

BIBB-FDZ
Daten- und Methodenberichte

Nr. 1 / 2010

BIBB-Übergangsstudie 2006

Version 1.1

Daniela Rohrbach-Schmidt

Bundesinstitut für Berufsbildung
- Forschungsdatenzentrum -
Robert-Schuman-Platz 3
53175 Bonn

Telefon: 0228 / 107 - 20 41
Fax: 0228 / 107 - 20 20
E-Mail: fdz@bibb.de

www.bibb-fdz.de



1	Einleitung	4
2	Die BIBB-Übergangsstudie 2006 im Überblick	4
2.1	Kurzbeschreibung und Aufbau der BIBB-Übergangsstudie 2006	4
2.2	Zentrale Eigenschaften der BIBB-Übergangsstudie 2006 im Überblick	6
3	Inhalte der Erhebung	7
3.1	Warm-up und Grundschule	8
3.2	Bildungswege und Berufsbiographie	8
3.3	Soziodemographische Angaben und Grundeinstellungen	9
3.4	Themen in der Zusammenschau	9
4	Datenerhebung	10
4.1	Grundgesamtheit und Stichprobenziehung	10
4.2	Realisierung der Stichprobe	12
4.3	Repräsentativität und Selektivität	16
5	Datenaufbereitung	17
5.1	Querschnittsatz	18
5.2	Ereignisdatensatz	22
5.3	Textvariablen	26
6	Datenzugang und Anonymisierung	29
6.1	Datenzugang	29
6.2	Bestimmungen des Datenschutzes	29
6.3	Anonymisierungen nach Datenprodukten	30
6.3.1	Anonymisierung: SUF	30
6.3.2	Formale Anonymisierung: Datenfernverarbeitung	31
6.3.2	Projektbezogene Anonymisierung: Gastrechnen im BIBB	32
7	Arbeitshilfen	32
7.1	Gewichtung	33
7.2	Deklarierung der fehlenden Werte nach dem BIBB-FDZ-Standard	34
7.3	Arbeiten mit dem Ereignisdaten und Zusammenführung mit den Querschnittinformationen (Holger Alda)	36
7.4	Visualisierungsprogramm	38
7.5	Thematisches und methodisches Forschungspotential	39
8	Literatur	42
9	Anhang	44
9.1	Ausfallanalyse für Kreiskennziffern	44
9.2	Übersicht über Zusatzvariablen	45
9.3	Beispielprogramm zur Auswertung der Übergangsstudie	50

¹ Wissenschaftliche Mitarbeiterin am Forschungsdatenzentrum im Bundesinstitut für Berufsbildung (BIBB-FDZ). Kontakt: rohrbach@bibb.de

Danksagung

Die Autorin dankt den Projektmitarbeitern der BIBB-Übergangsstudie 2006 im BIBB, Ursula Beicht, Michael Friedrich, und Joachim G. Ulrich für ihre wertvolle Hilfe bei der Erstellung des Handbuches.

Datenverfügbarkeit

Der in diesem Handbuch beschriebene Datensatz kann als Scientific Use File über unsere Internetseite www.bibb-fdz.de beantragt werden.

Hinweis zur Zählweise bei Versionsnummern

Änderungen gegenüber der Vorversion ohne größere inhaltliche Relevanz werden durch fortlaufende Nummern *nach* dem Punkt dokumentiert (zweite Ebene). Inhaltlich relevante Änderungen führen demgegenüber zu einer fortlaufenden Nummerierung auf der ersten Ebene.

1 Einleitung

Gegenstand des vorliegenden Daten- und Methodenberichts ist die detaillierte Beschreibung des Datensatzes "Bildungswege und Berufsbiografie von Jugendlichen und jungen Erwachsenen im Anschluss an allgemeinbildende Schulen" aus dem Jahr 2006 (Kurztitel: BIBB-Übergangsstudie 2006).

Die Daten wurden von infas, Institut für angewandte Sozialforschung (Bonn), im Auftrag des Bundesinstituts für Berufsbildung (BIBB) erhoben. Das Forschungsdatenzentrum des BIBB (BIBB-FDZ) hat weitere Aufbereitungsschritte an den Daten vorgenommen, die im vorliegenden Handbuch dokumentiert sind. Für die externe nicht-kommerzielle Forschung bestehen Datenzugangsmöglichkeiten über das BIBB-FDZ mittels der off-Site Nutzung eines Scientific-Use-Files, über Datenfernverarbeitung und / oder Gastaufenthalte. Ziel dieses Handbuches ist es, externen Wissenschaftlern eine Arbeitshilfe bei der Auswertung der im Bundesinstitut für Berufsbildung (BIBB) erhobenen und im Forschungsdatenzentrum im BIBB (BIBB-FDZ) aufbereiteten Datensätze zu sein.

Das Handbuch ist wie folgt gegliedert: Im nächsten Kapitel wird zunächst ein Überblick über die zentralen Merkmale der BIBB-Übergangsstudie 2006 gegeben. Kapitel 3 stellt die Inhalte der Befragung vor. Kapitel vier widmet sich den Details der Datenerhebung, bevor im fünften Kapitel die Datenaufbereitung des Querschnitt-, Ereignis- und Textvariablendatensatzes beschrieben wird. Anschließend wird aufgezeigt, in welcher Weise Anonymisierungen der Daten für die jeweiligen Zugangswege vorgenommen wurden. Kapitel sieben erläutert abschließend die Arbeitshilfen zu der BIBB-Übergangsstudie 2006.

2 Die BIBB-Übergangsstudie 2006 im Überblick

2.1 Kurbeschreibung und Aufbau der BIBB-Übergangsstudie 2006

Vor dem Hintergrund eines sich verschärfenden Ausbildungsstellenmarktes, einer Zunahme in der Schwierigkeit, Komplexität und z.T. auch Länge der Übergangsprozesse von der allgemeinbildenden Schule in die berufliche Bildung sowie dem Fehlen einer hinreichend detaillierten Datenquelle (vgl. Beicht und Friedrich 2008: 79 ff.) wurde die BIBB-Übergangsstudie 2006 im Bundesinstitut für Berufsbildung mit dem Ziel initiiert, umfassende Informationen zum Bildungs- und Berufsverlauf von Jugendlichen und jungen Erwachsenen in Deutschland zu gewinnen.

In dieser Studie wurden 7.230 Jugendliche der Geburtsjahrgänge 1982 bis 1988 befragt. Die Datenerhebung erfolgte von Juni bis August 2006 mittels computergestützter Telefoninterviews durch infas, Bonn. Es handelt sich um eine retrospektive Längsschnittdatenerhebung, in der die gesamte Bildungs- und Berufsbiografie beginnend mit der allgemeinbildenden Schulzeit erfasst wurde. Die Stichprobenziehung erfolgte auf Basis einer Telefonauswahlgrundlage, die neben den im Telefonverzeichnis gelisteten

Anschlüssen zufällig generierte Telefonnummern enthält, um auch nicht eingetragene Telefonanschlüsse zu erreichen (vgl. Gabler und Häder, 1998). Die Daten wurden durch Gewichtung auf Grundlage der amtlichen Statistik an die Strukturen der Grundgesamtheit angepasst. Die Ergebnisse sind für Jugendliche in den entsprechenden Altersgruppen in Deutschland repräsentativ.

Das BIBB-FDZ bietet drei separate Bestandteile der Daten der BIBB-Übergangsstudie 2006 an: Erstens einen reinen *Querschnitt Datensatz*, der vor allem die Informationen enthält, die über die einzelnen erfassten Bildungs- und Berufsetappen konstant sind (u.a. Soziodemographie, Wertvorstellungen, Grundschulinformationen, ehrenamtliche Tätigkeiten, persönliche Grundeinstellungen und Selbstwirksamkeitserwartungen, soziale und ethnische Herkunft). Zweitens wird ein *Ereignisdatensatz* angeboten, der für jeden der 7.230 Jugendlichen im Mittel vier, insgesamt 27.965, detaillierte biographische Informationen zu Bildungs- und Berufsaktivitäten (u.a. Schule, Ausbildung, Studium, Berufstätigkeit) – sowie sonstigen Aktivitäten (Zivildienst, freiwilliges soziales oder ökologisches Jahr, Suche nach Ausbildungs- oder Arbeitsstelle, Kinderbetreuung) enthält. Dieser Ereignisdatensatz ist im sogenannten ‚langen Format‘ aufgebaut, d.h. die Längsschnittinformationen sind zeilenweise abgelegt. Neben einer Personen-ID (systemfreies Kennzeichen) enthält der Datensatz für jede Person mindestens eine, in der Regel jedoch mehrere Spell-ID bzw. die zu jeder erfassten Aktivität erhobenen Variablen. Tabelle 1 gibt einen schematischen Überblick über die Datenorganisation des Ereignisdatensatzes.

Tabelle 1: Datenorganisation des Ereignisdatensatzes der BIBB-Übergangsstudie 2006 (langes Format)

Personen-ID	Spell-ID	Schulbesuch	Berufliche Ausbildung	...
1	1	x
1	2	.	x	...
1	3
2	1	x
2	2	.	x	...
...
$N_i=7.230$	$N_t=27.965$ mit $t \geq 1 < 19$			

Drittens werden jeweils die Textvariablen des Querschnitt –und Längsschnittteils separat und aus Datenschutzgründen bzw. eines höheren Reidentifikationsrisikos nur für Analysen im Gastwissenschaftleraufenthalt angeboten.

Über den Zugang und Informationen zu der BIBB-Übergangsstudie 2006 hinaus, stellt das BIBB-FDZ das Visualisierungsprogramm *LDEX, Visual Explorer for Longitudinal Data* für die BIBB-Übergangsstudie 2006 im Rahmen eines Gastwissenschaftleraufenthalts im BIBB zur Verfügung (s. Abschnitt 7.3). Das Programm wurde von den Projektverantwortlichen im BIBB in Auftrag gegeben und freundlicherweise für die Weiterverwendung frei gegeben.

2.2 Zentrale Eigenschaften der BIBB-Übergangsstudie 2006 im Überblick

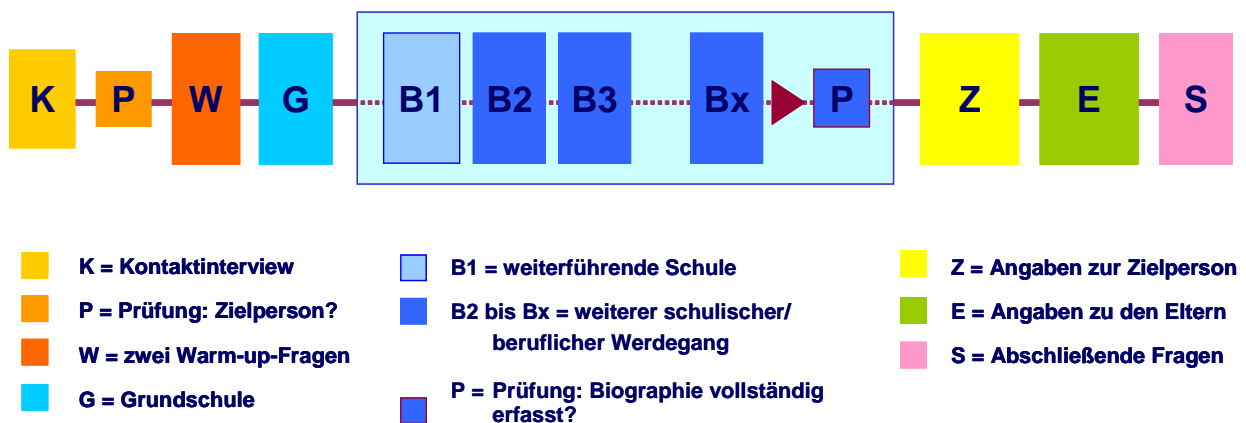
Befragungstitel	BIBB-Übergangsstudie 2006
Kurzbeschreibung	Studie zu Bildungswegen und Berufsbiographien von jungen Erwachsenen der Geburtsjahrgänge 1982 bis 1988 im Jahr 2006. Das FDZ bietet einen Querschnitt- und einen Ereignisdatensatz der BIBB-Übergangsstudie 2006 an.
Erhebungsdesign	Längsschnittdesign
Erhebungsjahr	2006
Befragungseinheit primärer Lebenszyklus	Jugendliche Übergang Schule zu Ausbildung
Stichworte	Bildungswege; Berufsbiographie; Schulleistung; Schulwahl; Berufswahl; Familiärer Hintergrund; Bewerbungsverhalten der Jugendlichen; erste Schwelle; berufliche Erstausbildung; Bildungs- und Warteschleifen; Übergangssystem; Migrationshintergrund
Zugangswege	SUF, DFV, GWA
Rücklauf	7230
Nettozahl	7961 (246579 (bereinigte Einsatzstichprobe), 200039 Kontakt zum Haushalt, Screening durchgeführt, 7961 Zielperson im Haushalt vorhanden)
Grundgesamtheit	Alle Jugendlichen und jungen Erwachsenen der Geburtsjahrgänge 1982 bis 1988, die zum Befragungszeitpunkt ihren Wohnsitz in Deutschland hatten
Auswahlverfahren	ADM-Telefonstichprobendesign; Last-Birthday-Verfahren
Gewichtung-Hochrechnung	Designgewichte (Haushaltstransformation); Populationsgewichte (Geschlecht, Geburtsjahrgang, höchster Schulabschluss, Bundesland)

Forschungsfragen	Größenordnung und Trendhypothesen zur relativen Häufigkeit von verschiedenen Bildungswegen; Persönliche, familiäre und schulische Einflussfaktoren auf die Bildungs- und Berufsverläufe von Jugendlichen
Datenaufbereitung	Anonymisierungen, (Re-)Klassifizierungen, Missing Values (FDZ-Standard), Datenbereinigungen
Verfahren	CATI
Repräsentative Region	Raumordnungsregion

3 Inhalte der Erhebung

Die Befragung erfasst die gesamte Bildungs- und Berufsbiographie von Personen der Geburtsjahrgänge 1982 bis 1988 zum Befragungszeitpunkt im Jahr 2006 und enthält somit Angaben von Jugendlichen und jungen Erwachsenen im Alter zwischen 18 und 24 Jahren (vgl. Abbildung 1).

Abbildung 1: Schematische Darstellung des Ablaufs der CATI-Interviews
(entnommen aus Beicht und Friedrich 2008, S. 84)



3.1 Warm-up und Grundschule

Den Beginn der eigentlichen Befragung² bilden Fragen zum Alter (Geburtsmonat und -jahr), individuellen Wertorientierungen (Wichtigkeit bestimmter Lebenseinstellungen) und einer Frage danach, ob sich die Befragten eher als Jugendliche oder als Erwachsene fühlen. Im Anschluss wurde nach dem Beginn und dem Ende des Grundschulbesuchs gefragt und erfasst, wie gerne die Befragten in die Grundschule gegangen sind. Das Geschlecht wurde zu Beginn des Interviews durch den Interviewer eingegeben.

3.2 Bildungswege und Berufsbiographie

Den Kern der Erhebung stellt der sich anschließende längsschnittliche Teil dar, der die Bildungs- und Berufsbiographie im Anschluss an die Grundschule zeitlich und inhaltlich sehr genau erfasst (vgl. Beicht und Friedrich 2008, S. 84-86; s.a. Abschnitt 4).

Hierfür wurden jeweils zehn Oberkategorien vorgelesen, denen die Befragten die Aktivitäten zuordnen mussten (Fragenblock 2000, siehe Fragebogen). Diese zehn Oberkategorien umfassen die wesentlichen Bereiche der schulisch-beruflichen Aktivitäten von Jugendlichen:

1. Schulbesuch³ (Frageblock 2100)
2. Berufliche Ausbildung, Lehre (Frageblock 2200)
3. Studium (auch Berufsakademie, Frageblock 2300)
4. Berufsvorbereitung, Berufsgrundbildung (Frageblock 2400)
5. Einstiegsqualifizierung (EQJ), Praktikum oder Teilqualifizierung im Betrieb (Frageblock 2450)
6. Erwerbstätigkeit, jobben (Frageblock 2600)
7. Langfristige Fortbildung, Umschulung in Vollzeit (Frageblock 2700)
8. Maßnahme der Bundesagentur für Arbeit (außer berufsvorbereitende Maßnahme, Frageblock 2900)
9. Wehr-/Zivildienst, Freiwilliges Soziales oder Ökologisches Jahr (Frageblock 2800)
10. Sonstiges oder zu Hause (Frageblock 2850)

Die Jugendlichen sollten dabei auch angeben, ob und welche parallelen Aktivitäten sie im Zeitraum nach der Grundschule bis zum Befragungszeitpunkt ausgeübt haben (z.B. Studium und Erwerbstätigkeit). Die Jugendlichen waren aufgefordert, alle Phasen, auch solche ohne Qualifizierung oder Berufstätigkeit bzw. Nebentätigkeiten (z.B. Jobben während des Studiums) anzugeben. Dabei waren Lücken von maximal zwei Monaten zugelassen.

² Vorgeschaltet war ein Kontaktinterview (siehe Fragebogen). Die Variablen des Kontaktinterviews sind nicht Bestandteil des für wissenschaftliche Zwecke zur Verfügung gestellten Datensatzes.

