

Forschungsdatenzentrum

der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt-
und Berufsforschung

FDZ

FDZ-Datenreport

Dokumentation zu Arbeitsmarktdaten

06/2012

DE

Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS)

Band I: Datenreport Welle 5

Marco Berg,
Ralph Cramer,
Christian Dickmann,
Reiner Gilberg,
Birgit Jesske,
Martin Kleudgen,
Arne Bethmann,
Benjamin Fuchs,
Mark Trappmann,
Anja Wurdack



Bundesagentur für Arbeit

Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS)

Datenreport Welle 5

Marco Berg, Ralph Cramer, Christian Dickmann, Reiner Gilberg, Birgit Jesske, Martin Kleudgen, infas Institut für angewandte Sozialwissenschaft GmbH - Arne Bethmann, Benjamin Fuchs, Mark Trappmann, Anja Wurdack, Institut für Arbeitsmarkt- und Berufsforschung (IAB)

Die FDZ-Datenreporte beschreiben die Daten des FDZ im Detail. Diese Reihe hat somit eine doppelte Funktion: Zum einen stellen Nutzerinnen und Nutzer fest, ob die angebotenen Daten für das Forschungsvorhaben geeignet sind, zum anderen dienen sie zur Vorbereitung der Auswertungen. Dieser Datenreport dokumentiert die Aufbereitung der Welle 5 von PASS. Das Dokument basiert auf dem Datenreport der Welle 4: Marco Berg, Ralph Cramer, Christian Dickmann, Reiner Gilberg, Birgit Jesske, Martin Kleudgen, (alle infas Institut für angewandte Sozialwissenschaft GmbH), Arne Bethmann, Benjamin Fuchs, Daniel Gebhardt (alle Institut für Arbeitsmarkt- und Berufsforschung (IAB): Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS) Band I: Datenreport Welle 4, FDZ Datenreport, 08/2011 (de), Nürnberg, aktualisierte Version vom 03.09.2012.

FDZ-Datenreporte (FDZ data reports) describe FDZ data in detail. As a result, this series of reports has a dual function: on the one hand, those using the reports can ascertain whether the data offered is suitable for their research task; on the other, the data can be used to prepare evaluations. This data report documents the data preparation of the fifth PASS wave and is based upon the fourth wave's data report: Marco Berg, Ralph Cramer, Christian Dickmann, Reiner Gilberg, Birgit Jesske, Martin Kleudgen, (all infas Institut für angewandte Sozialwissenschaft GmbH), Arne Bethmann, Benjamin Fuchs, Daniel Gebhardt (all Institut für Arbeitsmarkt- und Berufsforschung (IAB): Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS) Band I: Datenreport Welle 4, FDZ Datenreport, 08/2011 (de), Nuremberg, updated version 03.09.2012.

Inhaltsverzeichnis

1	Einleitung	8
1.1	Ziele und Fragestellungen des ‚Panel Arbeitsmarkt und soziale Sicherung‘	8
1.2	Instrumente und Befragungsprogramm	9
1.3	Besonderheiten und Neuerungen der Welle 5	10
2	Kennwerte	13
2.1	Stichprobenumfang	14
2.2	Response-Raten	20
2.3	Panelbereitschaft, Zuspielbereitschaft, Verknüpfung mit Prozessdaten	24
2.4	Abgespaltene Haushalte	25
3	Datensatzstruktur	27
4	Generierte Variablen	29
4.1	Vercodung offener Angaben	29
4.2	Harmonisierung	31
4.3	Dependent Interviewing	33
4.4	Einfache Generierungen	38
4.5	Konstruktvariablen	60
5	Datenaufbereitung	83
5.1	Strukturprüfungen und entfernte Interviews	85
5.2	Filterbereinigung	91
5.3	Plausibilitätsprüfungen	92
5.4	Rückwirkende Veränderungen der Wellen 1 bis 4	95
5.5	Anonymisierung	105
5.6	Arbeitslosengeld-II-Bezüge	109
5.7	Erwerbsbiografien	112
5.8	Ein-Euro-Job-Spell Datensatz (ee_spells)	116
6	Gewichtung Welle 5	118
6.1	Erweiterung der Stichprobe Welle 5	118
6.2	Integration der Aufstockungsstichproben	122
6.3	Designgewichte für die Bestands Haushalte der Welle 4	123
6.4	Designgewichte für die Zugangsstichprobe der Welle 5	124
6.5	Wiederteilnahme-Propensity Haushalte	124
6.6	Teilnahme-Propensity erstbefragte Split-Haushalte	128
6.7	Ausfallgewichtung für Haushalte aus der BA-Auffrischungs- sowie der BA-Bestandsaufstockungsstichprobe der Welle 5	130
6.8	Ausfallgewichtung für Haushalte aus der EWO-Aufstockungsstichprobe der Welle 5	133
6.9	Wiederteilnahme-Propensity Personen	134
6.10	Integration der Gewichte zum Gesamtgewicht vor Kalibrierung	138
6.11	Integration temporärer Ausfälle (Haushalte)	138
6.12	Kalibrierung zum Haushaltsgewicht Welle 5 Querschnitt	142
6.13	Kalibrierung zum Personengewicht Welle 5 Querschnitt	159
6.14	Schätzung der BA-Querschnittgewichte für Haushalte und Personen ohne Arbeitslosengeld-II-Bezug	178
7	Anhang: Kurzbeschreibung des Datensatzes	179

Tabellenverzeichnis

Tabelle 1:	Panelstichprobenumfang auf Haushaltsebene nach Wellen und Teilstichproben	16
Tabelle 2:	Panelstichprobenumfang auf Personenebene nach Wellen und Teilstichproben	17
Tabelle 3:	Panelstichprobenumfang fremdsprachige Interviews nach Wellen.....	18
Tabelle 4:	Response-Rate der Welle 5 auf Haushaltsebene nach Teilstichproben	21
Tabelle 5:	Durchschnittliche Realisierungsrate innerhalb der befragten Haushalte nach Wellen und Teilstichproben	22
Tabelle 6:	Anteil Personeninterviews in den Wellen 2 bis 5 mit panelbereiten Befragten der Vorwelle nach Teilstichproben.....	23
Tabelle 7:	Panelbereitschaft bei erstbefragten Haushalten*** nach Wellen	24
Tabelle 8:	Zustimmung zur Zuspiegelung von Prozessdaten in Personeninterviews (15 bis unter 65 Jahre), in denen in der jeweiligen Welle die Zuspieffrage gestellt wurde, nach Wellen.....	25
Tabelle 9:	Vercodung offener Angaben auf Haushaltsebene in Welle 5	29
Tabelle 10:	Vercodung offener Angaben auf Personenebene in Welle 5.....	30
Tabelle 11:	Harmonisierte Variablen im Personendatensatz (<i>PENDDAT</i>)	32
Tabelle 12:	Wellenübergreifend generierte, aber nicht vollständig harmonisierte Variablen im Personendatensatz (<i>PENDDAT</i>)	33
Tabelle 13a:	Fortgeschriebene Vorwelleninformationen in Welle 5, Haushaltsfragebogen.....	36
Tabelle 13b:	Fortgeschriebene Vorwelleninformationen in Welle 5, Personenfragebogen	37
Tabelle 14:	Typen der einfachen generierten Variablen in den Querschnittdatensätzen (<i>HHENDDAT</i> ; <i>PENDDAT</i>) bei Haushalten bzw. Personen, die schon in einer früheren Welle Angaben zum jeweiligen Thema gemacht haben.....	38
Tabelle 15:	Einfache Generierungen für Welle 5 im Haushaltsdatensatz (<i>HHENDDAT</i>) (alphabetische Reihenfolge)	40
Tabelle 16:	Einfache Generierungen für Welle 5 im Personendatensatz (<i>PENDDAT</i>) (alphabetische Reihenfolge)	42
Tabelle 17:	Einfache Generierungen für Welle 5 im Arbeitslosengeld-II-Spelldatensatz (<i>alg2_spells</i>) (Reihenfolge analog zur Reihenfolge im Datensatz)	50
Tabelle 18:	Einfache Generierungen für Welle 5 im BIO-Spelldatensatz (<i>bio_spells</i>) (Reihenfolge analog zur Reihenfolge im Datensatz)	53
Tabelle 19:	Einfache Generierungen für Welle 5 im Ein-Euro-Spelldatensatz (<i>ee_spells</i>) (Reihenfolge analog zur Reihenfolge im Datensatz).....	56
Tabelle 20:	Einfache Generierungen für Welle 5 im Personenregisterdatensatz (<i>p_register</i>) (alphabetische Reihenfolge)	57
Tabelle 21:	Übersicht über die Schritte der Datenaufbereitung in der Welle 5 des PASS.....	85
Tabelle 22:	Übersicht über die vergebenen Missing Codes	92
Tabelle 23:	Überarbeitung Einkommensvariablen	96

Tabelle 24:	Übersicht über rückwirkende Veränderungen im Haushaltsdatensatz (<i>HHENDDAT</i>)	101
Tabelle 25:	Übersicht über rückwirkende Veränderungen im Personendatensatz (<i>PENDDAT</i>)	102
Tabelle 26:	Übersicht über rückwirkende Korrekturen in Spelldatensätzen (<i>bio_spells, alg2_spells, ee_spells</i>).....	103
Tabelle 27:	Übersicht über rückwirkende Veränderungen in Registerdatensätzen (<i>hh_register; p_register</i>).....	104
Tabelle 28:	Übersicht über rückwirkende Veränderungen in den Gewichtungsdatensätzen (<i>hweights; pweights</i>).....	104
Tabelle 29:	Übersicht über die in Welle 5 anonymisierten Variablen im Personendatensatz (<i>PENDDAT</i>).....	106
Tabelle 30:	Übersicht über die in Welle 5 anonymisierten Variablen im BIO- Spelldatensatz (<i>bio_spells</i>)	109
Tabelle 31:	Querschnittvariablen im ALG-II-Spelldatensatz (<i>alg2_spells</i>).....	110
Tabelle 32:	ET-spezifische Querschnittvariablen im BIO-Spelldatensatz (<i>bio_spells</i>)	113
Tabelle 33:	AL-spezifische Querschnittvariablen im BIO-Spelldatensatz (<i>bio_spells</i>)	114
Tabelle 34:	Querschnittvariablen im EE-Spelldatensatz (<i>ee_spells</i>)	117
Tabelle 35:	Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der Wiederholerhaushalte.....	125
Tabelle 36:	Logit-Modelle zur Wiederteilnahme für Panelbereitschaft, Erreichbarkeit und Teilnahme	126
Tabelle 37:	Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der erstbefragten Split-Haushalte (Welle 4 und Welle 5)	128
Tabelle 38:	Logit-Modelle zur Erstteilnahme von Split-Welle 4-Haushalten für Erreichbarkeit und Teilnahme	129
Tabelle 39:	Logit-Modelle zur Erstteilnahme von Split-Welle 5-Haushalten für Erreichbarkeit und Teilnahme	129
Tabelle 40:	Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der BA-Auffrischungs- bzw. BA- Aufstockungsstichprobe Welle 5	130
Tabelle 41:	Logit-Modelle zur Erstteilnahme für Erreichbarkeit und Teilnahme der BA-Auffrischungs- bzw. BA-Aufstockungsstichprobe Welle 5.....	132
Tabelle 42:	Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der EWO-Aufstockungsstichprobe Welle 5	133
Tabelle 43:	Logit-Modelle zur Erstteilnahme für Erreichbarkeit und Teilnahme der EWO-Aufstockungsstichprobe Welle 5	134
Tabelle 44:	Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der Wiederholer	135
Tabelle 45:	Logit-Modelle zur Wiederteilnahme für Panelbereitschaft, Erreichbarkeit und Teilnahme	136
Tabelle 46:	Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der temporären Ausfälle	139
Tabelle 47:	Logit-Modelle temporärer Ausfälle	141

Tabelle 48:	Soll-Verteilungen und Verteilungen nach Kalibrierung (BA-Stichprobe, Haushalte)	145
Tabelle 49:	Kenngößen der Verteilung der Gewichte	147
Tabelle 50:	Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungsstichprobe, Haushalte)	148
Tabelle 51:	Kenngößen der Verteilung der Gewichte	153
Tabelle 52:	Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Haushalte)	154
Tabelle 53:	Kenngößen der Verteilung der Gewichte	159
Tabelle 54:	Soll-Verteilungen und Verteilungen nach Kalibrierung (BA-Stichprobe, Personen)	161
Tabelle 55:	Kenngößen der Verteilung der Gewichte	162
Tabelle 56:	Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungsstichprobe, Personen)	164
Tabelle 57:	Kenngößen der Verteilung der Gewichte	169
Tabelle 58:	Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen)	172
Tabelle 59:	Kenngößen der Verteilung der Gewichte	177

Abbildungsverzeichnis

Abbildung 1:	Realisierte Panelstichprobe Haushalte und Personen nach Erhebungswellen	19
Abbildung 2:	Datensatzstruktur des PASS in Welle 5	28
Abbildung 3:	Übersicht über generierte Variablen auf Personenebene in Welle 5	59

Datenverfügbarkeit

Der in diesem Beitrag beschriebene Datensatz ist für die Fachöffentlichkeit zugänglich.
Nähere Informationen dazu auf der Internetseite <http://fdz.iab.de/>

1 Einleitung

1.1 Ziele und Fragestellungen des ‚Panel Arbeitsmarkt und soziale Sicherung‘

Mit dem Panel ‚Arbeitsmarkt und soziale Sicherung‘ (PASS) baut das Institut für Arbeitsmarkt- und Berufsforschung (IAB) einen neuen Datensatz für die Arbeitsmarkt-, Sozialstaats- und Armutforschung in Deutschland auf, der eine neue empirische Grundlage für Wissenschaft und Politikberatung schafft.

Die Studie ist im Rahmen der SGB-II-Forschung¹ am IAB angesiedelt, deren gesetzlicher Auftrag es ist, die Wirkungen der Leistungen zur Eingliederung und der Leistungen zur Sicherung des Lebensunterhalts des SGB II zu untersuchen. Sie erlaubt jedoch aufgrund des komplexen Stichprobendesigns auch die Beantwortung von Fragen, die über diesen Rahmen weit hinausgehen. Bei der Entwicklung der neuen Studie waren vor allem fünf Kernfragestellungen leitend, die in Achatz et al. (2007) ausführlich erläutert werden:

1. Welche Wege führen aus dem Bezug von Arbeitslosengeld II?
2. Wie verändert sich die soziale Lage der betroffenen Haushalte im Leistungsbezug?
3. Wie entwickeln sich Bewältigungsprozesse und Handlungsorientierungen betroffener Personen?
4. Wie gestalten sich Kontakte zu den Trägern der Grundsicherung, und wie sieht die institutionelle Handlungspraxis aus?
5. Welche Erwerbsverlaufsmuster oder Haushaltsdynamiken führen in den Bezug von Arbeitslosengeld II?

Dieser Datenreport gibt einen Überblick über die fünfte Befragungswelle, für die zwischen Februar 2011 und September 2011 15.607 Personen in 10.235 Haushalten² befragt wurden. Darunter befanden sich 9.693 Personen und 6.547 Haushalte, mit denen bereits zum wiederholten Mal ein Interview im Rahmen von PASS geführt wurde³.

¹ Sozialgesetzbuch (SGB) Zweites Buch (II) – Grundsicherung für Arbeitsuchende.

² Die Zahlen umfassen nur auswertbare Interviews. Für wiederholt befragte Haushalte wurden auch solche berücksichtigt, für die nur ein Haushaltsinterview ohne Personen- bzw. Senioreninterview abgeschlossen werden konnte.

³ In der Welle 4 wurde die Bestandsstichprobe sowohl für Arbeitslosengeld II-Empfänger als auch der allgemeinen Bevölkerungstichprobe aus neuen Postleitzahlgebieten aufgestockt.

Der vorliegende wellenspezifische Datenreport⁴ der Welle 5 dokumentiert die wellenbezogenen Aspekte der Studie. Nach einem Kurzüberblick zu den Neuerungen und Besonderheiten der Welle 5 (Kapitel 1.3) werden die zentralen Kennzahlen zu Stichprobe und Ausschöpfungsquoten der Welle 5 berichtet (Kapitel 2). Weiterhin werden die Schritte der Datenaufbereitung und die dabei getroffenen Entscheidungen erläutert (Kapitel 5) sowie ein Überblick über die generierten Variablen gegeben (Kapitel 4). Zusätzlich wird das Gewichtungsverfahren vorgestellt (Kapitel 6). In den gesonderten Tabellenbänden werden die Häufigkeiten aller im Scientific Use File enthaltenen Variablen, unterteilt nach den jeweiligen Datensätzen, aufgelistet, die in Welle 5 erhoben wurden (Band II bis Band V).

1.2 Instrumente und Befragungsprogramm

Im PASS werden sowohl auf der Ebene der Haushalte als auch auf der Ebene der Personen Informationen mittels jeweils eigener Fragebögen erhoben. Zunächst wird mit jedem Haushalt ein Haushaltsinterview geführt, in dem Informationen, die den gesamten Haushalt betreffen, erhoben werden. Die Zielperson für dieses Haushaltsinterview⁵ wird bereits in der Kontaktphase ausgewählt, welche den eigentlichen Interviews vorausgeht. An das Haushaltsinterview schließen sich Personeninterviews mit den einzelnen Haushaltsmitgliedern an. Ziel ist es, mit allen im Haushalt lebenden Personen ab 15 Jahre ein Personeninterview zu führen – Haushaltsmitglieder ab 65 Jahre erhalten dabei eine Kurzfassung des Personenfragebogens (Seniorenbogen), in dem für diese Altersgruppe irrelevante Fragen nicht enthalten sind.

Erhebungsinstrumente und Befragungsprogramm der Welle 5 basieren auf der Welle 4 des PASS, wobei jedoch Einzelfragen und -module überarbeitet oder neu entwickelt wurden (siehe für einen Überblick Kapitel 1.3).

⁴ Mit der Dokumentation der Welle 3 wurde der Report erstmalig in zwei Teile aufgeteilt: ein wellenspezifischer Datenreport (inkl. Codebuch) und ein wellenübergreifender User Guide. Der wellenübergreifende User Guide wird in Verantwortung des PASS-Projektteams beim IAB erstellt. Die Dokumentation des wellenspezifischen Datenreports ab der Welle 3 wird von infas erstellt. Grundlage dafür bildete der Datenreport der Welle 2. Der wellenübergreifende User Guide soll der Dokumentation der Studie insgesamt dienen. Hier werden ausführlich die Ziele und das Design vom PASS beschrieben wie auch die Inhalte und Instrumente der Befragung vorgestellt. Des Weiteren werden die Struktur des Scientific Use Files und das Konzept der Variablentypen und -benennung erläutert. Abschließend wird anhand von Beispielen die Verwendung der verschiedenen Datensätze erläutert.

⁵ Die Zielperson für das Haushaltsinterview sollte sich möglichst gut mit allgemeinen Fragen zum Haushalt auskennen. Die Auswahl erfolgte nach bestimmten Regeln und ist ausführlich im Methodenbericht dokumentiert (Jesske & Quandt, 2011).

Die Erhebungsinstrumente vom PASS sind so angelegt, dass sie sowohl die wiederholte Befragung von Haushalten und Personen, die bereits an einer der Vorwellen teilgenommen haben, als auch die erstmalige Befragung ermöglichen⁶. Um bei der erneuten Befragung Seam-Effekte⁷ zu vermeiden und die Datenqualität zu erhöhen, wurden bereits seit der 3. Welle bei bestimmten Fragen per Dependent Interviewing Informationen, die der Befragte beim letzten Interview gegeben hatte, fortgeschrieben. Außerdem wurde weiterhin auf die erneute Erhebung unveränderlicher Merkmale weitgehend verzichtet. Auf Haushaltsebene existiert ab der Welle 4 abweichend zum Vorgehen der Wellen 1 bis 3 ein integrierter Fragebogen für wiederbefragte Haushalte (HHalt) und für erstmals befragte Haushalte (HHneu)⁸.

Die einzelnen Instrumente und das Befragungsprogramm werden ausführlich im wellenübergreifenden User Guide dargestellt. Der nachfolgende Abschnitt liefert einen Überblick zu den Besonderheiten und Neuerungen der Welle 5.

1.3 Besonderheiten und Neuerungen der Welle 5

Nutzern, die bereits mit den Daten der Panelwellen gearbeitet haben, möchten wir an dieser Stelle einen schnellen Überblick über Besonderheiten der PASS-Welle 5 geben.

Die Besonderheiten und Neuerungen in Welle 5 betreffen das Frageprogramm für den Haushalts- und Personenfragebogen (Änderung der Referenzzeiträume, Modifikation von Einzelfragen und neue Fragenmodule)⁹, die Stichprobe und die Aufbereitung der Daten.

1.3.1 Personenfragebogen

Im Personenfragebogen werden die seit Welle 2 erhobenen erwerbsbiografischen Informationen fortgeschrieben¹⁰. Dabei wurde in Welle 5 die in Welle 4 geänderte Abfragelogik der chronologischen Retrospektiverhebung beibehalten (s.a. Kap. 1.3.1 in Berg et al., FDZ Datenreport 08/2011).

⁶ Zu den erstmals befragten Haushalten gehören: (1) Haushalte der Zugangsstichprobe und Aufstockungsstichprobe der aktuellen Welle sowie (2) Haushalte, die sich von bereits in den Vorwellen befragten Haushalten abgespalten haben (Split-Haushalte) (weitere Erläuterungen vgl. Methodenbericht der Welle 4 (Jesske & Quandt, 2011)).

⁷ Die Anzahl der Veränderungen an der Schnittstelle (Seam) von einem Interview eines Panels auf das nächste ist häufig deutlich höher als die innerhalb eines Interviews beobachteten Veränderungen (vgl. Jäckle 2008).

⁸ Dabei werden Split-Haushalte wie neue Haushalte für die Abfrage behandelt.

⁹ Kleinere Änderungen im Frageprogramm (Neuaufnahme, Umformulierung bzw. Löschen von Einzelfragen) werden in diesem Rahmen nicht vollständig aufgeführt.

¹⁰ Dies geschieht u.a. mit Hilfe der so genannten „Dependent Interviewing“-Methode. Beim Dependent Interviewing werden Angaben, die wiederholt befragte Personen im Vorwelleninterview gemacht haben, in den Interviewtext des aktuellen Interviews eingespielt, um zu prüfen, ob diese Informationen aktualisiert werden müssen.

In Welle 5 wurde im Modul zur Erwerbsbiografie eine strukturelle Änderung vorgenommen. Das aktuelle Brutto- und Nettoeinkommen wurde nicht mehr summarisch über alle andauernden Erwerbstätigkeiten hinweg abgefragt, sondern bezogen auf die jeweilige Tätigkeit (ET2800-ET3900). Dies führt zur Generierung neuer Variablen, die in Kapitel 4 näher erläutert werden. Die bisherigen Variablen zum aktuellen Brutto- und Nettoeinkommen (PEK0100b-PEK1200) entfallen dadurch. Eine summarische Erhebung gibt es in Welle 5 nur noch für Sonderzahlungen aus dem letzten Jahr (PEK1360b) und für staatliche Zahlungen für Erwerbstätige (PEK2100).

Ebenfalls im Erwerbsmodul wieder aufgenommen wurden aus Welle 3 eine Nachfrage zum Zeitpunkt der Entfristung einer Tätigkeit (ET1753) sowie eine Variable, in der nachgefragt wird, aus welchen Quellen die Befragten mit neuer Tätigkeit von eben dieser zuerst erfahren haben (ET2400).

Weitere Ergänzungen im Personenfragebogen der Welle 5 betreffen:

- Ein Schwerpunktmodul zu Netzwerken, das bereits in Welle 3 zum Einsatz kam (zusätzlich zu den in jeder Welle gestellten Fragen (PSK0100-PSK0400) werden Fragen zu Netzwerkpersonen außerhalb des Haushalts (PSK0205-PSK0270) und zu sozialen Ressourcen (PSK0280a-j und PSK0285a-f) gestellt.
- Das Modul „Arbeitsuche“, bei dem für Befragte, die keine Tätigkeit suchen, die Itemliste der Gründe, warum dies der Fall ist, aus Welle 1 wieder aufgenommen wurde (PAS0850a-k).
- Das Modul „Einstellungen (Rollenbilder)“, bei dem aus Welle 2 Fragen zur Rollenverteilung der Geschlechter (PEO0400a-d) und zum Umgang mit Geld in Partnerschaften (PEO0415, PEO0420, PEO0430, PEO0440, PEO0450) erneut zum Einsatz kamen.
- 21 Items zur Abfrage der Persönlichkeitsmerkmale nach „Big Five“ (Big-Five-Inventar (BFI-K) nach Rammstedt und John (2005)) (PEO1400a-s).
- Das Modul „Einstellungen“, das um den Unterpunkt „Familie und Beruf“, für den ein Set von neuen Fragen entwickelt wurde (PEO0800a-b, PEO0900a-b, PEO1000a-b, PEO1100a-b), sowie im Unterpunkt „Arbeitszeit“ um die Fragen zur gewünschten eigenen Arbeitszeit (PEO1200) und die des Partners (PEO1300) erweitert wurde.
- Fragen zur Wohnortverbundenheit (PSK0070a-c und PSK0080).
- Die Frage zur Fortschreibung von Ein-Euro-Jobs (PEE0600).

Darüber hinaus wurde der Personenfragebogen im Face-to-Face-Feld um ein Modul zur „Stellenannahmefähigkeit“ erweitert. In diesem Modul wurde erhoben, unter welchen Bedingungen die Befragten bereit wären, eine ihnen angebotene neue Stelle anzunehmen. Die Abfrage erfolgte dabei im faktoriellen Survey-Design anhand von Vignetten¹¹.

¹¹ Vignetten enthalten Situationsbeschreibungen bzw. Fallbeispiele, die sich aus verschiedenen Merkmalen zusammensetzen und die dem Befragten statt einzelner Items vorgelegt werden. Zwi-

Es wurden fünf fiktive Stellenangebote (Vignetten) variiert, die sich im Hinblick auf das Einkommen, den Arbeitsumfang und die fachlichen Anforderungen sowie die innerbetrieblichen Aufstiegsmöglichkeiten, die Vertragsart (Befristung) oder die Entfernung zum aktuellen Wohnort unterscheiden. Die Bewertung durch die Befragten erfolgte in Hinblick auf die Attraktivität der Stellenangebote, die Wahrscheinlichkeit, mit der das Angebot angenommen werden würde, sowie die Bereitschaft, alleine oder gegebenenfalls gemeinsam mit Partner an den neuen Ort umzuziehen.

Neben den Ergänzungen und Erweiterungen gab es im Personenfragebogen folgende Kürzungen:

- Im Modul „Einstellungen Leben“ wurde die Frage zur generalisierten Selbstwirksamkeitserwartungen (PEO0100a-e) heraus rotiert und wird in Welle 6 erneut integriert.
- Im Modul „Migration“ die Fragen zur Sprache im Freundes- und Bekanntenkreis (PMI1110, PMI1120 und PMI1130).
- Im Modul „Religion“ werden die Fragen zur Religionszugehörigkeit (PD0200 und PD0300) und Religiösität (PD0400) nur an neubefragte Personen gerichtet.
- Im Modul „Freizeit“ werden die standardisierten Items PA0950a-r ersetzt durch eine offene Nachfrage zur Freizeitbeschäftigungen (PA1100 und PA1200) und Gründen für nicht wahrgenommene Freizeitbeschäftigungen (PA1300).

1.3.2 Haushaltsfragebogen

Im Haushaltsfragebogen wurden geringfügige Änderungen in Welle 5 vorgenommen.

Neu hinzugekommen ist eine standardisierte Itemliste von Gründen, warum das eigene Kind derzeit nicht (überwiegend) in einer Betreuungseinrichtung bzw. von einer Tagesmutter betreut wird (HKI0260*b-o).

Bei den Fragen zur Wohnsituation entfallen die Nachfragen zum Zustand der Wohnung (HW2000) und zum Einzugsjahr (HW0900).

Im Modul „Deprivation“ wurden ausgewählte Items gestrichen: Vorhandensein einer Heizung (HLS0500a und HLS0500b), Vorhandensein eines Gefrierschranks (HLS1300a und HLS1300b) und Verwendung von rezeptfreien Medikamenten (HLS2400a und HLS2400b).

schen den Fallbeispielen werden die in ihrem Einfluss besonders interessierenden Merkmale in ihren Ausprägungen variiert.

1.3.3 Stichprobe und Datenaufbereitung

Wie in den Vorwellen, wurde auch in Welle 5 eine sog. Zugangsstichprobe für die BA-Teilstichprobe¹² gezogen. Ziel ist es, die Repräsentativität der BA-Stichprobe im Querschnitt sicherzustellen und im Zeitverlauf ausreichend neue Übergänge in den Arbeitslosengeld II-Bezug beobachten zu können. Für die Auffrischung werden Bedarfsgemeinschaften gezogen, die im Juli 2010 Arbeitslosengeld II bezogen haben, nicht jedoch zum Zeitpunkt der Stichprobenziehung der 1., 2., 3. oder Welle 4 (vgl. Kapitel 2.1 und zum Auffrischungskonzept Trappmann et al 2009: 11 ff.). Zusätzlich fand in Welle 5 eine Bestandsauffrischung der bestehenden Stichproben statt, in dem 100 neue PLZ-Bereiche ausgewählt wurden. Die Bestandsauffrischung umfasst sowohl die BA-Stichprobe als auch die Bevölkerungstichprobe. Jedoch wurde die Bevölkerungstichprobe anders als in Welle 1 aus den Einwohnermeldeeregistern gezogen. Eine genaue Beschreibung zum Vorgehen findet sich im Kapitel 6.3. Alle in der Welle 5 neubefragten Haushalte können wiederum über den Stichprobenindikator (*sample*) identifiziert werden.

Die Datenaufbereitung erfolgte wiederum in enger Abstimmung mit dem IAB. Grundsätzliche Vorgehensweisen z. B. zur Fortschreibung der Datensätze und zur Bereinigung von Problemen in den Haushaltsstrukturen wurden im Aufbereitungsprozess diskutiert und durch das IAB entschieden.

Das Konzept für die Integration der Spelldatensätze im Erwerbsmodul sowie die notwendigen Aufbereitungsschritte wurden gemeinsam mit dem IAB diskutiert und verabschiedet. Das Vorgehen wird im Kapitel 5.7 dokumentiert.

2 Kennwerte

Im folgenden Kapitel wird ein kurzer Überblick über zentrale Kennziffern der Studie, wie Brutto- und Nettostichprobengröße oder Ausschöpfungsquoten, gegeben. Diese werden für die Panelstichprobe über den Verlauf der bisherigen drei Wellen dargestellt und sowohl für die beiden ursprünglichen Teilstichproben und die Zugangsstichproben getrennt als auch für die Gesamtstudie ausgewiesen.

- Teilstichprobe 1 (BA-Stichprobe) bezeichnet dabei im Folgenden die Leistungsempfängerstichprobe aus den Prozessdaten der Bundesagentur für Arbeit.
- Teilstichprobe 2 (MICROM-Stichprobe) verweist auf die geschichtete Bevölkerungstichprobe.
- Zugangsstichprobe 1 (BA-Stichprobe) benennt die Stichprobe aus den SGB II-Zugängen zwischen Welle 1 und Welle 2.
- Zugangsstichprobe 2 (BA-Stichprobe) benennt die Stichprobe aus den SGB II-Zugängen zwischen Welle 2 und Welle 3.

¹² Die Welle 1 des PASS besteht aus zwei Teilstichproben: (1) eine aus den Prozessdaten der BA gezogene Stichprobe von Arbeitslosengeld II-Empfängerhaushalten und (2) eine aus einer Datenbank des kommerziellen Anbieters MICROM gezogene, statusgeschichtete allg. Bevölkerungstichprobe.

- Zugangsstichprobe 3 (BA-Stichprobe) benennt die Stichprobe aus den SGB II-Zugängen zwischen Welle 3 und Welle 4.
- Zugangsstichprobe 4 (BA-Stichprobe) benennt die Stichprobe aus den SGB II-Zugängen zwischen Welle 4 und Welle 5.
- Bestandsauffrischung/Aufstockung 1 (EWO-Stichprobe) benennt die Stichprobe aus den Einwohnermeldeamtszugängen in zehn neuen PLZ-Bereichen Welle 5
- Bestandsauffrischung/Aufstockung 2 (BA-Stichprobe) benennt die Stichprobe aus den SGB II-Zugängen in 100 neuen PLZ-Bereichen Welle 5.

2.1 Stichprobenumfang

Ausgangsbasis für den Stichprobenumfang in einem Panel bilden die befragten Haushalte der Erstbefragung. Im PASS beinhaltet die Bruttopanelstichprobe die befragten Haushalte aus der Welle 1, aber auch die erstbefragten Haushalte aus den Zugangsstichproben in den Wellen 2, 3, 4 und 5. Dabei muss berücksichtigt werden, dass nur solche erstbefragten Haushalte auch bei den Wiederholungsbefragungen zur Verfügung stehen, für welche die Panelbereitschaft¹³ vorliegt. Die Panelbereitschaft wird nur bei der Erstbefragung erhoben. In den Folgewellen ist eine erneute Bereitschaftserklärung für diese Haushalte nicht notwendig. Der Zugang zum Panel wird neben der Panelbereitschaftserklärung bereits in der ersten Befragung durch die grundsätzliche Teilnahmebereitschaft, d.h. durch die Realisierung eines Interviews bedingt. Maßnahmen, die im Rahmen vom PASS einen möglichst selektionsfreien Zugang zum Panel sicherstellen sollten, werden ausführlich im Methoden- und Feldbericht der Wellen 1 bis 5 beschrieben¹⁴.

PASS ist in der Welle 1 mit 12.794 realisierten Haushaltsinterviews gestartet, von denen 12.000 Haushalte ihre Panelbereitschaft erklärt haben. Die panelbereiten Haushalte der Welle 1 bilden den Stichprobenumfang für den Start der ersten Wiederholungsbefragung.

Das Panelkonzept im PASS sieht vor, dass durch Auszug von Personen aus Panelhaushalten neue Haushalte bzw. abgespaltene Haushalte (Splithaushalte) entstehen, die als eigene Haushalte gezählt werden, sobald dort ein Haushaltsinterview realisiert werden konnte. Damit nimmt die Anzahl der Haushalte gegenüber der Ausgangsstichprobe zu. Das Vorgehen im Panelkonzept vom PASS wird im Abschnitt „Abspaltene Haushalte“ näher beschrieben. Neben der Erweiterung des Panels kann es aufgrund von Panelmortalität auch zum Verlust von Haushalten kommen. Haushalte, in denen alle Befragungspersonen verstorben oder ins Ausland verzogen sind, werden in den Folgewellen nicht mehr zum Panelbrutto gezählt. Panelverluste entstehen ebenfalls dadurch, dass für einen Haushalt über zwei Wellen infolge kein Haushaltsinterview realisiert werden konnte. Dieser Fall konnte

¹³ Die Panelbereitschaft wird durch den Haushaltsbevollmächtigten erteilt und zählt damit für alle Haushaltsmitglieder. Panelbereite Haushalte haben der Speicherung ihrer Adresse zum Zweck von Wiederholungsbefragungen im Rahmen der Studie zugestimmt.

¹⁴ Vgl. Hartmann et al. (2008); Büngeler et al. (2009); Büngeler et al. (2010), Jesske & Quandt (2011), Jesske & Schulz (im Erscheinen 2012).

erstmalig am Ende der Welle 3 eintreten und wirkt sich auf das Panelbrutto in der Wellen 4¹⁵ und 5 aus. Die Bruttoeinsatzstichprobe für die Welle 5 vom PASS umfasste insgesamt 9.155 Panelhaushalte. Dazu kommen in jeder Welle erstbefragte Haushalte aus der Zugangsstichprobe und den Splithaushalten sowie in Welle 5 aus den zusätzlichen Aufstockungsstichproben.

Die Fallzahlen für den Bruttostichprobenumfang der jeweiligen Erhebungswellen und Teilstichproben wird in der nachfolgenden Tabelle ausgewiesen. In der Welle 5 konnte in 6.547 Haushalten der Panelstichprobe mindestens ein Interview geführt werden. Dazu kommen 753 erstbefragte Haushalte aus der Zugangsstichprobe, von denen 702 ihre Panelbereitschaft erklärt haben, und 2.831 aus den Aufstockungsstichproben, mit 2.672 Panelbereiten. Zu den erstbefragten Haushalten der Welle 5 gehören ebenso 104 Splithaushalte, die sich aus den sechs Teilstichproben der bisherigen Wellen 1 bis 4 ergeben.

¹⁵ Zusätzlich auf das Panelbrutto in der Welle 4 wirkte sich der Wechsel des Erhebungsinstituts aus. Damit die Adressen der Panelteilnehmer vom IAB an infas übergeben werden konnten, war die Erlaubnis der Zielperson zur Weitergabe notwendig. Ausführlichere Erläuterungen zu diesem Verfahren und den Ergebnissen finden sich im Methodenbericht der Welle 4 (Jesske & Quandt, 2011).

Tabelle 1: Panelstichprobenumfang auf Haushaltsebene nach Wellen und Teilstichproben¹

		Stichprobe								
	n	BA	Microm	BA-Zugang 1	BA-Zugang 2	BA-Zugang 3	BA-Zugang 4	Aufstockung EWO	Aufstockung BA	Gesamt
Welle 1	HH-Interview realisiert	6.804	5.990							12.794
	davon: HH panelbereit	6.452	5.548							12.000
Welle 2	Panel-HH brutto	6.520	5.611							12.131
	HH-Interview realisiert	3.491	3.897	1.041						8.429
	davon: HH panelbereit	3.360	3.766	1.003						8.129
Welle 3	Panel-HH brutto	5.851	5.150	1.010						12.011
	HH-Interview realisiert	3.754	3.901	694	1.186					9.535
	davon: HH panelbereit	3.576	3.777	669	1.145					9.167
Welle 4*	Panel-HH brutto	3.926	3.628	863	1.069					9.486
	HH-Interview realisiert	2.815	2.977	563	745	748				7.848
	davon: HH panelbereit	2.754	2.933	554	727	723				7.691
Welle 5**	Panel-HH brutto	3.376	3.319	674	956	726				9.051
	HH-Interview realisiert	2.382	2.680	464	608	517	753	1.510	1.321	10.235
	davon: HH panelbereit	2.347	2.633	456	598	512	702	1.415	1.257	9.920

Quelle: HH-Register und PENDDAT; Scientific Use File IAB

* Reduktion der Bruttostichprobe durch Widerrufsverfahren

** Erhöhung der Bruttostichprobe durch Aufstockung

¹ Die Registerfiles des Scientific Use Files umfassen immer die Nettostichprobe realisierter Interviews der jeweiligen Wellen. Im Falle von Split-Haushalten ist es möglich, dass es nachträglich zur Erweiterung des Panel-HH brutto der Vorwelle kommt, wenn der Split-Haushalt in der Vorwelle zwar identifiziert wurde, aber noch nicht realisiert werden konnte.

Den 10.235 realisierten Haushaltsinterviews in der Welle 5 entsprechen 15.607 realisierte Personeninterviews. Die Verteilung der befragten Personen über die Teilstichproben und die jeweiligen Erhebungswellen ist der nachfolgenden Tabelle zu entnehmen.

Tabelle 2: Panelstichprobenumfang auf Personenebene nach Wellen und Teilstichproben

	Personeninterview realisiert	Welle 1	Welle 2	Welle 3	Welle 4*	Welle 5**
		abs.	abs.	abs.	abs.	abs.
BA		9.386	4.753	4.913	3.958	3.394
Microm		9.568	6.392	6.207	5.016	4.511
Stichprobe	BA-Zugang 1		1.342	898	786	653
	BA-Zugang 2			1.421	983	822
	BA-Zugang 3				1.025	760
	BA-Zugang 4					1.019
	Aufstockung EWO					2.589
	Aufstockung BA					1.859
	Gesamt		18.954	12.487	13.439	11.768

Quelle: P_Register; Scientific Use File IAB

* Reduktion der Bruttostichprobe durch Widerrufsverfahren

** Erhöhung der Bruttostichprobe durch Aufstockung

Für Befragte ohne ausreichende Deutschkenntnisse wurden zwei weitere Interviewsprachen: Türkisch und Russisch angeboten. Tabelle 3 gibt an, wie viele Haushalte bzw. Personen in den beiden Erhebungssprachen befragt wurden.

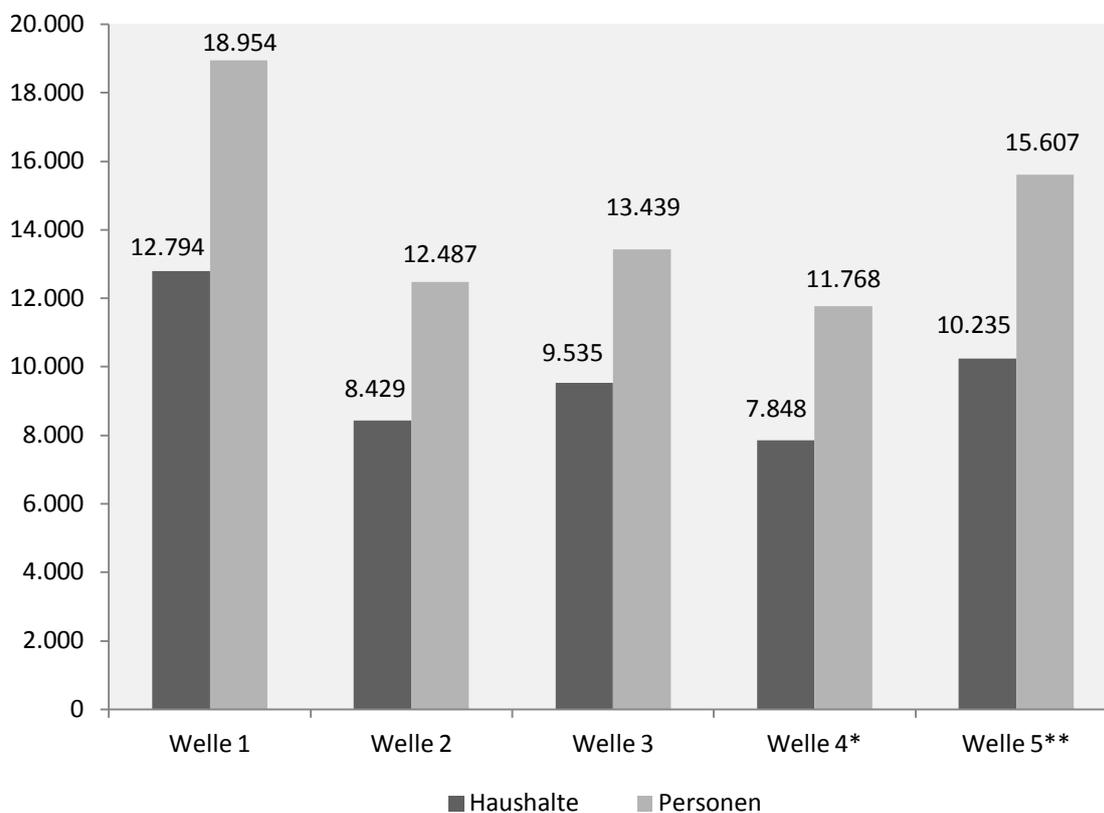
Tabelle 3: Panelstichprobenumfang fremdsprachige Interviews nach Wellen

		Russisch	Türkisch
		abs.	abs.
Welle 1	Haushalte	275	163
	Personen	432	305
Welle 2	Haushalte	156	39
	Personen	219	31
Welle 3	Haushalte	210	69
	Personen	330	109
Welle 4	Haushalte	179	42
	Personen	285	78
Welle 5	Haushalte	159	36
	Personen	259	58

Quelle: PENDDAT; Scientific Use File IAB

Damit ergibt sich für den Gesamtbestand der realisierten Panelstichprobe folgendes Bild im Überblick für Haushalte und Personen über die fünf Erhebungswellen.

Abbildung 1: Realisierte Panelstichprobe Haushalte und Personen nach Erhebungswellen



* Reduktion der Bruttostichprobe durch Widerrufsverfahren

** Erweiterung der Bruttostichprobe durch Aufstockung

2.2 Response-Raten

Die Berechnung der Response-Rate erfolgt nach AAPOR-Standards (AAPOR 2006). Berichtet wird die Ausschöpfungsquote RR1, in der auch alle Fälle ungeklärter Eligibilität in den Nenner eingehen und die somit den geringsten Wert aller Ausschöpfungsquoten annimmt¹. Als Ausschöpfungsquote auf der Haushaltsebene wird der Anteil der auswertbaren Haushaltsinterviews an der Summe aus auswertbaren Haushaltsinterviews und nicht neutralen Ausfälle berechnet. Als neutrale Ausfälle gelten nur Haushalte, in denen alle Personen verstorben und Haushalte, die dauerhaft ins Ausland verzogen sind. Haushalte gelten als auswertbar, wenn zumindest ein vollständiges Haushaltsinterview vorliegt. Neuhaushalte gelten nur dann als auswertbar, wenn neben dem Haushaltsinterview auch mindestens ein vollständiges Personeninterview vorliegt.

¹ Dies wird in Deutschland sehr uneinheitlich gehandhabt. Häufig werden sehr viele nicht befragte Personen oder Haushalte als „ineligible“ verbucht und bei der Berechnung der Ausschöpfung aus dem Nenner gestrichen. Im Falle einer Ziehung aus Registern darf aber weder ein Haushalt, der nicht unter der erwarteten Adresse wohnt, noch ein Haushalt, der behauptet, nicht zur Zielgruppe zu gehören, als neutraler Ausfall verbucht werden. Zudem beschränkt sich die Population vom PASS nicht auf deutschsprachige Befragte und auf befragungsfähige Personen, so dass auch die Ausfallgründe „spricht kein Deutsch“ oder „Person ist krank/nicht befragungsfähig“ nicht als neutrale Ausfälle gelten können.

Auf der Haushaltsebene ergaben sich für die Welle 4 folgende Ausschöpfungsquoten:

Tabelle 4: Response-Rate der Welle 5 auf Haushaltsebene nach Teilstichproben
Stichprobe

Welle 5		BA	Microm	BA-Zugang 1	BA-Zugang 2	BA-Zugang 3	BA-Zugang 4	Aufstockung EWO	Aufstockung BA	Gesamt
HH brutto	abs.	3.376	3.319	674	956	726	2.792	6.237	5.428	23.508
	%	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
neutrale Ausfälle	abs.	27	21	3	2	3	18	80	27	181
	%	0,8	0,6	0,4	0,2	0,4	0,6	1,3	0,5	0,8
HH brutto bereinigt*		3.349	3.298	671	954	723	2.774	6.157	5.401	23.327
	%	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
HH-Interview realisiert	abs.	2.382	2.680	464	608	517	753	1.510	1.321	10.235
	%	71,1	81,3	69,2	63,7	71,5	27,1	24,5	24,5	43,9
davon: HH panelbereit	abs.						702	1.415	1.257	
	%						25,3	23,0	23,3	

* HH brutto - neutrale Ausfälle

Quelle: HH-Register; Scientific Use File IAB - für BA-Zugang 4 und Aufstockungsstichproben: Methodendatensatz infas

In einer Haushaltserhebung kann zwischen der Response-Rate auf Haushaltsebene und der Response-Rate innerhalb von Haushalten unterschieden werden.

