Daten- und Methodenbericht September 2021

Heidrun Schneider | Sandra Vietgen | Marten Wallis

DZHW-Studienberechtigtenpanel 2012

Daten- und Methodenbericht zu den Erhebungen des Studienberechtigtenjahrgangs 2012 (1. bis 2. Befragungswelle)



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/de/



Autor*innen

Heidrun Schneider Telefon +49 (0)511 450670-155 E-Mail: h.schneider@dzhw.eu

Sandra Vietgen Telefon +49 (0)511 450670-405 E-Mail: vietgen@dzhw.eu

Marten Wallis Telefon +49 (0)511 450670-434 E-Mail: wallis@dzhw.eu

Impressum

Herausgeber

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

Geschäftsführerinnen:

Prof. Dr. Monika Jungbauer-Gans Karen Schlüter Vorsitzender des Aufsichtsrats: Ministerialdirigent Peter Greisler

Registergericht:

Amtsgericht Hannover | B 210251 Umsatzsteuer-Identifikationsnummer: DE291239300

September 2021

Inhaltsverzeichnis

Tab	ellen-/	Abbildungsverzeichnis			
1	Einleitung				
2	Datennutzungshinweise				
3	Inhalt und Anlage der Studie				
4	Grund	gesamtheit und Stichprobenverfahren	8		
5	Gewicl	ntung	10		
	5.1 Vor	gehen und Anwendungshinweise	10		
	5.2 Gev	vichtung des Datensatzes	11		
6	Anony	misierung	14		
7	Literat	ur	18		
Tal	oellen-	/Abbildungsverzeichnis			
Abbildung 1: Tabelle 1:		Datenzugangswege und Analysepotential Anlage und konzeptionelle Veränderungen der DZHW-			
		Studienberechtigtenbefragungsreihe von 1976 bis 2012			
	ldung 2:	Kohorten-Panel-Design der DZHW-Studienberechtigtenbefragungsreihe			
Tabelle 2: Abbildung 3:		Bereitgestellte Gewichte zum DZHW-Studienberechtigtenpanel 2012	. 11		
AUUI	iuulig 5:	Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten des DZHW-Studienberechtigtenpanels 2012	15		
Tabe	elle 3:	Maßnahmen der statistischen Anonymisierung der Daten des DZHW-	. 13		
		Studienherechtigennanels 2012 nach Zugangsweg	16		

1 Einleitung

Die DZHW-Studienberechtigtenpanels stellen eine Untersuchungsreihe zu den Ausbildungs- und Studienentscheidungen beziehungsweise Ausbildungs- und Studienverläufen von Schulabgänger*innen mit Abitur oder Fachhochschulreife dar. ¹ Sie werden durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)² durchgeführt, vom Bundesministerium für Bildung und Forschung (BMBF) finanziert und dienen – in Ergänzung zur amtlichen Hochschulstatistik – dem nationalen Bildungsmonitoring sowie der Beantwortung von Fragestellungen der Hochschul- und Wissenschaftsforschung. Seit 1976 werden alle ein bis drei Jahre neue Studienberechtigtenjahrgänge befragt.

Im Rahmen der Tätigkeit des 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. Sie werden über verschiedene Zugangswege als Scientific Use Files (SUF) für die wissenschaftliche Sekundärnutzung zur Verfügung gestellt. Neben den Datensätzen der Erhebungen werden auch die Dokumentationsmaterialien zu den Datensätzen und zur Durchführung der Studien bereitgestellt.³

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation der ersten zwei Befragungswellen des Studienberechtigtenjahrgangs 2012 (doi: 10.21249/DZHW:gsl2012:2.0.0). Weitere Dokumentationsmaterialien zur Studie (Datensatzreports, Fragebögen, Variablenfragebögen etc.) können frei im Metadatensuchsystem des FDZ-DZHW (https://metadata.fdz.dzhw.eu) heruntergeladen werden.

In Kapitel 2 wird ein Überblick über die zentralen Informationen zur Nutzung der Daten dieser Studie gegeben. Kapitel 3 stellt Inhalt und Anlage der DZHW-Studienberechtigtenbefragungsreihe bis zum Jahr 2012 vor. In Kapitel 4 wird die Grundgesamtheit und das Stichprobenverfahren beschrieben und in den Kapiteln 5 und 6 folgt die Darstellung der vorgenommenen Gewichtung und Anonymisierung.

¹Aktuelle Informationen zum DZHW-Studienberechtigtenpanel können über die Website des Projektes (https://bildungswege.dzhw.eu/) abgerufen werden.

²Das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW, https://www.dzhw.eu/) entstand im August 2013 durch eine Ausgründung aus der HIS Hochschul-Informations-System GmbH. Im nachfolgenden Text wird durchgängig der Begriff DZHW verwendet, auch wenn die Studie teilweise vor der Ausgründung 2013 durchgeführt wurde.

³Informationen zu verfügbaren Datensätzen und Dokumentationen werden auf der Website des FDZ-DZHW (https://fdz.dzhw.eu) zur Verfügung gestellt.

2 Datennutzungshinweise

[Voraussetzungen der Datennutzung] Die Dokumentation der Daten des DZHW-Studienberechtigtenpanels 2012 ist öffentlich verfügbar unter dem Digital Object Identifier (DOI) 10.21249/DZHW:gsl2012:2.0.0. Die Daten selbst werden durch das FDZ des DZHW entsprechend der europäischen Datenschutzgrundverordnung (EU-DSGVO) anonymisiert bereitgestellt und ausschließlich zur wissenschaftlichen Nutzung freigegeben. Das FDZ bietet ein Scientific Use File (SUF) für die wissenschaftliche Sekundärnutzung an. Dies kann über die Website des FDZ beantragt werden (https://www.fdz.dzhw.eu/de/form/data-usage-application). Für die Nutzung von SUF wird ein Datennutzungsvertrag abgeschlossen.

[Datenzugang] Das SUF wird über zwei Zugangswege angeboten, die hinsichtlich des Speicherortes, der Möglichkeit der eigenständigen Verknüpfung mit externen Daten und der Kontrollmöglichkeiten des FDZ unterschiedlich restriktiv sind.