³ Die institutionelle Vielfalt von Schulformen in Deutschland erschwert die korrekte Einordnung des Schulbesuchs durch die Befragten. Aus diesem Grund wurden Interviewer und Befragter mit einer Liste mit allen bekannten länderspezifischen Schulformen und Zuordnungsregel durch das BIBB zur Verfügung gestellt und im Erhebungsinstrument berücksichtigt.

Für die detaillierte Erfassung der genannten Aktivitäten leitete das CATI-Programm anschließend nacheinander in den jeweiligen Frageblock weiter. Um Fehlzugeordnungen durch die Befragten zu vermeiden, wurde vor den detaillierten Fragen bei jeder Aktivität nochmals nachgefragt und ggfs. in einen anderen Frageblock umgeleitet (dies trat in 3,5% der genannten Aktivitäten auf).

Zu jeder Aktivität wurde zunächst die Dauer der Aktivität monatsgetreu⁴ erfasst. Neben diesen kalendarischen Informationen wurden anschließend detaillierte Informationen zu der konkreten Ausgestaltung der jeweiligen Phase erhoben. Hierzu gehören (je nach Station):

- Angaben zur genauen Art der jeweiligen Aktivität (etwa der konkrete Schultyp, die Ausbildungsart),
- Entscheidungsgründe für bestimmte Stationen
- subjektive Bewertungen im Hinblick auf die fachliche und persönliche Entwicklung sowie die Auswirkungen auf den weiteren beruflichen Werdegang
- Leistungen (Noten),
- Such- und Bewerbungsaktivitäten.

Nach Abschluss der Detailfragen zu jeder genannten biographischen Phase wurden die einzelnen registrierten Stationen den Befragten nochmals in der biographischen Abfolge vorgelesen und Gelegenheit zur eventuellen Korrektur der Zeitangaben bzw. Ergänzung vergessener Aktivitäten gegeben⁵ (Prüfmodul, Frageblock 3000, siehe Fragebogen).

Die Anzahl der durchlaufenden Stationen variiert erwartungsgemäß mit dem Alter und der Typik der durchlaufenen Stationen. Für die meisten Befragten wurden drei oder vier Aktivitäten (Ereignisse) für den Zeitraum nach der Grundschule bis zum Befragungszeitpunkt erfasst (vgl. ebd.: 98, s. Abschnitt 3.3).

3.3 Soziodemographische Angaben und Grundeinstellungen

Den letzten Teil der Befragung bildete die Erfassung wichtiger soziodemographischer Merkmale der Jugendlichen und ihrer Eltern. In diesem Abschnitt des Interviews wurden detaillierte Informationen zum Migrationshintergrund der Jugendlichen (u.a. Geburtsland, Jahre in Deutschland, erste gelernte und meist gesprochene Sprache), sowie zur familiären Situation bzw. zu der Bildung und dem Beruf der Eltern zum Zeitpunkt, als der Befragte 15 Jahre alt war, erfragt. Abschließend wurden zudem inner- und außerschulische ehrenamtliche Tätigkeiten während der Schulzeit, persönliche Grundeinstellungen und Selbstwirksamkeitserwartungen der Jugendlichen erhoben.

3.4 Themen in der Zusammenschau

Die folgende Tabelle 2 gibt den Inhalt der Befragung anhand von Themenbereichen wieder.

⁴ Ungenaue Zeitangaben (z.B. ‚Herbst‘) wurden später in Monatsangaben umgewandelt.

⁵ In Interviewerschulungen wurde besonders auf die Wichtigkeit der Erfassung von kürzeren Zeiträumen (z.B. Warten auf einen Ausbildungs- /Studienplatz) hingewiesen.

Tabelle 2: Themen- und Variablenübersicht über die BIBB-Übergangsstudie 2006

Themenblock	Fragebogen seiten (von – bis)
Eingangsfragen und Grundschule (200)	8-10
Biographie im Anschluss an die Grundschule (2000)	11-102
0. Strukturblock: Aktivitäten (2000)	11-13
1. Schulbesuch (2100)	14-27
2. Berufliche Ausbildung, Lehre (2200)	28-39
3. Studium (2300)	40-48
4. Berufsvorbereitung, -grundbildungsjahr (2400)	49-58
5. Einstiegsqualifizierung, Praktikum und Teilqualifizierung (2450)	59-66
6. Erwerbstätigkeit, Jobben (2600)	67-76
7. Langfristige Fortbildung oder Umschulung (2700)	77-80
8. Maßnahmen der BA (2900)	81-87
9. Wehr-/Zivildienst, Freiwilliges Soziales oder Ökologisches Jahr (2800)	88-94
10. Sonstiges oder zu Hause (2850)	95-101
Prüfmodul „Zeitachse für alle Aktivitäten (3000)	102
Angaben zur Person (4000)	103-106
Angaben zu den Eltern und abschließende Fragen (5000)	107-118

4 Datenerhebung

4.1 Grundgesamtheit und Stichprobenziehung

Die Durchführung der Erhebung erfolgte durch das Institut für angewandte Sozialwissenschaften GmbH – infas (Bonn) im Auftrag des BIBB.

Grundgesamtheit der BIBB-Übergangsstudie 2006 sind alle Jugendliche und junge Erwachsene der Geburtsjahrgänge 1982 bis 1988, die zum Zeitpunkt der Befragung ihren Wohnsitz in Deutschland hatten. Zudem umfasst die Grundgesamtheit solche Jugendlichen der genannten Geburtsjahrgänge, die über ausreichende deutsche Sprachkenntnisse verfügen, um das Interview durchführen zu können⁶. Gegenstand der Befragung sind alle Jugendlichen, also auch diejenigen, die die allgemeinbildenden Schulen noch besuchen.

⁶ Unter den Interviewern waren Personen mit fundierten Kenntnissen in den wichtigsten Sprachen der Jugendlichen mit Migrationshintergrund. Diese sollten das Kontaktinterview übernehmen, falls der erstkontaktierte Gesprächspartner nicht über ausreichende Deutschkenntnisse verfügte. Dies sollte garantieren, dass die Kontaktaufnahme zu einem Jugendlichen mit Migrationshintergrund nicht schon im Vorfeld an den fehlenden Sprachkenntnissen der Eltern oder anderer Haushaltszugehöriger scheitert (vgl. Beicht und Friedrich 2008: 88).

Die retrospektive und chronologische Erfassung von zum Teil sehr unterschiedlichen Bildungs- und Berufsbiographien und des sonstigen Frageprogramms erfolgte mittels computerunterstützter Telefonbefragung (CATI)⁷. Für die inhaltliche Prüfung des Erhebungsinstrumentariums und der Funktionalität der CATI-Programmierung wurde im Mai 2006 zunächst ein Pretest mit 35 Befragten durchgeführt. Dabei zeigte sich die besondere Schwierigkeit, die spezielle Altersgruppe im Rahmen einer Zufallsstichprobe von Telefonhaushalten zu erreichen. Im Ergebnis führte der Pretest zu einer starken Kürzung des Einführungstextes. Zudem wurden die Interviewer im Hinblick auf die Vermittlung von Informationen über die Bedeutung und den Inhalt der Studie sowie zum Auswahlverfahren bzw. Datenschutz geschult.

Die Haupterhebung fand vom 30. Mai bis zum 20. August 2006 statt. Die Stichprobenziehung erfolgte in einem zweistufigen Verfahren entsprechend dem üblichen ADM-Telefonstichprobendesign (Gabler und Häder 1998, Heckel 2002)⁸. Im ersten Auswahlschritt wurde unter Berücksichtigung der regionalen Merkmale Kreis und BIK-Gemeindegroßenklasse (zehnstufig) eine Zufallsstichprobe aus registrierten Festnetznummern gezogen. Diese wurde zu einem Anteil von 22 Prozent durch künstlich erzeugte Telefonnummern (auf Basis vorhandener Anfangs-Nummernblöcke) ergänzt. Im zweiten Auswahlschritt wurde im Rahmen eines Kontaktinterviews geprüft, ob im Zielhaushalt eine oder mehrere Personen der Geburtsjahrgänge 1982 bis 1988 leben. Im

Falle mehrerer Personen mit diesem Merkmal wurde die zu befragende Person mit der Last-Birthday-Methode ermittelt. Die Interviewer sollten außerdem soweit als möglich die Telefonnummern (auch Mobilfunknummern) der (v.a. älteren) Zielpersonen erfragen, wenn diese zwar im elterlichen Haushalt gemeldet, aber selten oder gar nicht mehr dort anwesend waren.

Eine Bruttostichprobe im Umfang von rund 320.000 Telefonnummern wurde eingesetzt, um eine Nettofallzahl von rund 1.000 Befragten pro Geburtsjahrgang zu erzielen⁹. Telefonnummern von nicht erreichten Haushalten wurden über die gesamte Feldzeit wiederholt gewählt. Ebenso wurde über den gesamten Zeitraum der Datenerhebung mehrfach die Bundeslandverteilung in der realisierten Stichprobe mit der tatsächlichen Verteilung der Bundesländer abgeglichen, so dass entsprechende Korrekturen bei der letzten Tranche vorgenommen werden konnten.

⁷ Für die Abwägung von Vor- und Nachteilen einer CATI-Befragung in der BIBB-Übergangsstudie siehe Schiel et al. 2006: S. 1 ff.

⁸ Für weitere Details zur Stichprobenziehung siehe Schiel et al. 2006.

⁹ Es wurde davon ausgegangen, dass zum Befragungszeitpunkt etwa in 14 Prozent der Haushalte (5,6 Mio.) Personen im Alter zwischen 18 und 24 leben und zudem mit relativ vielen Ausfällen aufgrund der nicht geschalteten Rufnummern gerechnet werden musste. Die Nummern wurden in sechs Tranchen eingeteilt und nacheinander eingesetzt, um Selektivitätseffekte zu minimieren.

Parallel zur Hauptstudie wurde von Anfang August an eine Non-Response-Studie durchgeführt, bei der vormalige Verweigerer erneut kontaktiert und ein Screening¹⁰ sowie ggfs. sogar ein Interview (n=155) durchgeführt werden konnten.

Die durch das Stichprobenverfahren bedingten unterschiedlichen Auswahlwahrscheinlichkeiten auf Haushalts- und Personenebene sind in der Gewichtungvariable berücksichtigt (siehe Abschnitt 7.1).

4.2 Realisierung der Stichprobe

Von den rund 320.000 Telefonnummern der Bruttoeinsatzstichprobe führten knapp 77 Prozent zu einem Haushalt. Davon konnte in rund 81 Prozent der Fälle (200.039) ein Screening durchgeführt werden. In vier Prozent der gescreeenten Haushalte (7.961) lebte nach Auskunft der Kontaktperson eine Zielperson der Jahrgänge 1982 bis 1988¹¹. Von diesen konnte ein Großteil (93%) erfolgreich für ein Interview gewonnen werden. Für die übrigen sieben Prozent (n=558) liegen die Antworten zum Kurzfragebogen vor (s.u.). Nach einer Prüfung verbleiben 7.230 vollständige und auswertbare Interviews (Rücklauf für Zielperson im Haushalt 90,8%).

Die Ausfallstatistik ist der folgenden Tabelle 3 zu entnehmen.

¹⁰ In 60 Prozent der vormaligen Verweigerer-Haushalte, die dann doch ein Screening durchliefen, lebte keine Zielperson.

¹¹ Zur Frage, inwieweit der gemessene Anteil von vier Prozent als Prävalenzrate von Haushalten mit mindestens einer Zielperson herangezogen werden kann siehe Schiel et al. 2006.

Tabelle 3: Übersicht über die Ausschöpfung der eingesetzten Bruttostichprobe (entnommen aus Beicht und Friedrich 2008, 92)

Ausschöpfung der Bruttostichprobe	Anzahl	
	absolut	in %
Insgesamt eingesetzte Telefonnummern (Bruttoeinsatzstichprobe)	320.557	100,0
keine gültigen Telefonnummern (nicht geschaltete Nummern, Firmenanschluss, Fax, Modem, kein Kontakt nach mehr als 10 Versuchen)	73.978	23,1
Telefonnummern, die zu einem Haushalt führten (bereinigte Einsatzstichprobe)	246.579	76,9
Bereinigte Einsatzstichprobe	246.579	100,0
kein Kontakt zum Haushalt, kein Screening möglich (Anrufbeantworter, nicht abgehoben, besetzt)	16.453	6,7
Kontakt zu Haushalt, Screening abgebrochen (Verweigerung, keine Klärung möglich, ob eine Zielperson im Haushalt lebt)	28.699	11,6
Kontakt zum Haushalt, aber Zielperson nicht befragbar (wegen mangelnder Sprachkenntnisse, aus gesundheitlichen Gründen,)	1.388	0,6
Kontakt zum Haushalt, Screening durchgeführt	200.039	81,1
Kontakt zum Haushalt, Screening durchgeführt	200.039	100,0
keine Zielperson im Haushalt vorhanden	192.078	96,0
Zielperson im Haushalt vorhanden	7.961	4,0
Zielperson im Haushalt vorhanden	7.961	100,0
Zielperson oder Kontaktperson verweigert (Kurzinterview liegt vor)	558	7,0
Interview unterbrochen, konnte in Feldzeit nicht fortgeführt werden	4	0,1
Interview abgebrochen	153	1,9
Interviews vollständig, nach Prüfung aber nicht auswertbar	16	0,2
Auswertbare Interviews	7.230	90,8

Die Interviews dauerten im Mittel etwa 34 Minuten bei relativ hoher Varianz (Standardabweichung 12 Minuten), einem kürzesten Interview von knapp 12 Minuten und längsten Interview von 107 Minuten Länge. Die großen Unterschiede gehen vor allem auf die Unterschiede in der Anzahl der erfassten Aktivitäten zurück. Diese nehmen durchschnittlich mit dem Alter des Befragten zu.

Tabelle 4 weist die Häufigkeiten in der Anzahl der erfassten Aktivitäten bzw. Ereignisse. Für die meisten Befragten wurden drei oder vier Aktivitäten für den Zeitraum nach der Grundschule bis zum Befragungszeitpunkt erfasst (vgl. Beicht und Friedrich 2008: 98). 91 Prozent der erfassten Bildungsbiographien bestehen aus ein bis sechs erfassten Aktivitäten.

In den verbleibenden neun Prozent der Biographien konnten sieben und bis zu 18 Aktivitäten im Anschluss an die Grundschule erfasst werden.

Tabelle 4: Häufigkeiten in der Anzahl der erfassten Aktivitäten

Anzahl der Aktivitäten	Häufigkeiten	Prozent	Kumulierte Häufigkeiten (%)
3	1644	22,7	22,7
4	1508	20,9	43,6
2	1061	14,7	58,3
5	995	13,8	72,0
1	711	9,8	81,9
6	659	9,1	91,0
7	326	4,5	95,5
8	155	2,1	97,6
9	76	1,1	98,7
Andere über 9, max. t=18	95	1,3	100,0
Total	7230	100,0	

Hinweise: ungewichtete Häufigkeiten.

Von allen genannten Aktivitäten betreffen 39 Prozent den Schulbesuch (vgl. Tabelle 5, vgl. Schiel et al. 2006: 21). Rund 16 Prozent der erfassten Aktivitäten beziehen sich auf Erwerbstätigkeit, 14 Prozent auf eine schulische oder betriebliche Ausbildung und 13 Prozent auf sonstige Tätigkeiten oder Tätigkeiten zu Hause. Die übrigen Aktivitäten machen jeweils weniger als fünf Prozent aller erfassten Aktivitäten aus.

Tabelle 5: Anzahl der genannten Aktivitätsarten

Aktivitätsarten	Häufigkeiten	Prozent	Kumulierte Häufigkeiten (%)
Schule besucht	10849 ^a	38,8	38,8
berufliche Ausbildung, Lehre	3948	14,1	52,9
Studium, Berufsakademie	1224	4,4	57,3
Berufsvorbereitung, Berufsgrundbildung	1316	4,7	62,0
Einstiegsqualifizierung	1134	4,1	66,0

Erwerbstätig oder Jobben	4491	16,1	82,1
langfristige Fortbildung, Umschulung	124	0,4	82,5
Trainings-, Weiterbildungsmaßnahme, ABM	250	0,9	83,4
Wehrdienst, Zivildienst, freiw. soz./ökol. Jahr	998	3,6	87,0
Sonstiges oder zu Hause	3631	13,0	100,0
Total	27965	100,0	

Hinweise: ^a Inklusive der 19 nachträglich zusätzlich gebildeten Schulspeils. Ungewichtete Häufigkeiten.

Zum Zeitpunkt der Befragung¹² ist die größte Gruppe der Befragten (25,5%) voll oder teilweise erwerbstätig (s. Tabelle 6). Rund 22 Prozent absolvieren zum Befragungszeitpunkt (auch) eine schulische oder betriebliche Berufsausbildung, 16,7 Prozent entfallen auf sonstige Tätigkeiten, knapp 16 Prozent besuchen eine Schule. Knapp 13 Prozent sind 2006 Studenten.