Mit Response-Rate innerhalb von Haushalten bezeichnen wir den durchschnittlichen Anteil aller Personen ab 15 Jahre innerhalb der Haushalte mit einem auswertbaren Haushaltsinterview, für die ein vollständiges Personeninterview vorliegt.

Im Durchschnitt ergeben sich innerhalb der befragten Haushalte folgende Realisierungsraten:

Tabelle 5: Durchschnittliche Realisierungsrate innerhalb der befragten Haushalte nach Wellen und Teilstichproben

	Welle 1	Welle 2	Welle 3	Welle 4	Welle 5
	%	%	%	%	%
BA	85,6	85,5	83,1	88,4	88,7
Microm	84,2	85,1	83,6	88,0	88,3
BA-Zugang 1		86,2	84,3	90,2	89,5
BA-Zugang 2			84,2	88,3	89,3
BA-Zugang 3				89,6	91,2
BA-Zugang 4					88,9
Aufstockung EWO					84,4
Aufstockung BA					90,0
Gesamt	84,9	85,4	83,5	88,5	88,3

Quelle: P_Register; Scientific Use File IAB

Neben den Response-Raten auf der Haushaltsebene und innerhalb der Haushalte sind im Folgenden die Wiederbefragungsquoten auf der Personenebene angegeben. Dabei handelt es sich um den Anteil von panelbereiten Personen, mit denen in der Folgewelle ein Interview realisiert werden konnte.

Tabelle 6: Anteil Personeninterviews in den Wellen 2 bis 5 mit panelbereiten Befragten der Vorwelle nach Teilstichproben
Stichprobe

			BA	Microm	BA-Zugang 1	BA-Zugang 2	BA-Zugang 3	Gesamt
Welle 2	panelbereite Befragungspersonen W1	abs.	8.925	8.938				17.863
	in W2 wiederholt befragte Personen	abs.	4.274	5.829				10.103
	<i>Anteil</i>	<i>%</i>	<i>47,9</i>	<i>65,2</i>				<i>56,6</i>
Welle 3	panelbereite Befragungspersonen W2	abs.	4.686	6.292	1.298			12.276
	in W3 wiederholt befragte Personen	abs.	3.365	4.956	820			9.141
	<i>Anteil</i>	<i>%</i>	<i>71,8</i>	<i>78,8</i>	<i>63,2</i>			<i>74,5</i>
Welle 4*	panelbereite Befragungspersonen W3	abs.	4.844	6.100	894	1.380		13.218
	in W4 wiederholt befragte Personen	abs.	3.287	4.347	626	854		9.114
	<i>Anteil</i>	<i>%</i>	<i>67,9</i>	<i>71,3</i>	<i>70,0</i>	<i>61,9</i>		<i>69,0</i>
Welle 5	panelbereite Befragungspersonen W4	abs.	3.946	5.004	785	979	993	11.707
	in W5 wiederholt befragte Personen	abs.	2.972	4.151	570	714	702	9.109
	<i>Anteil</i>	<i>%</i>	<i>75,3</i>	<i>83,0</i>	<i>72,6</i>	<i>72,9</i>	<i>70,7</i>	<i>77,8</i>

Quelle: PENDDAT; Scientific Use File IAB

* Reduktion der Bruttostichprobe durch Widerrufsverfahren zwischen Welle 3 und 4

2.3 Panelbereitschaft, Zuspielbereitschaft, Verknüpfung mit Prozessdaten

Zur Speicherung der Adressen zum Zweck der Wiederbefragung in der Folgewelle und zum Zusammenspielen der Erhebungsdaten mit Prozessdaten der Bundesagentur für Arbeit ist jeweils die Einverständniserklärung der Befragten notwendig.

Die Panelbereitschaft wurde schon im Kapitel 2.1 beim Stichprobenumfang näher erläutert. Die Panelbereitschaft für in einer Welle erstmals befragte Haushalte¹⁸ im PASS sieht wie folgt aus:

Tabelle 7: Panelbereitschaft bei erstbefragten Haushalten* nach Wellen**

	Realisierte HH-Interviews mit erstmalig befragten HH	Realisierte HH-Interviews mit erstmals befragten HH mit Panelbereitschaft	Anteil mit Panelbereitschaft
	abs.	abs.	%
Welle 1	12.794	12.000	93,8
Welle 2	1.086	1.048	96,5
Welle 3	1.327	1.285	96,8
Welle 4*	903	866	95,9
Welle 5**	3.688	3.476	94,3

* Reduktion der Bruttostichprobe durch Widerrufungsverfahren

** Erhöhung der Bruttostichprobe durch Aufstockung

***erstmals befragte HH aus Auffrischung, Aufstockung und Split

Quelle: PENDDAT und HH_Register; Scientific Use File IAB

Die Panelbereitschaft wird bei erstmals befragten Haushalten in jeder Welle im Anschluss an das erste Personeninterview erfasst. Die Angabe dieser Person wird dann auf den Haushalt übertragen. Ist die Person panelbereit, wird der Haushalt als panelbereit gewertet. Ist die Person nicht panelbereit, wird der Haushalt als nicht panelbereit gewertet (s.dazu auch Kap. 2.1)¹⁹.

¹⁸ Bei allen Haushalten der Welle 1 handelt es sich um erstmals befragte Haushalte. Ab Welle 2 handelt es sich nur bei Haushalten der jeweiligen Auffrischungstichproben um erstmals befragte Haushalte sowie bei den erstmals teilnehmenden Split-Haushalten. Ab Welle 2 sind erstmals befragte Haushalte daher in der Minderheit – bei einem Großteil der in diesen Wellen realisierten Haushaltsinterviews handelt es sich um Interviews mit Haushalten, die bereits zu einem früheren Zeitpunkt befragt wurden

¹⁹ Die Angabe zur Panelbereitschaft wird also von einer Person für den gesamten Haushalt gemacht. Die auf Haushaltsebene vorliegende Information wurde im Zuge der Datenaufbereitung in den Personendatensatz (PENDDAT) integriert. Die einzelnen, in einem Haushalt befragten Personen übernahmen den entsprechenden Wert, der für den Haushalt vorlag. Analog wurde auch in Welle 2 vorgegangen. In Welle 1 wurde die Panelbereitschaft dagegen nach jedem Personen- und Senioreninterview personenspezifisch erhoben – hier können also unterschiedliche Werte innerhalb eines Haushalts vorkommen. Ein Haushalt, in dem mindestens eine panelbereite Person lebte, wurde dabei als panelbereit gewertet.

Im Gegensatz zur Panelbereitschaft wurde die Erlaubnis zur Zuspiegelung von Prozessdaten der Bundesagentur für Arbeit an die Befragungsdaten für jeden Befragten erhoben, der mit dem Personenfragebogen interviewt wurde. Für Personen ab 65 Jahre entfällt die Nachfrage, da sie im Seniorenfragebogen nicht enthalten ist. Die Zuspiegelbereitschaft wird dabei nicht in jeder Welle neu erhoben²⁰.

Tabelle 8 gibt einen Überblick über die Zuspiegelbereitschaft in den einzelnen Wellen. Aufgeführt sind nur diejenigen Interviews, in denen die Frage zur Zuspiegelbereitschaft in der betreffenden Welle im Rahmen des Personenfragebogens gestellt wurde.

Tabelle 8: Zustimmung zur Zuspiegelung von Prozessdaten in Personeninterviews (15 bis unter 65 Jahre), in denen in der jeweiligen Welle die Zuspiefrage gestellt wurde, nach Wellen

	Realisierte Personen- Interviews der Welle, in der die Zuspiefrage gestellt wurde	Realisierte Personen- Interviews der Welle, in der die Zuspiegelbereitschaft erteilt wurde	Anteil mit Zuspiegelbereitschaft
	abs.	abs.	%
Welle 1	17.249	13.766	79,8
Welle 2	3.358	2.560	76,2
Welle 3	2.656	2.128	80,1
Welle 4*	2.032	1.774	87,3
Welle 5**	5.145	4.414	85,8

* Reduktion der Bruttostichprobe durch Widerrufsverfahren

** Erhöhung der Bruttostichprobe durch Aufstockung

Basis: Personen 15 bis 64 Jahre

Quelle: PENDDAT; Scientific Use File IAB

2.4 Abgespaltene Haushalte

PASS ist als dynamisches Panel angelegt. Personen, die in Befragungshaushalte einziehen oder dort geboren werden, werden ebenfalls befragt, sofern sie 15 Jahre oder älter sind. Auch Personen, die aus den Befragungshaushalten ausziehen bzw. ein Jahr oder länger nicht im Haushalt leben, sollen weiterhin befragt werden. Ihre neuen Haushalte werden als Abspaltungen des ursprünglichen Befragungshaushalts verstanden. Diese abgespaltenen Haushaltsteile (oder Splithaushalte) werden zu Befragungshaushalten des

Bei wiederbefragten Haushalten wurde im Rahmen der Adressaktualisierung nach dem ersten Personeninterview angekündigt, dass sich im kommenden Jahr wieder ein Interviewer melden wird. Widerspruch der Befragte dieser Ankündigung nicht explizit, wurde der Haushalt als weiterhin panelbereit gewertet und die Variable panel im Personendatensatz (PENDDAT) entsprechend fortgeschrieben.

²⁰ Aufgrund von Umstellungen bei der Filterung konnte es vorkommen, dass in Welle 2 und 3 die Frage nach der Zuspiegelbereitschaft erneut gestellt wurde, sofern die befragte Person ihre Zuspiegelbereitschaft in einer der Vorwellen noch nicht erklärt hatte.

PASS. Alle in ihnen lebenden Personen ab 15 Jahre werden zu Zielpersonen für Personeninterviews. Sollte sich in einer Folgewelle wiederum ein Haushaltsteil aus dem Splithaushalt abspalten, wird auch dieser zu einem Befragungshaushalt im PASS, unabhängig davon, ob darin noch Personen aus einer der ursprünglichen Stichproben leben („infinite degree contagion model“, Rendtel und Harms 2009, 267). Ins Ausland verzogene Personen werden dagegen nicht weiter befragt, da sie nicht mehr zur Population gehören und auch inhaltlich die SGB II-spezifischen Fragestellungen nicht mehr zutreffen. Personen, die für weniger als ein Jahr nicht im Haushalt leben, werden weiterhin als Haushaltsmitglieder gezählt und bilden ebenfalls keinen neuen PASS-Haushalt.

Aus den Befragungen der Welle 1 bis 5 liegen insgesamt 447 abgespaltene Haushalte vor, von denen in Welle 5 283 befragt werden konnten. Darunter befanden sich 83 neue Splithaushalte aus der Welle 5 sowie 21 erstbefragte Splithaushalte, die bereits in Welle 4 identifiziert werden konnten. Weitere Informationen zu den Splithaushalten finden sich im Methodenbericht der Erhebungswelle 5 (Jesske & Schulz, 2012).

Die befragten Splithaushalte können in den Datensätzen durch den Vergleich der aktuellen (*hnr*) und der ursprünglichen Haushaltsnummer (*uhnr*) identifiziert werden, die sich in diesen Fällen unterscheidet. Die Ursprungshaushaltsnummer (*uhnr*) enthält dabei die Haushaltsnummer des Panelhaushalts, von dem sich der neue Haushalt abgespalten hat. Splithaushalte übernehmen das Stichprobenkennzeichen (*sample*), die Informationen zum Jahr der Stichprobenziehung (*jahrsamp*), die Primary Sampling Unit (*psu*) und deren Stratifizierung (*strpsu*) von ihrem Ursprungshaushalt.

3 Datensatzstruktur

Die übliche Struktur für die Aufbereitung eines Paneldatensatzes, wie sie z. B. in Erhebungen wie dem Sozio-oekonomischen Panel (SOEP) oder dem British Household Panel Survey (BHPS) praktiziert wird, ist eine Ablage der Informationen zu Personen und Haushalten in jährlichen Einzeldatensätzen. Diese können ggf. um spezifische, auch wellenübergreifende Datensätze ergänzt werden, wie z. B. für Register- oder Spelldaten.

Diese Datenstruktur erlaubt es, die Informationen platzsparend zu speichern. Welche Variablen in welchen Jahren erhoben wurden, ist nach dem Blick in die Datensätze unmittelbar zu erkennen. Auch die Zuschlüsselung von Informationen – über Schlüsselvariablen wie Haushalts- oder Personennummern – ist vergleichsweise einfach. Allerdings ist diese für Paneldaten übliche Struktur auch mit Nachteilen verbunden, welche die Arbeit mit diesen Datensätzen deutlich erschweren. Sollen Analysen nicht nur im Quer- sondern auch im Längsschnitt durchgeführt werden, müssen zunächst alle relevanten Variablen aus den Einzeldatensätzen der jeweiligen Wellen in einen gemeinsamen Datensatz integriert werden, wobei darauf geachtet werden muss, dass es sich inhaltlich tatsächlich um dieselben Konstrukte handelt. Für typische Längsschnittanalysen muss der erstellte, wellenübergreifende Datensatz dann in das Long-Format umstrukturiert werden. Im Gegensatz zum Wide-Format, in dem in der Datenmatrix für jede Beobachtungseinheit (z. B. ein Haushalt oder eine Person) genau eine Zeile vorhanden ist und dann für jede Erhebungswelle mehrere Datensätze existieren, werden im Long-Format alle einer Beobachtungseinheit zugeordneten Erhebungswellen untereinander angeordnet. Statt die Informationen in wellenspezifischen Variablen in derselben Zeile anzuordnen, werden im Long-Format die Informationen in wellenspezifischen Zeilen der Beobachtungseinheiten der jeweils gleichen Variable zugeordnet.

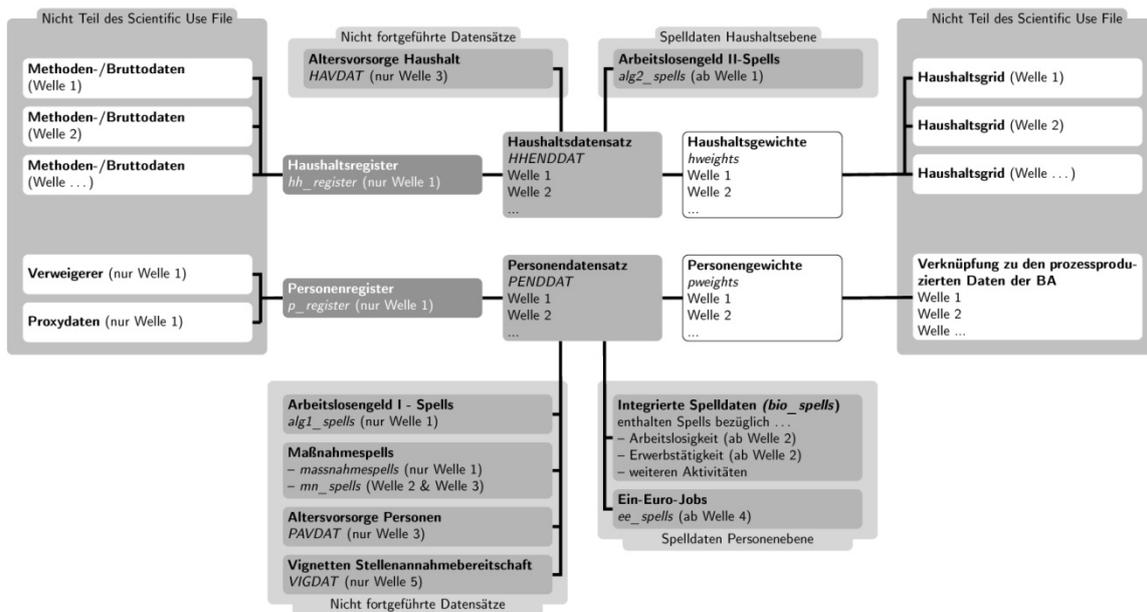
Auch die Aufbereitung im Long-Format ist mit Vor- und Nachteilen verbunden. Der entscheidende Vorteil dieser Variante ist, dass die Daten bereits in der für viele Längsschnittanalysen (wie z. B. Verweildaueranalysen) benötigten Struktur vorliegen. Der zusätzliche Aufwand für die Erstellung eines wellenübergreifenden Files entfällt. Der Umstieg vom Long- ins Wide-Format ist zudem vergleichsweise einfach zu realisieren. STATA bietet dazu bspw. mit dem „reshape“-Befehl eine Möglichkeit, ohne großen Aufwand zwischen den beiden Formaten zu wechseln. Das bis vor wenigen Jahren noch zentrale Argument gegen diese Art der Datensatzstruktur ist der deutlich höhere Bedarf an Speicherplatz, der sich insbesondere daraus ergibt, dass auch Variablen, die nur in einer oder in wenigen Befragungswellen erhoben wurden, immer eine komplette, alle Wellen übergreifende Spalte im Datensatz benötigen. Zudem werden die Long-Files mit zunehmender Laufzeit des Panels auch allein durch das wiederholte Aneinanderreihen der Einzelwellen relativ groß, wodurch der Speicherbedarf und die Dauer einzelner Operationen mit den Daten deutlich steigen. Die breite Verfügbarkeit schneller Prozessoren und großer Speicherkapazitäten selbst auf einfachen Desktop-PCs lässt dieses Gegenargument jedoch mittlerweile als vernachlässigbar erscheinen. Einen weiteren Nachteil stellt das Zuschlüsseln von Informationen dar. Im Gegensatz zu jenen im Wide-Format aufbereiteten Datensätzen benötigt man jetzt eine zusätzliche Schlüsselvariable, mit der eine Beobachtung eindeutig identifiziert werden kann. Dies kann ein Wellenindikator in den Haushalts- oder Personendatensätzen sein oder aber die Spellnummer in den

Spelldatensätzen, die ebenfalls im Long-Format vorliegen. Weiterhin ist nicht auf den ersten Blick ersichtlich, welche Variablen für welche Wellen erhoben wurden, da alle jemals erfragten Variablen im Datensatz vorhanden sind. Diese Variablen sind für die Wellen, in denen sie nicht erhoben wurden, mit einem Sondercode (-9) gefüllt.

In der Abwägung der Vor- und Nachteile des Long-Formats für die Nutzer überwiegen u.E. deutlich die Vorteile. Entsprechend wurden der Haushalts- und der Personendatensatz des PASS (*HHENDDAT*; *PENDDAT*) sowie die entsprechenden Gewichtungsdaten (*hweights*; *pweights*) im Long-Format aufbereitet.

Weiterhin enthält der Scientific Use File auf Haushaltsebene die in Spellform aufbereiteten Daten zum Arbeitslosengeld II-Bezug des Haushalts (*alg2_spells*). Auf Personenebene existiert ab Welle 4 ein integrierter biografischer Spelldatensatz (*bio_spells*), dieser integriert und ersetzt die bis Welle 3 vorliegenden Spelldatensätze *et_spells*, *al_spells* und *lu_spells*. Darüber hinaus liegt ab Welle 4 ein Ein-Euro-Spelldatensatz vor (*ee_spells*). Das Haushalts- und das Personenregister (*hh_register*; *p_register*) liegen im Wide-Format vor. In der Welle 5 wurde der Scientific Use File auf der Personenebene um einen Datensatz für das Vignettenmodul (*VIGDAT*) erweitert.

Abbildung 2: Datensatzstruktur des PASS in Welle 5



4 Generierte Variablen

4.1 Vercodung offener Angaben

Einige Items der Befragung wurden als geschlossene Items mit offener Restkategorie bzw. als offene Items erhoben. In diesen Fällen wurden i.d.R. zusätzliche Variablen generiert²¹, die sich von der ursprünglichen Version der Variable lediglich dahingehend unterscheiden, dass die offenen Angaben soweit als möglich den entsprechenden Kategorien zugecodet wurden. In einigen Fällen wurden zudem auch neue Kategorien auf Basis der offenen Angaben gebildet. Die Bezeichnung dieser Variablen unterscheidet sich von der ursprünglichen Variable häufig lediglich dadurch, dass die „0“ in der letzten Ziffer durch eine „1“ ersetzt wurde. Die Items zu Geburtsland, Staatsangehörigkeit und den Zuzugsländern von Eltern/Großeltern wurden zudem anonymisiert und tragen entsprechende Variablennamen²². Die Tabelle 9 und 10 geben einen Überblick über die offenen Angaben, die in der Welle 5 vercodet wurden²³.

Tabelle 9: Vercodung offener Angaben auf Haushaltsebene in Welle 5

fester Variablenname	zugecodet zu Variable	Datensatz	Bezeichnung
<i>HD1100a-o</i>	<i>HD1101a-o</i>	<i>HHENDDAT</i>	Erwerbsstatus der HH-Mitglieder, ggf. Proxy-Angabe
<i>HW0880a-i</i>	<i>HW0881a-j</i>	<i>HHENDDAT</i>	anderer, nicht in Liste enthaltener Auszugsgrund
<i>AL21300a-h</i>	<i>AL21301a-h</i>	<i>alg2_spells</i>	anderer, nicht in Liste enthaltener Grund für Kürzung
–	<i>AL21401a-h</i>		
<i>AL22100a-h</i>	<i>AL21501a-h</i>		
	<i>AL21601a-h</i>		
	<i>AL21701a-h</i>		
	<i>AL21801a-h</i>		
	<i>AL21851a-h</i>		
	<i>AL21901a-h</i>		
	<i>AL22001a-h</i>		
	<i>AL22101a-h</i>		
	<i>AL22102a-h</i>		
	<i>AL22103a-h</i>		
<i>AL22200a – AL22200h</i>	<i>AL22201a-h</i>	<i>alg2_spells</i>	anderer, nicht in Liste enthaltener Grund, warum kein ALG II mehr bezogen wurde
<i>AL20550a-h</i>	<i>AL20551a-h</i>	<i>alg2_spells</i>	anderer, nicht in Liste enthaltener Grund, warum ALG-II-Bezug begann

²¹ Andere offene Angaben wurden dagegen nicht vercodet, bspw. der Name des Trägers der Grundsicherung (PTK0100).

²² ogebland (Geburtsland); ostaatatan (Staatsangehörigkeit); ozulanda bis ozulandf (Zuzugsländer Eltern/Großeltern).

²³ Variablen, für die zwar in Vorwellen offene Angaben erhoben und vercodet wurden, jedoch nicht mehr in der aktuellen Welle, sind (mit Ausnahme des ALG-II-Spelldatensatzes) nicht aufgeführt. Diese Variablen sind für die Beobachtungen in den Wellen ohne Erhebung dieser Variablen mit -9 gefüllt (Item in Welle nicht erhoben) und in den Datenreports der Erhebungswellen dokumentiert.

Tabelle 10: Vercodung offener Angaben auf Personenebene in Welle 5

fester Variablenname	zugecodet zu Variable	Datensatz	Bezeichnung
<i>PB0230 (Code 6)</i>	<i>PB0231</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener deutscher Schulabschluss (Fortschreibung)
<i>PB0230 (Code 7)</i>	<i>PB0231</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener ausländischer Schulabschluss (Fortschreibung)
<i>PB0400 (Code 9)</i>	<i>PB0401</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener deutscher Schulabschluss (Ersterhebung o. keine Angabe in VW)
<i>PB0400 (Code 10)</i>	<i>PB0401</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener ausländischer Schulabschluss (Ersterhebung o. keine Angabe in VW)
<i>PB1000</i>	<i>PB1001</i>	<i>PENNDAT</i>	anderer, nicht in Liste enthaltener ausländischer Schulabschluss (Ersterhebung o. keine Angabe in VW)
<i>PB1300a-j (Code 9)</i>	<i>PB1301a-j</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener deutscher Berufsabschluss (Fortschreibung o. Ersterhebung)
<i>PB1300a-j (Code 10)</i>	<i>PB1301a-j</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener ausländischer Berufsabschluss (Fortschreibung o. Ersterhebung)
<i>PB1600</i>	<i>PB1601</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener Abschluss, dem der ausländische Abschluss entspricht
<i>AL0600</i>	<i>AL0601</i>	<i>bio_spells</i>	anderer, nicht in Liste enthaltener Grund, weshalb nicht mehr arbeitslos gemeldet
<i>BIO0100</i>	<i>BIO0101</i>	<i>bio_spells</i>	anderer, nicht in Liste enthaltener Aktivitätstyp
<i>EE0300a-h</i>	<i>EE0301a-h</i>	<i>ee_spells</i>	anderer Grund, warum keine Teilnahme am Ein-Euro-Job
<i>EE1000a-e</i>	<i>EE1001a-e</i>	<i>ee_spells</i>	anderer Grund, warum Ein-Euro-Job vorzeitig beendet
<i>PTK0320a-g</i>	<i>PTK0321a-g</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener Grund, weshalb keine Arbeit gesucht werden muss
<i>PEE0200a-d</i>	<i>PEE0201a-e</i>	<i>PENDDAT</i>	andere Quelle der Kenntnis von Ein-Euro-Jobs
<i>PAS0900a-g</i>	<i>PAS0901a-g</i> <i>PAS0901i</i>	<i>PENDDAT</i>	andere, nicht in Liste enthaltene Stellen, bei denen sich ZP über offene Stellen informiert hat
<i>PG0900a-f</i>	<i>PG0901a-g</i>	<i>PENDDAT</i>	andere, nicht in Liste enthaltene gesundheitl. Einschränkung
<i>PG1300</i>	<i>PG1301</i>	<i>PENDDAT</i>	andere, nicht in Liste enthaltene Krankenversicherung
<i>PP1300a-e</i>	<i>PP1301a-e</i>	<i>PENDDAT</i>	andere private Pfllegetätigkeiten
<i>PMI0200</i>	<i>ogebland</i>	<i>PENDDAT</i>	anderes, nicht in Liste enthaltene Geburtsland
<i>PMI0500</i>	<i>ostaatan</i>	<i>PENDDAT</i>	andere, nicht in Liste enthaltene Staatsangehörigkeit
<i>PMI1000a-f</i>	<i>ozulanda-f</i>	<i>PENDDAT</i>	anderes, nicht in der Liste enthaltene Zugzugsland des Elternteils/Großelternteils

Tabelle 10: Vercodung offener Angaben auf Personenebene in Welle 5 (Forts.)

fester Variablenname	zugecodet zu Variable	Datensatz	Bezeichnung
<i>PA1100²⁴</i>	<i>freiz1-3</i>	<i>PENDDAT</i>	Erste bis dritte ausgeübte Freizeitbeschäftigung
<i>PA1200</i>	<i>frwunsch</i>	<i>PENDDAT</i>	Gewünschte Freizeitbeschäftigung
<i>PA1300a-f</i>	<i>PA1301a-g</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener Grund für die Nichtausübung der Freizeitbeschäftigung
<i>PSH0200 (Code 9)</i>	<i>PSH0201</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener deutscher Schulabschluss der Mutter
<i>PSH0200 (Code 10)</i>	<i>PSH0201</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener ausländischer Schulabschluss der Mutter
<i>PSH0300a-i (Code 7)</i>	<i>PSH0301a-i</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener deutscher Berufsabschluss der Mutter
<i>PSH0300a-i (Code 8)</i>	<i>PSH0301a-i</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener ausländischer Berufsabschluss der Mutter
<i>PSH0500 (Code 9)</i>	<i>PSH0501</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener deutscher Schulabschluss des Vaters
<i>PSH0500 (Code 10)</i>	<i>PSH0501</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener ausländischer Schulabschluss des Vaters
<i>PSH0600a-i (Code 7)</i>	<i>PSH0601a-i</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener deutscher Berufsabschluss des Vaters
<i>PSH0600a-i (Code 8)</i>	<i>PSH0601a-i</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener ausländischer Berufsabschluss des Vaters

4.2 Harmonisierung

Bei einigen Variablen gab es über die Wellen hinweg Veränderungen an den Erhebungsinstrumenten. Vor allem die Integration des Erwerbsbiografiemoduls in Welle 2 führte dazu, dass zentrale Informationen zum Erwerbsstatus, der aktuellen Haupterbstätigkeit, dem Status der Nichterwerbstätigkeit und dem Bezug von Arbeitslosengeld I anders erhoben wurden als noch in der Welle 1. Informationen werden seitdem nicht mehr nur bezogen auf den Befragungszeitpunkt erhoben, sondern in Episodenform für bestimmte Zeiträume erfasst.

Um in solchen Fällen wellenübergreifende Analysen zu vereinfachen, werden für wichtige Konstrukte über die Wellen hinweg harmonisierte Variablen generiert. Harmonisierungen stellen damit eine Sondergruppe der generierten Variablen dar (siehe Abschnitt 4.4), die dazu dienen, unterschiedlich erhobene Konstrukte im Nachhinein zu vereinheitlichen.

Veränderungen zwischen den Wellen können das gesamte Erhebungskonzept, Kategorien und die befragten Gruppen betreffen. Harmonisierte Variablen berücksichtigen deshalb unterschiedliche Quellvariablen, die sich aus veränderten Erhebungskonzepten ergeben, Veränderungen an Kategorien sowie an den befragten Gruppen. Es wurde damit versucht, diese soweit möglich über die Wellen hinweg zu vereinheitlichen, bevor auf ihrer Basis die Generierung durchgeführt wurde.

²⁴ Die Variable PA1100 selbst ist im PENDDAT nicht enthalten, da sie keine weitere Informationen enthält als den Umstand, ob eine Zielperson eine offene Angabe gemacht hat oder die Frage mit wn oder ka beantwortet hat. Die Angaben von wn und ka in PA1100 wurden in die Variablen freiz1-3 übernommen.

Eine Harmonisierung wurde bisher für die einfache Einteilung der Stellung im Beruf (stibkz) vorgenommen. Allerdings ist damit zu rechnen, dass sich die Zahl der notwendigen Harmonisierungen mit zunehmender Laufzeit des Panels erhöht.

Tabelle 11: Harmonisierte Variablen im Personendatensatz (PENDDAT)

Variable	thematischer Bereich	Bezeichnung
<i>stibkz</i>	Erwerbstätigkeit	jetzige Stellung im Beruf, einfache Einteilung, harmonisiert (anonymisiert)

Während die ausdrücklich harmonisierten Variablen – neben Veränderungen am Erhebungskonzept – auch Veränderungen an Kategorien und den befragten Gruppen über die Wellen berücksichtigen, werden von einem zweiten Variablentyp Veränderungen an den befragten Gruppen nicht explizit berücksichtigt. Diese Variablen werden zwar für alle Wellen generiert, können jedoch je nach Welle Informationen für verschiedene Befragtengruppen enthalten. Diese Unterschiede ergeben sich aus Überarbeitungen der Filterführung, die zwischen den Wellen vorgenommen wurden und sich auf die jeweiligen Quellvariablen einer generierten Variable auswirken.

Wellenübergreifende Variablen dieses Typs treten somit neben die eigentlichen Harmonisierungen und vereinheitlichen einzelne Aspekte zwischen den Wellen. Im Gegensatz zu den harmonisierten Variablen werden sie in jeder Welle jeweils für alle Gruppen generiert, für die in dieser Welle die jeweiligen Quellvariablen erhoben wurden. Für Auswertungen im Querschnitt einer spezifischen Welle sind sie somit einfach nutzbar. Im Längsschnitt müssen diese Unterschiede jedoch beachtet werden, bevor Aussagen über Veränderungen zwischen den Wellen getroffen werden können.

Aus diesem Grund sollte vor der Arbeit mit den wellenübergreifenden, jedoch nicht harmonisierten Variablen geprüft werden, ob Unterschiede in den befragten Gruppen problematisch für die jeweiligen Auswertungen sein könnten und ob ggf. eine Vereinheitlichung notwendig ist²⁵.

²⁵ Bspw. wurden in Welle 1 andere Befragtengruppen nach einer Erwerbstätigkeit gefragt als in den Folgewellen. Entsprechend haben auch jeweils andere Gruppen Angaben zur StiB, zur beruflichen Tätigkeit, zur Arbeitszeit, zur Befristung etc. gemacht.

Insbesondere die folgenden, wellenübergreifenden Variablen weisen Unterschiede hinsichtlich der Gruppen auf, für die sie generiert sind:

Tabelle 12: Wellenübergreifend generierte, aber nicht vollständig harmonisierte Variablen im Personendatensatz (PENDDAT)

Variable	thematischer Bereich	Bezeichnung
<i>isco88</i>	Erwerbstätigkeit	ISCO 88 (ZUMA-Verkodung), aktuelle Erwerbstätigkeit, generiert
<i>kldb</i>	Erwerbstätigkeit	Klassifizierung der Berufe 1992, aktuelle Erwerbstätigkeit
<i>azhpt2</i>	Erwerbstätigkeit	Akt. tatsächl. Arb.zeit Haupterwerb. (ohne Minijobs, inkl. kat. Ang.), gen.
<i>azges2</i>	Erwerbstätigkeit	Akt. tatsächl. Arb.zeit insgesamt (ohne Minijobs, inkl. kat. Ang.), gen.
<i>befrist</i>	Erwerbstätigkeit	Jetzige Tätigkeit: Befristung? Generiert (alle Wellen)
<i>mps</i>	Erwerbstätigkeit	Magnitude-Prestigeskala, aktuelle Erwerbstätigkeit, generiert
<i>siops</i>	Erwerbstätigkeit	Standard International Occupational Prestige Scale, aktuelle Erwerbstätigkeit, generiert
<i>isei</i>	Erwerbstätigkeit	International Socio-Economic Index, aktuelle Erwerbstätigkeit, generiert
<i>egp</i>	Erwerbstätigkeit	Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), aktueller Beruf, generiert
<i>esec</i>	Erwerbstätigkeit	European Socio-economic Classification (ESeC), aktueller Beruf, generiert
<i>stib</i>	Erwerbstätigkeit	Stellung im Beruf, Kennziffer, aktuelle Erwerbstätigkeit, generiert
<i>netges</i>	Erwerbstätigkeit	Akt. Nettoeinkommen insgesamt (ohne Minijobs, inkl. kat. Ang.), gen.
<i>alg1abez</i>	Leistungsbezug	Aktueller ALG1-Bezug, generiert
<i>aktmassn</i>	Maßnahmeteilnahme	Aktuelle Teilnahme an einem vom Arbeitsamt finanzierten/ geförderten Programm, generiert

4.3 Dependent Interviewing

An verschiedenen Stellen wurden sowohl in den Haushaltsinterviews als auch in den Personeninterviews Informationen per Dependent Interviewing, d.h. in Abhängigkeit von Angaben in der Vorwelle, erhoben. Dabei wurden Angaben aus der letzten Befragung zur Steuerung der Filterführung genutzt oder direkt als Teil des Fragetexts im aktuellen Interview eingeblendet.

Mit der Nutzung von Vorwelleninformationen wurden insbesondere zwei Ziele verfolgt. Erstens sollten an bestimmten Stellen nur Veränderungen seit der Vorwelle erfasst werden, teilweise in Abhängigkeit davon, ob schon in der Vorwelle Informationen zu einem bestimmten Fragekomplex vorlagen²⁶. An diesen Stellen wurden Vorwelleninformationen zur Filtersteuerung genutzt. Zweitens sollten dem Befragten inhaltliche Informationen gegeben werden. An den Stellen, an denen Veränderungen seit der Vorwelle erfasst werden sollten, wurde deshalb das Interviewdatum der Vorwelle in den Fragetext eingebunden,

²⁶ Bspw. wurden Personen nur einmal nach ihrem höchsten Schulabschluss gefragt. Haben sie einmal eine entsprechende Angabe gemacht, werden in den folgenden Wellen nur noch neue Schulabschlüsse erfragt, die seit der jeweils letzten Befragung erworben wurden.

um den Berichtszeitraum deutlicher abzugrenzen²⁷. An anderen Stellen, insbesondere bei der Fortschreibung von Episodeninformationen²⁸, wurden auch Antworten des Befragten aus der Vorwelle in die Fragetexte integriert, die dazu dienten, den Befragten an die Angaben in der Vorwelle zu erinnern. Damit sollte verhindert werden, dass Statuswechsel berichtet werden, die in der Realität nicht aufgetreten sind, sondern ein Artefakt der offenen Erhebung und damit von Erinnerungsfehlern oder ungenauen Angaben sind.

Betrachtet man also die Informationen einer einzelnen Welle im Datensatz, so liegen infolge des Dependent Interviewings für einige Befragte nur unvollständige Informationen vor, die lediglich die Veränderung zwischen zwei Befragungszeitpunkten widerspiegeln. Für Befragte, die erstmals zu einem Thema befragt wurden, können dagegen Informationen vorliegen, die bezogen auf diese Welle vollständig sind²⁹.

Im Zuge der Datenaufbereitung werden die erhobenen Veränderungen mit Informationen aus der Vorwelle kombiniert, um auch Variablen bzw. Datensätze mit vollständigen Informationen zu erzeugen. Mit den neu erhobenen Episodeninformationen werden die Spells in den bestehenden Spelldatensätzen fortgeschrieben. In den Querschnittdatensätzen (*HHENDDAT*, *PENDDAT*) werden dagegen generierte Variablen erstellt, in denen die Informationen der Vorwelle mit den erhobenen Veränderungen kombiniert werden.

Im Folgenden geben die

²⁷ Sollten bspw. nur neue Schulabschlüsse seit dem letzten Interview berichtet werden, wurde zunächst gefragt: „Haben Sie seit unserem letzten Interview am [Einblendung Interviewdatum der Vorwelle] einen allgemeinbildenden Schulabschluss gemacht?“

²⁸ Beispiele sind die Fortschreibung des Arbeitslosengeld-II-Bezugs aus der Vorwelle im Haushaltsinterview der jeweils aktuellen Welle bzw. die Fortschreibung von Erwerbstätigkeiten oder Arbeitslosigkeiten im Personeninterview.

²⁹ Personen, die erstmals zu ihrem Schulabschluss befragt werden, berichten den jeweils höchsten Schulabschluss. Die Information zum höchsten Schulabschluss liegt in dieser Welle also vollständig in den erhobenen Variablen vor. In der Folgewelle wird nur noch nach neu erworbenen Schulabschlüssen gefragt. Wurde bspw. ein Schulabschluss neu erhoben, so liegt diese Information zwar in den erhobenen Variablen vor, es ist allerdings unklar, ob es sich tatsächlich um den höchsten Schulabschluss handelt. Die Information in der Folgewelle ist in diesem Sinne in den berichteten Variablen unvollständig.

Tabelle 13a und 13b einen kurzen Überblick über alle relevanten Stellen in den Fragebögen und zeigen, in welcher Variable die fortgeschriebene Information zu finden ist. Die Fälle, in denen die Fortschreibung in generierten Variablen erfolgt, sind zusätzlich in Kapitel 4.4 des vorliegenden Datenreports aufgeführt.

Tabelle 13a: Fortgeschriebene Vorwelleninformationen in Welle 5, Haushaltsfragebogen

Haushaltsfragebogen für wiederholt befragte Haushalte (HHalt)			
Konstrukt	Fragenr.	Anmerkung	Fortschreibung in Var.
Wohnsituation		Wohnform, Art des Mietverhältnisses und die Art des Wohnheims während des Interviews fortgeschrieben	HHENDDAT: <i>HW0200 bis HW0400</i>
Haushaltsstruktur		Haushaltsgröße während des Interviews fortgeschrieben	HHENDDAT: <i>HA0100</i>
		Geschlecht der Personen im HH während des Interviews ggf. korrigiert	HHENDDAT: <i>HD0100a bis HD0100o</i>
		Alter der Personen im HH während des Interviews fortgeschrieben	HHENDDAT: <i>HD0200a bis HD0200o</i>
		Verwandtschaftsbeziehungen während des Interviews fortgeschrieben	nicht im SUF ausgeliefert
Wohnfläche der Wohnung in qm Arbeitslosengeld-II-Bezug	HW1000	in generierter Variable fortgeschrieben	HHENDDAT: <i>wohnfl</i>
	Modul „Arbeitslosengeld 2“	in Arbeitslosengeld-II-Spell-datensatz fortgeschrieben	alg2_spells: Variablen des Arbeitslosengeld-II-Spell-datensatzes
		Informationen zum aktuellen Arbeitslosengeld-II-Bezug des HH	HHENDDAT: <i>alg2abez</i> PENDDAT: <i>hhalg2</i>
		Information zum Arbeitslosengeld-II-Bezug der Bedarfsgemeinschaft	<i>p_register:</i> <i>bgbez5; bgbez5</i>

Tabelle 13b Fortgeschriebene Vorwelleninformationen in Welle 5, Personenfragebogen

Personenfragebogen			
Konstrukt	Fragenr.	Anmerkung	Fortschreibung in Var.
höchster allgemeinbildender Schulabschluss	PB0220- PB1100	in generierter Variable fortgeschrieben	PENDDAT: <i>schul1</i> (ohne offene Ang.) <i>schul2</i> (mit offenen Ang.)
Jahr des höchsten Schulabschlusses	PB0410	in generierter Variable fortgeschrieben	PENDDAT: <i>schulabj</i>
Berufsabschluss	PB1200- PB1600	höchster Berufsabschluss in generierter Variable fortgeschrieben	PENDDAT: <i>beruf1</i> (ohne offene Ang.) <i>beruf2</i> (mit offenen Ang.)
Jahr des Berufsabschlusses	PB1310	in generierter Variable fortgeschrieben	<i>berabj</i>
Zeiten von fortgeschriebenen Aktivitäten im BIO-Spelldatensatz	BIO0200, BIO0800, BIO0300	in BIO-Spelldatensatz für Aufsatzspells fortgeschrieben	<i>bio_spells</i> <i>BIO0400, BIO0500, BIO0600</i>
		in BIO-Spelldatensatz für Aufsatzspells fortgeschrieben Informationen zur aktuellen Erwerbstätigkeit in generierten Variablen fortgeschrieben	<i>bio_spells:</i> ET2300, ET2700 PENDDAT: <i>isco88; kldb;stib; stibkz; arbeitszeit; befrist; mps; siops; isej; egp; esec</i>
Zeiten von Arbeitslosengeld-I-Bezug bei fortgeschriebenen Arbeitslosigkeits-Spells		Informationen zum aktuellen (Nicht-)Erwerbstätigkeitsstatus in generierten Variablen fortgeschrieben	PENDDAT: <i>etakt; alakt; statakt</i>
		Informationen zum aktuellen Arbeitslosengeld-I-Bezug	<i>bio_spells:</i> AL0700, AL0800, AL0900, <i>AL1000, AL1100, AL1200</i>
Zeiten von fortgeschriebenen Aktivitäten im EE-Spelldatensatz		in BIO-Spelldatensatz für Aufsatzspells fortgeschrieben	<i>bio_spells:</i> AL0600, AL0601 PENDDAT: <i>alg1abez</i>
			<i>ee_spells:</i> <i>EE0800a, EE0800b</i>
Angaben zum vorzeitigen Ende im EE-Spelldatensatz			<i>ee_spells:</i> <i>EE0900,</i> <i>EE1000a-EE1000e, EE1001a-EE1001e</i>

Von diesen Merkmalen, bei denen in der Vergangenheit erhobene Informationen mit Angaben zu Veränderungen zwischen den Befragungszeitpunkten fortgeschrieben bzw. aktualisiert werden, sind die so genannten „unveränderlichen Merkmale“ zu unterscheiden. Bei ihnen wird davon ausgegangen, dass sie sich im Zeitverlauf nicht verändern. Entsprechend werden sie im PASS nur einmal erhoben, auch wenn teilweise spätere Korrekturen möglich sind. Da Angaben zu diesen Merkmalen in der Regel nur für den Zeitpunkt der ersten Befragung in den erhobenen Variablen verfügbar sind, werden sie danach in Form generierter Variablen zur Verfügung gestellt (siehe dazu Kapitel 4.4, User Guide PASS Wave 3).

4.4 Einfache Generierungen

Die einfachen Generierungen umfassen Variablen, bei denen mehrere, aus erhebungstechnischen Gründen getrennte Items eines Konstrukts zusammengefasst oder Informationen der aktuellen Welle mit Vorwelleninformation kombiniert (siehe Kapitel 4.3) (wie z. B. beim höchsten Bildungsabschluss) oder wichtige Informationen aus anderen Teildatensätzen zugespielt wurden (z. B. Indikatoren für den aktuellen Bezug von Arbeitslosengeld I oder Arbeitslosengeld II).

Bei Haushalten bzw. Personen, die erstmals zu einem Thema befragt werden, können die einfachen generierten Variablen immer auf Basis der in der aktuellen Welle erhobenen Informationen erzeugt werden. Bei Haushalten bzw. Personen, die bereits in einer früheren Welle Angaben zu einem Thema gemacht haben, können sie allerdings in den Querschnittdatensätzen (*HHENDDAT*; *PENDDAT*) hinsichtlich der Herkunft der zu ihrer Generierung jeweils notwendigen Variablen unterschieden werden. Die drei verschiedenen Typen der einfachen generierten Variablen sind in Tabelle 14 aufgeführt.

Tabelle 14: Typen der einfachen generierten Variablen in den Querschnittdatensätzen (*HHENDDAT*; *PENDDAT*) bei Haushalten bzw. Personen, die schon in einer früheren Welle Angaben zum jeweiligen Thema gemacht haben

Typ	Generierung auf Basis v. Quelldaten		Beschreibung
	aus Welle der erstmaligen Erfassung des Themas bei HH/Person	aktuelle Welle	
<i>unveränderlich (uv)</i>	ja	nein	Die bei erstmaliger Erhebung erfassten Informationen werden in die Folgewelle grundsätzlich übernommen – außer in der aktuellen Welle wurden Eingabefehler korrigiert. <u>Bsp:</u> <i>zpsex</i> (Geschlecht)
<i>fortgeschrieben (fs)</i>	ja	ja	Die in der Vorwelle aktuellen Informationen werden mit Angaben der aktuellen Welle kombiniert und ggf. fortgeschrieben. <u>Bsp:</u> <i>schul1</i> (Höchster Schulabschluss)
<i>unabhängig neu (neu)</i>	nein	ja	Die Variable wird in jeder Welle unabhängig von Angaben aus der Vorwelle aus den Angaben der aktuellen Welle neu generiert. <u>Bsp:</u> <i>hhincome</i> (Haushaltsnettoeinkommen)

Zum Typ „unveränderlich (uv)“ muss mit Blick auf die einfachen Generierungen für *PENDDAT* eine vertiefende Erläuterung vorgenommen werden. Eine erstmalige Erfassung eines Themas findet bei einer Person u.U. nicht nur in der ersten Welle, in der die Person ein Personen-/Senioreninterview gibt, statt. Zwei Personengruppen werden erneut

als erstbefragte Person behandelt, auch wenn sie zum wiederholten Mal ein Personen-/Senioreninterview geben.

