

Lara Berroth | Kolja Briedis | Ute Hoffstätter | Vanessa K. Jänsch | Kai Mühleck |
Tina Oestreich | Isabelle Schiffer-Fiedler | Ulrike Schwabe | Carola Teichmann |
Anne Weber | Antje Wegner

Nacaps 2020

Daten- und Methodenbericht zur Datenpaketversion
1.0.0 der National Academics Panel Study 2020
(1.-2. Befragungswelle)

nacaps 

fdz.DZHW.
Forschungsdatenzentrum
für Hochschul- und Wissenschaftsforschung

Dieses Werk steht unter der Creative Commons Namensnennung Nicht kommerziell-Share Alike 4.0 International (CC BY-NC-SA 4.0) Deed

<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.de>

Bibliographische Angaben

Bertho, L., Briedis, K., Hoffstätter, U., Jänsch, V. K., Mühleck, K., Oestreich, T., Schiffer-Fiedler, I., Schwabe, U., Teichmann, C., Weber, A. & Wegner, A. (2026). Nacaps 2020. Daten- und Methodenbericht zur Datenpaketversion 1.0.0 der National Academics Panel Study 2020 (1.-2. Befragungswelle). Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). <https://doi.org/10.21249/DZHW:nac2020-dmr-de:1.0.0>

Zugehöriges Datenpaket: <https://doi.org/10.21249/DZHW:nac2020:1.0.0>

Impressum

Herausgegeben von

Deutsches Zentrum für Hochschul- und
Wissenschaftsforschung GmbH (DZHW)
Lange Laube 12 | 30159 Hannover | www.dzhw.eu
Postfach 2920 | 30029 Hannover
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960

Geschäftsführung

Dr. Marcus Beiner
Regina Oelfke

Vorsitzender des Aufsichtsrats

Ministerialdirigent Peter Greisler

Registergericht

Amtsgericht Hannover | HRB 210251
Umsatzsteuer-Identifikationsnummer:
DE291239300

Inhaltsverzeichnis

Tabellen-/Abbildungsverzeichnis	II
1 Inhalt und Anlage der Studie	1
2 Erhebungsinstrument	3
2.1 Erstbefragung.....	3
2.2 Folgebefragung	4
3 Grundgesamtheit und Stichprobenverfahren	6
4 Durchführung der Erhebung	7
4.1 Erstbefragung.....	7
4.2 Folgebefragung	8
5 Rücklauf	10
5.1 Erstbefragung.....	10
5.2 Folgebefragung	12
6 Datenaufbereitung	13
6.1 Datenübertragung.....	13
6.2 Datenprüfung und -bereinigung	13
6.3 Generierung von Variablen	14
6.4 Datenstruktur und Datenformat.....	15
6.5 Vergabe von Variablennamen und Variablen-/Wertelabels	15
6.6 Codierung fehlender Werte	16
7 Repräsentativität der Daten	19
8 Gewichtung	21
8.1 Vorgehen und Anwendungshinweise	21
8.2 Modellierung des Ausfallgewichts für Welle 2.....	22
8.3 Kalibrierung.....	22
8.4 Trimmung.....	23
8.5 Verteilung der Gewichte	23
9 Anonymisierung	24
10 Literatur	26
11 Anhang	28

11.1 Dokumentation der Herkunft sekundär genutzter Fragen 28

11.2 Logit-Regressionen zur Erstellung der Panelausfallgewichte 29

Tabellen-/Abbildungsverzeichnis

Abbildung 1: Nacaps-Erhebungsdesign..... 1

Tabelle 1: Versandtage der Erhebungsunterlagen der Erstbefragung 7

Tabelle 2: Versandtage der Erhebungsunterlagen der Folgewellen..... 9

Tabelle 3: Beteiligung der Hochschulen 10

Tabelle 4: Brutto- und Nettorücklaufquote der ersten Befragungswelle 11

Abbildung 2: Relativer Anteil beendeter Fragebögen (in %) über den Verlauf der Feldzeit 11

Tabelle 5: Brutto- und Nettorücklaufquote der Nacaps-Folgebefragungswellen (nach AAPOR 2016) 12

Tabelle 6: Themengebiete und Kürzel für Variablennamen in Nacaps 2020..... 16

Tabelle 7: Systematik des FDZ-DZHW für fehlende Werte 18

Tabelle 8: Vergleich der Grundgesamtheit laut Promovierendenstatistik mit Nacaps 2020 nach Fächergruppen 19

Tabelle 9: Vergleich der Grundgesamtheit laut Promovierendenstatistik mit Nacaps 2020 nach Fächergruppen und Geschlecht 20

Tabelle 10: Bereitgestellte Gewichte zu Nacaps 2020 21

Tabelle 11: Verteilung der Gewichte in Nacaps 2020..... 23

Tabelle 12: Zugangswege im FDZ-DZHW 24

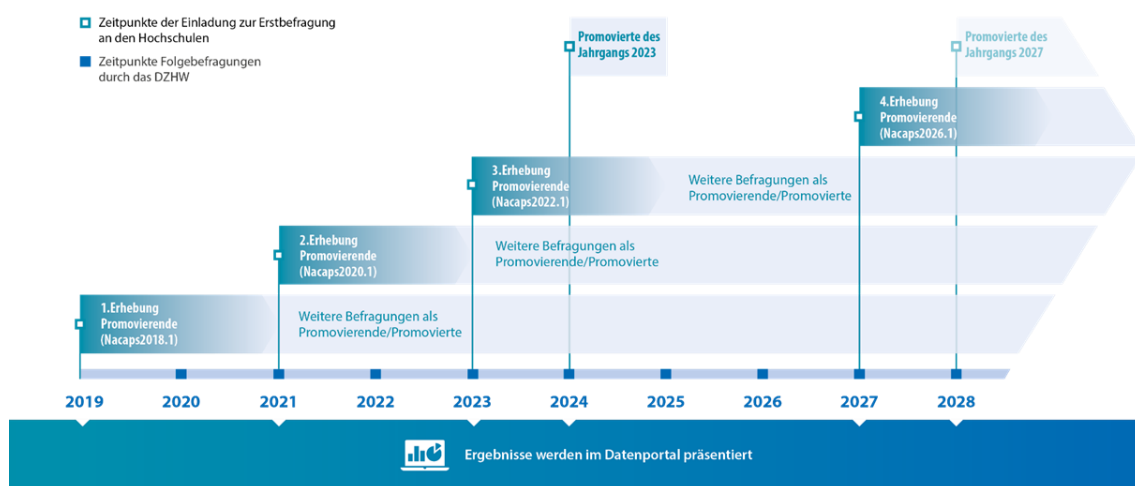
Tabelle 13: Logit-Regression zur Erstellung des Ausfallgewichts für Welle 2 29

1 Inhalt und Anlage der Studie

Nacaps 2020 ist Teil der Studienreihe *National Academics Panel Study (Nacaps)*, einer Längsschnittstudie, die erstmalig systematisch anhand standardisierter Online-Befragungen von Promovierenden und Promovierten Daten zu Karriere- und Lebensverläufen von Hochqualifizierten in Deutschland erhebt – für Forschung, Hochschulen und Wissenschaftspolitik (Briedis et al., 2025; Wegner & Briedis, 2020).¹ Nacaps wird vom Deutschen Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) durchgeführt und wurde bis einschließlich 2024 vom Bundesministerium für Bildung und Forschung (BMBF) gefördert.

Die Nacaps-Studienreihe weist ein Multi-Kohorten-Panel-Design auf, bei dem im Abstand von zwei bzw. vier Jahren Promovierendenkohorten kontaktiert und in der Folge jährlich befragt werden (vgl. Abbildung 1). Bislang gibt es die Promovierendenkohorten 2018, 2020 und 2022, die jeweils im Folgejahr zum ersten Mal befragt wurden. Zur Auffrischung und partiellen Ergänzung der Promovierenden-erhebungen wurden im Jahr 2024 zusätzlich Promovierte des Jahres 2023 einmalig befragt. Geplant ist zudem eine neue Promovierendenkohorte, die im Jahr 2027 zum ersten Mal befragt werden soll.

Abbildung 1: Nacaps-Erhebungsdesign



Nacaps 2020 bildet die zweite Promovierendenkohorte der Studienreihe. Die Grundgesamtheit beinhaltet alle Personen, die zwischen dem 02. Dezember 2018 und dem 01. Dezember 2020 an einer deutschen Hochschule zur Promotion registriert worden sind. Die erste Befragungswelle fand von Februar bis April 2021 statt. Die Folgebefragungswellen erfolgten im jährlichen Abstand jeweils im Frühjahr/Sommer.

Neben Grunddaten für die Bildungsberichterstattung enthalten die Daten detaillierte Informationen über die Lern- und Entwicklungsbedingungen während der Promotionsphase sowie über die Karriere- und Lebensverläufe von Promovierten und Promotionsabbrecher*innen im Anschluss an die Promotionsphase. Darüber hinaus beinhalten die Daten eine Reihe von Persönlichkeitsmerkmalen (u. a. Big Five, Selbstwirksamkeit, internale/externale Kontrollüberzeugungen), Angaben zur

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

Gesundheitssituation sowie sozio- und bildungsbiografische Hintergrundinformationen. Dadurch ergibt sich ein großes, bisher nicht vorhandenes Analysepotenzial für die Hochschul- und Wissenschaftsforschung.

2 Erhebungsinstrument

In allen Befragungswellen von Nacaps 2020 wurde als Erhebungsinstrument ein standardisierter Online-Fragebogen in deutscher und englischer Sprache eingesetzt. Die Programmierung des Fragebogens und die Durchführung der Befragung erfolgten mit der DZHW-Onlinebefragungssoftware *Zofar*. Um einen Kohortenvergleich zwischen Nacaps 2018 und Nacaps 2020 zu ermöglichen, entspricht das Erhebungsinstrument in großen Teilen dem der Kohorte 2018.

Auf umfangreiche Pretests wurde verzichtet. Diese wurden bereits bei der Entwicklung des Erhebungsinstrumentes für die vorherige Kohorte durchgeführt und dokumentiert (vgl. Briedis et al., 2025).

Die Inhalte der Erhebungsinstrumente werden zunächst für die Erstbefragung (vgl. Kapitel 2.1) und anschließend für die Folgebefragung (vgl. Kapitel 2.2) dargestellt. Darauf hingewiesen sei, dass einige Fragen der Erhebungsinstrumente aus anderen Instrumenten – teils unmodifiziert, teils modifiziert – sekundär genutzt wurden, vgl. dazu Anhang 11.1.

