

Ute Hoffstätter / Andreas Sarcletti

19. Sozialerhebung

Daten- und Methodenbericht zur Studierendenbefragung
2009

Version 1.0.0

Daten- und Methodenbericht

Juni 2017

Autor(inn)en:
Ute Hoffstätter
Dr. Andreas Sarcletti

Unter Mitarbeit von:
Andreas Daniel (Kapitel 7)

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
Dr. Bernhard Hartung

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	2
Tabellenverzeichnis	2
I Einleitung	1
II Übersicht zur 19. Sozialerhebung (2009)	2
III Datennutzungshinweise	4
1 Inhalt und Anlage der Studie	8
2 Erhebungsinstrument	11
2.1 Inhalte des Erhebungsinstrumentes	11
2.2 Pretest	12
3 Grundgesamtheit und Stichprobenverfahren	13
4 Durchführung der Erhebung	14
5 Rücklauf	15
6 Datenaufbereitung	16
6.1 Datenübertragung	16
6.2 Codierung offener Angaben	17
6.3 Datenprüfung und Datenbereinigung	18
6.4 Generierung von Variablen	20
6.5 Erstellung des Datensatzes	20
6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels	20
6.7 Codierung fehlender Werte	21
7 Gewichtung	22
7.1 Vorgehen und Anwendungshinweise	22
7.2 Gewichtung des Datensatzes	24
8 Anonymisierung	25
9 Literaturverzeichnis	30

Abbildungsverzeichnis

Abbildung 1: Datenzugangswege und Analysepotential.....	5
Abbildung 2: Rücklauf der 19. Sozialerhebung (2009) im Zeitverlauf.....	16
Abbildung 3: Datenzugangsweg, statistischer Anonymisierungsgrad und Analysepotential der Daten der 19. Sozialerhebung (2009)	26

Tabellenverzeichnis

Tabelle 1: Brutto- und Nettostichprobe sowie Rücklaufquote der 19. Sozialerhebung (2009)	15
Tabelle 2: Vercodete Merkmale und verwendete Codierlisten in der 19. Sozialerhebung (2009)	18
Tabelle 3: Themengebiete im Variablennamen der 19. Sozialerhebung (2009)	21
Tabelle 4: Systematik fehlender Werte des FDZ-DZHW	22
Tabelle 5: Bereitgestellte Gewichte zur 19. Sozialerhebung (2009)	23
Tabelle 6: Überblick zur Anonymisierung der 19. Sozialerhebung (2009)	28

I Einleitung

Die Sozialerhebung des Deutschen Studentenwerks (DSW) ist eine seit 1951 bestehende Untersuchungsreihe zur wirtschaftlichen und sozialen Situation der Studierenden in Deutschland.¹ Sie wird seit 1982² (10. Sozialerhebung) im Auftrag des DSW durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)³ durchgeführt. Seit der 6. Sozialerhebung (1967/1968) wird die Studie durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert. Die Sozialerhebung dient – in Ergänzung zur amtlichen Hochschulstatistik – dem nationalen Bildungsmonitoring.

Im Rahmen des vom BMBF geförderten Projektes zum Aufbau eines Forschungsdatenzentrums für Hochschul- und Wissenschaftsforschung am DZHW (FDZ-DZHW) werden die Daten einiger jüngerer Erhebungen der Reihe nachträglich zum Zweck der Nachnutzung aufbereitet und dokumentiert.⁴ Sie werden über verschiedene Zugangswege als *Scientific Use Files* (SUF) für die wissenschaftliche Sekundärnutzung bzw. als *Campus Use Files* (CUF) für Lehr- und Übungszwecke zur Verfügung gestellt. Neben den Datensätzen der Erhebungen werden auch Dokumentationsmaterialien zu den Datensätzen und zur Durchführung der Studien bereitgestellt.⁵

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation zur 19. Sozialerhebung (doi: 10.21249/DZHW:ssy19:1.0.0).⁶ Weitere Dokumentationsmaterialien zur Studie (Datensatzreport, Fragebogen, Filterführungsdiagramm etc.) können frei im Metadaten-System des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden.

Abschnitt II dieses Berichts stellt die Eckdaten der 19. Sozialerhebung im Überblick dar. Die zentralen Informationen zur Nutzung der Daten dieser Studie folgen in Abschnitt III. Kapitel 1 stellt Inhalt und Anlage der Sozialerhebung bis 2009⁷ im Allgemeinen und der 19. Sozialerhebung im Speziellen vor. Die weitere Gliederung des Berichts orientiert sich im Wesentlichen am Ablauf des Forschungsprozesses. In Kapitel 2 wird das Erhebungsinstrument beschrieben. In den Kapiteln 3 bis 6 werden der Erhebungsprozess (Stichprobenziehung, Erhebungsablauf, Rücklauf) und die Datenaufbereitung dargestellt. In den Kapiteln 7 und 8 folgt die Beschreibung der vorgenommenen Gewichtung und Anonymisierung.

¹ Für aktuelle Informationen zur Sozialerhebung siehe die Website des Projekts (<http://www.sozialerhebung.de>).

² Die 1. (1951) und 2. Sozialerhebung (1953) wurde jeweils vom Verband Deutscher Studentenwerke durch Gerhard Kath durchgeführt. Die 3. (1956) bis 9. Sozialerhebung (1979) wurde vom Deutschen Studentenwerk (DSW) ebenfalls durch Gerhard Kath (Geschäftsführer des Studentenwerks Frankfurt am Main) durchgeführt. Bis zur 5. Sozialerhebung (1963) wurde die Sozialerhebung vom Bundesministerium des Innern finanziell gefördert.

³ Das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung (DZHW, <http://www.dzhw.eu>) entstand im August 2013 durch eine Ausgründung aus der HIS Hochschul-Informationssystem GmbH. Im nachfolgenden Text wird stets der Begriff DZHW verwendet, auch wenn die Studie vor der Ausgründung 2013 durchgeführt wurde. Auf allen Originaldokumenten der 19. Sozialerhebung (Fragebogen, Flyer etc.) sowie in den Berichten zum Projekt ist entsprechend das HIS-Institut für Hochschulforschung (HIS-HF) als Akteur gekennzeichnet.

⁴ Da zum Erhebungszeitpunkt der Daten keine Datennachnutzung vorgesehen war, sind einige Informationen zur Erhebung nicht mit dem Fokus einer späteren Datennachnutzung dokumentiert worden und teilweise nicht mehr rekonstruierbar. An entsprechenden Stellen ist dies im Text angemerkt.

⁵ Informationen zu verfügbaren Datensätzen und Dokumentationen können im Metadatensuchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu/>) heruntergeladen werden.

⁶ Berücksichtigt wird hier nur die Stichprobe deutscher Studierender und Bildungsinländer(innen), die mittels eines Papierfragebogens befragt wurden. Die im Rahmen der 19. Sozialerhebung ebenfalls erhobenen Daten der Bildungsausländer(innen) sowie eine online erhobene Teilstichprobe werden hier nicht dokumentiert. Zu Details vgl. Kapitel 1.

⁷ Es werden nur die Sozialerhebungen bis zur hier dokumentierten 19. Sozialerhebung (2009) berücksichtigt. Die 20. Sozialerhebung (2012) und die 21. Sozialerhebung (2016) werden ebenfalls zur Datennachnutzung aufbereitet werden.

II Übersicht zur 19. Sozialerhebung (2009)

Studienreihe	Sozialerhebung des Deutschen Studentenwerks (DSW)
Erhebung	19. Sozialerhebung (2009)
Erhebende Institution	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)
Förderung	Bundesministerium für Bildung und Forschung (BMBF)
Projektmitarbeiter(innen) (Projektleitung)	Lars Borchert, <u>Wolfgang Isserstedt</u> , Maren Kandulla, Michael Leszczensky, Elke Middendorff
Themen	Angaben zum Studium Zeitbudget Finanzierung des Studiums und Erwerbstätigkeit Wohnsituation Beratungs- und Informationsbedarf Studienbezogene Auslandserfahrungen
Erhebungsdesign	Querschnittserhebung (Trend-Design)
Grundgesamtheit	Studierende mit deutscher Staatsangehörigkeit und Bildungs- inländer(innen), die im Sommersemester 2009 an einer staatli- chen oder staatlich anerkannten Hochschule in der Bundesrepublik Deutschland eingeschrieben waren (mit Aus- nahme der Verwaltungsfachhochschulen, Hochschulen des Fern- studiums und Universitäten der Bundeswehr)
Stichprobe	Einfach (disproportional) geschichtete Zufallsstichprobe
Erhebungsmethode	Standardisierte postalische Befragung
Erhebungszeitraum	1. Mai 2009 bis 31. Juli 2009
Auswertbare Fälle	n = 16.370
Rücklaufquote	32,0 %
Datenprodukte und Zugangswege	CUF: Download SUF: Download, Remote-Desktop, On-Site
DOI	10.21249/DZHW:ssy19:1.0.0
Weitere Informationen	https://fdz.dzhw.eu

Projektpublikationen*

Isserstedt, W., Middendorff, E., Kandulla, M., Borchert, L. & Leszczensky, M. (2010). *Die wirtschaftliche und soziale Lage der Studierenden in der Bundesrepublik Deutschland 2009. 19. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch HIS Hochschul-Informationssystem* (Bundesministerium für Bildung und Forschung (BMBF), Hrsg.) (Wissenschaft Ideen zünden!), Bonn.

Isserstedt, W. & Kandulla, M. (2010). *Internationalisierung des Studiums - Ausländische Studierende in Deutschland - Deutsche Studierende im Ausland. Ergebnisse der 19. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch HIS Hochschul-Informationssystem* (Bundesministerium für Bildung und Forschung (BMBF), Hrsg.) (Wissenschaft Ideen zünden!), Bonn.

* Alle Projektpublikationen werden auf der Website des Projektes (www.sozialerhebung.de) zum Download bereitgestellt.

Publikationen zum Datensatz (Auswahl)

Sarcelletti, A. (2015). Bachelor students' transition to postgraduate studies. Do students with and without migration background have different plans? *Beiträge zur Hochschulforschung*, 37 (2), 116-139.

Thomsen, S. & Haaren, F. von. (2014). *Did tuition fees in Germany constrain students' budgets? New evidence from a natural experiment* (IZA Discussion Paper no. 8623). Bonn: Forschungsinstitut zur Zukunft der Arbeit (IZA).

Kulik, M., Haaren, F. von & Thomsen, S. L. (2013). *Studiengebühren: Kein Effekt auf das Ausgabeverhalten der Studierenden* (niw info spezial Nr. 1). Hannover: Niedersächsisches Institut für Wirtschaftsforschung e. V. (NIW).

III Datennutzungshinweise

[Voraussetzungen der Datennutzung] Die Daten der 19. Sozialerhebung werden durch das FDZ-DZHW entsprechend dem Bundesdatenschutzgesetz (vgl. § 40 Abs. 1 und Abs. 2 BDSG) ausschließlich zur wissenschaftlichen Nutzung und anonymisiert bereitgestellt.⁸ Das FDZ bietet ein *Scientific Use File* (SUF) für die wissenschaftliche Sekundärnutzung und ein *Campus Use File* (CUF) für Lehr- und Übungszwecke an.

