

Elke Middendorff / Ute Hoffstätter

17. Sozialerhebung

Daten- und Methodenbericht zur
Studierendenbefragung 2003

Daten- und Methodenbericht

November 2020

Autorinnen:
Dr. Elke Middendorff
Ute Hoffstätter

Unter Mitarbeit von:
Dr. Andreas Daniel (Kapitel 3)

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

Geschäftsführung:
Prof. Dr. Monika Jungbauer-Gans
Karen Schlüter

Vorsitzender des Aufsichtsrats:
Ministerialdirigent Peter Greisler

Registergericht:
Amtsgericht Hannover | B 210251

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

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



Inhaltsverzeichnis

| | |
|---------------------------------------|----|
| Tabellenverzeichnis | IV |
| 1 Einleitung | 5 |
| 2 Datennutzungshinweise | 6 |
| 3 Gewichtung | 7 |
| Vorgehen und Anwendungshinweise | 7 |
| Gewichtung des Datensatzes | 8 |
| 4 Anonymisierung | 9 |
| Literaturverzeichnis | 12 |

Tabellenverzeichnis

| | | |
|------------|--|----|
| Tabelle 1: | Bereitgestellte Gewichte zur 17. Sozialerhebung (2003) | 8 |
| Tabelle 2: | Überblick zur Anonymisierung der 17. Sozialerhebung (2003) | 11 |

1 Einleitung

Die Sozialerhebung des Deutschen Studentenwerks (DSW) ist eine seit 1951 bestehende Untersuchungsreihe zur wirtschaftlichen und sozialen Lage der Studierenden in Deutschland.¹ Sie wird seit 1982² (10. Sozialerhebung) im Auftrag des bzw. seit der 21. Sozialerhebung in Kooperation mit dem DSW durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)³ durchgeführt. Das Bundesministerium für Bildung und Forschung (BMBF) fördert die Studie seit der 6. Sozialerhebung (1967/1968). Die Sozialerhebung dient – in Ergänzung zur amtlichen Hochschulstatistik – unter anderem dem nationalen und internationalen Bildungsmonitoring. Darüber hinaus liefert sie wichtiges Steuerungswissen für hochschul- und sozialpolitische Fragen sowie belastbare und umfassende Daten für die Forschung.

Im Rahmen der Tätigkeit des vom BMBF geförderten Forschungsdatenzentrums für Hochschul- und Wissenschaftsforschung am DZHW (FDZ-DZHW) werden die Daten einiger jüngerer Erhebungen dieser Reihe nachträglich zum Zweck der Nachnutzung aufbereitet und dokumentiert.⁴ Die 17. Sozialerhebung wird als *Scientific Use File* (SUF) für die wissenschaftliche Sekundärnutzung zur Verfügung gestellt. Neben dem Datensatz der Erhebung wird auch Dokumentationsmaterial zum Datensatz und zur Durchführung der Studie bereitgestellt.⁵

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation zur 17. Sozialerhebung (doi: [10.21249/DZHW:ssy17:2.0.0](https://doi.org/10.21249/DZHW:ssy17:2.0.0)).⁶ Im Unterschied zur im November 2019 releasen ersten Version enthält die jetzt veröffentlichte zweite Version des SUF auch die Datensätze der 362 Bildungsinländer*innen, die gleichzeitig und mit demselben Instrument wie die deutschen Studierenden befragt worden waren. Von Seiten des Primärforschungsprojektes wurden deren Daten in einem eigenen Datensatz abgelegt und getrennt analysiert. Beide Datensätze wurden jetzt zusammengefügt und neu gewichtet. (s. Kapitel 3). Die zentralen Informationen zur Nutzung der Daten dieser Studie folgen in Kapitel 2. Kapitel 3 und 4 enthalten die Beschreibung der vorgenommenen Gewichtung bzw. der Anonymisierung.

Weitere Dokumentationsmaterialien zur Studie (Datensatzreport, Fragebogen etc.) können frei im Metadaten-System des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden.