Tabelle 6: Aktueller Status der Befragten: Andauern der Aktivitäten zum Befragungszeitpunkt

Aktivitätsarten	Häufigkeiten	Prozent gesamt	Fallzahl innerhalb^a	Prozent innerhalb^b
Schule besucht	1223	15,7	7729 / 7211	17,0
berufliche Ausbildung, Lehre	1696	21,8	3767 / 3568	47,5
Studium, Berufsakademie	1004	12,9	1218 / 1135	88,5
Berufsvorbereitung, Berufsgrundbildung	139	1,8	1189 / 1165	11,9
Einstiegsqualifizierung	151	1,9	935 / 906	16,7
Erwerbstätig oder Jobben	1988	25,5	3850 / 3208	62,0
langfristige Fortbildung, Umschulung	45	0,6	115 / 113	39,8
Trainings-, Weiterbildungsmaßnahme, ABM	49	0,6	221 / 213	23,0
Wehrdienst, Zivildienst, freiw. soz./ökol. Jahr	193	2,5	992 / 984	19,6
Sonstiges oder zu Hause	1300	16,7	3067 / 2762	47,1
Gesamt	7788	100,0		

Hinweise: ungewichtete Häufigkeiten.^a Fallzahlen für Befragte, die jemals die Aktivität ausgeübt haben jeweils inklusive / exklusive parallele Aktivitäten. ^b Basierend auf Fallzahlen inklusive parallele Aktivitäten.

¹² Bei jeder erfassten Aktivität wurden die Jugendlichen gefragt, ob diese aktuell noch andauert (Fragen 2106, 2203, 2303, 2406, 2452, 2601, 2702, 2902, 2803, 2852, 2854 und 2855). Bei der Darstellung in Tabelle 6 muss insbesondere berücksichtigt werden, dass bei der Erhebung auch parallele Aktivitäten erfasst wurden. Die auf Aktivitäten bezogenen Fallzahlen für Personen können auch zu einem Zeitpunkt die Personenfallzahl (n=7230) überschreiten.

4.3 Repräsentativität und Selektivität

Insgesamt ist die Verteilung der realisierten Stichprobe über die Bundesländer (Beicht und Friedrich 2008, S. 92) und das Geschlecht sehr gut angepasst, während Abweichungen gegenüber der amtlichen Statistik (Angaben des Statistischen Bundesamtes aus dem Statistischen Jahrbuch 2005) im Hinblick auf den Geburtsjahrgang und die Bildung bestehen (s. Tabelle 7).

Vermutlich aufgrund der besseren Erreichbarkeit von Jüngeren und höherer Mobilität von Älteren weichen die älteren Geburtsjahrgänge nach unten, die jüngeren nach oben ab. Eine Analyse der Interviewabbrüche (n=153) zeigt, dass diese sich nicht in bestimmten Jahrgängen konzentrieren (vgl. Schiel et al. 2006). Die realisierte Stichprobe besteht im Vergleich zur amtlichen Statistik zu einem höheren Anteil aus Personen mit einem hohen Abschluss und weniger Personen mit einem niedrigen Abschluss (umfragetypischer Bildungsbias).

Tabelle 7: Verteilungen der Personen der Geburtsjahrgänge 1982 bis 1988 nach sozio-demografischen Merkmalen in der Grundgesamtheit und in der realisierten Nettostichprobe (entnommen aus Beicht und Friedrich 2008, 92)

Soziodemografische Merkmale	Grundgesamtheit* in Prozent	Realisierte Stichprobe in Prozent	Differenz der Prozentpunkte
Bundesland	100,0	100,0	
Baden-Württemberg	13,0	13,4	0,4
Bayern	14,8	15,5	0,7
Berlin	4,5	2,5	-2,0
Brandenburg	3,6	3,1	-0,5
Bremen	0,8	0,6	-0,2
Hamburg	2,1	1,4	-0,7
Hessen	6,9	7,9	1,0
Mecklenburg-Vorpommern	2,6	1,7	-0,8
Niedersachsen	9,2	10,1	0,9
Nordrhein-Westfalen	20,9	23,2	2,3
Rheinland-Pfalz	4,7	5,2	0,5
Saarland	1,2	1,5	0,3
Sachsen	5,9	5,1	-0,8
Sachsen-Anhalt	3,4	3,1	-0,3
Schleswig-Holstein	3,0	2,6	-0,4
Thüringen	3,4	3,3	-0,1
Geschlecht	100,0	100,0	
männlich	51,2	51,4	0,2
weiblich	48,8	48,6	-0,2

Geburtsjahrgang	100,0	100,0	
1988	14,7	18,3	3,6
1987	14,4	21,3	6,9
1986	14,2	18,4	4,2
1985	13,8	13,0	-0,8
1984	13,9	12,2	-1,7
1983	14,2	9,8	-4,4
1982	14,7	7,0	-7,7
höchster Schulabschluss	100,0	100,0	
noch in schulischer Ausbildung	3,9	11,8	7,9
Hauptschulabschluss	26,4	11,8	-14,6
Realschulabschluss	35,4	36,7	1,3
(Fach-)Hochschulreife	27,8	38,5	10,7
ohne Angabe	1,6	0,7	-0,9
ohne Abschluss	4,8	0,6	-4,2

Quelle: Statistisches Bundesamt: Statistisches Jahrbuch 2005. Bei den Angaben zu den Bundesländern handelt es sich um die amtliche Verteilung der 18- bis unter 25-Jährigen. Bei den Angaben zum Schulabschluss wurde die amtliche Verteilung bereinigt um die 15- bis unter 18-Jährigen, die komplett aus der Gruppe „noch in schulischer Ausbildung“ abgezogen wurden.

Eine auf Basis der Ergebnisse zum Kurzinterview (Verweigerer) durchgeführte Selektivitätsanalyse mit Hilfe einer multivariaten logistischen Regression erbringt, dass die soziodemographischen Variablen (Geschlecht, Jahrgang, Staatsangehörigkeit, Schulabschluss), die Anzahl der Zielpersonen im Haushalt und das Merkmal Bewerbung bzw. Beginn einer Berufsausbildung keine starken Prädiktoren für die Verweigerung des Interviews sind (vgl. Schiel et al. 2006). Vielmehr wird die Teilnahmebereitschaft durch unbeobachtete Merkmale der Befragten, also solche die mit den untersuchten Variablen nur bedingt kovariieren, beeinflusst. Möglicherweise sind dies Merkmale der unterschiedlichen Erreichbarkeit der Zielpersonen (Beicht und Friedrich 2008, S. 96).

Für die Ausgleichung der Unterschiede in den demographischen Kennwerten der Stichprobe und der Grundgesamtheit (gemäß der der amtlichen Statistik) wurde ein Populationsgewicht erstellt (siehe Abschnitt 7.1), welches im Datensatz zur Verfügung steht.

5 Datenaufbereitung

Durch die sorgfältige Entwicklung von Filterführung und Prüfmodulen durch die Projektmitarbeiter im BIBB und deren programmiertechnische Umsetzung in ein CATI-Programm durch infas (Bonn) konnten viele Datenfehler im Vorfeld vermieden werden.

Insbesondere führt das in das Interview integrierte Prüfmodul zur Vollständigkeit der erfassten Biographie (Frageblock 3000, s. Fragebogen) und zu einer guten Qualität des Rohdatensatzes. Vom Umfrageinstitut wurde dem BIBB ein personenbezogener Querschnittdatensatz und ein auf Aktivitäten bezogener Ereignisdatensatz übergeben. Diese Datensätze wurden zudem im Projektteam der BIBB-Übergangsstudie 2006 weiteren Qualitätsprüfungen unterzogen und um weitere Variablen ergänzt. Schließlich wurden im BIBB-FDZ weitere Datenaufbereitungen durchgeführt. Hierzu gehören die Hinzufügung von weiteren Hilfsvariablen (s. 5.1. und 5.2), Anonymisierungen (Kapitel 6) sowie die Standarddeklarierung der fehlenden Werte (vgl. Abschnitt 7.2).

Im BIBB-FDZ sind drei Datensatzbestandteile der BIBB-Übergangsstudie im Angebot: Ein Querschnittdatensatz (vgl. 5.1), ein Ereignisdatensatz (vgl. 5.2.) und je ein Datensatz mit den Textvariablen aus dem Querschnitt- und dem Längsschnittteil der BIBB-Übergangsstudie (vgl. 5.3). Die Textvariablen stehen aufgrund des höheren Reidentifikationsrisikos nur für Auswertungen im Gastwissenschaftleraufenthalt zur Verfügung.

5.1 Querschnittdatensatz

In dem Querschnittdatensatz stehen die Identifikatoren aus dem Originaldatensatz für eine Auswertung nicht zur Verfügung bzw. wurden durch ein neu gebildetes, systemfreies Personenkennzeichen (*lfd_neu*) ersetzt.

Der Querschnittdatensatz enthält alle Spell-invarianten Originalvariablen¹³ (Frageblöcke 200, 4000 und 5000 - Eingangsfragen und Grundschule, Angaben zur Person und Eltern, vgl. Tabelle 2) außer der Anzahl der Telefonnummern des Haushaltes (Frage 5031). Allerdings werden die offen abgefragten Angaben der Jugendlichen nur für Auswertungen im Gastwissenschaftleraufenthalt und in der Datenfernverarbeitung zur Verfügung gestellt (s. Abschnitt 5.3 und 6.3).

Wenn nicht explizit erwähnt, sind die direkt erhobenen Variablen im Datensatz nach der Fragennummer im Fragebogen mit einem vorangestellten *q* benannt.

Neben den direkt erhobenen enthält der Datensatz eine Reihe von geänderten bzw. nachträglich gebildeten Variablen vor allem aus dem Querschnitt- z.T. aber auch dem Längsschnittteil. Diese werden im Folgenden in der Reihenfolge erläutert, in der sie bzw. die Ausgangsvariablen bei der Befragung erhoben wurden.

Im Frageblock 200 wurde die Angaben zur Frage 103 (Geburtsmonat und -jahr) in zwei getrennten Variablen abgelegt (*gebmonat*, *gebjahr*).

¹³ Die Variablen des Screeningteils sind nicht Bestandteil des für wissenschaftliche Zwecke zur Verfügung gestellten Datensatzes.

Im Frageblock 2100 (Schulbesuch) des Ereignisteils wurden die Jugendlichen nach ihrem erreichten allgemeinbildenden Schulabschluss gefragt (Frage 2113). Hiermit wurde eine Variable ‚höchster allgemeinbildender Abschluss‘ (*hbild*) beim Verlassen der allgemeinbildenden Schule¹⁴ erzeugt und dem Querschnittdatensatz beigelegt. Auf Basis der Spell-Identifikationsnummer des Ereignisdatensatzes wurde dem Querschnittdatensatz außerdem die Variable Anzahl der Aktivitäten pro Person (*n_spells*) beigelegt.

Im Anschluss an den Ereignisteil wurden die Jugendlichen gebeten, Angaben zu Migrationshintergrund, der Bildung und dem Beruf der Eltern zu machen (Frageblock 5000). Für den Scientific-Use-File wurden die Angaben bei den Fragen zum Geburtsland (4000), und Staatsangehörigkeit (4008) jeweils in gruppierte Variablen rekodiert (*q4000_gr*, *q4008_de-q4008_98*, für Wertelabels s. Anhang 9.2). Bei dem Beruf des Vaters wurde erhoben, welche berufliche Tätigkeit im Hauptberuf der leibliche Vater (oder alternativ der Stief-, Pflege- oder Adoptivvater) ausgeübt hat, als der Befragte 15 Jahre alt war (Frage 5014). In den Fällen, in denen Jugendliche nicht durchgängig oder die meiste Zeit mit ihrem Vater zusammengelebt haben (Fragen 5002 und 5003) oder zu diesem Zeitpunkt keine Erwerbstätigkeit vorlag bzw. keine Angaben hierzu gemacht wurden (Fragen 5012, 5013) wurde nach der beruflichen Tätigkeit der Mutter gefragt (Frage 5022). Die erfassten beruflichen Tätigkeiten der Väter (*VBR*) und Mütter (*MBR*) wurden von GESIS-ZUMA wie auch für die Befragten (vgl. 5.2) in Berufskennziffern nach der Klassifikation der Berufe (KLDB) in der Version von 1992 und der Internationalen Standardklassifikation der Berufe (ISCO) in der Version von 1988 rekodiert¹⁵ (KldB1992: *VBR_KLDB92*, *MBR_KLDB92*; ISCO88: *VBR_I88*, *MBR_I88*). Die Variablen liegen den Datensatzversionen auf 4- oder 2-Stellerebene bei (vgl. 6.3). Die 2-Steller wurden im BIBB-FDZ mit den Berufsgruppen (KldB) bzw. Hauptuntergruppen (ISCO-88) gelabelt.

Zusätzlich zu der Bereitstellung der Berufskennziffern beauftragte das BIBB GESIS-ZUMA, die Berufskennziffern (ISCO88) der Eltern (und der Befragten, vgl. 5.2) in Status-/Prestigewerte nach Treimann (*VBR_SIOPS*, *MBR_SIOPS*), nach Ganzeboom (*VBR_ISEI*, *MBR_ISEI*), nach Wegener (*VBR_MPS*, *MBR_MPS*¹⁶) in und in das soziologische Klassenschema nach Erikson, Goldthorpe und Portocarero (*VBR_EGP*, *MBR_EGP*, 11-stufig, für Wertelabels s. Anhang 9.2) zu rekodieren (vgl. Geis 2009, s.a. Ganzeboom 2003, Ganzeboom 2001). Die hierfür notwendigen Informationen zur Selbstständigkeit und der Anzahl der Mitarbeiter wurden aus den Variablen zur beruflichen Stellung des Vaters (Frage 5015) und der Mutter (Frage 5023) bzw. der Selbstständigengruppe (Frage 5019 und 5027) generiert¹⁷. Die beiden Hilfsvariablen zur Selbstständigkeit des Vaters und der Mutter

¹⁴ Da Schulabschlüsse auch im späteren Bildungsverlauf noch nachgeholt werden und zum Zeitpunkt des Interviews noch andauern können, ist bei der Bildung der Variablen ein Referenzzeitpunkt festzulegen.

¹⁵ Für Väter liegen *n*=6187/6024 und für Mütter *n*=804/794 gültige Berufskennziffern in der KldB1992 bzw. ISCO88 vor.

¹⁶ Für die Rekodierung von ISCO88 in MPS siehe Christoph (2005).

¹⁷ Die Mitarbeiterzahl wurde nicht direkt erhoben. Auf Basis der Angaben zur Selbstständigengruppe wurde folgende Zuordnung von Ausprägungen bei der Frage 5019 und 5027 zur Mitarbeiteranzahl vorgenommen: 31=5, 32=11, 33=1, 34=0.

(*semp_l_v*, *semp_l_m*) und zur Mitarbeiterzahl (*supvis_v*, *supvis_m*) sind zusammen mit den generierten EGP-Variablen für die Eltern (*egp_v*, *egp_m*, vgl. Anhang 9.2) im Datensatz enthalten. Außerdem hat GESIS-ZUMA für die Rekodierung in EGP die originalen ISCO88-Variablen (*VBR_I88*, *MBR_I88*) in EGP-ISCO-Variablen rekodiert (*i88_v*, *i88_m*, siehe Ganzeboom und Treiman 2002). Die beiden EGP-ISCO-Hilfsvariablen sind nicht im Scientific-Use-File enthalten. Zusätzlich wurde dem Datensatz von GESIS-ZUMA eine Variable *alter* (Alter in Jahren zum Befragungszeitpunkt) beigefügt.

Schließlich liegen je nach Anonymisierungsversion (s. Abschnitt 6.3) im Querschnittdatensatz detaillierte regionale Variablen für den Wohnort zum Befragungszeitpunkt¹⁸ vor: Bundesland (*bula*), Raumordnungsregionen (*ror*) und Arbeitsamtsbezirk (*aaz*, letztere beiden jeweils nur in der Datenfernverarbeitung und an den Gastwissenschaftlerarbeitsplätzen im BIBB). Die Raumordnungsregionen wurde mittels eines im BIBB-FDZ vorliegenden Umsteigeschlüssels auf Basis der Regionalinformation auf Kreisebene erstellt¹⁹. Die 97 Raumordnungsregionen setzen sich aus den Kreisen und kreisfreien Städten zusammen, die Grenzen von Raumordnungsregionen verlaufen nicht innerhalb von Kreisen. Die Raumordnungsregionen liegen für 6677 Fälle der BIBB-Übergangsstudie 2006 vor, in 553 Fällen war aufgrund fehlender Kreiskennziffern keine Zuordnung möglich. Eine Ausfallanalyse für Kreiskennziffern zeigt, dass das Fehlen der Kreiskennziffer weitestgehend neutral gegenüber der Struktur des Samples ist (siehe Anhang, Abschnitt 9.1). Für alle 7230 Fälle liegen Regionalinformationen auf der Ebene von Arbeitsamtsbezirken und Bundesländern vor.

In Ergänzung der direkt erhobenen Spell-invarianten Originalvariablen wurde dem Querschnittdatensatz vom Umfrageinstitut eine Gewichtungvariable (*gewicht*, s. Abschnitt 7.1) zugespielt.

Insgesamt enthält der Datensatz damit 156 (GWA/DFV) bzw. 144 (SUF) Variablen. Tabelle 8 weist die im Querschnittdatensatz der BIBB-Übergangsstudie 2006 enthaltenen Variablen tabellarisch aus.

¹⁸ Die Ausgangsvariable wurde automatisch auf Basis der Informationen aus der Feldphase generiert.