2.1 Erstbefragung

Der Fragebogen der ersten Befragungswelle war in fünf thematische Blöcke gegliedert. Es wurden Angaben zur Promotion (A), Promotionsthema und die Promotionsbedingungen (B), Arbeitsalltag in der Promotionsphase (C), persönliche Lebenssituation und Karriereziele (D) und abschließend Soziodemografie und vorangegangene Bildungsetappen (E) erfasst. In einem letzten Schritt wurden Aspekte wie Panelbereitschaft, Newsletter, Verlosung und Abfrage der Kontaktdaten erfasst.

Zu Beginn der Befragung wurden die Befragten zudem nach ihrem derzeitigen Promotionsstatus² gefragt. Je nach Promotionsstatus wurden im weiteren Verlauf der Befragung verschiedene Fragen angezeigt, die genau auf die Situation der Befragten zugeschnitten waren.

[Angaben zur Promotion] Der Themenblock Angaben zur Promotion umfasste Fragen zur Motivation bei Promotionsbeginn, zum inhaltlichen Beginn und 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 zum Befragungszeitpunkt bereits abgeschlossen war, wurde ferner nach dem Datum des Abschlusses gefragt. War die Promotion zum Befragungszeitpunkt abgebrochen, 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, zu Kooperationen mit externen Organisationen (mit einem Schwerpunkt auf Kooperationen mit Unternehmen), zur Form der Promotion (Monographie vs. kumulative Promotion), zu Abbruchintentionen, zum Vorhandensein einer Promotions- bzw. Betreuungsvereinbarung sowie zur Anzahl an Betreuenden. Ferner umfasste er Fragen zu den Betreuenden der Promotion (Affiliation, Geschlecht, formaler Betreuungsstatus, Position,

² Es gab fünf Antwortmöglichkeiten: „Ich promoviere“, „Ich habe das Promotionsverfahren abgeschlossen“, „Ich habe mein Promotionsvorhaben unterbrochen“, „Ich habe mein Promotionsvorhaben abgebrochen“ und „Ich habe meine Dissertationsschrift bzw. alle relevanten Arbeiten zur Begutachtung eingereicht“.

Reputation, Austauschhäufigkeit, Betreuungsstil, Zufriedenheit sowie Sympathieeinschätzung) sowie zur guten wissenschaftlichen Praxis.

[Arbeitsalltag in der Promotionsphase] Der Themenblock Arbeitsalltag in der Promotionsphase setzte sich zusammen aus Fragen zu Finanzierungsquellen, zur Beschäftigung (wie Befristung, Stellenumfang), zu den Qualifizierungszielen sowie zum thematischen Bezug der Beschäftigung zum Promotionsprojekt. Weiterhin wurde nach wissenschaftlichen Vortrags- und Publikationstätigkeiten, 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 (geplanten und realisierten) 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 allgemeine Risikobereitschaft, die allgemeinen Kontrollüberzeugungen, die allgemeinen Selbstwirksamkeitserwartungen sowie Zielverfolgungsstrategien über das TEN/FLEX-Instrument (vgl. Brandtstädter & Renner, 1990) erfasst. Ferner wurden Instrumente aus der Kohorte Nacaps 2018 wieder eingesetzt, die die Wichtigkeit von bestimmten Stelleneigenschaften, die Attraktivität von verschiedenen Beschäftigungsmöglichkeiten, 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 Fragen nach Geburtsort und Staatsangehörigkeiten, Schulabschluss, Ausbildungsabschluss und beruflicher Stellung der Eltern. Außerdem wurden der Ort und die Note der Studienberechtigung sowie Angaben zum promotionsrelevanten Studium und zu einer fachnahen Erwerbstätigkeit vor Promotionsbeginn erhoben.

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

2.2 Folgebefragung

In der zweiten Befragungswelle waren bestimmte Kernthemen enthalten. Zum Teil wurden dabei auch Angaben aus der ersten Welle in die Befragung eingespielt („Preload“), um daran anknüpfen zu können. Die genauen Inhalte der Kernthemen werden im Folgenden genauer erläutert.

[Allgemeine Angaben zur Promotion] Einleitend wurde nach dem aktuellen Promotionsstatus (promovierend, unterbrochen, Dissertationsschrift bzw. alle relevanten Arbeiten zur Begutachtung eingereicht, abgeschlossen, abgebrochen) gefragt. Sofern die Promotion im Anschluss an die vorausgehende Befragung abgeschlossen wurde, wurde ferner nach dem Datum des Abschlusses, nach Änderungen des Promotionsthemas, der Dissertationsart (Monographie vs. kumulative Promotion) und -veröffentlichung sowie nach der Promotionsnote und der selbsteingeschätzten Reputation der erstbegutachtenden Person und der Institution gefragt; sofern die Promotion abgebrochen/unterbrochen wurde, wurden die Gründe und das Datum des Abbruchs erhoben. Für Befragte, die ihre Promotion im Anschluss an die Vorwelle abgeschlossen oder abgebrochen haben, wurden zudem (nochmals) die Promotionshochschule und das (angestrebte) Promotionsfach erhoben.

[Promotionsthema/-bedingungen] Die Befragung enthielt Fragen zum subjektiv wahrgenommenen Promotionsfortschritt, zu Erfolgserwartung und Abbruchintentionen (inkl. Gründen), zur Anzahl der Betreuer*innen, zur Hauptbetreuungsperson (Wechsel, Affiliation, Geschlecht, formaler Betreuungsstatus, Position, Austauschhäufigkeit, Betreuungsstil, Zufriedenheit sowie Sympathieeinschätzung) und zum Besuch von Kursen und zu Lehrverpflichtungen. Zudem wurden ein aus mehreren Fragen bestehendes Instrument zur Messung der Lernumwelt in der Promotionsphase sowie Fragen zur Kooperation mit außeruniversitären Forschungseinrichtungen eingesetzt.

[Wissenschaftliche Aktivitäten] Darüber hinaus wurden die wissenschaftlichen Aktivitäten während der Promotion bzw. der letzten zwölf Monate detailliert erfasst mit Fragen zu Anzahl und Format wissenschaftlicher Publikationen und Beiträgen in Publikums- und Online-Medien, Anzahl von Konferenzbesuchen und -beiträgen, Gutachten sowie Anträgen auf Drittmittel und Patente (gestellt/bewilligt).

[Mobilitätserfahrungen] Des Weiteren wurden die promotionsbedingten oder zu anderen wissenschaftlichen Zwecken absolvierten Auslandsaufenthalte (Aufenthaltsart, Land, Datum des Beginns, Dauer) seit Beginn der Promotion anhand eines Tableaus erhoben. Diese Fragen wurden ausschließlich den Personen gestellt, die in der jeweiligen Welle angegeben haben, dass sie ihre Promotion abgeschlossen oder abgebrochen haben.

[Finanzierung und Erwerbstätigkeit] Zudem wurden Fragen zur Finanzierung des Lebensunterhalts Promovierender und Unterbrecher*innen erhoben. Dazu gehören die Finanzierungsquellen, die monatlichen Netto-Einkünfte sowie die selbsteingeschätzte Finanzierungssicherheit. Bei Promovierten und Abbrecher*innen wurde einleitend der Erwerbsstatus erhoben. Bei Nicht-Erwerbstätigen wurden sodann die Finanzierungsquellen und die Brutto-Einkünfte erfragt. Bei allen Erwerbstätigen (Promovierende, Unterbrecher*innen, Promovierte und Abbrecher*innen) folgten detailliertere Fragen zu u. a. Beschäftigungszeitraum, Befristung, Stellenumfang, Anzahl der Verträge, monatlichem Brutto-Erwerbseinkommen, Betriebsgröße, Sektor/Branche, Adäquanz, beruflicher Stellung, wissenschaftlicher Personalkategorie, Führungsverantwortung, beruflicher Zufriedenheit und Arbeitsort. Zudem wurden mehrere Fragen zum Thema Wissenschaftsbezug der Erwerbstätigkeit gestellt. Von Promovierenden wurden darüber hinaus Informationen zum Promotionsbezug der Erwerbstätigkeit (z. B. Promotion als arbeitsvertraglich festgelegtes Qualifikationsziel, Arbeitszeitanteile, thematischer Bezug) gesammelt. Zudem wurden Gründe für und gegen die Gründung einer selbstständigen beruflichen Existenz sowie der thematische Bezug der (künftigen) Selbstständigkeit zur Promotion erhoben.

[Persönliche Lebenssituation] Die Befragung enthielt verschiedene Fragen zur Partner- und Elternschaft, zum allgemeinen Gesundheitszustand sowie zum körperlichen und seelischen Befinden. Außerdem wurde der Wohnort erhoben.

[Einstellungen, Selbsteinschätzung und Ziele] Darüber hinaus wurden die allgemeine Lebenszufriedenheit, die Zufriedenheit mit der Vereinbarkeit von Arbeits- und Privatleben, die allgemeinen Selbstwirksamkeitserwartungen und Zielverfolgungsstrategien über das TEN/FLEX-Instrument (vgl. Brandtstädter & Renner, 1990) erfasst. Ferner wurden Instrumente entwickelt und eingesetzt, die 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, die Selbstständigkeitspläne und die spezifischen Selbstwirksamkeitserwartungen mit Blick auf eine Tätigkeit innerhalb oder außerhalb der Wissenschaft erfassen.

[Kontaktdaten] Die Befragung schloss mit einer Abfrage bzw. Aktualisierung der Kontaktdaten und bot die Möglichkeit zur Angabe von Kommentaren.

3 Grundgesamtheit und Stichprobenverfahren

[Grundgesamtheit] Nacaps orientiert sich grundsätzlich an der Promovierenden-Definition des Hochschulstatistikgesetzes sowie der 2017 eingeführten Promovierendenstatistik des Statistischen Bundesamts: „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). Zentrales Kriterium ist also die Registrierung zur Promotion an einer promotionsberechtigten Hochschule.