¹ Weiterführende Informationen, Berichte und Materialien zur Sozialerhebung stehen auf der Website des Projekts zur Verfügung (<http://www.sozialerhebung.de>).

² Die 1. (1951) und 2. Sozialerhebung (1953) wurden vom Studentenwerk Frankfurt am Main im Auftrag des Verbands Deutscher Studentenwerke durchgeführt. Das Studentenwerk führte auch die 3. (1956) bis 9. Sozialerhebung (1979) durch, die vom Deutschen Studentenwerk (DSW) beauftragt wurden. Einen detaillierten Überblick über Akteure, Methoden, Themen und projektbezogene Publikationen der Untersuchungsreihe von ihren Anfängen bis zur 21. Sozialerhebung bietet ein Working Paper von Middendorff (2019). (<http://www.sozialerhebung.de/Hintergrund/geschichte>).

³ Das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung (DZHW, <http://www.dzhw.eu>) entstand im August 2013 durch eine Ausgründung aus der HIS Hochschul-Informationen-System GmbH. Im nachfolgenden Text wird stets der Begriff DZHW verwendet, auch wenn die Studie vor der Ausgründung 2013 durchgeführt wurde. Auf allen Originaldokumenten der 17. Sozialerhebung (Fragebogen, Flyer etc.) sowie in den Berichten zum Projekt ist entsprechend die HIS GmbH (HIS) 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/#!/de/studies/stu-ssy17S?version=2.0.0>) heruntergeladen werden.

⁶ Berücksichtigt wird hier nur die Stichprobe deutscher und bildungsinländischer Studierender, die mittels eines Papierfragebogens befragt wurden. Die im Rahmen der 17. Sozialerhebung ebenfalls und erstmals getrennt erhobenen Daten der Bildungsausländer(innen) werden hier nicht dokumentiert.

2 Datennutzungshinweise

[Voraussetzungen der Datennutzung] Die Daten der 17. Sozialerhebung werden durch das FDZ des DZHW entsprechend der europäischen Datenschutzgrundverordnung (EU-DSGVO) anonymisiert bereitgestellt und ausschließlich zur wissenschaftlichen Nutzung freigegeben.⁷ Das FDZ bietet ein *Scientific Use File* (SUF) für die wissenschaftliche Sekundärnutzung 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 an einer wissenschaftlichen Einrichtung müssen gemeinsam mit einer/einem betreuenden Mitarbeiter(in) einen Datennutzungsvertrag abschließen. 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.

[Datenzugang] Das SUF der 17. Sozialerhebung wird via Download angeboten.

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

[Datenprodukte] Über den *Digital Object Identifier* (DOI) [10.21249/DZHW:ssy17:2.0.0](https://doi.org/10.21249/DZHW:ssy17:2.0.0) ist eine Website mit zentralen Informationen zur Studie, weiteren Dokumentationsmaterialien sowie einer Übersicht der zur Verfügung stehenden Datenprodukte zur Studie erreichbar.

[Gebühren der Datenbereitstellung] Das SUF wird derzeit (Stand: November 2020) kostenfrei zur Verfügung gestellt. Änderungen bzw. die aktuelle Gebührenordnung können auf der Website des FDZ (<https://fdz.dzhw.eu>) eingesehen werden.

[Pflichten der Datennutzer(innen)] Die Datennutzer(innen) sind verpflichtet, folgende Regeln⁸ einzuhalten:

- **Wissenschaftliche Nutzung:** Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.
- **De-Anonymisierungsverbot:** Jeder Versuch der Re-Identifikation von Analyseeinheiten (z. B. Personen, Haushalten, Institutionen) ist verboten.
- **Gebot zur Mitteilung von Sicherheitslücken:** Falls Datennutzer(innen) Kenntnis von Sicherheitslücken hinsichtlich Datenschutz bzw. Datensicherheit erlangen, müssen diese dem FDZ unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** SUF dürfen nur durch die Person(en) genutzt werden, die den Datennutzungsvertrag abgeschlossen hat/haben.