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

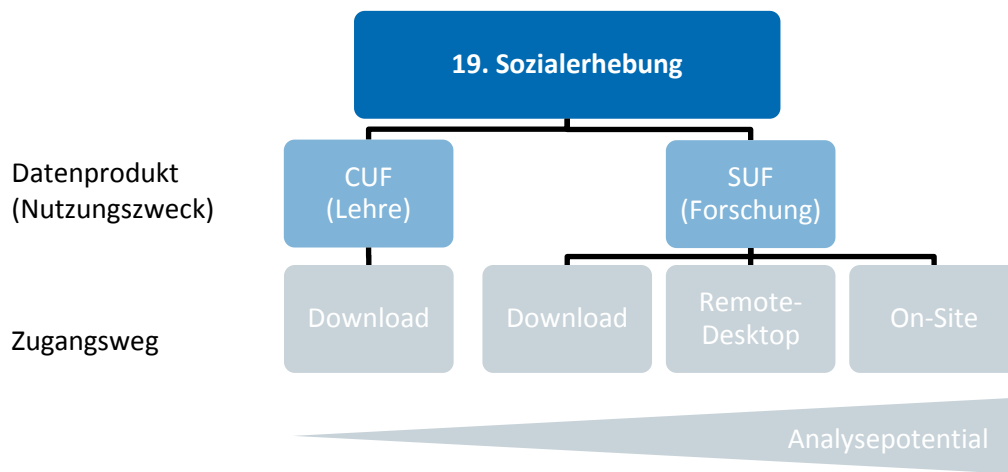
[Datenzugang] Das CUF der 19. Sozialerhebung kann nach Bereitstellung am lokalen Computer genutzt werden. Das SUF wird über drei Zugangswege angeboten, die hinsichtlich des Speicherortes, der Möglichkeit der eigenständigen Verknüpfung mit externen Daten und der Kontrollmöglichkeiten des FDZ unterschiedlich restriktiv sind.

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

⁸ Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, S. 6), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (LIfBi) (vgl. Koberg, 2016, 699ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner, Müller, Schmucker und Roß (2012, S. 9) orientieren. Das FDZ-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 können: Rechtlich-institutionelle Maßnahmen, informationelle Maßnahmen, technische Maßnahmen und statistische Maßnahmen.

Die bereitgestellten Daten weisen je nach Zugangsweg einen unterschiedlich hohen Informationsgehalt auf und unterscheiden sich damit hinsichtlich ihres Analysepotentials (vgl. Abbildung 1). Dabei gilt: Je stärker der Datenzugriff der Nutzer(innen) durch technische und organisatorische Maßnahmen kontrolliert wird, desto mehr Informationen können für die Datennutzer(innen) bereitgestellt werden.⁹ Mit diesem Vorgehen wird ein Höchstmaß an Nutzbarkeit und gleichzeitig ein bestmöglicher Schutz der bereitgestellten Daten sichergestellt.

Abbildung 1: Datenzugangswege und Analysepotential



[Datenprodukte] Über den *Digital Object Identifier* (DOI) 10.21249/DZHW:ssy19:1.0.0 ist eine Website mit zentralen Informationen zur Studie, weiteren Dokumentationsmaterialien sowie einer Übersicht der zur Verfügung stehenden Datenprodukte zur Studie erreichbar.

Die Daten der 19. Sozialerhebung werden über jeden im FDZ-DZHW angebotenen Zugangsweg – jeweils mit zugangswegspezifischem Analysepotential (vgl. Abbildung 1) – bereitgestellt. Das Download-CUF und das Download-SUF sind jeweils im Stata- und im SPSS-Format verfügbar. Für die Zugangswege Remote-Desktop und On-Site werden die Datensätze standardmäßig im Stata-Format bereitgestellt.

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

[Pflichten der Datennutzer(innen)] Die Datennutzer(innen) sind verpflichtet, folgende Regeln¹⁰ einzuhalten:

- **Wissenschaftliche Nutzung:** Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.

⁹ Zu den verschiedenen Anonymisierungsgraden und Analysepotentialen des CUF und der verschiedenen SUF-Varianten vgl. Kapitel 8.

¹⁰ Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

- **De-Anonymisierungsverbot:** Jeder Versuch der Re-Identifikation von Analyseeinheiten (z. B. Personen, Haushalten, Institutionen) ist verboten.
- **Gebot zur Mitteilung von Sicherheitslücken:** Falls Datennutzer(innen) Kenntnis von Sicherheitslücken hinsichtlich Datenschutz bzw. Datensicherheit erlangen, müssen diese dem FDZ-DZHW unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** SUF dürfen nur durch die Person genutzt werden, die den Datennutzungsvertrag abgeschlossen hat. CUF dürfen ausschließlich im Rahmen der angegebenen Lehrveranstaltung weitergegeben werden.
- **Löschungsgebot:** Download-SUF sind nach Ablauf der vereinbarten Nutzungsdauer (in der Regel drei Jahre) von jeglichen Rechnern, Servern und Datenträgern zu löschen. Ebenso müssen alle Sicherungskopien, modifizierten Datensätze (z. B. Arbeits-, Auszugs- oder Hilfsdateien) sowie Ausdrucke vernichtet werden.
- **Bereitstellung/Meldung von Publikationen:** Jede Art von Publikation, die aus der Arbeit mit Daten des FDZ-DZHW hervorgeht, muss dem FDZ im Voraus gemeldet und nach Veröffentlichung unverzüglich in elektronischer Form zur Verfügung gestellt werden. Informationen zu bereits vorhandenen Veröffentlichungen können dem Metadatensuchsystem¹¹ entnommen werden.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen laut folgenden Vorgaben zitiert werden.

[Zitation]

- **Verwendeter Datensatz:**
Isserstedt, W., Middendorff, E., Kandulla, M., Borchert, L. & Leszczensky, M. (2010). 19. Sozialerhebung. Aufbereitet durch Hoffstätter, U. & Sarcletti, A., doi: 10.21249/DZHW:ssy19:1.0.0, DATENSATZNAME¹², released 2017. Hannover: FDZ-DZHW.

¹¹ <https://metadata.fdz.dzhw.eu>

¹² An dieser Stelle bitte den genauen Namen der verwendeten Datensatzversion angeben, z. B. ssy19_d_1-0-0.dta für das Download-SUF der 19. Sozialerhebung (Version 1-0-0).

■ **Daten- und Methodenbericht:**

Hoffstätter, U. & Sarcletti, A. (2017). 19. Sozialerhebung. Daten- und Methodenbericht zur Studierendenbefragung 2009. Version 1.0.0. Hannover: FDZ-DZHW.

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

Diese Arbeit nutzt die Daten der 19. Sozialerhebung (2009). Die Daten wurden vom Forschungsdatenzentrum des Deutschen Zentrums für Hochschul- und Wissenschaftsforschung (FDZ-DZHW) unter der doi: 10.21249/DZHW:ssy19:1.0.0 veröffentlicht.¹³

¹³ Bei englischsprachigen Publikationen: This scientific work uses data of the 19th Social Survey (2009), conducted by the German Center for Higher Education Research and Science Studies (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung; DZHW). The data were published by the research data center of the DZHW, doi: 10.21249/DZHW:ssy19:1.0.0.

1 Inhalt und Anlage der Studie

[Studienreihe] Seit 1951¹⁴ führt das Deutsche Studentenwerk die Sozialerhebung durch, in der – meist in dreijährigem Abstand – Studierende standardisiert zu ihrem Studium befragt werden. Die Sozialerhebung ist damit die am längsten in Deutschland bestehende sozialwissenschaftliche Studie aus dem Bereich der Studierendenforschung. Der Fokus liegt stets auf der wirtschaftlichen und sozialen Lage der Studierenden. Die Studierenden der jeweiligen Stichprobe werden jeweils nur einmal befragt, das heißt es handelt sich um eine Querschnittsbefragung.

Die Grundgesamtheit besteht seit der 13. Sozialerhebung (1991) aus den Studierenden, die im Sommersemester des jeweiligen Erhebungsjahres an einer staatlichen oder staatlich anerkannten Hochschule in der Bundesrepublik Deutschland eingeschrieben sind. Studierende der Verwaltungsfachhochschulen, der Hochschulen des Fernstudiums und der Universitäten der Bundeswehr gehören nicht zur Grundgesamtheit.

Die ersten sechs Sozialerhebungen (1951 bis 1963) waren Vollerhebungen (ohne beurlaubte ausländische Studierende). Seit der 7. Sozialerhebung wird jeweils nur eine Stichprobe gezogen. Seit der 10. Sozialerhebung (1982) werden die Erhebungen durch das DZHW durchgeführt. Seitdem werden auch Bildungsausländer(innen), das heißt ausländische Studierende ohne in Deutschland erworbene Hochschulzugangsberechtigung, mit in die Grundgesamtheit einbezogen (Middendorff, 2016, S. 5). Zuvor war die Sozialerhebung auf Studierende mit deutscher Staatsangehörigkeit und Bildungsinländer(innen)¹⁵ beschränkt.¹⁶ Seit der 17. Sozialerhebung (2003) wird die Stichprobe der Bildungsausländer(innen) als gesonderte Stichprobe gezogen und die Bildungsausländer(innen) werden seitdem mit einem eigenen zweisprachigen (deutsch und englisch) Fragebogen befragt (Middendorff, 2016, S. 5).

[Stichprobe der 19. Sozialerhebung] Für die hier beschriebene 19. Sozialerhebung wurde eine Stichprobe aus der Gruppe der Studierenden mit deutscher Staatsangehörigkeit und Bildungsinländer(innen) sowie eine gesonderte Stichprobe aus der Gruppe der Bildungsausländer(innen) gezogen. Aufgrund der Unterschiedlichkeit von Erhebungsinstrument und untersuchter Studierendengruppe werden im hier dokumentierten SUF (bzw. CUF) nur die Daten der erstgenannten Stichprobe berücksichtigt.¹⁷

Bis einschließlich zur 17. Sozialerhebung (2003) wurde die Sozialerhebung ausschließlich mit einem Papierfragebogen durchgeführt. In der 18. und der hier vorliegenden 19. sowie der schon durchgeführten 20. Sozialerhebung wurde im Rahmen eines Methodentests ein Teil der Stichprobe dazu genutzt, die Studierenden online zu befragen¹⁸ (Middendorff, 2016, S. 12). Die Daten dieser Substichprobe der 19. Sozialerhebung dienten lediglich methodischen For-

¹⁴ Für einen detaillierten Überblick über die Geschichte der Sozialerhebung siehe Middendorff (2016).

¹⁵ Bildungsinländer(innen) sind Studierende mit ausländischer Staatsangehörigkeit, die ihre Hochschulzugangsberechtigung in Deutschland erworben haben.

¹⁶ In der 6. Sozialerhebung wurde zum ersten Mal versucht, ausländische Studierende mit in die Studie einzubeziehen. Aufgrund mangelnder Datenqualität wurden die Daten jedoch nicht ausgewertet und die Befragung ausländischer Studierender bis einschließlich der 9. Sozialerhebung (1979) wieder eingestellt (vgl. Middendorff (2016, S. 5)).

¹⁷ In die Stichprobe der Studierenden mit deutscher Staatsangehörigkeit und Bildungsinländer(innen) wurde jeder 27. im Sommersemester 2009 immatrikulierte Studierende einbezogen. Für die Stichprobe der Bildungsausländer(innen) wurde jeder 18. im Sommersemester 2009 immatrikulierte Studierende einbezogen.

¹⁸ Da die Methodentests verschiedene Vorteile der Umstellung auf eine Online-Befragung aufgezeigt haben (vgl. Middendorff (2016, S. 12)), wird die Sozialerhebung seit 2016 (21. Sozialerhebung) ausschließlich als Online-Befragung durchgeführt.

schungszwecken. Sie waren keine Grundlage für die Berichterstattung zum Projekt und sind nicht Bestandteil des hier dokumentierten SUF bzw. CUF. Somit sind in das hier dokumentierte SUF und CUF der 19. Sozialerhebung nur die Daten der Bildungsinländer(innen) und Studierenden mit deutscher Staatsangehörigkeit einbezogen, die mittels des Papierfragebogens befragt wurden.

