierende aus diesen Fächern in Nacaps gegenüber der herangezogenen Datenquelle jedoch unterrepräsentiert (-6%); in den Rechts-, Wirtschafts- und Sozialwissenschaften (+3%) und der Humanmedizin/den Gesundheitswissenschaften (+5%) sind Promovierende in Nacaps dagegen überrepräsentiert.

¹³ Bei einem Abgleich sind allerdings nur geringe Abweichungen zu der Fächerverteilung in Nacaps zu beobachten. Lediglich im Bereich Humanmedizin/Gesundheitswissenschaften sind Promovierende in Nacaps überrepräsentiert (+4%).

8 Anonymisierung

[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten¹⁴, 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 2017.¹⁵ Danach sind personenbezogene Daten für die Weitergabe zur wissenschaftlichen Sekundärnutzung (ohne Vorliegen einer Einverständniserklärung zur Sekundärnutzung der personenbezogenen Daten) in der Regel derart aufzubereiten, dass „die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden können“ (Art. 4 Abs. 5 DSGVO; s. auch Art. 89 DSGVO sowie Erwägungsgrund 26 DSGVO). Das heißt, für die Weitergabe von Daten aus wissenschaftlichen Forschungsprojekten an Dritte sind die Daten derart zu anonymisieren, dass kein Bezug zur Person mehr hergestellt werden kann.

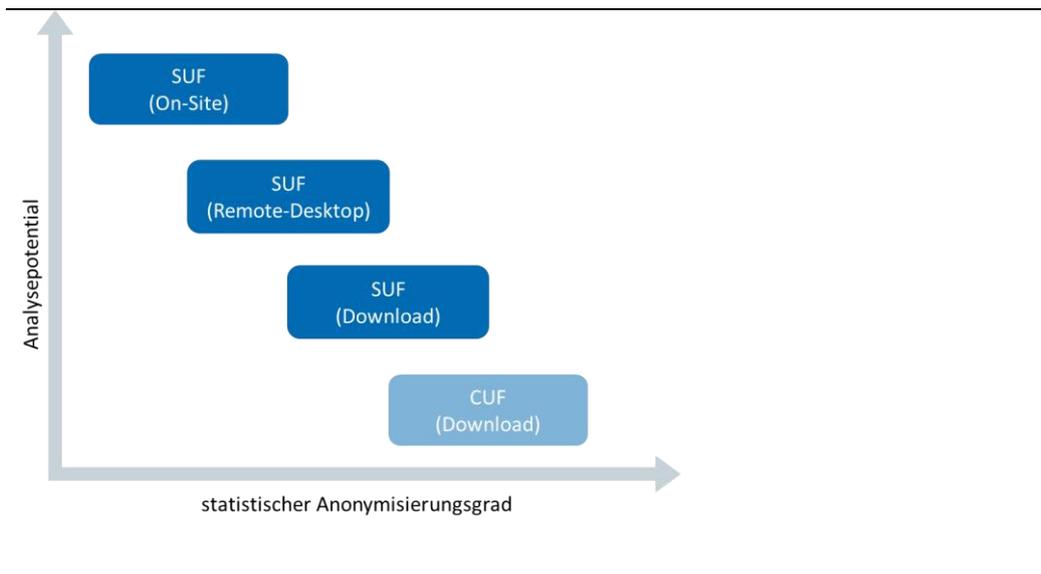
[Datenzugang, Anonymisierungsgrad und Analysepotential] Das FDZ des DZHW stellt für Nacaps 2018 ein SUF für die wissenschaftliche Sekundärnutzung und ein CUF für Lehr- und Übungszwecke zur Verfügung. Die Anonymität der Befragten wird dabei über eine Kombination aus statistischen Maßnahmen und technischen Zugriffsbeschränkungen sichergestellt. Je stärker der Datenzugang technisch kontrolliert wird, desto geringer ist das Risiko einer De-Anonymisierung der Daten, desto weniger müssen die Daten mittels statistischer Maßnahmen um Informationen reduziert werden und desto größer bleibt ihr Analysepotential.