Zum einen handelt es sich um Personen, die in einen Haushalt wiedereinziehen. Personen, die aus ihrem bisherigen Haushalt in einen Splithaushalt umziehen (siehe auch Kapitel 2.4), nehmen ihre Preloadinformationen mit. Somit können sie auch im Splithaushalt korrekt als erst- oder wiederzubefragende Personen gesteuert werden. Wenn eine Person hingegen aus einem Splithaushalt in einen in einer Vorwelle bewohnten Haushalt des Panels zurückzieht, findet keine Übernahme des Preloads für diese Person aus dem Splithaushalt in den Ursprungshaushalt statt. Die wiedereinziehenden Personen werden daher wie neubefragte Personen behandelt. Dieser Sachverhalt tritt seit Welle 3 auf, da es in Welle 2 erstmals zu Auszügen aus wiederbefragten Haushalten und somit seit Welle 3 erstmals zu Wiedereinzügen von in einer Vorwelle ausgezogenen Personen kommen kann.

Zum anderen wird für eine Person nur dann personenbezogener Preload für das Dependent Interviewing (siehe Kapitel 4.3) erstellt, wenn sie in einer der beiden direkten Vorwellen ein Personen-/Senioreninterview gegeben hat. Hintergrund ist, dass eine Grenze gezogen werden soll, bis zu welchem Zeitpunkt sich eine Person an in Spellform erhobene Ereignisse zurückerinnern soll. Bei Personen, die ein Personen-/Senioreninterview letztmals in der dritten Vorgängerwelle oder früher gegeben haben, läge der Referenzzeitpunkt vor dem für erstbefragte Personen relevanten Datum. Um die Belastung der Zielperson zu begrenzen und in der Annahme, dass die Validität der erhobenen Informationen jenseits dieser Grenze zu stark bedroht ist, werden Personen, deren Referenzzeitpunkt für Auskünfte über Spellereignisse vor dem für erstbefragte Personen relevanten Datum liegt, wie erstbefragte Personen behandelt.³⁰ Dieser Sachverhalt tritt seit Welle 4 auf, da hier erstmals ein früheres Personeninterview mehr als zwei Wellen zurückliegen kann.

Für diese beiden Personengruppen werden die den „unveränderlichen“ Generierungen zugrunde liegenden Informationen erneut erhoben (z. B. im Modul „Soziale Herkunft“), da diese Personen erneut als erstbefragte Personen gesteuert werden. Auch in der Datenaufbereitung werden die erhobenen Informationen genau wie die Angaben von tatsächlich zum ersten Mal im Rahmen des PASS befragten Personen weiterverarbeitet. Diese Generierungen, z. B. bei den Statusangaben zu Mutter und Vater, basieren also auf der aktuellen Welle. Eine Übernahme von Informationen aus der Vorwelle findet nicht statt, genauso wenig eine Plausibilisierung mit den früheren Angaben. Zwar kann grundsätzlich angenommen werden, dass die Angaben der Zielpersonen, die zu den „unveränderlichen“ Generierungen weiterverarbeitet werden, im Falle der erneuten Erhebung konsistent sind gegenüber früheren Angaben. Inkonsistenzen und damit Abweichungen gegenüber Informationen aus den Vorwellen können jedoch nicht grundsätzlich ausgeschlossen wer-

³⁰ Hiervon ausgenommen ist die Information, ob eine Person in einem früheren Interview bereits zu ihrer Zuspieldbereitschaft befragt worden ist. Diese Preloadinformation wird generiert, unabhängig davon, wie lange ein früheres Personeninterview zurückliegt. Es wird auf diese Weise vermieden, dass Personen, die in einer Vorwelle ihre Einwilligung gegeben haben, diese Frage *RegP0100* in einer folgenden Welle verneinen und ihre Einwilligung damit faktisch zurückziehen. Die Möglichkeit für die Zielperson, ihre Zuspieldbereitschaft zu widerrufen, bleibt von dieser Entscheidung unberührt.

den. Personen, die zu einer der beiden beschriebenen Gruppen zählen, können in *PENDDAT* darüber identifiziert werden, dass sie in mehr als einer Welle in der Kennvariable *altbefr* als erstbefragt (Code „0“ bzw. für Welle 1 Code „-9“) geflaggt sind.

Die einfachen generierten Variablen sind in den datensatzspezifischen Tabelle 15 bis 20 dargestellt. Die einzelnen Variablen werden dabei jeweils kurz beschrieben. Darüber hinaus werden die für die Generierung der Variable in Welle 5 notwendigen Quellvariablen aufgeführt³¹. Für die Querschnittdatensätze (*HHENDDAT*; *PENDDAT*) ist zudem jeweils angegeben, um welchen Typ der in Tabelle 16 dargestellten einfacheren Generierung es sich handelt (uv; fs; neu). Für die Spelldatensätze ist diese Unterteilung nicht sinnvoll, da es hier keine wellenspezifischen Beobachtungen gibt. Stattdessen werden die generierten Variablen jeweils auf Spellebene neu erzeugt, wenn der Spell in der aktuellen Welle neu angelegt wurde oder mit in der aktuellen Welle erhobenen Angaben fortgeschrieben wurde. Auch die Registerdatensätze folgen einer anderen Logik, so dass auch hier auf eine weitere Differenzierung verzichtet wurde.

Tabelle 15: Einfache Generierungen für Welle 5 im Haushaltsdatensatz (*HHENDDAT*) (alphabetische Reihenfolge)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>alg2abez</i>	<i>aktueller ALG-II-Bezug des HH, generiert</i> Indikator für aktuellen Arbeitslosengeld-II-Bezug des Haushalts (neu)	<i>zensiert</i> ; <i>AL20300</i> ; <i>AL20400</i> ; <i>AL20500 (alg2_spells)</i> ; Information zu weiteren Arbeitslosengeld-II-Bezügen (<i>AL22700</i>); <i>hintjahr (HHENDDAT)</i>
<i>bik</i>	<i>BIK-Regionsgrößenklassen (GKBIK10), generiert</i> Die Information zur Regionsgrößenklasse wurde von infas generiert, indem von der in den Adressdaten vorliegenden Postleitzahl auf <i>GKBIK10</i> umgeschlüsselt wurde (neu)	vom Erhebungsinstitut geliefert
<i>blneualt</i>	<i>alte oder neue Bundesländer, generiert</i> Zusammenfassung der Bundesländer zu alten Bundesländern (ohne Berlin) und neuen Bundesländern (mit Berlin). Das Bundesland wurde von infas auf Basis der in den Adressdaten vorliegenden Postleitzahlen bestimmt (neu)	vom Erhebungsinstitut generierte und gelieferte Information zum Bundesland, in dem Haushalt zum Befragungszeitpunkt lebt

³¹ Wie die Variablen in den Querschnittdatensätzen (*HHENDDAT*; *PENDDAT*) für Beobachtungen der Vorwellen generiert wurden, ist in den jeweiligen Datenreports dokumentiert. Ebenfalls in der Dokumentation der jeweiligen Welle beschrieben ist die Generierung der wellenspezifischen Variablen in den Registerdatensätzen. Die generierten Variablen in den Spelldatensätzen wurden immer in den bereits fortgeschriebenen Datensätzen erzeugt. Wurde ein Spell nicht fortgeschrieben, blieben die entsprechenden generierten Variablen unverändert (ggf. mit Ausnahme, dass im Zensierungsindikator ein Sondercode gesetzt wurde, falls der Spell aus technischen Gründen nicht fortgeführt werden konnte). Wurde ein Spell fortgeschrieben, wurde immer auf die aktuellsten Informationen zurückgegriffen, d.h. auf die mit den Informationen aus der aktuellen Welle gefüllten Variablen bzw. die für die aktuelle Welle relevanten Querschnittvariablen in den Spells.

Tabelle 15: Einfache Generierungen für Welle 5 im Haushaltsdatensatz (HHENDDAT) (alphabetische Reihenfolge) (Fortsetzung 1)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>hhinckat</i>	<i>kategorisiertes Haushaltseinkommen pro Monat (in Euro), generiert</i> Zusammenfassung der mehrstufig erhobenen kategorialen Angaben zum Haushaltseinkommen zu einer Variablen (neu)	HEK0700; HEK0800; HEK0900; HEK1000; HEK1100 (HHENDDAT)
<i>hhincome</i>	<i>Haushaltseinkommen pro Monat (in Euro) inkl. kategorisierter Angaben, generiert</i> Generierung einer integrierten Variable aus kategorisierter und offener Angabe zum Haushaltsnettoeinkommen (neu)	HEK0600; HEK0700; HEK0800; HEK0900; HEK1000; HEK1100 (HHENDDAT)
<i>hintdat</i>	<i>Interviewdatum Haushaltsinterview</i> Generierte Variable mit dem Datum, an dem das Haushaltsinterview geführt wurde im Format JJMMTT (neu)	<i>hintjahr; hintmon; hinttag</i> (HHENDDAT)
<i>kindu4</i>	<i>Steuervariable: Kind unter 4 Jahre im HH</i> Die Variable zeigt an, dass in der Welle mind. eine Person im Haushalt unter vier Jahre alt ist. Da die Generierung nur auf den Altersangaben im Haushaltsdatensatz basiert, ist unerheblich, ob diese Person im Alter von vier Jahren auch tatsächlich Kind einer anderen im Haushalt lebenden Person ist (neu)	HD0200a - HD0200o (HHENDDAT)
<i>kindu13</i>	<i>Steuervariable: Kind unter 13 Jahre im HH</i> Die Variable zeigt an, dass in der Welle mind. eine Person im Haushalt unter 13 Jahre alt ist. Da die Generierung nur auf den Altersangaben im Haushaltsdatensatz basiert, ist unerheblich, ob diese Person im Alter von 13 Jahren auch tatsächlich Kind einer anderen im Haushalt lebenden Person ist (neu)	HD0200a - HD0200o (HHENDDAT)
<i>kindu15</i>	<i>Steuervariable: Kind unter 15 Jahre im HH</i> Die Variable zeigt an, dass in der Welle mind. eine Person im Haushalt unter 15 Jahre alt ist. Da die Generierung nur auf den Altersangaben im Haushaltsdatensatz basiert, ist unerheblich, ob diese Person im Alter von 15 Jahren auch tatsächlich Kind einer anderen im Haushalt lebenden Person ist. Fehlte die offene Altersangabe, wurde die kategoriale Nachfrage zu den Altersgruppen in die Generierung einbezogen (neu)	HD0200a - HD0200o; kategoriale Nachfrage zu Altersgruppe (bei KA in HD0200) (HHENDDAT)
<i>wohnfl</i>	<i>Wohnfläche in qm, generiert</i> Information zur Wohnfläche der aktuellen Wohnung des Haushalts. Bei wiederbefragten Haushalten wurde die Wohnungsgröße ab der Welle 2 nur erhoben, wenn seit der Vorwelle ein Umzug oder eine Veränderung der Wohnung stattgefunden hatte (fs)	<u>Bei Ersterfassung:</u> HW1000 (HHENDDAT) <u>Bei wiederholter Erfassung:</u> <i>wohnfl</i> aus Vorwelle; HW1000; (HHENDDAT)

Tabelle 16: Einfache Generierungen für Welle 5 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
akt1euro	<i>akt. Teiln. an einem 1-Euro-Job, generiert</i> Indikator: Befragungsperson nimmt zum Interviewzeitpunkt an einem Ein-Euro-Job teil (neu)	zensiert (ee_spells)
alakt	<i>aktuell arbeitslos gemeldet, generiert (ab Welle 2)</i> Zeigt an, dass die ZP zum Zeitpunkt des Personeninterviews der jeweiligen Welle arbeitslos gemeldet war (neu)	zensiert; spinteg; BIO0101 (bio_spells)
alg1abez	<i>aktueller ALG1-Bezug, generiert</i> Indikator: Befragungsperson bezieht zum Interviewzeitpunkt Arbeitslosengeld I. In der Welle 5 wurden die Zeiten seit Januar 2009 erhoben, in denen der Befragte arbeitslos gemeldet war. Für jede Episode wurde zudem erfragt, ob und in welchem Zeitraum Arbeitslosengeld I bezogen wurde (neu)	AL0700; AL1000; AL1100; AL1200 (bio_spells)
apartner	<i>Steuervariable: Unverh. Partner im HH</i> Indikator: Befragungsperson hat einen unspezifischen oder unverheirateten Partner im HH (neu)	Informationen zu Beziehungen im Haushalt (Haushaltsgrid); PD0500 - PD0900 (PENDDAT)
azhpt1	<i>akt. vertragl. Arb.zeit Haupterwerb. (ohne Minijobs), gen.</i> Wöchentliche, vertraglich vereinbarte Arbeitszeit der Haupterwerbstätigkeit, die der Befragte zum Interviewzeitpunkt ausübt, generiert aus offenen Angaben zur Arbeitszeit (neu)	ET2003 (bio_spells)
azhpt2	<i>akt. tatsächl. Arb.zeit Haupterwerb. (ohne Minijobs, inkl. kat. Ang.), gen.</i> Wöchentliche tatsächliche Arbeitszeit der Haupterwerbstätigkeit, die der Befragte zum Interviewzeitpunkt ausübt, generiert aus offenen Angaben zur Arbeitszeit und kategorialer Nachfrage bei unregelmäßigen Arbeitszeiten (neu)	ET2103; ET2203 (bio_spells)
azges1	<i>akt. vertragl. Arb.zeit insgesamt (ohne Minijobs), gen.</i> Wöchentliche, vertraglich vereinbarte Arbeitszeit aller Erwerbstätigkeiten, die der Befragte zum Interviewzeitpunkt ausübt, generiert aus offenen Angaben zur Arbeitszeit (neu)	ET2003 (bio_spells)
azges2	<i>akt. tatsächl. Arb.zeit insgesamt (ohne Minijobs, inkl. kat. Ang.), gen.</i> Wöchentliche, tatsächliche Arbeitszeit aller Erwerbstätigkeiten, die der Befragte zum Interviewzeitpunkt ausübt, generiert aus offenen Angaben zur Arbeitszeit und kategorialer Nachfrage bei unregelmäßigen Arbeitszeiten (neu)	ET2103; ET2203 (bio_spells)

Tabelle 16: Einfache Generierungen für Welle 5 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 1)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>befrist</i>	<i>jetzige Tätigkeit: Befristung?</i> <i>generiert (alle Wellen)</i> Indikator: Die Erwerbstätigkeit, die der Befragte zum Interviewzeitpunkt ausübt, ist befristet (neu)	<i>PET2510a; PET2510b</i> (PENDDAT)
<i>begjeewt</i>	<i>Anfangsjahr der ersten Erwerbstätigkeit, generiert</i> Jahr, in dem der Befragte erstmals hauptberuflich tätig war. Für die Generierung wurden Informationen zur ersten Erwerbstätigkeit mit Informationen aus den Erwerbstätigkeitsepisoden kombiniert, wenn der Befragte die erste Erwerbstätigkeit bereits bei der Erhebung der Erwerbstätigkeitsepisoden seit Januar 2009 genannt hat (uv)	<u>Bei Ersterfassung:</u> <i>bjahr (bio_spells); PET3200b</i> (PENDDAT) <u>Nach Ersterfassung:</u> <i>begjeewt</i> aus Vorwelle (PENDDAT)
<i>begmeewt</i>	<i>Anfangsmonat der ersten Erwerbstätigkeit, generiert</i> Monat, in dem der Befragte erstmals hauptberuflich tätig war (Generierung siehe <i>begjeewt</i>) (uv)	<u>Bei Ersterfassung</u> <i>bmonat (bio_spells); PET3200a</i> (PENDDAT) <u>Nach Ersterfassung:</u> <i>begmeewt</i> aus Vorwelle (PENDDAT)
<i>berabj</i>	<i>Jahr des höchsten Berufsabschlusses</i> Jahr, in dem der Befragte seinen zum Interviewzeitpunkt höchsten Berufsabschluss erworben hat (fs) <u>Anmerkung:</u> Die Abschlussjahre der in der Welle 1 berichteten Berufsabschlüsse wurden in der Welle 2 nacherhoben.	<u>Bei Ersterfassung:</u> <i>PB1310aj-kj</i> (PENDDAT) <u>Bei wiederholter Erfassung:</u> <i>berabj</i> aus Vorwelle; <i>PB1310aj-kj</i> (PENDDAT)
<i>beruf1</i>	<i>höchster beruflicher Abschluss, ohne ausl. Abschl. und off. Angaben, generiert</i> Identifikation des zum jeweiligen Interviewzeitpunkt höchsten Berufsabschlusses durch Hierarchisierung der von den Befragungspersonen genannten beruflichen Abschlüsse, ohne offene Angaben (fs)	<u>Bei Ersterfassung:</u> <i>PB0100; PB0200; PB0300; PB1200b; PB1200c; PB1300a-j;</i> (PENDDAT) <u>Bei wiederholter Erfassung:</u> <i>beruf1</i> aus Vorwelle; <i>PB0100; PB0200; PB1200a; PB1300a-j</i> (PENDDAT)
<i>beruf2</i>	<i>höchster beruflicher Abschluss, mit ausl. Abschl. und off. Angaben, generiert</i> Wie <i>beruf1</i> mit folgenden Unterschieden: 1. Einbeziehen von offenen Angaben; 2. Einbeziehen von Angaben zu ausländischen Abschlüssen; 3. Differenzierung der Studienabschlüsse nicht nach FH/Uni, sondern nach Abschlussniveau (Bachelor; Master/Diplom; Promotion) (fs)	<u>Bei Ersterfassung:</u> <i>PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601</i> (PENDDAT) <u>Bei wiederholter Erfassung:</u> <i>beruf2</i> aus Vorwelle; <i>PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601</i> (PENDDAT)
<i>brges</i>	<i>akt. Bruttoeinkommen insgesamt (ohne Minijobs, inkl. kat. Ang.), gen.</i> Enthält die über alle ETs (>400 Euro) kumulierte Angabe zum Bruttoeinkommen. Generiert aus offenen Angaben zum Bruttoeinkommen und kategorialer Nachfrage bei wn/ka-Angabe bei der offenen Abfrage (neu)	<i>ET2800; ET2900; ET3000; ET3100; ET3200; ET3300</i> (<i>bio_spells</i>)

Tabelle 16: Einfache Generierungen für Welle 5 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 2)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>brutto</i>	<i>Bruttoeinkommen der aktuellen Haupterwerbstätigkeit inkl. kategorisierter Angaben, generiert</i> Generierung einer integrierten Variable aus kategorisierter und offener Angabe zum Bruttoeinkommen (neu)	<i>ET2800; ET2900; ET3000; ET3100; ET3200; ET3300 (bio_spells)</i>
<i>bruttokat</i>	<i>kategorisiertes Bruttoeinkommen der aktuellen Haupterwerbstätigkeit, generiert</i> Zusammenfassung der mehrstufig erhobenen kategorialen Bruttoeinkommensangaben zu einer Variable (neu)	<i>ET2800; ET2900; ET3000; ET3100; ET3200; ET3300 (bio_spells)</i>
<i>ejhrlewt</i>	<i>Endzeitpunkt letzte Erwerbstätigkeit (Jahr)</i> Jahr, in dem der Befragte zuletzt erwerbstätig war. Für die Generierung wurden Informationen aus den Erwerbstätigkeitsepisoden mit den Angaben zur letzten Erwerbstätigkeit kombiniert, wenn der Befragte seit Januar 2009 nicht erwerbstätig war (fs)	<u>Bei Ersterfassung:</u> <i>PET1200b (PENDDAT); ejahr; emonat (bio_spells)</i> <u>Bei wiederholter Erfassung:</u> <i>ejhrlewt aus Vorwelle (PENDDAT); ejahr; emonat (bio_spells)</i>
<i>ekin1517</i>	<i>Steuervariable: eigenes Kind zwischen 15 und 17 Jahre im HH</i> Indikator, dass Befragungsperson ein unspezifisches, leibliches oder Stief-/Adoptivkind im Alter zwischen 15 und 17 Jahre im Haushalt hat (neu)	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)
<i>ekind</i>	<i>Steuervariable: eigenes Kind im HH</i> Indikator, dass Befragungsperson ein unspezifisches, leibliches oder Stief-/Adoptivkind beliebigen Alters im Haushalt hat (neu) In seltenen Haushaltskonstellationen kann es vorkommen, dass eine Person laut <i>ekind</i> im HH lebende Kinder hat, ihre <i>pnr</i> jedoch in den Zeigern <i>zmhh</i> und <i>zvhh</i> des <i>p_registers</i> nicht erscheint. Dies kann z. B. auftreten bei gleichgeschlechtlichen Beziehungen mit Kindern oder, wenn in einem HH neben dem aktuellen Partner auch noch der ehemalige Partner lebt.	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)
<i>ekin614</i>	<i>Steuervariable: eigenes Kind zwischen 6 und 14 Jahre im HH</i> Indikator, dass Befragungsperson ein unspezifisches, leibliches oder Stief-/Adoptivkind im Alter zwischen 6 und 14 Jahre im Haushalt hat (neu)	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)
<i>ekinu15</i>	<i>Steuervariable: eigenes Kind unter 15 Jahre im HH</i> Indikator, dass Befragungsperson ein unspezifisches, leibliches oder Stief-/Adoptivkind im Alter von unter 15 Jahre im Haushalt hat (neu)	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)
<i>ekinu18</i>	<i>Steuervariable: eigenes Kind unter 18 Jahre im HH</i> Indikator, dass Befragungsperson ein unspezifisches, leibliches oder Stief-/Adoptivkind im Alter von unter 18 Jahre im Haushalt hat (neu)	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)

Tabelle 16: Einfache Generierungen für Welle 5 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 3)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>epartner</i>	<i>Steuervariable: Ehepartner oder eingetr. Lebenspartner im HH</i> Indikator, dass Befragungsperson einen Ehepartner oder einen gleichgeschlechtlichen, eingetragenen Lebenspartner im Haushalt hat (neu)	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)
<i>etakt</i>	<i>aktuell erwerbstätig (>400 Euro pro Monat), generiert (ab Welle 2)</i> Zeigt an, dass die ZP zum Zeitpunkt des Personeninterviews der jeweiligen Welle einen andauernden ET-Spell (d.h. eine ET > 400 Euro) hatte (neu)	<i>zensiert, spinteg, BIO0101 (bio_spells)</i>
<i>famstand</i>	<i>Familienstand, generiert</i> Generierung einer integrierten Familienstandsvariable aus den Angaben im Personenbogen und der aus dem Haushaltsdatensatz generierten Steuerungsvariable <i>epartner</i> (neu)	<i>epartner; PD0500; PD0700 (PENDDAT)</i>
<i>gebhalbj</i>	<i>Geburtshalbjahr, generiert</i> Indikator, ob Geburtstag im 1. oder 2. Halbjahr des Geburtsjahres liegt (neu)	Informationen zum Geburtsmonat
<i>hhalg2</i>	<i>Steuervariable: ALG-II-Bezug aktuell</i> Indikator, für den Arbeitslosengeld-II-Bezug des Haushalts zum Zeitpunkt des Haushaltsinterviews (neu)	<i>HA0250b (HHENDDAT); AL20400; AL20500 (alg2_spells)</i>
<i>kindzges</i>	<i>Anzahl eigene Kinder gesamt (inner- und außerhalb HH), generiert</i> Gesamtzahl der Kinder der Befragungsperson als Summe der Kinder, die im Haushalt leben und der Anzahl der Kinder außerhalb des Haushalts (neu)	Informationen zu Beziehungen im Haushalt (Haushaltsgrid); <i>PD0900; PD1000; PD1100 (PENDDAT)</i>
<i>kindzihh</i>	<i>Anzahl eigene Kinder innerhalb des HH, generiert</i> Auf Basis der Angaben im HH-Bogen generierte Variable für die Zahl der Kinder, die eine Person im HH hat (Summe aus der Zahl der Personen in der HH-(Halb-) Matrix, die als Kinder der Befragungsperson zählen und der Anzahl der Personen in der HH-(Halb-)Matrix, für die angegeben wird, dass es sich bei der HH-Person um einen Eltern- teil handelt) (neu) <i>Anmerkung: Bei Verwendung der Variable sollte deren Personenbezogenheit beachtet werden. Dies bedeutet, dass ein Kind, das mit seinen Eltern in einem Haushalt lebt, sowohl beim Vater als auch bei der Mutter als „Kind im HH“ gezählt wird. Eine Aggregation dieser Variable über die Mitglieder des Haushalts ist also nicht sinnvoll.</i>	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)

Tabelle 16: Einfache Generierungen für Welle 5 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 4)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>mberuf1</i>	<p><i>höchster ber. Abschluss der Mutter, inkl. Mut. im HH, ohne of. Ang., gen.</i></p> <p>Der Berufsabschluss der Mutter wurde in der Welle 1 nur dann erhoben, wenn diese nicht im Haushalt wohnt. Wohnte sie im Haushalt, wurden die Angabe zum Berufsabschluss aus ihrem Personeninterview ergänzt.</p> <p>Ab der Welle 2 wurde die Information zum Berufsabschluss der Mutter bei allen Neubefragten Personen erhoben, unabhängig davon, ob die Mutter im Haushalt lebte oder nicht.</p> <p>Ab der Welle 2 wiederholt Befragte wurden die Werte aus der generierten Variable <i>mberuf1</i> aus der Vorwelle übernommen (uv)</p>	<p><u>Bei Ersterfassung:</u> <i>PSH0300a-i</i> (PENDDAT)</p> <p><u>Nach Ersterfassung:</u> <i>mberuf1</i> aus Vorwelle (PENDDAT)</p>
<i>mberuf2</i>	<p><i>höchster ber. Abschluss der Mutter, inkl. Mut. im HH, mit of. Ang., gen.</i></p> <p>Wie <i>mberuf1</i> mit dem Unterschied, dass bei der Generierung von <i>mberuf2</i> auch die gemachten offenen Angaben berücksichtigt wurden (uv)</p>	<p><u>Bei Ersterfassung:</u> <i>PSH0301a-i</i> (PENDDAT)</p> <p><u>Nach Ersterfassung:</u> <i>mberuf2</i> aus Vorwelle (PENDDAT)</p>
<i>mhh</i>	<p><i>Steuervariable: Mutter lebt im HH</i></p> <p>Indikator, der anzeigt, dass die unspezifische, leibliche, Stief-/Adoptivmutter der Befragungsperson im Haushalt lebt (neu)</p>	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)
<i>migration</i>	<p><i>Migrationshintergrund d. Befragten, generiert</i></p> <p>Generierte Variable für eine vierstufige Unterscheidung des Migrationshintergrunds: ohne Migrationshintergrund; eigene Migration (1. Generation); Migration mind. eines Elternteils, aber nicht selbst migriert (2. Generation); Migration mind. eines Großelternteils, aber nicht selbst migriert und kein Elternteil migriert (3. Generation). (uv)</p> <p><u>Anmerkung:</u> Das Generierungskonzept der Variable wurde ab der Welle 2 überarbeitet. Statt nur die Information, ob die Befragungsperson in Deutschland geboren ist, und die Angaben, welche Vorfahren nach Deutschland zugezogen sind, zur Generierung zu verwenden, wird nun auch die Information, ob und ggf. welcher Vorfahr außerhalb von Deutschland geboren ist, einbezogen. Um eine konsistente Logik über die Wellen zu gewährleisten, wurde die Variable für Welle 1 ebenfalls neu generiert.</p>	<p><u>Bei Ersterfassung:</u> <i>PMI0100; PMI0700; PMI0800a-f; PMI0900a-f</i> (PENDDAT)</p> <p><u>Nach Ersterfassung:</u> <i>migration</i> aus Vorwelle (PENDDAT)</p>

Tabelle 16: Einfache Generierungen für Welle 5 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 5)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>mschul2</i>	<i>höchster allg. Schulabschluss d. Mutter, inkl. Mut. im HH, inkl. of. Ang., gen.</i> Wie <i>mschul1</i> mit dem Unterschied, dass bei der Generierung von <i>mberuf2</i> auch die gemachten offenen Angaben berücksichtigt wurden (uv)	<u>Bei Ersterfassung:</u> <i>PSH0201 (PENDDAT)</i> <u>Nach Ersterfassung:</u> <i>mschul2</i> aus Vorwelle (PENDDAT)
<i>mschul1</i>	<i>höchster allg. Schulabschluss d. Mutter, inkl. Mutter im HH, o. offene Ang., gen.</i> Der höchste Schulabschluss der Mutter wurde in der Welle 1 nur dann erhoben, wenn diese nicht im Haushalt wohnt. Wohnte sie im Haushalt, wurden die Angabe zum Schulabschluss aus ihrem Personeninterview ergänzt (uv) Ab der Welle 2 wurde die Information zum höchsten Schulabschluss der Mutter bei allen Neubefragten Personen erhoben, unabhängig davon ob die Mutter im Haushalt lebte oder nicht.	<u>Bei Ersterfassung:</u> <i>PSH0200 (PENDDAT)</i> <u>Nach Ersterfassung:</u> <i>mschul1</i> aus Vorwelle (PENDDAT)
<i>mstib</i>	<i>Stellung im Beruf der Mutter, Kennziffer, generiert</i> Detaillierte Stellung der Mutter im Beruf, generiert aus den Einzelvariablen (uv)	<u>Bei Ersterfassung:</u> <i>PSH0320; PSH0330; PSH0340; PSH0360; PSH0370; PSH0380 (PENDDAT)</i> <u>Nach Ersterfassung:</u> <i>mstib (PENDDAT)</i>
<i>netges</i>	<i>akt. Nettoeinkommen insgesamt (ohne Minijobs, inkl. kat. Ang.), gen.</i> Enthält die über alle ETs (>400 Euro) kumulierte Angabe zum Nettoeinkommen. Generiert aus offenen Angaben zum Nettoeinkommen und ab Welle 2 kategorialer Nachfrage bei wn/ka-Angabe bei der offenen Abfrage (neu)	<i>ET3400; ET3500; ET3600; ET3700; ET3800; ET3900 (bio_spells)</i>
<i>netto</i>	<i>Nettoeinkommen der aktuellen Haupterwerbstätigkeit inkl. kategorisierter Angaben, generiert</i> Generierung einer integrierten Variable aus kategorisierter und offener Angabe zum Nettoeinkommen (neu)	<i>ET3400; ET3500; ET3600; ET3700; ET3800; ET3900 (bio_spells)</i>
<i>nettokat</i>	<i>kategorisiertes Nettoeinkommen der aktuellen Haupterwerbstätigkeit, generiert</i> Zusammenfassung der mehrstufig erhobenen kategorialen Nettoeinkommensangaben zu einer Variable (neu)	<i>ET3400; ET3500; ET3600; ET3700; ET3800; ET3900 (bio_spells)</i>
<i>palter</i>	<i>Alter (aus PD010), generiert</i> Alter der Befragungsperson, generiert auf Basis der Angaben zum Geburtsdatum und des Datums des Personeninterviews in der aktuellen Welle (neu)	<i>PD0100; pintjahr, pintmon, pinttag (PENDDAT)</i>
<i>panel</i>	<i>Panelbereitschaft</i> (neu)	Vom Erhebungsinstitut gelieferte Information zur Panelbereitschaft der Haushalte

Tabelle 16: Einfache Generierungen für Welle 5 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 6)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>pintdat</i>	<i>Interviewdatum Personeninterview</i> Generierte Variable mit dem Datum, an dem das Personeninterview geführt wurde im Format JJMMTT (neu)	<i>pintjahr, pintmon, pinttag</i> (PENDDAT)
<i>schul1</i>	<i>höchster Schulabschluss, ohne ausl. Abschl. und offene Angaben</i> Variable für den höchsten Schulabschluss; äquivalente west- und ostdeutsche Abschlüsse wurden zusammengefasst (z. B. EOS und Abitur); ohne offene Angaben (fs)	<u>Bei Ersterfassung:</u> <i>PB0200; PB0220; PB0230; PB0300; PB0400</i> (PENDDAT) <u>Bei wiederholter Erfassung:</u> <i>schul1</i> aus Vorwelle; <i>PB0200; PB0220; PB0230; PB0300; PB0400</i> (PENDDAT)
<i>schul2</i>	<i>höchster Schulabschluss, mit ausl. Abschl. und offenen Angaben</i> Wie <i>schul1</i> mit folgenden Unterschieden: 1. Einbeziehen von offenen Angaben; 2. Einbeziehen von Angaben zu ausländischen Abschlüssen (fs)	<u>Bei Ersterfassung:</u> <i>PB0200; PB0220; PB0231; PB0300; PB0401</i> (PENDDAT) <u>Bei wiederholter Erfassung:</u> <i>schul2</i> aus Vorwelle; <i>PB0200; PB0220; PB0231; PB0300; PB0401</i> (PENDDAT)
<i>schulabj</i>	<i>Jahr des höchsten Schulabschlusses</i> Jahr, in dem die Befragungsperson ihren höchsten Schulabschluss erworben hat (fs) <u>Anmerkung:</u> Wiederholt Befragte, für die bereits Informationen zum Schulabschluss aus einer Vorwelle vorlagen, wurden in der aktuellen Welle nicht zum Abschlussjahr gefragt, wenn sie seit der Vorwelle einen neuen Abschluss erworben haben. In diesem Fall wurde das Abschlussjahr in Abhängigkeit des Interviewjahrs und -monats geschätzt. Wurde das Interview in der Welle 5 vor Mai 2011 geführt, wurde davon ausgegangen, dass der Abschluss in 2010 erworben wurde, bei später geführten Interviews wurde 2011 als Abschlussjahr angenommen	<u>Bei Ersterfassung:</u> <i>PB0220; PB0230; PB0410; pintjahr; pintmon</i> (PENDDAT) <u>Bei wiederholter Erfassung:</u> <i>schulabj</i> aus Vorwelle; <i>PB0220; PB0230; PB0410; pintjahr; pintmon</i> (PENDDAT)
<i>statakt</i>	<i>aktueller Hauptstatus, generiert (ab Welle 2)</i> Zeigt an, welchen Hauptstatus die ZP zum Zeitpunkt des Personeninterviews der jeweiligen Welle hatte (neu)	<i>zensiert; spinteg; BIO0101; az2ges (bio_spells)</i>
<i>stib</i>	<i>Stellung im Beruf, Kennziffer, generiert</i> Generierung der detaillierten Kennziffer der Stellung im Beruf aus den Einzelvariablen. Generierung der Variable mit Informationen aus dem Erwerbstätigkeitsmodul (<i>ET0603-ET1203</i>). Bei mehreren aktuell andauernden Erwerbstätigkeitsepisoden wurde die Episode mit dem höheren Stundenumfang gewählt. Bei mehreren aktuell andauernden Episoden mit exakt gleichem Stundenumfang, diejenige die eher begonnen hat	<i>ET0603; ET0703; ET0803; ET0903; ET1003; ET1103; ET1203 (bio_spells)</i>

(neu)

Tabelle 16: Einfache Generierungen für Welle 5 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 7)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>stibewt</i>	<i>Stellung im Beruf, erste Erwerbstätigkeit, Kennziffer, generiert</i> Detaillierte Kennziffer der Stellung im Beruf, in dem der Befragte erstmals hauptberuflich tätig war. Für die Generierung wurden Informationen zur ersten Erwerbstätigkeit mit Informationen aus den Erwerbstätigkeitsepisoden kombiniert, wenn der Befragte angab, die erste Erwerbstätigkeit bereits bei der Erhebung der Erwerbstätigkeitsepisoden seit Januar 2009 genannt zu haben (uv)	<u>Bei Ersterfassung:</u> PET3300; PET3400; PET3500; PET3600; PET3700; PET3800; PET3900 (PENDDAT); ET0603; ET0703; ET0803; ET0903; ET1003; ET1103; ET1203 (bio_spells) <u>Nach Ersterfassung:</u> stibewt aus Vorwelle (PENDDAT)
<i>stibkz</i>	<i>jetzige Stellung im Beruf, einf. Eint., harmonisiert (anonymisiert)</i> Generierung der einfachen Kennziffer der Stellung im Beruf aus den Einzelvariablen (neu)	PET1510 (PENDDAT)
<i>stiblewt</i>	<i>Stellung im Beruf, letzte Erwerbstätigkeit, Kennziffer, generiert</i> Detaillierte Kennziffer der Stellung im Beruf, in dem der Befragte zuletzt erwerbstätig war. Für die Generierung wurden Informationen aus den Erwerbstätigkeitsepisoden mit den Angaben zur letzten Erwerbstätigkeit kombiniert, wenn der Befragte seit Januar 2009 nicht erwerbstätig war (fs)	<u>Bei Ersterfassung:</u> PET1210; PET1220; PET1230; PET1240; PET1250; PET1260; PET1270 (PENDDAT); ET0603; ET0703; ET0803; ET0903; ET1003; ET1103; ET1203 (bio_spells) <u>Bei wiederholter Erfassung:</u> stiblewt aus Vorwelle (PENDDAT); ET0603; ET0703; ET0803; ET0903; ET1003; ET1103; ET1203 (bio_spells)
<i>vberuf1</i>	<i>höchster ber. Abschluss des Vaters, inkl. Vater im HH, ohne of. Ang., gen.</i> Generierung für den höchsten Berufsabschluss des Vaters analog zu <i>mberuf1</i> (uv)	<u>Bei Ersterfassung:</u> PSH0600a-i (PENDDAT) <u>Nach Ersterfassung:</u> vberuf1 aus Vorwelle (PENDDAT)
<i>vberuf2</i>	<i>höchster ber. Abschluss des Vaters, inkl. Vater im HH, mit of. Ang., gen.</i> Generierung für den höchsten Berufsabschluss des Vaters (unter Berücksichtigung der offenen Angaben) analog zu <i>mberuf2</i> (uv)	<u>Bei Ersterfassung:</u> PSH0601a-i (PENDDAT) <u>Nach Ersterfassung:</u> vberuf2 aus Vorwelle (PENDDAT)
<i>vhh</i>	<i>Steuervariable: Vater lebt im HH</i> Indikator, der anzeigt, dass der unspezifische, leibliche, Stief-/Adoptivvater der Befragungsperson im Haushalt lebt (neu)	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)
<i>vschul1</i>	<i>höchster allg. Schulabschluss d. Vaters, inkl. Vater im HH, o. of Ang., gen</i> Generierung für den höchsten Schulabschluss des Vaters analog zu <i>mschul1</i> (uv)	<u>Bei Ersterfassung:</u> PSH0500 (PENDDAT) <u>Nach Ersterfassung:</u> vschul1 aus Vorwelle (PENDDAT)

Tabelle 16: Einfache Generierungen für Welle 5 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 8)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>vschul2</i>	<i>höchster allg. Schulabschluss d. Vater, inkl. Vater im HH, inkl. of. Ang., gen</i> Generierung für den höchsten Schulabschluss des Vaters (unter Berücksichtigung der offenen Angaben) analog zu <i>mschul2</i> (uv)	<u>Bei Ersterfassung:</u> <i>PSH0501</i> (PENDDAT) <u>Nach Ersterfassung:</u> <i>vschu2</i> aus Vorwelle (PENDDAT)
<i>vstib</i>	<i>Stellung im Beruf des Vaters, Kennziffer, generiert</i> Detaillierte Stellung im Beruf des Vaters, generiert aus den Einzelvariablen (uv)	<u>Bei Ersterfassung:</u> <i>PSH0620; PSH0630; PSH0640; PSH0660; PSH0670; PSH0680</i> (PENDDAT) <u>Nach Ersterfassung:</u> <i>vstib</i> aus Vorwelle (PENDDAT)

Tabelle 17: Einfache Generierungen für Welle 5 im Arbeitslosengeld-II-Spellldatensatz (*alg2_spells*) (Reihenfolge analog zur Reihenfolge im Datensatz)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>bmonat</i>	<i>ALG-II-Episode: Anfangsmonat, generiert</i> Beginnmonat der ALG-II-Episode, in der für die Generierung Jahreszeitangaben in konkrete Monatsangaben umgerechnet wurden. <u>Anmerkung:</u> Die generierten Datumsvariablen wurden plausibilitätsgeprüft und ggf. bereinigt. Die originalen Datumsangaben des Befragten sind ab der Welle 2 in den Quellvariablen erhalten. Die Jahreszeitangaben wurden nach folgenden Regeln in Monatswerte recodiert 21 Jahresanfang/Winter → Januar 24 Frühjahr/Ostern → April 27 Jahresmitte/Sommer → Juli 30 Herbst → Oktober 32 Jahresende → Dezember	<i>AL20100</i> (<i>alg2_spells</i>)
<i>bjahr</i>	<i>ALG-II-Episode: Anfangsjahr, generiert</i> Endjahr der Arbeitslosengeld-II-Episode.	<i>AL20200</i> (<i>alg2_spells</i>)
<i>emonat</i>	<u>Anmerkung:</u> siehe <i>bmonat</i> <i>ALG-II-Episode: Endmonat, generiert</i> Endmonat der Arbeitslosengeld-II-Episode in der für die Generierung Jahreszeitangaben in konkrete Monatsangaben umgerechnet wurden und bei rechtszensierten (d.h. zum Befragungszeitpunkt des Haushalts aktuell andauernden) Spells der Interviewmonat eingesetzt wurde.	<i>AL20300</i> (<i>alg2_spells</i>); <i>hintmon</i> (HHENDDAT)
<i>ejahr</i>	<u>Anmerkung:</u> siehe <i>bmonat</i> <i>ALG-II-Episode: Endjahr, generiert</i> Endjahr der Arbeitslosengeld-II-Episode, in der bei rechtszensierten (d.h. zum Befragungszeitpunkt des Haushalts aktuell andauernden) Spells das Interviewjahr eingesetzt wurde. <u>Anmerkung:</u> siehe <i>bmonat</i>	<i>AL20400</i> (<i>alg2_spells</i>); <i>hintjahr</i> (HHENDDAT)

Tabelle 17: Einfache Generierungen für Welle 5 im Arbeitslosengeld-II-Spelldatensatz (alg2_spells) (Reihenfolge analog zur Reihenfolge im Datensatz) (Forts. 1)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
alg2kbma- alg2kbmh	ALG-II: 1. Kürzung: Anfangsmonat, generiert Beginnmonat der Arbeitslosengeld-II-Kürzung, in der für die Generierung die Jahreszeitangaben in konkrete Monatsangaben umgerechnet wurden. <u>Anmerkung:</u> Die Arbeitslosengeld-II-Kürzungen sind in die Arbeitslosengeld-II-Bezugsspellings eingebettet. Über den Indikator am Ende der jeweiligen Variablen (a-h) können die Informationen zu den einzelnen Kürzungsepisoden unterschieden werden. Die generierten Datumsvariablen wurden plausibilitätsgeprüft und ggf. bereinigt. Die originalen Datumsangaben des Befragten sind ab der Welle 2 in den Quellvariablen enthalten.	1. Kürzung: AL21000a (alg2_spells) bis 8. Kürzung: AL21000h (alg2_spells)
alg2kbja- alg2kbjh	ALG II: 1. Kürzung: Anfangsjahr, generiert Beginnjahr der Arbeitslosengeld-II-Kürzung. <u>Anmerkung:</u> siehe alg2kma - alg2kbf	1. Kürzung: AL21100a (alg2_spells) bis 8. Kürzung: AL21100h (alg2_spells)
alg2kema- alg2kemh	ALG II: 1. Kürzung: Endmonat, generiert Endmonat der Arbeitslosengeld-II-Kürzung, in der für die Generierung die Jahreszeitangaben in konkrete Monatsangaben umgerechnet wurden. Wurde vom Befragten eine Dauer der Kürzung berichtet, wurde aus dieser auf Basis des generierten Beginndatums der Kürzung das jeweilige Enddatum berechnet. <u>Anmerkung:</u> siehe alg2kma - alg2kbf	1. Kürzung: alg2kbma; alg2kbja; AL21200a; AL21201a; AL21202a (alg2_spells) bis 8. Kürzung: alg2kbmh; alg2kbjh; AL21200h; AL21201h; AL21202h (alg2_spells)
alg2keja - alg2kejf	ALG II: 1. Kürzung: Endjahr, generiert Endjahr der Arbeitslosengeld-II-Kürzung. Wurde vom Befragten eine Dauer der Kürzung berichtet, wurde aus dieser auf Basis des generierten Beginndatums der Kürzung das jeweilige Enddatum berechnet. <u>Anmerkung:</u> siehe alg2kma - alg2kbf	1. Kürzung: alg2kbma; alg2kbja; AL21200a; AL21201a; AL21202a (alg2_spells) bis 8. Kürzung: alg2kbmh; alg2kbjh; AL21200f; AL21201f; AL21202f (alg2_spells)

Tabelle 17: Einfache Generierungen für Welle 5 im Arbeitslosengeld-II-Spelldatensatz (alg2_spells) (Reihenfolge analog zur Reihenfolge im Datensatz) (Forts. 2)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
AL22150a bis AL22150h	<p><i>ALG II: Kürzung: Welchem HH-Mitglied wurde gekürzt, generiert</i></p> <p>In dieser Variable wurde codiert, für welche Haushaltsmitglieder eine Kürzung des Arbeitslosengelds II stattgefunden hat. Es handelt sich um eine Stringvariable mit 15 Stellen. Von links beginnend, steht dabei jede Stelle dieser Variable für die Position einer Person im Haushaltsgrid. Die erste Stelle der Variable zeigt so bspw. an, ob in der jeweiligen Kürzungsepisode für die erste Person im Haushalt das Arbeitslosengeld II gekürzt wurde, die zweite Stelle, ob für die zweite Person gekürzt wurde usw. Da die Quellinformationen für die Generierung nur von der Welle 2 bis zur Welle 4 erhoben wurden, ist die Frage in allen in der ersten Welle und ab der fünften Welle berichteten Kürzungen an allen 15 Stellen mit „I“ (Item in Welle nicht erhoben) gefüllt (s. u.). Jede der 15 Stellen der Variable, die für eine der maximal 15 Personen in der Haushaltsstruktur steht, hat einen der folgenden Codes, der den Kürzungsstatus der jeweiligen Person angibt.</p> <p><u>Codes:</u></p> <p>1 - dem HH-Mitglied wurde das ALG II gekürzt 2 - dem HH-Mitglied wurde das ALG II nicht gekürzt W - weiß nicht K - keine Angabe T - trifft nicht zu (Filter) F - Frage irrtümlich nicht gestellt U - unplausibler Wert I - Item in Welle nicht erhoben</p>	Information, für welches Haushaltsmitglied in der jeweiligen Kürzungsepisode gekürzt wurde (nur bis Welle 4 erhoben)
zensiert	<p><i>ALG-II-Episode: Spell dauert z. Zeitpkt. d. letzten HH-Int. an (Rechtszens.), gen.</i></p> <p>Der Zensierungsindikator gibt an, ob ein Spell zum Zeitpunkt des letzten Haushaltsinterviews noch andauerte.</p> <p><u>Anmerkung:</u> Ein Spell gilt als zensiert, wenn eine der folgenden Bedingungen erfüllt ist:</p> <p>(a) Es handelt sich um einen in den Vorwellen zensierten Spell eines Haushalts, der in den Folgewellen bis hin zur aktuellen Welle nicht wiederbefragt wurde.</p> <p>(b) Ein in Welle 4 befragter Haushalt gibt an, dass ein ALG-II-Spell auch zum Interviewdatum der Welle 5 noch andauert. Oder ein Enddatum wird genannt, das identisch mit dem Interviewdatum der Welle 5 ist und in der Nachfrage wird bestätigt, dass der Bezug aktuell noch andauert.</p> <p>Der Code-5 wurde vergeben, wenn die Haushaltsauskunftsperson der Vorwelle in Welle 5 nicht mehr im HH lebt u.in Welle 5 nicht befragt wurde</p>	AL20300; AL20400, AL20500 (alg2_spells)