- Remote-Desktop: Die Daten werden auf einem Terminal-Server des FDZ bereitgestellt. Über eine besonders gesicherte Verbindung zwischen dem lokalen Computer der nutzenden Person und dem Terminal-Server des FDZ können die Daten mit der auf dem Terminal-Server vorhandenen Software analysiert werden. Das Übertragen der Daten auf den lokalen Computer ist nicht möglich. Analyseergebnisse werden erst nach einer Prüfung auf datenschutzrechtliche Unbedenklichkeit durch das FDZ freigegeben und zur Verfügung gestellt.
- On-Site: Die Daten werden in den Räumlichkeiten des FDZ in einer kontrollierten Umgebung an einem speziell gesicherten Computer zur Analyse bereitgestellt. Wie beim Remote-Desktop-Zugang werden Analyseergebnisse erst nach einer Prüfung auf datenschutzrechtliche Unbedenklichkeit durch das FDZ freigegeben und zur Verfügung gestellt.

Die bereitgestellten Daten weisen je nach Zugangsweg einen unterschiedlich hohen Informationsgehalt auf und unterscheiden sich damit hinsichtlich ihres Analysepotentials (siehe hierzu 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.⁵ Mit diesem Vorgehen wird ein Höchstmaß an Nutzbarkeit und gleichzeitig ein bestmöglicher Schutz der bereitgestellten Daten sichergestellt.

⁴ Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, 6), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (LlfBi) (vgl. Koberg (2016, S. 699)) 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 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.

⁵ Zu den verschiedenen Anonymisierungsgraden und Analysepotentialen der zwei SUF-Varianten vgl. Kapitel 6.

Abbildung 1:

Studienberechtigtenpanel 2012
(1. und 2. Welle)

Datenprodukt
(Nutzungszweck)

SUF
(Forschung)

Remote- Desktop

On-Site

Datenzugangswege und Analysepotential

[Gebühren der Datenbereitstellung] SUF werden derzeit (Stand: September 2021) kostenfrei zur Verfügung gestellt. Änderungen bzw. die aktuelle Gebührenordnung können auf der Website des FDZ (https://fdz.dzhw.eu) eingesehen werden.

[Pflichten der Datennutzer*innen] Die Datennutzer*innen sind verpflichtet, folgende Regeln⁶ einzuhalten:

- Wissenschaftliche Nutzung: Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.
- **De-Anonymisierungsverbot:** Jeder Versuch der Re-Identifikation von Analyseeinheiten (z.B. Personen, Haushalten, Institutionen) ist verboten.
- **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 Personen genutzt werden, die einen Datennutzungsvertrag abgeschlossen haben.
- Bereitstellung/Meldung von Publikationen: Jede Art von wissenschaftlichen Arbeiten (z.B. Publikationen, Abschlussarbeiten, Dissertationen), die aus der Arbeit mit Daten des FDZ hervorgehen, ist dem FDZ anzuzeigen. Dabei ist dem FDZ eine elektronische Version der Druckfassung zur Verfügung zu stellen.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z.B. Abschlussarbeiten) und Vorträgen laut der Vorgaben des FDZ zitiert werden.

 $^{^{}m 6}$ Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

3 Inhalt und Anlage der Studie

[Studienreihe] Das DZHW-Studienberechtigtenpanel 2012 ist Teil der DZHW-Studienberechtigtenbefragungsreihe, in der anhand von standardisierten Mehrfachbefragungen Informationen zu den nachschulischen Werdegängen von Schulabsolvent*innen mit (schulischer) Hochschulzugangsberechtigung erfasst werden. Die Untersuchungsreihe beginnt mit der Befragung des Studienberechtigtenjahrgangs 1976 in der damaligen BRD; seitdem wurde jeder zweite oder dritte Studienberechtigtenjahrgang (Kohorte) befragt. Die Schulabsolvent*innen in den neuen Bundesländern wurden erstmalig 1990 einbezogen. Die Grundgesamtheit aller DZHW-Studienberechtigtenbefragungen sind die Schulabsolvent*innen allgemeinbildender und berufsbildender Schulen eines Schuljahres mit bundesdeutscher Hochschulreife.

In der Regel⁸ werden für jeden Studienberechtigtenjahrgang mehrere Befragungswellen zu unterschiedlichen Zeitpunkten vor und nach Erwerb der Hochschulzugangsberechtigung durchgeführt, das heißt es handelt sich um ein kombiniertes Kohorten-Panel-Design. Tabelle 1 hebt die konzeptionellen Änderungen hervor. Bis 1986 verfolgte das DZHW das Ziel jeden Studienberechtigtenjahrgang kurz nach sowie 2 ½, 4 ½ und 12 ½ Jahre nach Schulabschluss zu befragen. Ab 1990 wurde die Befragung zu verspäteten oder korrigierten Ausbildungs- und Studienentscheidungen zweieinhalb Jahre nach Schulabschluss gestrichen.⁹ Seit dem Studienberechtigtenjahrgang 2005 werden die Studienberechtigtenjahrgänge zusätzlich bereits ein halbes Jahr *vor* Schulabschluss befragt.

Die verschiedenen Erhebungen wurden als schriftlich-postalische Paper-and-Pencil-Befragung (PAPI) durchgeführt. Die Erhebungsinstrumente aller Kohorten enthalten Fragen zu Bildungsintentionen, nachschulischem Werdegang, zum Studium, zum Übergang in den Beruf und zu soziodemographischen und bildungsbiographischen Merkmalen. Der thematische Fokus einer Befragungswelle orientiert sich an der jeweils aktuellen Bildungs-, Berufs- und Lebensphase der Befragten zum Befragungszeitpunkt. Über eine Spanne von fast 40 Jahren bis zur vorliegenden Befragung entstand so eine Zeitreihe von 19 Jahrgängen mit insgesamt 44 Befragungswellen (vgl. Abbildung 2).

⁷ Aktuelle Informationen zum Studienberechtigtenpanel sowie den bisher verfügbaren Scientific Use Files der Erhebungsreihe finden Sie unter https://metadata.fdz.dzhw.eu.