Während das CUF per Download bereitgestellt 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 4 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.

¹⁴ „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).

¹⁵ Die DSGVO gilt grundsätzlich innerhalb der EU und somit ebenfalls für das DZHW. Das BDSG in seiner Neufassung vom 30. Juni 2017 (Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680 (Datenschutz-Anpassungs- und Umsetzungsgesetz EU DSAnpUG-EU)) kommt teils zusätzlich zur Anwendung, da die DZHW GmbH juristisch als öffentliche Stelle des Bundes betrachtet wird (§ 2 Abs. 3 BDSG). Der Bund hält die absolute Mehrheit der Anteile der DZHW GmbH und das Institut erfüllt Aufgaben der öffentlichen Verwaltung des Bundes im weitesten Sinn.

Abbildung 4: Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten von Nacaps 2018



[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 bei Nacaps 2018 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. Um eine eindeutige Zuordnung der Daten zu unterbinden, wurden diese Schlüsselmerkmale – je nach Datenprodukt bzw. Zugangsweg – aggregiert oder gar nicht freigegeben. Beispielsweise sind für Nacaps 2018 besonders zentrale Quasi-Identifikatoren die Promotionshochschule und das Promotionsfach. Die Promotionshochschule wird über keinen Zugangsweg freigegeben und auch nicht in vercodeter bzw. aggregierter Form bereitgestellt. Das Promotionsfach wird im On-Site- und Remote-Desktop-SUF aggregiert zu DZHW-Fächergruppen freigegeben. Im Download-SUF und im Download-CUF erfolgt eine noch etwas stärkere Aggregation zu Destatis-Fächergruppen.

Darüber hinaus empfehlen Ebel und Meyermann, offene Angaben zu löschen „selbst wenn die jeweiligen Fragestellungen an sich unproblematisch sind. Denn es besteht die Gefahr, dass Studienteilnehmer/-innen bei eigentlich unbedenklichen Fragen mit offener Antwortmöglichkeit kritische Informationen preisgegeben haben, die zu einer Identifikation führen könnten“ (Ebel & Meyermann, 2015, S. 5). Entsprechend werden bei Nacaps 2018 keine offenen Angaben freigegeben. Teilweise wurden die offenen Angaben jedoch vercodet und werden in dieser Form (teilweise aggregiert) zur Verfügung gestellt.

Zuletzt wurde geprüft, ob in den Daten *sensible Informationen*, z. B. zur Gesundheit, sexuellen Orientierung und zu politischen Einstellungen, enthalten waren. Diese eignen sich zwar nicht notwendig zur Re-Identifikation von Individuen oder Institutionen, jedoch können die

Informationen im Falle einer De-Anonymisierung nutzbringend sein (vgl. Koberg, 2016a, S. 694) und sind daher besonders schützenswert (Art. 9 DSGVO, Erwägungsgrund 51 DSGVO). Bei Nacaps 2018 wurde für einen Großteil der abgefragten Gesundheitsinformationen bei den Befragten eine zusätzliche Einverständniserklärung eingeholt. Diese Gesundheitsinformationen dürfen im On-Site- und Remote-Desktop-SUF herausgegeben werden. Weitere Einzelinformationen zur Gesundheit, für die nicht explizit eine Einwilligung eingeholt wurde, werden nicht freigegeben.

Im CUF wurden zum einen im Vergleich zu den SUF-Varianten teils restriktivere statistische Anonymisierungsmaßnahmen auf Variablenebene vorgenommen, zum anderen wurde eine per Zufallsauswahl gewonnene Substichprobe der Daten (25 % der Befragten) gezogen.

Die nachfolgende Tabelle 11 stellt in Kurzform die durchgeführten statistischen Anonymisierungsmaßnahmen je nach Datenform bzw. Zugangsweg dar. Genauere Informationen zur Aggregation der einzelnen Variablen sind im Datensatzreport zu finden.