[Analysepotential] Zum Kernbestand der Sozialerhebung gehören Fragen zum Hochschulzugang, zu Strukturmerkmalen des Studiums und Studienverlaufs, zur sozialen und wirtschaftlichen Lage (Studienfinanzierung, Lebenshaltungskosten, Erwerbstätigkeit, Wohnsituation), zu Themen des Tätigkeitsfeldes der Studentenwerke (*Student Services*) sowie zu soziodemographischen Merkmalen, die in Ergänzung zur amtlichen Statistik erhoben werden (Familienstand, Elternschaft, soziale Herkunft, Migrationshintergrund). Zu den Themenbereichen des Kernbestandes sind Zeitreihenanalysen möglich. Die Schwerpunktsetzung auf Themen aus dem Bereich der wirtschaftlichen und sozialen Lage der Studierenden geschieht aufgrund der Annahme, dass neben den Bedingungen des Hochschulzugangs und den Studienbedingungen auch die sozialen und wirtschaftlichen Rahmenbedingungen von großer Bedeutung für den Studienverlauf und -erfolg sind (Apolinarski et al., 2014, S. 1). Seit der 10. Sozialerhebung werden in jeder Befragung jeweils ein bis zwei ergänzende Schwerpunkte gesetzt. Hierbei handelt es sich um Themen wie z. B. Studium mit Kind, Studium mit Behinderung oder chronischer Krankheit. Die Daten der Sozialerhebung sind auch für Analysen zum Zeitaufwand für Studium und Erwerbstätigkeit geeignet. Auswertungen zum BAföG-Bezug, zu studentischer Erwerbstätigkeit und zu studienbezogenen Auslandsaufenthalten sind ebenfalls möglich. In eingeschränktem Maße sind mit den Daten auch Analysen zu Studierenden mit Kind(ern) sowie zu Studierenden mit Migrationshintergrund möglich.

[Einordnung ins Forschungsfeld] Bei der Sozialerhebung handelt es sich aus mehreren Gründen um eine für die Studierendenforschung einmalige Studie. Zunächst ist hier die seit 1951 bestehende Zeitreihe zu nennen, so dass mit den Daten der Sozialerhebung Fragestellungen im Bereich des Kernbestands der Studie über einen sehr langen Zeitraum hinweg untersucht werden können. Des Weiteren handelt es sich um die einzige bundesweite Studierendenstichprobe, die geeignet ist, differenzierte Aussagen über die Studierenden aller Semester, verschiedenster Abschlussarten, unterschiedlichster Hochschulzugangsvoraussetzungen und – mit Ausnahme der Verwaltungsfachhochschulen, der Hochschulen des Fernstudiums und der Universitäten der Bundeswehr – aller Hochschularten zu treffen.

[Spezifika der 19. Sozialerhebung] Zum Zeitpunkt der 19. Sozialerhebung war die Studienstrukturreform im Rahmen des Bologna-Prozesses so weit vorangeschritten, dass in der Stichprobe zum ersten Mal ein erheblicher Anteil an Studierenden der „neuen“ Studiengänge (Bachelor, Master) enthalten war (47 Prozent). Im Datensatz der 18. Sozialerhebung (2006) betrug der Anteil an Bachelor- und Masterstudierenden noch 13 Prozent (Isserstedt et al., 2010, S. 6). Mit den Daten der 19. Sozialerhebung sind (im Gegensatz zur 18. Sozialerhebung) aufgrund der hohen Zahl von Studierenden in Bachelorstudiengängen auch fundierte Aussagen zu den Absichten dieser Gruppe, ein Masterstudium aufzunehmen, möglich. Eine weitere Besonderheit der 19. Sozialerhebung ist, dass von den Studierenden im Datensatz etwa 60 Prozent an einer Hochschule studierten, die allgemeine Studiengebühren oder -beiträge er-

hob.¹⁹ 2009 wurden in mehreren westdeutschen Bundesländern Studiengebühren oder -beiträge erhoben. Die Einführung von Studiengebühren/-beiträgen in einigen Bundesländern und der Verzicht darauf in anderen Bundesländern kann als natürliches Experiment aufgefasst werden. Es lassen sich somit verschiedene Analysen zu den Auswirkungen der Einführung von Studiengebühren/ -beiträgen durchführen. Da die allgemeinen Studiengebühren später in allen Ländern wieder abgeschafft wurden, ist die 19. Sozialerhebung die einzige, in der ein nennenswerter Anteil an Studierenden von Studiengebühren betroffen war.

¹⁹ Von den Studierenden an Hochschulen mit allgemeinen Studiengebühren oder -beiträgen war etwa ein Fünftel von der Zahlung der Gebühren/Beiträge befreit (Quelle: Daten der 19. Sozialerhebung).

2 Erhebungsinstrument

Für die hier dokumentierte 19. Sozialerhebung wurde ein standardisierter Papierfragebogen in deutscher Sprache als Erhebungsinstrument eingesetzt.²⁰ Kapitel 2.1 stellt die zentralen Inhalte des Erhebungsinstrumentes vor. Kapitel 2.2 beschreibt den zur Prüfung und Verbesserung des Fragebogens durchgeführten Pretest.

2.1 Inhalte des Erhebungsinstrumentes

[Charakteristika der Studienreihe] Im Fokus der 19. Sozialerhebung steht, wie bei den übrigen Erhebungen der Studienreihe, die soziale und wirtschaftliche Situation der Studierenden. Die Kernthemen sind dabei Strukturmerkmale des Studiums, die finanzielle Situation der Studierenden, ihre soziale Herkunft sowie demographische Angaben. Diese Kernthemen werden in allen Erhebungen abgefragt. Daneben werden meist ein oder mehrere Schwerpunktthemen (z. B. Computer- und Internetnutzung, Beratungs- und Informationsbedarf) erfasst.

Die Fragen bezüglich der Strukturmerkmale des Studiums erfassen Informationen zum derzeitigen Studium (z. B. Studienfach, angestrebter Abschluss, Hochschul- und Fachsemester), zum Studienverlauf (z. B. bereits erworbene Hochschulabschlüsse, Studienunterbrechungen, Fach- und Hochschulwechsel; in 19. Sozialerhebung: Fragen 1 bis 11) sowie Angaben zum zeitlichen Aufwand für Studium und Erwerbstätigkeit (Frage 13). Zudem werden studienbezogene Auslandserfahrungen, deren Finanzierung, Hinderungsgründe bezüglich eines studienbezogenen Auslandsaufenthalts (Fragen 45.1 bis 47) sowie Sprachkenntnisse (Frage 48) erfasst. Diese Angaben werden durch persönliche Einschätzungen zum Stellenwert des Studiums und zur subjektiv empfundenen Belastung durch das Studium ergänzt (Fragen 12 und 14).

Im Fokus jeder Sozialerhebung steht die finanzielle Situation der Studierenden. Ausgaben für Lebenshaltung und Studiengebühren/-beiträge werden betragsgenau erfragt (Fragen 20 und 21). Ebenfalls betragsgenau erfasst werden die Einnahmen aus eigener Erwerbstätigkeit, familiärer/partnerschaftlicher Unterstützung, Stipendien und weiteren Quellen (Frage 19). Detaillierte Angaben zur Ausbildungsförderung durch das Bundesausbildungsförderungsgesetz (BAföG) werden zusätzlich erhoben (Fragen 23 bis 25). Items zur subjektiven Einschätzung der finanziellen Situation (Frage 22) und zu Motiven der eigenen Erwerbstätigkeit (Frage 26.4) ergänzen diese Messungen.

Darüber hinaus werden verschiedene soziodemographische (Fragen 27 bis 33, 40 und 41) bzw. bildungsbiographische Merkmale (Fragen 15 bis 18) sowie der familiäre Hintergrund der Studierenden (Fragen 34 bis 39) erhoben.

[Spezifika der 19. Sozialerhebung] Ein Schwerpunktthema, das auch in den Fragebögen anderer Erhebungen der Studienreihe behandelt wird, ist der Beratungs- und Informationsbedarf der Studierenden (Frage 44). Hier werden detailliert Informationsbedarfe sowie die Nutzung von Beratungsangeboten zu verschiedenen Themenbereichen (z. B. Studienfinanzierung, gesundheitliche Probleme) erhoben. Zudem wurde im Erhebungsinstrument der 19. Sozialerhebung verschiedenen Veränderungen der institutionellen Rahmenbedingungen Rechnung getragen. Vor dem Hintergrund des Bologna-Prozesses wurden Angaben zu einem aktuell

²⁰ Der Fragebogen kann von der Website des FDZ heruntergeladen werden. Ebenso steht dort ein Filterführungsdiagramm des Fragebogens zur Verfügung.

bestehenden oder angestrebten Masterstudium (Frage 5.1 bis 6) erfasst. Ein (weiteres) Beispiel für bildungspolitisch aktuelle Fragestellungen ist der Themenkomplex Studiengebühren (Frage 21.1 bis 21.3), der detaillierter als in den Fragebögen vorheriger Sozialerhebungen behandelt wurde. Im Unterschied zu vorangegangenen Sozialerhebungen wurde in der 19. Sozialerhebung die Staatsangehörigkeit der Eltern erfragt und damit die Erfassung eines möglichen Migrationshintergrundes verbessert (Frage 38). Bezüglich der Kernthemen bietet die 19. Sozialerhebung Vergleichsmöglichkeiten zu vorherigen Sozialerhebungen. Zudem bestehen – allerdings nur mit den Daten der 16. bis 18. Sozialerhebung – hinsichtlich des wiederkehrenden Schwerpunktthemas Beratungs- und Informationsbedarf in eingeschränktem Maße ebenfalls Vergleichsmöglichkeiten.

2.2 Pretest

[Ziel und Verfahren] Das Erhebungsinstrument wurde im Vorfeld der Erhebung durch einen kognitiven Pretest geprüft. Dabei sollte erstens für die bereits in Erhebungsinstrumenten vorheriger Sozialerhebungen eingesetzten Fragen und Antwortvorgaben geprüft werden, ob sie von den Studierenden 2009 gleich perzipiert werden würden wie von den Studierenden vorangegangener Sozialerhebungen. Zweitens sollte für die neu eingesetzten oder modifizierten Fragen die Verständlichkeit und Beantwortbarkeit getestet sowie ermittelt werden, wie Studierende bei der Beantwortung ausgewählter Fragen (z. B. Zeitbudget, Einnahmen, Ausgaben) vorgehen. Die Aufnahme der neuen Fragen zog außerdem Veränderungen am Aufbau und Layout des Fragebogens sowie der Befragungsdauer nach sich, die evaluiert werden sollten. Zur Prüfung dieser verschiedenen Aspekte kam ein sogenanntes „Pretestverfahren im Feld“ zum Einsatz. Dieses Verfahren hat zum Ziel, dass die am Pretest teilnehmenden Personen „unter möglichst ähnlichen Bedingungen untersucht [werden], wie sie später für die eigentliche Erhebung vorgesehen sind“ (Häder, 2015, S. 396).

[Probanden] Über öffentlich zugängliche Aushänge an der örtlichen Universität (Leibniz Universität Hannover) wurden 25 Studierende mit deutscher Staatsangehörigkeit bzw. Bildungsinländer(innen) zur Teilnahme rekrutiert.

[Durchführung] Der Pretest fand im Januar 2009 statt, also vier Monate vor dem Feldstart. Dabei wurden die Testpersonen gebeten, den für die Befragung vorgesehenen standardisierten Fragebogen zu beantworten. Im Anschluss wurden Gruppendiskussionen mit jeweils vier bis acht Proband(inn)en und Mitarbeiter(inne)n der Sozialerhebung anhand eines Gesprächsleitfadens durchgeführt. Der Fokus lag dabei auf den neu eingesetzten oder modifizierten Fragen, wobei die Studierenden die Möglichkeit hatten, sich zu jeder Frage zu äußern. Zusätzlich erfolgten Nachfragen zur Ausfülldauer, zu Inhalt und Länge des Fragebogens, zu Aufbau und Layout, zur Verständlichkeit der Fragen und Ausfüllanweisungen sowie zur Vollständigkeit der Antwortmöglichkeiten. Auf Grundlage der Pretestergebnisse wurden die Formulierungen verschiedener Fragetexte präzisiert sowie die Reihenfolgen von Items und Antwortkategorien in einzelnen Itembatterien und Mehrfachnennungen überarbeitet. Der grundsätzliche Aufbau, der Umfang und das Layout der Fragebögen wurden unverändert beibehalten. Die Proband(inn)en erhielten eine Aufwandsentschädigung in Höhe von 20 Euro.