Tabelle 18: Einfache Generierungen für Welle 5 im BIO-Spell Datensatz (bio_spells) (Reihenfolge analog zur Reihenfolge im Datensatz)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>bmonat</i>	<p><i>Erwerbstätigkeit: Anfangsmonat, generiert</i> Beginnmonat der Erwerbstätigkeitsepisode, in der für die Generierung Jahreszeitangaben in konkrete Monatsangaben umgerechnet wurden.</p> <p><u>Anmerkung:</u> Die generierten Datumsvariablen wurden plausibilitätsgeprüft und ggf. bereinigt. Die originalen Datumsangaben des Befragten sind in den Quellvariablen erhalten. Die Jahreszeitangaben wurden nach folgenden Regeln in Monatswerte recodiert: 21 Jahresanfang/Winter → Januar 24 Frühjahr/Ostern → April 27 Jahresmitte/Sommer → Juli 30 Herbst → Oktober 32 Jahresende → Dezember</p>	<i>BIO0200 (bio_spells)</i>
<i>bjahr</i>	<p><i>Erwerbstätigkeit: Anfangsjahr, generiert</i> Beginnjahr der Erwerbstätigkeitsepisode.</p>	<i>BIO0300 (bio_spells)</i>
<i>emonat</i>	<p><u>Anmerkung:</u> siehe <i>bmonat</i> <i>Erwerbstätigkeit: Endmonat, generiert</i> Endmonat der Erwerbstätigkeitsepisode, in der für die Generierung Jahreszeitangaben in konkrete Monatsangaben umgerechnet wurden und bei rechtszensierten (d.h. zum Befragungszeitpunkt der Person aktuell andauernden) Spells der Interviewmonat eingesetzt wurde.</p>	<i>BIO0400, BIO0600 (bio_spells); pintmon (PENDDAT)</i>
<i>ejahr</i>	<p><u>Anmerkung:</u> siehe <i>bmonat</i> <i>Erwerbstätigkeit: Endjahr, generiert</i> Endjahr der Erwerbstätigkeitsepisode, in der bei rechtszensierten (d.h. zum Befragungszeitpunkt des Haushalts aktuell andauernden) Spells das Interviewjahr eingesetzt wurde.</p>	<i>BIO0500, BIO0600 (bio_spells); pintjahr (PENDDAT)</i>
<i>zensiert</i>	<p><u>Anmerkung:</u> siehe <i>bmonat</i> <i>Erwerbstätigkeit: ggf. noch andauernder Spell (Rechtszensierung)</i> Der Zensierungsindikator gibt an, ob ein Spell zum Zeitpunkt des Personeninterviews in der letzten Befragungswelle noch andauerte, d.h. es handelt sich um einen rechtszensierten Spell.</p> <p><u>Anmerkung:</u> Ein Spell gilt als zensiert, wenn eine der beiden folgenden Bedingungen erfüllt ist: Die Person gibt bezüglich Enddatum des BIO-Spells an, zum Zeitpunkt des Interviews die Tätigkeit noch auszuüben. Oder es wird ein Enddatum für die Tätigkeit genannt, welches identisch mit dem Interviewdatum ist und in der Nachfrage bestätigt, dass die Aktivität derzeit noch andauert.</p>	<i>BIO0400; BIO0500; BIO0600 (bio_spells)</i>

Tabelle 18: Einfache Generierungen für Welle 5 im BIO-Spellldatensatz (bio_spells) (Reihenfolge analog zur Reihenfolge im Datensatz) (Fortsetzung 1)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>stib</i>	<i>Stellung im Beruf, Kennziffer, generiert</i> Generierung der detaillierten Kennziffer der Stellung im Beruf aus den Einzelvariablen.	<u>Erfassung von</u> <u>Spellinformationen in Welle 5</u> ET0603; ET0703; ET0803; ET0903; ET1003; ET1103; ET1203 (bio_spells)
<i>az1</i>	<i>wöchentliche vertragliche Arbeitszeit</i>	Ansonsten bleibt der Wert der Vorwelle bestehen <u>Erfassung von</u> <u>Spellinformationen in Welle 5</u> ET2003 (bio_spells)
<i>az2</i>	<i>wöchentliche Arbeitszeit inkl. Angaben bei unregelmäßiger Arbeitszeit, gen.</i> Integrierte Variable zur wöchentlichen Arbeitszeit der Erwerbstätigkeit, zusammengefasst aus der offenen Angabe zur Arbeitszeit und der kategorialen Nachfrage. Für die geschlossenen Kategorien der Nachfrage wurden die Kategoriemittelwerte eingesetzt, für die offene Kategorie (40h und mehr) der Median der offenen angegebenen Wochenstunden.	Ansonsten bleibt der Wert der Vorwelle bestehen <u>Erfassung von</u> <u>Spellinformationen in Welle 5</u> ET2103; ET2203 (bio_spells) Ansonsten bleibt der Wert der Vorwelle bestehen
<i>alg1bm</i>	<i>ALG1-Bezug: Anfangsmonat, generiert</i> Beginnmonat des Arbeitslosengeld-I-Bezugs, in der für die Generierung die Jahreszeitangaben in konkrete Monatsangaben umgerechnet wurden. <u>Anmerkung:</u> Die Arbeitslosengeld-I-Bezüge sind in die Arbeitslosmeldungsspells eingebettet. Pro Arbeitslosmeldung ist maximal ein Arbeitslosengeld-I-Bezugszeitraum vorgesehen. Die generierten Datumsvariablen wurden plausibilitätsgeprüft und ggf. bereinigt. Die originalen Datumsangaben des Befragten sind in den Quellvariablen enthalten.	AL0800 (bio_spells)
<i>alg1bj</i>	Umrechnung der Monatsangaben siehe <i>bmonat</i> . <i>ALG1-Bezug: Anfangsjahr, generiert</i> Beginnjahr des Arbeitslosengeld-I-Bezugs. <u>Anmerkung:</u> siehe <i>alg1bm</i>	AL0900 (bio_spells)

Tabelle 18: Einfache Generierungen für Welle 5 im BIO-Spell Datensatz (bio_spells) (Reihenfolge analog zur Reihenfolge im Datensatz) (Fortsetzung 2)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>alg1em</i>	<p><i>ALG1-Bezug: Endmonat, generiert</i></p> <p>Endmonat des Arbeitslosengeld-I-Bezugs, in der für die Generierung die Jahreszeitangaben in konkrete Monatsangaben umgerechnet wurden und bei rechtszensierten (d.h. zum Befragungszeitpunkt der Person aktuell andauernden) Spells das Interviewdatum eingesetzt wurde.</p>	<p><i>AL1000; AL1200 (bio_spells); pintmon (PENDDAT)</i></p>
<i>alg1ej</i>	<p><u>Anmerkung:</u> siehe <i>alg2kma - alg2kbme</i></p> <p><i>ALG1-Bezug: Endjahr, generiert</i></p> <p>Endjahr des Arbeitslosengeld-I-Bezugs, in der bei rechtszensierten (d.h. zum Befragungszeitpunkt der Person aktuell andauernden) Spells das Interviewdatum eingesetzt wurde.</p>	<p><i>AL1100; AL1200 (bio_spells); pintjahr (PENDDAT)</i></p>
<i>alg1akt</i>	<p><u>Anmerkung:</u> siehe <i>alg2kma - alg2kbme</i></p> <p><i>ALG1-Bezug: ggw. noch andauernder Spell (Rechtszensierung)</i></p> <p>Der Zensierungsindikator gibt an, ob der ALG1-Bezug zum Zeitpunkt des Personeninterviews in der letzten Welle noch andauerte, d.h. es sich um einen rechtszensierten Spell handelt.</p> <p><u>Anmerkung:</u> Ein Spell gilt als zensiert, wenn eine der beiden folgenden Bedingungen erfüllt ist: Die Person gibt bezüglich Enddatum des Arbeitslosengeld-I-Bezug-Spells an, zum Zeitpunkt des Interviews noch Arbeitslosengeld I zu erhalten. Oder es wird ein Enddatum genannt, welches identisch mit dem Interviewdatum ist und in der Nachfrage bestätigt, dass der Bezug derzeit noch andauert. Die Generierung basiert auf den plausibilitätsgeprüften, generierten Datumsvariablen.</p>	<p><i>emonat; ejahr; AL1000; AL1100; AL1200 (bio_spells)</i></p>
<i>br</i>	<p><i>Bruttoeinkommen (inkl. kat. Ang.), generiert</i></p>	<p><i>ET2800; ET2900; ET3000; ET3100; ET3200; ET3300 (bio_spells)</i></p>
<i>net</i>	<p><i>Nettoeinkommen (inkl. kat. Ang.), generiert</i></p>	<p><i>ET3400; ET3500; ET3600; ET3700; ET3800; ET3900 (bio_spells)</i></p>

Tabelle 19: Einfache Generierungen für Welle 5 im Ein-Euro-Spell Datensatz (ee_spells) (Reihenfolge analog zur Reihenfolge im Datensatz)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>bmonat</i>	<p><i>Maßnahme: Anfangsmonat, generiert</i></p> <p>Beginnmonat der Maßnahmeepisode, in der für die Generierung Jahreszeitangaben in konkrete Monatsangaben umgerechnet wurden.</p> <p><u>Anmerkung:</u> Die generierten Datumsvariablen wurden plausibilitätsgeprüft und ggf. bereinigt. Die originalen Datumsangaben (mit Ausnahme der in den Wertebereichsprüfungen als unplausibel identifizierten Werte) des Befragten sind in den Quellvariablen erhalten.</p> <p>Die Jahreszeitangaben wurden nach folgenden Regeln in Monatswerte recodiert: 21 Jahresanfang/Winter → Januar 24 Frühjahr/Ostern → April 27 Jahresmitte/Sommer → Juli 30 Herbst → Oktober 32 Jahresende → Dezember</p>	<i>EE0600a (ee_spells)</i>
<i>bjahr</i>	<p><i>Maßnahme: Anfangsjahr, generiert</i></p> <p>Beginnjahr der Maßnahmeepisode.</p>	<i>EE0600b (ee_spells)</i>
<i>emonat</i>	<p><u>Anmerkung:</u> siehe <i>bmonat</i></p> <p><i>Maßnahme: Endmonat, generiert</i></p> <p>Endmonat der Maßnahmeepisode, in der für die Generierung Jahreszeitangaben in konkrete Monatsangaben umgerechnet wurden und bei rechtszensierten (d.h. zum Befragungszeitpunkt der Person aktuell andauernden) Spells das Interviewdatum eingesetzt wurde.</p>	<i>EE0600a; EE0600b; EE0700; EE0800a; EE0800b (ee_spells); pintmon, pintjahr (PENDDAT)</i>
<i>ejahr</i>	<p><u>Anmerkung:</u> siehe <i>bmonat</i></p> <p><i>Maßnahme: Endjahr, generiert</i></p> <p>Endjahr der Maßnahmeepisode, in der bei rechtszensierten (d.h. zum Befragungszeitpunkt der Person aktuell andauernden) Spells das Interviewdatum eingesetzt wurde.</p>	<i>EE0600a; EE0600b; EE0700; EE0800a; EE0800b (ee_spells); pintjahr; pintjahr (PENDDAT)</i>
<i>zensiert</i>	<p><u>Anmerkung:</u> siehe <i>bmonat</i></p> <p><i>Maßnahme: ggf. noch andauernder Spell (Rechtszensierung)</i></p> <p>Der Zensierungsindikator gibt an, ob ein Spell zum Zeitpunkt des Personeninterviews in der letzten Befragungswelle noch andauerte, d.h. es handelt sich um einen rechtszensierten Spell.</p>	<i>EE0700 (ee_spells)</i>

Tabelle 20: Einfache Generierungen für Welle 5 im Personenregisterdatensatz (p_register) (alphabetische Reihenfolge)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
<i>alter5</i>	<p><i>Alter der Person in Welle 5 (2011)</i></p> <p>Variable enthält die „beste“ verfügbare Altersinformation einer Person. Dabei handelt es sich entweder (a) um das aus dem Geburtsdatum, das in Welle 5 berichtet wurde, berechnete Alter oder (b), wenn kein Geburtsdatum aus der Welle 5 vorliegt, um die Altersangabe aus dem Haushaltsinterview. Die Information aus <i>alter5</i> wurde auch in den Haushaltsdatensatz übernommen und entspricht den Angaben in <i>HD0200a</i> bis <i>HD0200o</i>. Dieses Vorgehen ist konsistent mit dem Vorgehen im Feld. Bereits während der Feldzeit wurde die Altersvariable in der Datenbank mit den jeweils „besten“ Informationen gefüllt. Dort wird eine Variable in der Datenbank zunächst mit der Altersangabe laut Haushaltsinterview belegt. Wird ein Personeninterview geführt, wird diese Variable in der Datenbank mit dem Alter überschrieben, das auf Basis der Angaben des Personeninterviews berechnet wurde (Geburtsdatum, Personeninterviewdatum). Sowohl die im Haushaltsdatensatz als auch die im Personendatensatz ausgelieferten Altersangaben basieren auf dieser Variable der Datenbank. Die für Welle 5 im Haushaltsdatensatz enthaltene „beste“ Altersinformation wurde in den Plausibilitätsprüfungen und bei der Generierung der BG- und Haushaltstypen berücksichtigt.</p>	<p><i>PD0100; pintjahr; pintmon; pinttag (PENDDAT); HD0200a bis HD0200o (HHENDDAT)</i></p>
<i>erwprox5</i>	<p><i>Erwerbsstatus laut HH-Interview in Welle 5 (2011)</i></p> <p>Variable ist eine unveränderte Übernahme der <i>HD1101*</i> der aktuellen Welle aus <i>HHENDDAT</i>.</p>	<p><i>HD1101*</i></p>
<i>korrsex</i>	<p><i>Angabe zum Geschlecht wurde zwischen Befragungswellen korrigiert</i></p> <p>Variable gibt für Personen, die in mehr als einer Welle Mitglied eines Befragungshaushalts waren an, ob im Rahmen des Haushaltsinterviews das Geschlecht korrigiert wurde.</p>	<p><i>HD0100a bis HD0100o aller Wellen (HHENDDAT)</i></p>
<i>lastint</i>	<p><i>Befragungswelle des letzten Interviews auf Personenebene</i></p> <p>Variable gibt an, in welcher Welle das letzte Interview auf Personenebene (Personeninterview oder Senioreninterview) mit der Person geführt wurde.</p>	<p>Personeninterviews aller Wellen (<i>PENDDAT</i>)</p>
<i>neuj5</i>	<p><i>Jahr, seit dem P. neu im aktuellen HH ist, berichtet in Welle 5 (2011)</i></p> <p>Variable zeigt an, seit welchem Jahr die Person neu zum dem Haushalt gehört, dessen Haushaltsmitglied sie in der Welle 5 ist.</p> <p><u>Anmerkung:</u> Information zum Datum stammt aus dem Welle 5-Interview mit dem wiederbefragten Haushalt, in den die Person seit der Vorwelle zugezogen ist bzw. in dem sie geboren wurde.</p>	<p>Informationen zum Datum, seit dem eine Person neu im Haushalt ist. Erhoben im Haushaltsgrid</p>

Tabelle 20: Einfache Generierungen für Welle 5 im Personenregisterdatensatz (p_register) (alphabetische Reihenfolge) (Fortsetzung 1)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
neum5	<p>Monat, seit dem P. neu im aktuellen HH ist, berichtet in Welle 5 (2011)</p> <p>Variable zeigt an, seit welchem Monat die Person neu zum dem Haushalt gehört, dessen Haushaltsmitglied sie in der Welle 5 ist.</p>	Informationen zum Datum, seit dem eine Person neu im Haushalt ist. Erhoben im Haushaltsgrid
wegj5	<p><u>Anmerkung:</u> siehe <i>neuj5</i></p> <p>Jahr, seit dem P. nicht mehr im vorherigen HH ist, berichtet in Welle 5 (2011)</p> <p>Variable zeigt an, seit welchem Jahr die Person nicht mehr zum Haushalt der Vorwelle gehört.</p>	Informationen zum Datum, seit dem eine Person nicht mehr im Haushalt ist. Erhoben im Haushaltsgrid
wegm5	<p><u>Anmerkung:</u> Information zum Datum stammt aus dem Welle 5-Interview mit dem Haushalt, in dem die Person in der Vorwelle gelebt hat.</p> <p>Monat, seit dem P. nicht mehr im vorherigen HH ist, berichtet in Welle 5 (2011)</p> <p>Variable zeigt an, seit welchem Monat die Person nicht mehr zum Haushalt der Vorwelle gehört.</p>	Informationen zum Datum, seit dem eine Person nicht mehr im Haushalt ist. Erhoben im Haushaltsgrid
zdub5	<p><u>Anmerkung:</u> siehe <i>wegj4</i></p> <p>Zeiger: Personennr. der durch die ZP gedoppelten Person in Welle 5 (2011)</p> <p>Zeigt an, dass eine aus einem Ur-HH bekannte Person aktuell in einem Split-HH dieses Ur-HH lebt, ohne dass der Ur-HH den Auszug dieser Person gemeldet hat.</p>	Informationen zu allen Haushaltsmitgliedern eines Ur-HH und aller seiner Split-HH im Haushaltsgrid der aktuellen und der Vorwelle
zmhh5	<p><u>Anmerkung:</u> Eine detaillierte Erläuterung zu den Gründen der Einführung dieser Variable findet sich in Kapitel 5.4.1.2.</p> <p>Zeiger: Personnummer der Mutter der Zielperson im HH in Welle 5 (2011)</p> <p>Identifiziert die Personnummer der Mutter, sofern sie im Haushalt lebt. Als Mutter gelten: leibliche Mütter, Stiefmütter, Adoptivmütter oder Pflegemütter sowie Mütter mit nicht weiter spezifiziertem Status.</p>	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)
zparth5	<p>Zeiger: Personnummer des Partners der Zielperson im HH in Welle 5 (2011)</p> <p>Identifiziert die Personnummer eines im Haushalt lebenden Partners. Als Partner gelten: Ehepartner, eingetragene Lebenspartner, unverheiratete Partner und Partner mit nicht weiter spezifiziertem Status.</p>	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)
zupanel	<p>Befragungswelle des Zugangs der Person zum Panel</p> <p>Variable gibt an, in welcher Welle die Person erstmals Mitglied eines Befragungshaushaltes war.</p>	Informationen zu den im Haushalt lebenden Personen aller Wellen (Haushaltsgrid)

Tabelle 20: Einfache Generierungen für Welle 5 im Personenregisterdatensatz (p_register) (alphabetische Reihenfolge) (Fortsetzung 2)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 5
zvhh5	<p><i>Zeiger: Personennummer des Vaters der Zielperson im HH in Welle 5 (2011)</i></p> <p>Identifiziert die Personennummer des Vaters, sofern er im Haushalt lebt. Als Vater gelten: leibliche Väter, Stiefväter, Adoptivväter oder Pflegeväter sowie Väter mit nicht weiter spezifiziertem Status.</p>	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)

In den Datensätzen auf Personenebene sind eine Vielzahl generierter Variablen und Konstruktvariablen enthalten. Darunter auch Variablen (z. B. für die Stellung im Beruf), die in mehreren Datensätzen zu finden sind. Abbildung 3 gibt einen Überblick über die auf Personenebene enthaltenen einfachen und komplexen generierten Variablen.

Abbildung 3: Übersicht über generierte Variablen auf Personenebene in Welle 5

	PENDDAT					BIO-Spells Erwerbs- und Arbeitslosigkeitsbiografie	EE_Spells Ein-Euro-Job-Teilnahme
	Aktueller Status	Erwerbshistorie		Soziale Herkunft			
		letzte ET	erste ET	Mutter	Vater		
Bildung	berabj						
	beruf1			mberuf1	vberuf1		
	beruf2			mberuf2	vberuf2		
	schulabj						
	schul1			mschul1	vschul1		
	schul2			mschul2	vschul2		
Bildungsklassifikation	casmin			mcasmin	vcasmin		
	isced97			miscd97	visced97		
	bilzeit			mbilzeit	vbilzeit		
Informationen zum aktuellen Status	akt1euro						
	alakt						
	etakt						
	statakt					spelltyp	
Sozio-ökonomische Position	egp	egplewt	egpeewt	megp	vegp	egp	
	esec	eseclwlt	eseceewt	mesec	vesec	esec	
	isei	iseilewt	iseieewt	misei	visei	isei	
	mpps	mppslewt	mpseewt	mmpps	vmpps	mpps	
	siops	siopslewt	siopseewt	msiops	vsiops	siops	
Stellung im Beruf	stib	stiblewt	stibeewt	mstib	vstib	stib	
	stibkz						
Datierung der Erwerbstätigkeit			begmeewt			bmonat	bmonat
			begjeewt			bjahr	bjahr
		emonlewt				emonat	emonat
		ejhrlewt				ejahr	ejahr
Datierung der Arbeitslosigkeit						alg1bm	
						alg1bj	
						alg1em	
						alg1ej	
Informationen zur Erwerbstätigkeit	befrist						
	azhpt1					az1	
	azhpt2					az2	
	azges1						
Berufliche Tätigkeit	isco88	iscolewt	iscoeewt	misco	visco	isco88	
	klldb	klldblewt	klldbeewt	mkldb	vkldb	klldb	
Tätig in Branche	branche					branche	

Abbildung 3: Übersicht über generierte Variablen auf Personenebene in Welle 5 (Forts.)

	PENDDAT				BIO Spells	EE Spells	
	Aktueller Status	Erwerbshistorie		Soziale Herkunft		Erwerbs- und Arbeitslosigkeitsbiografie	Ein-Euro-Job-Teilnahme
		letzte ET	erste ET	Mutter	Vater		
Einkommen	netges						
	brges						
	netto						
	nettokat						
	brutto						
	bruttokat						
Leistungsbezug	alg1abez				alg1akt		
	hhalg2						
Haushaltskontext und Familienstand	hhgr						
	famstand						
	vhh						
	mhh						
	apartner						
	epartner						
	ekind						
	ekin614						
	ekinu15						
	ekinu18						
	ekin1517						
	kindzges						
	kindzihh						
	Migrationshintergrund	oggebland					
		ostaatan					
ozulanda							
ozulandb							
ozulandc							
ozulandd							
ozulande							
ozulandf							
migration							
Informationen zur Person		gebhalbj					
	palter						
	zpalthh						
	zpsex						
Allgemein	altbefr						
	fb_vers						
	panel						
	pintdat						
	RegP0100						
	sample						
Freizeitverhalten	freiz1						
	freiz2						
	freiz3						
	frwunsch						

4.5 Konstruktvariablen

Unter Konstruktvariablen fallen diejenigen Variablen, für deren Generierungen umfangreichere Umcodierungen und/oder Vercodungsarbeiten notwendig sind. In den meisten Fällen handelt es sich um bereits an anderer Stelle erprobte, theoretisch untermauerte und zumindest teilweise standardisierte sozial- oder wirtschaftswissenschaftliche Konstrukte. Beispiele hierfür sind die European Socio-economic Classification (ESeC), die International Standard Classification of Education (ISCED) oder das bedarfsgewichtete Haushaltseinkommen. Auf den folgenden Seiten werden die im PASS bereitgestellten Konstruktvariablen detailliert dargestellt, inkl. eines kurzen Überblicks über den theoretischen Hintergrund des Konstrukts und der wichtigsten Literaturverweise.

4.5.1 Personenebene

Bildung in Jahren

<u>Variablenname</u> <u>Variablenlabel</u> <u>Ausgangsvariablen</u> <u>Typ / Datensatz</u> <u>Bearbeitung</u> <u>Erläuterung</u>	<p><i>bilzeit</i></p> <p>Schul- und Berufsbildung in Jahren, generiert</p> <p><i>schul2; beruf2</i></p> <p>Bildung / Personendaten</p> <p>Bernhard Christoph</p> <p>Für viele Analyseverfahren ist die Verwendung einer linearen Bildungsvariable deutlich geeigneter als die einer kategorialen. Recht einfach lässt sich eine solche Umrechnung bei den Schulabschlüssen realisieren. Hier kann einfach die reale Dauer bis zum Bildungsabschluss verwendet werden. Dabei ist lediglich zu beachten, dass äquivalente Abschlüsse immer mit der gleichen Zeitdauer veranschlagt werden. So sollte z. B. eine Hochschulreife immer mit der gleichen Zeitdauer in die Variable eingehen, unabhängig davon, ob sie nach zwölf oder 13 Jahren erreicht wurde. Hier wurden die Schulabschlüsse wie folgt veranschlagt:</p> <table border="0" style="width: 100%;"> <tr> <td>Hauptschule; POS 8. Klasse; and. Abschluss:</td> <td style="text-align: right;">9 Jahre</td> </tr> <tr> <td>Mittlere Reife; POS 10. Klasse:</td> <td style="text-align: right;">10 Jahre</td> </tr> <tr> <td>Fachhochschulreife:</td> <td style="text-align: right;">12 Jahre</td> </tr> <tr> <td>Allg. oder fachgeb. Hochschulreife (incl. EOS)</td> <td style="text-align: right;">13 Jahre</td> </tr> </table> <p>Anders stellt sich die Situation bei den beruflichen Abschlüssen dar. Aufgrund der Unterschiedlichkeit der Bildungswege in der beruflichen Bildung und der mit diesen verbundenen – trotz vergleichbarer Dauer – teilweise recht stark divergierenden Einkommenserträge, verbietet sich hier eine einfache Umrechnung der Ausbildungsdauer. Dieses Problem kann dadurch umgangen werden, dass versucht wird, den mit einem bestimmten Berufsabschluss verknüpften Zuwachs an Humankapital zu operationalisieren (vgl. z. B. Helberger 1988).</p> <p>Ein ähnlicher Ansatz wurde in der vorliegenden Studie verfolgt. Dabei wurde jeweils nur der höchste Berufsabschluss der Befragungsperson berücksichtigt, und die für den jeweiligen Abschluss veranschlagten Zuwächse an Bildungsjahren wurden zur Dauer der Schulzeit hinzuaddiert.</p> <table border="0" style="width: 100%;"> <tr> <td>Anlernausbildung:</td> <td style="text-align: right;">+1 Jahr</td> </tr> <tr> <td>Lehre, Berufsfachschule, Schule des Gesundheitswesens:</td> <td style="text-align: right;">+1,5 Jahre</td> </tr> <tr> <td>Meister:</td> <td style="text-align: right;">+3 Jahre</td> </tr> <tr> <td>Berufsakademie:</td> <td style="text-align: right;">+3 Jahre</td> </tr> <tr> <td>Fachhochschule/Bachelor:</td> <td style="text-align: right;">+3 Jahre</td> </tr> <tr> <td>Uni/Master:</td> <td style="text-align: right;">+5 Jahre</td> </tr> <tr> <td>Promotion:</td> <td style="text-align: right;">+8 Jahre</td> </tr> <tr> <td>Anderer Abschluss, D:</td> <td style="text-align: right;">+1,5 Jahre</td> </tr> <tr> <td>Anderer Abschluss, Ausl.:</td> <td style="text-align: right;">+1,5 Jahre</td> </tr> </table> <p><u>Literatur</u></p> <p>Helberger (1988)</p>	Hauptschule; POS 8. Klasse; and. Abschluss:	9 Jahre	Mittlere Reife; POS 10. Klasse:	10 Jahre	Fachhochschulreife:	12 Jahre	Allg. oder fachgeb. Hochschulreife (incl. EOS)	13 Jahre	Anlernausbildung:	+1 Jahr	Lehre, Berufsfachschule, Schule des Gesundheitswesens:	+1,5 Jahre	Meister:	+3 Jahre	Berufsakademie:	+3 Jahre	Fachhochschule/Bachelor:	+3 Jahre	Uni/Master:	+5 Jahre	Promotion:	+8 Jahre	Anderer Abschluss, D:	+1,5 Jahre	Anderer Abschluss, Ausl.:	+1,5 Jahre
Hauptschule; POS 8. Klasse; and. Abschluss:	9 Jahre																										
Mittlere Reife; POS 10. Klasse:	10 Jahre																										
Fachhochschulreife:	12 Jahre																										
Allg. oder fachgeb. Hochschulreife (incl. EOS)	13 Jahre																										
Anlernausbildung:	+1 Jahr																										
Lehre, Berufsfachschule, Schule des Gesundheitswesens:	+1,5 Jahre																										
Meister:	+3 Jahre																										
Berufsakademie:	+3 Jahre																										
Fachhochschule/Bachelor:	+3 Jahre																										
Uni/Master:	+5 Jahre																										
Promotion:	+8 Jahre																										
Anderer Abschluss, D:	+1,5 Jahre																										
Anderer Abschluss, Ausl.:	+1,5 Jahre																										

Bildung in Jahren, Mutter

<u>Variablenname</u>	<i>mbilzeit</i>																
<u>Variablenlabel</u>	Schul- und Berufsbildung in Jahren, generiert																
<u>AusgangsvARIABLEN</u>	<i>mschul2; mberuf2</i>																
<u>Typ / Datensatz</u>	Bildung / Personendaten																
<u>Bearbeitung</u>	Bernhard Christoph																
<u>Erläuterung</u>	<p>Generelle Darstellung: vgl. Bildung in Jahren</p> <p>Bei den Bildungsabschlüssen der Eltern unterscheiden sich die für die Berufsbildung hinzuaddierten Jahre von denen beim Befragten, da die Berufsbildung der Eltern weniger detailliert erhoben wurde (insbes. im Bereich der tertiären Bildung). Die verwendeten Ausprägungen sind wie folgt:</p> <table border="0"> <tr> <td>Anlernausbildung:</td> <td>+1 Jahr</td> </tr> <tr> <td>Lehre, Berufsfachschule, Schule des Gesundheitswesens:</td> <td>+1,5 Jahre</td> </tr> <tr> <td>Meister:</td> <td>+3 Jahre</td> </tr> <tr> <td>Berufsakademie:</td> <td>+3 Jahre</td> </tr> <tr> <td>Fachhochschule:</td> <td>+3 Jahre</td> </tr> <tr> <td>Universität:</td> <td>+5 Jahre</td> </tr> <tr> <td>Anderer Abschluss, D:</td> <td>+1,5 Jahre</td> </tr> <tr> <td>Anderer Abschluss, Ausl.:</td> <td>+1,5 Jahre</td> </tr> </table>	Anlernausbildung:	+1 Jahr	Lehre, Berufsfachschule, Schule des Gesundheitswesens:	+1,5 Jahre	Meister:	+3 Jahre	Berufsakademie:	+3 Jahre	Fachhochschule:	+3 Jahre	Universität:	+5 Jahre	Anderer Abschluss, D:	+1,5 Jahre	Anderer Abschluss, Ausl.:	+1,5 Jahre
Anlernausbildung:	+1 Jahr																
Lehre, Berufsfachschule, Schule des Gesundheitswesens:	+1,5 Jahre																
Meister:	+3 Jahre																
Berufsakademie:	+3 Jahre																
Fachhochschule:	+3 Jahre																
Universität:	+5 Jahre																
Anderer Abschluss, D:	+1,5 Jahre																
Anderer Abschluss, Ausl.:	+1,5 Jahre																
<u>Literatur</u>	Helberger (1988)																

Bildung in Jahren, Vater

<u>Variablenname</u>	<i>vbilzeit</i>																
<u>Variablenlabel</u>	Schul- und Berufsbildung in Jahren, generiert																
<u>AusgangsvARIABLEN</u>	<i>vschul2; vberuf2</i>																
<u>Typ / Datensatz</u>	Bildung / Personendaten																
<u>Bearbeitung</u>	Bernhard Christoph																
<u>Erläuterung</u>	<p>Generelle Darstellung: vgl. Bildung in Jahren</p> <p>Bei den Bildungsabschlüssen der Eltern unterscheiden sich die für die Berufsbildung hinzuaddierten Jahre von denen beim Befragten, da die Berufsbildung der Eltern weniger detailliert erhoben wurde (insbes. im Bereich der tertiären Bildung). Die verwendeten Ausprägungen sind wie folgt:</p> <table border="0"> <tr> <td>Anlernausbildung:</td> <td>+1 Jahr</td> </tr> <tr> <td>Lehre, Berufsfachschule, Schule des Gesundheitswesens:</td> <td>+1,5 Jahre</td> </tr> <tr> <td>Meister:</td> <td>+3 Jahre</td> </tr> <tr> <td>Berufsakademie:</td> <td>+3 Jahre</td> </tr> <tr> <td>Fachhochschule:</td> <td>+3 Jahre</td> </tr> <tr> <td>Universität:</td> <td>+5 Jahre</td> </tr> <tr> <td>Anderer Abschluss, D:</td> <td>+1,5 Jahre</td> </tr> <tr> <td>Anderer Abschluss, Ausl.:</td> <td>+1,5 Jahre</td> </tr> </table>	Anlernausbildung:	+1 Jahr	Lehre, Berufsfachschule, Schule des Gesundheitswesens:	+1,5 Jahre	Meister:	+3 Jahre	Berufsakademie:	+3 Jahre	Fachhochschule:	+3 Jahre	Universität:	+5 Jahre	Anderer Abschluss, D:	+1,5 Jahre	Anderer Abschluss, Ausl.:	+1,5 Jahre
Anlernausbildung:	+1 Jahr																
Lehre, Berufsfachschule, Schule des Gesundheitswesens:	+1,5 Jahre																
Meister:	+3 Jahre																
Berufsakademie:	+3 Jahre																
Fachhochschule:	+3 Jahre																
Universität:	+5 Jahre																
Anderer Abschluss, D:	+1,5 Jahre																
Anderer Abschluss, Ausl.:	+1,5 Jahre																
<u>Literatur</u>	Helberger (1988)																

CASMIN

- Variablenname
- Variablenlabel
- AusgangsvARIABLEN
- Typ / Datensatz
- Bearbeitung
- Erläuterung

casmin

Bildung klassifiziert nach CASMIN, aktual. Version, generiert

schul2; beruf2

Bildung / Personendaten

Bernhard Christoph

Die CASMIN-Bildungsklassifikation wurde im Rahmen des CASMIN-Projekts (Comparative Analysis of Social Mobility in Industrial Nations) als international vergleichende Skala zur Erfassung von schulischen und beruflichen Bildungsabschlüssen entwickelt (König et al. 1987). Sie liegt mittlerweile in einer aktualisierten Fassung vor (Brauns & Steinmann 1999).

Die hier durchgeführte Umcodierung der Bildungsabschlüsse in CASMIN orientiert sich dabei – insbesondere, was die Behandlung problematischer Fälle angeht – an den in Lechert et al. (2006) sowie Granato (2000) beschriebenen Vorgehensweisen (selbstverständlich unter Berücksichtigung der leicht unterschiedlichen Erfassung der Bildungsvariablen im vorliegenden Datensatz). Die Details können der folgenden tabellarischen Darstellung entnommen werden. Zellen mit Kombinationen, die gültige CASMIN-Werte ergeben, sind hellgrau, solche mit definierten missing values dunkelgrau unterlegt.

<u>schul</u> <u>beruf</u>	<u>nicht</u> <u>erhob.</u>	<u>Schüler</u>	<u>n. gest.</u>	<u>TNZ</u>	<u>KA</u>	<u>WN</u>	<u>ohne</u> <u>Abschl.</u>	<u>Sonder-</u> <u>schule</u>	<u>HS</u>	<u>RS</u>	<u>FHR</u>	<u>Abi</u>	<u>And. dt.</u> <u>Abschl.</u>	<u>And. aus.</u> <u>Abschl.</u>
nicht erhob.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-
unplaus. Wert	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Schüler	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
nicht gest.	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
TNZ	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
KA	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
WN	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
ohne Abschl.	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Anlernausbild.	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Lehre	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Berufsfachs.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Schul. d. Ges.wes.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Meister	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
BA	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
FIH/Bachelor	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
Uni/Master	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
Dissert.	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
And. dt. Abschl.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
And. aus Abschl.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c

Literatur

Brauns et al. (1999); Granato (2000); König et al. (1987); Lechert et al. (2006)

MCASMIN

Variablenname

Variablenlabel

AusgangsvARIABLEN

Typ / Datensatz

Bearbeitung

Erläuterung

mcasmin

Bildung der Mutter, klassifiziert nach CASMIN, aktual. Version, generiert

mschul2; mberuf2

Bildung / Personendaten

Bernhard Christoph

Generelle Darstellung: vgl. CASMIN

Aufgrund der unterschiedlichen Erhebung der Bildungsvariablen für Befragte und deren Eltern weicht das Vercodungsschema von *mcasmin* und *vcasmin* leicht von dem bei *casmin* verwendeten ab. Zu Details vgl. die folgende Abbildung.

Schul- Beruf	nicht erhob.	Plat fehlt	Elternt. unbek.	nicht gest.	TNZ	KA	WN	ohne Abschl.	Sonder- Schule	HS	RS	FHR	Abi	And. dt. Abschl.	And. au. Abschl.
nicht erhob.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
unplaus. Wert	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Plat fehlt	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Elternt. unbek.	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
nicht gest.	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
TNZ	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
KA	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
WN	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
ohne Abschl.	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Anlern- ausbild.	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Lehre	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Meister	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
BA	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
FH	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
Uni	-	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
And. dt. Abschl.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
And. aus Abschl.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c

Literatur

Brauns et al. (1999); Granato (2000); König et al. (1987); Lechert et al. (2006)

VCASMIN

Variablenname

Variablenlabel

AusgangsvARIABLEN

Typ / Datensatz

Bearbeitung

Erläuterung

vcasmin

Bildung des Vaters klassifiziert nach CASMIN, aktual. Version, generiert

vschul2; vberuf2

Bildung / Personendaten

Bernhard Christoph

Generelle Darstellung: vgl. CASMIN

Aufgrund der unterschiedlichen Erhebung der Bildungsvariablen für Befragte und deren Eltern weicht das Vercodungsschema von *mcasmin* und *vcasmin* leicht von dem bei *casmin* verwendeten ab. Zu Details vgl. die folgende Abbildung.

Schul- Beruf	nicht erhob.	Plat fehlt	Eltern. unbek.	nicht gest.	TNZ	KA	WN	ohne Abschl.	Sonder- Schule	HS	RS	FHR	Abi	And. dt. Abschl.	And. au. Abschl.
nicht erhob. unplaus. Wert	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Plat fehlt	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Eltern. unbek. nicht gest.	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
TNZ	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
KA	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
WN	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
ohne Abschl.	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Anlern- ausbild.	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Lehre	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Meister	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
BA	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
FH	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
Uni	-	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
And. dt. Abschl.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
And. aus Abschl.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c

Literatur

Brauns et al. (1999); Granato (2000); König et al. (1987); Lechert et al. (2006)

ISCED 97

- Variablenname
- Variablenlabel
- Ausgangsvariablen
- Typ / Datensatz
- Bearbeitung
- Erläuterung

isced97

Bildung klassifiziert nach isced97, aktual. Version, generiert

schul2; beruf2

Bildung / Personendaten

Bernhard Christoph

Als Alternative zu CASMIN bietet sich die Bildungsklassifikation ISCED-97 (International Standard Classification of Education) der OECD an (OECD 1999, für eine Kurzdarstellung vgl. auch BMBF 2003).

Bei der Vercodung der ISCED-97 ist zu beachten, dass die Klassifikation auch Ausprägungen enthält, die sich anhand der vorhandenen Daten nicht sinnvoll zuordnen lassen. Die ISCED Stufen 0 (Vorschulerziehung/Kindergarten) und 1 (Grundschule) sind aufgrund des Alters der Stichprobenpersonen (mind. 15 Jahre) nicht adäquat. Stattdessen wurde eine eigene Gruppe für Personen mit einer niedrigeren Bildung als ISCED Stufe 2 (ISCED 2=Haupt- bzw. Realschulabschluss) generiert. Aus diesem Grund umfasst die hier vorgenommene Codierung lediglich die ISCED Stufen 2-6.

Die Details der Zuordnung können der folgenden Tabelle entnommen werden. Zellen mit Kombinationen, die gültige ISCED-Werte ergeben, sind hellgrau, solche mit definierten missing values dunkelgrau unterlegt.

Schul Beruf	nicht erhob.	Schüler	nicht gest.	TNZ	KA	WN	ohne Abschl.	Sonder- schule	HS	RS	FHR	Abi	And. dt. Abschl.	And. aus. Abschl.
nicht erhob.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-
unplaus. Wert	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Schüler	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
nicht gest.	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
TNZ	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
KA	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
WN	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
ohne Abschl.	-	-	-	-3	-2	-1	1	1	2	2	3a	3a	2	2
Anlern- ausbild.	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
Lehre	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Berufs- fachs.	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Schul. d. Ges.wes.	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
Meister	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
BA	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
FH/ Bachelor	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Uni/ Master	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Dissert.	-	-	-	6	6	6	6	6	6	6	6	6	6	6
And. dt. Abschl.	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
And. aus. Abschl.	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2

Literatur

BMBF (2003); OECD (1999)

MISCED 97

Variablenname

Variablenlabel

AusgangsvARIABLEN

Typ / Datensatz

Bearbeitung

Erläuterung

misc97

Bildung der Mutter klassifiziert nach isced97, aktual. Version, generiert

mschul2; mberuf2

Bildung / Personendaten

Bernhard Christoph

Zum theoretischen Hintergrund und zur Generierung vgl. ISCED-97.

Im Gegensatz zur Umsetzung von ISCED-97 bei den Befragten selbst lässt sich ISCED Stufe 6 für die Eltern nicht generieren, da die entsprechenden Abschlüsse (Promotion oder Habilitation) nicht separat erfragt wurden. Aus diesem Grund umfasst die hier vorgenommene Codierung lediglich die ISCED Stufen 2-5. Zu Details der Vercodung vgl. die folgende Abbildung.

Schul Beruf	nicht erhob.	Plat fehlt	Elternt. unbek.	nicht gest.	TNZ	KA	WN	ohne Abschl.	Sonder- Schule	HS	RS	FHR	Abi	And. dt. Abschl.	And. au. Abschl.
nicht erhob.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
unplaus. Wert	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Plat fehlt	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Elternt. unbek.	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
nicht gest.	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
TNZ	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
KA	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
WN	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
ohne Abschl.	-	-	-	-	-3	-2	-1	1	1	2	2	3a	3a	2	2
Anlern- ausbild.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
Lehre	-	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Meister	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
BA	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
FH	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Uni	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
And. dt. Abschl.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
And. aus Abschl.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2

Literatur

BMBF (2003); OECD (1999)

VISCED 97

Variablenname

Variablenlabel

Ausgangsvariablen

Typ / Datensatz

Bearbeitung

Erläuterung

visced97

Bildung des Vaters klassifiziert nach isced97, aktual. Version, generiert

vschul2; vberuf2

Bildung / Personendaten

Bernhard Christoph

Zum theoretischen Hintergrund und zur Generierung vgl. ISCED-97.

Im Gegensatz zur Umsetzung von ISCED-97 bei den Befragten selbst lässt sich ISCED Stufe 6 für die Eltern nicht generieren, da die entsprechenden Abschlüsse (Promotion oder Habilitation) nicht separat erfragt wurden. Aus diesem Grund umfasst die hier vorgenommene Codierung lediglich die ISCED Stufen 2-5. Zu Details der Vercodung vgl. die folgende Abbildung.