⁷ Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, S. 6 ff.), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (LifBi) (Koberg, 2016, S. 699 ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (Hochfellner, Müller, Schmucker und Roß, 2012, S. 9 f.) orientieren. Das FDZ des DZHW hat diesen Ansatz an die Anforderungen der eigenen Datenbestände angepasst und nutzt vier Kategorien von Maßnahmen zur Sicherstellung des Datenschutzes, die in unterschiedlicher Weise kombiniert werden können: Rechtlich-institutionelle Maßnahmen, informationelle Maßnahmen, technische Maßnahmen und statistische Maßnahmen.

⁸ Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

- **Löschungsgebot:** Download-SUF sind nach Ablauf der vereinbarten Nutzungsdauer (in der Regel 1,5 Jahre) von jeglichen Rechnern, Servern und Datenträgern zu löschen. Ebenso müssen alle Sicherungskopien, modifizierten Datensätze (z. B. Arbeits-, Auszugs- oder Hilfsdateien) sowie Ausdrücke vernichtet werden.
- **Bereitstellung/Meldung von Publikationen:** Jede Art von Publikation, die aus der Arbeit mit Daten des FDZ hervorgeht, ist dem FDZ unmittelbar nach Veröffentlichung anzuzeigen und – unabhängig vom Veröffentlichungsformat – als elektronische Version zur Verfügung zu stellen.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen gemäß der Vorgaben des FDZ zitiert werden.⁹

3 Gewichtung

Die Gewichtung der Daten dient dem Ausgleich von Verzerrungen der Stichprobe aufgrund des Stichprobendesigns sowie unterschiedlicher Mitwirkungsbereitschaft verschiedener Gruppen in der Grundgesamtheit. Sie erfolgt im Vergleich zur definierten Grundgesamtheit. Nach einer allgemeinen Einführung in die Vorgehensweise und einer Darstellung der erstellten Gewichte wird die Gewichtungsprozedur im Detail beschrieben.

3.1 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 für bestimmte relevante Subgruppen die Fallzahlen zu erhöhen.
- **Verzerrung durch Nonresponse:** Ausfallprozesse (z. B. Nichtteilnahmen, fehlende Erreichbarkeit, Verlust auf dem Postweg) führen zu einem verringerten Rücklauf und somit zu einer Differenz zwischen Brutto- und Nettostichprobe. Wenn diese Ausfallprozesse unsystematisch sind (Missing Completely at Random), können sie ignoriert werden.¹⁰ Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (Missing at Random, Missing Not at Random), der einer Modellierung bedarf.¹¹

[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.

⁹ vgl. Zitation unter [https://metadata.fdz.dzhw.eu/#!/de/data-sets/dat-ssy17-ds1\\$?version=2.0.0](https://metadata.fdz.dzhw.eu/#!/de/data-sets/dat-ssy17-ds1$?version=2.0.0)

¹⁰ Das trifft dann zu, wenn die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

¹¹ Siehe grundlegend zu den unterschiedlichen Formen von Ausfallprozessen Rubin (1976).

In einem letzten Schritt können die nonresponse-adjustierten Designgewichte anhand von Merkmalsverteilungen aus der Grundgesamtheit kalibriert werden (*Kalibrierung*).

Aufgrund des Stichprobendesigns der 17. 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. In Tabelle 1 sind die erstellten Gewichte dargestellt.

Tabelle 1: Bereitgestellte Gewichte zur 17. Sozialerhebung (2003)

| Variablenname | Beschreibung |
|---------------|---|
| gewide | Gewicht für Analysen auf Bundesebene |
| gewiow | Gewicht für Analysen auf der Ebene Ost- / Westdeutschland |
| gewireg | Gewicht für Analysen auf Regionalebene (Nord, Süd, Ost, West) |

[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.¹² Das Gewicht *gewide* ist für Auswertungen auf Bundesebene vorgesehen. Das Gewicht *gewiow* bietet Gewichte für Auswertungen differenziert nach Ost- und Westdeutschland, *gewireg* für Analysen auf Regionalebene (Nord, Süd, Ost, West). 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.