3 Grundgesamtheit und Stichprobenverfahren

[Grundgesamtheit] Die Grundgesamtheit der 19. Sozialerhebung umfasst Studierende mit deutscher Staatsangehörigkeit und Bildungsinländer(innen)²¹, die im Sommersemester 2009 an einer staatlich anerkannten Hochschule in der Bundesrepublik Deutschland eingeschrieben waren. Studierende an Verwaltungsfachhochschulen, Hochschulen des Fernstudiums sowie Studierende an den Universitäten der Bundeswehr waren dabei ausgenommen, da sich ihre Studiensituation grundsätzlich von der Studiensituation anderer Studierender unterscheidet.²² Zudem konnten aus terminlichen Gründen keine Studierenden von Hochschulen, die erst im Jahre 2009 gegründet wurden, berücksichtigt werden (Isserstedt et al., 2010, 39).

[Stichprobenverfahren] Aufgrund nicht zugänglicher Listen von Studierenden war es ausgeschlossen, eine einfache Zufallsstichprobe zu ziehen. Es wurde eine Zufallsstichprobe auf Ebene aller Hochschulen konzipiert, an denen die zur Grundgesamtheit gehörenden Studierenden immatrikuliert waren, da diese nur über die Hochschulen erreicht werden konnten. An den relevanten Hochschulen wurde aus der Gruppe von Studierenden mit deutscher Staatsangehörigkeit und Bildungsinländer(inne)n eine Zufallsstichprobe mit der Auswahlwahrscheinlichkeit von 1/27 (d. h. jede(r) 27. im Sommersemester 2009 immatrikulierte Studierende) gezogen.²³ An einzelnen Hochschulen wurde auf Anfrage der Studentenwerke eine größere Stichprobe (d. h. mit einer erhöhten Auswahlwahrscheinlichkeit) gezogen, um auch Auswertungen auf Ebene der Studentenwerke zu ermöglichen (disproportionale Ziehung).²⁴ Die disproportionale Ziehung wird durch eine entsprechende Designgewichtung ausgeglichen (siehe Kapitel 7).

²¹ Bildungsinländer(innen) sind Studierende mit ausländischer Staatsangehörigkeit, die ihre Hochschulzugangsberechtigung in Deutschland erworben haben.

²² Studierende an Fernhochschulen sind sehr häufig berufstätig und absolvieren das Studium nebenbei. Studierende an Verwaltungsfachhochschulen sind in der Regel Beamtenanwärter(innen) („Beamte auf Widerruf“). Studierende an den Universitäten der Bundeswehr sind in der Regel Offiziersanwärter(innen).

²³ Zum Zweck der internen methodischen Weiterentwicklung der Studie erhielten 20 Prozent der Stichprobe statt des Papierfragebogens einen Link zur Online-Version der Befragung. Diese Fälle sind nicht Bestandteil des SUF und werden daher in dieser Dokumentation nicht berücksichtigt (vgl. Kapitel 1).

²⁴ Stichprobenerhöhungen wurden an folgenden 15 Hochschulen durchgeführt: Universität Gießen, Fachhochschule Gießen-Friedberg und Fachhochschule Fulda (jeweils jede(r) 24.), Universität Göttingen (jede(r) 17.), Technische Universität München und Universität Tübingen (jeweils jede(r) 16.), Universität Hannover (jede(r) 15.), Universität Marburg und Universität Bremen (jeweils jede(r) 13.), Universität Kassel (jede(r) 12.), Technische Universität Darmstadt und Universität Wuppertal (jeweils jede(r) 11.), Hochschule München (jede(r) 10.), Technische Universität Braunschweig (jede(r) 9.) sowie Hochschule Darmstadt (jede(r) 8.).

4 Durchführung der Erhebung

[Kontaktaufnahme] Alle Hochschulen mit Studierenden der Grundgesamtheit wurden schriftlich vom DZHW kontaktiert und um Teilnahme an der 19. Sozialerhebung gebeten. Im Anschreiben wurde auf die Relevanz der Studie und ihre Geschichte aufmerksam gemacht. Zudem teilte das DZHW den Hochschulen die Kriterien mit, anhand derer sie die Zielpersonen für die 19. Sozialerhebung identifizieren sollten.²⁵ 218 der 303 angeschriebenen Hochschulen (71,9 %) erklärten sich zur Teilnahme bereit.²⁶ Da die Hochschulen die Kontaktdaten ihrer Studierenden aus Datenschutzgründen nicht herausgeben dürfen, werden im Rahmen der Sozialerhebungen die Erhebungsunterlagen durch die Hochschulen versendet. Das DZHW ermittelte dafür jeweils die Anzahl der benötigten Erhebungsunterlagen²⁷ und sendete diese an die teilnahmebereiten Hochschulen, die aus ihrem Studierendenverzeichnis die Adressen der Studierenden zogen und sie jeweils zwei Mal auf Adresstiketten druckten. Ein Etikett wurde für den Umschlag mit den Erhebungsunterlagen benötigt, ein zweites für die Erinnerungspostkarte. Beides verschickten die Hochschulen jeweils in einem vorgegebenen Zeitfenster an die Zielpersonen.

[Erhebungsunterlagen] Die Erhebungsunterlagen bestanden pro zu befragender Person aus einem Anschreiben des Deutschen Studentenwerks als Auftraggeber der Studie, dem Fragebogen, einem Flyer mit allgemeinen Informationen zur Sozialerhebung sowie exemplarischen Ergebnissen vorheriger Erhebungen und einem an das DZHW adressierten Freiumschlag zur Rücksendung des ausgefüllten Fragebogens. Darüber hinaus erhielten alle Studierenden der gezogenen Stichprobe zeitversetzt eine Erinnerungspostkarte.

[Feldphase] Der Erhebungszeitraum erstreckte sich von Anfang Mai bis Ende Juli 2009. Aufgrund des angewendeten Kontaktverfahrens über die Hochschulen konnte das DZHW keinen direkten Einfluss auf den genauen Versandzeitpunkt der Erhebungsunterlagen nehmen. Die Erinnerungspostkarten wurden etwa zwei Wochen nach Feldstart an die Studierenden verschickt. Sie enthielten auch eine Dankesformel, denn sie wurden an alle Personen der Stichprobe verschickt – also auch diejenigen, die sich bereits an der Befragung beteiligt hatten –, da den Hochschulen unbekannt war, welche Personen bereits einen Fragebogen an das DZHW zurückgeschickt hatten.

[Rücklaufsteigernde Maßnahmen] Neben der Erinnerungspostkarte und dem zusammen mit dem Anschreiben verschickten Flyer wurde die 19. Sozialerhebung durch Presse- und Öffentlichkeitsarbeit seitens des Deutschen Studentenwerks und der einzelnen Studentenwerke beworben. Es wurde mit Pressemitteilungen und Plakatwerbung auf die bevorstehende Befragung hingewiesen. Darüber hinaus stand die Projektwebsite (<http://www.sozialerhebung.de>) als Informationsquelle zur Verfügung. Sie enthielt die Frage-

²⁵ In der Regel (vgl. Kapitel 3) jeder 27. im Sommersemester immatrikulierte Studierende mit deutscher Staatsangehörigkeit oder Bildungsinländer(in).

²⁶ Kleinere, Kunst- und Musik- sowie private Hochschulen gaben überproportional keine positive Rückmeldung. Bezogen auf die Studierendenzahl sind dennoch knapp 95 Prozent an den 218 (von 303) Hochschulen, die sich zur Teilnahme bereit erklärt haben, immatrikuliert.

²⁷ Dazu wurden die Zahlen der amtlichen Statistik zu Studierenden an den Hochschulen des vorangegangenen Semesters (Wintersemester 2008/09) mit der jeweiligen Auswahlwahrscheinlichkeit (vgl. Kapitel 3) kombiniert.

bögen, Fotos und Textbausteine für Pressemitteilungen sowie weiteres Informationsmaterial zur aktuellen Studie sowie zu vorangegangenen Erhebungen.²⁸

5 Rücklauf

[Rücklauf] Ca. 53.008 Studierende wurden von den Hochschulen angeschrieben. Von diesen ca. 53.008 Studierenden haben ca. 51.143 (Bruttostichprobe²⁹) den Fragebogen erhalten.³⁰ Insgesamt wurden 16.541 Fragebögen an das DZHW zurückgeschickt. Von diesen wurden 171 entfernt, da sie nicht zur Grundgesamtheit gehörten (stichprobenneutrale Ausfälle) oder nicht auswertbar waren (stichprobenrelevante Ausfälle) (vgl. Kapitel 6.3). Somit verblieben in der Nettostichprobe 16.370 Fälle. Im Hinblick auf die Bruttostichprobe liegt die Rücklaufquote damit bei 32 Prozent (vgl. Tabelle 1).

Tabelle 1: Brutto- und Nettostichprobe sowie Rücklaufquote der 19. Sozialerhebung (2009)

Bruttostichprobe	51.143
Nettostichprobe	16.370
Rücklaufquote	32,0 %

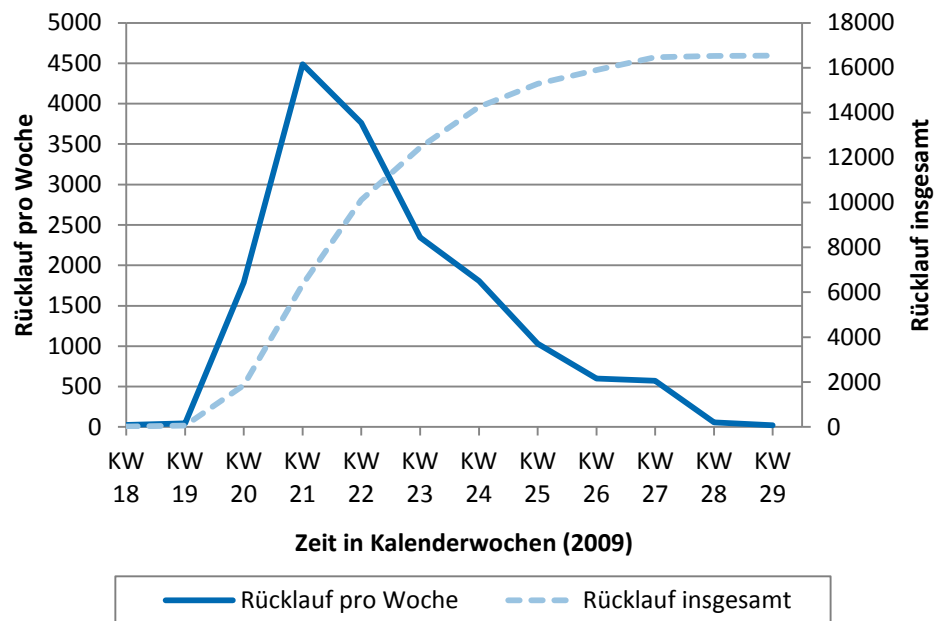
Abbildung 2: Rücklauf der 19. Sozialerhebung (2009) im Zeitverlauf stellt den Rücklauf der Fragebögen im Zeitverlauf dar. Insgesamt erstreckte sich die Rücklaufphase über zwölf Wochen. Ein Großteil der ausgefüllten Fragebögen hat das DZHW während der ersten Hälfte der Feldphase erreicht, in der auch die Erinnerungspostkarte verschickt wurde.

²⁸ Die Materialien zur 19. Sozialerhebung sind heute unter <http://www.sozialerhebung.de/sozialerhebung/archiv> (Stand: 01.06.2017) zu finden.

²⁹ Da eine genaue Differenzierung zwischen der (unbereinigten) Brutto-Ausgangsstichprobe und der bereinigten Bruttostichprobe nicht möglich ist, wird sich im Folgenden auf die Bezeichnung „Bruttostichprobe“ beschränkt.

³⁰ Einige Studierende konnten wegen fehlender oder veralteter Postadressen nicht erreicht werden. Es wurde um eine Rückmeldung seitens der Hochschulen über die Anzahl und Gründe nicht zustellbarer Sendungen an das DZHW gebeten. Dies erfolgte aber nur in seltenen Fällen. Von 45 Hochschulen wurde eine Anzahl unzustellbarer Postsendungen dokumentiert. Diese schwankten von 0 Prozent bis 14 Prozent mit einem Durchschnitt von 2,5 Prozent.