⁸ Die Ausnahmen bilden die Studienberechtigtenjahrgänge 1991, 1993, 1996 und 2004, die nur einmal befragt wurden.

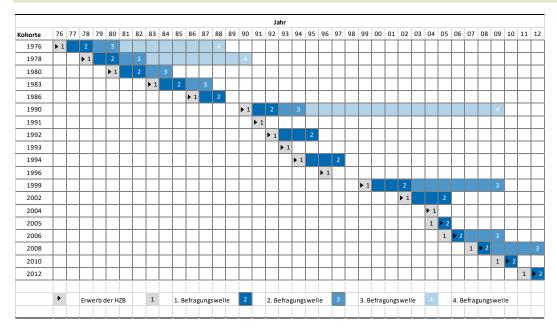
⁹ Eine weitere Ausnahme stellt die, in Kooperation mit dem Bundesinstitut für Berufsbildung (BIBB) durchgeführte, zusätzliche Befragungswelle des Studienberechtigtenjahrgangs 1990 dar, die 5 ½ Jahre nach Schulabschluss erfolgte.

Tabelle 1: Anlage und konzeptionelle Veränderungen der DZHW-Studienberechtigtenbefragungsreihe von 1976 bis 2012

		Studienberechtigtenjahrgänge ^a		
Befragungszeitpunkt	Thematischer Fokus	bis 1986 ^b	1990 bis 2004	2005 bis 2012
½ Jahr <i>vor</i> Schulabschluss	Ausbildungs- und Studi- en <i>absichten</i>	-	-	1. Welle
½ Jahr <i>nach</i> Schulabschluss	Ausbildungs- und Studien <i>entscheidungen</i>	1. Welle	1. Welle	2. Welle
2 ½ Jahre nach Schulab- schluss	verspätete oder korrigier- te Ausbildungs- und Studien <i>entscheidungen</i>	2. Welle	-	-
3 ½ bzw. 4 ½ Jahre nach Schulabschluss	Ausbildungs- und Studi- en <i>verläufe</i>	3. Welle	2. Welle	3. Welle ^c
10 ½ bis 20 ½ Jahre nach Schulabschluss	Erwerbseintritt und - verläufe	4. Welle	3. Welle	-

a Für einige Studienberechtigtenjahrgänge konnten nicht alle Wellen realisiert werden (siehe Abbildung 2).

Abbildung 2: Kohorten-Panel-Design der DZHW-Studienberechtigtenbefragungsreihe



Anmerkung: Die Befragungen der Jahrgänge 1976 bis 1986 richteten sich ausschließlich an Hochschulzugangsberechtigte aus den Bundesländern der damaligen Bundesrepublik Deutschland.

Quelle: Schneider und Franke (2014); eigene Überarbeitung

[Analysepotential] In allen Kohorten wird je Befragungswelle ein identischer Kern an Informationen erhoben. Auf dieser Grundlage können mit Zeitreihen- bzw. Kohortenvergleichen langfristige Trends

b Die Befragungen der Jahrgänge 1976 bis 1986 richteten sich ausschließlich an Hochschulzugangsberechtigte aus den Bundesländern der damaligen Bundesrepublik Deutschland.

c Für den Studienberechtigtenjahrgang 2012 konnte die 3. Welle erst 6 ½ Jahre nach Erwerb der Hochschulzugangsberechtigung realisiert werden.

im Übergang von der Schule in das weitere Ausbildungssystem und in den Arbeitsmarkt in den Blick genommen werden. Besonders hervorzuheben ist, dass in allen Befragungen von Studienberechtigtenjahrgängen, die mit mehr als einem halben Jahr Abstand nach Erwerb der Hochschulreife erfolgten, halbjahres- oder monatsgenaue kontinuierliche Verlaufsdaten zum individuellen Tätigkeitsverlauf seit dem Schulabschluss erfasst werden, die sich für Ereignisdaten- und Sequenzmusteranalysen eignen. Prinzipiell variiert innerhalb eines Befragungsjahrgangs der thematische Fokus von Welle zu Welle. Spezifische Fragen wurden jedoch wiederholt gestellt, sodass die Angaben derselben Befragten aus verschiedenen Wellen direkt miteinander verknüpft und intra-individuelle Veränderungen über den Zeitverlauf betrachtet werden können. Auf Basis des DZHW-Studienberechtigtenpanels können somit Querschnitts-, Trend-, Ereignisdaten-, Sequenzmuster- sowie eingeschränkt kausale Panelanalysen durchgeführt werden. Die Daten der Studienberechtigtenpanels werden in der Regel anhand der Merkmale Geschlecht, Bundesland, Art der Schule und der Hochschulreife gewichtet und an die Grundgesamtheit angeglichen. Zusätzlich werden paneltypische Ausfallprozesse in der Gewichtung der Daten berücksichtigt.

[Einordnung ins Forschungsfeld] Das Stichproben- und Erhebungsdesign sowie die damit verbundenen Analysemöglichkeiten unterscheiden das DZHW-Studienberechtigtenpanel von anderen in Deutschland durchgeführten Befragungen von Studienberechtigten. Keine andere Befragung ermöglicht bundesweite Analysen. Zusätzlich weisen andere Erhebungen in diesem Feld keine oder deutlich kürzere Zeitreihen auf. Andere Studienberechtigtenbefragungen sind beispielsweise die sächsische Abiturientenbefragung¹⁰ (durchgeführt seit 1996 alle zwei Jahre durch das Kompetenzzentrum für Bildungs- und Hochschulplanung, welches im Jahr 2015 in das Zentrum für Qualitätsanalyse (ZQA) der TU Dresden integriert wurde), die TOSCA-Studie¹¹ (inzwischen durchgeführt vom Hector-Institut für Empirische Bildungsforschung an der Universität Tübingen) sowie das Berliner-Studienberechtigtenpanel Best Up¹² (durchgeführt vom Deutschen Institut für Wirtschaftsforschung (DIW) und dem Wissenschaftszentrum Berlin (WZB)).