¹⁹ Die administrative Verwaltungsgliederung ist in Deutschland nach Bundesländern, Regierungsbezirken, Landkreisen bzw. kreisfreien Städten sowie Gemeinden eingeteilt. Diese ist in der achtstelligen Gemeindekennziffer festgelegt, wobei die ersten zwei Ziffern für das Bundesland, die dritte Ziffer für den Regierungsbezirk, Ziffern vier und fünf für Stadt- bzw. Landkreis und die letzten drei Ziffern für die Gemeinde stehen. Zur Identifikation der 440 Stadt- und Landkreise werden entsprechend die ersten fünf Ziffern benötigt. Raumordnungsregionen sind – mit Ausnahme der Stadtstaaten – großräumige, funktional abgegrenzte Raumeinheiten für die Raumordnungsberichterstattung des Bundes. Die Variable mit den Kreiskennziffern steht aus Datenschutzgründen über keinen Datenzugangsweg für Auswertungen zur Verfügung.

Tabelle 8: Variablenübersicht für den Querschnittdatensatz (ohne Textvariablen) der BIBB-Übergangsstudie 2006

Themenblock	Anzahl Variablen	Variablenbereich / -namen	Fragebo- genseiten (von – bis)
Eingangsfragen und Grundschule (200)	21	gebmonat-q205	8-10
Angaben zur Person (4000)	46 / 30 (SUF)	q4000-q4011_8	103-106
Angaben zu den Eltern und abschließende Fragen (5000)	62	q5000-q104	107-118
Zusätzlich gebildete Variablen:	27 / 31 (SUF)	lfd_neu (SUF: q4000gr) (SUF: q4008_de-q4008_98) VBR_KLDB92 (SUF: VBR_KLDB92_2digit) VBR_I88 (SUF: VBR_I88_2digit) VBR_SIOPS VBR_ISEI VBR_MPS i88_v (DFV/GWA) sempl_v supvis_v egp_v MBR_KLDB92 (SUF: MBR_KLDB92_2digit) MBR_I88 (SUF: MBR_I88_2digit) MBR_SIOPS MBR_ISEI MBR_MPS i88_m (DFV/GWA) sempl_m supvis_m egp_m alter n_spells hbild westost bula ror, aaz (DFV, GWA) gewicht	s. Text (5.1)
Gesamt	156 / 144 (SUF)		

5.2 Ereignisdatensatz

In dem Ereignisdatensatz stehen die Identifikatoren aus dem Originaldatensatz für eine Auswertung nicht zur Verfügung bzw. wurden durch ein neu gebildetes, systemfreies Personenkennzeichen (*lfd_neu*) ersetzt.

Der Ereignisdatensatz enthält alle direkt erhobenen Originalvariablen²⁰ des Längsschnittteils (2000er Block, vgl. Kapitel 3). Allerdings werden die offen abgefragten Angaben der Jugendlichen nur für Auswertungen im Gastwissenschaftleraufenthalt und in der Datenfernverarbeitung zur Verfügung gestellt (s. Abschnitt 5.3 und 6.3).

Wenn nicht explizit erwähnt, sind die direkt erhobenen Variablen im Datensatz nach der Fragennummer im Fragebogen mit einem vorangestellten *q* benannt.

Neben den direkt erhobenen enthält der Datensatz eine Reihe von geänderten bzw. nachträglich gebildeten Variablen. Diese werden im Folgenden in der Reihenfolge erläutert, in der sie bzw. die Ausgangsvariablen bei der Befragung erhoben wurden.

Zunächst enthält der Datensatz einige synthetische Variablen, die die Verschränkung der Quer- und Längsschnittinformationen organisieren (siehe Abschnitt 2.1 und 5.2). Die Variable *lfd_neu* ist eine systemfreie Personenkennziffer und *index_n* eine Spell-Identifikationsnummer in chronologischer Sortierung mit dem Wert ‚1‘ (immer allgemeinbildende Schule ab dem 5. Schuljahr) bzw. höheren Werten für jede nachfolgend genannte Aktivität eines Befragten. Die Werte über ‚1‘ sind also Personen-spezifisch, d.h. sie zählen die genannten Aktivitäten, informieren jedoch nicht darüber, um welche (typische) Station (vgl. Abschnitt 3.1.) es sich dabei handelt. Diese Information ist in der Variable *q2000b* enthalten. Wenn in den Rohdaten in einigen Fällen die erste erfasste Aktivität nicht die Station ‚allgemeinbildende Schule ab dem 5. Schuljahr‘ war und somit künstlich erzeugt werden mussten (*n*=19), ist dies in der Indikatorvariable *zuspell* (mit dem Wert 0 für „Originalspell“ und 1 für „zusätzlicher Spell“) festgehalten.

Die Spell-ID (*index_n*) korrespondiert mit zeilenweise abgelegten Informationen zu den einzelnen genannten biographischen Phasen (langes Format). Für die kalendarische Verortung eines jeden Spells wurden die Spell-spezifischen Informationen über Beginn und Ende (jeweils Monat und Jahr) einer jeden Aktivität aus den Zeilen ‚gelöst‘ und in vier übergreifenden Kalendervariablen abgelegt: *begmo*, *begja*, *endmo* und *endja*.

Während des CATI-Interviews wurden in 992 Fällen (rund 3,5 Prozent aller erfassten Aktivitäten) Falschzuordnungen der Jugendlichen identifiziert und es kam es zu einer Umleitung in einen anderen Frageblock. Die Variable *q1002* informiert darüber, aus welchem Frageblock die Fälle umgeleitet wurden (gültige Werte: 1-9).

²⁰ Die Variablen des Screeningteils sind nicht Bestandteil des für wissenschaftliche Zwecke zur Verfügung gestellten Datensatzes.

Im Frageblock 2100 (Schulbesuch) wurde nach dem Bundesland gefragt, in dem die genannte Schule liegt (Frage 2108). Nach Beendigung der Hauptstudie wurde festgestellt, dass aufgrund eines Programmierfehlers Informationen für diese Variable nur für Schüler, die noch zur Schule gehen erhoben wurde. Das Umfrageinstitut infas hat diese Information im Rahmen einer Nacherhebung²¹ ermittelt (infas 2007). Die Ausgangsvariable (*q2108*) und die neue, mit anderen Wertelabels versehene Variable (*q2108_neu*, vgl. Anhang 9.2) sind im Datensatz enthalten. Im Hinblick auf die abgefragten schulischen und beruflichen Qualifizierungen der Jugendlichen waren ebenso kleine Korrekturen bei den Originalvariablen nötig. Die Variable *q2100kor* (Frage 2001) enthält die Ausprägungen und die Verteilung der originalen Schulbesuchvariable unter Berücksichtigung der 19 zusätzlich generierten Schulspells (siehe Indikatorvariable *zuspell*). Im Frageblock zur beruflichen Ausbildung (Frageblock 2200) wurden die Jugendlichen nach der Art der von ihnen absolvierten Ausbildung (Frage 2200) gefragt. Wie auch in anderen BIBB-Umfragen stellt sich hier das Problem, dass Jugendliche betriebliche und schulische Ausbildungsformen falsch zuordnen (vgl. Hall 2009). Im Fall der BIBB-Übergangsstudie konnten die Angaben mit Rückgriff auf die erfassten Berufsbezeichnungen überwiegend korrigiert werden (Beicht und Friedrich 2008: S. 98-99)²². In Ergänzung zu der Originalvariable liegt dem Datensatz die bereinigte Variable *ausbild* (für Wertelabels s. Anhang 9.2) bei.

Im Frageblock zu beruflicher Ausbildung und Lehre (Block 2200) wurde offen erfasst, in welchem Beruf die Jugendlichen ausgebildet wurden bzw. werden (falls die Ausbildung zum Befragungszeitpunkt noch andauert (Frage 2203)). Im BIBB wurden die offenen Nennungen zum Ausbildungsberuf in Berufskennziffern nach der Klassifikation der Berufe (KLDB) in der Version von 1992 kodiert (*bkz_2201*²³, 4-Steller). Im Auftrag des BIBB wurde die Variable in die Internationale Standardklassifikation der Berufe (ISCO) in der Version von 1988 (*AUSB_188*) kodiert (jeweils 4-Steller). Die rekodierten Variablen sind je nach Datenzugangsart auf der Ebene des 4- bzw. 2-Stellers im Datensatz enthalten (vgl. Abschnitt 6.3). Die 2-Steller wurden im BIBB-FDZ mit den Berufsgruppen (KldB) bzw. Hauptuntergruppen (ISCO-88) gelabelt. Zusätzlich wurden die Berufskennziffern (ISCO88) für den Ausbildungsberuf des Befragten in Status-/Prestigewerte nach Treimann (*SIOPS: AUSB_SIOPS*), nach Ganzeboom (*ISEI: AUSB_ISEI*) und nach Wegener (*MPS: AUSB_MPS*) und in das soziologische Klassenschemas nach Erikson, Goldthorpe und Portocarero (*EGP*, 11-stufig, für Wertelabels siehe 9.2) rekodiert (vgl. Geis 2009, Christoph 2003, s.a. Ganzeboom 2003). Für letztere hat GESIS-ZUMA die Informationen zur

²¹ Dabei wurde nur für Jugendliche mit einer Erwerbstätigkeit, einem Studium und langfristiger Fortbildung / Umschulung Mobilität unterstellt und nacherhoben (1498 Haushalte). Hier konnte für 1389 Haushalte im Rahmen der Nacherhebung die Information nachermittelt werden. Bei den übrigen Fällen wurde das Bundesland des Wohnortes zum Befragungszeitpunkt eingesetzt (vgl. infas 2007).

²² Im Ereignisdatsatz steht auch die korrigierte Variable *ausbild* zur Verfügung.

²³ Ergänzend zu der BIBB-Variable liegt dem Datensatz eine um ZUMA-Codes ergänzte Variable zum Ausbildungsberuf in der Klassifikation der Berufe bei (*AUSB_KLDB92*).

beruflichen Stellung (Frage 2606) und der Selbstständigengruppe (Frage 2610) in der ggfs. zeitlich der Ausbildung nachfolgenden Erwerbstätigkeit herangezogen²⁴. Die beiden Hilfsvariablen zur Selbständigkeit des Befragten (*semp_l_a*) und zur Mitarbeiterzahl (*supvis_a*) sind zusammen mit der generierten EGP-Variable (*egp_a*) im Datensatz enthalten (für EGP-Wertelabels s. 9.2). Außerdem hat GESIS-ZUMA für die Rekodierung in EGP die originale ISCO88-Variable (*AUSB_I88*) in eine EGP-ISCO-Variable rekodiert (*i88_a*, siehe Ganzeboom und Treiman 2002). Die EGP-ISCO-Hilfsvariable ist nicht im Scientific-Use-File enthalten.

Im Frageblock 2300 (Studium) wurde bei der Variable zur Frage 2301 bzw. 2301a (Hauptfach, ggfs. zweites Hauptfach) festgestellt, dass die in den Daten enthaltenen Zuordnung von Werten und Wertelabels nicht mit derjenigen übereinstimmten, die im Fragebogen abgedruckt ist. Die Variablenliste für q2301 und q2301a kann dem Anhang 9.2 entnommen werden.

Im Frageblock 2600 (Erwerbstätigkeit, Jobben) wurde für die Berufstätigkeit zum Befragungszeitpunkt oder davor erfasst, welche berufliche Tätigkeit ausgeübt wird bzw. wurde (Frage 2604). Auch hierfür wurde im Auftrag des BIBB von GESIS-ZUMA eine Rekodierung der offenen Angaben in Berufskennziffern (KldB1992: *BBR_KLDB92*, ISCO88: *BBR_I88*) vorgenommen und gemäß des Anonymisierungskonzepts auf 4- bzw. 2-Stellerebene dem Datensatz beigefügt. Die 2-Steller wurden mit den Berufsgruppen (KldB) bzw. Hauptuntergruppen (ISCO-88) gelabelt. Analog zur Vorgehensweise bei dem Ausbildungsberuf wurde auch hier eine Rekodierung in SIOPS, ISEI, MPS und EGP (11-stufig) vorgenommen. Diese und die Hilfsvariablen über den Selbstständigkeit und die Mitarbeiterzahl bei der ausgeübten Erwerbstätigkeit (Fragen 2606 und 2610) sind dem Datensatz beigefügt (*BBR_SIOPS*, *BBR_ISEI*, *BBR_MPS*, *semp_l_b*, *supvis_b*, *egp_b*). Die EGP-ISCO-Hilfsvariable (*i88_b*, s.o.) ist nicht im Scientific-Use-File enthalten.

Schließlich fehlte bei einigen Variablen zu den (Abbruch-)Gründen für bestimmte Aktivitäten im Ausgangsdatsatz eine Indikatorvariable für die Fälle, die eine offene Angabe („aus anderem Grund, und zwar“) gemacht haben. In allen Datensatzversionen wurden Indikatorvariablen (1 „trifft zu „ und 2 „trifft nicht zu“) für diese Variablen eingefügt und mit *qFragebogennummer_sonst* benannt: Fragen 2109, 2112, 2213, 2308, 2408, 2411, 2457, 2460, 2707, 2904. Auf eine Auflistung dieser Variablen im Anhang wird verzichtet.

Insgesamt enthält der Datensatz damit 393 (GWA / DFV) bzw. 391 (SUF) Variablen. Tabelle 9 weist die im Ereignisdatsatz der BIBB-Übergangsstudie 2006 enthaltenen Variablen tabellarisch aus.

Für den Ereignisdatsatz stellt das BIBB-FDZ das Visualisierungsprogramm LDEX, Visual Explorer for Longitudinal Data für die BIBB-Übergangsstudie im Rahmen der Datenfernverarbeitung und Gastrechnen im BIBB zur Verfügung (s. Abschnitt 7.3).

²⁴Die Mitarbeiterzahl wurde dabei nicht direkt erhoben. Auf Basis der Angaben zur Selbstständigengruppe wurde folgende Zuordnung von Ausprägungen bei der Frage 2610 zur Mitarbeiteranzahl vorgenommen: 31=5, 32=11, 33=1, 34=0.

Tabelle 9: Variablenübersicht für den Ereignisdatensatz (ohne Textvariablen) der BIBB-Übergangsstudie 2006

Themenblock	Anzahl Variablen	Variablenbereich	Fragebogen s.
Biographie im Anschluss an die Grundschule (2000)	358	begmo-q2862	11-101
0. Strukturblock: Aktivitäten (2000)	6	begmo-q1002	11-13
1. Schulbesuch (2100)	57	q2100kor-q2123i	14-27
2. Berufl. Ausbildung, Lehre (2200)	52	q2200- q2224ha	28-39
3. Studium (2300)	36	q2300-q2315ha	40-48
4. Berufsvorbereitung u.a. (2400)	40	q2400-q2417ha	49-58
5. Einstiegsqualifizierung, Praktikum und Teilqualifizierung (2450)	35	q2450-q2464ha	59-66
6. Erwerbstätigkeit, Jobben (2600)	32	q2602-q2616ha	67-76
7. Langfristige Fortbildung oder Umschulung (2700)	20	q2700-q2708d	77-80
8. Maßnahmen der BA (2900)	27	q2900-q2190ha	81-87
9. Wehr-/Zivildienst, Freiwilliges Soz. od. Ökologisches Jahr (2800)	25	q2800-q2810ha	88-94
10. Sonstiges oder zu Hause (2850)	28	q2850-q2862	95-101

Zusätzlich gebildete Variablen²⁵:	35 / 33 (SUF)	lfd_neu index_n zuspell q1002 q2100kor q2108_neu ausbild bkz_2201 (SUF: bkz_2201_2digit) AUSB_KLDB92 (SUF: AUSB_KLDB92_2digit) AUSB_I88 (SUF: AUSB_I88_2digit) AUSB_SIOPS AUSB_ISEI AUSB_MPS i88_a (DFV, GWA) sempl_a supvis_a egp_a BBR_KLDB92 (SUF: BBR_KLDB92_2digit) BBR_I88 (SUF: BBR_I88_2digit) BBR_SIOPS BBR_ISEI BBR_MPS i88_b (DFV, GWA) sempl_b supvis_b egp_b q2109_sonst- q2904_sonst	s. Text (5.1)
Gesamt GWA, DFV / SUF	393 / 391 (SUF)		

5.3 Textvariablen

Zusätzlich zu den numerischen Variablen im Querschnitt- und des Ereignisdatensatz stehen die Textvariablen der BIBB-Übergangsstudie 2006 für Analysen zur Verfügung. Aus Datenschutzgründen bzw. eines höheren Reidentifikationsrisikos können die Textvariablen nur im Gastwissenschaftleraufenthalt analysiert werden. Die Textvariablen des Querschnitt- bzw. des Ereignisdatenteils sind als separate Datensätze verfügbar und enthalten die den Tabellen 10 und 11 zu entnehmenden Textvariablen.

²⁵ Der Scientific-Use-File enthält keine Textvariablen. Sofern es notwendig war, wurden Indikatorvariablen für die offenen Angaben eingefügt (s. Abschnitt 6.3). Diese Indikatorvariablen sind in den Abschnitten 5.1 und 5.2 nicht separat aufgeführt.