Für die Definition der Grundgesamtheit einer konkreten Nacaps-Kohorte erfolgt eine Eingrenzung auf die Personen, die sich in einem bestimmten Zeitraum zur Promotion registriert haben. Um die Anschlussfähigkeit mit der amtlichen Promovierendenstatistik zu gewährleisten, wird der dort geltende Stichtag 01. Dezember bei der Zeitraumfestlegung in Nacaps berücksichtigt (Statistisches Bundesamt, 2021a).

Da bereits in der Nacaps-Kohorte 2018 diejenigen Promovierenden befragt wurden, die zum Stichtag 01. Dezember 2018 zur Promotion registriert waren, wurden diese Personen bei der Definition der Grundgesamtheit der Nacaps-Kohorte 2020 ausgeschlossen. Demnach bestand die Grundgesamtheit bei Nacaps 2020 aus denjenigen Personen, die im Zeitraum zwischen dem 02. Dezember 2018 und dem 01. Dezember 2020 an einer promotionsberechtigten Hochschule in Deutschland zur Promotion registriert worden sind. Die Definition der Grundgesamtheit unterscheidet sich zwischen den Nacaps-Kohorten 2018 und 2020 also insofern, dass die Nacaps-Kohorte 2018 alle zum Stichtag Promovierenden umfasst, während bei der Nacaps-Kohorte 2020 nur diejenigen betrachtet werden, die nach dem Stichtag der Nacaps-Kohorte 2018 registriert worden sind (vgl. Briedis et al., 2025).

Nach Angaben des Statistischen Bundesamts belief sich die Anzahl der so definierten Promovierenden in Deutschland zum Stichtag 01. Dezember 2020 auf 58.005 Personen (Statistisches Bundesamt, 2020a, 2020b, 2021e).

[Vollerhebung] Im ersten Schritt wurden alle 156 promotionsberechtigten Hochschulen zu einer Teilnahme aufgefordert, anschließend wurden in einem zweiten Schritt alle (nach dem 01. Dezember 2018 neu registrierten) Promovierenden der 65 teilnehmenden Hochschulen zur Befragung eingeladen.³ Da weder eine Auswahl bzw. Ziehung von promotionsberechtigten Hochschulen noch von Promovierenden erfolgte, handelt es sich bei dem gewählten Verfahren um eine Vollerhebung.

Dennoch weist das gewählte Verfahren Ähnlichkeiten zu einer zweistufigen Stichprobenziehung auf. Demnach kann Unit-Nonresponse sowohl auf Ebene der Hochschulen (vergleichbar mit Primary Sampling Unit (PSU), primärer Stichprobeneinheit, bei zweistufigen Verfahren) als auch auf Ebene der Promovierenden (vergleichbar mit Secondary Sampling Unit (SSU), sekundärer Stichprobeneinheit, bei zweistufigen Verfahren) entstehen (Briedis et al., 2022). Die promotionsberechtigten Hochschulen fungieren daher als wichtige *Gatekeeper* für den Feldzugang.

³ 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. Das betrifft aber insgesamt nur wenige Einzelfälle.

4 Durchführung der Erhebung

4.1 Erstbefragung

[Kontaktaufnahme und Adresspflege] Von den insgesamt 65 teilnehmenden Hochschulen erfolgte an 62 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).

An den teilnehmenden Hochschulen war der Einladungs- und Remindersversand größtenteils an zentralen Stellen wie Graduiertenzentren und Evaluationsbüros angesiedelt, die den Promovierenden der Hochschulen 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, Serienbriefvorlagen für die Vorankündigung, das Einladungs- und die Reminderanschriften sowie eine Excel-Vorlage für die Verknüpfung der Anschriften mit den personalisierten Zugangscodes (Token). 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, die alle per E-Mail zugestellt wurden. 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 die Befragung noch nicht abgeschlossen hatten. Informationen zum Datenschutz wurden zu Beginn des Online-Fragebogens bei der Abfrage der Zustimmung zu den Datenschutzbestimmungen bereitgestellt. Alle Anschriften 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 22. Februar bis 18. April 2021. Vorankündigung, Einladungen und Reminder zur Befragungsteilnahme wurden an folgenden Tagen verschickt (vgl. Tabelle 1).

Tabelle 1: Versandtage der Erhebungsunterlagen der Erstbefragung

	Versandtage
Vorankündigung	08.02.2021
Einladung / Beginn der Feldphase	22.02.2021
1. Reminder	09.03.2021
2. Reminder	24.03.2021
3. Reminder	08.04.2021
Ende	18.04.2021

Diese Versandtage wurden von der großen Mehrheit der Hochschulen so umgesetzt. An einzelnen Hochschulen wurden die Einladungen und Erinnerungen einen Tag oder wenige Tage später als geplant verschickt. Bei fünf Hochschulen verzögerte sich der Feldstart. Feldstart war in diesen Fällen am 01.03., 03.03., 08.03., 09.03. bzw. 01.04. Bei den drei zuletzt gestarteten Hochschulen wurde aus diesem Grund auf eine dritte Erinnerung verzichtet.

[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 Fairphone 3+ (die aktuellste Version des Smartphones)
- 2 x BOSE „TV-Speaker“ (Soundbar)
- 10 x Wunschgutscheine im Wert von je 50 Euro

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, Beispieltex te für Newseinträge auf Webseiten und Bildmaterialien zur Verfügung gestellt. Zusätzlich wurden die Hochschulen gebeten, zwei Wochen 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 u. a. über die Projekt-Homepage (<https://www.nacaps.de/>) und den Twitter-Account (@Nacaps_Panel)⁴ insbesondere vor und während der Feldphase zielgruppenrelevante Beiträge geteilt. Manche Hochschulen nutzten auch ihre eigenen Social-Media-Kanäle, um die Befragung zu bewerben.

Des Weiteren wurden die Einladung und die drei Reminder planmäßig 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 an einem 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 in Nacaps 2020 neu teilnehmenden Hochschulen mit einem Willkommenspaket (Nacaps-Kaffeebecher, Kühlschrankmagnet, Post-It, Flyer, Poster usw.) begrüßt. Auf Anfrage war auch die Bereitstellung eines hochschuleigenen Datensatzes möglich, sofern dafür die datenschutzrechtlichen Voraussetzungen erfüllt waren und ein beidseitig unterschriebener Vertrag abgeschlossen wurde.

4.2 Folgebefragung

[Kontaktaufnahme und Adresspflege] Für die Folgebefragung konnten die Einladungsschreiben – im Unterschied zur Erstbefragung – direkt über die Nacaps-E-Mail-Adresse an die Befragten gesendet

⁴ Die Nutzung des Twitter-Accounts (später X) wurde im März 2025 eingestellt. Relevante Beiträge werden nun zusätzlich zur Projekt-Homepage über ein LinkedIn-Profil (@Nacaps-Panel) geteilt.

werden, da die Befragten in der Erstbefragung um Einwilligung zur Kontaktaufnahme (für weitere Befragungen) und die Angabe ihrer Adressen gebeten wurden.

Zudem fand auch eine Kontaktaktion vor dem Einladungsversand statt, die u. a. der Adresspflege diente. Im Verlauf der Befragungswelle wurden Widerrufsfälle, Wünsche zur Einstellung der Kontaktaufnahme, Adressänderungen und Aktualisierungen durch Abwesenheitsnotizen (im Fall einer neu angegebenen Adresse) oder Nichtzustellbarkeitsnachrichten (ggf. Rückgriff auf alternative Adresse) in die Adressdatenbank eingepflegt.

Ein Großteil der befragten Personen hat sich nach der Erstbefragung außerdem für den Nacaps-Newsletter angemeldet, der jährlich – meist zum Jahresende – an alle Abonnent*innen mit Informationen zum Projekt und relevanten Rezensionen, Veröffentlichungen oder Tagungsbeiträgen der Projektmitarbeiter*innen versandt wurde.

[Erhebungsunterlagen und Feldphase] Die Folgebefragung erfolgte wie die Erstbefragung online mit der DZHW-internen *Zofar*-Befragungssoftware. Der Versand der Einladungs- und Erinnerungsschreiben (Reminder) erfolgte per E-Mail mit personalisiertem Link zum Online-Fragebogen.

Der Ablauf der Feldphase der Folgebefragung ist datumsgenau in Tabelle 2 dargestellt. Sie startete Ende Februar 2022, vor Feldstart wurde ein postalisches Briefanschreiben als Vorankündigung an die Panelteilnehmer*innen mit gültiger Postadresse versandt. Die Feldphase endete nach etwa zwei Monaten gegen Ende April 2022.

Tabelle 2: Versandtage der Erhebungsunterlagen der Folgewellen

2. Welle	
Vorankündigung/Aktionen	Vorankündigung Postkarte, postalisch am 14.02.2022
Start	22.02.2022
1. Reminder	01.03.2022
2. Reminder	15.03.2022
3. Reminder	31.03.2022
Ende	20.04.2022

[Rücklaufsteigernde Maßnahmen] Eine Maßnahme zur Steigerung des Rücklaufs in der Folgebefragung ist die Incentivierung durch eine angekündigte Verlosung (technischer Sachpreise und/oder Gutscheine). Die Preise wurden in jedem Anschreiben genannt. Es wurden ein großer Sachpreis (Smartphone), zwei kleinere Sachpreise (Kopfhörer) und zehn Gutscheine mit einem Wert von jeweils 50 Euro verlost. Zudem sollten die zusätzlichen Kontaktaufnahmen (postalisches Vorankündigungsschreiben, Newsletterversand jährlich) mehr Personen zur Teilnahme motivieren.

5 Rücklauf

5.1 Erstbefragung

Die erste Befragungswelle von Nacaps 2020 erfolgte an 65 deutschen Hochschulen, wobei alle 156 promotionsberechtigten Hochschulen in Deutschland⁵ angefragt worden waren, an der Befragung teilzunehmen (vgl. Tabelle 3). Für eine detaillierte Betrachtung der teilnehmenden vs. nichtteilnehmenden Hochschulen vergleiche Kapitel 7.