Tabelle 11: Maßnahmen der statistischen Anonymisierung nach Zugangsweg

Merkmal	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (25%-Substichprobe)
Direkte Identifikatoren	keine Freigabe/ Generieren einer systemfreien ID	keine Freigabe/ Generieren einer systemfreien ID	keine Freigabe/ Generieren einer systemfreien ID	keine Freigabe/ Generieren einer systemfreien ID
Sprache des Fragebogens	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Preload Hochschule	keine Freigabe	keine Freigabe	keine Freigabe	keine Freigabe
Datenschutzangaben	keine Freigabe	keine Freigabe	keine Freigabe	keine Freigabe
Zeitpunkt des inhaltlichen Beginns mit der Promotion (Monat/Jahr)	Freigabe	Freigabe	Monat: Aggregation zu Quartalen; Missing, wenn vor 2005 Jahr: Aggregation, wenn vor 2005	Monat: Aggregation zu Quartalen; Missing, wenn vor 2005 Jahr: Aggregation, wenn vor 2005
Zeitpunkt der Zulassung zur Promotion (Monat/Jahr)	Freigabe	Freigabe	Monat: Aggregation zu Quartalen; Missing, wenn vor 2005 Jahr: Aggregation, wenn vor 2005	Monat: Aggregation zu Quartalen; Missing, wenn vor 2005 Jahr: Aggregation, wenn vor 2005
Zeitpunkt des Abschlusses der Promotion (Monat)	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Zeitpunkt des Abbruchs der Promotion (Monat/Quartal)	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Gemeinsamer Abschluss der Hochschulen im In-/Ausland	keine Freigabe	keine Freigabe	keine Freigabe	keine Freigabe
Promotionsfach (offene Angabe)	Vercodung/ Aggregation zu DZHW-Fächergruppen ^a	Vercodung/ Aggregation zu DZHW-Fächergruppen ^a	Aggregation zu Destatis-Fächergruppen ^b	Aggregation zu Destatis-Fächergruppen ^b
Art Promotionsprogramm	keine Freigabe	keine Freigabe	keine Freigabe	keine Freigabe
Gründe Abbruchgedanken: gesundheitliche Probleme	keine Freigabe	keine Freigabe	keine Freigabe	keine Freigabe
Einrichtung Betreuer*innen (Promotionshochschule/andere Einrichtung)	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Geschlecht Betreuer*innen	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Formaler Betreuungsstatus Betreuer*innen	Aggregation zu Kategorien: - „offiziell“	Aggregation zu Kategorien: - „offiziell“	keine Freigabe	keine Freigabe

Merkmal	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (25%-Substichprobe)
	- „nicht offiziell“	- „nicht offiziell“		
Position Betreuer*innen	keine Freigabe	keine Freigabe	keine Freigabe	keine Freigabe
Hauptbetreuer*in	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Sympathie Betreuer*in	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Stipendiengeber	Aggregation zu Kategorien: - „Stiftung/DFG“ - „Uni/FH/außer-universitäre Forschungseinrichtung/Sonstiges“	Aggregation zu Kategorien: - „Stiftung/DFG“ - „Uni/FH/außer-universitäre Forschungseinrichtung/Sonstiges“	keine Freigabe	keine Freigabe
Ideelle Promotionsförderung	Aggregation zu Variablen: - „Stiftung/DFG“ - „Uni/FH/außer-universitäre Forschungseinrichtung/Sonstiges“ - „keine der genannten Institutionen“	Aggregation zu Variablen: - „Stiftung/DFG“ - „Uni/FH/außer-universitäre Forschungseinrichtung/Sonstiges“ - „keine der genannten Institutionen“	keine Freigabe	keine Freigabe
Beginn des gegenwärtigen Vertrags (Monat)	Freigabe	Freigabe	Aggregation zu Kategorien: - „Januar/Februar“ - „März/April“ - „Mai/Juni“ - „Juli/August“ - „September/Oktober“ - „November/Dezember“	Aggregation zu Kategorien: - „Januar/Februar“ - „März/April“ - „Mai/Juni“ - „Juli/August“ - „September/Oktober“ - „November/Dezember“
Länder Auslandsaufenthalt (offene Angabe)	Vercodung zu Ländern ^c	Vercodung zu Ländern ^c	Aggregation zu Weltregionen ^c (Dummy-Variable pro Weltregion)	Aggregation zu Weltregionen ^c (Dummy-Variable pro Weltregion)
Hinderungsgrund Auslandsaufenthalt: gesundheitliche Probleme	keine Freigabe	keine Freigabe	keine Freigabe	keine Freigabe
Anzahl Kinder	Aggregation von 5 und mehr Kindern			
Geburtsjahr Kinder	1.-4. Kind: Freigabe 5.-10. Kind: keine Freigabe	1.-4. Kind: Aggregation zu Kategorien: - „bis 1999“ - „2000-2006“ - „2007-2012“ - „2013-2015“ - „ab 2016“ 5.-10. Kind: keine Freigabe	keine Freigabe	keine Freigabe
Allgemeiner Gesundheits-	Freigabe	Freigabe	keine Freigabe	keine Freigabe