[Spezifika des Studienberechtigtenpanels 2012] Neben den allgemeinen Charakteristika der Studienreihe weist die hier betrachtete Befragung des Studienberechtigtenjahrgangs 2012 folgende Spezifika auf. So wurden - wie seit der Pilotstudie mit dem Studienberechtigtenjahrgang 2005 beziehungsweise der Befragung des Studienberechtigtenjahrgangs 2006 üblich – in einer ersten Welle bereits ein halbes Jahr vor Schulabschluss die Ausbildungs- und Studienabsichten in den Blick genommen. Wie vorangegangene Befragungen von Studienberechtigten fand auch die erste Befragung der Studienberechtigten des Jahres 2012, wenn möglich, im Klassenverbund und während des Unterrichts statt. Länderübergreifend wurde durch den Beschluss der Kultusministerkonferenz (2000) möglich, die Schulzeit an allgemeinbildenden Gymnasien von insgesamt neun Jahren auf acht Jahre zu verkürzen, was zu doppelten Abiturjahrgängen zwischen 2007-2016 führte. Beim Studienberechtigtenjahrgang 2012 traf das auf die Schüler*innen allgemeinbildender Gymnasien der Bundesländer Baden-Württemberg, Berlin, Brandenburg und Bremen zu. Aufgrund dessen liegt ein Schwerpunkt DZHW-Studienberechtigtenpanels 2012 auf dem Vergleich von G8-Schulabsolvent*innen der doppelten Abiturjahrgänge des Jahres 2012. Weitere besondere Themen sind das Informationsverhalten bei der Wahl des nachschulischen Werdegangs, der Einfluss des sozialen Umfeldes sowie kultureller Erfahrungen auf die Bildungsentscheidungen. Dem wurde durch die Integration neuer Befragungsinstrumente Rechnung getragen.

 $^{10}\ Vgl.\ dazu:\ https://tu-dresden.de/zqa/forschung/Forschungsprojekte/saechsische-abiturientenstudie$

-

 $^{^{11} \} V{gl.} \ dazu: \ https://uni-tuebingen.de/fakultaeten/wirtschafts-und-sozialwissenschaftliche-fakultaet/faecher/fachbereich-sozialwissenschaften/hector-institut-fuer-empirische-bildungsforschung/forschung/aktuelle-studien/tosca/$

¹² Vgl. dazu: http://www.best-up.eu/bestup.asp

4 Grundgesamtheit und Stichprobenverfahren

[Grundgesamtheit] Die Grundgesamtheit der Befragung des Studienberechtigtenjahrgangs 2012 umfasst alle Schülerinnen und Schüler der Bundesrepublik Deutschland, die im Schuljahr 2011/2012 an allgemeinbildenden oder berufsbildenden Schulen die allgemeine oder fachgebundene Hochschulreife, die Fachhochschulreife, die fach- oder landesgebundene Fachhochschulreife erwarben.

[Stichprobenverfahren] Aufgrund fehlender oder nicht zugänglicher Listen von Studienberechtigten auf Individualebene war es nicht möglich, eine einfache Zufallsstichprobe zu ziehen. So mussten die Individuen über die Schulen rekrutiert werden. Als Stichprobendesign wurde ein disproportional geschichtetes Klumpenstichprobenverfahren als Spezialfall einer mehrstufigen Zufallsauswahl (vgl. Häder, 2015, S. 169) gewählt. Die Schulen bzw. Schulzweige¹³ stellen dabei die "primary sampling units" (Klumpen), die Studienberechtigten des Jahrgangs 2012 die "secondary sampling units" innerhalb dieser Klumpen dar. 14 Die Auswahlgesamtheit bildete eine Liste mit allen relevanten Schulen bzw. Schulzweigen. Die Schichtung erfolgte anhand der Merkmale Bundesland und Schultyp. Innerhalb der Schichten wurden Ziehungsquoten für die Klumpen und Individuen festgelegt, die sich zum einen an der Verteilung des Studienberechtigtenjahrgangs 2008/2009 laut amtlicher Statistik¹⁵ orientierten und zum anderen die Rücklaufquoten der letzten Erhebung, dem Studienberechtigtenpanel 2010, berücksichtigten. Im Zuge dessen erhielten Studienberechtigte aller Schulzweige aus Bay-Mecklenburg-Vorpommern, Saarland, Sachsen-Anhalt überproportionale Auswahlwahrscheinlichkeiten, um in diesen Gruppen ausreichend große Fallzahlen zu generieren. Darüber hinaus wurde aus diesem Grund jeweils für die Schulformen Abendgymnasium, Abendkolleg, (Höhere) Berufsfachschulen und Fachschulen pro Bundesland (sofern diese Schulform vorhanden ist) eine Mindestanzahl von 50 Schüler*innen festgelegt.

Im Rahmen der praktischen Durchführung des Ziehungsverfahrens erfolgte zunächst in jeder Schicht die Auswahl der Klumpen (Schulen) über eine einfache Zufallsauswahl. Die ausgewählten Schulen meldeten im Folgenden ihre Schulzweige und die Schülerzahlen zurück, sodass ein zweiter Auswahlschritt durchgeführt werden konnte, der die Erfüllung der Ziehungsquoten auf Individualebene zum Ziel hatte. In diesem zweiten Schritt wurden in jeder Schicht sukzessive so viele Schulen/Schulzweige zufällig gezogen, dass mindestens 97 der schichtspezifischen Ziehungsquote für die Studienberech-

¹³ Manche berufsbildenden Schulen bestehen aus verschiedenen Schulzweigen, von denen jedoch nur jene für die Stichprobe relevant waren, die zum Erwerb der Hochschulzugangsberechtigung führten. Daher wurden in manchen Fällen nur bestimmte Schulzweige berücksichtigt.