Tabelle 3: Beteiligung der Hochschulen

Beteiligte	Anzahl
Hochschulen mit Promotionsrecht	156
Teilnehmende Hochschulen	65
Rücklaufquote auf Ebene der Hochschulen	41,70 %

Auf Hochschulebene fand eine Vollerhebung statt, d. h. die 65 kooperierenden Hochschulen haben die definierte Grundgesamtheit – alle Promovierenden, die im Zeitraum zwischen dem 02. Dezember 2018 und 01. Dezember 2020 registriert worden sind – zur Teilnahme an der Befragung eingeladen.⁶ Tabelle 4 gibt einen Überblick über den Rücklauf in der ersten Befragungswelle.⁷ Insgesamt wurden 44.381 Einladungen zur Befragung erfolgreich verschickt [1]. Mit 14.162 Fragebögen, bei denen die Befragten den Datenschutzbestimmungen zugestimmt haben [2], ergibt sich somit eine (bereinigte) Rücklaufquote von 31,91 Prozent. Insgesamt wurden 11.898 Fragebögen inhaltlich abgeschlossen [3], dies entspricht 26,81 Prozent Rücklauf. Zudem haben 4.995 Befragte am Ende der Befragung zugestimmt, dass sie bereit sind an Folgebefragungen teilzunehmen [4]. Dies entspricht einer Panelbereitschaft der Teilnehmenden aus Welle 1 von 35,27 Prozent.

Die Rücklaufquoten auf Ebene der Promovierenden unterscheiden sich zwischen den Nacaps-Partnerhochschulen zum Teil erheblich (zwischen 7,1 % und 83,0 %). Aus Datenschutzgründen werden differenzierte Analysen auf Hochschulebene nicht berichtet.

⁵ Auch wenn die Akquise der Hochschulen bereits im Sommer 2020 begonnen hat, ist sichergestellt, dass dies der Anzahl der promotionsberechtigten Hochschulen zum Stichtag 01.12.2020 entspricht.

⁶ 8 der 65 Hochschulen haben davon abweichend zusätzlich auch Promovierende zur Befragung eingeladen, deren Registrierungsdatum außerhalb dieses Zeitraums lag, d. h. die also nicht zur Grundgesamtheit gehörten. Das lag daran, dass sie entweder nicht differenzieren konnten, wann die Registrierung erfolgt ist (2 Hochschulen) oder sie explizit alle Promovierenden befragen wollten (6 Hochschulen). Die Daten dieser zusätzlich eingeladenen Personen sind nicht in den veröffentlichten Daten enthalten, jedoch auf Anfrage via userservice@dzhw.eu beim FDZ erhältlich.

⁷ In den Auswertungen berücksichtigt sind nur Personen mit eindeutiger Zugehörigkeit zur Grundgesamtheit, wobei die Zugehörigkeit zur Grundgesamtheit über das in der Befragung von den Promovierenden angegebene Zulassungsdatum (Monat/Jahr) bestimmt wurde. Promovierende mit dem Zulassungsdatum Dezember 2018 zählen dabei zur Grundgesamtheit von Nacaps 2020, auch wenn ihr Zulassungsdatum potenziell auch der 01.12.2018 sein könnte (vgl. dazu Definition der Grundgesamtheit in Kapitel 3). Nur wenn die Promovierenden in der Befragung keine Angabe zu ihrem Zulassungsdatum gemacht haben, wurde die Angabe der Hochschule zur Grundgesamtheitszugehörigkeit verwendet.

Tabelle 4: Brutto- und Nettorücklaufquote der ersten Befragungswelle

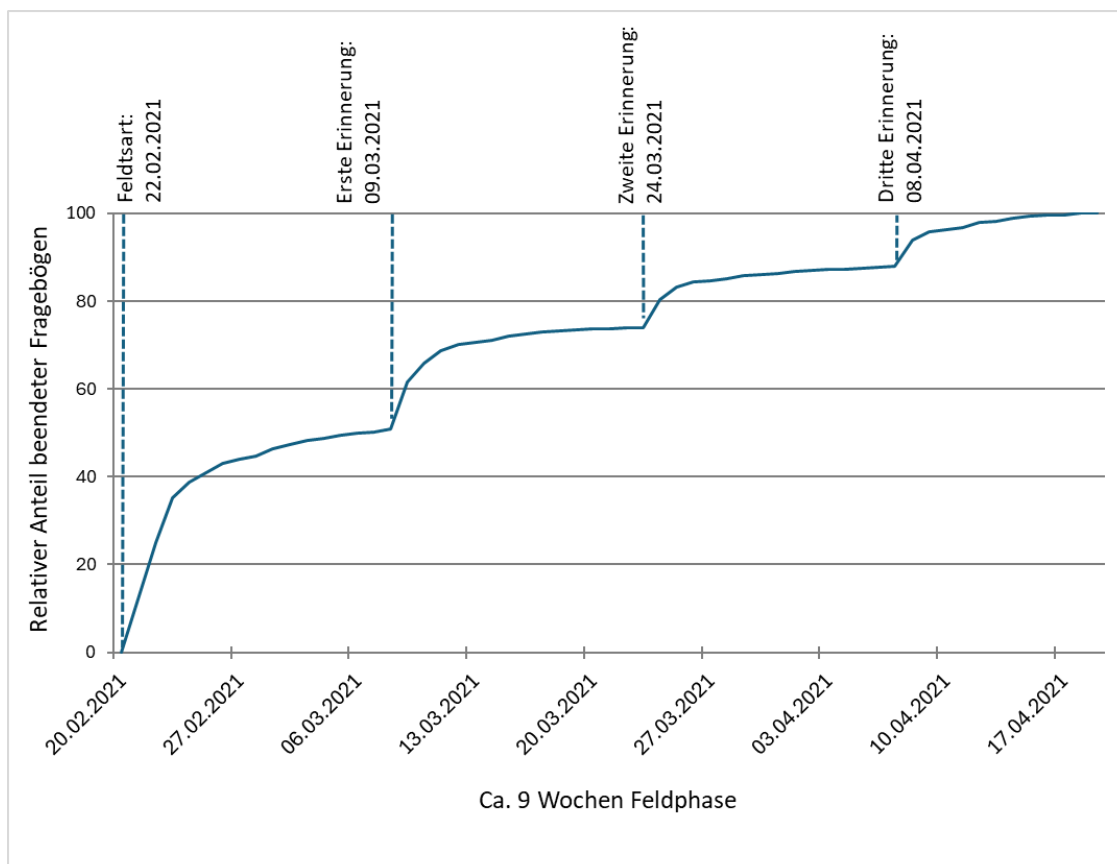
Welle 1	
Bereinigte versendete Einladungen [1]	44.381*
Datenschutz zugestimmt [2]	14.162
Fragebogen inhaltlich abgeschlossen (mind. bis Frage E16) [3]	11.898
Rücklaufquote 1 ([2]/[1])	31,91 %
Rücklaufquote 2 ([3]/[1])	26,81 %
Panelbereitschaft [4]	4.995

* Ausgeschlossen sind Personen, die nicht zur Grundgesamtheit gehören (vgl. Fußnote 6) und trotzdem teilgenommen haben. Personen, die nicht zur Grundgesamtheit gehören und nicht teilgenommen haben, können nicht ausgeschlossen werden, da die Zugehörigkeit zur Grundgesamtheit für diese Personen nicht bestimmt werden kann (vgl. Fußnote 7).

Es sei darauf hingewiesen, dass Fälle, die zum Zeitpunkt der Rücklaufberechnungen für Tabelle 4 bereits um komplette Löschung ihrer Daten gebeten hatten, nicht mehr in die Berechnungen eingegangen sind.

In Abbildung 2 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 rücklaufsteigernde Wirkung zeigen.

Abbildung 2: Relativer Anteil beendeter Fragebögen (in %) über den Verlauf der Feldzeit



5.2 Folgebefragung

Der Rücklauf der Folgebefragung wurde in der Tabelle zur Berechnung der Rücklaufquote der American Association for Public Opinion Research (AAPOR)⁸ erfasst.⁹ In Tabelle 5 sind die wesentlichen Rücklaufzahlen und die Rücklaufquote nach AAPOR RR1 (minimale Quote) und RR2 (inkl. teilweise ausgefüllter Fragebögen) dargestellt.

In Welle 2 wurden 4.799 Einladungen an alle Teilnehmer*innen aus Welle 1 versandt, die in Welle 1 ihre Einwilligung für weitere Befragungen sowie eine gültige E-Mail-Adresse angegeben¹⁰ und nicht in der Zwischenzeit ihre Teilnahme widerrufen hatten.

Für das weitere Panel gibt es keine generelle Ausfallregel, d. h. Befragte werden auch zur nächsten Folgebefragung eingeladen, wenn sie zwischenzeitlich an keiner der Wellen teilgenommen haben. Jedoch kommt es dennoch zu Ausfällen zwischen den Befragungswellen, da vereinzelt Personen ihre Teilnahme an der Studie widerrufen haben und somit nicht mehr kontaktiert wurden.

Es sei darauf hingewiesen, dass Fälle, die zum Zeitpunkt der Rücklaufberechnungen für Tabelle 5 bereits um komplette Löschung ihrer Daten gebeten hatten, nicht mehr in die Berechnungen eingegangen sind.

Tabelle 5: Brutto- und Nettorücklaufquote der Nacaps-Folgebefragungswellen (nach AAPOR 2016)

Daten	2. Welle
Versendete Einladungen [1]	4.799
Nicht zulässig (Dublette) [2]	60
Unzustellbar	45
Widerrufe (während der Feldphase)	2
Befragung geöffnet, aber nicht ausgefüllt [3]	108
Teilweise ausgefüllt (abgebrochen ab A01 und vor C01) [4]	155
Vollständig ausgefüllt (abgebrochen ab C01 bzw. nicht abgebrochen) [5]	2.836
Anzahl Fälle ([3]+[4]+[5])	3.099
davon verwertbare Fälle ([4]+[5])	2.991
Response Rate 1 (RR1) (minimale Quote) $((5)/((1)-[2]))$	59,84 %
Response Rate 2 (RR2) (inkl. teilweise ausgefüllter Fragebögen) $(([4]+[5])/((1)-[2]))$	63,11 %

⁸ <https://aapor.org/response-rates/> (zugegriffen am: 04.05.2026)

⁹ Für die Erstbefragung konnten keine Rücklaufzahlen nach AAPOR berechnet werden, da wesentliche Zahlen nur geschätzt werden konnten aufgrund fehlender Rückmeldung der Hochschulen und mangelnder Aufschlüsselung der Fälle nach Grundgesamtheit (Widerrufe, Nicht-Teilnahme, nicht zustellbare Einladungen).

¹⁰ Entsprechend konnten nicht alle Personen mit Panelbereitschaft (vgl. [4] in Tabelle 4) kontaktiert werden.