Merkmal	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (25%-Substichprobe)
zustand				
Anzahl Tage: gesundheitlich beeinträchtigt in letzten vier Wochen	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Amtlich anerkannte Behinderung	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Jahr der Anerkennung der Behinderung	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Grad der Behinderung	Aggregation zu Kategorien: - „bis 49 (keine Schwerbehinderung)“ - „ab 50 (Schwerbehinderung)“	Aggregation zu Kategorien: - „bis 49 (keine Schwerbehinderung)“ - „ab 50 (Schwerbehinderung)“	keine Freigabe	keine Freigabe
Programme für Menschen mit Behinderung an der Hochschule bekannt	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Teilnahme an Programmen für Menschen mit Behinderung	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Geschlecht	Freigabe	Freigabe	Aggregation der Kategorien "divers" und "keine Angabe" zu "nicht bestimmbar"	Aggregation der Kategorien "divers" und "keine Angabe" zu "nicht bestimmbar"
Geburtsjahr	1970-1994 einzeln ausweisen, ansonsten Aggregation zu Kategorien: - „bis 1959“ - „1960-1964“ - „1965-1969“ - „1995 und jünger“	1970-1994 einzeln ausweisen, ansonsten Aggregation zu Kategorien: - „bis 1959“ - „1960-1964“ - „1965-1969“ - „1995 und jünger“	Aggregation zu Kategorien: - „bis 1969“ - „1970-1979“ - „1980-1984“ - „1985-1986“ - „1987-1988“ - „1989-1990“ - „1991-1992“ - „1993-1994“ - „1995 und jünger“	Aggregation zu Kategorien: - „bis 1969“ - „1970-1979“ - „1980-1984“ - „1985-1986“ - „1987-1988“ - „1989-1990“ - „1991-1992“ - „1993-1994“ - „1995 und jünger“
Geburtsland (offene Angabe)	Vercodung zu Ländern ^c	Aggregation zu NEPS-Regionen ^d	Aggregation zu Weltregionen ^c	Aggregation zu Weltregionen ^c
Land, in dem am 30. Juni 1990 gelebt wurde (BRD/DDR/anderes Land)	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Land, in dem am 30. Juni 1990 gelebt wurde (offene Angabe)	Vercodung zu Ländern ^c	Aggregation zu NEPS-Regionen ^d	Aggregation zu Weltregionen ^c	Aggregation zu Weltregionen ^c
Jahr der Zuwanderung	Freigabe	Freigabe	Aggregation zu Kategorien:	Aggregation zu Kategorien:

Merkmal	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (25%-Substichprobe)
nach Deutschland			- „bis 1989“ - „1990-1999“ - „2000-2009“ - „ab 2010“	- „bis 1989“ - „1990-1999“ - „2000-2009“ - „ab 2010“
Staatsangehörigkeit (offene Angabe)	Vercodung zu Ländern ^c	Aggregation zu NEPS-Regionen ^d	Aggregation zu Weltregionen ^c (Dummy-Variable pro Weltregion)	Aggregation zu Weltregionen ^c (Dummy-Variable pro Weltregion)
Jahr deutsche Staatsangehörigkeit	Freigabe	Freigabe	Aggregation zu Kategorien: - „bis 1989“ - „1990-1999“ - „2000-2009“ - „ab 2010“	Aggregation zu Kategorien: - „bis 1989“ - „1990-1999“ - „2000-2009“ - „ab 2010“
Geburtsland Eltern (offene Angabe)	Vercodung zu Ländern ^c	Aggregation zu NEPS-Regionen ^d	Aggregation zu Weltregionen ^c	Aggregation zu Weltregionen ^c
Staatsangehörigkeit Eltern (offene Angabe)	Vercodung zu Ländern ^c	Aggregation zu NEPS-Regionen ^d	Aggregation zu Weltregionen ^c (Dummy-Variable pro Weltregion)	Aggregation zu Weltregionen ^c (Dummy-Variable pro Weltregion)
Land, in dem am 30. Juni 1990 gelebt wurde Eltern (BRD/DDR/anderes Land)	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Land, in dem am 30. Juni 1990 gelebt wurde Eltern (offene Angabe)	Vercodung zu Ländern ^c	Aggregation zu NEPS-Regionen ^d	Aggregation zu Weltregionen ^c	Aggregation zu Weltregionen ^c
Höchster Schulabschluss Eltern	Freigabe	Freigabe	Aggregation zu Kategorien: - „Allgemeine Hochschulreife/Abitur“ - „Sonstiges“ - „nicht zuordenbar“	Aggregation zu Kategorien: - „Allgemeine Hochschulreife/Abitur“ - „Sonstiges“ - „nicht zuordenbar“
Höchster Ausbildungsabschluss Eltern	Freigabe	Freigabe	Aggregation zu Kategorien: - „Promotion (Dokortitel)“ - „andere Abschlüsse“ - „nicht zuordenbar“	Aggregation zu Kategorien: - „Promotion (Dokortitel)“ - „andere Abschlüsse“ - „nicht zuordenbar“
Berufliche Stellung Eltern	Freigabe	Freigabe	Aggregation zu Kategorien: - „Angestellte(r)“ - „Arbeiter(in)“ - „andere“ - „nicht zuordenbar“	Aggregation zu Kategorien: - „Angestellte(r)“ - „Arbeiter(in)“ - „andere“ - „nicht zuordenbar“
Beruf der Eltern (offene Angabe)	keine Freigabe	keine Freigabe	keine Freigabe	keine Freigabe
Land Studien-	Freigabe	Aggregation zu	Aggregation zu	Aggregation zu