Tabelle 10: Textvariablen im Querschnittteil der BIBB-Übergangsstudie 2006

Fragebo- gennr.	Variableninhalt	Variablenname
4000	Textvariable: Herkunftsland: offene Nennungen (n=197)	q4000o
4005	Textvariable: zuerst gelernte Sprache: offene Nennungen (n=148)	q4005o
4006	Textvariable: mehrere gelernte Sprachen: offene Nennungen (n=27)	q4006o
4007	Textvariable: früher zuhause meistens gesprochene Sprache: offene Nennungen (n=8)	q4007o
4008	Textvariable: Staatsangehörigkeit 1: offene Nennungen“ (n=166)	q4008o1
4008	Textvariable: Staatsangehörigkeit 2: offene Nennungen“ (n=27)	q4008o2
5009	Textvariable: Ausbildung des Vaters „anderer Ausbildungsabschluss: offene Nennungen (n=109)	q5009o
5011	Textvariable: Ausbildung der Mutter „anderer Ausbildungsabschluss: offene Nennungen (n=117)	q5011o
5014	Textvariable: Berufliche Tätigkeit des Vaters: offene Nennungen (n=696)	q5014o
5022	Textvariable: Berufliche Tätigkeit der Mutter: offene Nennungen (n=848)	q5022o

Tabelle 11: Textvariablen im Ereignisdatenteil der BIBB-Übergangsstudie 2006

Fragebo- gennr.	Variableninhalt	Variablenname
2100	Textvariable: Besucher Schultyp eine andere Schule: offene Nennung (n=180)	q2100o
2102	Textvariable: Besucher Schultyp spezifisch, allgemeinbildende Schule (n=3)	q2102o
2103	Textvariable: Besucher Schultyp spezifisch, Fachschulen (n=101)	q2103o
2109	Textvariable: Gründe für Schulbesuch: andere, und zwar	q2109o
2112	Textvariable: Gründe für Abbruch des Schulbesuchs: offene Nennung (n=46)	q2112o
2201	Textvariable: Ausbildungsberuf: offene Nennungen (n=3940)	q2201o
2206	Textvariable: Ausbildungsstelle gefunden durch „anderes“: offene Nennung (n=169)	q2206o
2213	Textvariable: Gründe für Ausbildungsabbruch durch „sonstige Gründe“: offene Nennung (n=11)	q2213o
<i>Fortsetzung Tabelle 11</i>		
2214	Textvariable: Gründe für Ausbildungsabbruch „persönlich“, spezifische Nachfrage: offene Nennung (n=37)	q2214o
2215	Textvariable: Gründe für Ausbildungsabbruch „zu schwierig“, spezifische Nachfrage: offene Nennung (n=5)	q2215o
2301	Textvariable: Studienfach: offene Nennung (n=341)	q2301o
2301a	Textvariable: zweites Hauptfach: offene Nennung (n=48)	q2301ao

2308	Textvariable: Abbruchgrund Studium: offene Nennung (n=6)	q2308o
2309	Textvariable: Abbruchgrund „persönlich“, spezifische Nachfrage „anderer Grund“: offene Nennung (n=15)	q2309o
2310	Textvariable: Abbruchgrund „zu schwierig“, spezifische Nachfrage „anderer Grund“: offene Nennung (n=4)	q2310o
2311	Textvariable: Angestrebter Studienabschluss „andere“: offene Nennung (n=34)	q2311o
2401	Textvariable: Nachfrage Art der Berufsvorbereitung „Sonstiges“: offene Nennung (n=37)	q2401o
2403	Textvariable: Berufsfeld der Berufsvorbereitung: offene Nennung (n=168)	q2403o
2404	Textvariable: Fachrichtung der Berufsvorbereitung: offene Nennung (n=109)	q2404o
2408	Textvariable: Gründe für Lehrgang „aus anderem Grund“: offene Nennung (n=34)	q2408o
2411	Textvariable: Abbruchgrund: offene Nennung (n=5)	q2411o
2412	Textvariable: Abbruchgrund „persönlich“, spezifische Nachfrage „anderer Grund“: offene Nennung (n=15)	q2412o
2455	Textvariable: Bereich der Einstiegsqualifizierung: offene Nennungen (n=519)	q2455o
2456	Textvariable: Grund für Praktikum „anderer Grund“: offene Nennungen (n=66)	q2456o
2457	Textvariable: Grund für Einstiegsqualifizierung „aus anderem Grund“: offene Nennungen (n=30)	q2457o
2460	Textvariable: Abbruchgrund „sonstige Gründe“: offene Nennungen (n=6)	q2460o
2461	Textvariable: Abbruchgrund „persönlich“, spezifische Nachfrage „anderer Grund“: offene Nennung (n=3)	q2461o
2604	Textvariable: berufliche Tätigkeit: offene Nennungen (n=4470)	q2604o
2612	Textvariable: Weg in die Beschäftigung: offene Nennungen (n=223)	q2612o
2707	Textvariable: Abbruchgrund „sonstige Gründe“: offene Nennungen (n=3)	q2707o
2900	Textvariable: Art der Maßnahme „sonstige BA-Maßnahme“: offene Nennungen (n=33)	q2900o
2904	Textvariable: Grund für Maßnahme „aus anderem Grund“: offene Nennungen (n=30)	q2904o
2805	Textvariable: Grund für Verlängerung des Wehrdienstes „sonstige Grund“: offene Nennungen (n=60)	q2805o
2850	Textvariable: Beschäftigung während der Beschäftigungslosigkeit (sonstiges oder zu Hause) „anderes“: offene Nennungen (n=337)	q2850o

6 Datenzugang und Anonymisierung

6.1 Datenzugang

Der Querschnitts- und der Ereignisdatensatz der BIBB-Übergangsstudie 2006 können als Scientific-Use-File (SUF), über die Datenfernverarbeitung und an den Gastarbeitsplätzen im BIBB in Bonn analysiert werden. Diese drei Formen des Zugangs unterscheiden sich sowohl hinsichtlich der Anonymität der nutzbaren Daten als auch in der Art der Datenbereitstellung. Die Unterschiede zwischen den Datensatzversionen für die BIBB-Übergangsstudie sind im Abschnitt 6.3 dargestellt.

SUF bedienen Forschungsinteressen und werden in anonymisierter Aufbereitung auf einem geeigneten Datenträger den Nutzern überlassen. Die Nutzung dieser anonymisierten Mikrodatenfiles ist an Richtlinien gebunden und erfordert eine förmliche Genehmigung durch das BIBB.

Die kontrollierte Datenfernverarbeitung erlaubt die Analyse formal anonymisierter Originaldaten über die Verarbeitung lauffähiger, vom Nutzer erstellter Syntaxprogramme im BIBB-FDZ. Die Datennutzer können hierfür auf den Internetseiten des BIBB-FDZ Strukturdatensätze (Spieldaten) herunterladen, die in Aufbau und Merkmalsausprägungen dem Originalmaterial gleichen. Mittels dieser Spieldaten können Auswertungsprogramme in den Analyseprogrammen Stata oder SPSS erstellt werden, mit denen das BIBB-FDZ anschließend die Originaldaten auswertet. Der auf Vertraulichkeit geprüfte Output wird schließlich an die Nutzer zurückgeschickt.

GWA erlauben die Analyse schwach anonymisierter Mikrodaten an den abgeschotteten PC-Arbeitsstationen des BIBB-FDZ in einem eigenen Gästeraum. Über einen Gastaufenthalt lassen sich beispielsweise die aus den SUFs entfernten Sondervariablen bzw. vergrößerten Merkmale in ihrem originalen Wertebereich analysieren.

Die Nutzung der Daten erfordert für alle drei Datenzugangswege eine förmliche Beantragung. Die Nutzungsanträge für alle Zugänge können auf den Internetseiten des BIBB-FDZ (www.bibb-fdz.de) heruntergeladen werden.

6.2 Bestimmungen des Datenschutzes

Der Zugang zu den SUF-Daten des BIBB-FDZ erfolgt ausschließlich unter dem Gebot der Datensparsamkeit und der Einhaltung geltender Datenschutzregeln, insbesondere Artikel 89 der Verordnung (EU) 2016/679 (Datenschutz-Grundverordnung DS-GVO) und § 27 Bundesdatenschutzgesetz (BDSG 2018). Demnach dürfen für die Durchführung wissenschaftlicher Vorhaben Forschungsdaten zeitlich befristet an die unabhängige wissenschaftliche Forschung weitergegeben werden, wenn die Herstellung eines Bezugs zu einer Erhebungseinheit nicht möglich ist ("Anonymität"). Um dies ausnahmslos herzustellen,

sind vom Daten empfangenden Institut bzw. den Forschenden besondere technische und organisatorische Datensicherungsmaßnahmen einzurichten, die Unbefugten keinen Datenzugriff ermöglichen. Es wird empfohlen, dass die zugangs- und nutzungsberechtigten Forscher/-innen auf die Einhaltung von Datenschutzbestimmungen besonders verpflichtet sind (vgl. SUF-Nutzungsrichtlinien des BIBB-FDZ).

Die Mitarbeiter/-innen des BIBB-FDZ nehmen nur zum Zwecke der Beratung, der Verbesserung des BIBB-FDZ-Services sowie zur Gewährleistung der Einhaltung des Datenschutzes Einblick in Forschungsfragen, Methoden und Analysen der Wissenschaftler/-innen. Mitarbeiter/-innen des BIBB, die nicht dem BIBB-FDZ angehören, erhalten keinen Einblick in die Tätigkeiten der Wissenschaftler/-innen.

6.3 Anonymisierungen nach Datenprodukten

In allen Datensatzversionen stehen die Identifikatoren aus dem Originaldatensatz für eine Auswertung nicht zur Verfügung bzw. wurden durch ein neu gebildetes, systemfreies Personenkennzeichen (lfd_neu) ersetzt. Damit keine Rückschlüsse auf den originalen Personenidentifikator mehr gezogen werden können, wurde der Datensatz nach einer Zufallszahl neu sortiert. Anschließend wurden die Befragungspersonen mithilfe einer vierstelligen Nummer fortlaufend durchnummeriert.

Die für die einzelnen Datensatzversionen nach Zugangswegen vorgenommenen Anonymisierungen (Vergrößerungen und Löschungen) können der folgenden Auflistung entnommen werden:

6.3.1 Anonymisierung: SUF

Dem Scientific-Use-File (SUF) liegen jeweils eine anonymisierte Version des Querschnitt-(Q) und des Ereignisdatensatzes (E) exklusive den Textvariablen²⁶ bei. Auf den FDZ-Internetseiten (www.bibb-fdz.de) stehen formal anonymisierte Version der SUFs (Spieldaten) zum Download bereit, anhand derer sich Nutzer über die Variablen und Ausprägungen im SUF informieren können.

Tabelle 12 unten weist die anonymisierten Variablen im SUF aus.

²⁶ Die Textvariablen stehen für Auswertungen über diesen Datenzugangsweg nicht zur Verfügung. Die Textvariablen können aus Datenschutzgründen nur im Gastaufenthalt im BIBB analysiert werden.

Tabelle 12: Anonymisierte Variablen im SUF der BIBB-Übergangsstudie 2006

Frage- bogen- nr	Variableninhalt	Anonymisierung und Variablennamen (in Klammern)	Daten- satz
-	Wohnort (automatisch generiert)	Bundesland (bland), West-Ost-Dummy (westost)	Q
2201	Ausbildungsberuf (KldB92 und ISCO88 4-Steller)	Vergrößert auf KldB92, ISCO-88: 2-Steller (AUSB_KLDB92_2digit, AUSB_I88_2digit ²⁷)	E
2604	Berufliche Tätigkeit (Befragter, KldB92 und ISCO88 4-Steller)	Vergrößert auf KldB92, ISCO-88: 2-Steller (BBR_KLDB92_2digit, BBR_I88_2digit ³⁰)	E
5014	Berufliche Tätigkeit des Vaters (KldB92 und ISCO88 4-Steller)	Vergrößert auf KldB92, ISCO-88: 2-Steller (VBR_KLDB92_2digit, VBR_I88_2digit ³⁰)	Q
5022	Berufliche Tätigkeit der Mutter (KldB92 und ISCO88 4-Steller)	Vergrößert auf KldB92, ISCO-88: 2-Steller (MBR_KLDB92_2digit, MBR_I88_2digit ³⁰)	Q
4000	Herkunftsland	Deutschland, EU (außer DE), Nicht-EU-Land (q4000_gr)	Q
4008	Staatsangehörigkeit	Deutschland, EU (außer DE), Nicht-EU-Land (q4008_de-q4008_98)	Q
5031	Anzahl der Telefonnummern des Haushalts	gelöscht	Q
i88_a, i88_b,	EGP-ISCO-Hilfsvariablen zur Rekodierung in EGP-Klassen (siehe Text)	gelöscht	E
i88_v, i88_m	EGP-ISCO-Hilfsvariablen zur Rekodierung in EGP-Klassen (siehe Text)	gelöscht	Q

6.3.2 Formale Anonymisierung: Datenfernverarbeitung

Für die Datenfernverarbeitung stehen im Internet unter www.bibb-fdz.de formal anonymisierte Versionen des Scientific-Use-Files (SUF) des Querschnitt- (Q) und des Ereignisdatensatzes (E) der BIBB-Übergangsstudie 2006 zum Download zur Verfügung.

²⁷ In den Ausgangsvariablen (4-Steller) sind ZUMA-Zusatzcodes enthalten: 7510 „ZUMA-Zusatzcode: Schlosser o.n.A.“, 7520 „ZUMA-Zusatzcode: Elektroniker o.n.A.“, 7900 „ZUMA-Zusatzcode: Aufsichtskräfte in der Produktion“. Diese wurden im 2-Steller in die Kategorie 74 „Sonstige Handwerks- und verwandte Berufe“ rekodiert.

Spieldaten der schwächer anonymisierten Versionen von Zusatzvariablen (z.B. tiefer gegliederte Regional- oder Berufsangaben) können *im Einzelfall* vom BIBB-FDZ angefordert werden. Voraussetzung dafür ist mindestens ein gültiger Nutzungsvertrag für Gastwissenschaftleraufenthalte.

Tabelle 13: Anonymisierte Variablen bei der DFV der BIBB-Übergangsstudie 2006

Fragebo- genr.	Variableninhalt	Anonymisierung	Daten -satz
-	Wohnort (automatisch generiert)	Raumordnungsregion (ror), Arbeitsamtsbezirk (aab)	Q
5031	Anzahl der Telefonnummern des Haushalts	gelöscht	Q

6.3.2 Projektbezogene Anonymisierung: Gastrechnen im BIBB

An den Gastwissenschaftlerarbeitsplätzen im BIBB-FDZ stehen projektbezogen anonymisierte Versionen des Querschnitt- (Q) und des Ereignisdatensatzes (E) inklusive den Textvariablen zur Verfügung. Formal anonymisierte Versionen der anonymisierten Daten (SUFs) exklusive Textvariablen können im Internet unter www.bibb-fdz.de heruntergeladen werden. Spieldaten der schwächer anonymisierten Versionen von Zusatzvariablen (z.B. tiefer gegliederte Regional- oder Berufsangaben) können *im Einzelfall* vom BIBB-FDZ angefordert werden. Voraussetzung dafür ist mindestens ein gültiger Nutzungsvertrag für Gastwissenschaftleraufenthalte.

Tabelle 14: Anonymisierte Variablen im GWA der BIBB-Übergangsstudie 2006

Fragebo- genr.	Variableninhalt	Anonymisierung	Daten -satz
-	Wohnort (automatisch generiert)	Raumordnungsregion (ror), Arbeitsamtsbezirk (aab)	Q
5031	Anzahl der Telefonnummern des Haushalts	gelöscht	Q

7 Arbeitshilfen

Auf den Internetseiten der BIBB-Übergangsstudie im BIBB-FDZ (BIBB-FDZ, 2010) stellen wir verschiedene Datensatz-spezifische Arbeitshilfen zum Download zur Verfügung. Hierzu gehören neben diesem Datenhandbuch der Fragebogen der BIBB-Übergangsstudie 2006, Testdaten (vgl. Abschnitt 6.1), die Methodenberichte von infas zur Hauptstudie und zur Nacherhebung des Bundeslandes der besuchten Schule (vgl. Schiel u.a. 2006, infas 2007), und eine Zusammenstellung von wichtigen Veröffentlichungen mit den Daten.

In den folgenden Abschnitten informieren wir über die Bildung der Gewichtungvariable (7.1), die Deklaration der fehlenden Werte nach dem BIBB-FDZ-Standard (7.2), den Umgang mit Ereignisdaten (7.3), das im Gastwissenschaftleraufenthalt zur Verfügung gestellte Visualisierungsprogramm zu den Daten (7.4), und das thematische und methodische Forschungspotential der Daten (7.5).

7.1 Gewichtung

Für die Berücksichtigung der durch das Stichprobenverfahren bedingten unterschiedlichen Auswahlwahrscheinlichkeiten auf Haushalts- und Personenebene (s. Abschnitt 4.1) und für die Ausgleichung der Unterschiede in den demographischen Kennwerten der Stichprobe und der Grundgesamtheit (s. Abschnitt 4.3) wurde vom Befragungsinstitut ein Redressmentgewicht (*gewicht*) aus Populations- und Designkomponente erstellt und dem Querschnitt Datensatz beigelegt (vgl. Schiel et al. 2006, S. 19 ff.).