6 Datenaufbereitung

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. Der Rohdatensatz wurde auf einem geschützten Server gespeichert.

6.2 Datenprüfung und -bereinigung

[Prüfung von Wertebereichen numerischer offener Angaben während der Befragung] In der Erstbefragung wurden bei numerischen offenen Angaben innerhalb der Fragebogenprogrammierung keine Wertebereiche definiert, die nicht unter- bzw. überschritten werden sollten und es konnten auch Textangaben erfolgen. Die Felder waren lediglich mit einem Ausfüllhinweis versehen, dass nur Zahlen eingetragen werden sollen. In der Folgewelle wurde ebenfalls auf die Definition gültiger Wertebereiche verzichtet, jedoch Typen (numerisch oder string) festgelegt.

[Nachträgliche Bereinigung numerischer offener Angaben] Im Zuge der Datenaufbereitung der Erstbefragung wurde für jede Variable nachträglich überprüft, ob die erfasste Ausprägung in dem für diese Variable erwarteten Wertebereich lag. Haben die Befragten mehr als die in der Ausfüllanweisung geforderten Dezimalstellen angegeben, wurden die Werte nachträglich gerundet. Zudem erfolgte bei offenen Angaben, die eine Eingabe numerischer Werte erforderten, aber Textangaben enthielten (siehe vorheriger Abschnitt), eine Recodierung zu einem numerischen Wert, wenn eine eindeutige Zuordnung möglich war. Ansonsten wurde der Missingwert „-966 nicht bestimmbar“ (vgl. Kapitel 6.6) zugewiesen. Die bereinigten Angaben wurden jeweils in neu generierten Variablen mit dem Suffix „_g1“ gespeichert (vgl. Kapitel 6.3 und 6.5). In der Folgewelle erfolgte i. d. R. keine nachträgliche Bereinigung von Wertebereichen bei numerischen offenen Angaben. Unplausible oder unmögliche Angaben wurden beibehalten, damit von Nutzer*innenseite entschieden werden kann, wie damit umgegangen werden soll.¹¹ Eine Ausnahme besteht bei Variablen, bei denen laut Ausfüllanweisung nur ganze Zahlen bzw. eine Dezimalstelle angegeben werden sollten. In diesen Fällen wurden die abweichenden Werte teilweise nachträglich gerundet und die Originalangaben überschrieben.

[Filterprüfung] Weiterhin wurden Vollständigkeitsprüfungen hinsichtlich der Filterführung vorgenommen (Prüfung, dass keine Fragen fälschlicherweise übersprungen wurden); auf Filterverstöße (Prüfung, dass keine Fragen beantwortet wurden, die gemäß der Filterführung übersprungen werden sollten) wurde jedoch nicht getestet. 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 und die Entscheidung, welche Angabe für die Analysen verwendet werden soll, kann von den Datennutzer*innen selbst getroffen werden.¹² Eine Ausnahme besteht bei Fällen, bei denen zunächst Angaben bei den Gesundheitsfragen gemacht, der entsprechende Informed Consent durch Zurückspringen jedoch

¹¹ Darauf hingewiesen sei jedoch, dass die offen erfassten Informationen zu den Kindern (Anzahl und Geburtsjahre) für ihre Aggregation im Zuge der Anonymisierung (vgl. Kapitel 9) zum Teil zuvor plausibilisiert wurden (ohne Änderung der Originalvariablen).

¹² Eine Möglichkeit Filterverstöße für die Datennutzer*innen zu markieren, besteht in der Erstellung von Flag-Variablen. Zwar kamen bei einem Großteil der Variablen Filterverstöße vor, die gleichzeitig aber stets nur vereinzelte Fälle betreffen. Da diese zusätzlichen Flag-Variablen die Datensätze stark aufblähen würden, wurde nach Abwägung von Aufwand und Nutzen auf die Kennzeichnung von Filterverstößen mit Flag-Variablen verzichtet.

nachträglich zurückgezogen wurde. In diesen Fällen wurden die Originalangaben nachträglich zu „-989 filterbedingt fehlend“ (vgl. Kapitel 6.6) recodiert.

[Löschung von Fällen bzw. Angaben] Im Panelverlauf wurden einige Fälle bzw. Angaben aus dem Datensatz entfernt. Wenn Befragte nur Angaben aus einzelnen Wellen gelöscht haben wollten, wurden die Angaben für diese Wellen auf „-995 keine Teilnahme (Panel)“ (vgl. Kapitel 6.6) recodiert, aber die Fälle grundsätzlich im Datensatz belassen. Bei einem Wunsch nach vollständiger Löschung der Daten wurde dieser Fall komplett entfernt (vgl. Kapitel 5).

6.3 Generierung von Variablen

Im Rahmen der Datenaufbereitung wurden auch zusätzliche Variablen generiert (gekennzeichnet über das Suffix „_g#“ im Variablennamen, vgl. Kapitel 6.5). Dabei handelt es sich erstens um Variablen mit numerischen Codierungen von ursprünglich offenen Nennungen¹³, zweitens um bereinigte Variablen (vgl. Kapitel 6.2) sowie drittens um Variablen, die aus Datenschutz-/Anonymisierungsgründen erstellt wurden (vgl. Kapitel 9). Das Vorgehen bei der Generierung kann für jede in den Daten enthaltene generierte Variable dem Datensatzreport entnommen werden. Dabei wird teilweise Bezug auf bestimmte Listen genommen, die hier zu finden sind:

- Bei Variablen zur Hochschule:
 - Welle 1: Destatis-Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik, Promovierendenstatistik und Gasthörerstatistik WS 2019/2020 und SS 2020 (Statistisches Bundesamt, 2020c): https://www.statistik-nord.de/fileadmin/Dokumente/Erhebungen/Hochschulstatistiken/1_Studenten_2_Pr%C3%BCfungen/Schlüsselverzeichnis_Studenten_Pruefungs_Promovierenden_Gasthoerer_WS_2019.xlsx, Schlüssel 2.2 (Hochschulen nach Hochschularten und Bundesländern) und Schlüssel 2.3 (Hochschulen, numerisch) (zugegriffen am 29.04.2026)
 - Welle 2: Destatis-Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik, Promovierendenstatistik und Gasthörerstatistik WS 2020/2021 und SS 2021 (Statistisches Bundesamt, 2021b): https://www.statistik-nord.de/fileadmin/Dokumente/Erhebungen/Hochschulstatistiken/1_Studenten_2_Pr%C3%BCfungen/Schlüsselverzeichnis_WS_2020.xlsx, Schlüssel 2.2 (Hochschulen nach Hochschularten und Bundesländern) und Schlüssel 2.3 (Hochschulen, numerisch) (zugegriffen am 29.04.2026)
- Bei Variablen zum Studien- und Promotionsfach:
 - Welle 1: Destatis-Fächersystematik WS 2019/2020 (Statistisches Bundesamt, 2020d): https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00131465, S. 432-434 (zugegriffen am 29.04.2026)
 - Welle 2: Destatis-Fächersystematik WS 2020/2021 (Statistisches Bundesamt, 2021c): https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00135733, S. 434-436 (zugegriffen am 29.04.2026)
- Bei Variablen zu Land und Staatsangehörigkeit:
 - Welle 1: Destatis-Staats- und Gebietssystematik (Stand 01.01.2021) (Statistisches Bundesamt, 2021d): <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Staats-und-Gebietssystematik/Generische-Publikationen/2021-Staats-und-Gebietssystematik-pdf.pdf> (zugegriffen am 18.05.2026)
 - Welle 2: Destatis-Staats- und Gebietssystematik (Stand 01.01.2022) (Statistisches Bundesamt, 2022): <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Staats-und-Gebietssystematik/Generische-Publikationen/2022-Staats-und-Gebietssystematik-pdf.pdf> (zugegriffen am 18.05.2026)

¹³ Darauf hingewiesen sei in diesem Zusammenhang, dass halboffene Angaben grundsätzlich nicht codiert wurden und auch bei einigen weiteren offenen Angaben auf eine Codierung verzichtet wurde. Teilweise dienten sie vor allem dazu, mit Blick auf zukünftige Befragungen zu prüfen, welche Antwortkategorien ggf. bei einer geschlossenen Frage sinnvoll wären; teilweise waren sie ausschließlich Hilfsvariablen, die bei der Codierung anderer offener Angaben unterstützende Informationen liefern sollten. Für die Codierung der offenen Angaben zum Beruf der Eltern in Welle 1 (Frage E12) gab es bislang keine ausreichenden Kapazitäten.

- NEPS-recoding map of countries (Koberg, 2016b): https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/6-0-0/SC5_6-0-0_Anonymisation.pdf, S. 33-34 (zugegriffen am 11.05.2026)
- Bei Variablen zur Wahrung des Einkommens:
 - ISO 4217 Currency Codes: <https://www.iso.org/iso-4217-currency-codes.html> (zugegriffen am 16.04.2025)
 - Wechselkurs-/Wahrungsrechner der Europaischen Kommission: https://commission.europa.eu/funding-tenders/procedures-guidelines-tenders/information-contractors-and-beneficiaries/exchange-rate-infoeuro_de (zugegriffen am 16.04.2025)

6.4 Datenstruktur und Datenformat

Die Daten der ersten und zweiten Befragungswelle wurden zwar separat erhoben, jedoch in einem Datensatz zusammengefuhrt. Die Zuordnung der Falle erfolgte uber die im Rahmen der ersten Feldphase vergebenen Identifikationsnummern der Befragten. Fur jede befragte Person existiert eine Datenzeile (wide-Format). Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehorigen Fragen im Fragebogen. Generierte Variablen (vgl. Kapitel 6.3) wurden soweit moglich hinter der/den jeweiligen Ausgangsvariablen positioniert. Der Datensatz wird sowohl im Stata- als auch im SPSS-Format bereitgestellt.¹⁴

6.5 Vergabe von Variablennamen und Variablen-/Wertelabels

[Variablen- und Wertelabelvergabe] Fur Variablen- und Wertelabels wurden Kurzformen der Fragebogenformulierungen gewahlt. Die Variablen- und Wertelabel liegen auf Deutsch vor.