Merkmalsname	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (25%-Substichprobe)
berechtigung		Weltregionen ^c (Dummy-Variable pro Weltregion)	Weltregionen ^c (Dummy-Variable pro Weltregion)	Weltregionen ^c (Dummy-Variable pro Weltregion)
Jahr Studienberechtigung	1995-2017 einzeln ausweisen, ansonsten Aggregation zu Kategorien: - „bis 1979“ - „1980-1984“ - „1985-1989“ - „1990-1994“ - „2018-2019“	1995-2017 einzeln ausweisen, ansonsten Aggregation zu Kategorien: - „bis 1979“ - „1980-1984“ - „1985-1989“ - „1990-1994“ - „2018-2019“	Aggregation zu Kategorien: - „bis 1989“ - „1990-1999“ - „2000-2004“ - „2005-2006“ - „2007-2008“ - „2009-2010“ - „2011-2012“ - „2013-2019“	Aggregation zu Kategorien: - „bis 1989“ - „1990-1999“ - „2000-2004“ - „2005-2006“ - „2007-2008“ - „2009-2010“ - „2011-2012“ - „2013-2019“
Abschlussnote Studienberechtigung	Freigabe	Freigabe	keine Freigabe	keine Freigabe
Studienfach des Hochschulabschlusses, der zur Promotion berechtigt (offene Angabe)	Vercodung/ Aggregation zu Destatis-Studienbereichen ^b	Vercodung/ Aggregation zu Destatis-Studienbereichen ^b	Aggregation zu Destatis-Fächergruppen ^b	Aggregation zu Destatis-Fächergruppen ^b
Hochschule des Hochschulabschlusses, der zur Promotion berechtigt (offene Angabe)	Vercodung/ Aggregation zum Bundesland der Hochschule ^e Vercodung/ Aggregation zur Hochschulart (Universität vs. Fachhochschule) ^e	Aggregation zum Bundesland der Hochschule: Baden-Württemberg, Bayern, Hessen und Nordrhein-Westfalen einzeln ausweisen, ansonsten Bundesländergruppen: - „Berlin, Brandenburg und Mecklenburg-Vorpommern“ - „Bremen und Niedersachsen“ - „Hamburg und Schleswig-Holstein“ - „Rheinland-Pfalz und Saarland“ - „Sachsen, Sachsen-Anhalt und Thüringen“	Aggregation zum Bundesland der Hochschule: Baden-Württemberg, Bayern, Hessen und Nordrhein-Westfalen einzeln ausweisen, ansonsten Bundesländergruppen: - „Berlin, Brandenburg und Mecklenburg-Vorpommern“ - „Bremen und Niedersachsen“ - „Hamburg und Schleswig-Holstein“ - „Rheinland-Pfalz und Saarland“ - „Sachsen, Sachsen-Anhalt und Thüringen“	Aggregation zur Region der Hochschule: - „Norden“ - „Süden“ - „Osten“ - „Westen“
Land des Hochschulabschlusses, der zur Promotion berechtigt (offene Angabe)	Vercodung zu Ländern ^c	Aggregation zu Weltregionen ^c	Aggregation zu Weltregionen ^c	Aggregation zu Weltregionen ^c
Abschlussjahr des Hochschulabschlusses, der zur Promotion berechtigt	2010-2017 einzeln ausweisen, ansonsten Aggregation zu Kategorien:	2010-2017 einzeln ausweisen, ansonsten Aggregation zu Kategorien:	Aggregation zu Kategorien: - „bis 1999“ - „2000-2004“ - „2005-2009“	Aggregation zu Kategorien: - „bis 1999“ - „2000-2004“ - „2005-2009“

Merkmal	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (25%-Substichprobe)
	- „bis 1989“	- „bis 1989“	- „2010-2014“	- „2010-2014“
	- „1990-1994“	- „1990-1994“	- „2015-2019“	- „2015-2019“
	- „1995-1999“	- „1995-1999“		
	- „2000-2004“	- „2000-2004“		
	- „2005-2009“	- „2005-2009“		
	- „2018-2019“	- „2018-2019“		
Weitere offene Angaben	keine Freigabe/ keine Vercodung	keine Freigabe/ keine Vercodung	keine Freigabe/ keine Vercodung	keine Freigabe/ keine Vercodung

Sonderhinweis: Auch die offenen Angaben zur Promotionshochschule werden über keinen Zugangsweg freigegeben. Sie werden aus Datenschutzgründen auch nicht in vercodeter bzw. aggregierter Form bereitgestellt.

^a projekteigene Codierung (vgl. cl-dzhw-43)

^b Destatis-Fächersystematik WiSe 2017/2018 (Statistisches Bundesamt 2018) (vgl. cl-dzhw-43)

^c Destatis-Staats- und Gebietssystematik, Stand 01.04.2019 (Statistisches Bundesamt 2019b) (vgl. cl-dzhw-42)

^d Länderklassifikation des NEPS, https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/6-0-0/SC5_6-0-0_Anonymisation.pdf, S. 33-34 (Koberg 2016b) (vgl. cl-dzhw-42)

^e Destatis-Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2018/19 und SoSe 2019 (Statistisches Bundesamt 2019a)