Schulberuf	nicht erhob.	Platz fehlt	Elternt. unbek.	nicht gest.	TNZ	KA	WN	ohne Abschl.	Sonder-Schule	HS	RS	FHR	Abi	And. dt. Abschl.	And. au. Abschl.
nicht erhob.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
unplaus. Wert	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Platz fehlt	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Elternt. unbek.	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
nicht gest.	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
TNZ	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
KA	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
WN	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
ohne Abschl.	-	-	-	-	-3	-2	-1	1	1	2	2	3a	3a	2	2
Anlernausbild.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
Lehre	-	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Meister	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
BA	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
FH	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Uni	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
And. dt. Abschl.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
And. aus Abschl.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2

Literatur

BMBF (2003); OECD (1999)

International Standard Classification of Occupations 1988 (ISCO88); ZUMA-Verkodung

<u>Gebildet</u>	<u>Erwerbstätigkeit</u>	<u>Variablenname</u>	<u>Ausgangsvariablen</u>
	aktuelle	<i>isco88</i>	ET2500
	Spelldaten (<i>bio_spells</i>)	<i>isco88</i>	ET2500
	erste	<i>iscoeewt</i>	ET2500, PET1280, PET3950
	letzte	<i>iscolewt</i>	ET2500, PET1280
	des Vaters	<i>visco</i>	PSH0800
	der Mutter	<i>misco</i>	PSH0700
<u>Variablenlabel</u>	aktuelle Ewt.: ISCO 88 (ZUMA-Verkodung), generiert		
	Spelldaten (<i>bio_spells</i>): ISCO 88 (ZUMA-Verkodung), generiert		
	erste Ewt.: ISCO 88 (ZUMA-Verkodung), erste Erwerbstätigkeit, generiert		
	letzte Ewt.: ISCO 88 (ZUMA-Verkodung), letzte Erwerbstätigkeit, generiert		
	Vater: ISCO 88 (ZUMA-Verkodung) des Vaters, generiert		
	Mutter: ISCO 88 (ZUMA-Verkodung) der Mutter, generiert		
<u>Typ / Datensatz</u>	Beruf / Personendaten		
<u>Ansprechpartner</u>	Bernhard Christoph		
<u>Erläuterung</u>	Die International Standard Classification of Occupations (ISCO) ist eine von der ILO entwickelte, international vergleichbare Berufsklassifikation. Besonderheit der ISCO-88 ist, speziell im Vergleich zur hier ebenfalls zur Verfügung gestellten Klassifizierung der Berufe des statistischen Bundesamts, dass neben der beruflichen Tätigkeit auch das zur Ausübung dieser Tätigkeit i.d.R. notwendige Bildungsniveau bei der Zuordnung zu einem bestimmten Berufscodex berücksichtigt wird.		
<u>Literatur</u>	ILO (1990)		

Klassifizierung der Berufe 1992 (KIdB92)

<u>Gebildet</u>	<u>Erwerbstätigkeit</u>	<u>Variablenname</u>	<u>Ausgangsvariablen</u>
	aktuelle	<i>kldb_it</i>	ET2500
	Spelldaten (<i>bio_spells</i>)	<i>kldb</i>	ET2500
	erste	<i>kldbewt</i>	ET2500, PET1280, PET3950
	letzte	<i>kldblewt</i>	ET2500, PET1280
	des Vaters	<i>vkldb</i>	PSH0800
	der Mutter	<i>mkldb</i>	PSH0700
<u>Variablenlabel</u>	aktuelle Ewt.: Klassifizierung der Berufe 1992, aktuelle Erwerbstätigkeit Spelldaten (<i>bio_spells</i>): Klassifizierung der Berufe 1992 , generiert erste Ewt.: Klassifiz. d. Berufe 1992, erste Erwerbstätigkeit, gen. letzte Ewt.: Klassifizierung der Berufe 1992, letzte Erwerbst., gen. Vater: Klassifizierung der Berufe 1992 des Vaters, generiert Mutter: Klassifizierung der Berufe 1992 der Mutter, generiert		
<u>Typ / Datensatz</u>	Beruf / Personendaten		
<u>Ansprechpartner</u>	Bernhard Christoph		
<u>Erläuterung</u>	Die KIdB92 ist die aktuelle Version der Klassifizierung der Berufe in der Variante des Statistischen Bundesamts. Sie ist eine spezifisch auf Deutschland bezogene Berufsklassifikation. Die Klassifikation basiert ausschließlich auf der beruflichen Tätigkeit.		
<u>Literatur</u>	StBA (1992)		

Klassenschema nach Erikson, Goldthorpe und Portocarrero (EGP)

<u>Gebildet</u>	<u>Erwerbstätigkeit</u>	<u>Variablenname</u>	<u>Ausgangsvariablen</u>
	aktuelle	<i>egp</i>	<i>isco88, stib</i>
	Spelldaten (<i>bio_spells</i>)	<i>egp</i>	<i>isco88, stib</i>
	erste	<i>egpeewt</i>	<i>iscoeewt, stibeewt</i>
	letzte	<i>egplewt</i>	<i>iscolewt, stiblewt</i>
	des Vaters	<i>vegp</i>	<i>visco, vstib</i>
	der Mutter	<i>megp</i>	<i>misco, mstib</i>
<u>Variablenlabel</u>	aktuelle Ewt.: Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), akt. Beruf, gen. Spelldaten (<i>bio_spells</i>): Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), gen. erste Ewt.: Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), erste Ertwerbst., gen. letzte Ewt.: Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), letzte Erwerbst., gen. Vater: Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), Beruf d. Vaters, gen. Mutter: Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), Beruf d. Mutter, gen.		
<u>Typ / Datensatz</u>	sozioökonomische Position / Personendaten		
<u>Bearbeitung</u>	Bernhard Christoph		
<u>Erläuterung</u>	Das Klassenschema von Erikson, Goldthorpe und Portocarero (Erikson et al. 1979, 1982; Erikson & Goldthorpe 1992) ist eines der gebräuchlichsten Instrumente zur Operationalisierung der Klassenposition. Die Vercodung erfolgte hier ausschließlich auf Basis der ISCO-88-Berufsklassifikation und der Stellung im Beruf. Sie basiert auf einem früher erarbeiteten Vercodungsansatz von Christoph et al. (2005), wo sich auch eine detaillierte Darstellung des Verfahrens findet. Im Unterschied zum dort beschriebenen Vorgehen wurden hier allerdings – analog zur im Folgenden beschriebenen Vercodung der European Socio-Economic Classification (ESeC) – mithelfende Familienangehörige nicht als Selbständige, sondern als abhängig Erwerbstätige vercodet. Ein Unterschied zwischen den hier durchgeführten Codierungen von EGP und ESeC besteht darin, dass bei EGP solche Fälle auf „Missing“ (-7) gesetzt wurden, bei denen die berufliche Tätigkeit und die Stellung im Beruf nicht miteinander vereinbar schien (z. B. „Direktoren und Hauptgeschäftsführer“ [ISCO=1210], die angeblich „Angestellte mit einfacher Tätigkeit“ [StiB=51] zu sein). Im Falle von ESeC wurde aus Gründen der Kompatibilität mit der von uns übernommenen, stark standardisierten Vercodung dieses Instruments auf eine analoge Überprüfung verzichtet.		
<u>Literatur</u>	Christoph et al. (2005); Erikson & Goldthorpe (1992); Erikson et al. (1982); Erikson et al. (1979):		

European Socio-economic Classification (ESeC)

<u>Gebildet</u>	<u>Erwerbstätigkeit</u>	<u>Variablenname</u>	<u>Ausgangsvariablen</u>
	aktuelle	<i>esec</i>	<i>isco88, stib, PET2000, PET2700</i>
	Spelldaten (<i>bio_spells</i>)	<i>esec</i>	<i>isco88, stib, ET1100, ET1101, ET1102, ET1300, ET1301, ET1302,</i>
	erste	<i>eseceewt</i>	<i>iscoeewt, stibeewt, PET1261</i>
	letzte	<i>eseclewt</i>	<i>iscolewt, stiblewt, PET3801</i>
	des Vaters	<i>vesec</i>	<i>visco, vstib, PSH0670</i>
	der Mutter	<i>mesec</i>	<i>misco, mstib, PSH0370</i>
<u>Variablenlabel</u>	aktuelle Ewt.: European Socio-economic Classification (ESeC), akt. Beruf, gen. Spelldaten (<i>bio_spells</i>): European Socio-economic Classification (ESeC), gen. erste Ewt.: European Socio-economic Classification (ESeC), erste Erwerbst., gen. letzte Ewt.: European Socio-economic Classification (ESeC), letzte Erwerbst., gen. Vater: European Socio-economic Classification (ESeC), Beruf d. Vaters, gen. Mutter: European Socio-economic Classification (ESeC), Beruf d. Mutter, gen.		
<u>Typ / Datensatz</u>	sozioökonomische Position / Personendaten		
<u>Bearbeitung</u>	Bernhard Christoph		
<u>Erläuterung</u>	Die European Socio-economic Classification orientiert sich von ihrer theoretischen Konzeption her stark am EGP-Klassenschema. Im Unterschied zu Letzterem wurde bei ESeC jedoch hoher Wert auf eine international vergleichbare Operationalisierung und eine ausführliche Validierung der Klassifikation gelegt (vgl. für eine generelle Darstellung: Rose & Harrison 2007; für Deutschland Müller et al. 2006, 2007). Der Stata do-file für die Generierung von ESeC wurde uns dankenswerterweise von Heike Wirth von GESIS-ZUMA zur Verfügung gestellt (Fischer & Wirth 2007) und von uns lediglich für die vorliegende Studie angepasst. Dieser do-file stellt eine Stata-Umsetzung der von Harrison & Rose (2006) geschriebenen Standard-SPSS-Syntax zur Generierung von ESeC dar.		
<u>Literatur</u>	Fischer & Wirth (2007); Harrison & Rose (2006); Müller et al. (2006, 2007); Rose & Harrison (2007)		

Magnitude-Prestigeskala (MPS)

<u>Gebildet</u>	<u>Erwerbstätigkeit</u>	<u>Variablenname</u>	<u>Ausgangsvariablen</u>
	aktuelle	<i>mps</i>	<i>isco88</i>
	Spelldaten (<i>bio_spells</i>)	<i>mps</i>	<i>isco88</i>
	erste	<i>mpseewt</i>	<i>iscoeewt</i>
	letzte	<i>mpslewt</i>	<i>iscolewt</i>
	des Vaters	<i>vmmps</i>	<i>visco</i>
	der Mutter	<i>mmmps</i>	<i>misco</i>
<u>Variablenlabel</u>	aktuelle Ewt.: Magnitude-Prestigeskala, akt. Beruf, gen. Spelldaten (<i>bio_spells</i>): Magnitude-Prestigeskala, generiert erste Ewt.: Magnitude-Prestigeskala, erste Erwerbst., gen. letzte Ewt.: Magnitude-Prestigeskala, letzte Erwerbst., gen. Vater: Magnitude-Prestigeskala, Beruf d. Vaters, gen. Mutter: Magnitude-Prestigeskala, Beruf d. Mutter, gen.		
<u>Typ / Datensatz</u>	sozioökonomische Position / Personendaten		
<u>Ansprechpartner</u>	Bernhard Christoph		
<u>Erläuterung</u>	Die Magnitude-Prestigeskala [MPS] (Wegener 1985, 1988) ist bis heute die einzige spezifisch deutsche Prestigeskala auf Basis detaillierter Berufsangaben. Sie wurde ursprünglich für die ältere 1968er Version der International Standard Classification of Occupations (ISCO-68) entwickelt. Da die Berufsvercodung in der vorliegenden Studie lediglich anhand der aktuelleren ISCO-88-Klassifikation sowie der KIdB 1992 des Statistischen Bundesamts vorgenommen wurde, fand hier die auf ISCO-88 transferierte Variante der Skala (Christoph 2005) Verwendung. Die Zuspiegelung erfolgte im Rahmen der Berufsvercodung durch infas.		
<u>Literatur</u>	Christoph (2005); Wegener (1985, 1988)		

Standard International Occupational Prestige Scale (SIOPS/Treiman-Skala)

<u>Gebildet</u>	<u>Erwerbstätigkeit</u>	<u>Variablenname</u>	<u>Ausgangsvariablen</u>
	aktuelle	<i>siops</i>	<i>isco88</i>
	Spelldaten (<i>bio_spells</i>)	<i>siops</i>	<i>isco88</i>
	erste	<i>siopseewt</i>	<i>iscoeewt</i>
	letzte	<i>siopslewt</i>	<i>iscolewt</i>
	des Vaters	<i>vsiops</i>	<i>visco</i>
	der Mutter	<i>msiops</i>	<i>misco</i>
<u>Variablenlabel</u>	aktuelle Ewt.: Standard International Occupational Prestige Scale, akt. Beruf, gen. Spelldaten (<i>bio_spells</i>): Standard International Occupational Prestige Scale, generiert erste Ewt.: Standard International Occupational Prestige Scale, erste Erwerbst., gen. letzte Ewt.: Standard International Occupational Prestige Scale, letzte Erwerbst., gen. Vater: Standard International Occupational Prestige Scale, Beruf d. Vaters, gen. Mutter: Standard International Occupational Prestige Scale, Beruf d. Mutter, gen.		
<u>Typ / Datensatz</u>	sozioökonomische Position / Personendaten		
<u>Ansprechpartner</u>	Bernhard Christoph		
<u>Erläuterung</u>	Die ursprünglich von Treiman (1977) für ISCO-68 entwickelte Treiman-Prestigeskala ist die erste und bis heute einzige international vergleichende Berufsprestigeskala. Seit der Umsetzung der Skala auf ISCO-88 (Ganzeboom & Treiman 1996, 2003) bürgerte sich die dort von den Autoren verwendete Bezeichnung als „Standard International Occupational Prestige Scale“ ein. Die Zuspiegelung erfolgte im Rahmen der Berufsvercodung durch infas.		
<u>Literatur</u>	Ganzeboom & Treiman (1996, 2003); Treiman (1977)		

International Socio-Economic Index (ISEI)

<u>Gebildet</u>	<u>Erwerbstätigkeit</u>	<u>Variablenname</u>	<u>Ausgangsvariablen</u>
	aktuelle	<i>isei</i>	<i>isco88</i>
	Spelldaten (<i>bio_spells</i>)	<i>isei</i>	<i>isco88</i>
	erste	<i>iseieewt</i>	<i>iscoeewt</i>
	letzte	<i>iseilewt</i>	<i>iscolewt</i>
	des Vaters	<i>visei</i>	<i>visco</i>
	der Mutter	<i>misei</i>	<i>misco</i>
<u>Variablenlabel</u>	aktuelle Ewt.: International Socio-Economic Index, akt. Beruf, gen. Spelldaten (<i>bio_spells</i>): International Socio-Economic Index, generiert erste Ewt.: International Socio-Economic Index, erste Erwerbst., gen. letzte Ewt.: International Socio-Economic Index, letzte Erwerbst., gen. Vater: International Socio-Economic Index, Beruf d. Vaters, gen. Mutter: International Socio-Economic Index, Beruf d. Mutter, gen.		
<u>Typ / Datensatz</u>	sozioökonomische Position / Personendaten		
<u>Ansprechpartner</u>	Bernhard Christoph		
<u>Erläuterung</u>	Der International Socio-Economic Index ist sicherlich einer der gebräuchlichsten Indizes seiner Art. Dies verdankt er nicht zuletzt der Tatsache, dass ihm, im Gegensatz zu den meisten SEIs, ein eigenständiges theoretisches Konzept zugrunde liegt, in dem der Beruf bzw. dessen sozioökonomischer Status als „intervening variable“ zwischen Bildung und Einkommen begriffen wird. Der ISEI wurde zunächst für ISCO-68 entwickelt (Ganzeboom et al. 1992), später jedoch auch für ISCO-88 umgesetzt (Ganzeboom & Treiman 1996, 2003). Die Zuspiegelung erfolgte im Rahmen der Berufsvercodung durch infas.		
<u>Literatur</u>	Ganzeboom et al. (1992); Ganzeboom & Treiman (1996, 2003)		

Klassifikation der Wirtschaftszweige 2003 (WZ2003)

<u>Gebildet</u>	<u>Erwerbstätigkeit</u>	<u>Variablenname</u>	<u>Ausgangsvariablen</u>
	aktuelle	<i>branche</i>	ET2600
	Spelldaten (<i>bio_spells</i>)	<i>branche</i>	ET2600
<u>Variablenlabel</u>	aktuelle Ewt.: Jetzige Tätigkeit: Wirtschaftszweig/Branche (WZ2003) Spelldaten (<i>bio_spells</i>): Wirtschaftszweig/Branche (WZ 2003), generiert		
<u>Typ / Datensatz</u>	sozioökonomische Position / Personendaten		
<u>Ansprechpartner</u>	Bernhard Christoph		
<u>Erläuterung</u>	Die Vercodung der offenen Branchenangabe erfolgte anhand des 2-Stellers der Klassifikation der Wirtschaftszweige des statistischen Bundesamts (WZ2003). Auf der Ebene des Zweistellers entspricht diese Klassifikation weitgehend der europäischen Nomenclature générale des Activités économiques dans les Communautés Européennes (NACE) in der Revision 1.1.		
<u>Literatur</u>	StaBA (2002); EG (2002)		

Ausgeübte und gewünschte Freizeitbeschäftigungen von Jugendlichen

<u>Variablenname</u>	<i>freiz1, freiz2, freiz3, frwunsch</i>
<u>Variablenlabel</u>	freiz1: Freizeitbeschaeftigung 1, ausgeuebt freiz2: Freizeitbeschaeftigung 2, ausgeuebt freiz3: Freizeitbeschaeftigung 3, ausgeuebt frwunsch: Freizeitbeschaeftigung, gewuenscht
<u>Ausgangsvariablen</u>	PA1100 (für freiz1-freiz3); PA1200 (für frwunsch)

Typ / Datensatz
Bearbeitung
Erläuterung

Freizeit/Personendaten

Johanna Eckert (DJI), Arne Bethmann, Claudia Wenzig

Erläuterung:

Die Variablen freiz1, freiz2, freiz3 und frwunsch basieren auf einem neu entwickelten Kategorienschema zu Freizeitbeschäftigungen von Jugendlichen. Ausgangspunkt des Kategorienschemas bilden die offenen Angaben zu den drei beliebtesten Freizeitbeschäftigungen (PA1100) und zu der gewünschten Freizeitbeschäftigung (PA1200).

Die beliebtesten Freizeitbeschäftigungen wurden gemäß Fragetext in maximal drei Einzelvariablen überführt. Bei der Frage nach dem Freizeitwunsch wurde gemäß Fragetext nur eine Nennung berücksichtigt. Darüber hinaus gehende Nennungen wurden für die Codierung nicht einbezogen.

Das Schema wurde auf der Grundlage der offenen, bereinigten Angaben induktiv erarbeitet. Um eine Vergleichbarkeit zwischen den Wellen zu erreichen, enthält das neue Kategorienschema auch alle Freizeitbeschäftigungen, die in den Vorwellen geschlossen abgefragt wurden. Weiterhin ist das Schema so aufgebaut, dass es ggf. in den Folgewelle durch neue Oberkategorien wie auch Subkategorien erweitert werden kann.

Insgesamt umfasst das Kategorienschema 16 Oberkategorien zuzüglich der Kategorien „keine Freizeitbeschäftigung“ und „nicht zuordenbare Nennung“, wobei sich die Reihenfolge der 14 inhaltlichen Oberkategorien aus der Häufigkeit ihrer Nennungen ergibt. Die Oberkategorien können mit Hilfe von 77 Subkategorien ausdifferenziert werden.

	Oberkategorie/Variablenausprägung	Anzahl der Subkategorien
1000	Sport und Bewegung	31
2000	Zeit mit Familie, Freunden verbringen	4
3000	Computer, Spiel und Kommunikation	5
4000	Musik machen und hören	6
5000	Lesen	-
6000	Kultur, Kino, Fernsehen und Veranstaltungen	8
7000	Kreative Hobbies, Heimwerken, Kochen und Backen	11
8000	Ausgehen, Feiern, Disco	3
9000	Nichts tun, Ausruhen	-
10000	Shoppen und Einkaufen	-
11000	Reisen, Ausflüge, Touren machen und mobil sein	3
12000	Beschäftigung mit Haustieren	-
13000	Ehrenamtliches Engagement	4
14000	Lernen und Bildung	-
15000	Spiele und Denksport	2
16000	Nebenjob	-
99998	Keine Freizeitbeschäftigung	-
99999	nicht zuordenbare Nennung	-

Literatur

Johanna Eckert, Arne Bethmann, Claudia Wenzig (geplant): Manuelle Vercodung „Ausgeübte und gewünschte Freizeitbeschäftigungen von Jugendlichen“. PASS Welle 5 (2011). Abschlussbericht.

4.5.2 Haushalts- oder Bedarfsgemeinschaftsebene

Bedarfsgewichtetes Haushaltseinkommen, alte OECD-Gewichtung

<u>Variablenname</u>	<i>oecdinca</i>
<u>Variablenlabel</u>	bedarfsgewichtetes Haushaltseinkommen, alte OECD-Gewichtung (gerundet)
<u>Ausgangsvariablen</u>	<i>HD0200a-HD0200o; HA0100; hhincome</i>
<u>Typ / Datensatz</u>	sozioökonomische Position / Haushaltsdaten
<u>Bearbeitung</u>	Bernhard Christoph
<u>Erläuterung</u>	<p>Mittels eines bedarfsgewichteten Haushaltseinkommens (auch Äquivalenzeinkommen genannt) wird versucht, die in Mehrpersonenhaushalten durch gemeinsames Wirtschaften im Vergleich zu Einpersonenhaushalten erreichbaren Einsparungen zu berücksichtigen. Dies geschieht dadurch, dass in Mehrpersonenhaushalten bei der Berechnung des Pro-Kopf-Einkommens nicht die reale Haushaltsgröße als Divisor verwendet wird, sondern eine über den unterstellten Bedarf der Personen berechnete, in der Regel niedrigere Zahl (die bedarfsgewichtete Haushaltsgröße).</p> <p>Bei der alten OECD-Skala wird dabei nur für die erste (mind. 15-jährige) Person des Haushalts ein Bedarfsgewicht von 1,0 unterstellt. Alle weiteren Personen ab 15 erhalten ein Bedarfsgewicht von 0,7; Personen bis einschl. 14 gehen mit einem Gewicht von 0,5 in die bedarfsgewichtete Haushaltsgröße ein.</p> <p>Zur alten OECD-Skala vgl. OECD (1982) für einen Überblick zum Thema vgl. Hauser (1996).</p>
<u>Literatur</u>	Hauser (1996); OECD (1982)

Bedarfsgewichtetes Haushaltseinkommen, neue OECD-Gewichtung

<u>Variablenname</u>	<i>oecdincn</i>
<u>Variablenlabel</u>	bedarfsgewichtetes Haushaltseinkommen, neue OECD-Gewichtung (gerundet)
<u>Ausgangsvariablen</u>	<i>HD0200a-HD0200o; HA0100; hhincome</i>
<u>Typ / Datensatz</u>	sozioökonomische Position / Haushaltsdaten
<u>Bearbeitung</u>	Bernhard Christoph
<u>Erläuterung</u>	<p><u>Generelle Darstellung:</u> vgl. Bedarfsgewichtetes Haushaltseinkommen, alte OECD-Gewichtung.</p> <p>Bei der neuen OECD-Skala wird nur für die erste (mind. 15-jährige) Person des Haushalts ein Bedarfsgewicht von 1,0 unterstellt. Alle weiteren Personen ab 15 erhalten ein Bedarfsgewicht von 0,5; Personen bis einschl. 14 gehen mit einem Gewicht von 0,3 in die bedarfsgewichtete Haushaltsgröße ein.</p> <p>Zur neuen OECD-Skala vgl. Hagenaars et al. (1994).</p>
<u>Literatur</u>	Hagenaars et al. (1994)

Deprivationsindex, ungewichtet

<u>Variablenname</u>	<i>depindug2</i>
<u>Variablenlabel</u>	Alle Wellen: Deprivationsindex, ungewichtet (Itemsumme: 23)
<u>Ausgangsvariablen</u>	<i>HLS0100a-HLS0400a; HLS0100b-HLS0400b; HLS0600a-HLS1200a; HLS0600b-HLS1200b; HLS1400a-HLS2500a; HLS1400b-HLS2500b;</i>
<u>Typ / Datensatz</u>	materielle Situation / Haushaltsdaten
<u>Bearbeitung</u>	Bernhard Christoph
<u>Erläuterung</u>	<p>In der Armutforschung wird im Anschluss an Ringen (1988) i.d.R. zwischen direkter und indirekter Messung von Armut unterschieden. Bei der indirekten Messung konzentriert man sich dabei auf die zur Erreichung eines bestimmten Lebensstandards zur Verfügung stehenden Ressourcen, insbesondere das (bedarfsgewichtete Haushalts-) Einkommen. Aus diesem Grund spricht man hier auch vom Ressourcenansatz.</p> <p>Die direkte Messung versucht hingegen, die tatsächlich vorhandene Güterausstattung der Haushalte zu erfassen bzw. festzustellen, inwieweit die Haushalte aus finanziellen Gründen nicht in der Lage sind, sich bestimmte, als relevant erachtete Dinge zu leisten. Hier spricht man auch vom Deprivationsansatz (vgl.: hierzu z. B. Halleröd 1995).</p> <p>Als genereller Tenor der Forschung ist festzuhalten, dass der Ressourcen- und der Deprivationsansatz zum Teil unterschiedliche Personengruppen erfassen bzw. als arm klassifiziert. Für eine Abgrenzung der Armenpopulation im engeren Sinne wird daher auch oft vorgeschlagen, Maße für Einkommensarmut und Deprivation zu kombinieren und nur diejenigen, die von beiden Ansätzen als arm klassifiziert werden, als Armutpopulation im engeren Sinne zu betrachten (vgl. Halleröd 1995; Nolan & Whelan 1996; Andreß und Lipsmeier 2001).</p> <p>Dem Index liegt eine Liste von 23 Gütern oder Aktivitäten zugrunde, für die jeweils erhoben wird, ob der Haushalt der Befragten diese Güter hat bzw. an diesen Aktivitäten partizipiert. Der hier berechnete ungewichtete Index summiert lediglich die Anzahl der Items, bei denen dies nicht der Fall ist. Allerdings wird das Fehlen eines Items nur dann berücksichtigt, wenn es auf finanzielle Gründe zurückzuführen ist. Dies geschieht, um auszuschließen, dass bestimmte Konsumpräferenzen (z. B. der bewusste Verzicht auf ein Auto oder einen Fernseher) als Minderung des Lebensstandards fehlinterpretiert werden.</p> <p>Zudem gilt, dass ein Fehlen aus finanziellen Gründen nur dann angenommen wurde, wenn beide Fragen explizit in diesem Sinne beantwortet wurden. WN- bzw. KA-Angaben wurden also gewertet, als ob das entsprechende Gut vorhanden ist bzw. als ob es aus anderen als aus finanziellen Gründen fehlt. Diese Annahme ist sicherlich nicht in jedem Fall korrekt. Alternativ wäre es möglich gewesen, in allen Fällen, in denen die Frage nach (mindestens) einem Gut nicht beantwortet wurde, überhaupt keinen Indexwert zu berechnen ("istwise deletion"). Bei der Abfrage von 23 Gütern hätte dies allerdings sehr schnell zu einer hohen Zahl fehlender Indexwerte führen können, weshalb das oben beschriebene alternative Vorgehen gewählt wurde. Im Vergleich zur listwise deletion besteht hierbei allerdings die Gefahr, die Zahl der fehlenden Güter zu unterschätzen.</p> <p>Für die Wellen 1 bis 4 liegt mit der Variable <i>depindug</i> eine Version des ungewichteten Deprivationsindex vor, der auf 26 statt 23 Items basiert, nämlich zusätzlich zu den oben genannten Items auch auf HLS0500*, HLS1300* und HLS2600*. Diese drei Items werden seit Welle 5 nicht mehr erhoben. Daher wurde <i>depindug2</i> neu in den Datensatz eingefügt und rückwirkend seit Welle 1 generiert.</p>
<u>Literatur</u>	Andreß & Lipsmeier (2001); Halleröd (1995); Nolan & Whelan (1996); Ringen (1988)

Deprivationsindex, gewichtet

<u>Variablenname</u>	<i>depindg2</i>
<u>Variablenlabel</u>	Deprivationsindex, gew. (n. aus fin. Gründ. fehl. Items; gew. Itemssumme: 13,14)
<u>Ausgangsvariablen</u>	<i>HLS0100a-HLS0400a; HLS0100b-HLS0400b; HLS0600a-HLS1200a; HLS0600b-HLS1200b; HLS1400a-HLS2500a; HLS1400b-HLS2500b; PLS0100-PLS0400; PLS0600-PLS1200; PLS1400-PLS2500;</i>
<u>Typ / Datensatz</u>	Alle Wellen: Deprivationsindex, gewichtet (Itemssumme: 11.08)
<u>Bearbeitung</u>	Bernhard Christoph
<u>Erläuterung</u>	<p>Für eine generelle Darstellung: vgl. Deprivationsindex, ungewichtet</p> <p>Ein Problem, welches häufig an ungewichteten Indizes, wie dem oben dargestellten kritisiert wird, ist die identische Gewichtung aller in den Index eingehenden Items. Vergleicht man z. B. zwei Items wie die Frage nach dem Vorhandensein einer Toilette innerhalb der Wohnung oder nach einem Videorecorder bzw. DVD-Player, so wird unmittelbar deutlich, dass mit dem Fehlen dieser beiden Items jeweils sehr unterschiedliche Einschränkungen des Lebensstandards einhergehen. Eine Gewichtung der einzelnen Items scheint daher konzeptionell sinnvoll, selbst wenn empirische Untersuchungen gezeigt haben, dass sich gewichtete und ungewichtete Indexvarianten im Ergebnis i.d.R. nur geringfügig unterscheiden (vgl. Lipsmeier 1999).</p> <p>Wir haben uns hier entschieden, die Gewichtung der Items über den Anteil der Befragten, die ein Item als notwendig erachten, vorzunehmen. Dieses Vorgehen wurde nicht nur ausgewählt, weil es ein konzeptionell überzeugendes und übliches Verfahren ist (es findet sich z. B. bei Halleröd 1995), sondern auch, weil es sich vergleichsweise sparsam umsetzen ließ. Da für die im Personenbogen zu erhebenden Deprivationsgewichte eine hohe Stabilität über die Zeit unterstellt werden kann, müssen sie nur einmalig bzw. in vergleichbar langen Intervallen erhoben werden. Zudem war es aufgrund der großen Stichprobe des PASS möglich, die Befragten in mehrere zufällig ausgewählte Subsamples aufzuteilen, denen jeweils nur ein Teil der Items vorgelegt wurde.</p> <p>Alternative Arten der Gewichtung, wie z. B. eine Beschränkung des Indizes auf die von mind. einem bestimmten Anteil der Befragten als notwendig erachteten Items (z. B. Andreß & Lipsmeier 1995, Andreß et al. 1996) oder eine theoretische Beschränkung auf wenige, als grundlegend angesehene Items (z. B. Nolan & Whelan 1996) wurden hier nicht umgesetzt, lassen sich aber ggf. auf Basis der ausgelieferten Daten generieren. Für eine zusammenfassende Diskussion unterschiedlicher Vorgehensweisen bei der Indexgewichtung vgl. Andreß & Lipsmeier (2001, insbes. S. 28ff.)</p> <p>Für die Wellen 1 bis 4 liegt mit der Variable <i>depindg</i> eine Version des gewichteten Deprivationsindex vor, der auf 26 statt 23 Items basiert, nämlich zusätzlich zu den oben genannten Items auch auf HLS0500*, HLS1300* und HLS2600* sowie PLS0500, PLS1300 und PLS2600. Die drei HLS-Items werden seit Welle 5 nicht mehr erhoben. Daher wurde <i>depindg2</i> neu in den Datensatz eingefügt und rückwirkend seit Welle 1 generiert.</p>
<u>Literatur</u>	Andreß & Lipsmeier (1995, 2001); Andreß et al. (1996); Halleröd (1995); Lipsmeier (1999); Nolan & Whelan (1996)

Haushaltstypologie

Variablenname

hhtyp

Variablenlabel

Haushaltstyp, generiert

Ausgangsvariablen

Haushaltsinformationen zu Alter und Verwandtschaftsbeziehungen

Typ / Datensatz

Haushaltsstruktur / Haushaltsdaten

Bearbeitung

Daniel Gebhardt

Erläuterung

Für die Typologisierung von Haushalten existiert eine Vielzahl von Varianten und Vorschlägen (vgl. z. B. Lengerer et al. 2005 für die Haushaltstypologie des Mikrozensus, Porst (1984) sowie Beckmann & Trometer 1991 für die Typologie des ALLBUS und Frick et al. (o.J.) für das SOEP). Die Haushaltstypologie des PASS orientiert sich an der letztgenannten Typologie. Die entscheidenden Differenzierungskriterien bilden dabei Partnerschaften sowie Anzahl und Alter der Kinder sowie das Vorhandensein von Generationenbeziehungen. Während im SOEP jedoch lediglich die Stellung aller Haushaltsmitglieder zum Haushaltsvorstand die Basis für die Generierung des Haushaltstyps bildet, werden im PASS die Angaben zu den Beziehungen zwischen allen Haushaltsmitgliedern untereinander für die Generierung verwendet. Daneben werden das Alter der Haushaltsmitglieder laut Haushaltsinterview und die Haushaltsgröße bei der Generierung mit einbezogen.

Definition der Beziehungen bei der Generierung des Haushaltstyps:

- Paare: Ehepaare; eingetragene Lebenspartnerschaften; unverheiratete Partnerschaften sowie Partnerschaften mit nicht weiter spezifiziertem Status (fehlender Wert in Nachfrage nach Art der Partnerschaft).
- Kind einer Person: leibliches Kind; Stief-, Adoptiv- oder Pflegekind; Kind mit nicht weiter spezifiziertem Status (fehlender Wert in Nachfrage nach Art der Beziehung zum Kind).
- Elternteil einer Person: leibliches Elternteil; Stief-, Adoptiv- und Pflegeelternteil; Eltern mit nicht weiter spezifiziertem Status (fehlender Wert in Nachfrage nach Art der Elternschaft).

Definition der Haushaltstypen:

- 1-Personen Haushalt: Haushalt mit nur einer Person im Haushalt.
- Paar ohne Kinder: Haushalt besteht ausschließlich aus zwei in Partnerschaft lebenden Personen.
- 1-Eltern-Haushalt: Haushalt besteht ausschließlich aus einem Elternteil und dessen Kindern. Hinsichtlich des Alters der Kinder werden keine Einschränkungen vorgenommen.
- Paar mit Kindern < 16 Jahre: Haushalt besteht ausschließlich aus zwei in Partnerschaft lebenden Personen und deren jeweiligen bzw. gemeinsamen Kindern. Alle Kinder sind unter 16 Jahre alt.
- Paar mit Kindern >= 16 Jahre: Haushalt besteht ausschließlich aus zwei in Partnerschaft lebenden Personen und deren jeweiligen bzw. gemeinsamen Kindern. Alle Kinder sind 16 Jahre oder älter.
- Paar mit Kindern < 16 Jahre und >= 16 Jahre: Haushalt besteht ausschließlich aus zwei in Partnerschaft lebenden Personen und deren jeweiligen bzw. gemeinsamen Kindern. Es leben sowohl Kinder unter 16 Jahre als auch im Alter von 16 Jahren oder älter im Haushalt.
- Mehrgenerationenhaushalt: Haushalt besteht aus Angehörigen mindestens dreier linear aufeinanderfolgender Generationen. Der Haushalt besitzt einen Mehrgenerationenkern, d.h. mindestens eine Person im Haushalt ist sowohl Kind als auch Elternteil einer anderen Person im Haushalt. Ausgehend von den zu diesem Mehrgenerationenkern gehörenden Personen leben im Haushalt ausschließlich Eltern, Kinder, Geschwister, Partner sowie Geschwister der Partner.
- Anderer Haushaltstyp: Haushalt, der keinem der anderen definierten Haushaltstypen zugeordnet werden konnte.
- Generierung nicht möglich (fehlende Werte): Grundsätzlich alle Haushalte mit mindestens einem fehlenden (-1, -2, -4) oder unplausiblen Wert (-8) in der Oberkategorie in einer Beziehungsvariable oder bei der Altersangabe (Ausnahme: In Haushalten mit drei und weniger Mitgliedern wurde in eindeutigen Konstellationen auch bei fehlenden Altersangaben der Haushaltstyp gebildet).

Literatur

Beckmann & Trometer (1991); Frick et al. (o.J.); Lengerer et al. (2005); Porst (1984)

Bedarfsgemeinschaftsnummer, Welle 5

<u>Variablenname</u>	<i>bgnr5</i>
<u>Variablenlabel</u>	Nummer der Bedarfsgemeinschaft in Welle 5
<u>Ausgangsvariablen</u>	Haushaltsinformationen zu Alter und Verwandtschaftsbeziehungen
<u>Typ / Datensatz</u>	Bedarfsgemeinschaft / Personenregister
<u>Bearbeitung</u>	Gerrit Müller
<u>Erläuterung</u>	<p>Die Variable <i>bgnr5</i> wird auf Personenebene gebildet und ordnet jedem Haushaltsmitglied eine Nummer zu, welche die Zugehörigkeit zu einer bestimmten Bedarfsgemeinschaft anzeigt. Das heißt, Haushaltsmitglieder mit gleicher Nummer bilden gemeinsam eine BG. Die Variable <i>bgnr5</i> setzt sich aus der bekannten Haushaltsnummer sowie einem zweistelligen BG-Indikator zusammen.</p> <p>Zur Identifikation der BG-Zugehörigkeit werden ausschließlich Angaben zu den Verwandtschaftsbeziehungen zwischen den verschiedenen Haushaltsmitgliedern aus dem HH-Grid sowie zum Alter der Personen laut Haushaltsinterview verwendet. Die so identifizierten Bedarfsgemeinschaften sind daher als „synthetische“ BGs zu verstehen. Angaben zu tatsächlichem Leistungsbezug, Erwerbsfähigkeit bzw. Ausbildungsstatus der einzelnen Personen werden bei der Bildung der BGs nicht berücksichtigt. Vielmehr geht es um die Identifikation von Personengruppen innerhalb eines Haushalts, die im Falle einer Bedürftigkeit gemeinsam eine Bedarfsgemeinschaft im Sinne des SGB II bilden bzw. bilden würden. Diese künstliche Zuordnung ist notwendig, da das Konzept der Bedarfsgemeinschaft (BG) und der entsprechenden Zugehörigkeit von Personen zu dieser im Rahmen einer Befragung nicht direkt zu erheben ist.</p> <p>Inhaltlich erfolgt die Zuordnung von Personen zu Bedarfsgemeinschaften auf Basis der aktuell gültigen Fassung des § 7 Abs. 3 SGB II (zuletzt geändert am 26. März 2007). Danach bilden grundsätzlich alle Personen, die das 25. Lebensjahr vollendet und das 65. Lebensjahr noch nicht vollendet haben, jeweils eine eigene BG; es sei denn, diese Personen leben in einer Partnerschaft und/oder haben ein Kind/Kinder, das/die wiederum selbst das 25. Lebensjahr noch nicht vollendet hat/haben und ohne Partner bzw. eigene Kinder sind. In letzterem Falle bilden Personen jeweils gemeinsam mit ihrem Partner und Kind(ern) eine BG. Sollten zwei Personen mit einem gemeinsamen Kind im gleichen Haushalt leben, aber im HH-Grid nicht angeben, Partner voneinander zu sein, wird dies im Sinne des §7 Abs. (3a) trotzdem unterstellt und die entsprechenden Personen gemeinsam mit dem Kind/Kindern zu einer BG zugeordnet. Personen die das 15. Lebensjahr vollendet und das 25. Lebensjahr noch nicht vollendet haben werden grundsätzlich ihren Eltern zugeordnet, es sei denn sie leben bereits mit einem Partner (bzw. einem eigenen Kind) gemeinsam im Haushalt. Sollten Personen zwischen 15 und 25 Jahre ohne Eltern (oder Partner bzw. Kind) leben, so bilden sie jeweils eine eigene BG.</p> <p>Personen, die das 65. Lebensjahr vollendet haben, fallen nicht in den Geltungsbereich des SGB II und gehören daher keiner BG an (Code 0); es sei denn, sie leben gemeinsam mit einem Partner, der das 65. Lebensjahr (bzw. einem Kind, welches das 25. Lebensjahr) noch nicht vollendet hat in einem Haushalt. Kinder, die das 15. Lebensjahr noch nicht vollendet haben und ohne Eltern in einem Haushalt leben, gehören ebenfalls keiner BG an (Code 0). Sie fallen in den Geltungsbereich des SGB XII. Für Haushalte mit fehlenden Beziehungs- bzw. Altersangaben zu einzelnen Personen wurde keine BG-Zuordnung vorgenommen, sondern für alle Haushaltsmitglieder Code 99 vergeben. Diese Haushalte können approximativ als Ein-BG-Haushalte interpretiert werden.</p>
<u>Literatur</u>	Sozialgesetzbuch, Zweites Buch - Grundsicherung für Arbeitsuchende (SGB II)

Bedarfsgemeinschaftstypologie, Welle 5

<u>Variablenname</u>	<i>bgtyp5</i>
<u>Variablenlabel</u>	Typ der Bedarfsgemeinschaft in Welle 5
<u>AusgangsvARIABLEN</u>	Haushaltsinformationen zu Alter und Verwandtschaftsbeziehungen
<u>Typ / Datensatz</u>	Bedarfsgemeinschaft / Personenregister
<u>Bearbeitung</u>	Gerrit Müller
<u>Erläuterung</u>	<p>Der Bedarfsgemeinschaftstypologie liegt das gleiche Konzept der synthetischen Bedarfsgemeinschaft wie bei der Variable <i>bgnr5</i> zugrunde. Kinder gehören bis zur Vollendung des 25. Lebensjahres der BG ihrer Eltern an, sofern sie nicht selbst Partner oder Kind haben. Dies wird hier anders gehandhabt als in der BA-Statistik, wo oftmals noch Typisierungen auf Basis der Volljährigkeit (18. Lebensjahr) ausgewiesen werden. Alleinerziehenden-BGs, im Sinne unserer Typisierung, in denen das jüngste Kind zwischen 18 und 24 Jahre alt ist, werden in der BA-Statistik beispielsweise als alleinstehend ausgewiesen. Dieser Unterschied ist bei Vergleichen mit Zahlen der offiziellen Statistik unbedingt zu berücksichtigen.</p> <p>Code 0, keine BG, wurde vergeben, wenn eine oder mehrere Personen nicht in den Geltungsbereich des SGB II fallen (vgl. Code 0 bei <i>bgnr5</i>). Code -5, Generierung nicht möglich (fehlende Werte), wurde für Haushalte mit fehlenden Beziehungs- bzw. Altersangaben zu einzelnen Personen (vgl. Code 99 bei <i>bgnr5</i>) vergeben.</p>
<u>Literatur</u>	–

Arbeitslosengeld-II-Bezug der Bedarfsgemeinschaft zum Zeitpunkt der Stichprobenziehung, Welle 5

<u>Variablenname</u>	<i>bgbezs5</i>
<u>Variablenlabel</u>	ALG-II-Bezug der Bedarfsgemeinschaft bei Stichprobenziehung in Welle 5
<u>AusgangsvARIABLEN</u>	<i>HA0250*</i> , <i>HA0300</i> , <i>AL20100</i> , <i>AL20200</i> , <i>AL20300</i> , <i>AL20400</i> , <i>AL20604</i> , <i>AL20704*</i> , <i>HA0400</i> , <i>sample</i> , <i>hnr</i> , <i>bgnr5</i> , <i>hhgr</i>
<u>Typ / Datensatz</u>	Bedarfsgemeinschaft / Personenregister
<u>Bearbeitung</u>	Mark Trappmann
<u>Erläuterung</u>	Diese Variable gibt für jede Bedarfsgemeinschaft, die entsprechend dem bei Variable <i>bgnr5</i> beschriebenen Vorgehen gebildet wurde, an, ob diese Bedarfsgemeinschaft zum Zeitpunkt der Stichprobenziehung für Welle 5 Arbeitslosengeld II bezogen hat oder nicht.
<u>Literatur</u>	–

Arbeitslosengeld-II-Bezug der Bedarfsgemeinschaft zum Befragungszeitpunkt, Welle 5

<u>Variablenname</u>	<i>bgbezb5</i>
<u>Variablenlabel</u>	ALG-II-Bezug der Bedarfsgemeinschaft zum Befragungszeitpunkt in Welle 5 (2010)
<u>AusgangsvARIABLEN</u>	<i>AL20604</i> , <i>AL20704</i> , <i>zensiert (alg2_spells)</i> , <i>sample</i> , <i>hhgr</i> , <i>bgnr5</i>
<u>Typ / Datensatz</u>	Bedarfsgemeinschaft / Personenregister
<u>Bearbeitung</u>	Daniel Gebhardt
<u>Erläuterung</u>	Diese Variable gibt für jede Bedarfsgemeinschaft, die entsprechend dem bei Variable <i>bgnr5</i> beschriebenen Vorgehen gebildet wurde, an, ob diese Bedarfsgemeinschaft zum Befragungszeitpunkt in Welle 5 Arbeitslosengeld II bezogen hat oder nicht.
<u>Literatur</u>	–

Anzahl der Bedarfsgemeinschaften im Haushalt

<u>Variablenname</u>	<i>anzbg</i>
<u>Variablenlabel</u>	Anzahl der synthetischen Bedarfsgemeinschaften im HH, generiert
<u>Ausgangsvariablen</u>	<i>bgnr5, hnr</i>
<u>Typ / Datensatz</u>	Bedarfsgemeinschaft / Haushaltsdatensatz
<u>Bearbeitung</u>	Daniel Gebhardt
<u>Erläuterung</u>	Diese Variable gibt die Anzahl Bedarfsgemeinschaften an, die innerhalb des Haushalts existieren. Die Identifizierung der Bedarfsgemeinschaften erfolgte dabei entsprechend der bei den Erläuterungen zur Variable <i>bgnr5</i> beschriebenen Vorgehensweise.
<u>Literatur</u>	–

Anzahl der zum Stichprobenzeitpunkt Leistungen beziehenden Bedarfsgemeinschaften im Haushalt

<u>Variablenname</u>	<i>nbgbezug</i>
<u>Variablenlabel</u>	Anzahl zum Stichprobenzeitpunkt beziehender BGs im HH
<u>Ausgangsvariablen</u>	<i>bgbezs5, bgnr5, hnr</i>
<u>Typ / Datensatz</u>	Bedarfsgemeinschaft / Haushaltsdatensatz
<u>Bearbeitung</u>	Daniel Gebhardt
<u>Erläuterung</u>	Diese Variable gibt die Anzahl der Bedarfsgemeinschaften innerhalb des Haushalts an, die zum Zeitpunkt der Stichprobenziehungen SGB-II-Leistungen bezogen haben. Der Wert wurde durch Aggregation der laut der Variablen <i>bgbezs5</i> aus dem Personenregister beziehenden Bedarfsgemeinschaften über die Haushaltsnummer gebildet.
<u>Literatur</u>	–

5 Datenaufbereitung

Die Datenaufbereitung wird seit Welle 3 nicht vom IAB selbst, sondern von infas durchgeführt. Um die Konsistenz der Datenaufbereitung im Längsschnitt zu gewährleisten, wurden infas die relevanten Syntaxfiles der Datenaufbereitung aus Welle 2 zusammen mit den notwendigen Ausgangs- und Zwischendatensätzen sowie einer Dokumentation der Arbeitsschritte zur Verfügung gestellt. Wichtige Entscheidungen, bspw. zur Bereinigung von Strukturproblemen in den befragten Haushalten oder zur Entwicklung des in Welle 4 erstmals erstellten *bio_spells*-Datensatzes, wurden in Abstimmung mit dem IAB getroffen. Auch darüber hinaus stand das IAB für die Beantwortung von Fragen während der Zeit der Datenaufbereitung zur Verfügung.

Die in den Interviews der Welle 5 erhobenen Informationen liegen bei infas zunächst als ASCII-Daten vor. In einer ersten Stufe erstellte infas aus diesen Rohdaten zunächst die folgenden Datensätze³²:

- Haushaltsdatensatz für im Querschnitt erhobene Fragen
- Haushaltsdatensatz für im Längsschnitt erhobene Daten (Modul „Arbeitslosengeld 2“)
- Datensatz über die Fortschreibung der Haushaltszusammensetzung (Matrix)
- Datensatz über die Fortschreibung der Verwandtschaftsbeziehungen im Haushalt (Beziehungsmatrix)
- Personen-/Senioren datensatz für im Querschnitt erhobene Fragen in Welle 5 inklusive der später in Spellform umgebrochenen Fragen des Vignettenmoduls
- Personendatensatz für im Längsschnitt erhobene Daten I (Modul „Erwerbsbiografie [Spells]“)
- Personendatensatz für im Längsschnitt erhobene Daten II (Modul „Maßnahmen“)
- Datensatz für offene Texte (übergreifend über alle Haushalts-, Personen- und Seniorenbefragungen)

In einer zweiten Stufe werden die Daten einer ausführlichen formalen und inhaltlichen Prüfung unterzogen und zum Scientific Use File aufbereitet. Darüber hinaus liefert infas einen Bruttodatensatz sowie weitere Spezialdatensätze, die sich jedoch nicht direkt aus den eigentlichen Erhebungsinstrumenten ergeben.

Die sich anschließende Datenprüfung bei infas unterteilte sich in drei Schritte, die in den folgenden Abschnitten genauer dargestellt werden. Zunächst wurde in den wiederbefragten Haushalten die Haushaltsstruktur überprüft und ggf. Bereinigungen vorgenommen. Bei schweren Strukturproblemen wurden die entsprechenden Interviews entfernt (siehe dazu Kapitel 5.1). Im Anschluss erfolgte eine detaillierte Überprüfung und ggf. Bereinigung der Filterführung. Dabei wurden einerseits Filterfehler markiert und andererseits spezifische Codes für Missing Values gesetzt (siehe dazu Kapitel 5.2). Im Anschluss wurden bei ausgewählten Items inhaltliche Plausibilitätsprüfungen vorgenommen. Eindeutig unplausible bzw. widersprüchliche Angaben wurden durch einen spezifischen Missing Code als solche gekennzeichnet, wobei eine derartige Bereinigung der Daten allerdings sehr restriktiv gehandhabt wurde.