[Variablenbenennung] Fur die Variablenbenennung wurde ein Prafix-Stamm-Suffix-Schema verwendet. Das *Prafix* enthalt die Wellenkennung anhand der Buchstaben a (Welle 1) und b (Welle 2). Aus dem *Stamm* geht erstens der Themenbereich, dem die Variable zugeordnet ist, hervor (vgl. Tabelle 6 fur die verschiedenen Themenbereiche sowie das zugehorige Kurzel fur den Stamm des Variablennamens). 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 gehoren und dadurch die gleiche Themendifferenzierung und Nummerierung aufweisen (z. B. bei Itembatterien, Mehrfachnennungen oder Fragen, in denen geschlossene und offene Fragen kombiniert werden). Mit dem *Suffix* „_g#“ werden daruber hinaus generierte Variablen (vgl. Kapitel 6.3) gekennzeichnet.

¹⁴ Bei Nutzung von Stata/BE (maximale Variablenanzahl: 2.048) konnen Wellen-Subsets der Daten mit Hilfe der Wellenprafixe in den Variablennamen (vgl. Kapitel 6.5) geladen werden, z. B. die ID-Variable plus alle Variablen der ersten Welle uber: `use pid a* using "datasetname.dta"`

Tabelle 6: Themengebiete und Kürzel für Variablenamen in Nacaps 2020

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/sci- entific activities	Promotionsthema und -motive/Wis- senschaftliche Aktivitäten
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 Eigen- schaften
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

*In den veröffentlichten Daten sind Variablen mit diesem Stamm nicht mehr enthalten.

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 in Nacaps 2020 verwendeten Missingcodierungen sind in Tabelle 7 fett hervorgehoben. Sie lassen sich drei verschiedenen Gruppen zuordnen.¹⁵ In den ersten beiden Gruppen wird zwischen fehlenden Werten aufgrund von Nicht-Beantwortung von Fragen seitens der Befragten (Item-Nonresponse, Interviewabbruch oder Panel-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 9) für bestimmte Variablen vergeben wurde.

Schließlich sei darauf hingewiesen, dass in den aus dem Befragungssystem exportierten Rohdaten zunächst nur zwischen zwei Missingkategorien unterschieden wird: (1) Frage wurde gesehen, aber nicht beantwortet und (2) Frage wurde nicht gesehen. Fehlende Werte der ersten Kategorie werden im

¹⁵ Eine mögliche vierte Gruppe würde spezielle Missingcodierungen umfassen, die im Rahmen der Datenaufbereitung eines konkreten Datensatzes nur für einzelne Items vergeben wurden. Zudem gibt es auch eine fünfte Gruppe für „andere fehlende Werte“ wie Datenverlust.

Zuge der Datenaufbereitung zu dem FDZ-Missingwert -998 „keine Angabe“ bzw. zu -988 „trifft nicht zu“ codiert. Für die zweite Missingkategorie (Frage nicht gesehen) muss unterschieden werden, ob die Frage nicht gesehen wurde, weil die Person nicht in die Frage hineingefiltert wurde oder weil die Person die Befragung abgebrochen hat. Zunächst werden in der Datenaufbereitung alle Missings dieser Kategorie zu -989 „filterbedingt fehlend“ codiert. Im Anschluss werden die Interviewabbrüche codiert. Dafür wird auf Prozessdaten zurückgegriffen, die Informationen zur maximal erreichten sowie zur zuletzt aufgerufenen Seite im Fragebogen beinhalten. Durch diese nachträgliche, händische Codierung kann es vereinzelt zu Unstimmigkeiten¹⁶ kommen. Die Systematik der fehlenden Werte darf daher nicht als verlässliche, technische Gegebenheit verstanden werden.

Explizit hingewiesen sei außerdem auf die folgenden Sonderfälle bzw. Abweichungen von der Systematik:

- Bei den Variablen bdwr31a bis bdwr31d gibt es die Kategorie „trifft nicht zu (Antwortkategorie)“, die aktiv von den Befragten angeklickt werden konnte und der analog zur Fragebogenprogrammierung der Code 6 (nicht der Code -988) zugewiesen wurde.
- Bei den Variablen adtc07a und adtc07b, bdtc07a bis bdtc07d, bdtc27a und bdtc27b gibt es die Kategorie „trifft nicht zu“ die aktiv von den Befragten angeklickt werden konnte und der analog zur Fragebogenprogrammierung der Code 2 (nicht der Code -988) zugewiesen wurde. Auf den Zusatz „(Antwortkategorie)“ im Wertelabel wurde hier verzichtet.

¹⁶ Es kann vorkommen, dass Werte als -989 „filterbedingt fehlend“ codiert sind, die eigentlich ein Interviewabbruch sind, oder dass filterbedingt fehlende Werte fälschlicherweise als -996 „Interviewabbruch“ codiert sind. Es empfiehlt sich daher, für die Nachvollziehbarkeit der Filterführung immer auch den Fragebogen zu konsultieren und Interviewabbrüche nicht auszuwerten.

Tabelle 7: 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 Dieser Wert wird vergeben, wenn Angaben aufgrund unterschiedlicher Faktoren als nicht plausibel eingestuft werden.

^e Dieser Wert wird vergeben, wenn eine eindeutige Codierung nicht möglich ist, z. B. wenn eine offene Angabe nicht vercodet werden konnte.

7 Repräsentativität der Daten

Um einen Überblick über eine mögliche ungleiche Verteilung zwischen der realisierten Nacaps-Stichprobe und der Grundgesamtheit zu erhalten, wurde die Verteilung der Personen nach Fächergruppen zwischen der Nacaps-Stichprobe mit den Daten der Promovierendenstatistik des Statistischen Bundesamtes (Destatis) verglichen. Für die in der Promovierendenstatistik erfasste Grundgesamtheit liegt zwar eine öffentliche Statistik zum Stichtag 01. Dezember 2020 vor; diese umfasst alle Personen, deren Promotion zu diesem Zeitpunkt begonnen wurde, andauerte oder seit dem 01. Dezember des Vorjahres erfolgreich bzw. erfolglos abgeschlossen, abgebrochen oder unterbrochen wurde (Statistisches Bundesamt, 2021a). Die öffentliche Statistik weist jedoch nicht aus, in welchen Jahren die Promovierenden registriert wurden. Aus diesem Grund wurde für den Abgleich zwischen amtlicher Statistik und Nacaps eine Sonderauswertung beim Statistischen Bundesamt angefragt, in der nur die Promovierenden ausgewiesen wurden, die zwischen dem 02. Dezember 2018 und dem 01. Dezember 2020 registriert worden sind (Statistisches Bundesamt, 2021e). Anhand dieser Daten konnten die Verteilungen zwischen Nacaps-Stichprobe und der Grundgesamtheit verglichen werden (vgl. Tabelle 8).

Tabelle 8: Vergleich der Grundgesamtheit laut Promovierendenstatistik mit Nacaps 2020 nach Fächergruppen

Fächergruppen	Promovierendenstatistik	Nacaps 2020	
	in %	in %	Diff.
Geisteswissenschaften	7,4	8,6	1,2
Sport	0,4	0,5	0,1
Rechts-, Wirtschafts- und Sozialwissenschaften	15,3	18,4	3,1
Mathematik, Naturwissenschaften	25,8	31,3	5,5
Humanmedizin/Gesundheitswissenschaften	28,8	17,8	-11,0
Agrar-, Forst- und Ernährungswissenschaften, Veterinärmedizin	2,5	1,9	-0,6
Ingenieurwissenschaften	18,0	18,2	0,2
Kunst, Kunstwissenschaft	1,3	1,4	0,1
Außerhalb der Studienbereichsgliederung	0,5	2,0	1,5
	100,0	100,0	

Quellen: Sonderauswertung des Statistischen Bundesamtes (2021e) und Nacaps 2020

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 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 Frauen aus der Fächergruppe Rechts-, Wirtschafts- und Sozialwissenschaften leicht überrepräsentiert sind (vgl. Tabelle 9). In der Fächergruppe Mathematik/Naturwissenschaften sind dagegen

Männer und Frauen gleichermaßen überrepräsentiert. In der Fächergruppe Humanmedizin/Gesundheitswissenschaften sind wiederum sowohl Männer als auch Frauen davon betroffen, dass die Fächergruppe insgesamt unterrepräsentiert ist; 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 (50,5 % Männer, 49,0 % Frauen, 0,5 % divers) und den Informationen aus der amtlichen Statistik (52,0 % Männer und 48,0 % Frauen) eher gering.

Tabelle 9: Vergleich der Grundgesamtheit laut Promovierendenstatistik mit Nacaps 2020 nach Fächergruppen und Geschlecht

Fächergruppen	Promovierendenstatistik		Nacaps 2020	
	in % männlich	in % weiblich	in % männlich	in % weiblich
Geisteswissenschaften	6,2	8,7	7,6	10,1
Sport	0,4	0,4	0,6	0,4
Rechts-, Wirtschafts- und Sozialwissenschaften	14,7	15,9	16,5	21,1
Mathematik, Naturwissenschaften	27,6	23,8	33,2	29,8
Humanmedizin/Gesundheitswissenschaften	21,2	37,1	11,6	22,6
Agrar-, Forst- und Ernährungswissenschaften, Veterinärmedizin	1,9	3,3	1,4	2,4
Ingenieurwissenschaften	26,7	8,4	26,7	10
Kunst, Kunstwissenschaft	0,9	1,8	0,9	1,9
Außerhalb der Studienbereichsgliederung	0,4	0,6	1,5	1,8
	100,0	100,0	100,0	100,0

Quellen: Sonderauswertung des Statistischen Bundesamtes (2021e) und Nacaps 2020

Ergänzend wurde zudem die Verteilung der teilnehmenden und nichtteilnehmenden Hochschulen analysiert. Insgesamt haben sich 41,7 Prozent aller 156 promotionsberechtigten Hochschulen beteiligt (vgl. Kapitel 5.1). Hinsichtlich der regionalen Verteilung gibt es nur geringfügige Unterschiede zwischen den Regionen Nord, Ost, Süd und West. Hochschulen aus dem Süden sind mit einer Teilnahmequote von 30,2 Prozent etwas unter-, Hochschulen aus dem Norden dagegen etwas überrepräsentiert (51,6 %). Aus dem Westen haben 43,4 Prozent und aus dem Osten 44,8 Prozent der Hochschulen teilgenommen. Zudem sind größere Hochschulen überrepräsentiert, kleinere Hochschulen sind dagegen vergleichsweise selten vertreten. Gemessen an der Zahl der Professorinnen und Professoren als Indikator für die Größe einer Hochschule haben 19,2 Prozent der kleineren Hochschulen, 48,1 Prozent der mittelgroßen und 57,7 Prozent der großen Hochschulen an Nacaps teilgenommen.¹⁷ Von den (zumeist) kleineren Einrichtungen weisen private Hochschulen eine durchschnittliche Teilnahmequote auf (35,8 %), gering ist sie dagegen bei kirchlichen Hochschulen (11,1 %) sowie bei den Kunst- und Musikhochschulen (7,5 %) – was sich angesichts der Größe der Hochschulen jedoch nicht nennenswert auf die Fächerzusammensetzung des Samples auswirkt.