9 Literaturverzeichnis

- AAPOR (2016). Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys. https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf (Zugegriffen am: 09.06.2020).
- Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2012). Ein Messinstrument zur Erfassung subjektiver Kompetenzerwartungen: Allgemeine Selbstwirksamkeit Kurzskala (ASKU) Hg. v. GESIS - Leibniz-Institut für Sozialwissenschaften. (GESIS Working Papers,17).
- Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2014). Eine Single-Item-Skala zur Erfassung von Risikobereitschaft: Die Kurzskala Risikobereitschaft-1 (R-1). (GESIS-Working Papers, 2014/34). Mannheim: GESIS Leibniz-Institut für Sozialwissenschaften.
- Beierlein, C., Kovaleva, A., László, Z., Kemper, C. J., & Rammstedt, B. (2015). Kurzskala zur Erfassung der Allgemeinen Lebenszufriedenheit (L-1). Zusammenstellung sozialwissenschaftlicher Items und Skalen. doi:10.6102/zis229.
- Ebel, T., & Meyermann, A. (2015). Hinweise zur Anonymisierung von quantitativen Daten. Forschungsdaten Bildung informiert. Bd.3. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung. <http://www.forschungsdaten-bildung.de/fdb-informiert> (Zugegriffen am: 09.06.2020).
- Gagné, M., Forest, J., Vansteenkiste, M., Crevier-Braud, L., Van den Broeck, A., Aspel, A. K., Bellerose, J., Benabou, C., Chemolli, E., Güntert, S. T., Halvari, H., Indiyastuti, D. L., Johnson, P. A., Hauan Molstad, M., Naudin, M., Ndao, A., Hagen Olafsen, A., Roussel, P., Wang, Z., & Westbye, C. (2015). The Multidimensional Work Motivation Scale: Validation evidence in seven languages and nine countries. *European Journal of Work and Organizational Psychology* 24 (2), 178-196. doi: 10.1080/1359432X.2013.877892.
- GESIS - Leibniz-Institut für Sozialwissenschaften (2015). Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2014. GESIS Datenarchiv, Köln. ZA5240 Fragebogen.
- Hochfellner, D., Müller, D., Schmucker, A., & Roß, E. (2012). Datenschutz am Forschungsdatenzentrum. (FDZ-Methodenreport Nr. 6). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Koberg, T. (2016a). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer, & J. Skopek (Hrsg.), *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study* (S. 691–708). Wiesbaden: Springer VS. doi:10.1007/978-3-658-11994-2.
- Koberg, T. (2016b). Starting Cohort 5: First-Year Students (SC5). SUF Version 6.0.0. Anonymization Procedures. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/~Forschungsdaten/SC5/6-0-0/SC5_6-0-0_Anonymisation.pdf (Zugegriffen am: 05.06.2020).
- Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2012). Eine Kurzskala zur Messung von Kontrollüberzeugung: Die Skala Internale-Externale-Kontrollüberzeugung-4 (IE-4). Hg. v. GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS-Working Papers, 19).
- Krempkow, R., Sembritzki, T., Schürmann, R., & Winde, M. (2016). Personalentwicklung für den wissenschaftlichen Nachwuchs. Bedarf, Angebote und Perspektiven – eine empirische Bestandsaufnahme im Zeitvergleich. Essen.
- Lane, J., Heus, P., & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyber-infrastructure. *Transactions on Data Privacy*, 1(1), 2–16.