¹⁴ Beispiel für den Personenkreis eines Klumpens: Alle Studienberechtigten des Prüfungsjahrgangs 2012 des allgemeinbildenden Gymnasiums XY in Bundesland YZ.
¹⁵ Es handelt sich um die zum Zeitpunkt der Stichprobenziehung aktuellsten veröffentlichten Daten des Statistischen Bundesamtes

¹⁵ Es handelt sich um die zum Zeitpunkt der Stichprobenziehung aktuellsten veröffentlichten Daten des Statistischen Bundesamtes (2010); (2011). Genutzt werden konnten Daten der amtlichen Statistik über die Anzahl der Absolvent*innen allgemeinbildender und berufsbildender Schulen: Allgemeinbildende Schulen - Schuljahr 2009/10 Fachserie 11 Reihe 1; S. 269-280 und Berufliche Schulen, Schuljahr 2009/10 Fachserie 11 Reihe 2; S. 53-107. Die Daten der Schulabsolvent*innen beider Publikationen beziehen sich jeweils auf das Schuljahr 2008/09. Außerdem musste berücksichtigt werden, dass im Schuljahr 2011/2012 Baden-Württemberg, Berlin, Brandenburg und Bremen doppelte Abiturient*innenjahrgänge an allgemeinbildenden Gymnasien haben werden, was auf das Schuljahr 2008/2009 nicht zutraf.

tigten erreicht wurde. ¹⁶ Hiernach wurden die gezogenen Schulen (oder Schulzweige an diesen) komplett erhoben.

¹⁶ Die Prüfung, ob die Ziehungsquote in einer Schicht erreicht wurde, erfolgte nach jedem Ziehungsschritt. Erfüllte beispielsweise die kumulierte Schülerzahl von fünf Schulen die Ziehungsquote für die Schülerzahl in der spezifischen Schicht, so wurde nach der Auswahl der fünften Schule keine weitere Schule gezogen.

5 Gewichtung

Die Gewichtung der Daten dient dem Ausgleich von Verzerrungen der Stichprobe im Vergleich zur definierten Grundgesamtheit. Es folgt zunächst eine allgemeinere Einführung in die Vorgehensweise und die Darstellung der erstellten Gewichte. Im Anschluss wird die Gewichtungsprozedur im Detail beschrieben.

5.1 Vorgehen und Anwendungshinweise

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

Designbedingte Verzerrung: Disproportionalitäten werden bewusst erzeugt, um in bestimmten relevanten Subgruppen die Fallzahlen zu erhöhen.

Verzerrung durch Nonresponse: Ausfallprozesse (z. B. Nichtteilnahmen, fehlende Erreichbarkeit, Verlust auf dem Postweg) führen zu einem verringerten Rücklauf und somit zu einer Differenz zwischen Brutto- und Netto-Stichprobe. Wenn diese Ausfallprozesse unsystematisch sind (Missing Completely at Random), können sie ignoriert werden.¹⁷ Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (Missing at Random, Not Missing at Random), der einer Modellierung bedarf.¹⁸

[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 Querund 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).

Vor dem Hintergrund der Erläuterungen zur Grundgesamtheit und zum Stichprobenverfahren in Kapitel 4 wird ersichtlich, dass ein derartiges idealtypisches Vorgehen für die Daten des Studienberechtigtenpanels 2012 nicht durchführbar ist. Das Stichprobendesign ermöglicht keine Ableitung von exakten Auswahlwahrscheinlichkeiten. Da in der ersten Befragungswelle zudem keine Informationen über Nichtteilnehmer*innen vorliegen, können hier auch keine individuellen Ausfallgewichte gebildet werden.

Als Gewichte für die erste Welle werden ein geschätztes Designgewicht sowie Querschnittsgewichte berechnet. Das *Designgewicht* ergibt sich bei zufallsgesteuerten Auswahlverfahren direkt aus dem Stichprobendesign. Die *Querschnittsgewichte* basieren auf einer Zellgewichtung mit der die Daten an

_

¹⁷ Insofern die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

¹⁸ Siehe grundlegend zu den unterschiedlichen Formen von Ausfallprozessen Rubin (1976).

ausgewählte Merkmale der Grundgesamtheit¹⁹ (Geschlecht, Bundesland, Art der Schule, Art der Hochschulreife, G8G9) angepasst wurden. Diejenigen Zellen, welche (vorab definierte) Ausreißerwerte in den Gewichten (<0,3 und >4) enthalten, werden mittels Zellverschmelzungen nivelliert. Aus der Merkmalsverteilung der Grundgesamtheit werden zugleich Informationen zu Nichtteilnehmer*innen gewonnen, sodass hiermit ebenfalls eine Form der Nonresponse-Adjustierung erfolgt. In Tabelle 2 sind die im Datensatz bereitgestellten Gewichte dargestellt.

Tabelle 2: Bereitgestellte Gewichte zum DZHW-Studienberechtigtenpanel 2012

Variablenname	Beschreibung
gew1	Querschnittgewicht Welle 1 (jedes Bundesland gesondert + G8-Gymnasien separat)
gew2	Querschnittgewicht Welle 1 (bundesweites Gewicht mit Länderrelationen + G8-Gymnasien separat)
gewao	Längsschnittgewicht 2-Wellen-Panel (jedes Bundesland gesondert + G8- Gymnasien separat; ohne Studierneigung)
gewbo	Längsschnittgewicht 2-Wellen-Panel (bundesweites Gewicht mit Länderrelationen G8-Gymnasien separat; ohne Studierneigung)

[Hinweise zur Anwendung der Gewichte] Bei den erstellten Gewichten handelt es sich um probability weights, die in Stata mit Hilfe ado-spezifischer Optionen berücksichtigt werden können.²⁰ Die Gewichte gew1 und gew2 sind für Auswertungen der ersten Welle vorgesehen, gew1 für Auswertungen einzelner Bundesländer und gew2 für Bundesauswertungen. Die Gewichte gewao und gewbo sind für Auswertungen des 2-Wellen-Panels konzipiert, gewao für Auswertungen einzelner Bundesländer und gewbo für Bundesauswertungen.

Grundlegend ist zu beachten, dass Gewichte nur für jene Analysen sinnvolle Korrekturgrößen darstellen, in denen die zur Gewichtung verwendeten Variablen in dem Analysemodell enthalten sind oder mit diesen in einem Zusammenhang stehen. Aus diesem Grund sollten Gewichte immer mit Bedacht und Fokus auf die analysierte Fragestellung verwendet werden.