¹⁷ Zu berücksichtigen ist jedoch, dass angesichts der geringen Zahl an Hochschulen insgesamt bereits geringfügige Veränderungen der Teilnahmezahlen deutliche Auswirkungen auf die prozentuale Verteilung (insbesondere bei der regionalen Verteilung) haben. Aus diesem Grund wird u.a. darauf verzichtet, die Teilnahmequote auf Ebene der Bundesländer zu berichten.

8 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.

8.1 Vorgehen und Anwendungshinweise

[Ursachen für Verzerrungen von 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) führen zu einem verringerten Rücklauf und somit zu einer Differenz zwischen Brutto- und Nettostichprobe. Wenn diese Ausfallprozesse unsystematisch sind (*Missing Completely at Random*), können sie ignoriert werden.¹⁸ Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (*Missing at Random, Missing Not at Random*), der einer Modellierung bedarf.¹⁹

[Konzeptuelles Vorgehen] Da es sich bei Nacaps 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 den 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.

Weil das Studiendesign von Nacaps Ausfälle von Individuen in einzelnen Erhebungswellen zulässt, werden in den veröffentlichten Daten 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 im Sinne von sehr kleinen oder sehr großen Gewichten werden schließlich getrimmt.

Tabelle 10: Bereitgestellte Gewichte zu Nacaps 2020

Variablenname	Beschreibung
agewi	Gewichtungsfaktor 1. Welle (getrimmt)
bgeaus	unkalibriertes Ausfallgewicht 2. Welle (getrimmt)
bgewi	kalibriertes Ausfallgewicht 2. Welle (getrimmt)

¹⁸ Insofern die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

¹⁹ Siehe Rubin (1987) zu den unterschiedlichen Formen von Ausfallprozessen.

[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.²⁰ Das Gewicht `agewi` dient der Gewichtung der ersten Erhebungswelle, die Gewichte `bgeaus` und `bgewi` dienen der Gewichtung bei Analysen der Daten der zweiten Welle. Die bereitgestellten Gewichte sind nur für Querschnittsanalysen geeignet. Grundlegend ist zu beachten, dass Gewichte nur dann sinnvolle Korrekturgrößen darstellen, wenn das verwendete Analysemodell die zur Gewichtung herangezogenen Variablen enthält oder mit diesen in einem Zusammenhang steht. Aus diesem Grund müssen Gewichte immer mit Fokus auf die analysierte Fragestellung verwendet werden. Im Folgenden wird die Vorgehensweise bei der Erstellung der Gewichte für die Wellen 1 und 2 näher dargestellt.

8.2 Modellierung des Ausfallgewichts für Welle 2

[Propensity Score Matching] Für die zweite Welle wurde ein Ausfallgewicht mithilfe von *Propensity Scores* berechnet.²¹ Hierzu wurde anhand eines Logit-Schätzmodells die individuelle Teilnahmewahrscheinlichkeit einer Person für die entsprechende Welle vorhergesagt (RP). In einem explorativen Vorgehen wurden individuelle Merkmale identifiziert, die einen Beitrag zur Erklärung der Teilnahmewahrscheinlichkeit leisten. Um für jede Person die Teilnahmewahrscheinlichkeit schätzen zu können, wurden fehlende Werte bei allen Variablen als zusätzliche 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. Aus dem Modell konnte dann in einem zweiten Schritt die bedingte Teilnahmewahrscheinlichkeit abgeleitet werden, deren Kehrwert das Ausfallgewicht (`bgeaus`) für die zweite Welle darstellt: $w = 1/RP$.

[Schätzmodelle] Das Schätzmodell, das zur Erstellung des Ausfallgewichts genutzt wird (vgl. Anhang 11.2), beinhaltet als Prädiktoren die Variablen *Geschlecht*, *Geburtsjahr*, *Land der Hochschulzugangsberechtigung (HZB)*, *Promotionsstatus*, *Promotionsfach*, *Brutto-Einkommen* und *Tätigkeit in der Wissenschaft*.

8.3 Kalibrierung

[IPF-Raking-Verfahren] Das Ausfallgewicht (`bgeaus`) wurde mithilfe des IPF-Raking-Verfahrens (*Iterative Proportional Fitting*) an die Randverteilung bekannter Merkmale in der Grundgesamtheit angepasst, also kalibriert.²² 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 folgt dabei einem iterativen Prozess, bis keine Verbesserung mehr möglich ist. Da für die Erstbefragung kein Ausfallgewicht erstellt wurde, wurde für alle Teilnehmenden der Erstbefragung im ersten Schritt ein Ausfallgewicht mit dem Wert 1 erstellt, welches im zweiten Schritt anhand des IPF-Raking-Verfahrens kalibriert wurde. Nach Anwendung des IPF-Rakings liegt ein Gewichtungsfaktor für die erste Welle (`agewi`) und ein kalibriertes Ausfallgewicht für die zweite Welle (`bgewi`) vor.

[Merkmalsverteilung der Grundgesamtheit] Der Abgleich der Stichprobe mit der Grundgesamtheit erfolgte anhand der Merkmale Studienbereich des Promotionsfachs, Geschlecht und Region der Promotionshochschule (Ost/West). Die Informationen zur Grundgesamtheit wurden den Daten des Statistischen Bundesamtes entnommen (vgl. Kapitel 7). Personengruppen, deren Anteil nicht dem Anteil

²⁰ Siehe hierzu die Stata-Hilfe (Befehl: `help weights`).

²¹ Das Verfahren entspricht seiner Logik nach dem *Propensity Score Matching*, das auf Rosenbaum und Rubin (1983) zurückgeht (siehe auch Blumenstiel & Gummer, 2015).

²² Das IPF-Raking-Verfahren geht auf Deming und Stephan (1940) zurück (siehe auch Lomax & Norman, 2016) und wurde von Kolenikov (2014) im Stata-`.ado` `ipfraking` implementiert.

in der Grundgesamtheit entspricht, werden unter Verwendung des kalibrierten Ausfallgewichts so entsprechend hoch- bzw. runtergewichtet.

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

8.4 Trimmung

Um Extremwerte zu beseitigen, wurden alle Gewichte einer Trimmung nach Potter (1990) (vgl. auch Valliant et al., 2018, S. 388f.) unterzogen. Dem Verfahren liegt die Annahme zugrunde, dass die Gewichte einer Wahrscheinlichkeitsverteilung (Betaverteilung) folgen. All jene Gewichte, die über dem 99-Prozent-Quantil liegen, werden auf diese Grenze trunziert. Der Überschuss jenseits der Trunkierung wird im Folgenden unter den verbleibenden Gewichten verteilt.

8.5 Verteilung der Gewichte

Die Verteilungsmaße der bereitgestellten Gewichte lassen sich Tabelle 11 entnehmen:

Tabelle 11: Verteilung der Gewichte in Nacaps 2020

Gewicht	n	Min	Max	SD	Varianz
agewi	11.675	0,119	2,298	0,376	0,141
bgeaus	2.991	2,681	7,392	1,075	1,156
bgewi	2.946	0,163	3,204	0,596	0,355

9 Anonymisierung

[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten²³, 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 2017.²⁴ 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).

[Sicherstellung des Datenschutzes im FDZ-DZHW] Im FDZ-DZHW wird der Datenschutz der Befragten über rechtlich-organisatorische Maßnahmen, technische Maßnahmen und statistische Maßnahmen sichergestellt²⁵, die in unterschiedlicher Weise kombiniert werden können. Aus den Kombinationen ergeben sich vier potenzielle Zugangswege, die in Tabelle 12 im Überblick dargestellt sind.²⁶

Tabelle 12: Zugangswege im FDZ-DZHW

	Campus Use File (CUF): Download	Scientific Use File (SUF): Download	Scientific Use File (SUF): Remote-Desktop	Scientific Use File (SUF): On-Site
rechtlich-organisatorische Maßnahmen	sehr gering	gering	moderat	hoch
technische Maßnahmen	sehr gering	gering	moderat	hoch
statistische Maßnahmen	sehr hoch	hoch	moderat	gering

Je stärker der Datenzugang rechtlich-organisatorisch und technisch kontrolliert und beschränkt wird, desto geringer ist das Risiko einer De-Anonymisierung der Daten. In der Folge müssen die Daten umso

²³ „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 Abs. 1 DSGVO).

²⁴ 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.

²⁵ Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, S. 6ff.), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (LifBi) (vgl. Koberg, 2016a, S. 699ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner, Müller, Schmucker & Roß, 2012, S. 9f.) orientieren.

²⁶ Für genauere Informationen vgl. <https://www.fdz.dzhw.eu/de/datennutzung>

weniger mittels statistischer Maßnahmen anonymisiert, also um Informationen reduziert, werden – das heißt, desto größer bleibt ihr Analysepotenzial.

Für die Daten von Nacaps 2020 gibt es die Zugangswege „SUF: Remote-Desktop“ und „SUF: On-Site“.²⁷ Im Folgenden werden die konkreten durchgeführten statistischen Anonymisierungsmaßnahmen genauer erläutert.

[Statistische Anonymisierungsmaßnahmen] Im Rahmen der statistischen Anonymisierungsmaßnahmen wurden zunächst alle Informationen darauf geprüft, ob sich über sie Personen direkt identifizieren lassen. Diese direkten Identifikatoren, wie Namen, Adressen und E-Mail-Adressen, wurden bereits während der Feldphase in einem separaten Datensatz erfasst und sind somit per se nicht in den Daten enthalten. Die Original-Identifikationsnummer wurde entfernt und durch eine neue zufällig vergebene Identifikationsnummer ersetzt.

Darüber hinaus wurde der Empfehlung von Ebel und Meyermann gefolgt, 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).