- Lenzner, T., Neuert, C., & Otto, W. (2015). Kognitives Pretesting. Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). doi: 10.15465/gesis-sg_010.
- Lind, I., & Samjeske, K. (2010). Schlussbericht zum Projekt "Balancierung von Wissenschaft und Elternschaft" (BAWIE). Bonn.
- Mainhard, T., van der Rijst, R., & van Tartwijk, J. (2009). A model for the supervisor–doctoral student relationship. *Higher Education* 58, 359–373. <https://doi.org/10.1007/s10734-009-9199-8>.
- Prüfer, P., & Rexroth, M. (2000). Zwei-Phasen-Pretesting. (ZUMA-Arbeitsbericht 2000/08). Mannheim: Zentrum für Umfragen, Methoden und Analysen (ZUMA). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-200864> (Zugegriffen am: 05.06.2020).
- Roach, M., & Sauer mann, H. (2010). A taste for science? PhD scientists' academic orientation and selfselection into research careers in industry. *Research Policy* 39 (3), 422–434.
- Robert-Koch-Institut (2015). Gesundheit in Deutschland aktuell: GEDA 2014/2015. Berlin.
- Schupp, J., & Gerlitz, J.-Y. (2014). Big Five Inventory-SOEP (BFI-S). Zusammenstellung sozialwissenschaftlicher Items und Skalen. doi:10.6102/zis54.
- Statistisches Bundesamt. (2016). Promovierende in Deutschland. Wintersemester 2014/2015. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2018). Bildung und Kultur. Studierende an Hochschulen. Fächersystematik. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2019a). Schlüsselverzeichnisse für die Studenten- und Prüfungsstatistik, Promovierendenstatistik WS 2018/2019 und SS 2019. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2019b). Staats- und Gebietssystematik. Stand 01.04.2019. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2019c). Bildung und Kultur. Promovierendenstatistik: Analyse zu Vollständigkeit und Qualität der zweiten Erhebung 2018. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2019d). Bildung und Kultur. Prüfungen an Hochschulen 2018. (Fachserie 11/4.2). Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2019e). Bildung und Kultur. Studierende an Hochschulen. Wintersemester 2018/19. (Fachserie 11/4.1). Wiesbaden: Statistisches Bundesamt.
- Tremblay, M. A., Blanchard, C. M., Taylor, S., Pelletier, L. G., & Villeneuve, M. (2009). Work Extrinsic and Intrinsic Motivation Scale: Its value for organizational psychology research. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, 41(4), 213–226. doi:10.1037/a0015167.
- Vollmar, M. (2019). Neue Promovierendenstatistik: Analyse der ersten Erhebung 2017. *WISTA Wirtschaft und Statistik*, 1, 68-79.
- Weichbold, M. (2019). Pretest. In N. Bauer, & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 349-356). Wiesbaden: Springer VS.
- Wegner, A., & Briedis, K. (2020). National Academics Panel Study – Die Längsschnittstudie zu Promovierenden und Promovierten in Deutschland. Wissenschaftliches Projektkonzept. https://www.nacaps.de/files/downloads/nacaps_wisskonzept.pdf (Zugegriffen am: 19.06.2020).

Anhang: Nacaps-Partnerhochschulen

- Akademie der Bildenden Künste München
- Bergische Universität Wuppertal
- Bucerius Law School
- Christian-Albrechts-Universität zu Kiel
- Eberhard Karls Universität Tübingen
- Europa-Universität Flensburg
- Fernuniversität in Hagen
- Freie Universität Berlin
- Friedrich-Alexander-Universität Erlangen-Nürnberg
- Friedrich-Schiller-Universität Jena
- Heinrich-Heine-Universität Düsseldorf
- Helmut-Schmidt-Universität
- Hertie School of Governance
- Hochschule Fulda
- Hochschule für Bildende Künste Dresden
- Hochschule für Musik Franz Liszt Weimar
- Hochschule für Musik, Theater und Medien Hannover
- Hochschule Geisenheim
- Johann Wolfgang Goethe-Universität Frankfurt am Main
- Johannes Gutenberg-Universität Mainz
- Katholische Universität Eichstätt-Ingolstadt
- Kühne Logistics University Hamburg
- Musikhochschule Lübeck
- Otto-Friedrich-Universität Bamberg
- Otto-von-Guericke-Universität Magdeburg
- Pädagogische Hochschule Freiburg
- Pädagogische Hochschule Ludwigsburg
- Pädagogische Hochschule Weingarten
- Rheinisch-Westfälische Technische Hochschule Aachen
- Ruhr-Universität Bochum
- Ruprecht-Karls-Universität Heidelberg
- Staatliche Akademie der Bildenden Künste Stuttgart
- Stiftung Tierärztliche Hochschule Hannover

- Stiftung Universität Hildesheim
- Technische Universität Clausthal
- Technische Universität Dortmund
- Technische Universität Kaiserslautern
- Technische Universität München
- Universität Bielefeld
- Universität des Saarlandes
- Universität Duisburg-Essen
- Universität Hamburg
- Universität Kassel
- Universität Leipzig
- Universität Mannheim
- Universität Osnabrück
- Universität Passau
- Universität Potsdam
- Universität Trier
- Universität Ulm
- Universität Vechta
- WHU – Otto Beisheim School of Management
- Zeppelin Universität