3.2 Gewichtung des Datensatzes

[Designgewichtung] Aufgrund des Stichprobendesigns sind Studierende einiger Hochschulen überrepräsentiert. Die deshalb bestehende höhere Wahrscheinlichkeit für Studierende dieser Hochschulen, in die Stichprobe zu gelangen, wurde durch eine Designgewichtung ausgeglichen. Elemente, die mit höherer Wahrscheinlichkeit als andere in die Stichprobe eingehen, erhalten somit ein niedrigeres Gewicht und umgekehrt.

[Kalibrierung der Designgewichte] Eine Nonresponse-Adjustierung der Designgewichte war auf Individualebene nicht möglich. Es lagen jedoch Informationen zu folgenden Merkmalen der Grundgesamtheit¹³ vor, die zur Kalibrierung der Gewichte verwendet werden konnten: Region der Hochschule, Geschlecht, Fächergruppe, Hochschultyp, Personen mit deutscher

¹² Siehe hierzu die Stata-Hilfe (Befehl: *help weights*).

¹³ Alle Informationen, die zur Kalibrierung der Designgewichte verwendet wurden, leiten sich aus Daten des Statistischen Bundesamtes zum Semester vor der Erhebung (WiSe 2002/2003) ab, da die aktuelle Statistik zum Zeitpunkt der Gewichtung noch nicht vorlag.

Staatsangehörigkeit versus Bildungsinländer.¹⁴ Bei der Redressment-Gewichtung wurden einzelne, sehr kleine Zellbesetzungen, die zu sehr hohen Gewichten führen würden, zusammengefasst.

Die Kalibrierung erfolgte sowohl auf Bundesebene (gewide) als auch gesondert für jedes Bundesland. Im Prozess der Anonymisierung wurden die Bundesländer zu vier Regionen aggregiert (Süd-, West-, Nord- und Ostdeutschland). Das Regionalgewicht (gewireg) wird für regionale Analysen zur Verfügung gestellt, das Ost-West-Gewicht (gewiow) für Ost-West-Vergleiche.

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.

Für die in der Version 2.0.0 integrierten 362 Bildungsinländer*innen wurden alle drei bereitgestellten Gewichte gemäß dem oben beschriebenen Verfahren neu berechnet.

4 Anonymisierung

[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten¹⁶, 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 2017.¹⁷ Danach sind personenbezogene Daten für die Weitergabe zur wissenschaftlichen Sekundärnutzung (ohne Vorliegen einer Einverständniserklärung zur Sekundärnutzung der personenbezogenen Daten) in der Regel derart aufzubereiten, dass „die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden können“ (Art. 4 Abs. 5 DSGVO; s. auch Art. 89 DSGVO sowie Erwägungsgrund 26 DSGVO). Das heißt, für die Weitergabe von Daten aus wissenschaftlichen Forschungsprojekten an Dritte sind die Daten derart zu anonymisieren, dass kein Bezug zur Person mehr hergestellt werden kann.

[Datenzugang, Anonymisierungsgrad und Analysepotential] Das FDZ des DZHW stellt für die 17. Sozialerhebung ein SUF für die wissenschaftliche Sekundärnutzung zur Verfügung. Die

¹⁴ 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 (inklusive Pädagogische Hochschulen, Theologische Hochschulen, Kunst- und Musikhochschulen) versus Fachhochschule; Fächergruppe: entsprechend Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2002/2003).

¹⁶ „Personenbezogene Daten (sind) alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person“) beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden kann, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind“ (Art. 4 DSGVO, S. 1).

¹⁷ 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.