5.2 Gewichtung des Datensatzes

Im Folgenden wird die Vorgehensweise bei der Erstellung der Gewichte für den Datensatz näher dargestellt.

[Designgewichtung] Die Berechnung der Designgewichte ergibt sich im vorliegenden Studiendesign direkt aus dem Stichprobenplan. Für jede Schicht wird ein spezifisches Designgewicht bestimmt,

¹⁹ Genutzt wurden die zum Zeitpunkt der Stichprobenziehung aktuellsten veröffentlichten Daten des Statistischen Bundesamtes (2010); (2011) über die Anzahl der Absolvent*innen allgemeinbildender und berufsbildender Schulen: Allgemeinbildende Schulen - Schuljahr 2009/10 Fachserie 11 Reihe 1 und Berufliche Schulen - Schuljahr 2009/10 Fachserie 11 Reihe 2. Die Daten der Schulabsolvent*innen beider Publikationen beziehen sich jeweils auf das Schuljahr 2008/09.

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

wobei die Schichten sich anhand der Merkmale Bundesland und Schulart²¹ ergeben (vgl. Kapitel 4). Benötigt werden für jede Schicht die folgenden Informationen aus dem Stichprobenplan:

- Anzahl der Schulen bzw. Schulzweige in der Grundgesamt
- Anzahl der gezogenen Schulen bzw. Schulzweige je Schicht

Anhand dieser Informationen wird für jede Schicht die spezifische *Auswahlwahrscheinlichkeit* π bestimmt. Diese ergibt sich aus der Formel:

$$\pi = \frac{\text{n(Stichprobe)}}{\text{N(Grundgesamtheit)}}$$

Das Designgewicht folgt schließlich als Kehrwert der Auswahlwahrscheinlichkeit:

 $\frac{1}{\pi}$

Die Designgewichte werden abschließend noch auf die Stichprobengröße normiert, indem die jeweiligen Gewichte durch den Mittelwert aller Gewichte geteilt werden.

[Querschnittsgewichtung] Für die erste Welle wurden mit einer Zellgewichtung Querschnittsgewichte erstellt. Hierbei wurden die Daten anhand der Merkmale Geschlecht, Schulgruppen²², G8G9 und Bundesland an die Grundgesamtheit angepasst, um mit Blick auf diese Merkmale Verzerrungen in der realisierten Stichprobe gegenüber der Verteilung in der Grundgesamtheit auszugleichen.²³ Diese Verzerrungen können sowohl aus Unit- oder Item-Nonresponse resultieren als auch aus dem disproportionalen Stichprobendesign.

Für jede Zelle der Gewichtungsmatrix wurde anhand der Sollzahlen zunächst der Anteil bestimmt, den diese Zelle in der Grundgesamtheit repräsentiert (pN). Für die realisierte Stichprobe wurde ebenfalls eine solche Quote gebildet (pn=Anteil realisierte Fälle in der jeweiligen Zelle geteilt durch Gesamtfallzahl in der Stichprobe). Der Gewichtungsfaktor ergibt sich, indem aus beiden Anteilswerten ein Quotient gebildet wird (pN/pn). Dieses Vorgehen wurde zweimal durchgeführt. Beim Gewicht gew1 wurden die Gewichtungsfaktoren für jedes Land gesondert berechnet. Bezugsgruppe für die Bestimmung der Anteilswerte pN und pn war dann die Fallzahl in Grundgesamtheit bzw. Stichprobe im jeweiligen Bundesland. Das Gewicht gew2 berücksichtigt zusätzlich die Relationen der Länder untereinander, da die Anteilswerte hier bezogen auf die bundesweiten Fallzahlen bestimmt wurden.

Für die zweite Erhebungswelle wurde kein Querschnittsgewicht erstellt, da keine neu hinzugekommenen Individuen befragt wurden (Auffrischungsstichprobe).

[Längsschnittgewichtung] Die Gewichte *gewao* und *gewbo* des Zwei-Wellen-Panels wurden ebenfalls mit einer Zellgewichtung erstellt. Hierbei wurden die Daten anhand der Merkmale Geschlecht, Schulgruppen²⁴, G8G9 und Bundesland an die Grundgesamtheit angepasst, um mit Blick auf diese Merkmale Verzerrungen in der realisierten Stichprobe gegenüber der Verteilung in der Grundge-

²¹ Hierbei handelt es sich um die neun Schularten, die auch maßgeblich für die Stichprobenziehung waren (vgl. Kapitel 4).

²² Folgende Schulgruppen wurden unterschieden: (1) Gymnasium/Gesamtschule/Waldorfschule, (2) Abendgymnasium/Kolleg, (3) Fachgymnasium/andere berufsbildende Schulen mit Erwerb der allgemeinen Hochschulreife, (4) Fach- und Berufsoberschulen (mit Erwerb der Fachhochschulreife), (5) sonstige berufsbildende Schulen (mit Erwerb der Fachhochschulreife).

²³ Grundlage bildeten die Daten des Statistischen Bundesamtes (2013); (2014) zu Schulabsolvent*innen mit Studienberechtigung des Schuljahres 2011/12. Diese wurden der entsprechenden Fachserie 11 des Schuljahres 2012/13 (Reihe 1 für allgemeinbildende Schulen sowie Fachserie 11 Reihe 2 für berufsbildende Schulen) entnommen.

²⁴ Folgende Schulgruppen wurden unterschieden: (1) Gymnasium/Gesamtschule/Waldorfschule, (2) Abendgymnasium/Kolleg, (3) Fachgymnasium/andere berufsbildende Schulen mit Erwerb der allgemeinen Hochschulreife, (4) Fach- und Berufsoberschulen (mit Erwerb der Fachhochschulreife), (5) sonstige berufsbildende Schulen (mit Erwerb der Fachhochschulreife).

samtheit auszugleichen.²⁵ Diese Verzerrungen können sowohl aus Unit- oder Item-Nonresponse resultieren als auch aus dem disproportionalen Stichprobendesign.