Für die Populationskomponente wurden die Merkmale Geschlecht, Geburtsjahrgang, höchster Schulabschluss und Bundesland einbezogen. Das Designgewicht wurde aus der Anzahl der Zielpersonen im Haushalt (Jugendliche im Alter von 18 bis 24 Jahre, q103) geteilt durch die Anzahl der Festnetznummern (q5031) berechnet.

Die Gewichte liegen im Datensatz in einer Variable (*gewicht*) vor.

Beicht und Friedrich (2008, S. 96-97) stellen die Verteilung der normierten Auf- und Abwertungsfaktoren des Gesamtgewichts dar. Danach liegen die Gewichtungsfaktoren am häufigsten (37% alle Fälle) in dem Bereich der normierten Werte von 1 bis 1,5 (beidseitig). Dies entspricht Gewichtungsfaktoren von 0,67 bis 1,5. In 10 Prozent der Fälle liegen die normierten Abwertungsfaktoren bei Werten von fünf und mehr (Gewichte von 0,18 oder weniger) und in zwei Prozent der Fälle bei Werten der Aufwertungsfaktoren von fünf und mehr.

Die z.T. hohen Gewichtungsfaktoren in der BIBB-Übergangsstudie können bei Analysen bestimmter Teilgruppen zu starken Unterschieden zwischen gewichteten und ungewichteten Analysen führen. Grundsätzlich empfehlen wir den Vergleich und ggfs. auch die Berichterstattung von gewichteten und ungewichteten Punktschätzern.

Beicht und Friedrich (2008, S. 97-98) konstatieren, dass die hohen Gewichtungsfaktoren unbestritten eine Schwäche des Datensatzes darstellen und diskutieren, inwieweit es für zukünftige Befragungen von Jugendlichen einer bestimmten Altersgruppe im BIBB vorteilhafter sein könnte, auf eine Zufallssteuerung der Stichprobe zu verzichten und stattdessen eine nach Erreichbarkeit geschichtete Auswahl vorzunehmen.

7.2 Deklaration der fehlenden Werte nach dem BIBB-FDZ-Standard

Fehlende und ungültige Werte wurden nach der Standarddeklaration fehlender Werte des BIBB-FDZ (vgl. Tab. 11) für alle Variablen der BIBB-Übergangsstudie 2006 umkodiert, um den Umgang mit den fehlenden Werten zu erleichtern. In vielen Fällen wurden Filtermissings neu angelegt, da sie im Ursprungsdatensatz nicht vorhanden waren. Im Längsschnitteil wurde auf eine Deklaration der Filtermissings verzichtet. Die Ereignisdatenstruktur ist sehr komplex und z.T. mit mehreren, ineinander genetzten Bereinigungsschleifen versehen. Je nach Auswertungsinteresse könnten die Fallzahlen der gefilterten Fälle variieren. Eine Setzung der Filtermissings käme vor diesem Hintergrund ggfs. einem Eingriff in die Logik der Daten gleich.

In Datensätzen des BIBB-FDZ werden grundsätzlich keine MISSING VALUES definiert. Das entscheiden Anwender selbstständig.

Tabelle 16 gibt die Werte und Wertelabels für die fehlenden Werte bei allen Variablen in der BIBB-Übergangsstudie wider.

Tabelle 16: Standarddeklaration des BIBB-FDZ für verschiedene Arten fehlender Antwortwerte

Wert	Wertelabel
-9	keine Angabe
-8	weiß nicht
-7	trifft nicht zu
-6	ungültig (z.B. Ausschluss durch Datenbereinigung)
-5	Verweigert
-1	Filter

Von der Standarddeklaration wurde in einigen wenigen Fällen insofern abgewichen, als dass bei dichotomen Variablen mit den Ausprägungen 1 „trifft zu“ und 0 bzw. 2 „trifft nicht zu“ letztere nicht in den Wert „-7“ umkodiert wurden (q2206_1-q2206_98, q2213_1-q2213_sonst, q2308_1-q2308_sonst, q2408_1-q2408_sonst, q2411_1-q2411_sonst, q2457_1-q2457_sonst, q2460_1-q2460_sonst, q2704_1-q2704_4, q2707_1-q2707_sonst, q2904_1-q2904_sonst).

In der Variable q205 (wie gerne zur Grundschule gegangen) waren 10 fehlende Werte enthalten, die sich nicht mit einer Filterregel zuordnen ließen. Diese wurden in -6 (ungültig) rekodiert.

In der Variable q5028 waren 97 Fälle mit dem Wert 99 („keine Angabe“) enthalten ist, welcher im Fragebogen nicht vorhanden ist. Die Fälle wurden gemäß der FDZ-Standarddeklaration in -9 („keine Angabe“) rekodiert.

In folgenden Fällen wurden die Standard-Wertelabels variablenspezifisch ergänzt:

- **q205** (Wie gerne zur Grundschule gegangen): -7 "trifft nicht zu (keine Grundschule besucht)" [recode Wert 6]
- **q2102** (Schultyp): -7 "trifft nicht zu (keine Schule besucht)" [recode Wert 95], -6 "ungültig (genannte Schule nicht auf der Liste)" [recode Wert 96]
- **q2120_1-q2120i_1** (berufl. bzw. schulische Pläne): -6 "ungültig (zu Feldbeginn nicht erhoben)" [recode Wert 9, nicht im Fragebogen vorhanden]
- **q2212, q2311, q2410, q2459, q2706**: -8 "weiß (noch) nicht"
- **q5000** (Vater in Deutschland geboren): -8 "weiß nicht (Vater unbekannt)" [recode Wert 8]
- **AUSB_KLDB92, AUSB_KLDB92_2digit (SUF), BBR_KLDB92, BBR_KLDB92_2digit (SUF), VBR_KLDB92, VBR_KLDB92_2digit (SUF), MBR_KLDB92, MBR_KLDB92_2digit (SUF)**, (KldB92-Variablen für den Ausbildungsberuf und den ausgeübten beruflichen Tätigkeit des Befragten, KldB92-Variablen für die ausgeübte berufliche Tätigkeit des Vaters und der Mutter): -8 "weiß nicht - Beruf nicht bekannt/benennbar", -7 "trifft nicht zu (keine Berufstätigkeit)", -6 "ungültig (unzureichende Angabe - nicht codierbar)"
- **AUSB_I88, AUSB_I88_2digit (SUF), i88_a, BBR_I88, BBR_I88_2digit (SUF), i88_b, VBR_I88, VBR_I88_2digit (SUF), MBR_I88, MBR_I88_2digit (SUF)**, (ISCO88-Variablen für den Ausbildungsberuf und die ausgeübte berufliche Tätigkeit des Befragten, ISCO88-Variablen für die ausgeübte berufliche Tätigkeit des Vaters und der Mutter): -7 "trifft nicht zu (keine Berufstätigkeit)", -6 "ungültig (unzureichende Angabe - nicht codierbar)"
- **ror** (Raumordnungsregion Wohnort des Befragten): -6 "ungültig (nicht zuzuordnen)"

In folgenden Fällen war es notwendig, zusätzliche Codes für fehlende Werte zu vergeben.

- **bkz_2201, bkz_2201_2digit (SUF)**, BIBB-KldB92-Variable für den Ausbildungsberuf des Befragten): -12 "nicht identifizierbar", -11 "nicht auf der Liste", -10 "nicht lesbar"
- **AUSB_KLDB92, AUSB_KLDB92_2digit (SUF), BBR_KLDB92, BBR_KLDB92_2digit (SUF), VBR_KLDB92, VBR_KLDB92_2digit (SUF), MBR_KLDB92, MBR_KLDB92_2digit (SUF)**, (KldB92-Variablen für den Ausbildungsberuf und die ausgeübte berufliche Tätigkeit des Befragten, KldB92-Variablen für die ausgeübte berufliche Tätigkeit des Vaters und der Mutter): -11 "Hausfrau, Hausmann", -10 "In schulischer Ausbildung"
- **AUSB_I88, AUSB_I88_2digit (SUF), i88_a, BBR_I88, BBR_I88_2digit (SUF), i88_b, VBR_I88, VBR_I88_2digit (SUF), MBR_I88, MBR_I88_2digit (SUF)**, (ISCO88-Variablen für den Ausbildungsberuf und die ausgeübte berufliche Tätigkeit des Befragten, ISCO88-Variablen für die ausgeübte berufliche Tätigkeit des Vaters und der Mutter): 14 "arbeitssuchend", -13 "Offiziere", -12 "Soldaten", -11 "Hausfrau, Hausmann", -10 "In schulischer Ausbildung"
- **VBR_SIOPS, MBR_SIOPS, VBR_ISEI, MBR_ISEI, VBR_MPS, MBR_MPS, i88_v, i88_m, egp_v, egp_m** (SIOPS, ISEI-, MPS-, EGP-ISCO-, und EGP-Variablen für die ausgeübte berufliche Tätigkeit von Vater und Mutter des Befragten): -14 "arbeitssuchend", -13 "Offiziere", -12 "Soldaten", -11 "Hausfrau, Hausmann", -10 "In schulischer Ausbildung"
- Bei hbild: -10 „ausländischer abschluss, nicht zuzuordnen“ bei hbild (q2113)

7.3 Arbeiten mit dem Ereignisdaten und Zusammenführung mit den Querschnittinformationen (Holger Alda)

Es gibt zwei Möglichkeiten, den Quer- und Längsschnitteil der BIBB-Übergangsstudie zusammenzuführen. Das konkrete Vorgehen kann dabei abhängig sein von der Datenorganisation für ein bestimmtes methodisches Vorgehen. So gibt es beispielsweise ereignisanalytische Verfahren, bei deren Durchführung die Daten in Form von Spells vorliegen müssen. In der Regel dürfte es aber ausreichend (und einfacher) sein, die interessierenden Informationen des Ereignisdatensatzes im Sinne der Forschungsfragestellung aufzubereiten, für jede Person eine Datenzeile zu behalten und die aufbereiteten Längsschnitt-Informationen mit dem Querschnitt-Teil zusammenzuführen. Anschließend kann wieder mit üblichen und vertrauten Auswertungsmethoden für Querschnitt-Datensätze gearbeitet werden.

In diesem Abschnitt wird das anhand eines fiktiven Auswertungsbeispiels erläutert. Auf der Ebene von STATA-Programmcodes wird beispielhaft gezeigt, wie sich einzelne Ereignisse aus dem Ereignis-Datensatz zunächst auf die Personenebene schreiben lassen und anschließend mit den Informationen aus dem Querschnittdatensatz zusammengefügt werden. Die entsprechenden Programmcodes befinden sich im Anhang 2. Im nachfolgenden Text werden Nummern für die einzelnen durchzuführenden Datenarbeiten vergeben. Die Nummern korrespondieren mit den Programmcodes im Anhang.

Die Grundidee der Arbeit mit dem Ereignisdatensatz ist wie folgt. Zuerst werden alle interessierenden Ereignisse im Ereignisdatensatz identifiziert, und zwar in der jeweiligen Datenzeile. Daran anschließend wird das Ergebnis in eine neue (Hilfs-)Variable geschrieben und anschließend auf die Personenebene übertragen. Sobald alle ereignisanalytischen Datenaufbereitungsarbeiten abgeschlossen und alle interessierenden Variablen in allen Datenzeilen zu einer Person stehen, wird jede Person im Datensatz mit einer Ereigniszeile einmal gehalten. Der letzte Schritt besteht dann darin, diesen Arbeitsdatensatz mit den Querschnittdaten zusammenzuführen.

Die fiktive Fragestellung der Beispielprogramme in STATA lautet: *wann haben wie viele Personen in der Übergangsstudie 2006 eine betriebliche Berufsausbildung mit mindestens der Note „befriedigend“ in der Abschlussprüfung abgeschlossen, die vorher irgendeine Form der Berufsvorbereitung/ der Berufgrundbildung absolviert haben? Die Ergebnisse sind nach Männern und Frauen sowie dem Alter der Personen zu differenzieren.*

Zur Beantwortung der Frage müssen aus dem Ereignisdatensatz im ersten Schritt abgeschlossene Berufsausbildungen mit mindestens der Abschlussnote „befriedigend“ identifiziert werden. Anschließend ist das Abschlussdatum zu bestimmen und in jede Datenzeile für eine Person zu übertragen. Um nach Männern und Frauen sowie nach dem Alter differenzieren zu können, sind die Ergebnisse aus dem Ereignisdatensatz anschließend mit dem Querschnitt-Datensatz zusammenzuführen.

Nach dem Setzen der Parameter im Kopf des do-files und dem Einlesen des Ereignis-Datensatz (1) wird der Ereignis-Datensatz auf Personen beschränkt, die irgendeine Form der Berufsvorbereitung absolvierten (2). Dazu wird eine Hilfsvariable gebildet, die solche Ereignisse zunächst auf der Ereignisebene identifiziert und diese dann anschließend allen Ereignissen der entsprechenden Personen zuordnet. Im Ergebnis haben nun alle Personen, die irgendwann einmal eine Berufsvorbereitung absolviert haben, in *allen* ihren Ereignissen in der Variable *treffer* eine eins stehen (alle anderen: missing). So lässt sich der Datensatz einfach auf diesen Personenkreis beschränken.

Der Datensatz enthält von den ursprünglich etwa 27 Tsd. Ereignissen noch 5660 von insgesamt 1120 Personen (`count if index_n == 1` ergibt im Ergebnis die Anzahl der Personen im Datensatz). Alle diese Personen haben einmal eine Berufsvorbereitung absolviert. Aus Vereinfachungs- und Darstellungsgründen wird inhaltlich nicht überprüft, ob die Berufsvorbereitung zeitlich vor dem Abschluss einer betrieblichen Berufsausbildung stattgefunden hat²⁸. Im nächsten Schritt werden diejenigen identifiziert, die eine Berufsausbildung abgeschlossen haben und für diese Personen die Abschlussnote überprüft und klassifiziert (3).

Zu diesem Zeitpunkt steht die interessierende Variable lediglich auf der Ereignisebene zur Verfügung. Sie wird zunächst auf alle Ereignismeldungen der Personen übertragen (4). Darüber hinaus interessiert laut Beispiel-Fragestellung das Abschlussdatum der Berufsausbildung. Hier wird analog vorgegangen, d.h. zuerst wird der interessierende Variablenwert auf der Ereignisebene in eine Hilfsvariable geschrieben, die dann abschließend als Konstante auf die Personenebene aggregiert wird (5). Nur die interessierenden Variablen werden im Datensatz gehalten. Anschließend wird – um Probleme beim Mergen mit dem Querschnittdatensatz zu vermeiden – jede Person nur mit einer Datenzeile gehalten (was ausreicht, das ja in allen Meldungen einer Person die interessierenden Ereignisse gespeichert sind) (6). Um die Dauern besser berechnen zu können, wird aus dem Ende-Monat und –Jahr anschließend ein Datum im Stata-Format gebildet (dafür ist die Angabe eines Tages erforderlich, hier wurde die Monatsmitte (15.) gewählt) (7).

Es schließt sich das Zusammenfügen mit den Querschnitt-Daten an (8). In Stata lässt sich über die Variable `_merge` aufgrund des besonderen Zuschnitts des Längsschnittdatensatzes automatisch die Personen identifizieren, die eine Berufsvorbereitung absolviert haben (`_merge == 3`) (9). Diese können nach Bedarf unterschieden werden in solche, die eine Berufsausbildung abgeschlossen haben (`gutabschluss ungleich missing`) oder nicht (`gutabschluss gleich missing`). Nun wird noch die Dauer berechnet, die der Abschluss vom Referenzdatum – hier der 15. August 2006 – entfernt ist (10). Anschließend können die zu erstellenden Auswertungen vergleichsweise einfach angefordert werden (11).

²⁸ Zumindest theoretisch könnte ja nach einer betrieblichen Berufsausbildung eine (weitere) Maßnahme zur Berufsbildungs-Vorbereitung erfolgen.

Die missings am Schluss wurden angefordert, um zu erfahren, wie viel Prozent der Personen mit einer absolvierten Berufsvorbereitung am 15. August 2006 auch einen qualifizierten Abschluss erworben haben. In den Zeilen davor wird berechnet, welche Abschlussnote sie dabei erzielten und wie viele Tage dieses Ereignis vom 15. August 2006 entfernt ist.

Die Beispielauswertung verdeutlicht ein typisches Charakteristikum der Arbeit mit Ereignisdaten. Sie ermöglichen eine Fülle von Auswertungen bei relativ hohen Freiheitsgraden an empirischen Fragestellungen. Dabei sind gewisse Vorbereitungen bei der Datenorganisation und Datensatzaufbereitung vorzunehmen um im Ergebnis einen möglichst einfach auszuwertenden Analysedatensatz zu bekommen. Selbstverständlich ist die hier beschriebene Vorgehensweise nur eine von vielen Möglichkeiten sich das Potential der BIBB-Übergangsstudie 2006 zu erschließen.

7.4 Visualisierungsprogramm

Über den Zugang und Informationen zu der BIBB-Übergangsstudie 2006 hinaus, stellt das BIBB-FDZ das Visualisierungsprogramm LDEX, Visual Explorer for Longitudinal Data für die BIBB-Übergangsstudie im Rahmen des Gastrechnens im BIBB zur Verfügung.