Anonymität der Befragten wird dabei über eine Kombination aus statistischen Maßnahmen und technischen Zugriffsbeschränkungen sichergestellt.

Das SUF wird via Download angeboten. Im Folgenden werden die durchgeführten statistischen Anonymisierungsmaßnahmen für den Zugangsweg Download-SUF erläutert.

[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 17. Sozialerhebung nur die letztgenannte erfasst, und auch nur dann, wenn die Befragten ihre Bereitschaft erklärten, am HISBUS-Panel teilzunehmen. Die E-Mail-Adressen wurden unmittelbar nach Öffnung des Rückumschlags vom Fragebogen getrennt und in einem separaten Datensatz ohne die Möglichkeit des Rückbezugs zum Datensatz der 17. Sozialerhebung erfasst. Sie sind nicht mit dem Datensatz der 17. Sozialerhebung verknüpfbar und somit nicht im SUF enthalten. Des Weiteren wurde, um einen Rückbezug auf die Originaldaten zu verhindern, die Original-Identifikationsnummer aus dem Datensatz entfernt und durch eine neue, nach dem Zufallsprinzip 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.¹⁸ Für die 17. Sozialerhebung wurden beispielsweise folgende Merkmale als Quasi-Identifikatoren eingestuft: Name sowie Art und Ort der Hochschule, Studienfach, Abschlussart, Alter und Staatsangehörigkeit. Um eine eindeutige Zuordnung der Daten der 17. Sozialerhebung zu betroffenen Personen zu unterbinden, wurden diese Schlüsselmerkmale aggregiert oder gelöscht (s. Tabelle 2).

Ebel und Meyermann (2015) empfehlen, offene Angaben in jedem Fall 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). Die offenen Angaben waren größtenteils bereits im Rahmen der Datenaufbereitung durch das Primärforschungsprojekt vercodet worden und werden in dieser Form zur Verfügung gestellt. Teilweise wurden jedoch – in Abhängigkeit von der Sensibilität der enthaltenen Informationen – die vom Primärforschungsprojekt vorgenommenen Codierungen zusätzlich aggregiert. Nicht codierte offene Angaben wurden im SUF gelöscht.

Zuletzt wurde geprüft, ob in den Daten *sensible Informationen*, z. B. zur Gesundheit, sexuellen Orientierung oder zu politischen Einstellungen, enthalten waren. Diese eignen sich zwar nicht unmittelbar 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 (Art. 9 DSGVO, Erwägungsgrund 51 DSGVO). In der 17. Sozialerhebung wurden gesundheitsbezogenen Informationen erhoben, für die bei den Befragten kein zusätzliches Einverständnis für die Sekundärnutzung eingeholt wurde. Daher wurden diese Antworten gelöscht. Die nachfolgende Tabelle 2 stellt in Kurzform die durchgeführten statistischen Anonymisierungsmaßnahmen dar. Variablen, die im Download-SUF aus Datenschutzgründen *nicht* verfügbar sind, enden mit dem Suffix „_a“.

¹⁸ 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.