³² Im Rahmen der Datenaufbereitung zum Einsatz kamen die Softwarepakete Stata Version 11 und PASW Version 18.

Die folgende Tabelle gibt eine Übersicht über alle im Rahmen der Datenaufbereitung durchgeführten Arbeitsschritte und ihre Abfolge:

Tabelle 21: Übersicht über die Schritte der Datenaufbereitung in der Welle 5 des PASS

Nr.	Arbeitsschritt
1	Einlese der erhobenen Rohdaten in Arbeitsdatensätze
2	Prüfung der Haushaltsstruktur (siehe Kapitel 5.1)
3	Entfernen problematischer Interviews (Haushalts- und/oder Personenebene) (siehe Kapitel 5.1)
4	Integration Personen- und Seniorendatensatz
5	Korrektur der Haushaltsstruktur bei wiederbefragten Haushalten (siehe Kapitel 5.1)
6	Filterbereinigung Haushaltsebene (siehe Kapitel 5.2)
7	Erstellung und Plausibilitätsprüfung eines Haushaltsgriddatensatzes (siehe Kapitel 5.3)
8	Generierung der synthetischen Bedarfsgemeinschaften (siehe Variablenbeschreibung 4.5)
9	Neugenerierung von Steuerungsvariablen auf Basis der filtergeprüften Haushaltsdaten und des plausibilitätsgeprüften Haushaltsgriddatensatzes
10	Filterbereinigung Personenebene (siehe Kapitel 5.2)
11	Zucodung offener Angaben (siehe Kapitel 4.1)
12	Plausibilitätsprüfung der Haushalts- und Personendaten (ohne Spelldaten) (siehe Kapitel 5.3)
13	Aufbereitung, Plausibilitätsprüfung und Erstellung der Spelldatensätze (siehe Kapitel 5.6 bis 5.8 sowie auch Kapitel 5.3)
14	Einfache Generierungen (siehe Kapitel 4.4)
15	Komplexe Generierungen (siehe Kapitel 4.5)
16	Generierung der Datenstruktur für Scientific Use File (Haushalts-, Personen-, Registerdatensätze)
17	Anonymisierung (siehe Kapitel 5.5)

5.1 Strukturprüfungen und entfernte Interviews

Der Filterbereinigung vorgelagert fand eine Strukturprüfung statt. Dabei sollten zunächst Interviews identifiziert werden, die im Sinne des PASS als nicht realisiert gelten und die aus diesem Grund ggf. aus den Datensätzen entfernt wurden. Darüber hinaus erfolgte ein Abgleich der Struktur der wiederbefragten Haushalte mit der Vorwelle, um unplausible oder problematische Veränderungen der Haushaltszusammensetzung bzw. Fehler bei der Zuweisung der Personeninterviews erkennen und ggf. korrigieren zu können. Für die Betrachtung der Haushalte im Längsschnitt ist die konsistente Zuordnung und eindeutige Identifizierbarkeit der Personen über die Wellen hinweg zentral. Eine feste Personennummer darf nicht in verschiedenen Wellen unterschiedlichen Personen zugeordnet sein. War die korrekte Haushaltszusammensetzung unklar, wurden alle Interviews der Welle 5 dieses Haushalts entfernt. Wurde auf der Personenebene ein Interview mit der falschen Person geführt, ohne dass weiterführende Probleme in der Haushaltszusammensetzung erkennbar waren, so wurde lediglich das Personeninterview entfernt.

Um problematische Fälle zu identifizieren, wurden verschiedene Prüfungen durchgeführt. Die entsprechenden Fälle wurden in einem formalisierten Verfahren zwischen infas und dem IAB diskutiert und abschließend wurde eine endgültige Entscheidung im IAB getroffen, wie mit diesen Fällen zu verfahren ist. Es sollte beachtet werden, dass die folgende Darstellung den Umfang der durchgeführten Prüfungen angibt. Es führen nicht alle Prü-

fungen in allen Wellen zur Identifikation von Problemen, sondern das Ergebnis einer Prüfung besteht üblicherweise darin, dass ein geprüfter Sachverhalt in geringer Fallzahl oder gar nicht auftritt. Zudem werden bekannte Fehlerquellen möglichst schon während des Interviews abgefangen. So sieht das Erhebungsinstrument beispielsweise vor, dass aus einem Panel-HH nicht alle bisher dort bekannten Zielpersonen gleichzeitig ausziehen können und dass unter den nach Auszügen verbleibenden Personen mindestens eine Person 15 Jahre oder älter sein muss.

- Durch einen Vergleich der Vornamen zwischen der aktuellen und der Vorwelle wurden Fälle erkannt, bei denen Veränderungen der Haushaltszusammensetzung nicht korrekt aufgenommen wurden. Statt Ein- und Auszüge an den entsprechenden Stellen des Haushaltsinterviews aufzunehmen, kam es vor, dass Interviewer Haushaltsmitglieder umbenannten sowie Veränderungen an Alter oder Geschlecht der Personen vornahmen. Einzelfallprüfungen wurden vorgenommen, wenn sich Veränderungen an den Vornamen nicht auf Korrekturen der Schreibweise zurückführen ließen, wenn sich das in der Vorwelle berichtete Geburtsjahr um mehr als ein Jahr von dem in der aktuellen Welle berichteten unterschied oder wenn das Geschlecht nicht mit der aus der Vorwelle bekannten Angabe übereinstimmte. Dabei wurde entschieden, ob es sich bei der Veränderung lediglich um eine Korrektur des Vornamens, des Alters oder des Geschlechts handelte oder aber, ob der Interviewer eine unzulässige Veränderung der Haushaltsstruktur vorgenommen hatte.
- Weiterhin wurde geprüft, ob mehrere Personen mit demselben Geburtsdatum im Haushalt leben. Im Haushaltskontext der beiden Wellen wurde hier jeweils entschieden, ob es sich um plausible oder unplausible Fälle handelte. Anschließend wurden die übrigen Fälle einer weiteren Prüfung unterzogen. Dabei wurden Haushalte identifiziert, bei denen ein Geburtsdatum in der aktuellen Welle und in der Vorwelle von Personen auf unterschiedlichen Positionen der Haushaltsstruktur berichtet wurde. Hier lag die Vermutung nahe, dass das entsprechende Personeninterview in der aktuellen Welle von einer anderen Person als in der Vorwelle geführt wurde. Im Kontext der Haushalts- und Personeninformationen der aktuellen und der Vorwelle wurden jeweils Einzelfallentscheidungen in Bezug auf die jeweiligen Haushalts- und Personeninterviews getroffen.
- Generell gilt, dass das Geburtsdatum aus dem Personen-/Senioreninterview der aktuellen Welle alle anderen Altersangaben zu dieser Person z. B. aus dem Haushaltsgrid sticht und Basis aller Generierungen ist, die u.a. auf dem Alter basieren. In einer speziellen Konstellation wird die Angabe des Geburtsdatums in *PD0100* jedoch korrigiert. Wenn sich das Geburtsjahr einer Person gemäß *PD0100* deutlich ändert, die Angaben zum Tag und Monat jedoch unverändert bleiben, das bisher bekannte Geburtsdatum laut *PD0100* sich noch nie geändert hat und mindestens zwei Angaben zum Geburtsdatum aus *PD0100* aus Vorwellen vorliegen, dann wird das Geburtsjahr unter Berücksichtigung der gesamten Haushaltskonstellation auf den aus den Vorwellen bekannten Wert zurückgesetzt. Ein theoretisches Beispiel ist eine Person, deren Geburtsdatum aus mindestens zwei Vorwellen als 01.02.1972 bekannt ist, deren Geburtsdatum nun als 01.02.1992 erfasst wird und die damit nun jünger wäre als die ebenfalls im Haushalt lebenden Kinder. Eine solche Konstellation würde ohne Korrektur zu einer

Unplausibilität in der Beziehungsstruktur führen und in der Konsequenz dazu, dass z. B. die synthetischen Bedarfsgemeinschaften nicht generiert werden können. Daher wird die Angabe im Beispielfall in der aktuellen Welle auf den 01.02.1972 korrigiert.

- Um Haushalte zu identifizieren, die im Sinne des PASS als nicht realisiert gelten, wurden die Datensätze auf Haushalts- und Personenebene zusammengespielt. Personeninterviews ohne vollständiges Haushaltsinterview wurden ebenso markiert wie Haushaltsinterviews, für die kein Interview auf Personenebene vorlag³³.
- Breiten Raum nehmen auch die Prüfungen von Ein- und Auszügen ein. Panelhaushalte, in denen Auszüge berichtet wurden, wurden grundsätzlich auf ihren Haushaltskontext sichtgeprüft und mit den realisierten abgesplitteten Haushalten in Bezug gesetzt. Es wurde geprüft, ob der verbleibende Haushaltskontext des Panelhaushalts lebensweltlich plausibel ist. Interviews aus Panelhaushalten, in denen alle Haushaltspersonen den Haushalt verlassen, bis auf einzelne Kinder im Alter von unter 15 Jahren, wurden sowohl mit Blick auf den Panelhaushalt als auch mit Blick auf die abgesplitteten Haushalte verworfen. Auch wurde beim Auszug von mehr als einer Person geprüft, ob die Ausziehenden einen gemeinsamen Splithaushalt bilden oder mehrere verschiedene und ob das jeweils plausibel ist. Als unplausibel betrachtet wurden Fälle, in denen ein Partner zusammen mit jungen Kindern den Panelhaushalt verlässt, die Ausziehenden aber laut Feldinformationen mehrere verschiedene Splithaushalte bilden, sprich die jungen Kinder angeblich eigene Haushalte bilden. Hier wurde im Falle der Nichtrealisierung des Splithaushalts entschieden, die Auszüge als plausibel anzuerkennen, alle Ausgezogenen jedoch nachträglich in einem gemeinsamen Splithaushalt zusammenzuführen.
- In Einzelfällen wurde beobachtet, dass laut Befragung im Panelhaushalt einzelne Personen einen abgesplitteten Haushalt bilden, sich im Splithaushalt dann aber alle Mitglieder des Panelhaushalts wiederfinden. In einer abgewandelten Variante finden sich im abgesplitteten Haushalt nicht alle Mitglieder des Panelhaushalts wieder, jedoch zumindest eine Person des Panelhaushalts, die im dortigen Interview als nicht ausgezogen bzw. als in einen anderen Splithaushalt als den betrachteten verzogen berichtet wurde. Auch hier wurde differenziert entschieden, welche berichteten Auszüge als gültig anerkannt und welche als unplausibel verworfen wurden. Wenn ein berichteter Auszug nachträglich als unplausibel verworfen wurde, wurde die angeblich ausgezogene Person nachträglich wieder in den Haushaltskontext des Panelhaushalts integriert.
- In Splithaushalten wird bei Personen, die laut Feld nicht aus dem Panelhaushalt bekannt sind, sondern über den Splithaushalt neu zum PASS hinzukommen, geprüft, ob sie nicht doch aus dem Panelhaushalt stammen. Es bestehen zwei Konstellationen, die solche Fälle begünstigen. Zum einen kommt es vor, dass im Panelhaushalt bei mehreren ausziehenden Personen berichtet wird, dass die Splitpersonen mehr als einen Splithaushalt bilden. In diesem Fall wird dynamisch für das laufende Feld

³³ Bei Neubefragten Haushalten, für die zwar ein Haushalts-, aber kein gültiges Personeninterview vorlag, wurden die Haushaltsinterviews analog zum Vorgehen in der Welle 1 entfernt. Bei wiederbefragten und abgespaltenen Haushalten wurden die Haushaltsinterviews dagegen beibehalten.

Preload für alle durch den Panelhaushalt benannten Splithaushalte erzeugt. Wenn jedoch tatsächlich Personen, die laut Panelhaushalt in verschiedenen Splithaushalten leben, in einem gemeinsamen Splithaushalt angetroffen werden, haben jene Personen, die durch den Panelhaushalt nicht diesem, sondern einem anderen Splithaushalt zugerechnet wurden, in diesem Splithaushalt keinen Preload und werden als neue Person erfasst.

- Zum anderen ist es möglich, dass Personen aus einem Panelhaushalt aus- oder in einen Haushalt einziehen, der in einer Vorwelle als Splithaushalt entstand und seinerzeit bereits realisiert wurde. Es findet also nach der damaligen Abspaltung des Splithaushalts ein weiterer Umzug aus dem ursprünglichen Panelhaushalt in diesen Splithaushalt statt. Unabhängig davon, ob der Panelhaushalt, aus dem der betroffene Splithaushalt hervorgegangen ist, in der Welle des erneuten Umzugs vom Panel- in den Splithaushalt realisiert wurde, ist es feldseitig nicht vorgesehen, solche Fälle steuern zu können. Hierzu müssten dem Splithaushalt die Personeninformationen aller Personen des Panelhaushalts (und ggf. aller Personen in weiteren Splithaushalten dieses Panelhaushalts) als Preload hinterlegt werden. Die wenigen zu erwartenden Fälle, in denen eine solche Konstellation eintritt, rechtfertigen einen solchen Aufwand im Feld nicht. Statt dessen müssen solche Fälle in den Strukturprüfungen entdeckt werden. In diesem Zusammenhang ist anzumerken, dass Splithaushalte in Bezug auf die Strukturprüfungen auch in den Folgewellen ihrer Erstrealisierung in ihrer Eigenschaft als Splithaushalt betrachtet werden müssen, auch wenn sie nach der Erstrealisierung in der Feldsteuerung als Panelhaushalte gelten.

In beiden Fällen wird die PNR der betroffenen Personen im Splithaushalt nachträglich korrigiert. Zu beachten ist auch, dass solche Personen in ihren Personen-/Senioreninterviews als Neubefragte gesteuert werden, obwohl sie u.U. altbefragte Personen sind. Diese Abweichung wird grundsätzlich nicht korrigiert, siehe hierzu auch Kapitel 4.4.

- In Panelhaushalten, die ab Welle 2 von einem Auszug berichtet haben, kann es ab Welle 3 auch zu Wiedereinzügen von früher bereits einmal dem Haushalt angehörigen Personen kommen. Die Anforderung, diese Einziehenden als wiedereinziehend zu erkennen und wieder auf ihrer früheren Haushaltsposition einzuordnen statt ihnen eine neue, bisher noch nicht vergebene Haushaltsposition zuzuweisen, ist Bestandteil des Haushaltsgrids. Es wurde nachträglich geprüft, ob diese Anforderung in allen Fällen im Feld erfüllt worden ist. Bei Personen, die in der aktuellen Welle nachträglich anhand eines Abgleichs von Vorname, Alter und Geschlecht mit den zuvor aus dem Haushalt Ausgezogenen als wiedereinziehende Person identifiziert wurden, musste in die Haushaltsstruktur eingegriffen werden. Notwendig waren dann ein nachträgliches Ändern der Personennummer der zu verschiebenden Person wie auch ein Verschieben der personenbezogenen Angaben im Haushaltsinterview z. B. zur Kinderbetreuung oder zu den Gründen der ALG-II-Kürzung auf die im Rahmen der Strukturprüfung als richtig definierten Position. Umgekehrt wird auch geprüft, ob Personen, die im Feld als wiedereinziehend markiert werden, tatsächlich mit der in einer Vorwelle ausgezogenen Person identisch sind. Falls nicht, handelt es sich um den Einzug einer neu zum PASS hinzukommenden Person. Auch dann werden die beschriebenen Änderungen in der Haushaltsstruktur vorgenommen.

- Bei Wiedereinzügen wird geprüft, ob der Splithaushalt, in dem die Person vor dem Wiedereinzug in den Panelhaushalt lebte, in der aktuellen Welle realisiert wurde und ob in diesem Fall die wiedereinziehende Person im Splithaushalt als ausgezogen bezeichnet wird. Auch Personen, die in einer Vorwelle in ihren Panelhaushalt zurückgezogen sind, müssen weiterhin mit Blick auf ihren Status im Splithaushalt überprüft werden, solange der Splithaushalt zur aktuellen Panelstichprobe zählt. Wenn eine wiedereinziehende Person weiterhin in ihrem Splithaushalt als aktuelles Haushaltsmitglied geführt wird, wurde in der Datenaufbereitung für die betroffenen Fälle entschieden, ob dies als plausibel anerkannt wurde oder ob die Haushaltsstruktur des Panel- oder des Splithaushalts bereinigt werden musste.
- Nicht nur Wiedereinzüge können dazu führen, dass Personen als aktuelles Haushaltsmitglied mehrerer Haushalte gelten. Es kann auch vorkommen, dass eine Person als Mitglied eines Splithaushalts benannt wird, obwohl sie im Panelhaushalt nicht als ausgezogen benannt wurde. Unter Prüfung der Haushaltskonstellationen der beteiligten Haushalte kann dies im Einzelfall als plausibel anerkannt werden. Solche Fälle werden in den *zdub**-Variablen im Personenregister dokumentiert. Weitere Erläuterungen hierzu sind in Kapitel 4.4 und Kapitel 5.4.1.2 enthalten.
- Im Verhältnis zwischen einem Panelhaushalt und seinen Splithaushalten können weitere Punkte auftreten. Es ist z. B. denkbar, dass Personen, die über einen Splithaushalt neu zum PASS hinzugekommen sind, in den Panelhaushalt umziehen. Oder es ist denkbar, dass Personen von einem Splithaushalt in einen anderen umziehen. Grundsätzlich müssen alle Personen in einem Panelhaushalt und alle mit ihm verbundenen Splithaushalte als Netzwerk betrachtet werden. Die Strukturprüfungen sind so ausgelegt, dass Personenbewegungen zwischen den Haushalten eines solchen Netzwerks entdeckt werden, unabhängig von der Richtung, in der sich eine Person im Netzwerk bewegt.
- Haushaltsstrukturprüfungen prüfen im Wesentlichen nicht die Haushaltsstruktur als solche auf Plausibilität, sondern sie betrachten die Veränderungen zwischen den Wellen. Daher kann die Haushaltsstruktur erstbefragter Haushalte nur in geringem Umfang geprüft werden. Bei erstbefragten Haushalten wird anhand der Angaben zu Vorname, Alter und Geschlecht betrachtet, ob einzelne Haushaltsmitglieder mehrfach aufgelistet werden. In diesem Fall wird von den doppelt berichteten Personen nur die zuerst berichtete Haushaltsposition behalten, die weiteren Haushaltspositionen werden hingegen verworfen. Dies zieht ggf. weitere Eingriffe in die Haushaltsstruktur nach sich. Wenn z. B. in einem erstbefragten Haushalt mit laut Befragung vier Personen die Personen auf den Positionen 2 und 3 identisch sind, wird nicht nur Person 3 entfernt, sondern auch Person 4 nachträglich auf Position 3 verschoben. Denn in einem erstbefragten Haushalt mit X Haushaltsmitgliedern sollen immer die Positionen 1 bis X durchgehend und ohne Lücken besetzt sein. Wie auch bei den nachträglich als wiedereinziehend erkannten Personen erfordert ein nachträgliches Ändern der Personennummer der zu verschiebenden Person auch ein Verschieben der personenbezogenen Angaben im Haushaltsinterview.

- Durch die Rückmeldung eines Feldinterviewers wurde in Welle 4 ein Haushalt entdeckt, der doppelt in der Panelstichprobe enthalten war. Der Haushalt 10015439 ist als identischer Haushalt 15044862 bereits seit Welle 1 doppelt in der Stichprobe. Beide Haushalte wurden jeweils in Welle 1 und in Welle 3 realisiert und jeweils in Welle 2 nicht realisiert. In Welle 4 wurde der Haushalt 10015439 realisiert. Da dem CAPI-Interviewer dieses Points „beide“ Haushalte zugeteilt worden waren, ist die Dopplung aufgefallen. Die Haushaltszusammensetzung blieb in beiden Haushalten über die Wellen hinweg unverändert. Der in Welle 4 nicht realisierte Haushalt 15044862 wird aus der Welle 5-Stichprobe gelöscht. Es findet keine rückwirkende Entfernung der Dublette in den Wellen 1 bis 3 statt, da bspw. Auswirkungen auf die Gewichtung die Folge wären. Im *hh_register* ist der Dublettenhaushalt in Variable *hnettod4* mit dem Code 26 gekennzeichnet, der den Grund für die Nichtrealisierung transparent macht. Im *p_register* sind die Haushaltsmitglieder des Dublettenhaushalts in der Variable *pnettod4* entsprechend mit Code 56 gekennzeichnet.
- Auch beim Umgang mit den Fällen, die sich in den Strukturprüfungen als problematisch herausgestellt haben, wurden Einzelfallentscheidungen getroffen. Maßgeblich dabei war jeweils, als wie schwerwiegend das betroffene Problem angesehen wurde. In Fällen, in denen die korrekte Haushaltszusammensetzung in der Welle 5 unklar war, wurden alle Interviews der Welle 5 entfernt. Sie werden in der Welle 6 wie Haushalte behandelt, die in der Welle 5 nicht teilgenommen haben. Wurden in nachträglich entfernten Haushaltsinterviews Auszüge gemeldet, wurden auch die abgesplitteten Haushalte verworfen. Dies betraf sowohl die in der aktuellen Welle realisierten Interviews in diesen Splithaushalten als auch die Stichprobe der Folgewelle. Splithaushalte, die aus einem verworfenen Interview eines Panelhaushalts heraus entstanden sind, werden nachträglich als nicht entstanden gewertet und zählen nicht zur Panelstichprobe der Folgewelle. Bestand lediglich ein Zuordnungsproblem auf Personenebene, d.h. wurde vermutet, dass ein Personeninterview in der Welle 5 mit einer falschen Person geführt wurde, dann wurde nur das betroffene Personen- bzw. Senioreninterview entfernt. Handelte es sich um ein Strukturproblem, das keine schwerwiegenden Auswirkungen hatte und bspw. durch das Entfernen eines Personeninterviews behoben werden konnte, wurden zusätzliche Korrekturen der Vornamen, Alters- und Geschlechtsangaben auf Haushaltsebene vorgenommen. Die betroffene fehlerhafte Information wurde dann auf den letzten gültigen Wert aus der Vorwelle zurückgesetzt bzw. im Falle des Alters auf den Vorwellenwert + die Zahl der Jahre seit der letzten gültigen Realisierung in diesem Haushalt.

Darüber hinaus wurden alle Interviews von Personen entfernt, für deren Haushalt kein vollständiges Haushaltsinterview vorlag. Im umgekehrten Fall, d.h. bei Haushalten, für die kein Interview auf Personenebene vorhanden war, wurde zwischen wiederbefragten Haushalten und Haushalten der Auffrischungstichprobe unterschieden. Die als nicht realisiert geltenden Haushalte aus der Auffrischungstichprobe wurden analog zum Vorgehen in den Vorwellen entfernt. Bei wiederbefragten Haushalten ohne Interviews auf Personenebene wurde das Haushaltsinterview dagegen nicht gelöscht.

Einen Hinweis auf entfernte Interviews geben die *Nettovariablen* (*hnettok5*, *hnettod5*, *pnettok5*, *pnettod5*) in den Haushalts- und Personenregisterdatensätzen. Über die

entsprechenden Variablen im Haushaltsregister ist nachvollziehbar, für welche wiederbefragten Haushalte die Haushaltsinterviews nachträglich entfernt wurden. Mittels Nettovariablen des Personenregisters kann überprüft werden, in welchen Fällen lediglich einzelne Interviews auf Personenebene oder alle Interviews des Haushalts gelöscht wurden. Bei Haushalten aus der Auffrischungsstichprobe der Welle 5, für die nicht mindestens ein gültiges Haushalts- und Personeninterview vorliegt, sind gelöschte Interviews nicht in den Registerdatensätzen nachvollziehbar, da diese Haushalte nicht in die Datensätze aufgenommen wurden.

5.2 Filterbereinigung

In der Filterbereinigung wurde die in den jeweiligen Instrumenten vorgesehene Filterführung mithilfe eines Statistikprogramms nachvollzogen. Wurden bestimmte Fragen gestellt, obwohl der Wert der relevanten Filtervariable etwas anderes verlangt hätte (wurden also z. B. detaillierte Angaben zur Berufsausbildung erhoben, obwohl die/der Befragte angegeben hatte, über keinen Berufsabschluss zu verfügen), wurden die Variablen auf den Missing Code „-3“ (trifft nicht zu) gesetzt, den sie auch bei einer korrekten Ausführung der Filterführung erhalten hätten.³⁴ Des Weiteren kam es vor, dass in Einzelfällen Items nicht erhoben wurden, obwohl dies laut der entsprechenden Filtervariable notwendig gewesen wäre (z. B. wenn keine weiteren Angaben zur Berufsausbildung erhoben wurden, obwohl die/der Befragte angegeben hatte, eine solche Ausbildung gemacht zu haben). In diesen Fällen wurde der spezifische Missing Code „-4“ (Frage irrtümlich nicht gestellt) vergeben. Die Vergabe des Codes „-4“ kann auch Folge der in Kapitel 5.1 beschriebenen Haushaltsstrukturprüfungen sein. Wenn der Auszug einer Person nachträglich als unplausibel verworfen und die Person nachträglich als noch zu ihrem bisherigen Haushalt gehörend gewertet wird, dann bedeutet dies auch, dass personenbezogene Angaben zu diesen Personen im Haushaltsinterview nachträglich als irrtümlich nicht erhoben codiert werden müssen. Der Code „-4“ weist also nicht in allen Fällen auf ein Problem des Erhebungsinstruments hin. Wenn Code „-4“ bei einer Frage vergeben wird, die filterrelevant für Folgefragen ist, dann werden die Folgefragen ebenfalls mit „-4“ vercodet, sofern diese Folgefragen tatsächlich nicht erhoben wurden. Wurden Folgefragen hingegen dennoch erhoben, weil z. B. mehrere Filterfragen auf diese Folgefrage verweisen und eine andere Filterfrage die Folgefrage korrekt ausgelöst hat, bleibt der dort erhobene Wert erhalten.

Daneben wurden im Rahmen der Filterprüfung auch die vom Feldinstitut vergebenen Missing Codes sowie die System Missings durch für alle Variablen einheitliche Werte ersetzt. Die vergebenen Werte sind in Tabelle 22 im Überblick dargestellt. „-1“ und „-2“ sind die einheitliche Umcodierung der in der Befragung erhobenen „WN“- und „KA“-Werte. „-3“ ist der generelle „TNZ“-Code für aufgrund der Filterführung nicht gestellte Fragen. Der Code „-4“ wurde, wie bereits erläutert, vergeben, wenn eine Frage durch einen Filterfehler nicht gestellt wurde. Bei „-5“ bis „-7“ handelt es sich um fragespezifische Codes. Hierbei kann es sich sowohl um spezifische Missing Codes (z. B. „TNZ, steht dem Arbeitsmarkt nicht zur Verfügung“) als auch um Sonderkategorien für gültige Werte (z. B. eine Katego-

³⁴ Die Bereinigung erfolgte dabei, wie in solchen Fällen üblich, immer von den zuerst erhobenen zu den später erfragten Items.

rie für Einkommen oberhalb von € 99.999,- bei der offenen Einkommensangabe) handeln. Diese Codes wurden nur im Bedarfsfall vergeben.

Tabelle 22: Übersicht über die vergebenen Missing Codes

Code	Erläuterung
-1	„weiß nicht“
-2	„keine Angabe“
-3	„trifft nicht zu (Filter)“ (Frage wurde aufgrund der Filterführung nicht gestellt)
-4	„Frage irrtümlich nicht gestellt“ (Frage hätte jedoch gestellt werden sollen)
-5	Fragespezifischer Code Nr. 1, wurde nur bei Bedarf vergeben
-6	Fragespezifischer Code Nr. 2, wurde nur bei Bedarf vergeben
-7	Fragespezifischer Code Nr. 3, wurde nur bei Bedarf vergeben
-8	„unplausibler Wert“
-9	„Item in Welle nicht erhoben“
-10	„Item in Fragebogenversion nicht erhoben“ ³⁵

Der Wert „-8“ ist ein spezifischer Missing Code, der im Rahmen der Plausibilitätsprüfung vergeben wurde (zur Plausibilitätsprüfung vgl. Kapitel 5.3). Der Missing Code „-9“ wird seit der Welle 2 erstmals benötigt. Er wird immer dann vergeben, wenn ein Item in einer bestimmten Welle nicht erhoben wurde. Aufgrund der bereits dargestellten Aufbereitung der Daten im Long-Format erhalten also Variablen, die ab der Welle 2 in keiner Fragebogenversion mehr erfragt wurden, für die Beobachtungen dieser Welle den Wert „-9“. Ebenso verhält es sich mit Beobachtungen der Welle 1. Erstmals nach der Welle 1 erhobene Variablen werden für Beobachtungen der Wellen ohne Erhebung nachträglich mit „-9“ vercodet. Mit dem Code „-10“ können Unterschiede zwischen den Fragebogenversionen berücksichtigt werden, also zwischen Personen- und Seniorenbogen bzw. bis Welle 3 auch zwischen den beiden Versionen des Haushaltsfragebogens.

5.3 Plausibilitätsprüfungen

Im Rahmen der Plausibilitätsprüfung wurde eine umfangreiche Liste theoretisch möglicher Widersprüche in den Angaben der Befragten kontrolliert. Dazu wurde die Liste der in den Vorwellen durchgeführten Prüfungen für die aktuelle Welle angepasst und sinnvoll erweitert. Daneben wurden auch die Haushaltsstruktur sowie die Spelldaten plausibilitätsgeprüft – insbesondere auf unzulässige Überschneidungen innerhalb der einzelnen Spelldaten. Dabei wurden grundsätzlich nur die im Querschnitt der Welle 5 erhobenen Daten für die Prüfung herangezogen. Eine Prüfung im Längsschnitt, also zwischen den in der aktuellen und der Vorwelle gemachten Angaben, wurde nicht durchgeführt.

Im Einzelnen wurden die folgenden Schritte ausgeführt:

³⁵ Ab Welle 4 dient Code „-10“ ausschließlich der Unterscheidung zwischen Personen- und Seniorenfragebogen. Bis einschließlich Welle 3 wurde zusätzlich auch auf der Haushaltsebene zwischen erst- und wiederbefragten Haushalten unterschieden. Die Unterscheidung auf Haushaltsebene entfällt ab Welle 4 infolge der Zusammenfassung der bisher getrennten Fragebogenversionen zu einem einzigen gemeinsamen Haushaltsfragebogen.

1. Widerspruchsprüfung: Generell wurden Widersprüche nur dann korrigiert, wenn entweder die Unplausibilität als besonders gravierend bzw. der Eingriff als vergleichsweise mild einzustufen war. Letzteres galt z. B., wenn nur eine geringe Zahl von Fällen betroffen war oder weil nur ein Missing Code (z. B. „-3“) durch einen neuen („-8“) ersetzt wurde. Für die Bereinigung unplausibler Angaben wurden zwei Strategien verwendet: eine direkte Korrektur der unplausiblen Angaben oder die Vergabe eines spezifischen Missing Codes.
 - Eine Korrektur unplausibler Angaben wurde ausschließlich dann vorgenommen, wenn mit an Sicherheit grenzender Wahrscheinlichkeit eine Fehleingabe des Interviewers vorlag. Ein Beispiel ist die Angabe einer monatlichen Gesamtmiete von € 9.998,-. Hier wurde im Rahmen der Datenbereinigung davon ausgegangen, dass es sich um eine Fehleingabe des fünfstellig erhobenen Missing Codes „99998“ (WN) handelte. Diese und ähnliche Angaben wurden in die entsprechenden Missingkategorien umcodiert. Hätten die umcodierten Missingkategorien einen Filter bei Folgefragen ausgelöst, wie dies z. B. bei der kategorialen Abfrage von Einkommensangaben der Fall ist, so wurden die kategorialen Nachfragen nachträglich auf Code „-4“ (Frage irrtümlich nicht gestellt) umgesetzt.
 - Allerdings war es eher selten der Fall, dass sich ein Wert mit hinreichender Sicherheit als Fehleingabe erkennen ließ. In den meisten Fällen konnte lediglich ein Widerspruch zwischen zwei Angaben festgestellt werden, ohne jedoch spezifische Fehleingaben oder ähnliches identifizieren zu können, die zu der unplausiblen Angabe geführt hatten. Daher wurde in diesen Fällen keine Korrektur vorgenommen, sondern der spezifische Missing Value Code „-8“ vergeben. Ob der Code dabei an eine oder beide der am Widerspruch beteiligten Variablen vergeben wurde, ist im Einzelfall entschieden worden.
2. Plausibilitätsprüfung der Haushaltsstruktur: Auf Grundlage der im Haushaltsinterview erhobenen Information zu den Verwandtschaftsbeziehungen der Haushaltsmitglieder untereinander sowie der Angaben zu Alter, Geschlecht und Vorname wurde eine Plausibilitätsprüfung der Haushaltsstruktur durchgeführt. Im Vorfeld dieser Prüfung wurden die Beziehungsinformationen im Haushalt durch die im Personeninterview berichteten Partnerschaften ergänzt.
 - Um unplausible Haushaltskonstellationen zu identifizieren, wurden zunächst die Beziehungsinformationen mit den demografischen Angaben zu den einzelnen Haushaltsmitgliedern kombiniert. Für die in diesen Prüfungen als unplausibel identifizierten Haushalte wurden Einzelfallentscheidungen unter Einbeziehung der gesamten Haushaltsstruktur und anderer Interviewinformationen (z. B. zum Familienstand im Personeninterview) getroffen. Unplausible Beziehungskonstellationen wurden dabei als solche markiert („-8“) bzw. auf Basis hinzugezogener Informationen über den Haushaltskontext korrigiert, sofern es sich mit an Sicherheit grenzender Wahrscheinlichkeit um eine fehlerhafte Angabe handelte. Ein Beispiel: Bei

gleichgeschlechtlichen Personen, die beide leibliche Eltern einer dritten Person im Haushalt waren, wurde auf Basis des Vornamens das Geschlecht korrigiert. Deuteten auch die Vornamen darauf hin, dass beide Personen das gleiche Geschlecht haben und lagen keine weiteren relevanten Hinweise vor, wurde vor dem Hintergrund der Haushaltskonstellation die Verwandtschaftsbeziehung als unplausibel gekennzeichnet.

- In einem zweiten Teil wurden Prüfungen durchgeführt, bei denen jeweils drei Verwandtschaftsverhältnisse in Beziehung zueinander gesetzt und auf Plausibilität geprüft wurden. Ein Beispiel für eine in dieser Prüfung als unplausibel klassifizierte Beziehungskonstellation wäre: Person A ist Ehepartner von Person B. Person A ist leibliches Elternteil von Person C. Person C ist Geschwister von Person B. Wurde im Rahmen dieser Prüfung eine solche oder ähnlich unplausible Kombination von Beziehungen identifiziert, so wurde auch hier versucht, die Beziehung im Rahmen des Haushaltskontexts zu plausibilisieren. Im genannten Fall wurde die Beziehungsinformation korrigiert, indem Person C als nicht weiter spezifiziertes Kind von B codiert wurde. Ziel ist es, möglichst viele der so identifizierten Unplausibilitäten inhaltlich zu korrigieren, da eine plausible und vollständig vorliegende Beziehungskonstellation notwendige Voraussetzung für die Generierung der BG ist.
3. Auch die Spelldatensätze wurden einer Reihe von Plausibilitätsprüfungen unterzogen, welche in Kapitel 5.6 bis 5.8 vertieft dargestellt werden.

5.4 Rückwirkende Veränderungen der Wellen 1 bis 4

5.4.1 Konzeptionelle Überarbeitungen

Im Rahmen der Arbeiten am SUF der Welle 4 haben an einer Reihe von generierten Variablen konzeptionelle Anpassungen stattgefunden. Dies ist auf drei verschiedene Gründe zurückzuführen.

Zum einen waren Änderungen in der Erhebungslogik zu berücksichtigen. Dies betrifft erstens die arbeitsmarktpolitischen Maßnahmen, an denen die Zielpersonen teilgenommen haben. Während in den Wellen 1 bis 3 ein umfangreiches Spektrum an Maßnahmen erhoben wurde, beschränkt sich das Erkenntnisinteresse ab Welle 4 auf Ein-Euro-Jobs. Zweitens ist das Konzept zur Erhebung der Erwerbstätigkeiten betroffen. Dieses stellt sich im Wellenverlauf so dar:

- Welle 1: Panelkonzept, d.h. nur Erhebung am aktuellen Rand
- Welle 2/Welle 3: modulare Erhebung von ET-/AL-Episoden³⁶ + Auffüllen von Lücken von > 3 Monaten und am aktuellen Rand
- ab Welle 4: integrierte Erhebung von ET-/AL-/LU-Episoden

Zum anderen waren konzeptionelle Mängel in der Abgrenzung von Haupt- und Nebentätigkeiten bei den generierten Variablen zum Einkommen und zur Arbeitszeit zu beheben. Weiterhin mussten im Personenregister sowie in den *bio_spells* Entscheidungen getroffen werden mit Blick auf das aktuelle Erhebungskonzept.

Auf diese Überarbeitungen ist bereits in Kap. 5.4 des Datenreports zur Welle 4 des PASS detailliert eingegangen worden (vgl. Berg et.al, FDZ Datenreport 08/2011). Zwei Themenbereiche werden nun erneut angesprochen. Dies betrifft zum einen die generierten Einkommensvariablen. Im *PENDDAT* können die Variablen *brutto(kat)* und *netto(kat)* in Welle 5 erstmals seit Welle 1 wieder generiert werden, im *bio_spells*-Spelldatensatz liegen in Welle 5 erstmals die Variablen *br* und *net* vor. Um die Funktion der neu bzw. wieder verfügbaren Variablen zu verdeutlichen, werden die entsprechenden Ausführungen aus dem Datenreport zur Welle 4 an dieser Stelle wiederholt.

Zum anderen wird erneut der Umgang mit Dublettenpersonen thematisiert. In Welle 4 sind erstmals Konstellationen aufgetreten, bei denen eine Person gleichzeitig in zwei Haushalten lebt. In Welle 5 ist es erstmals vorgekommen, dass solche Dublettenpersonen selbst auf der Personenebene ein Interview gegeben haben. Vor diesem Hintergrund wurden die konzeptionellen Überlegungen zum Umgang mit den Dublettenpersonen weiterentwickelt. Eine Anpassung der SUF-Daten der Vorwellen war dennoch nicht notwendig, da die konzeptionellen Änderungen sich auf dem Umgang mit Interviews auf der Personenebene beziehen. Die Ausführungen zu den Dublettenpersonen an dieser Stelle ersetzen somit das entsprechende Kapitel aus dem Datenreport zur Welle 4.

³⁶ Hier und im Folgenden: ET = Erwerbstätigkeiten; AL = Arbeitslosigkeiten; LU = Lücken (d.h. Aktivitäten, die nicht ET oder AL sind).

5.4.1.1 Einkommensvariablen im PENDDAT und in BIO-Spells

In Wellen 1 bis 4 beziehen sich die Variablen zur aktuellen Erwerbstätigkeit auf die Haupterwerbstätigkeit³⁷. Ausgenommen sind hier die Angaben zum Brutto-/Nettoeinkommen in Wellen 2 bis 4 – diese beziehen sich auf alle aktuell andauernden Erwerbstätigkeiten > 400 Euro (Unschärfe hinsichtlich Minijoblöhnen). Spellspezifische Angaben liegen nicht vor und werden erst ab Welle 5 erhoben. Die Angaben werden nur als Gesamtwert über alle Erwerbstätigkeiten erhoben. Das führt zu zwei Teilproblemen:

- I. Die generierten Variablen zu Arbeitsstunden und Brutto/Netto beziehen sich ab Welle 2 auf verschiedene Erwerbstätigkeiten (Haupt-ET bzw. alle ET). Werden auf dieser Basis Stundenlöhne berechnet, kommt es zu Fehlern bei ZP mit mehreren ET.
- II. Die unterschiedlichen Bezüge sind aus den Variablenlabels nicht ersichtlich.

Die generierten Variablen zu Einkommen und Arbeitszeit werden deshalb in Welle 4 entsprechend überarbeitet.

Das Erhebungskonzept der Einkommensvariablen hat sich zwischen Welle 1 und 2 stark verändert, ohne dass dies zur Bildung neuer Variablen geführt hat: *brutto (bruttokat)* und *netto (nettokat)* geben in Welle 1 das Einkommen der Haupttätigkeit wieder, ab Welle 2 das Einkommen aller nicht geringfügigen Tätigkeiten. Dies ist inkonsistent und führt potentiell zu Fehlern in der Auswertung. Dieses Problem soll mit der Überarbeitung korrigiert werden:

Tabelle 23: Überarbeitung Einkommensvariablen³⁸

Variable	Inhalt	Datensatz	Generiert für					Basis	
			W1	W2	W3	W4	W5	OFFA	KatA
bruttokat	HauptET, Brutto	PENDDAT	1	0	0	0	1	0	1
brutto	HauptET, Brutto	PENDDAT	1	0	0	0	1	1	1
nettokat	HauptET, Netto	PENDDAT	1	0	0	0	1	0	1
netto	HauptET, Netto	PENDDAT	1	0	0	0	1	1	1
brges	GesET, Brutto	PENDDAT	0	1	1	1	1	1	1
netges	GesET, Netto	PENDDAT	1	1	1	1	1	1	1*
br	SpellET, Brutto	BIO-Spells	0	0	0	0	1	1	1
net	SpellET, Netto	BIO-Spells	0	0	0	0	1	1	1

³⁷ Welle 2 bis Welle 3; es handelt sich um die zensierte ET im ET-Spell Datensatz. Bei mehreren zensierten Spells wurde der Spell mit dem größten Stundenumfang ausgewählt. Bei mehreren zensierten Spells mit gleichem Stundenumfang wurde der am längsten andauernde Spell ausgewählt. Für Senioren wurde nur eine Erwerbstätigkeit erhoben.

³⁸ In Welle 1 existiert nur für den Nettolohn der Haupterwerbstätigkeit eine kategoriale Nachfrage, nicht aber für die weiteren Tätigkeiten. Das wird bei der Generierung von *netges* akzeptiert. Fehlen die Angaben (MV) zum Nettoeinkommen der weiteren Tätigkeiten, kann die Variable *netges* nicht generiert werden.

Überarbeitete Variablen (bereits im Datensatz in Wellen 1 bis 3)

bruttokat (akt. Bruttoeinkommen Haupterwerb. (ohne Minijobs, kategorisiert), gen.)

brutto (akt. Bruttoeinkommen Haupterwerb. (ohne Minijobs, inkl. kat. Ang.), gen.)

nettokat (akt. Nettoeinkommen Haupterwerb. (ohne Minijobs, kategorisiert), gen.)

netto (akt. Nettoeinkommen Haupterwerb. (ohne Minijobs, inkl. kat. Ang.), gen.)

Diese Variablen sind in Welle 1 auf die jeweilige Haupt-ET bezogen. Ab Welle 2 wurden sie aber mit den kumulierten Angaben für alle ET (>400 Euro) gefüllt, da nur diese erhoben wurden. Die Variablenlabel werden ab Welle 4 entsprechend angepasst. Für Wellen 2 bis 4 werden die Variablen mit -9 gefüllt, da eine zu Welle 1 analoge Generierung nicht möglich ist.

Neue Variablen in W4

brges (akt. Bruttoeinkommen insgesamt (ohne Minijobs, inkl. kat. Ang.), gen.)

Diese Variable enthält die über alle ET (>400 Euro) kumulierte Angabe zum Bruttoeinkommen. Für Welle 1 ist die Variable in dieser Form nicht generierbar, da nur für die Haupt-ET das Bruttoeinkommen erhoben wurde. Für Welle 2 und Welle 3 ist die Variable inhaltlich identisch zur Variable *brutto*, die im SUF Welle 3 ausgeliefert wurde (d.h. vor der oben beschriebenen Überarbeitung). In Wellen 2 bis 4 wurde nur das kumulierte Bruttoeinkommen erhoben – die in Welle 2/Welle 3 genutzten Quellvariablen enthalten also schon die entsprechenden Angaben zum Gesamteinkommen aus ET > 400 Euro. Für Welle 4 soll die Variable analog zu Welle 2/Welle 3 erstellt werden. Ab Welle 5 erfolgt die Generierung auf Basis von spellspezifischen Einkommensangaben.

netges (akt. Nettoeinkommen insgesamt (ohne Minijobs, inkl. kat. Ang.), gen.)

Diese Variable enthält die über alle ET (>400 Euro) kumulierte Angabe zum Nettoeinkommen. Für Welle 1 ist die Variable generierbar, in dem die offene und kategoriale Nettoeinkommensangabe der Haupterwerbstätigkeit mit der Angabe für die weiteren Tätigkeiten kombiniert wird (hier fehlt allerdings die kategoriale Nachfrage). Für Wellen 2 und 3 ist die Variable identisch zur Variable *netto*, die im SUF Welle 3 ausgeliefert wurde. In Wellen 2 bis 4 wurde nur das kumulierte Nettoeinkommen erhoben – die in Welle 2/Welle 3 genutzten Quellvariablen enthalten also schon die entsprechenden Angaben zum Gesamteinkommen aus ET > 400 Euro. Für Welle 4 wurde die Variable analog zu Wellen 2 und 3 erstellt. Ab Welle 5 erfolgt die Generierung auf Basis von spellspezifischen Einkommensangaben.

5.4.1.2 Dublettenzeiger im p_register

zdub* (Zeiger: Personennr. der durch die ZP gedoppelten Person in Welle X (20XY))

Die Datenstruktur des PASS (bspw. im Personenregister) sieht vor, dass eine Personennummer in jeder Welle nur jeweils einem Haushalt zugeordnet werden kann.

Deshalb müssen Personen, die faktisch mehr als einem Haushalt angehören oder bei denen ein Haushaltswechsel (Auszug) nicht ordentlich berichtet wurde, gesondert behandelt werden.

Dazu wird im Personenregister eine wellenspezifische Zeigervariable (*zdub**) angelegt, mit denen diese Fälle markiert werden.

Zwei Problemtypen sind zu unterscheiden:

1. Echte Dubletten

Echte Dubletten sind Personen, die in einer Welle faktisch von zwei Haushalten beansprucht werden³⁹. Die beteiligten Haushalte wurden befragt und die Person taucht in den jeweiligen Haushaltsstrukturen auf.

Wurden in der aktuellen Welle mit Dublette und Original Interviews auf Personenebene geführt, so wird das Interview der Dublette entfernt und auch nicht für die Preload-Generierung der Folgewelle genutzt. Die Markierung erfolgt, wie bei den anderen in der Datenaufbereitung gelöschten Personeninterviews auch, in den *pnetto**-Variablen der jeweiligen Welle. In der Gewichtung wird immer nur eine der beiden Beobachtungen der Person in der aktuellen Welle genutzt. Eine besondere Behandlung dieser Fälle in der Gewichtung ist also nicht notwendig.