Das Programm wurde von den Projektverantwortlichen im BIBB in Auftrag gegeben und freundlicherweise für die Weiterverwendung frei gegeben. Autor des Programms ist Ralf Künster. Für die befragten Jugendlichen lässt sich hiermit für die zentralen soziodemographischen und bildungsbezogenen Merkmale eine Grafik der Statusverteilung, also der erfassten Aktivitäten in chronologischem Verlauf erstellen. Die folgende Beispielabbildung zeigt die Grafik der Statusverteilung für alle Jugendlichen ohne weitere Differenzierung nach soziodemographischen und bildungsbezogenen Merkmalen.

Der Benutzer hat die Möglichkeit, folgende Auswahlmöglichkeiten bzw. deren Kombination, um die Fallauswahl zu spezifizieren:

- Fälle (Person_ID)
- Kohorte: alle (1982-1988), Jahrgänge 1982 bis 1988 einzeln
- Migrationshintergrund²⁹: alle, ja, nein
- Region: alle, alte Bundesländer, neue Bundesländer

²⁹ Die Operationalisierung entspricht derjenigen der BIBB-Projektmitarbeiter der BIBB-Übergangsstudie 2006 (vgl. Beicht und Ulrich 2008, S. 102). Demnach wurde kein Migrationshintergrund angenommen, wenn Jugendliche die deutsche Staatsbürgerschaft hatten, wenn sie in der Familie zuerst ausschließlich die deutsche Sprache gelernt haben und Vater und Mutter in Deutschland geboren waren. Wenn eine der Bedingungen nicht zutraf, wurde ein Migrationshintergrund angenommen.

- Schulabschluss³⁰: alle, Hauptschule und niedriger, mittlerer Abschluss, Abitur, FHS Reife, noch kein Abschluss, keine Angabe
- Schulabgangsjahrgang: alle, Jahrgänge 1996 bis 2006 einzeln

Neben der Spezifizierung von Subgruppen für die Abbildung der Statusverteilungen können in dem Programm Anpassungen der Grafikeinstellungen vorgenommen werden. Diese betreffen die Auswahl folgender Merkmale bzw. Ausprägungen:

- Zeitbezug:
 - Historische Zeit → Zeitachsenstart / Zeitachsenende: *Eingabe in Monaten seit Januar 1900 (jeweils 2-stellige Jahresangabe*12 + Monat)*
 - Alter → Zeitachsenstart / Zeitachsenende: *Eingabe in Monaten seit Geburt*
 - Ende der allgemeinbildenden Schule → Zeitachsenstart / Zeitachsenende: *Eingabe in vom Schulabgangszeitpunkt abweichenden Monaten (positive und negative Werte möglich)*
- Gewichtung: Gewichtete Grafik, ungewichtete Grafik
- Priorität bei Parallelität: 10 Aktivitätsarten (vgl. Abschnitt 3.2)
- Reihenfolge der Anzeige: 10 Aktivitätsarten (vgl. Abschnitt 3.2)

Die Visualisierung der Statusverteilungen im Zeit- bzw. Kohortenverlauf ist zunächst ein Hilfsmittel, die z.T. komplizierten Muster in den Verlaufsdaten anschaulich abzubilden. Hiermit ermöglicht sie u.U. ein besseres Verständnis über die zugrunde liegenden Muster der abgebildeten Verlaufsmodelle. Sie kann damit auch ein Analysetool für die Identifizierung wichtiger Aspekte der Daten und damit Ausgangspunkt für statistische Analysen sein.

7.5 Thematisches und methodisches Forschungspotential

Die BIBB-Übergangsstudie 2006 ist bislang eine der wenigen aktuellen und für Deutschland repräsentativen Längsschnittstudien zum Bildungs- und Berufsverhalten von Jugendlichen und jungen Erwachsenen. Darüber hinaus erlaubt sie im Gegensatz zu vielen anderen Studien aufgrund der sorgfältigen Erhebung und Nachkontrolle die Unterscheidung zwischen wichtigen Differenzierungen im Bereich der Berufsausbildung (Beicht 2009, vgl. Konietzka 2007) – betrieblichen und schulischen Ausbildungen, verschiedene Institutionen im Übergangssystem (Berufsvorbereitung, Berufsgrundbildung, Einstiegsqualifizierung), und die Art der Hochschulbildung (u.a. Hauptfächer). Im Hinblick auf Ausbildungschancen und

³⁰ Der Schulabschluss wurde unter Verwendung der Angaben bei den Fragen 2113 und 2106 (Schulbesuch dauert noch an) verwendet. Die Zuordnungen der Merkmale der Variablen zu den Kategorien sind: „Hauptschule und niedriger“: Abschluss der Sonder- bzw. Förderschule (1), Hauptschulabschluss (2), qualifizierender Hauptschulabschluss (3), ausländischer Schulabschluss, nicht den deutschen Abschlüssen zuzuordnen (8), ohne Abschluss (9); „mittlerer Abschluss“: mittlere Reife/mittlerer Abschluss/Realschulabschluss (4), erweiterter Realschulabschluss / erweiterter Sekundarabschluss I (5); „Abitur, FHS Reife“: Fachhochschulreife (6), allgemeine Hochschulreife / Abitur (7); „noch kein Abschluss“: wenn Schulbesuch noch andauert (Frage 2106=ja); „keine Angabe“: verweigert (97), weiß nicht (98).

Karriereverläufe sind neben den detaillierten Informationen zur genauen Tätigkeit zudem wichtige individuelle Informationen zu den Entscheidungsgründen für bestimmte Stationen, subjektive Bewertungen im Hinblick auf die fachliche und persönliche Entwicklung sowie die Auswirkungen (auch von Abbrüchen) auf den weiteren beruflichen Werdegang enthalten. Diese subjektiven Einschätzungen über den biographischen Verlauf werden ergänzt durch Leistungsvariablen (Schulnoten) sowie detaillierte Informationen zu Such- und Bewerbungsaktivitäten.

Neben den Informationen zum Bildungs- und Berufsverlauf der Jugendlichen enthält der Datensatz Angaben zu den Wertorientierungen, persönlichen Grundeinstellungen und Selbstwirksamkeitserwartungen der Jugendlichen.

Ein besonderes Merkmal der BIBB-Übergangsstudie 2006 ist die detaillierte Erfassung des ethnischen und sozialen Hintergrunds der Jugendlichen. Zum Migrationshintergrund wurden erstens auf die Jugendlichen bezogene Informationen erhoben, darunter das Geburtsland der Jugendlichen, die Zeit, seit der sie in Deutschland sind, die erste gelernte und meist gesprochene Sprache und ihre Staatsangehörigkeit. Zweitens wurden die Jugendlichen gebeten, zum Migrationshintergrund der Eltern Auskunft zu geben. Hier wurde erfasst, ob Vater und Mutter in Deutschland geboren wurden. Der Datensatz erlaubt damit einen Vergleich unterschiedlicher Migrationserfahrungen von verschiedenen Herkunftsgruppen für den Bildungs- und Berufsverlauf. Der soziale Hintergrund der Jugendlichen wurde ebenfalls anhand mehrerer Indikatoren und vergleichsweise aufwendig erfragt (schulischer und beruflicher Bildungsabschluss der Eltern, Beruf und berufliche Stellung des Vaters oder der Mutter). Dem Datensatz sind zusätzlich die gängigen soziologischen Klassifikationen der sozialen Herkunft von der Projektgruppe beigelegt (SIOPS, ISEI, MPS, EGP). Die Daten erlauben damit ein reiches Spektrum an Analysen einzelner Bildungsübergänge und von Bildungssequenzen der Jugendlichen in einer Lebensverlaufsperspektive unter Berücksichtigung primärer und sekundärer Herkunftseffekte.

Im Gastwissenschaftleraufenthalt und bei der Datenfernverarbeitung können überdies Auswertungen auf kleinräumlicher Ebene (Raumordnungsregionen, Arbeitsamtbezirke) vorgenommen werden.

Im Hinblick auf die Ausschöpfung des Datenpotentials in methodischer Hinsicht u.a. verschiedene Verfahren von Verlaufsdatenanalysen angegangen werden: Beicht und Ulrich (2008, S. 165ff.) haben z.B. Sequenzmusteranalysen verwendet, um typische Sequenzen der nicht studienberechtigten Jugendlichen zu identifizieren. Für die Frage, wie lange bestimmte Übergänge – etwa die von der Schule in eine Ausbildung oder ein Studium – dauern, können z.B. Kaplan-Meier-Schätzer berechnet werden (vgl. ebd., S. 195 ff.). Um zu bestimmen, welche Faktoren die Dauer bestimmen, können Cox-Regressionen eine geeignete Analyseverfahren sein (vgl. ebd., S. 201ff.). Eine weitere Analysestrategie könnte sein, die Determinanten der verschiedenen Verlaufstypen mit Hilfe von multinomialen Logitmodellen zu untersuchen. Wenn die Übergänge in einzelne Aktivitäten im Forschungsinteresse stehen, können Panelmodelle der Wahrscheinlichkeiten, einen bestimmten Übergang zu absolvieren, eine geeignete Methode sein.

Im Folgenden sind bereits veröffentlichte Arbeiten aufgeführt, die auf Auswertungen der BIBB-Übergangsstudie 2006 basieren.

Beicht, U., 2009: Verbesserung der Ausbildungschancen oder sinnlose Warteschleife? Zur Bedeutung und Wirksamkeit von Bildungsgängen am Übergang Schule – Berufsausbildung. BIBB REPORT 11/09. BIBB: Bonn.

Beicht, U., Granato, M., 2009: Übergänge in eine berufliche Ausbildung. Geringere Chancen und schwierige Wege für junge Menschen mit Migrationshintergrund - Expertise des Gesprächskreises Migration und Integration der Friedrich-Ebert-Stiftung. In: Friedrich-Ebert-Stiftung (Hrsg.): WISO Diskurs, Bonn: bub Bonner Universitäts-Buchdruckerei.

Beicht, U., Granato, M., Ulrich, J. G., 2009: Übergänge von Jugendlichen mit Migrationshintergrund in eine vollqualifizierende Berufsausbildung. In: Bundesinstitut für Berufsbildung (Hrsg.): Datenreport zum Berufsbildungsbericht 2009. Informationen und Analysen zur Entwicklung der beruflichen Bildung, Kapitel A5.8.1, S. 164-170.

Beicht, U., Ulrich, J. G., 2009: Determinanten von Ausbildungslosigkeit und typische bildungsbiographische Verläufe von Jugendlichen ohne Berufsausbildung. In: Bundesinstitut für Berufsbildung (Hrsg.): Datenreport zum Berufsbildungsbericht 2009. Informationen und Analysen zur Entwicklung der beruflichen Bildung, S. 217-222.

Beicht, U., Ulrich, J. G., 2008: Welche Jugendlichen bleiben ohne Berufsausbildung? Analyse wichtiger Einflussfaktoren unter besonderer Berücksichtigung der Bildungsbiografie. BIBB REPORT 6/08. BIBB: Bonn.

Beicht, U., Ulrich, J. G., 2008: Übergänge von der allgemeinbildenden Schule in eine vollqualifizierende Ausbildung - Ergänzende Analysen für den zweiten nationalen Bildungsbericht zum Schwerpunktthema "Übergänge im Bildungssystem und zwischen Bildungssystem und Arbeitsmarkt" auf Basis der BIBB-Übergangsstudie 2006 Arbeitspapier. Bundesinstitut für Berufsbildung, Bonn, April 2008.

Beicht, U., Friedrich, M., 2008: Anlage und Methode der BIBB-Übergangsstudie. In: Ausbildungschancen und Verbleib von Schulabsolventen / Beicht, U., Friedrich, M., Ulrich, J. G. [Hrsg.]. - Bielefeld: Bertelsmann, 2008, S. 79-100. (Berichte zur beruflichen Bildung)

Beicht, U., Ulrich, J. G., 2008: Ergebnisse der BIBB-Übergangsstudie. In: Ausbildungschancen und Verbleib von Schulabsolventen / Beicht, U., Friedrich, M., Ulrich, J. G. [Hrsg.]. - Bielefeld: Bertelsmann, 2008, S. 101-292. (Berichte zur beruflichen Bildung)

Beicht, U., Friedrich, M., Ulrich, J. G., 2008: Zusammenfassung und Diskussion zentraler Ergebnisse (der BIBB-Übergangsstudie). In: Ausbildungschancen und Verbleib von Schulabsolventen / Beicht, U., Friedrich, M., Ulrich, J. G. [Hrsg.]. - Bielefeld: Bertelsmann, 2008, S. 293-332. (Berichte zur beruflichen Bildung).

Beicht, U., Ulrich, J. G., 2008: Ausbildungsverlauf und Übergang in Beschäftigung. Teilnehmer/innen an betrieblicher und schulischer Berufsausbildung im Vergleich. In: Berufsbildung in Wissenschaft und Praxis, 37, H. 3, 19-23.

Beicht, U., Friedrich, M., Ulrich, J. G. 2007: Deutlich längere Dauer bis zum Ausbildungseinstieg. BIBB REPORT 2/07. BIBB: Bonn.

Bundesministerium für Bildung und Forschung, 2008: Einflussfaktoren auf den Übergangsprozess in Berufsausbildung (Beicht, U., Friedrich, M., Ulrich, J. G.). In: Berufsbildungsbericht 2008, Kapitel 1.2.4, S. 82 ff.

Friedrich, M., Eberhard, V., Ulrich, J. G., 2008: Übergänge der Jugendlichen von der Schule in die Berufsausbildung : theoretische Bezüge und Ergebnisse bisheriger Studien. In: Ausbildungschancen und Verbleib von Schulabsolventen / Beicht, U., Friedrich, M., Ulrich, J. G., [Hrsg.]. - Bielefeld: Bertelsmann, 2008, S. 59-78. (Berichte zur beruflichen Bildung).

Bundesministerium für Bildung und Forschung, 2008: Wege nach Verlassen allgemein bildender Schulen (Beicht, U., Friedrich, M., Ulrich, J. G.). In: Berufsbildungsbericht 2007, Kapitel 1.3.2, S. 66 ff.

Beicht, U., Friedrich, M., Ulrich, J. G., 2007: Steiniger Weg in die Berufsausbildung - Werdegang von Jugendlichen nach Beendigung der allgemein bildenden Schule. Berufsbildung in Wissenschaft und Praxis, 36, H. 2, 5-9.

8 Literatur

Beicht, U., 2009: Verbesserung der Ausbildungschancen oder sinnlose Warteschleife? Zur Bedeutung und Wirksamkeit von Bildungsgängen am Übergang Schule – Berufsausbildung. BIBB REPORT 11/09. BIBB: Bonn.

Beicht, U., Friedrich, M., 2008: Anlage und Methode der BIBB-Übergangsstudie. In: Ausbildungschancen und Verbleib von Schulabsolventen / Beicht, U., Friedrich, M., Ulrich, J. G. [Hrsg.]. - Bielefeld: Bertelsmann, 2008, S. 79-100. (Berichte zur beruflichen Bildung)

Beicht, U., Ulrich, J. G., 2008: Ergebnisse der BIBB-Übergangsstudie. In: Ausbildungschancen und Verbleib von Schulabsolventen / Beicht, U., Friedrich, M., Ulrich, J. G. [Hrsg.]. - Bielefeld: Bertelsmann, 2008, S. 101-292. (Berichte zur beruflichen Bildung)

Beicht, U., Friedrich, M., Ulrich, J. G., 2008: Zusammenfassung und Diskussion zentraler Ergebnisse (der BIBB-Übergangsstudie). In: Ausbildungschancen und Verbleib von Schulabsolventen / Beicht, U., Friedrich, M., Ulrich, J. G. [Hrsg.]. - Bielefeld: Bertelsmann, 2008, S. 293-332. (Berichte zur beruflichen Bildung).

BIBB-FDZ, 2009: Datenübersicht. Internetseite des BIBB-FDZ. URL: www.bibb-fdz.de (letzter Zugriff 30.11.2009).

Christoph, B., 2005: Zur Messung des Berufsprestiges: Aktualisierung der Magnitude-Prestigeskala auf die Berufsklassifikation ISCO88. ZUMA-Nachrichten Nr. 57 (November 2005), S. 79-128.

Friedrich, M., Eberhard, V., Ulrich, J. G., 2008: Übergänge der Jugendlichen von der Schule in die Berufsausbildung : theoretische Bezüge und Ergebnisse bisheriger Studien. In: Ausbildungschancen und Verbleib von Schulabsolventen / Beicht, U., Friedrich, M., Ulrich, J. G., [Hrsg.]. - Bielefeld: Bertelsmann, 2008, S. 59-78. (Berichte zur beruflichen Bildung).

Gabler, S. Häder, S., 1998: Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. In: *Telefonstichproben in Deutschland / Gabler, S. Häder, S., Hoffmeyer-Zlotnik, J.* [Hrsg.]. Opladen: Westdeutscher Verlag, 1998, S. 69-88.

Ganzeboom, H.B.G., 2001: Tools from deriving status measures from ISKO-88. URL: <http://home.fsw.vu.nl/hbg.ganzeboom/isko88/index.htm>, letzter Zugriff am 31.01.2010.

Ganzeboom, H.B.G., Treiman, D. J., 2002. International Stratification and Mobility File, Standard. Recode Module ISKOE.P. INC. Revision 2002/08. URL: <http://home.scw.vu.nl/>.