Anschließend wurden die Quasi-Identifikatoren bestimmt, also Informationen, die in Kombination oder durch die Anspielung externer Informationen geeignet sind, eine Person indirekt zu identifizieren. Bei Nacaps 2020 sind dies beispielsweise bestimmte Informationen zu Promotionseckdaten (z. B. Promotionshochschule, -fach oder -note) bzw. zum Promotionsverlauf (z. B. Zeitpunkt des Abschlusses), zu Promotionsprogrammen und Promotionsförderung, zum/zur Betreuer*in der Promotion, zu Auslandsaufenthalten und wissenschaftlichen Aktivitäten (z. B. Publikationen und Anträge), zur Erwerbstätigkeit (z. B. Branche, Stellung, Arbeitsort), zu Bildungsetappen vor der Promotion, zur Soziodemographie (z. B. Geburtsjahr, Geburtsland/Staatsangehörigkeit oder Informationen zu den Eltern) und zur persönlichen Lebenssituation (z. B. Kinder, Hauptwohnsitz). Um eine eindeutige Zuordnung der Daten zu einer Person zu unterbinden, wurde für diese Schlüsselmerkmale – je nach Zugangsweg – festgelegt, ob sie ggf. nur in aggregierter Form oder gar nicht freigegeben werden dürfen.

Zuletzt wurde geprüft, ob in den Daten sensible Informationen, z. B. zur Gesundheit, sexuellen Orientierung und zu politischen Einstellungen, enthalten waren. Diese Informationen eignen sich zwar nicht zwingend zur Re-Identifikation von Individuen oder Institutionen, stellen jedoch im Falle einer De-Anonymisierung ein potenziell erhöhtes Schadensrisiko für die Befragten dar (vgl. Koberg, 2016a, S. 694) und sind daher besonders schützenswert (Art. 9 DSGVO, Erwägungsgrund 51 DSGVO). Bei Nacaps 2020 sind dies beispielsweise Informationen zur Gesundheit. Auch für die sensiblen Informationen wurde – je nach Zugangsweg – festgelegt, ob sie ggf. nur in aggregierter Form oder gar nicht freigegeben werden dürfen.

Dem Datensatzreport kann entnommen werden, welche Variablen in den bereitgestellten Daten enthalten sind und über welche(n) Zugangsweg(e) sie jeweils nutzbar sind.

²⁷ Für folgende Datenpaketversionen sind zusätzlich auch die Zugangswege „SUF: Download“ und „CUF: Download“ vorgesehen.

10 Literatur

- AAPOR (2016). Standard definitions. Final dispositions of case codes and outcome rates for surveys. <https://aapor.org/wp-content/uploads/2022/11/Standard-Definitions20169theditionfinal.pdf> (zugegriffen am: 19.05.2026).
- Blumenstiel, J. E. & Gummer, T. (2015). Prävention, Korrektur oder beides? In J. Schupp & C. Wolf (Hrsg.), *Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen* (S. 13–44). Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-10459-7_1.
- Brandtstädter, J. & Renner, G. (1990). Tenacious goal pursuit and flexible goal adjustment: Explication and age-related analysis of assimilative and accommodative strategies of coping. *Psychology and Aging*, 5(1), 58–67. <https://doi.org/10.1037/0882-7974.5.1.58>.
- Briedis, K., Lietz, A., Mühleck, K., Ruß, U., Scheller, P., Schiffer-Fiedler, I., Schwabe, U., Seifert, M., Teichmann, C., de Vogel, S., Birkelbach, R., Hoffstätter, U., Oestreich, T., Steins, H. & Weber, A. (2025). *Nacaps 2018. Daten- und Methodenbericht zur Datenpaketversion 3.0.0 der National Academics Panel Study 2018 (1.-6. Befragungswelle)*. Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). <https://doi.org/10.21249/DZHW:nac2018-dmr-de:3.0.0>.
- Briedis, K., Hoffstätter, U. & Schwabe, U. (2022). The German National Academics Panel Study (Nacaps): A new longitudinal survey on early researchers' careers and life courses. *Soziale Welt*, 73(4), 736–758. <https://doi.org/10.5771/0038-6073-2022-4-736>.
- Deming, W. E. & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 11(4), 427–444. <https://doi.org/10.1214/aoms/1177731829>.
- 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. <https://doi.org/10.25656/01:21970>.
- 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. https://doi.org/10.1007/978-3-658-11994-2_38.
- 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: 11.05.2026).
- Kolenikov, S. (2014). Calibrating survey data using iterative proportional fitting (Raking). *The Stata Journal: Promoting communications on statistics and Stata*, 14(1), 22–59. <https://doi.org/10.1177/1536867X1401400104>.
- Lane, J., Heus, P. & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. *Transactions on Data Privacy*, 1(1), 2–16.
- Lomax, N. & Norman, P. (2016). Estimating population attribute values in a table: “Get me started in” iterative proportional fitting. *The Professional Geographer*, 68(3), 451–461. <https://doi.org/10.1080/00330124.2015.1099449>.
- Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. In: *Proceedings of the Survey Research Methods Section*, S. 225–230.

- Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley & Sons.
- Statistisches Bundesamt (2020a). *Bildung und Kultur. Studierende an Hochschulen. Sommersemester 2019*. (Fachserie 11/4.1). Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2020b). *Bildung und Kultur. Studierende an Hochschulen. Wintersemester 2019/2020*. (Fachserie 11/4.1). Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2020c). *Schlüsselverzeichnisse für die Studenten- und Prüfungsstatistik, Promovierendenstatistik und Gasthörerstatistik WS 2019/2020 und SS 2020*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2020d). *Bildung und Kultur. Studierende an Hochschulen. Wintersemester 2019/2020*. (Fachserie 11/4.1). *Fächersystematik (Fächergruppen, Studienbereiche und Studienfächer)*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2021a). *Bildung und Kultur. Statistik der Promovierenden 2020*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2021b). *Schlüsselverzeichnisse für die Studenten- und Prüfungsstatistik, Promovierendenstatistik und Gasthörerstatistik WS 2020/2021 und SS 2021*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2021c). *Bildung und Kultur. Studierende an Hochschulen. Wintersemester 2020/2021*. (Fachserie 11/4.1). *Fächersystematik (Fächergruppen, Studienbereiche und Studienfächer)*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2021d). *Staats- und Gebietssystematik. Stand 01.01.2021*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2021e). *Sonderauswertung der Promovierendenstatistik 2020. Erstregistrierungen zwischen dem 02. Dezember 2018 und dem 01. Dezember 2020 [Unveröffentlichte Sonderauswertung]*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2022). *Staats- und Gebietssystematik. Stand 01.01.2022*. Wiesbaden: Statistisches Bundesamt.
- Valliant, R., Dever, J. A. & Kreuter, F. (2018). *Practical tools for designing and weighting survey samples*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-93632-1>.
- 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: 12.05.2026).

11 Anhang

11.1 Dokumentation der Herkunft sekundär genutzter Fragen

Aufgrund des Umfangs erfolgt die Dokumentation der Herkunft sekundär genutzter Fragen in einer separaten Datei. Diese ist unter dem folgenden Link abrufbar:

[https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-nac2020\\$-1.0.0/attachments/nac2020_QuestionOrigin.xlsx](https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-nac2020$-1.0.0/attachments/nac2020_QuestionOrigin.xlsx)

11.2 Logit-Regressionen zur Erstellung der Panelausfallgewichte

Tabelle 13: Logit-Regression zur Erstellung des Ausfallgewichts für Welle 2

		Odds Ratio	Std. Err	z	P> z
Geschlecht (ref. weiblich)	männlich	0,97	0,04	-0,79	0,428
	divers	1,05	0,35	0,16	0,871
	keine Angabe	0,25	0,19	-1,87	0,061
Geburtsjahr (ref. 1991-1993)	bis 1984	0,97	0,09	-0,31	0,757
	1985-1990	0,98	0,06	-0,38	0,706
	1994 und jünger	1,08	0,06	1,40	0,161
	keine Angabe	0,37	0,11	-3,22	0,001
Land der HZB (ref. Deutschland)	anderes Land	0,54	0,03	-10,17	0,000
	keine Angabe	0,09	0,06	-3,41	0,001
Promotionsstatus (ref. promovierend)	Dissertation eingereicht	0,74	0,11	-2,05	0,040
	abgeschlossen	1,40	0,24	1,94	0,053
	unterbrochen	0,97	0,27	-0,10	0,924
	abgebrochen	0,61	0,23	-1,30	0,192
	keine Angabe	<i>Omitted</i>			
Fächergruppe (ref. Geisteswissenschaften)	Sport	0,92	0,27	-0,30	0,766
	Rechts-, Wirtschafts- und Sozialwiss.	0,92	0,08	-1,01	0,314
	Mathematik, Naturwiss.	0,83	0,07	-2,20	0,028
	Humanmed./Gesundheitswiss.	0,65	0,06	-4,71	0,000
	Agrar-, Forst-, Ernährungswiss., Veterinärm.	0,66	0,12	-2,31	0,021
	Ingenieurwissenschaften	0,78	0,07	-2,69	0,007
	Kunst, Kunstwissenschaft	0,80	0,16	-1,12	0,262
	keine Angabe	0,73	0,14	-1,59	0,111
aktuelles/letztes Bruttoeinkommen (ref. kein Einkommen)	bis 1429 €	0,60	0,33	-0,93	0,350
	1430 € - 1810 €	0,58	0,32	-0,98	0,327
	1811 € - 2399 €	0,65	0,36	-0,79	0,431
	über 2400 €	0,58	0,32	-0,98	0,327
	keine Angabe	0,43	0,24	-1,50	0,134
Berufstätigkeit in der Wissenschaft (ref. nein)	ja	1,23	0,07	3,79	0,000
	keine Angabe	<i>Omitted</i>			
	Konstante	0,69	0,38	-0,67	0,506

N=12.487; Pseudo R²=0,043

Die Stichprobe für das Regressionsmodell enthält auch Fälle, die aufgrund von Widerruf oder des Wunsches nach Datenlöschung aus den veröffentlichten Daten entfernt wurden (vgl. Kapitel 5). Fälle mit „keine Angabe“ beim Promotionsstatus oder bei der Berufstätigkeit in der Wissenschaft wurden aus dem Modell ausgeschlossen, da für sie die Nicht-Teilnahme perfekt vorhergesagt wird.