Abbildung 2: Rücklauf der 19. Sozialerhebung (2009) im Zeitverlauf



6 Datenaufbereitung

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

6.1 Datenübertragung

Zur weiteren Verarbeitung wurden die Angaben der Befragten aus den Papierfragebögen auf Basis eines Codeplans in ein computerlesbares Format übertragen. Zuvor wurden auf den Papierfragebögen numerische Codierungen für die offenen Angaben vermerkt (vgl. Kapitel 6.2) sowie manuelle Vorkorrekturen zur Erleichterung der Datenübertragung vorgenommen (vgl. Kapitel 6.3).

[Erstellung eines Codeplans] Auf Basis des Fragebogens der Befragung wurde ein Codeplan erstellt. Dabei wurde vermerkt, welcher Frage bzw. Teilfrage eine Variable zugeordnet ist, welchen Namen diese Variable trägt und welche numerischen Codierungen für die standardi-

sierten Antworten der Befragten verwendet werden. Um die Erfassungsreihenfolge festzulegen, wurden die Variablen zusätzlich nummeriert.³¹

[Datenerfassung] Für die Datenübertragung wurden der Codeplan, weitere Anweisungen zur Datenerfassung sowie die vorbereiteten Papierfragebögen an einen externen Dienstleister übergeben. Die Erfassung der Angaben erfolgte dort manuell durch Schreibkräfte. Die Qualität der Datenerfassung wurde stichprobenartig geprüft. Im Rahmen der Datenprüfung und -bereinigung (siehe Abschnitt 6.3) fielen Erfassungsfehler, die durch Überspringen von Variablen entstanden, meist durch Wertebereichsverletzungen in nachgelagerten Variablen auf und wurden entsprechend korrigiert.

6.2 Codierung offener Angaben

Vor der Datenerfassung erfolgte eine Codierung eines Teils der (halb-)offenen Angaben. Dabei wurden diesen anhand einer Codierliste numerische Codierungen zugeordnet. Je nach Variable wurden unterschiedliche Codierlisten verwendet. Es handelt sich um Klassifikationsschlüssel der amtlichen Statistik (z. B. Schlüsselverzeichnis der Studenten- und Prüfungsstatistik oder amtliche Codes für Staaten) oder um bereits in vorherigen Sozialerhebungen eingesetzte Schlüssel. Für wenige Variablen wurden neue Codierlisten auf Basis der in den Daten der 19. Sozialerhebung vorkommenden Nennungen entwickelt. Für einige halboffene Fragen wurden keine neuen Variablen mit numerischen Codierungen erstellt, sondern die Nennungen nur – sofern möglich – den vorhandenen (geschlossenen) Antwortkategorien zugeordnet. Einzelne offene Fragen wurden nicht vercodet. Dies war der Fall, wenn

- sie für die Auswertungen im Rahmen der 19. Sozialerhebung nicht benötigt wurden.
- lediglich relevant war, ob überhaupt eine Nennung erfolgte.
- die Fragen durch geschlossen erfasste Fragen annähernd ersetzt werden konnten.
- eine Vercodung aufgrund mangelnder personeller und zeitlicher Ressourcen nicht möglich war.³²

Die durch das Primärforschungsprojekt vorgenommenen Codierungsentscheidungen wurden unverändert beibehalten.

In Tabelle 2 sind die codierten Merkmale sowie die jeweils verwendete Codierliste dargestellt. Der Datensatz beinhaltet ausschließlich die codierten numerischen Variablen, die offenen Nennungen selbst sind nicht im Datensatz enthalten. Die Ausprägungen der einzelnen Variablen sind im Datensatzreport dokumentiert.

³¹ Die Daten wurden in einem einfachen spaltenorientierten Textformat ohne eine die Variablennamen enthaltende Kopfzeile erfasst. Der Codeplan musste daher festlegen, in welcher Reihenfolge die Daten zu erfassen sind, damit die zu einer Variable zugehörigen Daten in der richtigen Spalte eingetragen werden konnten.

³² Die offenen Angaben zu den im Folgenden genannten Fragen wurden elektronisch erfasst. Es wurde jedoch auf eine Vercodung verzichtet und nur in der Variable dokumentiert, ob eine offene Angabe erfolgt war (genannt) oder nicht (nicht genannt):

- Frage 21.3 (sonstige Finanzierungsquelle der Studiengebühren),
- Frage 25 (sonstiger Grund für kein BAföG(-Antrag)),
- Frage 26 (sonstige Tätigkeit, um Geld zu verdienen), Frage 44.1 und Frage 44.2 (sonstiges Thema Beratungsbedarf und -nutzung) sowie
- Frage 37 (Berufe der Eltern).

Alternativ zu den Berufen der Eltern können der Erwerbsstatus (Frage 34), der schulische (Frage 35) und berufliche (Frage 36) Abschluss sowie die berufliche Position (Frage 39) der Eltern für die Bestimmung der sozialen Herkunft verwendet werden. Bei den studienbezogenen Auslandsaufenthalten (Frage 45) wird die Art des studienbezogenen Auslandsaufenthalts (45.1) gar nicht dokumentiert und die sonstigen Finanzierungsquellen des studienbezogenen Auslandsaufenthalts (45.4) sind nur in der Form „genannt“ versus „nicht genannt“ verfügbar.

Tabelle 2: Vercodete Merkmale und verwendete Codierlisten in der 19. Sozialerhebung (2009)

Merkmal	Codierschema	Codierschema-ID
Studienfach	Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2008/2009 und SoSe 2009)	cl-destatis-studienfach-2009
Studienabschluss	Projekteigene Codierung	cl-dzhw-14
Hochschule	Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2008/2009 und SoSe 2009)	cl-destatis-hochschule-2009
Bundesland	Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2008/2009 und SoSe 2009)	cl-destatis-bundesland-1990
Ausland / Staatsangehörigkeit	Destatis Staatsangehörigkeits- und Gebietsschlüssel (Stand: 01.01.2009) (mit Anpassungen)	cl-destatis-ausland-2009
Befreiungsgrund von Studiengebühren	Projekteigene Codierung	cl-dzhw-21
Stipendiengeber	Projekteigene Codierung	cl-dzhw-10
Sonstige offene Abfragen	Zuordnung zu vorgegebenen Kategorien oder projekteigene Codierung	

6.3 Datenprüfung und Datenbereinigung

[Manuelle Vorkorrektur] Bereits vor der Übertragung der Daten wurden auf den Papierfragebögen eine manuelle Prüfung und gegebenenfalls eine Anpassung von Angaben der Befragten durchgeführt.³³ Dies sollte zum einen die Erfassbarkeit der Daten erleichtern. Dafür wurde die Form, aber nicht der Inhalt der gemachten Angaben, verändert. Beispielsweise wurden schwer lesbare Angaben oder Streichungen der Befragten verdeutlicht.

Zum anderen zielte die manuelle Prüfung darauf ab, schon vor der softwaregestützten Korrektur (siehe unten) erste Fehler oder Inkonsistenzen in den Angaben der Befragten zu bereinigen. Dafür wurden vorab verschiedene Konsistenzregeln aufgestellt, die auf den Papierfragebögen zu überprüfen waren. Bei festgestellten Inkonsistenzen wurden die betroffenen Angaben möglichst plausibilisiert und bei fehlenden Angaben – sofern möglich – aus anderen Nennungen im Fragebogen rekonstruiert. Zudem erfolgten inhaltliche Konsistenzprüfungen. Beispielsweise fand ein Abgleich der Angaben zu bereits erworbenen Hochschulabschlüssen (Frage 3) mit dem rechtlichen Status der besuchten Bildungseinrichtungen statt.³⁴ Zudem fand ein Abgleich des Zeitraums der Erwerbstätigkeit zwischen der zeitlichen Angabe selbst und der korrekten Einordnung der Tätigkeit in den entsprechenden Zeitraum (Frage 26) statt. Festgestellte Inkonsistenzen wurden – falls möglich – durch den Abgleich mit anderen Nennungen im Fragebogen aufgelöst oder andernfalls ein entsprechender Missingcode (vgl. Kapitel 6.7) vergeben.

³³ Die Zahl der vorgenommenen Korrekturen wurde nicht zentral, sondern nur auf den Papierfragebögen dokumentiert und ist daher nicht mehr systematisch rekonstruierbar.

³⁴ Wenn von den Befragten der Erwerb eines Hochschulabschlusses an einer Berufsakademie angegeben worden war, wurde dies nicht als Erwerb eines Hochschulabschlusses erfasst, da die Berufsakademien erst mit der Umwandlung in die Duale Hochschule Baden-Württemberg am 1. März 2009 den Status einer Hochschule erlangten.

[Softwaregestützte Korrektur] Im Anschluss an die Datenübertragung erfolgte eine umfassende Prüfung und Korrektur der Daten mit Hilfe einer DZHW-eigenen Software. Dabei sollten zum einen Fehler bei der vorherigen manuellen Vorkorrektur und Datenübertragung, zum anderen weitere inkonsistente Angaben der Befragten, die bei der Vorkorrektur nicht geprüft werden konnten, identifiziert werden.

Zu diesem Zweck wurden die erfassten Fragebogen-Daten in eine Datenbank eingelesen. Anschließend wurden anhand formaler Regeln gültige Wertebereiche und Antwortkombinationen definiert und geprüft. Folgende Typen von Prüfungen wurden vorgenommen:

- *Prüfung von Wertebereichen:* Es wurde geprüft, ob die erfasste Ausprägung einer Variablen in dem für diese Variable definierten Wertebereich lag.
- *Prüfung der Einhaltung der Filterführung:* Es wurde auf der einen Seite geprüft, ob aufgrund der im Fragebogen vorgesehenen Filterführung für die jeweilige befragte Person Angaben zu erwarten gewesen wären (Vollständigkeitsprüfung), und auf der anderen Seite, ob für die jeweilige Person keine Angaben hätten erfolgen dürfen (Filterverstöße).³⁵
- *Konsistenzprüfung:* Es wurde die Konsistenz der Angaben innerhalb des Fragebogens geprüft. Neben verschiedenen Merkmalskombinationen, die bereits bei der manuellen Vorkorrektur überprüft wurden, erfolgte hier eine noch detailliertere Definition von Merkmalskombinationen, die erfüllt sein mussten.³⁶

Insgesamt wurden mehrere hundert Konsistenzregeln getestet. Bei fehlenden, fehlerhaften oder unplausiblen Werten wurde zunächst mit Hilfe des Papierfragebogens geprüft, ob der entsprechende Wert falsch (bzw. nicht) übertragen worden war. Ansonsten wurde versucht, den korrekten Wert anhand anderer Angaben im Fragebogen zu erschließen. Im Zweifelsfall wurde ein spezifischer Missingcode vergeben (vgl. Kapitel 6.7). Fehlerkorrekturen wurden dokumentiert³⁷ und von mindestens einer weiteren Person geprüft.

[Löschung von Fällen] Ein Fall wurde gelöscht, wenn rund die Hälfte der Fragen oder zentrale Fragen (z. B. zum Studienfach) nicht beantwortet worden waren oder wenn zu viele Unstimmigkeiten erkennbar waren. Diese Fälle wurden als nicht auswertbar eingestuft. Darüber hinaus wurden Fälle identifiziert, die nicht zur Grundgesamtheit gehörten.³⁸ Diese wurden ebenfalls aus dem Datensatz entfernt. Insgesamt wurden 171 Fälle gelöscht.³⁹

³⁵ Die Eingangsfiler der einzelnen Fragen sind im Datensatzreport dokumentiert. Sie definieren für die jeweilige Variable, welche Befragtengruppe die zugehörige Frage beantworten sollte, sofern sie nicht an alle Personen gestellt wurde.

³⁶ Besonders detailliert wurden die Einnahmen und Ausgaben sowie das Zeitbudget geprüft.

³⁷ Die Dokumentation der Fehlerkorrekturen erfolgte handschriftlich auf den Papierfragebögen und ist daher nicht systematisch rekonstruierbar.