Tabelle 2: Überblick zur Anonymisierung der 17. Sozialerhebung (2003)¹⁹

| Merkmal | Download-SUF |
|---|---|
| Direkte Identifikatoren | Löschung |
| Original-ID | Löschung und Vergabe einer zufälligen ID |
| Studienfächer | Aggregation zu Fächergruppen ^a |
| Abschlussart | Zusammenfassung: „kirchliche Prüfung“ und „anderer Abschluss“ zu „sonstiger Abschluss (auch kirchlich)“ |
| Semesterzahl bis Fach-/Abschlusswechsel | 1 bis 10 einzeln ausgewiesen; Aggregation: „mehr als 10“ |
| Hochschule | Aggregation zu Hochschulart ^b |
| Bundesland des aktuellen Studiums | Aggregation zu Regionen ^b |
| Gründe für Studienunterbrechung | Zusammenfassung: „gesundheitliche Probleme“ zu „sonstige Gründe“ |
| Wartezeit bis Studienbeginn (in Monaten) | Aggregation: „bis 1 Jahr“, „mehr als 1 Jahr bis 2 Jahre“, „mehr als 2 Jahre bis 3 Jahre“, „mehr als 3 bis 5 Jahre“, „mehr als 5 bis 10 Jahre“ und „mehr als 10 Jahre“ |
| andere Finanzierungsquellen | Einzeln ausgewiesen: „Kindergeld“ und „Leistungen für/wegen eigene(r) Kinder“; ansonsten Aggregation zu „sonstige Quellen“ |
| Förderungsform (BAföG) | Löschung |
| Tätigkeit zum Gelderwerb in vorlesungsfreier bzw. Vorlesungszeit | Aggregation ^c |
| Alter (in Jahren) | bis 39 einzeln ausgewiesen, ansonsten Aggregation: „40 bis 49“, „50 Jahre und älter“ |
| Anzahl der Kinder | 1 und 2 einzeln ausgewiesen; Aggregation: „3 Kinder und mehr“ |
| andere als deutsche Staatsangehörigkeit / vorherige Staatsangehörigkeit | Aggregation: EU15, sonstiges Europa, weitere Weltregionen ^d |
| (Sonstige) gesundheitliche/private Informationen | Löschung |
| (Sonstige) offene Angaben | Vercodung/Löschung |

^a Aggregation orientiert an Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2002/2003 und SoSe 2003 von Destatis., vgl. Codierliste cl-dzhw-33, die hier hinterlegt ist:

[https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-ssy17-ins1\\$?version=2.0.0](https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-ssy17-ins1$?version=2.0.0).

^b Aggregation zu Hochschularten Fachhochschule und Universität (inklusive Pädagogische Hochschulen, Theologische Hochschulen, Kunst- und Musikhochschulen) sowie Aggregation zu den Regionen Süd-, West-, Nord- und Ostdeutschland, vgl. Codierliste cl-dzhw-34, die hier hinterlegt ist:

[https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-ssy17-ins1\\$?version=2.0.0](https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-ssy17-ins1$?version=2.0.0).

^c Die nachcodierten offenen Angaben von Tätigkeiten, die sich im weiteren Sinne auf die Bundeswehr bzw. medizinische Tests, Blutspenden u. Ä. bezogen sind den „sonstigen Tätigkeiten“ zugeordnet worden.

^d Aggregation zu EU und sonstiges Europa entsprechend dem Stand zum Sommersemester 2003 bzw. entsprechend der Unterscheidung von Weltregionen der Vereinten Nationen, vgl. Codierliste cl-dzhw-32, die hier hinterlegt ist: [https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-ssy17-ins1\\$?version=2.0.0](https://metadata.fdz.dzhw.eu/#!/de/instruments/ins-ssy17-ins1$?version=2.0.0).

¹⁹ Detaillierte Informationen zu den anonymisierten Variablen sind dem Datensatzreport sowie dem Metadaten-suchsystem ([https://metadata.fdz.dzhw.eu/#!/de/studies/stu-ssy17\\$?version=2.0.0](https://metadata.fdz.dzhw.eu/#!/de/studies/stu-ssy17$?version=2.0.0)) zu entnehmen.

Literaturverzeichnis

- Ebel, T. & Meyermann, A. (2015). *Hinweise zur Anonymisierung von quantitativen Daten* (Forschungsdaten Bildung informiert Nr. 3). Verbund Forschungsdaten Bildung.
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer & J. Skopek (Hrsg.), *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study* (S. 691–708). Wiesbaden: Springer VS. doi:10.1007/978-3-658-11994-2
- Middendorff, E. (2019). *Die Sozialerhebungen des Deutschen Studenten-werks 1951 – 2016. Ein historischer Überblick über Akteure, Wellen, Methoden, Themen und projektbezogene Publikationen* (Working Paper). Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(2), 581–592.