Für jede Zelle der Gewichtungsmatrix wurde anhand der Sollzahlen zunächst der Anteil bestimmt, den diese Zelle in der Grundgesamtheit repräsentiert (pN). Für die realisierte Stichprobe wurde ebenfalls eine solche Quote gebildet (pn=Anteil realisierte Fälle in der jeweiligen Zelle geteilt durch Gesamtfallzahl in der Stichprobe). Der Gewichtungsfaktor ergibt sich, indem aus beiden Anteilswerten ein Quotient gebildet wird (pN/pn). Dieses Vorgehen wurde zweimal durchgeführt. Beim Gewicht *gewao* wurden die Gewichtungsfaktoren für jedes Land gesondert berechnet. Bezugsgruppe für die Bestimmung der Anteilswerte pN und pn war dann die Fallzahl in Grundgesamtheit bzw. Stichprobe im jeweiligen Bundesland. Das Gewicht *gewbo* berücksichtigt zusätzlich die Relationen der Länder untereinander, da die Anteilswerte hier bezogen auf die bundesweiten Fallzahlen bestimmt wurden.

²⁵ Grundlage bildeten die Daten des Statistischen Bundesamtes (2013); (2014) zu Schulabsolvent*innen mit Studienberechtigung des Schuljahres 2011/12. Diese wurden der entsprechenden Fachserie 11 des Schuljahres 2012/13 (Reihe 1 für allgemeinbildende Schulen sowie Fachserie 11 Reihe 2 für berufsbildende Schulen) entnommen.

-

6 Anonymisierung

[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten²⁶, 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 2017.²⁷ 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.

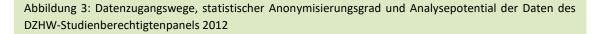
[Datenzugang, Anonymisierungsgrad und Analysepotential] Das FDZ des DZHW stellt für das Studienberechtigtenpanel 2012 ein SUF für die wissenschaftliche Sekundärnutzung zur Verfügung. Die Anonymität der Befragten wird dabei über eine Kombination aus statistischen Maßnahmen und technischen Zugriffsbeschränkungen 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.

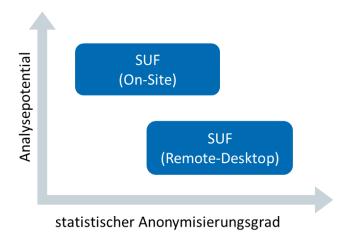
Das SUF wird über zwei verschiedene Zugangswege angeboten: Remote-Desktop und On-Site (für weiterführende Informationen vgl. Kapitel 2). Für jeden Zugangsweg wird eine andere SUF-Variante bereitgestellt, die unterschiedlich stark anonymisiert worden ist und entsprechend weniger oder mehr Informationen 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 vom Zugangsweg erläutert.

²⁶ "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 Kennummer, 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).

²⁷ Die DSGVO gilt grundsätzlich innerhalb der EU und somit ebenfalls für das DZHW. Das BDSG in seiner Neufassung vom 30. Juni

[&]quot;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 (Datenschutzanpassungs- und Umsetzungsgesetz EU DSAnpUG-EU)) kommt teils zusätzlich 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.





[Statistische Anonymisierungsmaßnahmen] Im Rahmen der Anonymisierung sind zunächst alle Informationen, mit denen sich Personen oder Institutionen direkt identifizieren lassen, zu löschen. Diese sogenannten direkten Identifikatoren, wie Namen, Adressen oder E-Mail-Adressen, wurden im Rahmen des Studienberechtigtenpanels 2012 bereits während der Feldphase in einem separaten Datensatz erfasst und nur genutzt, um eine Kontaktaufnahme zu ermöglichen. Diese Angaben sind somit nicht in den verschiedenen SUF-Varianten enthalten. Des Weiteren wurde, um einen Rückbezug auf die Originaldaten zu verhindern, die Original-Identifikationsnummer 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. ²⁸ Für das Studienberechtigtenpanel 2012 wurden beispielsweise folgende Quasi-Identifikatoren identifiziert: regionale Informationen (Geburtsland, Hochschule oder Arbeitsort), Staatsangehörigkeit, Sprache im Elternhaus, Schulart, schulische Prüfungsfächer, Hochschule, Studienfach, Abschlussart, Berufsangaben. Um eine eindeutige Zuordnung der Studienberechtigtendaten zu unterbinden, wurden diese Schlüsselmerkmale – je nach Zugangsweg – aggregiert oder gelöscht (vgl. Tabelle 3). Beispielsweise ist das Merkmal Studienfach im SUF für die On-Site-Nutzung uneingeschränkt verfügbar. Im Remote-Desktop-SUF hingegen wird das Merkmal zu Studienbereichen aggregiert.

Darüber hinaus empfehlen Ebel und Meyermann offene Angaben zu löschen "selbst wenn die jeweiligen Fragestellungen an sich unproblematisch sind. Denn es besteht die Gefahr, dass Studienteilnehmer*innen bei eigentlich unbedenklichen Fragen mit offener Antwortmöglichkeit kritische Informationen preisgegeben haben, die zu einer Identifikation führen könnten" (Ebel & Meyermann, 2015, S. 5). Die offenen Angaben wurden größtenteils bereits im Rahmen der Datenaufbereitung durch das Primärforschungsprojekt vercodet und werden daher in allen SUF-Varianten zur Verfügung gestellt. Teilweise wurden jedoch – in Abhängigkeit von der Sensibilität der enthaltenen Informationen und vom Zugangsweg – die vom Primärforschungsprojekt vorgenommenen Codierungen zusätzlich aggregiert. Nicht codierte offene Angaben wurden in allen SUF-Varianten 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 notwendig zur Re-Identifikation von Individuen oder Institutionen, jedoch können die Informationen im Falle

_

²⁸ 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.

einer De-Anonymisierung nutzbringend sein (Koberg, 2016, S. 694) und sind daher besonders schützenswert (Art. 9 DSGVO, Erwägungsgrund 51 DSGVO). Im Studienberechtigtenpanel 2012 wurden Gesundheitsinformationen erhoben, für die bei den Befragten kein zusätzliches Einverständnis für die Sekundärnutzung eingeholt wurde. Daher wurden diese Antworten in allen SUF-Varianten gelöscht. Die nachfolgende Tabelle 3 stellt in Kurzform die durchgeführten statistischen Anonymisierungsmaßnahmen je nach Zugangsweg dar. Variablen, die in den beiden SUF-Varianten aus Datenschutzgründen nicht verfügbar sind, enden mit dem Suffix "_a".