³⁸ Dies kam beispielsweise vor, wenn die Hochschulen versehentlich Personen angeschrieben hatten, die weder eine deutsche Staatsangehörigkeit aufweisen noch ihre Hochschulzugangsberechtigung in Deutschland erworben hatten. Diese Gruppe, die Bildungsausländer(innen), wurde mit einem gesonderten Fragebogen befragt und es wurde eine gesonderte Stichprobe gezogen (zu Details vgl. Kapitel 1). Die für deutsche Studierende oder studierende Bildungsinländer(innen) vorgesehenen Fragebögen wurden aussortiert, wenn sie versehentlich von Bildungsausländer(inne)n beantwortet worden waren.

³⁹ Dies entspricht 1,0 Prozent der Fälle. Es ist nicht rekonstruierbar, wie viele dieser Ausfälle stichprobenneutral und wie viele Ausfälle stichprobenrelevant sind.

6.4 Generierung von Variablen

Neben den Variablen, die die codierten Antworten der Befragten enthalten, beinhaltet der Datensatz der 19. Sozialerhebung auch generierte Variablen. Dabei handelt es sich zum einen um Variablen mit numerischen Codierungen von ursprünglich offenen Nennungen (vgl. Kapitel 6.2). Zum anderen wurden Variablen aus Datenschutzgründen verändert und es wurden im Forschungsfeld häufiger benötigte Variablen aus den Werten einer oder mehrerer Quellvariablen generiert (z. B. Aggregation der Studienfächer zu Studienbereichen und Fächergruppen oder Ableitung von Hochschulart und Hochschulort aus den Hochschulvariablen). Ein großer Teil der generierten Variablen wurde bereits durch das Primärforschungsprojekt erstellt. Der Variablenname einer generierten Variable ist im Datensatz durch das Suffix „_g#“ gekennzeichnet. Eine Übersicht aller für die 19. Sozialerhebung generierten Variablen sowie eine detaillierte Dokumentation der einzelnen Variablen mit Angabe ihrer jeweiligen Ausprägungen und Berechnungsvorschriften sind im Datensatzreport zu finden.

Generierte Variablen wurden im Datensatz – sofern möglich – nach der jeweiligen Ausgangsvariable positioniert. Wurde eine Variable aus verschiedenen Ausgangsvariablen generiert, wurde sie hinter jene Variable sortiert, welche ihr thematisch am ehesten entspricht. Falls eine thematische Zuordnung nicht möglich war, wurde die generierte Variable an das Ende des Datensatzes gestellt. Generierte Variablen, die aus Gründen der Anonymisierung bereitgestellt werden, wurden bis auf wenige Ausnahmen durch das FDZ erstellt.

6.5 Erstellung des Datensatzes

[Datenstruktur und Dateiformat] Der Datensatz enthält die Befragungsdaten sowie die zusätzlich generierten Variablen. 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 (vgl. Abschnitt III).

6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels

[Variablen- und Wertelabelvergabe] Für Variablen- und Wertelabels wurden Formulierungen des Fragebogens übernommen oder prägnante Kurzformen dieser Formulierungen gewählt. Dabei basieren die Variablenlabels in der Regel auf dem entsprechenden Fragetext. Grundlage für die Wertelabels sind je nach Fragetyp die Texte der Antwortoptionen bzw. eine Kombination der Texte von Frage und Antwortoption. Bei generierten Variablen, denen bestimmte Klassifikationen zugrunde liegen, wurden für die Wertelabels die Bezeichnungen der Schlüssel der Klassifikation wortgetreu übernommen. Die Variablen- und Wertelabels liegen auf Deutsch und auf Englisch vor. Im SPSS-Format existiert für jede Sprache ein eigener Datensatz. Im Stata-Format wurden zweisprachige Labels im gleichen Datensatz hinterlegt.

[Variablenbenennung] Für die Variablenbenennung wurde im FDZ-DZHW eine einheitliche Benennungssystematik erstellt. Mit Ausnahme der Identifikatorvariablen werden die Variablennamen nach einem Präfix-Stamm-Suffix-Schema gebildet. Die Variablennamen liefern Metainformationen zur entsprechenden Variable. Dies erleichtert die automatisierte Verarbeitung. Das Präfix der Variable enthält bei Längsschnittbefragungen mit mehr als einem Befragungszeitpunkt (Panel) die Wellenkennung anhand eines Buchstabens. Da es sich bei der

19. Sozialerhebung um eine Querschnittsbefragung handelt, entfällt hier das Präfix. Im Stamm geht der Themenbereich, dem die Variable zugeordnet ist, aus einem dreistelligen englischen Buchstabenkürzel hervor. Tabelle 3 stellt die verschiedenen Themenbereiche der 19. Sozialerhebung sowie das zugehörige Kürzel für den Stamm des Variablennamens dar. Das anhand eines Unterstrichs vom Stamm abgetrennte Suffix enthält verschiedene Zusatzinformationen, wie die Kenntlichmachung von generierten Variablen, Panelvariablen sowie verschiedenen Datenzugangswegen.

Detaillierte Informationen zur Variablenbenennung in der 19. Sozialerhebung befinden sich im Datensatzreport.

Tabelle 3: Themengebiete im Variablennamen der 19. Sozialerhebung (2009)

Themengebiete-Kürzel	Themengebiet (englisch)	Themengebiet (deutsch)
stu	studies	Studium
tim	time usage (studies/job)	Zeitaufwand (Studium/Erwerbstätigkeit)
ped	prior education	Vorbildung
fin	financing (of living during studies)	Finanzierung (des Lebensunterhalts während des Studiums)
baf	BAföG (German Federal Grant on Training and Education Promotion)	BAföG (Bundesausbildungsförderungsgesetz)
job	job (during studies)	Erwerbstätigkeit (im Studium)
dem	demographic information	Demographische Daten
par	parents	Eltern
liv	living (accommodation)	Wohnform
nut	nutrition	Ernährung
cin	counseling and information need	Beratungs- und Informationsbedarf
abr	experiences abroad	Auslandserfahrungen
lan	language (skills)	Sprach(kenntniss)e(n)

6.7 Codierung fehlender Werte

Zur Codierung fehlender Werte wurde im FDZ-DZHW eine übergreifende Systematik erstellt, um über verschiedene Datensätze des DZHW hinweg eine einheitliche Missingcodierung gewährleisten zu können. Fehlende Angaben werden dabei durch dreistellige negative Werte codiert. Tabelle 4 stellt die verwendete Missingsystematik dar. Die in der 19. Sozialerhebung verwendeten Missingcodierungen sind hervorgehoben.

Sie lassen sich vier verschiedenen Gruppen zuordnen. In den ersten beiden Gruppen wird zwischen fehlenden Werten aufgrund von Nicht-Beantwortung von Fragen seitens der Befragten (Nonresponse) und fehlenden Werten aufgrund der Filterführung bzw. für Befragte nicht relevanten Fragen unterschieden (nicht zutreffend). Die dritte Gruppe beinhaltet Missingcodierungen, die durch das Primärforschungsprojekt oder das FDZ im Zuge der Datenaufbereitung vergeben wurden (editierter fehlender Wert). Zu dieser Gruppe gehört auch die Codierung für fehlende Werte, die aufgrund von Anonymisierungsmaßnahmen (vgl. Kapitel 8) für bestimmte Variablen vergeben wurde. Die vierte Gruppe umfasst spezielle Missingcodierungen, die im Rahmen der Erstellung eines konkreten Datensatzes nur für einzelne Items

vergeben wurden. Im Datensatz der 19. Sozialerhebung gibt es keine solchen Item-spezifischen fehlenden Werte.

Tabelle 4: Systematik fehlender Werte des FDZ-DZHW

Wertebereich	Code	Wertelabel
-999 bis -990: Nonresponse	-999	weiß nicht
	-998	keine Angabe
	-997	keine Angabe (Antwortkategorie)
	-996	Interviewabbruch
	-995	keine Teilnahme (Panel)
	-994	verweigert
-989 bis -970: Nicht zutreffend	-989	filterbedingt fehlend
	-988	trifft nicht zu
	-987	designbedingt fehlend (Fragebogensplit)
	-986	designbedingt fehlend (Welle) ^a
	-985	designbedingt fehlend (Kohorte) ^b
-969 bis -950: Editierter fehlender Wert	-969	unbekannter fehlender Wert ^c
	-968	unplausibler Wert ^d
	-967	anonymisiert
	-966	nicht bestimmbar^e
	-965	ungültige Mehrfachnennung
-949 bis -930: Item-spezifische fehlende Werte^f	<i>(wurden in der 19. Sozialerhebung nicht vergeben)</i>	
-929 bis -920: Andere fehlende Werte	-929	Datenverlust

a Dieser Wert wird nur für Episoden-Datensätze im Long-Format vergeben.

b Dieser Wert wird nur in gepoolten Datensätzen vergeben.

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

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

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

f Die Ausprägungen dieser Missingkategorie sind definitionsgemäß für jeden Datensatz spezifisch.

7 Gewichtung

Die Gewichtung der Daten dient dem Ausgleich von Verzerrungen der Stichprobe aufgrund des Stichprobendesigns (vgl. Kapitel 3) und im Vergleich zur definierten Grundgesamtheit. Im Folgenden erfolgt zunächst eine allgemeine Einführung in die Vorgehensweise und eine Darstellung der erstellten Gewichte. Im Anschluss wird die Gewichtungsprozedur im Detail beschrieben.

7.1 Vorgehen und Anwendungshinweise

[Ursachen für die Verzerrungen der Stichproben] Maßgeblich für die Verzerrungen von Stichproben sind zwei Prozesse:

- **Designbedingte Verzerrung:** Disproportionalitäten werden bewusst erzeugt, um in bestimmten relevanten Subgruppen die Fallzahlen zu erhöhen (vgl. Kapitel 3).

- **Verzerrung durch Nonresponse:** Ausfallprozesse (z. B. Nichtteilnahmen, fehlende Erreichbarkeit, Verlust auf dem Postweg) führen zu einem verringerten Rücklauf und somit zu einer Differenz zwischen Brutto- und Nettostichprobe (vgl. Kapitel 5). Wenn diese Ausfallsprozesse unsystematisch sind (Missing Completely at Random), können sie ignoriert werden.⁴⁰ Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (Missing at Random, Missing Not at Random), der einer Modellierung bedarf.⁴¹

[Konzeptuelles Vorgehen] Im Zuge einer Gewichtungsprozedur sollten idealerweise zunächst designbedingte Disproportionalitäten ausgeglichen werden. Die hierfür benötigten *Designgewichte* ergeben sich bei zufallsgesteuerten Auswahlverfahren direkt aus dem Stichprobendesign. Im Anschluss sollte eine Adjustierung der Designgewichte mit Hilfe von *Nonresponsegewichten* im Quer- und Längsschnitt erfolgen, die auf der Grundlage von Informationen über Teilnehmer(innen) und Nichtteilnehmer(innen) auf Individualebene erzeugt werden. In einem letzten Schritt können die nonresponse-adjustierten Designgewichte anhand von Merkmalsverteilungen aus der Grundgesamtheit kalibriert werden (*Kalibrierung*).

Aufgrund des Stichprobendesigns (vgl. Kapitel 3) der 19. Sozialerhebung wird in einem ersten Schritt ein Designgewicht gebildet, um die ungleichen Inklusionswahrscheinlichkeiten auszugleichen. Da auf individueller Ebene keine Informationen zu Nichtteilnehmer(inne)n vorliegen, kann keine Nonresponse-Adjustierung des Designgewichts auf Individualebene erfolgen. Das Designgewicht wird in einem letzten Schritt anhand einer Merkmalsverteilung der Grundgesamtheit kalibriert. Da hier Informationen über Teilnehmer(innen) und Nichtteilnehmer(innen) auf aggregierter Ebene vorliegen, erfolgt hier zugleich eine Form der Nonresponse-Adjustierung.