Liegt in der aktuellen Welle nur für Dublette oder Original ein Interview auf Personenebene vor, wird dieses Interview nicht entfernt, d.h. wenn es zum Dubletteninterview keine konkurrierende Information aus dem Interview mit Original gibt, bleibt das Dubletteninterview im SUF.

Die Informationen aus dem als gültig gewerteten Personeninterview werden sowohl für die Dublette als auch für das Original auch für die Preload-Generierung der Folgewelle genutzt. Dies geschieht insbesondere mit Blick auf die fortzuschreibenden Spellinformationen. Bei echten Dubletten ist es denkbar, dass diese im Wellenverlauf mal in ihrem Ur-HH und mal in ihrem Splithaushalt ein Personeninterview geben. Würde man den personenbezogenen Preload für die Folgewelle abhängig davon generieren, ob die Person die Angaben als Original oder als Dublette gemacht hat, käme es im Wellenverlauf zu einer mehrfachen Erhebung von biografischen Informationen. Diese müssten dann entweder nachträglich in der Datenaufbereitung zusammengeführt werden oder sie würden als redundante Information in den Daten abgelegt werden. Mit der Aufbereitung der personenbezogenen Preloadinformationen unabhängig vom Haushalt, in dessen Rahmen die Angaben gemacht wurden, werden solche Probleme vermieden. Der haushaltsbezogene Preload unterscheidet sich hingegen zwischen den beiden aktuellen Haushalten der Dublettenperson.

Die Person behält unabhängig davon, ob sie als Original oder als Dublette ein Personeninterview gibt, in den meisten SUF-Datensätzen ihre aus dem Ur-HH bekannte Personennummer *pnr* bei. Im Personenquerschnitt ist diese Vorgehensweise möglich, da dort über die Haushaltsnummer *hnr* ersichtlich ist, in welchem Haushalt die Dublettenperson das Personeninterview gegeben hat. In den Spelldatensätzen ist diese Vorgehensweise sogar zwingend, da dort die Biografie der Person fortgeschrieben wird und die Biografie dieser Person nicht auf zwei verschiedene Personennummern aufgeteilt werden soll.

³⁹ Ob es sich um dieselbe Person handelt, wird in der Prüfung der Haushaltsstrukturen sichergestellt. Basis sind demografische Angaben (Name, Alter, Geschlecht, Geburtsdatum).

Im Personenregister werden Original und Dublette in zwei Datenzeilen dokumentiert. Es wird eine wellenspezifische Zeigervariable *zdub** integriert, die von einer Dublette auf das Original zeigt (unabhängig vom Befragungsstatus von Dublette und Original auf Personenebene). Für die Beobachtung der Dublette im Personenregister enthält die Zeigervariable also die feste Personennummer des Originals, d.h. sie kann nur für Beobachtungen, die Dubletten sind, mit einer Personennummer gefüllt sein. Handelt es sich bei einer Beobachtung nicht um eine Dublette, ist die Variable mit „0“ gefüllt (analog zum Vorgehen in den anderen Zeigervariablen) bzw. mit „-6“, wenn der Haushalt der Person in der aktuellen Welle nicht befragt wurde oder die Person nicht mehr Teil eines Befragungshaushalts ist (analog zur Vergabe des -6 Codes in den übrigen Variablen des Personenregisters). Im Personenregister ist eine Dublettenperson also zweifach enthalten. Zum einen als Original: Dort ist die *pnr* die feste Personennummer des Originals, unter der die Person seit Zugang zum Panel bekannt ist, *zdub** ist gleich 0. Zum anderen als Dublette: Dort ist die *pnr* neu generiert aus der *hnr* des Haushalts, in dem die Person Dublette ist und der Position der Dublette in diesem Haushalt. *zdub** enthält die feste Personennummer des Originals.

In dem Haushalt, in dem die Person Dublette ist, wird zudem im *hh_register* die in der Kennung *pnrzp** hinterlegte Personennummer auf die im *p_register* hinterlegte Personennummer der Dublette abgeändert, sofern die Dublettenperson in einer Welle der HBV dieses Haushalts ist.

In den Folgewellen führt das Aussetzen eines der beiden Haushalte nicht zur Auflösung der Dublette.

Auf mehreren SUF-Datensätzen aufsetzende Analysen können somit trotz des Auftretens von Dublettenpersonen weitgehend in der gewohnten Form durchgeführt werden. Bei der Nutzung des *p_register* ist jedoch zu beachten: Für Matchvorgänge mit dem *p_register* über die Personennummer muss zunächst eine Matchvariable generiert werden, die *zdub** entspricht, sofern diese größer 0 ist und ansonsten der *pnr*. Zudem müssen nichtbefragte Fälle über *pnetto** abgegrenzt werden, um z. B. zu vermeiden, dass Informationen des Originals zugespült werden, wenn in einer Welle die Dublette das Personeninterview gegeben hat.

2. Potentielle Dubletten

Eine Person ist als Mitglied eines Haushalts bekannt, der bereits früher im PASS befragt wurde (=Ur-HH). Obwohl dieser Haushalt in der aktuellen Welle nicht befragt wurde, taucht die Person nun in einem anderen Haushalt auf (=Dubletten-HH). Da der Ur-HH dieser Person seit dem Auftauchen der Person in dem Dubletten-HH, nicht wieder befragt wurde, ist unklar, ob die neu angelegte Person eine Dublette oder einen regulären Auszug darstellt (der lediglich noch nicht erfasst wurde). Diese Person stellt also eine potentielle Dublette des Originals im Ur-HH dar.

Bei potentiellen Dubletten wird davon ausgegangen, dass es sich um einen noch nicht berichteten Auszug aus dem Ur-HH handelt. Entsprechend wird der potentiellen Dublette im SUF die feste Personnummer des Originals zugewiesen, d.h. die Person wird behandelt, als ob sie vom Ur-HH in den Dubletten-HH umgezogen wäre. In der aktuellen Welle auf Personenebene geführte Interviews bleiben erhalten.

Da nicht sicher eine Dublette vorliegt, sondern stattdessen die Personnummer der betroffenen Person geändert wird, ist die Zeigervariable in diesen Fällen nicht mit einer Personnummer belegt.

Für den Preload der Folgewelle wird wie unter (1) vorgegangen. Der personenbezogene Preload wird also haushaltsübergreifend fortgeschrieben, unabhängig davon, ob es sich um eine echte oder eine potentielle Dublette handelt.

In der Folgewelle kann sich herausstellen, dass es sich um eine echte Dublette (siehe 1) handelt. In diesem Fall wird im *p_register* nachträglich eine zweite Zeile für diese Person eingefügt, deren *pnr* neu generiert ist aus der *hnr* des Haushalts, in dem die Person Dublette ist und der Position der Dublette in diesem Haushalt und bei der *zdub** mit der festen Personnummer des Originals gefüllt ist. Dies geschieht dann auch rückwirkend für alle Wellen, in denen die nun als Dublettenperson erkannte Person in ihrem aus dem Ur-HH heraus entstandenen Haushalt lebt. Dadurch gibt es ab dem SUF der Welle 5 *zdub**-Variablen im *p_register* für alle Wellen ab Welle 2, auch wenn die erste echte Dublettenperson erst in Welle 4 beobachtet wurde. In diesen Fällen wird ggf. auch *pnrzp** im *hh_register* rückwirkend geändert.

Ausprägungen der zu generierenden Variable:

- 6 HH n. befr./ZP kein Mitgl. eines BefrHH
- 0 ZP ist keine Dublette von anderer Pers.
- (Feste Personnummer des „Originals“, wenn ZP eine Dublette ist.)

5.4.2 Fehlerkorrekturen

Im Rahmen der Datenaufbereitung für den Scientific Use File der Welle 5 wurden auch Veränderungen an den bereits ausgelieferten Wellen des PASS vorgenommen. Dabei handelt es sich um Korrekturen von Fehlern, die seit der Fertigstellung des Scientific Use Files der Welle 4 festgestellt wurden. Die Tabelle 24 bis 28 geben einen Überblick über die rückwirkenden Veränderungen an bereits ausgelieferten Wellen des PASS⁴⁰.

⁴⁰ Anpassungen an Werte- oder Variablenlabels sind hier nur berücksichtigt, wenn sie dadurch die Interpretation von Variablen oder Werten verändert.

Tabelle 24: Übersicht über rückwirkende Veränderungen im Haushaltsdatensatz (HHENDDAT)

geänderte Variable	betroffener Datensatz	geänderte Welle	Art der Änderung	Beschreibung der Änderung
<i>depindug2</i> <i>depindg2</i>	HHENDDAT	1-4	Eingefügt	Siehe Kapitel 4.5.1
<i>HD1101*</i>	HHENDDAT	4	Korrektur	Code 14 war bislang mit "Mini-Job, geringfügig beschäftigt (>= 400 Euro)" gelabelt, korrekt muss es heißen "Mini-Job, geringfügig beschäftigt (<= 400 Euro)".
<i>HW0881</i>	HHENDDAT	2+4	Korrektur	Bislang wurde der Umgang mit Angaben wie „weiß nicht“ und „keine Angabe“ aus den offenen Texten uneinheitlich gehandhabt. Teils wurden diese Fälle in den Variablen inklusive der offenen Angaben auf Code -1 bzw. -2 gesetzt, teilweise wurde der die Erhebung des offenen Textes auslösende „Sonstiges“-Code beibehalten. Dies wurde nun dahingehend vereinheitlicht, dass in solchen Fällen stets der die Erhebung des offenen Textes auslösende „Sonstiges“-Code beibehalten, da die „Sonstiges“-Angabe den höheren Informationsgehalt hat. Es wurden in Welle 2 56 Fälle und in W4 ein Fall korrigiert.

Tabelle 25: Übersicht über rückwirkende Veränderungen im Personendatensatz (PENDDAT)

geänderte Variable	betroffener Datensatz	geänderte Welle	Art der Änderung	Beschreibung der Änderung
<i>PB0401</i> <i>ostaatan(na)</i> <i>PG0901*</i> <i>PG1301</i> <i>PSH0201</i> <i>mschul2</i> <i>PSH0301*</i> <i>mberuf2</i> <i>PSH0501</i> <i>vschul2</i> <i>PTK0321*</i> <i>PEE0201*</i>	<i>PENDDAT</i>	1-4	Korrektur	<p>Bislang wurde der Umgang mit Angaben wie „weiß nicht“ und „keine Angabe“ aus den offenen Texten uneinheitlich gehandhabt. Teils wurden diese Fälle in den Variablen inklusive der offenen Angaben auf Code -1 bzw. -2 gesetzt, teilweise wurde der die Erhebung des offenen Textes auslösende „Sonstiges“-Code beibehalten. Dies wurde nun dahingehend vereinheitlicht, dass in solchen Fällen stets der die Erhebung des offenen Textes auslösende „Sonstiges“-Code beibehalten, da die „Sonstiges“-Angabe den höheren Informationsgehalt hat.</p> <p>Die einzelnen Variablen waren in den Wellen 1-4 mit den nachstehenden Fallzahlen betroffen: <i>PB0401</i>: W3: 4 <i>ostaatan(na)</i>: W3: 1 <i>PG0901*</i>: W2: 53; W4: 3 <i>PG1301</i>: W2: 2; W3: 3 <i>PSH0201</i>: W1: 1; W3: 3; W4: 1 <i>mschul2</i>: W1: 1; W2: 1; W3: 4; W4: 3 <i>PSH0301*</i>: W1: 1; W3: 1 <i>mberuf2</i>: W1: 1; W3: 1; W4: 1 <i>PSH0501</i>: W1: 1; W2: 1; W3: 4; W4: 4 <i>vschul2</i>: W1: 1; W2: 1; W3: 5; W4: 9 <i>PTK0321*</i>: W3: 5; W4: 1 <i>PEE0201*</i>: W4: 1</p>
<i>migration</i>	<i>PENDDAT</i>	1-4	Korrektur	<p>In der bisherigen Generierung gab es mehrere Fehler. Personen ohne eigenen Migrationshintergrund wurden nur dann auf Code „3 - Mind. 1 Elternteil zugezogen“ gesetzt, wenn genau ein Elternteil zugewandert ist. Wenn dagegen beide Elternteile zugewandert sind, standen diese Personen auf Code „1 - Kein Migrationshintergrund“. Zudem wurde nicht korrekt geprüft, ob für beide Elternteile nur Missingwerte zur Frage nach dem Zuzug aus dem Ausland vorliegen. Der Zuzug der Großmutter mütterlicherseits wurde unvollständig berücksichtigt.</p>
<i>hhalg2</i>	<i>PENDDAT</i>	1-4	Entfällt	<p>Diese Variable beinhaltet die Steuerinformation über den ALG-II-Bezug im Haushalt für den Personenbogen zur Feldzeit und sollte für Analysen nicht verwendet werden.</p>

Tabelle 26: Übersicht über rückwirkende Korrekturen in Spelldatensätzen (*bio_spells*, *alg2_spells*, *ee_spells*)

geänderte Variable	betroffener Datensatz	geänderte Welle	Art der Änderung	Beschreibung der Änderung
<i>branche</i>	<i>bio_spells</i>	2-4	Korrektur	Undefinierte Werte 8 und 42 wurden auf -5 recodiert.
<i>ET0552</i>	<i>bio_spells</i>	4	Korrektur	Falls ET0602 = -6 wurde in ET0552 der Missingwert von -3 auf -6 umgesetzt
<i>ET1952</i>	<i>bio_spells</i>	4	Korrektur	Falls ET0602 = -6 wurde in ET1952 der Missingwert von -3 auf -6 umgesetzt
<i>AL0601</i>	<i>bio_spells</i>	2-4	Korrektur	Falls AL0600 = -6 wurde in AL0601 der Missingwert von -3 auf -6 umgesetzt
<i>BIO0101</i> <i>AL0601</i> <i>LU0101</i>	<i>bio_spells</i>	2-4	Korrektur	Bislang wurde der Umgang mit Angaben wie „weiß nicht“ und „keine Angabe“ aus den offenen Texten uneinheitlich gehandhabt. Teils wurden diese Fälle in den Variablen inklusive der offenen Angaben auf Code -1 bzw. -2 gesetzt, teilweise wurde der die Erhebung des offenen Textes auslösende „Sonstiges“-Code beibehalten. Dies wurde nun dahingehend vereinheitlicht, dass in solchen Fällen stets der die Erhebung des offenen Textes auslösende „Sonstiges“-Code beibehalten, da die „Sonstiges“-Angabe den höheren Informationsgehalt hat.
<i>AL21300a-</i> <i>AL22103a</i> <i>AL21300b-</i> <i>AL22103b</i> <i>AL21300c-</i> <i>AL22103c</i>	<i>alg2_spells</i>	1-4	Korrektur	Bislang wurde der Umgang mit Angaben wie „weiß nicht“ und „keine Angabe“ aus den offenen Texten uneinheitlich gehandhabt. Teils wurden diese Fälle in den Variablen inklusive der offenen Angaben auf Code -1 bzw. -2 gesetzt, teilweise wurde der die Erhebung des offenen Textes auslösende „Sonstiges“-Code beibehalten. Dies wurde nun dahingehend vereinheitlicht, dass in solchen Fällen stets der die Erhebung des offenen Textes auslösende „Sonstiges“-Code beibehalten, da die „Sonstiges“-Angabe den höheren Informationsgehalt hat.

Tabelle 27: Übersicht über rückwirkende Veränderungen in Registerdatensätzen (*hh_register*; *p_register*)

geänderte Variable	betroffener Datensatz	geänderte Welle	Art der Änderung	Beschreibung der Änderung
<i>weg*4</i>	<i>p_register</i>	4	Korrektur	Bei vier Personen muss das W4-Auszugsdatum nachgetragen werden. Diese ZP sind in W4 in ihren Ur-HH als ausgezogen gemeldet worden und der Split-HH wurde auch in W4 realisiert. Da der Ur-HH jedoch ein abgebrochenes Interview geliefert hat, wurden die Angaben nicht in den SUF übernommen.
<i>zupanel</i>	<i>p_register</i>	4	Korrektur	Der in W4 bekannten Dublettenperson wurde der Wert „zupanel==4“ zugewiesen, obwohl sie über ihren Ur-Haushalt bereits seit Welle 2 zum Panel gehört.
<i>erwprox</i>	<i>p_register</i>	4	Korrektur	Code 14 war bislang mit „Mini-Job, geringfügig beschäftigt (>= 400 Euro)“ gelabelt, korrekt muss es heißen „Mini-Job, geringfügig beschäftigt (<= 400 Euro)“.

Tabelle 28: Übersicht über rückwirkende Veränderungen in den Gewichtungsdatensätzen (*hweights*; *pweights*)

Geänderte Variable	Betroffener Datensatz	Geänderte Welle	Art der Änderung	Beschreibung der Änderung
-	-	-	-	-

5.5 Anonymisierung

Vom IAB als besondere Dienststelle der Bundesagentur für Arbeit (BA) erhobene Daten stellen immer Sozialdaten dar, weshalb die Ansprüche an den Datenschutz entsprechend hoch sind. Daher wurde es notwendig, einige Variablen nur in vergrößerter Form in den Scientific Use File aufzunehmen. Diese Variablen sind im Variablenlabel generell mit dem Zusatz „anonymisiert“ gekennzeichnet. Aus dem gleichen Grund war es auch notwendig, auf verfügbare Regionalinformationen mit Ausnahme der Bundesländer und einer davon abgeleiteten Ost-/West-Information zu verzichten. Auch die Angaben zu den Verwandtschaftsbeziehungen im Haushalt sowie die Vornamen der Haushaltsmitglieder sind aus Datenschutzgründen nicht Teil des Scientific Use Files. Hinweise auf die Haushaltsstruktur geben allerdings generierte Variablen, bspw. zum Haushalts- und Bedarfsgemeinschaftstyp (*hhtyp*⁴¹, *bgtyp*⁴²), Indikatorvariablen zu Partnern im Haushalt (*apartner*, *epartner*⁴³), Zeigervariablen auf Eltern und Partner im Haushalt (*zmhh*; *zvhh*; *zparthh*⁴⁴) sowie verschiedene Indikatorvariablen, die anzeigen ob Eltern (*mhh*; *vhh*⁴⁵) oder Kinder der Zielperson (z. B. *ekind*⁴⁶) im Haushalt leben.

Tabelle 29 gibt einen Überblick über die betroffenen Variablen und die vorgenommenen Anonymisierungsschritte⁴⁷ im Personendatensatz. In Tabelle 30 sind die anonymisierten Variablen des Erwerbstätigkeitsspelldatensatzes dargestellt.

⁴¹ Enthalten im Haushaltsdatensatz (*HHENDDAT*), siehe Kapitel 4.5.2.

⁴² Wellenspezifische Variablen enthalten im Personenregister (*p_register*), siehe Kapitel 4.4.

⁴³ Enthalten im Personendatensatz (*PENDDAT*), siehe Kapitel 4.4.

⁴⁴ Wellenspezifische Variablen enthalten im Personenregister (*p_register*), siehe Kapitel 4.4.

⁴⁵ Enthalten im Personendatensatz (*PENDDAT*), siehe Kapitel 4.4.

⁴⁶ Enthalten im Personendatensatz (*PENDDAT*), siehe Kapitel 4.4.

⁴⁷ Falls für Ihr Forschungsvorhaben die Verwendung der nicht anonymisierten Versionen einer oder mehrerer dieser Variablen unabdingbar sein sollte, möchten wir Sie bitten, sich an das Forschungsdatenzentrum zu wenden, um eine geeignete Datenzugangsmöglichkeit zu finden. Wie diese aussehen kann, wird vom Forschungsvorhaben und den dafür benötigten Variablen abhängen.

Tabelle 29: Übersicht über die in Welle 5 anonymisierten Variablen im Personendatensatz (PENDDAT)

Varname	Variablenlabel	Vorgehen
<i>PD0100</i>	Geburtsjahr (Geburtsdatum, anonymisiert)	Das genaue Geburtsdatum wurde zum Geburtsjahr verkürzt.
<i>gebhalbj</i>	Geburtshalbjahr, generiert	Das genaue Geburtsdatum wurde zu einem Indikator für das 1. bzw. 2. Halbjahr verkürzt.
<i>PET1210</i>	Letzte Stellung im Beruf, einfache Einteilung (vor Januar 2005) (anon.)	In der Befragung wurden aus erhebungstechnischen Gründen Berufs- und Zeitsoldaten separat erhoben. Aufgrund der geringen Fallzahlen und da diese ohnehin üblicherweise nicht im Rahmen der Stellung im Beruf erhoben werden, wurde diese Gruppe mit den Beamten und Richtern zu einer gemeinsamen Gruppe zusammengeführt.
<i>PET1250</i>	Letzte StiB Beamter: det. Erfassung, inkl. Soldaten (vor Januar 2005) (anon.)	Diese Variable enthält zusätzliche Fälle. Die Berufs- und Zeitsoldaten aus <i>PET1240</i> wurden der entsprechenden Beamtenkategorie zugeschlagen. Die Variable für Berufs- und Zeitsoldaten <i>PET1240</i> wird nicht ausgeliefert.
<i>PET1211</i>	Letzte Stellung im Beruf, einfache Einteilung (inkl. Spellinfo) (anon.), gen.	Vorgehen analog zu <i>PET1210</i> .
<i>PET1251</i>	Letzte StiB Beamter: det. Erfassung, inkl. Soldaten (inkl. Spellinfo) (anon.), gen.	Vorgehen analog zu <i>PET1250</i> . Die Variable für Berufs- und Zeitsoldaten <i>PET1240</i> wird nicht ausgeliefert.
<i>stiblewt</i>	Stellung im Beruf, letzte Erwerbstätigkeit, Kennziffer, generiert	Bei der Generierung der Stellung im Beruf, werden die Berufs- und Zeitsoldaten der entsprechenden Beamtenkategorie zugeordnet.
<i>PET1510</i>	Jetzige Stellung im Beruf, einf. Eint., Erh. ab Welle 2 (anon.)	Vorgehen analog zu <i>PET1210</i> .

Tabelle29: Übersicht über die in Welle 5 anonymisierten Variablen im Personendatensatz (PENDDAT) (Fortsetzung 1)

Varname	Variablenlabel	Vorgehen
<i>PET1900</i>	Jetzige StIB Beamter: det. Erfassung, inkl. Soldaten (anon.)	Vorgehen analog zu <i>PET1250</i> . Die in den Senioreninterviews erhobene Variable für Berufs- und Zeitsoldaten <i>PET1800</i> wird nicht ausgeliefert. Für die Personeninterviews wird aus den Erwerbstätigkeitsspeils <i>ET090*</i> keine generierte Variable für Berufs- und Zeitsoldaten in den Personendatensatz übernommen.
<i>stibkz</i>	Jetzige Stellung im Beruf, einf. Eint., harmonisiert (anonymisiert)	Bei der Generierung der Stellung im Beruf werden die Berufs- und Zeitsoldaten der entsprechenden Beamtenkategorie zugeordnet.
<i>stib</i>	Stellung im Beruf, Kennziffer, generiert	Vorgehen analog zu <i>stiblewt</i> .
<i>PET3300</i>	Erste Stellung im Beruf, einfache Einteilung (anon.)	Vorgehen analog zu <i>PET1210</i> .
<i>PET3700</i>	Erste StIB Beamter: det. Erfassung, inkl. Soldaten	Vorgehen analog zu <i>PET1250</i> . Die Variable für Berufs- und Zeitsoldaten <i>PET3600</i> wird nicht ausgeliefert.
<i>PET3301</i>	Erste Stellung im Beruf, einf. Eint. (zus.gef., inkl. Spellinfo) (anon.), gen.	Vorgehen analog zu <i>PET1210</i> .
<i>PET3701</i>	Erste StIB Beamter: det. Erf., inkl. Soldaten, (zus.gef., inkl. Spellinfo) (anon.), gen.	Vorgehen analog zu <i>PET1250</i> . Die Variable für Berufs- und Zeitsoldaten <i>PET3600</i> wird nicht ausgeliefert.
<i>stibeewt</i>	Stellung im Beruf, erste Erwerbstätigkeit, Kennziffer, generiert	Vorgehen analog zu <i>stiblewt</i> .
<i>PSH0320</i>	Damalige Stellung im Beruf der Mutter, einfache Einteilung (anon.)	Vorgehen analog zu <i>PET1210</i> .
<i>PSH0360</i>	Damalige StIB d. Mutter Beamte, inkl. Soldaten: det. Erfassung (anon.)	Vorgehen analog zu <i>PET1250</i> . Die Variable für Berufs- und Zeitsoldaten <i>PSH0350</i> wird nicht ausgeliefert.
<i>mstib</i>	Stellung im Beruf der Mutter, Kennziffer, generiert	Vorgehen analog zu <i>stiblewt</i> .
<i>PSH0620</i>	Damalige Stellung im Beruf des Vaters, einfache Einteilung (anon.)	Vorgehen analog zu <i>PET1210</i> .

Tabelle29: Übersicht über die in Welle 5 anonymisierten Variablen im Personendatensatz (PENDDAT) (Fortsetzung 2)

Varname	Variablenlabel	Vorgehen
<i>PSH0660</i>	Damalige StiB d. Vaters Beamter, incl. Soldaten: det. Erfassung (anon.)	Vorgehen analog zu <i>PET1250</i> . Die Variable für Berufs- und Zeitsoldaten <i>PSH0650</i> wird nicht ausgeliefert.
<i>vstib</i>	Stellung im Beruf des Vaters, Kennziffer, generiert	Vorgehen analog zu <i>stiblewt</i> .
<i>PMI0200</i>	Nicht in D geboren: Geburtsland	Länder mit zu geringen Fallzahlen wurden zu größeren Kategorien zusammengefasst.
<i>ogebland</i>	Geburtsland, inkl. offene Angaben, Kategorien (anonymisiert)	Vorgehen analog zu <i>PMI0200</i> .
<i>PMI0500</i>	Keine D Staatsangehörigkeit: Welche Staatsangehörigkeit? (anonymisiert)	Staatsangehörigkeiten von Ländern mit zu geringen Fallzahlen wurden zu größeren Kategorien zusammengefasst.
<i>ostaatan</i>	Staatsangehörigkeit, inkl. offene Angaben, Kategorien (anonymisiert)	Vorgehen analog zu <i>PMI0500</i> .
<i>PMI1000a</i>	Vater: Zuzugsland (anonymisiert)	Zuzugsländer mit zu geringen Fallzahlen wurden zu größeren Kategorien zusammengefasst.
<i>PMI1000b</i>	Mutter: Zuzugsland (anonymisiert)	Vorgehen analog zu <i>PMI1000a</i> .
<i>PMI1000c</i>	Vater des Vaters: Zuzugsland (anonymisiert)	Vorgehen analog zu <i>PMI1000a</i> .
<i>PMI1000d</i>	Mutter des Vaters: Zuzugsland (anonymisiert)	Vorgehen analog zu <i>PMI1000a</i> .
<i>PMI1000e</i>	Vater der Mutter: Zuzugsland (anonymisiert)	Vorgehen analog zu <i>PMI1000a</i> .
<i>PMI1000f</i>	Mutter der Mutter: Zuzugsland (anonymisiert)	Vorgehen analog zu <i>PMI1000a</i> .
<i>ozulanda</i>	Vater: Zuzugsland, inkl. offene Angaben, Kategorien (anonymisiert)	Vorgehen analog zu <i>PMI1000a</i> .
<i>ozulandb</i>	Mutter: Zuzugsland, inkl. offene Angaben, Kategorien (anonymisiert)	Vorgehen analog zu <i>PMI1000a</i> .

Tabelle29: Übersicht über die in Welle 5 anonymisierten Variablen im Personendatensatz (PENDDAT) (Fortsetzung 3)

Varname	Variablenlabel	Vorgehen
<i>ozulandc</i>	Vater des Vaters: Zuzugsland, inkl. offene Angaben, Kategorien (anonymisiert)	Vorgehen analog zu <i>PMI1000a</i> .
<i>ozulandd</i>	Mutter des Vaters: Zuzugsland, inkl. offene Angaben, Kategorien (anonymisiert)	Vorgehen analog zu <i>PMI1000a</i> .
<i>ozulande</i>	Vater der Mutter: Zuzugsland, inkl. offene Angaben, Kategorien (anonymisiert)	Vorgehen analog zu <i>PMI1000a</i> .
<i>ozulandf</i>	Mutter der Mutter: Zuzugsland, inkl. offene Angaben, Kategorien (anonymisiert)	Vorgehen analog zu <i>PMI1000a</i> .

Tabelle 30: Übersicht über die in Welle 5 anonymisierten Variablen im BIO-Spelldatensatz (bio_spells)

Varname	Variablenlabel	Vorgehen
<i>ET0601</i>	Stellung im Beruf, einfache Einteilung (anon.)	Vorgehen analog zu <i>PET1210</i> .
<i>ET1001</i>	StiB Beamter: detaillierte Erfassung (anon.)	Vorgehen analog zu <i>PET1250</i> . Die Variable für Berufs- und Zeitsoldaten wird nicht ausgeliefert.
<i>stib</i>	Stellung im Beruf, Kennziffer, generiert	Vorgehen analog zu <i>stiblewt</i> .

5.6 Arbeitslosengeld-II-Bezüge

Bereits in den Wellen 1 bis 4 wurden die Arbeitslosengeld-II-Bezüge auf Haushaltsebene in Episodenform erfasst. Dieses Konzept wurde in Welle 5 mit einem leicht überarbeiteten Fragenprogramm fortgeführt.

5.6.1 Konzept für die Fortschreibung der in der Vorwelle aktuell andauernden Arbeitslosengeld-II-Spells

Für die Fortschreibung, der in der Vorwelle andauernden und damit im Spelldatensatz rechtszensierten Arbeitslosengeld-II-Bezüge sind Dependent-Interviewing-Fragen enthalten. Hierüber setzen Haushalte mit andauernden Spells zur Vorwelle bei diesen andauernden Spells wieder in die Befragung ein.

Die in Welle 5 erstmals befragten Haushalte der Auffrischungstichprobe wurden nach dem Bezug von Arbeitslosengeld II in der Zeit seit der letzten Veränderung des Haushalts gefragt. Lag diese Veränderung vor Januar 2009 oder wurde dazu keine Angabe gemacht, wurden die Arbeitslosengeld-II-Bezüge des Haushalts seit Januar 2009 erhoben.

5.6.2 Struktur des Arbeitslosengeld-II-Spelldatensatzes

Durch die Integration der in Welle 5 berichteten Arbeitslosengeld-II-Bezüge verändern sich die Struktur und der Inhalt des Arbeitslosengeld-II-Spelldatensatzes. Dabei ist zu unterscheiden zwischen (1) neuen Variablen, die auf eine bestimmte Welle bezogen sind, (2) neuen Variablen, die nicht auf eine bestimmte Welle bezogen sind und (3) in Welle 5 nicht mehr erhobenen Variablen.

1. Auch in Welle 5 wurden mit *AL20604*, *AL20704a bis AL20704o*, *AL20804* und *AL20904* neue, wellenspezifische Querschnittvariablen in den Arbeitslosengeld-II-Spelldatensatz eingefügt. Diese Variablen sind auf den Interviewzeitpunkt der Welle 5 bezogen. Für die Befragungszeitpunkte der Vorwellen existieren ebenfalls Querschnittvariablen, in denen die auf die jeweilige Welle bezogenen analogen Informationen abgelegt sind. Tabelle gibt einen Überblick über die im Arbeitslosengeld-II-Spelldatensatz enthaltenen Querschnittinformationen.

Tabelle 31: Querschnittvariablen im ALG-II-Spelldatensatz (*alg2_spells*)

	Querschnittvariable mit Information bezogen auf				
	Welle 1	Welle 2	Welle 3	Welle 4	Welle 5
Erhält der HH ALG II für alle HH-Mitglieder?	<i>AL20600</i>	<i>AL20601</i>	<i>AL20602</i>	<i>AL20603</i>	<i>AL20604</i>
Erhält der HH ALG II für die Personen 1 bis 15?	<i>AL20700a bis AL20700o</i>	<i>AL20701a bis AL20701o</i>	<i>AL20702a bis AL20702o</i>	<i>AL20703a bis AL20703o</i>	<i>AL20704a bis AL20704o</i>
Höhe des monatlichen ALG-II-Bezugs?	<i>AL20800</i>	<i>AL20801</i>	<i>AL20802</i>	<i>AL20803</i>	<i>AL20804</i>
Hat eine Kürzung des ALG II begonnen?	<i>AL20900</i>	<i>AL20901</i>	<i>AL20902</i>	<i>AL20903</i>	<i>AL20904</i>

2. Liegen in Welle 5 im Vergleich zu Welle 4 nicht vor.
3. Liegen in Welle 5 im Vergleich zu Welle 4 nicht vor.

5.6.3 Plausibilitätsprüfung und –bereinigung des Arbeitslosengeld-II-Spelldatensatzes

Wie in Welle 1 bis 4 wurden die Information zu Arbeitslosengeld-II-Bezügen auch in der Welle 5 einer Reihe von Plausibilitätsprüfungen unterzogen. Unzulässige Überschneidungen und Datierungen der Arbeitslosengeld-II-Spells bzw. der Kürzungsepisoden wurden dabei ggf. bereinigt. Veränderungen wurden dabei grundsätzlich nur in den generierten Datumsvariablen (*bmonat*; *bjahr*; *emonat*; *ejahr*) des Arbeitslosengeld-II-Spells, der Kürzungsepisoden (*alg2kbn*; *alg2kbnj*; *alg2ken*; *alg2kej*) sowie im Zensierungsindikator des Arbeitslosengeld-II-Spells (*zensiert*) vorgenommen. Konnten Unplausibilitäten nicht durch Korrektur der Datumsvariablen bereinigt werden, wurden in geringem Umfang Arbeitslosengeld-II-Spells zusammengeführt bzw. Arbeitslosengeld-II-Spells bzw. Kürzungsepisoden gänzlich gelöscht.

5.6.4 Fortschreibung des Arbeitslosengeld-II-Spelldatensatzes

Im Anschluss an die Umstrukturierung der in Welle 5 berichteten Arbeitslosengeld-II-Episoden in das Spellformat und die Plausibilitätsprüfungen und -bereinigungen, im Zuge derer unzulässige Überschneidungen und Spells mit unplausiblen Datierungen bereinigt wurden, werden die zum Befragungszeitpunkt der Vorwelle andauernden Arbeitslosengeld-II-Bezüge mit den in Welle 5 erhobenen Informationen fortgeschrieben. Dabei sind zwei Varianten zu unterscheiden, bei denen in der ersten (1) lediglich der Zensierungsindikator *zensiert* verändert wird. Die zweite Variante (2) stellt eine Fortschreibung des in der Vorwelle zensierten Spells mit Welle 5-Informationen im engeren Sinne dar. Dabei werden neben dem Zensierungsindikator auch die generierten und erhobenen Enddaten, die wellenspezifischen Querschnittinformationen (siehe oben) sowie Informationen zu neuen Kürzungsepisoden in den in der Vorwelle andauernden Arbeitslosengeld-II-Spell integriert. Neben der Fortschreibung von in der Vorwelle zensierten Spells werden dem Spelldatensatz neue, in Welle 5 berichtete Spells zugespielt (3). Diese drei Varianten sollen im Folgenden kurz dargestellt werden:

1. *Fälle, in denen der Haushalt in Welle 5 einem andauernden Arbeitslosengeld-II-Bezug zum Befragungszeit der Vorwelle widerspricht.*

Hat der Haushalt der Information, dass zum Zeitpunkt der Vorwelle ein aktuell andauernder Arbeitslosengeld-II-Bezug vorlag, in der Fortschreibungsfrage explizit bzw. durch Angabe eines vor dem Interviewdatum der Vorwelle liegenden Enddatums implizit widersprochen, so wurde *zensiert* auf „2“ (nein) gesetzt. Es wird davon ausgegangen, dass die zum Interviewdatum der Vorwelle gegebene Information korrekt war. Da über das Interviewdatum der Vorwelle hinaus keine verlässlichen Aussagen über die weitere Dauer des Bezugs möglich sind, wird angenommen, dass dieser im Interviewmonat der Vorwelle geendet hat. Die berichteten und generierten Variablen zum Enddatum des Spells (*AL20300*, *AL20400* sowie *emonat*, *ejahr*) sowie die Nachfrage, ob ein Spell noch andauert (*AL20500*), bleiben dabei unverändert.⁴⁸ Das generierte

⁴⁸ Hier gilt dasselbe. Verändert wird nur der Zensierungsindikator. Das berichtete Enddatum, die Nachfrage für andauernde Spells und das generierte Enddatum bleiben unverändert.

Enddatum des Arbeitslosengeld-II-Spells (*emonat; ejahr*) wurde bereits in der Vorwelle auf das Interviewdatum der Vorwelle gesetzt.

2. *Fälle, in denen der Haushalt Angaben zum Enddatum eines in der Vorwelle aktuell andauernden Bezugs macht.*

Liegen in Welle 5 Angaben zum Enddatum eines in der Vorwelle zensierten Arbeitslosengeld-II-Bezugs vor, so wurde der in der Vorwelle zensierte Spell mit den aktuellen Angaben fortgeschrieben. Zunächst wurden das erhobene Enddatum (*AL20300; AL20400*), das generierte Enddatum (*emonat; ejahr*), die Nachfrage, ob der Arbeitslosengeld-II-Bezug aktuell noch andauert (*AL20500*) sowie der Zensierungsindikator (zensiert) mit den in der Vorwelle erhobenen Informationen überschrieben. Weiterhin wurden die in der Welle 5 berichteten Kürzungsepisoden sowie die auf die Welle 5 bezogenen Querschnittinformationen (*AL20604; AL20704a bis AL20704o, AL20804, AL20904*) übernommen.

3. *In Welle 5 neu berichtete Arbeitslosengeld-II-Spells, die keine in der Vorwelle zensierte Episode fortschreiben.*

In Welle 5 neu berichtete Spells wurden dem Arbeitslosengeld-II-Spelldatensatz angehängt. Danach wurde der Spellzähler neu generiert, um eine lückenlose Variable *spellnr* zu schaffen.

5.7 Erwerbsbiografien

Bereits in der 2. und 3. Welle wurden die Erwerbstätigkeits-, Arbeitslosigkeits- und Lückenphasen auf Personenebene in Episodenform erfasst. Dieses Konzept der modularen Spellerhebung wurde in Welle 4 umgewandelt zu einer integrierten Erhebung der Erwerbsbiografie. Für Personen, die in Welle 5 erstmals nach ihrer Erwerbsbiografie gefragt wurden, wurde das Referenzdatum für den Beginn des Retrospektivintervalls angepasst. In Welle 5 sollten hier alle Erwerbstätigkeits- und Arbeitslosigkeitsepisoden seit Januar 2009 berichtet werden. Personen, die bereits in der Vorwelle zu ihrer Erwerbsbiografie befragt wurden, sollten dagegen alle neuen Episoden seit dem Datum des letzten Interviews berichten.

5.7.1 Konzept für die Fortschreibung der in der Vorwelle aktuell andauernden Spells

In der Welle 5 wurden andauernden ET-, AL- und Lückenspells fortgeschrieben. Für die Fortschreibung der in der Vorwelle andauernden und damit im Spelldatensatz rechtszensierten Spells sind in den Personenfragebögen Dependent-Interviewing-Fragen enthalten.

5.7.2 Struktur des BIO-Spell Datensatzes

Der BIO-Spell Datensatz orientiert sich seit Welle 4 hinsichtlich seiner Struktur an den modularen Spell Datensätzen ET, AL und LU der Welle 2 bis 3. ET-spezifische Variablen haben dabei im BIO-Spell Datensatz ihren Variablennamen im Vergleich zum ET-SUF der Welle 3 behalten, analog auch AL- bzw. LU-spezifische Variablen. Zwischen ET, AL und LU gleiche Variablen wurden ab Welle 4 vereinheitlicht (BIO0100, BIO0101, BIO0200, BIO0300, BIO0400, BIO0500, BIO0600) bzw. lagen bereits einheitlich in den ursprünglichen Datensätzen des SUF Welle 3 vor (bmonat, bjahr, emonat, ejahr, zensiert). Darüber hinaus liegen Variablen für den Aktivitätstyp (spelltyp), für die Spellintegration (spintegr) und eine übergreifende Spellnummer (spellnr) vor.

Durch die Integration der in Welle 5 berichteten Erwerbstätigkeits- und Arbeitslosigkeitsphasen in den BIO-Spell Datensatz werden neue ET- bzw. AL-spezifische Variablen ergänzt. Dabei ist zu unterscheiden zwischen (1) neuen Variablen, die auf eine bestimmte Welle bezogen sind, und (2) neuen Variablen, die nicht auf eine bestimmte Welle bezogen sind.

1. Im BIO-Spell Datensatz werden die ET-spezifischen Variablen *ET0600 bis ET2200* als wellenspezifische, auf Welle 2 bezogene Querschnittinformationen verstanden, die Variablen *ET0601 bis ET2201* auf Welle 3, *ET0552 bis ET2202* auf Welle 4 und *ET0553 bis ET2203* auf Welle 5 bezogene Querschnittinformationen. Tabelle 32 gibt einen Überblick über die im BIO-Spell Datensatz enthaltenen ET-spezifischen Querschnittinformationen.

Tabelle 32: ET-spezifische Querschnittvariablen im BIO-Spell Datensatz (bio_spells)

	Querschnittvariable mit Information bezogen auf...			
	Welle 2	Welle 3	Welle 4	Welle 5
Stellung im Beruf (einfache und detaillierte Einteilung)	<i>ET0600</i>	<i>ET0601</i>	<i>ET0552</i>	<i>ET0553</i>
	<i>ET0700</i>	<i>ET0701</i>	<i>ET0602</i>	<i>ET0603</i>
	<i>ET0800</i>	<i>ET0801</i>	<i>ET0702</i>	<i>ET0703</i>
	<i>ET1000</i>	<i>ET1001</i>	<i>ET0802</i>	<i>ET0803</i>
	<i>ET1100</i>	<i>ET1101</i>	<i>ET1002</i>	<i>ET1003</i>
	<i>ET1200</i>	<i>ET1201</i>	<i>ET1102</i>	<i>ET1103</i>
			<i>ET1202</i>	<i>ET1203</i>
Aufsichtungsfunktion; Zahl der beaufsichtig- ten Arbeitskräfte	<i>ET1300</i>	<i>ET1301</i>	<i>ET1302</i>	<i>ET1303</i>
	<i>ET1400</i>	<i>ET1401</i>	<i>ET1402</i>	<i>ET1403</i>
Entfristung einer an- fangs unbefristeten Tätigkeit	<i>ET1700</i>	<i>ET1701</i>	<i>ET1702</i>	<i>ET1703</i>
				<i>ET1753a</i>
				<i>ET1753b</i>
Arbeitszeit (vertraglich; tatsäch- lich; durchschnittlich bei unregelmäßiger Arbeitszeit)	<i>ET2000</i>	<i>ET2001</i>	<i>ET1952</i>	<i>ET1953</i>
	<i>ET2100</i>	<i>ET2101</i>	<i>ET2002</i>	<i>ET2003</i>
	<i>ET2200</i>	<i>ET2201</i>	<i>ET2102</i>	<i>ET2103</i>
			<i>ET2202</i>	<i>ET2203</i>

Im BIO-Spelldatensatz gibt es darüber hinaus eine AL-spezifische Variable, die als wellenspezifische Querschnittinformation zu verstehen ist (AL1300 für Welle 2; AL1301 für Welle 3, AL1302 für Welle 4 und AL1003 für Welle 5). Tabelle 33 gibt einen Überblick über die im Spelldatensatz enthaltenen Querschnittinformationen.

Tabelle 33: AL-spezifische Querschnittvariablen im BIO-Spelldatensatz (bio_spells)

	Querschnittvariable mit Information bezogen auf ...			
	Welle 2	Welle 3	Welle 4	Welle 5
Höhe des monatlichen ALG1-Bezugs?	AL1300	AL1301	AL1302	AL1303

- Die nicht wellenspezifischen ET-Variable *ET2800-ET3900* sind in Welle 5 erstmals erhoben und in den BIO-Spelldatensatz integriert worden.

5.7.3 Plausibilitätsprüfung und –bereinigung der Spelldatensätze

Auf Personenebene orientierten sich die Plausibilitätsprüfungen und Bereinigungsregeln an der Welle 2 bis Welle 4. Die Prüfungen fanden analog zu Welle 4 nur innerhalb eines Spelltyps statt. Spelltypübergreifende Prüfungen wurden nicht durchgeführt. Bereinigungen und Umcodierungen wurden, wie bei den Arbeitslosengeld-II-Spellldaten, nur in den generierten Datumsvariablen vorgenommen. Auch hier wurden Jahreszeitangaben in konkrete Monatsangaben umcodiert, „-8“ Werte für unplausible Angaben gesetzt oder Datumsangaben ersetzt oder plausibilisiert. Da lediglich die generierten Datumsvariablen editiert wurden, stehen dem Nutzer mit den Datumsvariablen BIO0200-BIO0500 und AL0800-AL1100 die ursprünglich erhobenen Informationen zur Verfügung und erlauben so die Durchführung eigener Prüfungen und Bereinigungen.

Darüber hinaus erschien es in einigen Fällen geboten, ganze Episoden zu löschen. So wurden bspw. offensichtlich doppelt erfasste Spells entfernt. Ebenfalls gelöscht wurden Episoden, die komplett außerhalb des abgefragten Zeitraums liegen und trotzdem erhoben wurden.

5.7.4 Fortschreibung der Spelldatensätze

Im Anschluss an die Umstrukturierung der in Welle 5 berichteten Episoden ins Spellformat und die Plausibilitätsprüfungen und -bereinigungen, im Zuge derer unzulässige Überschneidungen und Spells mit unplausiblen Datierungen bereinigt wurden, werden die zum Befragungszeitpunkt der Vorwelle andauernden Spells mit den in Welle 5 erhobenen Informationen fortgeschrieben.

Dabei sind drei Varianten zu unterscheiden, bei denen in der ersten (1) lediglich der Zensierungsindikator *zensiert* verändert wird. Die zweite Variante (2) stellt eine Fortschreibung des in der Vorwelle zensierten Spells mit Welle 5-Informationen im engeren Sinne dar. Dabei werden neben dem Zensierungsindikator auch die generierten und erhobenen Enddaten und die wellenspezifischen Querschnittinformationen (siehe oben) in den in der Vorwelle andauernden Spell integriert.