Ganzeboom, H. B. G., und Treiman, D. J., 2003: Three Internationally Standardised Measures for Comparative Research on Occupational Status. S. 159-193 in: *Hoffmeyer-Zlotnik J. H.P./Wolf, C.* (Hg.), 2003: *Advances in Cross National Comparison. A European Book for Demographic and Socio-Economic Variables.* New York u.a.: Kluwer Acad./Plenum Publ.

Geis, A., 2009: Handbuch für Berufsvercodung. GESIS Center for Survey Design and Methodology. Mannheim: GESIS. URL: http://www.gesis.org/fileadmin/upload/dienstleistung/tools_standards/hdb_0903.pdf (letzter Zugriff am 30.11.2009).

GESIS, 2009: Datenzugänge im GESIS Datenarchiv – Studienarchiv. URL: <http://www.gesis.org/dienstleistungen/daten/recherche-datenzugang/datenarchiv-service/> (letzter Zugriff am 30.11.2009).

Hall, A., 2009: Die BIBB /BAuA-Erwerbstätigenbefragung 2006. Methodik und Frageprogramm im Vergleich zur BIBB / IAB-Erhebung 1998. Wissenschaftliche Diskussionspapiere, 107 (2009), BIBB.

Heckel, Christiane, 2002: Erstellung der ADM-Telefonauswahlgrundlage. In: *Telefonstichproben, methodische Innovationen und Anwendungen in Deutschland / Gabler, S. Häder, S.* [Hrsg.]. Münster: , S. 11-31.

infas Institut für angewandte Sozialwissenschaft GmbH, 2007: Bildungswege und Berufsbiographie von Jugendlichen und jungen Erwachsenen im Anschluss an allgemein bildende Schulen. Dokumentation der Nacherhebung. Bonn: infas Institut für angewandte Sozialwissenschaft GmbH.

Konietzka, D., 2007: Berufliche Ausbildung und der Übergang in den Arbeitsmarkt .In: *Bildung als Privileg? Theoretische Erklärungen und empirische Befunde zu den Ursachen der Bildungsungleichheit / Becker, R. , Lauterbach, W.* [Hrsg.]. Wiesbaden: VS Verlag, S. 273–302.

Schiel, S., Gilbert, R., Aust, F., Schröder, H. 2006: Bildungswege und Berufsbiographien von Jugendlichen und jungen Erwachsenen im Anschluss an allgemeinbildende Schulen. Methodenbericht des Instituts für angewandte Sozialwissenschaft infas, Bonn. Download unter www.bibb-fdz.de.

Statistisches Bundesamt, 2005: Statistisches Jahrbuch 2005. Für die Bundesrepublik Deutschland. Wiesbaden: Statistisches Bundesamt.

9 Anhang

9.1 Ausfallanalyse für Kreiskennziffern

Die Kreiskennziffern werden benötigt, um die an der BIBB-Übergangsstudie teilnehmenden Personen Raumordnungsregionen zuzuordnen. Die Raumordnungsregionen sind die tiefste regionale Ebene, auf der sich die BIBB-Übergangsstudie repräsentativ für die befragten Geburtskohorten auswerten lässt. In der Regel wurden die Kreiskennziffern vom Erhebungsinstitut anhand der Telefonvorwahlen bestimmt. Um etwaige Verzerrungen auszugleichen, wurden darüber hinaus zufällig generierte Telefonnummern eingesetzt. In solchen Fällen ist es nicht ohne weiteres möglich, anhand der Vorwahl die Kreiskennziffer zu bestimmen.

Für insgesamt 553 Personen fehlen die tiefgegliederten Regionalinformationen (Arbeitsagenturbezirke und Bundesland sind dann vorhanden), und so stellt sich die Frage, inwiefern regionale Analysen unter Verzerrungen aufgrund von selektiven Ausfällen leiden. Aus diesem Grund werden nachfolgend die Ergebnisse einer kleinen Ausfallanalyse kurz dargestellt.

Die Ausfallanalyse besteht in einer logistischen Regression, in der die prozentuale Risikoveränderung einer fehlenden Kreiskennziffer anhand konditionaler Merkmale der Befragten für ansonsten durchschnittliche Personen im Sample berechnet wird (so genannte marginale Effekte). Berücksichtigt werden die Merkmale Geschlecht, Alter, Herkunft, Kinder ja/nein, alte/neue Bundesländer und das Redressmentgewicht. Die Tabelle A1 zeigt das Ergebnis.

Tabelle A1: Ergebnisse einer logistischen Regression auf fehlende Kreiskennziffern (gleich eins) für Personen in der BIBB-Übergangsstudie 2006

Merkmal	Koeffizient	z-Wert	Mittelwert
männlich	0,011	1,88	0,513
<i>Geburtsjahr (Referenz: 1988)</i>			
1982	0,004	0,31	0,069
1983	-0,021	-1,83	0,098
1984	0,000	0,01	0,122
1985	0,013	1,23	0,129
1986	0,001	0,14	0,184
1987	0,006	0,65	0,213
<i>Herkunft (Referenz: D, GR, TUR)</i>			
Italien	0,120	1,02	0,001
ehem. Jugoslawien	0,031	0,57	0,003
Polen	-0,027	-1,10	0,014
Russland	-0,026	-1,76	0,047
anderes Land	-0,030	-1,61	0,027
Kinder vorhanden	-0,018	-0,88	0,017
alte Bundesländer	-0,119	-13,80	0,813
Redressmentgewicht	0,0045	2,02	1,000

Anmerkungen: Koeffizient ist dx/dy (marginaler Effekt), z-Werte werden auf der Basis heteroskedastie-konsistenter Standardfehler berechnet, Anzahl der Beobachtungen: 7226.

Die Ergebnisse zeigen wenn nur mäßige Effekte, die nicht von persönlichen Merkmalen der Befragten abhängig sind. Lediglich in den alten Bundesländern ist ein um etwa zwölf Prozent geringeres Ausfallrisiko für die Regionalinformation erkennbar. Die Gewichtungsfaktoren haben einen kleinen positiven Effekt auf die Wahrscheinlichkeit einer fehlenden regionalen Information. Mit einer Zunahme des Gewichts um 1,15 (Standardabweichung der Variable) nimmt das Ausfallrisiko um etwa ein halbes Prozent zu. Zusammenfassend lässt sich also festhalten, dass das Fehlen der Kreiskennziffer weitestgehend neutral gegenüber der Struktur des Samples ist.

9.2 Übersicht über Zusatzvariablen

Zusatzvariablen im Querschnitt Datensatz

Häufigkeitsauszählung für die Variable Herkunftsland gruppiert (q4000_gr, SUF)

Herkunftsland gruppiert	Freq.	Percent	Cum.
Weiß nicht	1	0.01	0.01
Verweigert	2	0.03	0.04
Deutschland	6,523	90.22	90.26
Anderes EU-Land	117	1.62	91.88
Nicht-EU	390	5.39	97.28
anderes Land (offene Nennung)	197	2.72	100.00
Total	7,230	100.00	

Anderes EU-Land: q4000=2 (Griechenland), 3 (Italien), 5 (Polen)

Nicht-EU: q4000=4 (ehem. Jugoslawien³¹), 6 (Russland, Kasachstan od. andere hem. Sowjetrepublik), 7 (Türkei).

Häufigkeitsauszählung für die Variable Staatsangehörigkeit gruppiert (q4008_de, SUF)

Staatsangehörigkeit:	Freq.	Percent	Cum.
deutsch			
nicht genannt	246	3.40	3.40
genannt	6,984	96.60	100.00
Total	7,230	100.00	

Häufigkeitsauszählung für die Variable Staatsangehörigkeit gruppiert (q4008_eu, SUF)

Staatsangehörigkeit:	Freq.	Percent	Cum.
----------------------	-------	---------	------

³¹ Fragebogentext „Ehemaliges Jugoslawien (Bosnien-Herzegowina, Kroatien, Mazedonien, Montenegro, Serbien, Slowenien). Slowenien ist Mitglied der EU, die Fälle können aber nicht identifiziert werden.

andere
europäische1

nicht genannt	7,127	98.58	98.58
genannt	103	1.42	100.00
Total	7,230	100.00	

1: q4008= 3 (griechisch), 4 (italienisch), 7 (polnisch), 9 (slowenisch), 10 (spanisch)

Häufigkeitsauszählung für die Variable Staatsangehörigkeit gruppiert (q4008_neu, SUF)

Staatsangehörigkeit:
nicht europäische1

	Freq.	Percent	Cum.
nicht genannt	7,090	98.06	98.06
genannt	140	1.94	100.00
Total	7,230	100.00	

1: q4008= 2 (bosnisch), 5 (kroatisch), 6 (mazedonisch), 8 (russisch), 11 (türkisch)

Häufigkeitsauszählung für die Variable Staatsangehörigkeit gruppiert (q4008_o1, SUF)

Staatsangehörigkeit:
2. offene Nennung

	Freq.	Percent	Cum.
nicht genannt	7,046	97.46	97.46
genannt	184	2.54	100.00
Total	7,230	100.00	

Häufigkeitsauszählung für die Variable Staatsangehörigkeit gruppiert (q4008_o2, SUF)

Staatsangehörigkeit:
1. offene Nennung

	Freq.	Percent	Cum.
nicht genannt	7,230	100.00	100.00
Total	7,230	100.00	

Häufigkeitsauszählung für die Variable Staatsangehörigkeit gruppiert (q4008_97, SUF)

Staatsangehörigkeit:
verweigert

	Freq.	Percent	Cum.
nicht genannt	7,229	99.99	99.99
genannt	1	0.01	100.00
Total	7,230	100.00	

Häufigkeitsauszählung für die Variable Staatsangehörigkeit gruppiert (q4008_98, SUF)

Staatsangehörigkeit:
weiß nicht

	Freq.	Percent	Cum.
nicht genannt	7,229	99.99	99.99

genannt	1	0.01	100.00
Total	7,230	100.00	

**Werte und Wertelabels der EGP-Variablen für ausgeübte Berufe Vater und Mutter
(*egp_v*, *egp_m*)**

1	i: higher managerial
2	ii: lower managerial
3	iiia: routine clerical
4	iiib: routine service-sales
5	iva: self-empl with empl
6	ivb: self-empl no empl
7	v: manual supervisors
8	vi: skilled worker
9	vii: unskilled worker
10	viii: agricul labour
11	ix: Self-empl farmer

Zusatzvariablen im Ereignisdatensatz

Werte und Wertelabels *q2108_neu* (Bundesland der besuchten Schule (Nacherhebung))

-9	keine Angabe
-8	weiß nicht
-7	trifft nicht zu
-6	ungültig
-5	verweigert
1	Schleswig-Holstein
2	Hamburg
3	Niedersachsen
4	Bremen
5	Nordrhein-Westfalen
6	Hessen
7	Rheinland-Pfalz
8	Baden-Württemberg
9	Bayern
10	Saarland
11	Berlin
12	Brandenburg
13	Mecklenburg-Vorpommern
14	Sachsen
15	Sachsen-Anhalt
16	Thüringen
17	außerhalb Deutschlands

Werte und Wertelabels *ausbild* (Art der Ausbildung, q2200 bereinigt)

-9	keine Angabe
-8	Weiße nicht
-7	trifft nicht zu
-6	ungültig
-5	verweigert
1	BBiG: betrieblich
2	BBiG: außerbetrieblich
3	BBiG: schulisch
4	Schulberuf, Beamtenausbildung
8	keine Berufsausbildung
9	keine Angabe

Werte und Wertelabels der EGP-Variablen für Ausbildungsberuf und ausgeübter Beruf des Befragten (*egp_a*, *egp_v*)

1	i: higher managerial
2	ii: lower managerial
3	iiia: routine clerical
4	iiib: routine service-sales
5	iva: self-empl with empl
6	ivb: self-empl no empl
7	v: manual supervisors
8	vi: skilled worker
9	vii: unskilled worker
10	viii: agricul labour
11	ix: Self-empl farmer

Werte und Wertelabels für *q2301* und *q2301a* (Frage 2301, Hauptfach Studium, weicht von Fragebogen ab)

Werte Wertelabels

-9	keine Angabe
-8	weiß nicht
-7	trifft nicht zu
-6	ungültig
-5	verweigert
1	Wirtschaftswissenschaften
2	Rechtswissenschaften
3	Politik-Sozialwissenschaften
4	Sozialwesen
5	Verwaltungswissenschaft
6	sonstiges wirtschafts- oder sozialwissenschaftliches
7	Informatik
8	Mathematik
9	Biologie
10	sonstiges naturwissenschaftliches Fach
11	Germanistik
12	Erziehungswissenschaften
13	Psychologie
14	sonstiges sprach- oder kulturwissenschaftliches Fach
15	Humanmedizin (einschl. Zahnmedizin)
16	Maschinenbau, Verfahrenstechnik
17	Elektrotechnik

18 ArchitekturInnenarchitektur
19 Bauingenieurwesen
20 sonstiges ingenieurwissenschaftliches Fach
21 Kunst, Kunstwissenschaften
22 Agrar-, Forst- und Erziehungswissenschaften
23 Sport, Sportwissenschaft
24 Veterinärmedizin
25 Sonstiges Fach (offene Angabe)

95 Nein, kein weiteres Hauptfach (nur q2301a)

9.3 Beispielprogramm zur Auswertung der Übergangsstudie

```
*****
*** Beispielprogramm zur Auswertung der Übergangsstudie (STATA do-file) ***
*****
```

```
*****
*** (1) ***
*****
```

```
clear
set more off
set logtype text
set mem 300m
set matsize 2000
cap log close
```

```
cd "<Laufwerk>"
use <Laufwerk>\Uebergangsstudie2006_Ereignisdatensatz_Gesamt.dta, clear
log using log\beispielsauswertung_übergangsstudie.log, replace
```

```
quietly mvdecode _all, mv(-2,-7,-8,-9)
```

```
*****
*** (2) ***
*****
```

```
*** Berufsfachschulen, in denen ein Berufsabschluss erreicht werden kann
(q2400 == 4), werden nicht berücksichtigt ***
```

```
generate hilfe=1 if q2400 >= 1 & q2400 <= 3
```

```
bysort lfd: egen treffer=max(hilfe)
keep if treffer==1
drop treffer hilfe
```

```

*****
*** (3) ***
*****

*** Personen, die eine betriebliche Berufsausbildung abgeschlossen haben **

generate hilfe = 1 if q2200 == 1 | q2200 == 2
replace hilfe = . if q2211 != 1

generate gutabschluss_hilfe=1 if hilfe==1 & q2217>=1 & q2217<=3
replace gutabschluss_hilfe=0 if hilfe==1 & q2217==4 | q2217==5

label define lgutabschluss_hilfe 1 "Abschlussnote drei oder besser" 0
"Abschlussnote vier oder schlechter"
label variable gutabschluss_hilfe "gute Abschlussnote bei Prüfung?"
label values gutabschluss_hilfe lgutabschluss_hilfe

*****
*** (4) ***
*****

bysort lfd: egen gutabschluss = max(gutabschluss_hilfe)
label variable gutabschluss "gute Abschlussnote bei Prüfung?"
label values gutabschluss lgutabschluss_hilfe

*****
*** (5) ***
*****

generate monat_hilfe = endmo if gutabschluss_hilfe != .
generate jahr_hilfe = endja if gutabschluss_hilfe != .

bysort lfd: egen monat = max(monat_hilfe)
bysort lfd: egen jahr = max(jahr_hilfe)

label variable monat "Abschlussmonat der Berufsausbildung"
label variable jahr "Abschlussjahr der Berufsausbildung"

*****
*** (6) ***
*****

keep lfd gutabschluss monat jahr
sort lfd
by lfd: keep if _n==1

*****
*** (7) ***
*****

mvdecode monat jahr, mv(98, 99)
generate tag=15
generate abschlussdatum = mdy(monat, tag, jahr)
format abschlussdatum %d
quietly compress
sort lfd
save beispieleauswertung_ereignisdatensatz.dta, replace

```

```

*****
*** (8) ***
*****

use "<Laufwerk>\Uebergangsstudie2006_Querschnitt_SUF.dta", clear
sort lfd
merge lfd using beispieldauswertung_ereignisdatensatz.dta

*****
*** (9) ***
*****

tab _merge
keep if _merge == 3
drop _merge

*****
*** (10) ***
*****

generate referenzdatum = 17028
format referenzdatum %d
generate dauer = referenzdatum - abschlussdatum

*****
*** (11) ***
*****

table gutabschluss, contents (n dauer mean dauer) by(q200)

bysort q200: tab gutabschluss
bysort q200: tab gutabschluss, missing

log close

```

Impressum

BIBB-FDZ Daten- und Methodenberichte
Nr. 1/2010
Autorin: Daniela Rohrbach-Schmidt

Herausgeber:
Bundesinstitut für Berufsbildung
Forschungsdatenzentrum
Robert-Schuman-Platz 3
53175 Bonn

Redaktion: Holger Alda
Redaktionsassistentin: Tanja Stierner

Downloads unter:
www.bibb-fdz.de

Tel.: +49-228-107-2041
Fax: +49-228-107-2020
E-Mail: fdz@bibb.de

ISSN-Nr.: 2190-300X

Der Inhalt dieses Werkes steht unter einer [Creative Commons Lizenz](http://creativecommons.org/licenses/by-nc-nd/4.0/) (Lizenztyp: Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 4.0 Deutschland). Weitere Informationen finden Sie unter www.bibb.de/cc-lizenz.