Tabelle 5: Bereitgestellte Gewichte zur 19. Sozialerhebung (2009)

Variablenname	Beschreibung
wgt01	Gewicht für Analysen auf Bundesebene
wgt02	Gewicht für Analysen auf Regionalebene (Nord, Süd, Ost, West)
wgt03	Gewicht für Analysen auf der Ebene Ost- versus Westdeutschland

[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.⁴² Das Gewicht wgt01 ist für Auswertungen auf Bundesebene vorgesehen. Das Gewicht wgt02 bietet Gewichte für die Regionen Nord-, Süd-, Ost- und Westdeutschland. Das Gewicht wgt03 enthält Gewichte für die Regionen Ostdeutschland (neue Bundesländer inklusive Berlin) und Westdeutschland (alte Bundesländer ohne Berlin). Die Gewichte wgt02 und wgt03 dürfen jeweils nur in Auswertungen bezüglich *einer* der jeweiligen Regionen verwendet

⁴⁰ Insofern die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

⁴¹ Siehe grundlegend zu den unterschiedlichen Formen von Ausfallprozessen Rubin (1976).

⁴² Siehe hierzu die Stata-Hilfe (Befehl: *help weights*).

werden.⁴³ 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 des Gewichtes näher dargestellt.

7.2 Gewichtung des Datensatzes

[Designgewichtung] Wie in Kapitel 3 beschrieben, sind Studierende einiger Hochschulen aufgrund des Stichprobendesigns überrepräsentiert. Folglich weisen diese Individuen eine höhere Auswahlwahrscheinlichkeit in die Stichprobe zu gelangen auf. Dies kann durch die Verwendung der inversen Auswahlwahrscheinlichkeit (π_i) korrigiert werden. Das Designgewicht wgt_d_i ist somit:

$$wgt_d_i = \frac{1}{\pi_i}$$

Elemente, die mit höherer Wahrscheinlichkeit als andere in die Stichprobe eingehen, erhalten somit ein niedrigeres Gewicht und umgekehrt.

[Kalibrierung der Designgewichte] Wie beschrieben, war eine Nonresponse-Adjustierung der Designgewichte auf Individualebene nicht möglich. Es lagen jedoch folgende demographische Merkmale aus der Grundgesamtheit⁴⁴ vor, die zur Kalibrierung der Gewichte verwendet werden konnten: Region, Geschlecht, Fächergruppe, Hochschultyp, Personen mit deutscher Staatsangehörigkeit versus Bildungsinländer.⁴⁵ Die Kalibrierung erfolgte sowohl auf Bundesebene ($wgt01$) als auch gesondert für jede der vier Regionen ($wgt02$) sowie Ost- und Westdeutschland ($wgt03$). Die jeweiligen Regionalgewichte wurden in einer Variablen zusammengefasst ($wgt02$). Die Gewichte für Ost- und Westdeutschland wurden ebenfalls in einer Variablen zusammengefasst ($wgt03$).

Da die Merkmalsträger in der Grundgesamtheit ebenfalls Informationen über Nichtteilnehmer(innen) enthielten, erfolgte durch die Verwendung der Redressmentgewichte zusätzlich eine Art Nonresponse-Adjustierung im Hinblick auf die verwendeten Merkmale. Die Kalibrierung wurde mittels des Raking-Algorithmus⁴⁶ durchgeführt.

[Trimmung der Gewichte] Die initial berechneten Gewichte wiesen einen kleinen Teil an Gewichtungsfaktoren auf, die Ausreißer darstellten. Um diese zu beseitigen, wurden alle Gewichte einer Trimmung nach Potter (1990) (vgl. auch Valliant, Dever & Kreuter, 2013, S. 388f.) unterzogen. Dem Verfahren liegt die Annahme zugrunde, dass die Gewichte einer Wahrscheinlichkeitsverteilung (Beta-Verteilung) folgen. All jene Gewichte, die über dem 99-Prozent-

⁴³ Die Regionen können über die Variable `stu16_g6r` (Nord-Süd-Ost-West-Aggregation für Remote-Desktop- und On-Site-Datensatz) bzw. `stu16_g5` (Ost-West-Aggregation für Download-Datensatz) selektiert werden (z. B. mittels des Stata-Befehls `bysort`).

⁴⁴ Alle Informationen, die zur Kalibrierung der Designgewichte verwendet wurden, leiten sich aus Daten des vergangenen Semesters (WiSe 2008/2009) des Statistischen Bundesamtes ab, da die aktuelle Statistik zum Zeitpunkt der Gewichtung noch nicht vorlag.

⁴⁵ Die Gewichtung wurde entlang folgender Ausprägungen durchgeführt: Geschlecht: weiblich versus männlich; Region: Nord (Bremen, Hamburg, Niedersachsen, Schleswig-Holstein), Süd (Baden-Württemberg, Bayern), Ost (Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt, Thüringen), West (Hessen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland); Hochschultyp: Universität versus Fachhochschule (inklusive Pädagogische Hochschulen, Theologische Hochschulen, Kunst- und Musikhochschulen); Fächergruppe: entsprechend Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2008/2009 und SoSe 2009).

⁴⁶ Raking ist auch unter dem Begriff des iterative proportional fitting (ipf) bekannt (Kolenikov (2014)).

Quantil liegen, werden auf diese Grenze trunziert. Der Überschuss jenseits der Trunkierung wird im Folgenden unter den verbleibenden Gewichten verteilt.

8 Anonymisierung

[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten, die in freiwilligen Befragungen durch das DZHW erhoben werden, gilt das Bundesdatenschutzgesetz (BDSG).⁴⁷ Danach dürfen personenbezogene Daten, die im Rahmen wissenschaftlicher Forschung erhoben worden sind, ausschließlich zum Zweck wissenschaftlicher Forschung verarbeitet oder genutzt werden (vgl. § 40 Abs. 1 BDSG). Darüber hinaus sind personenbezogene Daten im Forschungskontext zum Schutz der Befragten zu anonymisieren (vgl. § 40 Abs. 2 BDSG). Im BDSG wird der Vorgang der Anonymisierung definiert als „das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können.“ (§ 3 Abs. 6 BDSG)⁴⁸ Das heißt, für die Weitergabe von Daten aus wissenschaftlichen Forschungsprojekten an Dritte sind die Daten entweder absolut zu anonymisieren, so dass kein Bezug zur Person mehr hergestellt werden kann, oder mindestens faktisch zu anonymisieren, so dass die Herstellung eines Personenbezugs mit einem unverhältnismäßig hohen Aufwand verbunden und daher die Wahrscheinlichkeit der Re-Identifikation einer Person minimal ist.

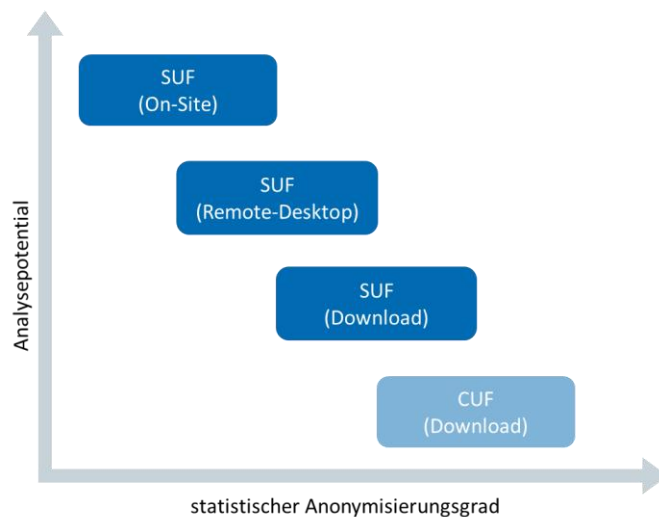
[Datenzugang, Anonymisierungsgrad und Analysepotential] Das FDZ-DZHW stellt für die 19. Sozialerhebung (2009) ein faktisch anonymisiertes SUF für die wissenschaftliche Sekundärnutzung und ein absolut anonymisiertes CUF für Lehr- und Übungszwecke zur Verfügung. Die Anonymität der Befragten wird dabei über eine Kombination aus statistischen Maßnahmen und technischen Zugriffsbeschränkungen sichergestellt. Je stärker der Datenzugang technisch kontrolliert wird, desto geringer ist das Risiko einer De-Anonymisierung der Daten, desto weniger müssen die Daten mittels statistischer Maßnahmen um Informationen reduziert werden und desto größer bleibt ihr Analysepotential.

Während das CUF nach einer Registrierung direkt durch das FDZ-DZHW übermittelt wird, wird das SUF über drei verschiedene Zugangswege angeboten: Download, Remote-Desktop und On-Site (für weiterführende Informationen vgl. Abschnitt III). Für jeden Zugangsweg wird eine andere SUF-Variante bereitgestellt, die unterschiedlich stark anonymisiert worden ist und entsprechend weniger oder mehr Informationen umfasst. Abbildung 3 gibt einen Überblick über den jeweiligen Grad der statistischen Anonymisierung und dem damit verbundenen Analysepotential. Im Folgenden werden die durchgeführten statistischen Anonymisierungsmaßnahmen in Abhängigkeit von Datenprodukt (SUF/CUF) und Zugangsweg erläutert.

⁴⁷ Das BDSG kommt zur Anwendung, da die DZHW GmbH juristisch als öffentliche Stelle des Bundes betrachtet wird (vgl. § 2 Abs. 3 BDSG). Der Bund hält die absolute Mehrheit der Anteile der DZHW GmbH und das Institut erfüllt Aufgaben der öffentlichen Verwaltung des Bundes im weitesten Sinn. Zur Auslegung einzelner rechtlicher Aspekte wird ergänzend die Europäische Datenschutzrichtlinie hinzugezogen.

⁴⁸ Wenn die Einzelangaben **nicht mehr** einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können, handelt es sich um „absolute Anonymisierung“. Wenn die Einzelangaben nur mit einem **unverhältnismäßig großen Aufwand** an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können, spricht man von faktischer Anonymisierung.

Abbildung 3: Datenzugangsweg, statistischer Anonymisierungsgrad und Analysepotential der Daten der 19. Sozialerhebung (2009)



[Statistische Anonymisierungsmaßnahmen] Im Rahmen der Anonymisierung sind zunächst alle Informationen, mit denen sich Personen oder Institutionen direkt identifizieren lassen, zu löschen. Von diesen sogenannten *direkten Identifikatoren*, wie Namen, Adressen oder E-Mail-Adressen, wurde im Rahmen der 19. Sozialerhebung nur die letztgenannte erfasst, und diese auch nur dann, wenn Personen auf der letzten Fragebogenseite ihre Bereitschaft erklärten, am HISBUS-Panel⁴⁹ teilzunehmen. Die E-Mail-Adressen wurden unmittelbar nach Öffnung des Rückumschlags vom Fragebogen abgetrennt und in einem separaten Datensatz ohne die Möglichkeit des Rückbezugs zum Datensatz der 19. Sozialerhebung erfasst. Sie sind weder im CUF noch in den verschiedenen SUF-Varianten enthalten. Des Weiteren wurden, um einen Rückbezug auf die Originaldaten zu verhindern, die Original-Identifikationsnummer der Befragten und der (aktuellen und vorherigen) Hochschule aus dem Datensatz 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.⁵⁰ Für die 19. Sozialerhebung wurden die folgenden Quasi-Identifikatoren identifiziert, die sowohl in externen Datenquellen⁵¹ als auch in der 19. Sozialerhebung vorliegen: Name sowie Art und Ort der Hochschule, Studienfach, Abschlussart, Höhe der Studiengebühren/-beiträge, Alter und Staatsangehörigkeit. Um eine eindeutige Zuordnung der Daten der 19. Sozialerhebung zu unterbinden, wurden diese Schlüsselmerkmale – je nach Datenprodukt bzw. Zugangsweg – aggregiert oder gelöscht (vgl. Tabelle 6). Beispielsweise ist das Merkmal „Höhe der Studiengebühren“ im SUF für die On-Site Nutzung uneingeschränkt

⁴⁹ Beim HISBUS-Panel handelt es sich um seit 2002 regelmäßig vom DZHW durchgeführte Online-Befragungen von Studierenden zu bildungspolitisch bedeutenden Themen. Die befragten Studierenden haben sich freiwillig bereit erklärt, am Panel teilzunehmen und werden aus anderen Befragungen des DZHW rekrutiert. Nähere Informationen zum HISBUS-Panel sind unter <http://www.hisbus.de> verfügbar.