Tabelle 3: Maßnahmen der statistischen Anonymisierung der Daten des DZHW-Studienberechtigenpanels 2012 nach Zugangsweg²⁹

Merkmal	On-Site-SUF	Remote-Desktop-SUF
Direkte Identifikatoren	Löschung	Löschung
Original-ID (Befragte und Schulen)	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID
Schulart	Aggregation von geringbesetzten Schularten	Aggregation zu allgemeinbildenden und beruflichen Schulen
Prüfungsfächer/Schwerpunktfach (Schule, Berufsschule)	Freigabe	Aggregation zu 15 Fächergruppen
Voraussichtliches Studienfach	Freigabe	Freigabe
Studienfach	Freigabe	Aggregation zu Studienbereichen ^a
Hochschule	Information zur Hochschulart	Information zur Hochschulart ^b
Hochschulort/Aufenthaltsort im Dezember 2012	Aggregation zum Bundesland oder Bundeslandgruppen und Ausland	Aggregation zum Bundesland oder Bundeslandgruppen und Ausland
Angestrebter Studienabschluss	Aggregation von geringbesetzten Abschlussarten	Aggregation von geringbesetzten Abschlussarten
Beruf (Praktikumsberuf, (voraussichtlicher) Ausbildungsberuf, berufliche Tätigkeit, Berufe der Eltern)	Freigabe	Aggregation zu Berufsordnungen ^c
Staatsangehörigkeit, Geburts- land, Geburtsland der Eltern	Aggregation zu maximal 30 Kategorien	7 Staaten einzeln ausgewiesen; weitere Staaten zu Weltregionen ^d
Sprache im Elternhaus	Erste Sprache: 20 Sprachen einzeln ausgewiesen; andere Sprachen zu "Sonstiges" Zweite Sprache: 3 Sprachen einzeln ausgewiesen; andere Sprachen zu "Sonstiges" Dritte Sprache: nur "genannt"/"nicht genannt"	Erste Sprache: 10 Sprachen einzeln ausgewiesen; andere Sprachen zu "Sonstiges" Zweite Sprache: 3 Sprachen einzeln ausgewiesen; andere Sprachen zu "Sonstiges" Dritte Sprache: nur "genannt"/"nicht genannt"
Merkmale zur Gesundheit	Löschung	Löschung

²⁹ Detaillierte Informationen zu den anonymisierten Variablen sind dem Datensatzreport sowie dem Metadatensuchsystem (https://metadata.fdz.dzhw.eu) zu entnehmen.

_

Anzahl der Geschwister	Aggregation zu "1", "2" und "3 und mehr"	Löschung
Alter der Geschwister	Aggregation zu "älter", "gleichalt" oder "jünger"	Aggregation zu "älter", "gleichalt" oder "jünger"
Probleme bei Wahl des nachschulischen Werdegangs	Aggregation: Zusammenfassung der Gründe "Gesundheitsprobleme" und "Sonstiges"	Aggregation: Zusammenfassung der Gründe "Gesundheitsprobleme" und "Sonstiges"
Gründe, die gegen Studium gesprochen haben	Aggregation: Zusammenfassung der Gründe "Krankheit" und "Sonstiges"	Aggregation: Zusammenfassung der Gründe "Krankheit" und "Sonstiges"
Bedeutung für gewählten Werdegang: Gesundheit	Löschung	Löschung

a Nach Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2011/12 von Destatis.

b Nur Unterscheidung von Fachhochschule vs. Universität (inklusive Pädagogische Hochschulen, Theologische Hochschulen, Kunstund Musikhochschulen und Verwaltungsfachhochschulen).

c Nach Klassifikation der Berufe von 1992 von Destatis.

d Nach Klassifikation der Vereinten Nationen.

7 Literatur

- Ebel, T. & Meyermann, A. (2015). *Hinweise zur Anonymisierung von quantitativen Daten* (forschungsdaten bildung informiert Nr. 3). Verbund Forschungsdaten Bildung. Verfügbar unter https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert_nr-7.pdf
- Häder, M. (2015). *Empirische Sozialforschung. Eine Einführung* (3. Aufl.). Wiesbaden: Springer Fachmedien. doi:10.1007/978-3-531-19675-6
- Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. (2012). *Datenschutz am Forschungsdatenzent-rum* (FDZ-Methodenreport Nr. 6). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. von 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
- Kultusministerkonferenz. (2000). Vereinbarung zur Gestaltung der gymnasialen Oberstufe in der Sekundarstufe II. (Beschluss der Kultusministerkonferenz vom 07.07.1972 idF vom 16.06.2000).
- Lane, J., Heus, P. & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. *Transactions on Data Privacy*, 1(1), pp. 2–16.
- Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(2), 581-592.
- Schneider, H. & Franke, B. (2014). *Bildungsentscheidungen von Studienberechtigten. Studienberechtigte 2012 ein halbes Jahr vor und ein halbes Jahr nach Schulabschluss* (Forum Hochschule Nr. 6). Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).
- Statistisches Bundesamt. (2010). *Bildung und Kultur. Berufliche Schulen*. Schuljahr 2009/2010 (Fachserie 11 Reihe 2). Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2011). *Bildung und Kultur. Allgemeinbildende Schulen*. Schuljahr 2009/2010 (Fachserie 11 Reihe 1). Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2013). *Bildung und Kultur. Allgemeinbildende Schulen*. Schuljahr 2012/2013 (Fachserie 11 Reihe 1). Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2014). *Bildung und Kultur. Berufliche Schulen*. Schuljahr 2012/2013 (Fachserie 11 Reihe 2). Wiesbaden: Statistisches Bundesamt.