⁵⁰ Dabei ist darauf hinzuweisen, dass die Identifikation einer Person bereits durch die Stichprobenauswahl erschwert wird, da eine Ungewissheit darüber besteht, ob eine befragte Person eine einzigartige Merkmalskombination in der Population aufweist.

⁵¹ z.B. Studenten- und Prüfungsstatistik des statistischen Bundesamtes, Alumninetzwerke der Hochschulen oder auch Berufszetwerke.

nutzbar. Im Remote-Desktop-SUF hingegen wird das Merkmal in 100-Euro-Schritten zu elf Gruppen (0 bis 100, über 100 bis 200, ..., über 1.000) und im Download-SUF sowie im CUF zu fünf Gruppen (0 bis 250, über 250 bis 500, ..., über 1.000) aggregiert. Offene Angaben sind ebenfalls Quasi-Identifikatoren (vgl. Ebel, 2015, S. 3). Sie wurden größtenteils bereits im Rahmen der Datenaufbereitung durch das Primärforschungsprojekt vercodet und werden in der Regel⁵² im CUF sowie in allen SUF-Varianten zur Verfügung gestellt. Teilweise wurden – in Abhängigkeit von der Sensibilität der enthaltenen Informationen und vom Zugangsweg – die vom Primärforschungsprojekt vorgenommenen Codierungen zusätzlich aggregiert. Nicht codierte offene Angaben (vgl. Kapitel 6.2) wurden im CUF und in allen SUF-Varianten gelöscht.

Zuletzt wurde geprüft, ob in den Daten *sensible Informationen*, z.B. zur Gesundheit, sexuellen Orientierung und zu politischen Einstellungen, enthalten waren. Diese eignen sich zwar nicht zur Re-Identifikation von Individuen oder Institutionen, jedoch können die Informationen im Falle einer De-Anonymisierung nutzbringend sein (Koberg, 2016, S. 694) und sind daher besonders schützenswert (vgl. §3 Abs. 9 BDSG, Art. 8 Abs. 1 und 2a EG-DSRL). In der 19. Sozialerhebung wurden Gesundheitsinformationen erhoben, für die bei den Befragten kein zusätzliches Einverständnis für die Sekundärnutzung eingeholt wurde. Daher wurden diese Antworten im CUF und allen SUF-Varianten gelöscht.

Zur Realisierung der absoluten Anonymität der Befragten im CUF wurde zusätzlich zu den vergleichsweise restriktiven statistischen Anonymisierungsmaßnahmen des Download-SUF eine 10-Prozent-Substichprobe der Download-SUF-Daten gezogen. Die nachfolgende Tabelle stellt in Kurzform die durchgeführten statistischen Anonymisierungsmaßnahmen je nach Datenform bzw. Zugangsweg dar.

⁵² Eine Ausnahme stellt die Angabe zum „Stipendiengeber“ (Frage 19) dar. Dieses Merkmal wird im CUF und im Download-SUF nicht zur Verfügung gestellt.

Tabelle 6: Überblick zur Anonymisierung der 19. Sozialerhebung (2009)⁵³

Merkmal	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF
Direkte Identifikatoren	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID
Studienfächer	Freigabe	Aggregation zu Studienbereichen ^a	Aggregation zu Fächergruppen ^a	Aggregation zu Fächergruppen ^a
Abschlussart	Freigabe	Freigabe	Aggregation ^b	Aggregation ^b
Semesterzahl bis Fach-/ Hochschulwechsel	Freigabe	Freigabe	1-10 einzeln ausgewiesen, ansonsten Aggregation (mehr als 10)	1-10 einzeln ausgewiesen, ansonsten Aggregation (mehr als 10)
Hochschule	Aggregation zu Hochschulart	Aggregation zu Hochschulart	Aggregation zu Hochschulart ^c	Aggregation zu Hochschulart ^c
Hochschulort	Aggregation zu vier Bundesländer-Gruppen	Aggregation zu vier Bundesländer-Gruppen	Aggregation zu neuen bzw. alten Bundesländern	Aggregation zu neuen bzw. alten Bundesländern
Anzahl Unterbrechungssemester	Freigabe	Freigabe	1-4 einzeln ausgewiesen, ansonsten Aggregation: 5-10, mehr als 10	1-4 einzeln ausgewiesen, ansonsten Aggregation: 5-10, mehr als 10
Wartezeit bis Studienbeginn (in Monaten)	Freigabe	Freigabe	Aggregation (Jahre): ≤1, >1 bis ≤2, >2 bis ≤3, >3 bis ≤5, >5 bis ≤10, >10	Aggregation (Jahre): ≤1, >1 bis ≤2, >2 bis ≤3, >3 bis ≤5, >5 bis ≤10, >10
andere Finanzierungsquellen	Freigabe	Aggregation ^d	Aggregation ^e	Aggregation ^e
Stipendiengeber	Freigabe	Aggregation zu sechs Kategorien	Löschung	Löschung
Grund für Befreiung von Studiengebühren	Aggregation zu vier Kategorien	Aggregation zu vier Kategorien	Aggregation zu vier Kategorien	Aggregation zu vier Kategorien
Höhe der Studiengebühren (in Euro)	Freigabe	Aggregation: 0 bis 100, >100 bis 200, ..., >1.000	Aggregation: 0 bis 250, >250 bis 500, ..., >1.000	Aggregation: 0 bis 250, >250 bis 500, ..., >1.000
Förderungsform (BAföG)	Freigabe	Freigabe	Löschung	Löschung
Alter (in Jahren)	Freigabe	0 bis 39 einzeln ausgewiesen, ansonsten Aggregation: 40-44, 45-50, ..., ≥75	0 bis 39 einzeln ausgewiesen, ansonsten Aggregation: 40-44, 45-50, ..., ≥75	0 bis 39 einzeln ausgewiesen, ansonsten Aggregation: 40-44, 45-50, ..., ≥75
Anzahl der Kinder	Freigabe	1 und 2 einzeln ausgewiesen,	1 und 2 einzeln ausgewiesen,	1 und 2 einzeln ausgewiesen,

⁵³ Detaillierte Informationen zu den anonymisierten Variablen sind dem Datensatzreport sowie dem Metadaten-suchsystem (<https://metadata.fdz.dzhw.eu>) zu entnehmen.

		ansonsten Aggregation: mehr als 2	ansonsten Aggregation: mehr als 2	ansonsten Aggregation: mehr als 2
Alter des jüngsten Kindes (in Jahren)	Freigabe	Aggregation: 1 bis 2, 3 bis 5, 6 bis 11, 12 bis 17, 18 und älter	Aggregation: 1 bis 5, 6 bis 17, 18 und älter	Aggregation: 1 bis 5, 6 bis 17, 18 und älter
Anzahl der Geschwister in Ausbildung	Freigabe	1, 2 und 3 einzeln ausgewiesen. Ansonsten Aggregation: 4 oder mehr	1 und 2 einzeln ausgewiesen. Ansonsten Aggregation: 3 oder mehr	1 und 2 einzeln ausgewiesen. Ansonsten Aggregation: 3 oder mehr
andere als deutsche Staatsangehörigkeit	Freigabe	4 Staaten einzeln ausgewiesen, ansonsten Aggregation zu Weltregionen	Aggregation zu Weltregionen	Aggregation zu Weltregionen
vorherige Staatsangehörigkeit	Freigabe	5 Staaten einzeln ausgewiesen, ansonsten Aggregation zu Weltregionen	Aggregation zu Weltregionen	Aggregation zu Weltregionen
andere als deutsche Staatsangehörigkeit der Eltern	Freigabe	4 Staaten einzeln ausgewiesen, ansonsten Aggregation zu Weltregionen	Aggregation zu Weltregionen	Aggregation zu Weltregionen
Erstes Land Auslandsstudium/ -praktikum	Freigabe	Freigabe	5 Staaten einzeln ausgewiesen, ansonsten Aggregation zu Weltregionen	5 Staaten einzeln ausgewiesen, ansonsten Aggregation zu Weltregionen
andere studienbezogene Auslandsaufenthalte (Länder)	Freigabe	Freigabe	Aggregation zu Weltregionen	Aggregation zu Weltregionen
Merkmale zur Gesundheit	Löschung	Löschung	Löschung	Löschung

^a Nach Schlüsselerzeichnis der Studenten- und Prüfungsstatistik WiSe 2008/2009 und SoSe 2009 von Destatis.

^b „Kirchliche Prüfung“ und „anderer Abschluss“ zu „sonstiger Abschluss (auch kirchlich)“ zusammengefasst.

^c Nur Unterscheidung von Fachhochschule und Universität (inklusive Pädagogische Hochschulen, Theologische Hochschulen, Kunst- und Musikhochschulen).

^d Einzeln ausgewiesen: „Kindergeld“ und „Leistungen für/wegen eigene(r) Kinder“; ansonsten Aggregation: „sonstige Sozialleistungen“ und „sonstige Quellen“.

^e Einzeln ausgewiesen: „Kindergeld“; ansonsten Aggregation: „alle anderen hier genannten Quellen“.

9 Literaturverzeichnis

- Apolinarski, B., Buck, D., Kandulla, M., Middendorff, E., Naumann, H. & Poskowsky, J. (2014). *Daten- und Methodenbericht zum Scientific Use File der 20. Sozialerhebung. Version 1.0.0* (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Hrsg.) (Dokumentation). Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).
- Ebel, T. (2015). *Empfehlungen zur Anonymisierung quantitativer Daten*. Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften.
- Häder, M. (2015). *Empirische Sozialforschung. Eine Einführung* (3. Aufl.). Wiesbaden: Springer Fachmedien Wiesbaden.
- Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. (2012). *FDZ-Methodenreport. Datenschutz am Forschungsdatenzentrum* (Nr. 06). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Isserstedt, W. & Kandulla, M. (2010). *Internationalisierung des Studiums - Ausländische Studierende in Deutschland - Deutsche Studierende im Ausland. Ergebnisse der 19. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch HIS Hochschul-Informationssystem* (Bundesministerium für Bildung und Forschung (BMBF), Hrsg.) (Wissenschaft Ideen zünden!), Bonn.
- Isserstedt, W., Middendorff, E., Kandulla, M., Borchert, L. & Leszczensky, M. (2010). *Die wirtschaftliche und soziale Lage der Studierenden in der Bundesrepublik Deutschland 2009. 19. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch HIS Hochschul-Informationssystem* (Bundesministerium für Bildung und Forschung (BMBF), Hrsg.) (Wissenschaft Ideen zünden!), Bonn.
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer & J. Skopek (Hrsg.), *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study* (S. 691-708). Wiesbaden: Springer VS.
- Kolenikov, S. (2014). Calibrating survey data using iterative proportional fitting (raking). *The Stata Journal*, 14 (1), 22-59.
- Kulik, M., Haaren, F. von & Thomsen, S. L. (2013). *Studiengebühren: Kein Effekt auf das Ausgabeverhalten der Studierenden* (niw info spezial Nr. 1). Hannover: Niedersächsisches Institut für Wirtschaftsforschung e. V. (NIW).
- 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.
- Middendorff, E. (2016). *Die Sozialerhebungen des Deutschen Studentenwerks 1951 – 2016. Ein historischer Überblick über Akteure, Wellen und Themen*. Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).
- Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. *Proceedings of the Survey Research Methods Section*, 225-230.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63 (2), 581-592.
- Sarcelletti, A. (2015). Bachelor students' transition to postgraduate studies. Do students with and without migration background have different plans? *Beiträge zur Hochschulforschung*, 37 (2), 116-139.
- Thomsen, S. & Haaren, F. von. (2014). *Did tuition fees in Germany constrain students' budgets? New evidence from a natural experiment* (IZA Discussion Paper no. 8623). Bonn: Forschungsinstitut zur Zukunft der Arbeit (IZA).
- Valliant, R., Dever, J. A. & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York (NY): Springer New York.