Neben der Fortschreibung von in der Vorwelle zensierten Spells werden dem Spelldatensatz neue, in Welle 5 berichtete Spells zugespielt (3). Diese drei Varianten sollen im Folgenden kurz dargestellt werden:

1. *Fälle, in denen die Person in Welle 5 einem andauernden Spell zum Befragungszeitpunkt der Vorwelle widerspricht.*

Hat die Person der Information, dass zum Zeitpunkt der Vorwelle ein aktuell andauernder Spell vorlag, in der Fortschreibungsfrage explizit bzw. durch Angabe eines vor dem Interviewdatum der Vorwelle liegenden Enddatums implizit widersprochen, so wurde der Zensierungsindikator *zensiert* auf „2“ (nein) gesetzt. Es wird davon ausgegangen, dass die zum Interviewdatum der Vorwelle gegebene Information korrekt war. Da über das Interviewdatum der Vorwelle hinaus keine verlässlichen Aussagen über die weitere Dauer des Spells möglich sind, wird angenommen, dass dieser im Interviewmonat der Vorwelle geendet hat. Die berichteten und generierten Variablen zum Enddatum des Spells (*BIO0400*, *BIO0500* sowie *emonat*, *ejahr*) sowie die Nachfrage, ob ein Spell noch andauert (*BIO0600*), bleiben dabei unverändert⁴⁹. Das generierte Enddatum des Spells (*emonat*; *ejahr*) wurde bereits in der Vorwelle auf das Interviewdatum der Vorwelle gesetzt.

2. *Fälle, in denen die Person Angaben zum Enddatum eines in der Vorwelle aktuell andauernden Spells macht.*

Liegen in Welle 5 Angaben zum Enddatum eines in der Vorwelle zensierten Spells vor, so wurde der in der Vorwelle zensierte Spell mit den aktuellen Angaben fortgeschrieben. Bei ET-Spells wurden das erhobene Enddatum (*BIO0400*; *BIO0500*), das generierte Enddatum (*emonat*; *ejahr*), die Nachfrage, ob der Spell aktuell noch andauert (*BIO0600*), der Grund für die Auflösung des Arbeitsvertrags (*ET2300*), die generierten Variablen zur Stellung im Beruf und der wöchentlichen Arbeitszeit (*stib*, *az1*, *az2*) sowie der Zensierungsindikator (*zensiert*) mit den in Welle 5 erhobenen Informationen überschrieben. Weiterhin wurden die auf die Welle 5 bezogenen Querschnittinformationen (*ET0553 bis ET2203*) übernommen.

⁴⁹ Das berichtete Enddatum ist also weiterhin mit dem Interviewdatum der Welle, in dem der Spell zensiert war bzw. dem Sondercode „0“ für andauernde Spells gefüllt. Auch die Nachfrage, ob der Spell noch andauert (für den Fall, dass das Enddatum dem Interviewdatum entspricht), wird nicht verändert. Die generierten Datumsvariablen enthalten weiterhin die letzte gültige Information, die hier dem Interviewdatum der Welle entspricht, in welcher der Spell zensiert war.

Bei AL-Spells wurden das erhobene Enddatum (*BIO0400; BIO0500*), das generierte Enddatum (*emonat; ejahr*), die Nachfrage, ob der Spell aktuell noch andauert (*BIO0600*), der Grund für die Beendigung der Arbeitslosigkeit (*AL0600, AL0601*) sowie der Zensierungsindikator (zensiert) mit den in Welle 5 erhobenen Informationen überschrieben. Weiterhin wurden die auf die Welle 5 bezogene Querschnittinformation (*AL1303*) übernommen. Bei den AL-Spell-Dateien gibt es zudem noch die Besonderheit, dass innerhalb des AL-Spells die Arbeitslosengeld-I-Bezugsepisode (ALG1-Bezug) erhoben wird. Welche Informationen fortgeschrieben werden, ist abhängig davon, ob in diesem AL-Spell zu einem früheren Zeitpunkt schon einmal ALG1 bezogen wurde und ob dieser Bezug in der Vorwelle noch andauert hat:

- Lag in der Vorwelle in dem fortzuschreibenden AL-Spell auch ein andauernder Bezug von ALG1 vor, wurde das erhobene Enddatum des Bezugs (*AL1000, AL1100*), der Indikator, ob der Bezug noch andauert (*AL1200*), das generierte Enddatum des Bezugs (*alg1em, alg1ej*) sowie der Zensierungsindikator des Bezugs (*alg1akt*) mit den in Welle 5 erhobenen Informationen überschrieben.
- War in dem fortzuschreibenden AL-Spell in Vorwelle noch nie ALG1 bezogen worden, wurden alle Informationen zum ALG1-Bezug mit den in Welle 5 erhobenen Informationen überschrieben. Neben dem Indikator, ob im AL-Spell schon einmal ALG1 bezogen wurde (*AL0700*), wurden das berichtete Beginn- und Enddatum (*AL0800, AL0900, AL1000, AL1100*), der Indikator für andauernden Bezug (*AL1200*) sowie die entsprechenden generierten Variablen (*alg1bm, alg1bj, alg1em, alg1ej, alg1akt*) mit den neu erhobenen Informationen ersetzt.
- Hat in dem fortzuschreibenden AL-Spell in der Vergangenheit ein ALG1-Bezug stattgefunden, der aber bereits in der Vorwelle beendet war, wurden in diesen Spells keine Veränderungen vorgenommen.

3. In Welle 5 neu berichtete Spells, die keine in der Vorwelle zensierte Episode fortzuschreiben.

In Welle 5 neu berichtete Spells wurden dem BIO-Spell-Dateiensatz angehängt. Danach wurde der Spellzähler neu generiert, um eine lückenlose Variable *spellnr* zu schaffen.

Das Fortschreiben der Spell-Dateiensätze hat keine Auswirkung auf die Spellnummern des Vorwelle-SUF. Bereits im SUF zu Welle 4 enthaltene Spells (*spellnret, spellnrl, spellnrlu, spellnr*) behalten ihre Spellnummer. Die neuen Spells aus Welle 5 werden an den entsprechenden Datensatz angehängt und die Spellnummern fortgeschrieben.

5.8 Ein-Euro-Job-Spell-Dateiensatz (*ee_spells*)

In der Welle 4 wurde das Konzept zur Erhebung der Maßnahmeteilnahme grundlegend überarbeitet. Anstelle des MN-Spell-Dateiensatzes trat ab Welle 4 der Ein-Euro-Spell-Dateiensatz (*ee_spells*). Dieser wurde in Welle 5 erstmals fortgeschrieben. Als Referenzzeitpunkt, ab dem Ein-Euro-Jobs erfragt wurden, diente für Welle 5 Januar 2010.

5.8.1 Konzept für die Fortschreibung der in der Vorwelle aktuell andauernden Spells

In der Welle 5 wurden andauernde EE-Spells fortgeschrieben. Für die Fortschreibung der in der Vorwelle andauernden und damit im Spelldatensatz rechtszensierten Spells sind in den Personenfragebögen Dependent-Interviewing-Fragen enthalten.

5.8.2 Struktur des EE-Spellldatensatzes

Durch die Integration der in Welle 5 berichteten Ein-Euro-Jobs in den EE-Spellldatensatz werden neue Variablen ergänzt, die auf eine bestimmte Welle bezogen sind. Tabelle 34 gibt einen Überblick über die im EE-Spellldatensatz enthaltenen Querschnittinformationen.

Tabelle 34: Querschnittvariablen im EE-Spellldatensatz (ee_spells)

Querschnittvariable mit Information bezogen auf...	Welle 4	Welle 5
	Arbeitszeit pro Woche im EEJ	EE1100
EEJ das gleiche was festangestellte Kollegen machen	EE1200	EE1201
Welche Art der Ausbildung für EEJ erforderlich	EE1300	EE1301
Ganze Zeit Arbeit oder auch Schulung/Kurs?	EE1400	EE1401
Beurteilung EEJ	EE1500a-EE1500h	EE1501a-EE1501h

Im EEJ-Spellldatensatz ist zu beachten, dass Spells auch vorhanden sind, wenn der Ein-Euro-Job nicht durchgeführt wurde, d.h. es nicht zur Teilnahme kam.

5.8.3 Plausibilitätsprüfung und –bereinigung des EEJ-Spellldatensatzes

Der EEJ-Spellldatensatz zur Teilnahme an Ein-Euro-Jobs wurde plausibilitätsgeprüft und bereinigt. Die Plausibilitätsprüfungen beinhalteten Prüfungen der Zeitangaben, des Referenzzeitpunkts für in Welle 5 neu aufgenommene Spells (1/2010) sowie Prüfungen auf logische Inkonsistenzen bei Befragten mit mehreren Ein-Euro-Job-Spells.

Bereinigungen und Umcodierungen wurden nur in den generierten Datumsvariablen (bmonat, bjahr, emonat, ejahr) vorgenommen. Jahreszeitangaben wurden dort in konkrete Monatsangaben umcodiert, „-8“ Werte für unplausible Angaben gesetzt oder Datumsangaben ersetzt oder plausibilisiert. Danach wurde ein Spellzähler *spellnr* generiert. Die Generierung wurde analog zu den chronologischen Zählern in den BIO-Spellldatensätzen vorgenommen, wobei nicht teilgenommene Spells bei der Sortierung nicht miteinbezogen wurden und somit ihre ursprüngliche *spellnr* beibehalten haben. EE-Spells aus Welle 4 haben ihre *spellnr* aus Welle 4 auch im SUF Welle 5 beibehalten.

6 Gewichtung Welle 5

Die Konstruktion der Gewichte für die fünfte Welle erfolgte im Wesentlichen analog zur vierten Welle (vgl. Berg et.al, FDZ Datenreport 08/2011). Ausgangspunkt für die Gewichtung der fünften Welle und des Längsschnitts von Welle 4 auf Welle 5 waren nunmehr die Querschnittsgewichte aus Welle 4 für Haushalte bzw. Personen. Die zwei Gewichte jedes Haushalts und die zwei Gewichte jeder Person wurden erneut fortgeschrieben. Dieses Kapitel des Datenreports stellt insbesondere die technischen Details und Modelle bei der Erzeugung der Gewichte für Welle 5 dar. Ein Überblick über das Gewichtungskonzept von PASS findet sich in Kapitel 8 (Trappmann (2011)) des User Guides von PASS (Bethmann und Gebhardt (Hg.)(2011). Anwendungsbeispiele sind dort in Kapitel 9.4 zu finden (Gebhardt und Trappmann 2011).

6.1 Erweiterung der Stichprobe Welle 5

In Welle 5 wurde PASS durch folgende neue Stichproben ergänzt:

1.) Eine Stichprobe von Neuzugängen zum SGB II im Juli 2010

Das Vorgehen war hier wie in den vorangegangenen Jahren. In den 300 ursprünglich ausgewählten PLZ wurden Neuzugänge mit derselben Ziehungswahrscheinlichkeit ausgewählt, mit der in Welle 1 der Bestand in dieser PLZ ausgewählt wurde.

2.) Eine Bestandsauffrischung (Aufstockung) der bestehenden Stichproben

Da die 300 bisher verwendeten Points teilweise sehr stark ausgeschöpft sind, wurden 100 neue Points für

i) die Bestandsauffrischung (Aufstockung) der BA-Stichprobe(n) und

ii) die Bestandsauffrischung (Aufstockung) der Bevölkerungsstichprobe (EWO-Aufstockung) gezogen.

Dabei wurde eine möglichst hohe Vergleichbarkeit der Prozesse mit der ursprünglichen Ziehung angestrebt. Da die Bevölkerungsstichprobe nicht mehr aus den Microm-Adressen gezogen wurde, sondern aus den Registern der Einwohnermeldeämter, wurde das in der ersten Welle verwendete Konzept (Rudolph & Trappmann 2007) modifiziert. Auf eine disproportionale Schichtung der Bevölkerungsstichprobe nach sozio-ökonomischem Status wurde ganz verzichtet. Darüber hinaus traten insbesondere zwei Schwierigkeiten auf, die eine vollständige Übertragung des alten Konzepts auf das neue Stichprobenkonzept verhinderten:

i) Im Rahmen einer Einwohnermeldeamtsstichprobe können nur Personen, keine Haushalte gezogen werden.

ii) Erstrecken sich ausgewählte PLZ über mehrere Gemeinden, so war zu entscheiden, aus welcher Gemeinde die Adressen gezogen werden sollten.

6.1.1 Auswahl weiterer Primary Sampling Units (PSUs)

Sampling Points, kurz PSU (Primary Sampling Units), in PASS sind Postleitzahlen (PLZ). Die PSUs sind in beiden Teilstichproben gleich. Die Auswahl der Sampling Points erfolgte proportional zur Größe der Points (pps-Ziehung), da sich dann bei gleicher Stichprobengröße innerhalb jedes Points die Größe der Points aus der zweistufigen Auswahl herauskürzt. Die Größe eines Points (das sogenannte Bedeutungsgewicht) wird idealerweise als Anzahl der Haushalte im Point definiert, da die zu erhebenden Einheiten in PASS Haushalte sind. Da jedoch in Einwohnermeldeamtsstichproben nur Personen (ab 18 Jahren) und keine Haushalte ausgewählt werden können, musste in diesem Fall auf Personen zurückgegriffen werden (s.u.).

Die Auswahl der PSUs erfolgte zudem geschichtet. Die Schichtung erfolgte nach Ortsgrößenklassen (PASS verwendet die zehnstufige BIK-Gemeindegrößenklasse) und Kreisen. Dabei erhielt jede Zelle aus Ortsgrößenklasse x Kreis ein Bedeutungsgewicht, das der Anzahl der Einheiten in der Zelle entspricht. Während sich jede Gemeinde in Deutschland eindeutig einem Feld in dem durch Ortsgrößenklasse und Kreis aufgespannten Allokationstableau zuordnen lässt, ist dies für die PLZ, die die PSUs in PASS darstellen, nicht der Fall. Jede PLZ musste daher einer Zelle zugeordnet werden. Dies erfolgte so, dass die PLZ der Zelle, in der die meisten Einheiten leben, zugeordnet wurde. Um die Designgewichte anschließend unverzerrt schätzen zu können, war es dann jedoch auch notwendig, die dort lebenden Einheiten dem Bedeutungsgewicht dieser Zelle zuzuschlagen. Insofern war es notwendig, die Zahl der Einheiten pro PLZ und pro Gemeinde zu kennen.

Da man innerhalb der PLZ faktisch nur eine Stichprobe von Personen ab 18 Jahren ziehen kann, ist das Bedeutungsgewicht, d.h. die Zahl der Einheiten idealerweise auch als Personen ab 18 Jahren definiert. Für die Ziehung wurde allerdings auf die den Mitgliedsinstituten vom ADM e.V. zur Verfügung gestellten Daten und den infas von infas Geodaten vorliegenden Daten zurückgegriffen, in denen lediglich die Zahl der Personen ab 15 Jahren pro PLZ und pro Gemeinde enthalten sind. Deshalb wurde für die Schichtung das Bedeutungsgewicht über die Zahl der 15-Jährigen und älteren Personen bestimmt. Auf dieser Basis erfolgte die Auswahl von 100 PLZ, die proportional zum Bedeutungsgewicht auf die Zellen des Allokationstableaus verteilt wurden. Die Allokation erfolgte mit Hilfe des Cox-Algorithmus.

6.1.2 Auswahl der Gemeinden für die Aufstockung der Bevölkerungsstichprobe

Die Auswahl der Personen erfolgte aus den Registern der Meldeämter von Gemeinden. Bei den PSUs in PASS handelt es sich aber um PLZ. Aus diesen PLZ kann keine einfache Personenstichprobe gezogen werden, wenn PLZ Gemeindegrenzen überschreiten. Der Aufwand, aus allen Gemeinden der PLZ Meldeamtsstichproben zu ziehen, wäre im ländlichen Raum teils beträchtlich. Von den 100 gezogenen PLZ erstreckten sich 19 über mehr als eine Gemeinde, teilweise über mehr als zehn Gemeinden.

Insofern war es notwendig, eine weitere Auswahlstufe zwischenschalten. Da die Verteilung der Personen auf die Gemeinden innerhalb der PLZ bekannt war (auch hier musste als Näherungswert die Verteilung der 15-Jährigen und Älteren verwendet werden, da die

Verteilung der 18-Jährigen und Älteren nicht vorlag), konnte eine bevölkerungsproportionale Auswahl einer Gemeinde innerhalb der PLZ durchgeführt werden. Die Auswahlwahrscheinlichkeit der Gemeinden innerhalb der PLZ wurde also proportional zum Anteil der Personen in PLZ x Gemeinde an allen Personen in der PLZ bestimmt und so eine Gemeinde für die Ziehung der Personenadressen ausgewählt. Diese zusätzliche Auswahlstufe erschwert zwar die Varianzschätzung, erhält jedoch die geringe Varianz der Designgewichte, ohne einen Bias einzuführen.

6.1.3 Auswahl der Haushalte für die Aufstockung der Bevölkerungsstichprobe

Die Auswahl der Personen erfolgte in den Gemeinden, in denen die gezogenen PLZ liegen. Dazu wurden die Gemeinden angeschrieben und gebeten, eine Zufallsziehung von Personen ab 18 Jahren aus den Melderegistern ausschließlich in den ausgewählten PLZ zu ziehen.

Daraus ergibt sich eine Personenstichprobe mit folgender Auswahlwahrscheinlichkeit:

$$(1) q_{ib} = (Q_i / Q) \times k \times (q_i / Q_i) = k \times (q_i / Q)$$

p_{ia} Auswahlwahrscheinlichkeit Haushalt a in Point i

q_{ib} Auswahlwahrscheinlichkeit Person b in Point i

Q GG Personen ab 18 Jahren in Bevölkerung

Q_i GG Personen ab 18 Jahren in Point i

k Anzahl Sample Points

q_i Stichprobengröße Personen ab 18 Jahren in Point i

q_{ia} Zahl der Personen ab 18 Jahren im Haushalt a in Point i

Aus Formel (1) kürzt sich Q_i heraus. Dadurch wird das Ideal einer gleichen Auswahlwahrscheinlichkeit der Personen in allen Points erreicht, wenn die Zahl der ausgewählten Personen q_i in jedem Point gleich groß ist.

Um die Auswahlwahrscheinlichkeit eines Haushalts zu bestimmen, ist zusätzlich die Information notwendig, wie viele Personen ab 18 Jahren in dem Haushalt leben. Diese Information lässt sich allerdings erst ex-post aus dem Survey, d.h. nur für die realisierten Fälle, ermitteln.

Es gilt:

$$(2) p_{ia} = (Q_i / Q) \times k \times (q_i \times q_{ia} / Q_i) = k \times (q_i \times q_{ia} / Q)$$

Der negative (weil vergrößernde) Effekt unterschiedlicher Haushaltsgrößen q_{ia} auf die Varianz der Ziehungswahrscheinlichkeiten ist ohne a priori-Kenntnis der Gruppierung der Personen zu Haushalten nicht vermeidbar.

Die genannte Zahl der Personen ab 18 Jahren pro PLZ x Gemeinde lag allerdings nicht vor, so dass stattdessen die Zahl der Personen ab 15 Jahren verwendet werden musste.

Empirisch sollte jedoch die Korrelation zwischen der Zahl der Personen ab 18 Jahren und der Zahl der Personen ab 15 Jahren auf PLZ-Ebene nahe genug bei 1 liegen, so dass die varianzreduzierende Eigenschaft des pps-Designs weitgehend erhalten bleibt.

6.1.4 Designgewichte für die Aufstockung der Bevölkerungsstichprobe

Die Designgewichte für die EWO-Aufstockungsstichprobe (sample = 6) sind als Kehrwert der in 6.1.2 dargestellten Auswahlwahrscheinlichkeiten auf den verschiedenen Stufen des Ziehungsdesigns definiert. Die Auswahlwahrscheinlichkeiten bestimmen sich dabei über drei Auswahlstufen. Die Auswahlwahrscheinlichkeit des Points, die Auswahlwahrscheinlichkeit der Gemeinde in den Fällen, in denen sich die PLZ über mehrere Gemeinden erstreckt (in allen anderen Fällen = 1) und die Auswahlwahrscheinlichkeit der Person im Point. Die Multiplikation dieser drei Auswahlwahrscheinlichkeiten ergibt die Auswahlwahrscheinlichkeit der ausgewählten Person in der Bruttostichprobe.

Die Transformation der Personenstichprobe in eine Haushaltsstichprobe stellt einen zusätzlichen Schritt bei der Aufstockung der Bevölkerungsstichprobe dar, der nur für die realisierten Fälle durchgeführt werden kann. Dieser zusätzliche Gewichtungsschritt, der die unterschiedlichen Auswahlwahrscheinlichkeiten auf Grund der unterschiedlichen (reduzierten) Haushaltsgröße korrigiert, wurde nach der Berechnung der Teilnahmepropensities durchgeführt, d.h. nach dem Übergang von der Bruttostichprobe zur Nettostichprobe, indem die Auswahlwahrscheinlichkeit der Personen mit der geschätzten Teilnahmepropensity und der Zahl der Zielpersonen im Haushalt multipliziert wurde.

6.1.5 Auswahl der BA-Aufstockungsstichprobe

Im IAB wurde parallel zu Ziehung der Einwohnermeldeamtsstichprobe eine Bestandsauffrischung (Aufstockung) der ALG II-Stichproben in denselben PLZ gezogen (sample = 7). Da die Auswahlwahrscheinlichkeit einer PLZ bereits bestimmt war und der Bevölkerungsgröße in diesem Point entsprach, die Zahl der Leistungsempfänger im Point aber zusätzlich mit der Bedürftigenquote des Points variiert, wären bei identischen q_i innerhalb der Points stark unterschiedliche Auswahlwahrscheinlichkeiten entstanden. Insbesondere hätten dann Leistungsempfänger in Gebieten mit hoher Bedürftigenquote geringere Auswahlwahrscheinlichkeiten gehabt. Dies war nicht erwünscht.

In Welle 1 wurde bereits eine Lösung für dieses Problem entwickelt (vgl. Rudolph und Trappmann 2007, 82). Die Zahl der auszuwählenden Leistungsempfänger pro Point wird als Produkt des festen $q_i=q$ (Stichprobengröße Personen pro Point) in der Bevölkerungsstichprobe mit dem Quotienten aus Leistungsempfängerquote im Point und der Leistungsempfängerquote im Bund gewählt. Zudem bietet die Datenbasis für die Ziehung der Leistungsempfängerstichprobe eine Zusammenfassung der Personen zu Bedarfsgemeinschaften, die näherungsweise Haushalten entsprechen. Somit konnten in dieser Quelle innerhalb jeder PLZ die interessierenden Einheiten direkt gezogen werden. Formel 4.4 in

Rudolph und Trappmann (2007) gibt die Auswahlwahrscheinlichkeiten einer Bedarfsgemeinschaft in dieser Stichprobe wieder.

Um sich über das Konzept einer Bestandsaufstockung der BA-Stichproben klarzuwerden, ist es nützlich, sich zunächst zu vergegenwärtigen, aus welcher Grundgesamtheit die bisherigen Prozessdatenstichproben ausgewählt wurden: Die BA-Stichprobe der Welle 1 ist eine Zufallsstichprobe von Haushalten mit Grundsicherungsbezug zum Stichtag im Juli 2006. Die Auffrischungen der Wellen 2, 3, 4 (und nun 5; s.o.) sind jeweils Zufallsstichproben von Haushalten mit Grundsicherungsbezug im Juli 2007 bzw. 2008 bzw. 2009 bzw. 2010, in denen keine Person lebt, die zu einem der vorherigen Stichtage Grundsicherungsleistungen bezogen hat („Neuzugänge“). Die gesamte BA-Stichprobe ist in Welle 5 somit eine Zufallsstichprobe von Haushalten, die in (mindestens) einem Juli der Jahre 2006, 2007, 2008, 2009 oder 2010 Grundsicherungsleistungen bezogen haben.

Ist es sinnvoll genau diesen Bestand aufzufrischen? Dies ist sicherlich nicht der Fall. Leistungsempfänger des Jahres 2006, die den Bezug inzwischen verlassen haben, sind für die SGB II – Forschung nicht von ausreichendem Interesse. Zudem können sie mit der ursprünglichen Stichprobe von mehr als 6.000 Haushalten mit Bezug im Juli 2006 hinreichend untersucht werden.

Nun kommt eine zweite Eigenschaft dieses spezifischen Auffrischungskonzepts zum Tragen. Alle Haushalte der fünf genannten BA-Stichproben (Welle 1 plus 4 Auffrischungen), die auch zum letzten Stichtag (Juli 2010) noch Grundsicherungsleistungen beziehen, stellen eine Zufallsstichprobe des Bestands zu diesem Zeitpunkt dar. Dieser Bestand wurde nun mit der Bestandsaufstockung aufgefrischt, um auch in Zukunft ausreichende statistische Power für Analysen insbesondere zu relevanten Subpopulationen (z. B. Jugendliche (U25), Personen mit Migrationshintergrund, Alleinerziehende, Aufstocker) im Leistungsbezug zu haben.

Analytisch ist es für die Berechnung der Designgewichte sinnvoll, die Fälle aus der Bestandsaufstockung BA in solche zu unterscheiden, die zur Population des Bestands (Welle 4) gehören und solche, die zur Population des Zugangs (zwischen Welle 4 und Welle 5) gehören. Diese Unterscheidung wurde vorgenommen, indem die Fälle aus der Bestandsaufstockung, bei denen es sich um Zugänge handelt, auf $sample = 8$ (Zugangsstichprobe) gesetzt wurden. Sie wurden bei der Designgewichtung wie Zugänge behandelt.

6.2 Integration der Aufstockungstichproben

Für die Gewichtung ist die beschriebene Form der Aufstockung nicht trivial, da mehrfach Gewichte integriert werden müssen. Diese Integration erfolgte wie bisher nach den Propensity-Modellen, aber vor der Kalibrierung.

Eine vergleichsweise wenig rechenintensive Möglichkeit, bei der nicht je Einzelfall die Auswahlwahrscheinlichkeiten in den jeweils anderen Stichproben (rückwirkend) bestimmt werden müssen, besteht darin, zunächst die Fälle der Bestandstichproben zu identifizieren, die nicht mehr zur Grundgesamtheit der Bestandsaufstockungen gehören. Die Ge-

wichte dieser Teilmenge sind durch die Aufstockung nicht beeinträchtigt. Die verbleibende Bestandsstichprobe und die Bestandsaufstockungen sind nun jeweils zwei Zufallsstichproben derselben Grundgesamtheit mit bekannten Inklusionswahrscheinlichkeiten. Auf die Fusion dieser beiden Stichproben kann somit das Konzept der Konvexkombination (Spieß und Rendtel 2000) angewendet werden. Die Integration der Bestandsaufstockungen erfolgte also durch Konvexkombination der jeweiligen Stichproben.

Die Gewichte der vereinigten Bevölkerungsstichproben sollen die Microm-Stichprobe und die EWO-Aufstockung auf die Haushalte der Bundesrepublik hochrechnen. Deshalb wurden zunächst für die Microm-Bestandsstichprobe und die EWO-Aufstockung jeweils nach dem bisherigen Konzept separate Gewichte berechnet. Für das Bevölkerungsgewicht vor Kalibrierung wurde dann der Bestand Microm mit der EWO-Aufstockung (sample = 6) über eine Konvexkombination integriert.

Die Gewichte der vereinigten BA-Stichproben sollen die Ursprungs-BA-Stichprobe, die bis dahin vier Auffrischungen durch Neuzugänge und die nun beschriebene Bestandsaufstockung weiterhin auf alle Haushalte hochrechnen, die in einem Juli der Jahre 2006-2010 Leistungen der Grundsicherung bezogen haben, und die Teilmenge hieraus, die im Juli 2010 noch Leistungen der Grundsicherung bezieht, soll auf alle Grundsicherungsempfänger im Juli 2010 hochgerechnet werden.

Für die BA-Stichprobe ohne die Bestandsaufstockung und für die Bestandsaufstockung wurden zunächst jeweils nach dem bisherigen Konzept separate Gewichte berechnet. Die BA-Bestandsaufstockung (sample = 7) musste dann mit noch ALG II beziehenden Fällen aus allen BA-Stichproben der Wellen 1-4 integriert werden (Konvexkombination). Fälle aus den BA-Stichproben der Wellen 1-4, die kein ALG II mehr beziehen, behalten an dieser Stelle ihr Gewicht. Hieraus ergab sich das neue BA-Gewicht vor Kalibrierung.

Da die neue Zugangsstichprobe BA (sample = 8) und die Bestände disjunkt sind, behalten bei der Integration alle Fälle ihr Gewicht. Eine Fusion der Stichproben für die Berechnung des BA-Gewichts vor Kalibrierung war nicht notwendig.

Im Anschluss daran erfolgte wie bisher die Integration von Bevölkerungsgewichten und BA-Gewichten zu Gesamtgewichten.

6.3 Designgewichte für die Bestandshaushalte der Welle 4

Unter Berücksichtigung von Zuzügen aus dem Inland wurden für die vierte Welle nunmehr aus den Haushaltsquerschnittsgewichten der dritten Welle neue „Haushaltsdesigngewichte“ für die Welle 4 generiert. Dies erfolgte erneut über das für Welle 2 beschriebene „Weight-Share“-Verfahren (vgl. Gebhardt et.al, FDZ Datenreport 06/2009). Geburten, Todesfälle oder Auszüge haben keinen Einfluss auf das Gewicht, Zuzüge aus dem Inland erhöhen jedoch die Inklusionswahrscheinlichkeit eines Haushalts, da auch die zugezogenen Personen die Chance hatten, in Welle 1, Welle 2 oder Welle 3 (nur Auffrischungsstichprobe BA) ins Sample zu gelangen. Das neue Designgewicht für Teilstichprobe i dw_{i,hh_4} errechnet sich also aus dem alten Querschnittsgewicht wq_{i,hh_3} :

$$1/dw_{i, hh_4} = 1/wq_{i, hh_3} + (n_{\text{sample } i} / n_{\text{population } i})$$

Das neue Designgewicht ist nur ein Zwischenschritt und daher in der Datenlieferung der Welle 4 nicht enthalten.

6.4 Designgewichte für die Zugangsstichprobe der Welle 5

In der Welle 5 wurde das Panel erneut durch Nachziehung von Neuzugängen zum Leistungsbezug aufgefrischt. Eine Ziehungschance hatten alle Haushalte, die im Juli 2010 Leistungen bezogen haben, im selben Monat 2009, 2008, 2007 und 2006 jedoch keine Auswahlwahrscheinlichkeit für die Prozessdatenstichprobe hatten. Diese Auffrischung ließ sich realisieren, indem nur Bedarfsgemeinschaften ausgewählt wurden, in denen kein Mitglied im Juli der vorangegangenen Jahre Leistungen bezogen hatte. Nachgezogen wurde wiederum in den 300 Points der ersten Welle und den 100 Aufstockungspoints der Welle 5. Die Größe der Stichprobe war in Analogie zum speziellen pps-Verfahren bei der Ziehung der ersten Prozessdatenstichprobe, das in Rudolph und Trappmann (2007) beschrieben ist, proportional zum Anteil der Neuzugänge an der Bevölkerung im Point (zum Zeitpunkt der Auswahl der Points). Die Berechnung der Designgewichte ist dort ebenfalls beschrieben. Das Designgewicht der Zugangsstichprobe befindet sich für Fälle mit *sample = 8* in der Variable *dw_ba*.

6.5 Wiederteilnahme-Propensity Haushalte

In diesem Schritt wurde wiederum analog zum Vorgehen für Welle 4 für jeden Haushalt, der in der Welle 4 teilgenommen hatte, die Wahrscheinlichkeit zur Wiederteilnahme an Welle 5 über Logit-Modelle für Panelbereitschaft, Erreichbarkeit und Teilnahme geschätzt. Auch Haushalte mit letzter Teilnahme in Welle 3, ohne Teilnahme in Welle 4 (temporäre Ausfälle), wurden in der Modellierung für Welle 5 berücksichtigt. Neben Variablen aus dem Haushaltsinterview und dem Personeninterview des Haushaltsvorstands der Vorwelle gingen hier auch Variablen ein, die mit der Feldarbeit zusammenhängen, z.B. die Zahl der Kontaktversuche. Die vorhergesagten Wahrscheinlichkeiten aller drei Modelle wurden multipliziert. Der Kehrwert dieses Produkts ist für jede Welle in der Variable *hpbleib* zu finden. Das Längsschnittgewicht eines Haushalts aus einer der Stichproben der Welle 1 für den gesamten möglichen Zeitraum $[t_1, t_2, t_3, t_4, t_5]$ zwischen allen fünf Wellen lässt sich dann als Produkt aus Querschnittgewicht zu t_1 , *hpbleib* (Welle 1 zu Welle 2) und *hpbleib* (Welle 2 zu Welle 3 usw.) berechnen (vgl. hierzu auch Kapitel 9.4 des User Guides (Bethmann und Gebhardt 2011)).

Tabelle 35: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der Wiederholerhaushalte

Variablenkürzel und Referenzkategorie	Erläuterung
alter_1 alter_2 alter_4 alter_5 Referenzkategorie	Haushaltsauskunftsperson (HBV) jünger als 30 Jahre HBV 30-39 Jahre alt HBV 50-64 Jahre alt HBV 65 Jahre und älter HBV 40-49 Jahre alt
sex_1 Referenzkategorie	HBV Männlich HBV Weiblich
nichtdeutsch Referenzkategorie	HBV hat nicht-deutsche Staatsangehörigkeit HBV hat deutsche Staatsangehörigkeit oder fehlende Angabe
schulbil_1 schulbil_2 schulbil_4 Referenzkategorie	Schulabschluss HBV: ohne Abschluss Schulabschluss HBV: Volks-/Hauptschule Schulabschluss HBV: Fach-/Hochschulreife Schulabschluss HBV: Realschule/mittlerer Abschluss/noch Schüler
gesundheit_3 gesundheit_4 gesundheit_5 Referenzkategorie	Subjektive Beurteilung des Gesundheitszustands HBV: Zufriedenstellend Subjektive Beurteilung des Gesundheitszustands HBV: Weniger gut Subjektive Beurteilung des Gesundheitszustands HBV: Schlecht Subjektive Beurteilung des Gesundheitszustands HBV: Sehr gut bis Gut
zufrieden_1 zufrieden_2 zufrieden_3 Referenzkategorie	Generelle Lebenszufriedenheit HBV: Skalenwert 0-2 Generelle Lebenszufriedenheit HBV: Skalenwert 3-5 Generelle Lebenszufriedenheit HBV: Skalenwert 6-8 Generelle Lebenszufriedenheit HBV: Skalenwert 9-10
anz_0_3 anz_4_6 anz_7_14 anz_65 Referenzkategorie	Anzahl Personen im Haushalt im Alter 0-3 Jahre Anzahl Personen im Haushalt im Alter 4-6 Jahre Anzahl Personen im Haushalt im Alter 7-14 Jahre Anzahl Personen im Haushalt im Alter 65 Jahre und älter Anzahl Personen im Haushalt im Alter 15-64 Jahre
eigentum Referenzkategorie	Art des Wohneigentums: Eigentümer Art des Wohneigentums: Mieter, fehlende Angabe
wnka_1 wnka_3 Referenzkategorie	Anzahl w.n.- und k.A.-Angaben in Haushalts- und Personeninterview des HBV: keine Anzahl w.n.- und k.A.-Angaben in Haushalts- und Personeninterview des HBV: 11 und mehr Anzahl w.n.- und k.A.-Angaben in Haushalts- und Personeninterview des HBV: 1-10
hhincome_1 hhincome_2 hhincome_4 Referenzkategorie	Haushaltseinkommen: bis 870 Euro Haushaltseinkommen: 871-1.400 Euro Haushaltseinkommen: mehr als 2.200 Euro Haushaltseinkommen: 1.401-2.200 Euro
alg2_1 Referenzkategorie	ALG II-Bezug des Haushalts: bezieht aktuell ALG II ALG II-Bezug des Haushalts: bezieht aktuell kein ALG II
stichprobe_ba Referenzkategorie	BA-Stichproben Microm-Stichprobe
anzkon_1 anzkon_3 anzkon_4 Referenzkategorie	Anzahl Kontaktversuche CATI/CAPI: 1 Kontaktversuch Anzahl Kontaktversuche CATI/CAPI: 4-9 Kontaktversuche Anzahl Kontaktversuche CATI/CAPI: 10 und mehr Kontaktversuche Anzahl Kontaktversuche CATI/CAPI: 2-3 Kontaktversuche

Tabelle 35: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der Wiederholerhaushalte (Fortsetzung)

Variablenkürzel und Referenzkategorie	Erläuterung
blinealt_2 Referenzkategorie	Neue Bundesländer Alte Bundesländer
bundesld_1 bundesld_2 bundesld_3 bundesld_4 bundesld_6 bundesld_7 bundesld_8 bundesld_9 bundesld_10 bundesld_11 bundesld_12 bundesld_13 bundesld_14 bundesld_15 bundesld_16 Referenzkategorie	Bundesland: Schleswig-Holstein Bundesland: Hamburg Bundesland: Niedersachsen Bundesland: Bremen Bundesland: Hessen Bundesland: Rheinland-Pfalz Bundesland: Baden-Württemberg Bundesland: Bayern Bundesland: Saarland Bundesland: Berlin Bundesland: Brandenburg Bundesland: Mecklenburg-Vorpommern Bundesland: Sachsen Bundesland: Sachsen-Anhalt Bundesland: Thüringen Bundesland: Nordrhein-Westfalen
bik_1 bik_2 bik_3 bik_4 bik_5 bik_6 bik_7 bik_8 bik_9 Referenzkategorie	BIK-Größenklasse der Gemeinde: unter 2.000 Einwohner BIK-Größenklasse der Gemeinde: 2.000 bis u. 5.000 Einwohner BIK-Größenklasse der Gemeinde: 5.000 bis u. 20.000 Einwohner BIK-Größenklasse der Gemeinde: 20.000 bis u. 50.000 Einwohner BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner STYP 2/3/4 BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner STYP 1 BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner STYP 2/3/4 BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner STYP 1 BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner STYP 2/3/4 BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner STYP 1

Tabelle 36: Logit-Modelle zur Wiederteilnahme für Panelbereitschaft, Erreichbarkeit und Teilnahme

	Panelbereitschaft		Kontakt		Teilnahme	
	Coef.	p	Coef.	p	Coef.	p
alter_1	-.5061303	0.436	-.505075	0.002	-.4977961	0.000
alter_2	-.8166375	0.160	-.0665502	0.696	-.1431984	0.146
alter_4	-.3234635	0.578	.2821844	0.095	.2729491	0.003
alter_5	1.268026	0.339	-.5361496	0.342	.1959219	0.443
sex_1	-.0451475	0.909	-.2183901	0.062	.0593323	0.383
nichtdeutsch	-.4626094	0.366	-.005286	0.980	-.3594525	0.002
schulbil_1	-.5154318	0.462	.1356257	0.623	-.3029301	0.042
schulbil_2	.2333508	0.632	-.02989	0.824	-.1522689	0.057
schulbil_4	-.5157242	0.291	.1970259	0.207	.1040593	0.246
gesundheit_3	-.4551552	0.310	-.3100665	0.019	.0972873	0.215
gesundheit_4	-.3220093	0.564	.0304291	0.862	.1669037	0.092
gesundheit_5	.0792266	0.926	-.1879924	0.428	.0177493	0.899

Tabelle 36: Logit-Modelle zur Wiederteilnahme für Panelbereitschaft, Erreichbarkeit und Teilnahme (Fortsetzung)

	Panelbereitschaft		Kontakt		Teilnahme	
	Coef.	p	Coef.	p	Coef.	p
zufrieden_1	-.5080861	0.636	.2813634	0.400	-.227833	0.204
zufrieden_2	-.3751184	0.643	-.0589571	0.768	.0437246	0.705
zufrieden_3	-.6125906	0.423	.0826564	0.643	.0637325	0.513
anz_0_3	-.0802828	0.881	-.0478796	0.760	.1097202	0.227
anz_4_6	-.2770277	0.610	.0272309	0.880	.0811451	0.435
anz_7_14	-.0220602	0.953	-.0371046	0.727	.0142988	0.808
anz_65	-1.315313	0.003	.7436965	0.057	.2930039	0.057
eigentum	-1.054304	0.059	.6599214	0.002	.115887	0.237
wnka_1	-1.030373	0.050	.1257656	0.292	.2129241	0.002
wnka_3	-2.071453	0.000	.160941	0.375	.1815859	0.080
hhincome_1	-.718591	0.248	-.2762657	0.108	.0264033	0.796
hhincome_2	-.6636692	0.231	-.095299	0.547	.0818454	0.374
hhincome_4	.5868573	0.492	.3668512	0.072	.3587214	0.001
alg2_1	-.216014	0.622	.1574547	0.237	.1499516	0.068
stichprobe_ba	-2.763789	0.013	-.4007235	0.014	-.135382	0.122
blneualt_2	1.33845	0.017				
bundesld_1			-.737057	0.003	-.4660793	0.004
bundesld_2			-.246651	0.564	.0866672	0.741
bundesld_3			.2156854	0.386	-.2355021	0.063
bundesld_4			-.5211735	0.303	.0639558	0.860
bundesld_6			-.5169035	0.029	-.3099269	0.033
bundesld_7			.0978482	0.769	-.554962	0.001
bundesld_8			-.3620773	0.097	-.4724849	0.000
bundesld_9			-.1783082	0.417	-.2047443	0.088
bundesld_10			-.2324686	0.636	-.4898657	0.067
bundesld_11			-.5371693	0.030	-.1109548	0.480
bundesld_12			.2199858	0.524	-.2081159	0.235
bundesld_13			.1195693	0.752	.3258527	0.190
bundesld_14			-.4243011	0.103	.5221029	0.004
bundesld_15			-.0007189	0.998	-.3191506	0.042
bundesld_16			.3127281	0.407	.0391422	0.844
bik_1			.3933182	0.706	.1801713	0.701
bik_2			.7116941	0.114	-.2224798	0.237
bik_3			.0281186	0.917	.1902125	0.194
bik_4			-.4132198	0.058	.2128612	0.120
bik_5			-.2750239	0.262	.1748974	0.232
bik_6			-.1094303	0.733	.0633053	0.755
bik_7			-.2606259	0.230	.2240069	0.073
bik_8			-.1401822	0.445	.0418463	0.691
bik_9			.223133	0.461	.0839234	0.573
anzkon_1			-.5785529	0.006		
anzkon_3			-.3378294	0.044		
anzkon_4			-1.23855	0.000		
cons	10.71506	0.000	4.128693	0.000	1.504246	0.000
n	7848		7818		7459	
Log likelihood	-166.02321		-1315.2448		-3139.9694	
PseudoR2	0.1570		0.0971		0.0447	

6.6 Teilnahme-Propensity erstbefragte Split-Haushalte

In diesem Schritt wurden die Teilnahme-Propensities für neue Split-Haushalte, d.h. Haushalte, die aufgrund des Auszugs einer Person der Panelstichprobe in einen neuen Haushalt in das Panel aufgenommen werden, berechnet. Hierbei wurden nur solche Split-Haushalte berücksichtigt, die in den vorausgegangenen Wellen noch nicht befragt wurden. Modelliert wurden also die Teilnahmepropensities für die erstbefragten Split-Haushalte, getrennt nach dem Kriterium in Welle 4 entstandener Split-Haushalt (Split-W4-Haushalte) und in Welle 5 entstandener Split-Haushalt (Split-WELLE 5-Haushalte). Die Wahrscheinlichkeit zur Wiederteilnahme wurde jeweils über Logit-Modelle für Erreichbarkeit und Teilnahme geschätzt. Fehlende zeitstabile Informationen über die Haushaltsauskunftsperson (HBV) wurden, falls notwendig, aus der Vorwelle ergänzt. Die vorhergesagten Wahrscheinlichkeiten der beiden Modelle wurden multipliziert. Der Kehrwert des Produkts ist für die Split-Haushalte ebenfalls in der Variable *hpleib* zu finden.

Tabelle 37: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der erstbefragten Split-Haushalte (Welle 4 und Welle 5)

Variablenkürzel und Referenzkategorie	Erläuterung
alter_1 alter_2 alter_4 alter_5 Referenzkategorie	Haushaltsauskunftsperson (HBV) jünger als 30 Jahre HBV 30-39 Jahre alt HBV 50-64 Jahre alt HBV 65 Jahre und älter HBV 40-49 Jahre alt
sex_1 Referenzkategorie	HBV Männlich HBV Weiblich
nichtdeutsch Referenzkategorie	HBV hat nicht-deutsche Staatsangehörigkeit HBV hat deutsche Staatsangehörigkeit oder fehlende Angabe
schulbil_1 schulbil_2 schulbil_4 Referenzkategorie	Schulabschluss HBV: ohne Abschluss Schulabschluss HBV: Volks-/Hauptschule Schulabschluss HBV: Fach-/Hochschulreife Schulabschluss HBV: Realschule/mittlerer Abschluss/noch Schüler
stichprobe1 stichprobe3 stichprobe4 stichprobe5 Referenzkategorie	BA-Stichprobe Auffrischungsstichprobe (BA) Welle 2 Auffrischungsstichprobe (BA) Welle 3 Auffrischungsstichprobe (BA) Welle 4 Microm-Stichprobe
anzkon_1 anzkon_3 anzkon_4 Referenzkategorie	Anzahl Kontaktversuche CATI/CAPI: 1 Kontaktversuch Anzahl Kontaktversuche CATI/CAPI: 4-9 Kontaktversuche Anzahl Kontaktversuche CATI/CAPI: 10 und mehr Kontaktversuche Anzahl Kontaktversuche CATI/CAPI: 2-3 Kontaktversuche

Tabelle 38: Logit-Modelle zur Erstteilnahme von Split-Welle 4-Haushalten für Erreichbarkeit und Teilnahme

	Kontakt		Teilnahme	
	Coef.	p	Coef.	p
alter_1	-.3169316	0.624	-.2019633	0.781
alter_2	.807183	0.350	-.597887	0.474
alter_4	.7431398	0.148	-1.09883	0.081
alter_5	.3700588	0.772	-.0205646	0.986
sex_1	-1.3948	0.004	-.1425814	0.806
nichtdeutsch	1.206919	0.304		
schulbil_1	-.7347448	0.446	.3429	0.773
schulbil_2	-.2239878	0.652	.1362247	0.809
schulbil_4	.5523845	0.384	-.5304055	0.471
anzkon_1	-1.720018	0.001		
anzkon_3	.9066009	0.234		
anzkon_4	-.7229558	0.561		
stichprobe1	-.9706771	0.055	-.3784742	0.477
stichprobe3	.4640556	0.686	-.1573479	0.858
stichprobe4	-1.772916	0.070		
cons	3.069574	0.000	-1.177896	0.024
n	208		154	
Log likelihood	-75.427106		-58.556792	
Pseudo R²	0.1858		0.0454	

Tabelle 39: Logit-Modelle zur Erstteilnahme von Split-Welle 5-Haushalten für Erreichbarkeit und Teilnahme

	Kontakt		Teilnahme	
	Coef.	p	Coef.	p
alter_1	-.169106	0.798	-.7357389	0.130
alter_2	-.2135482	0.731	-.2502378	0.553
alter_4	-.4650614	0.358	-.1883364	0.550
alter_5	-.8130421	0.372	-.8585624	0.305
sex_1	.8770132	0.062	-.1019223	0.718
nichtdeutsch			-.1523006	0.802
schulbil_1	-1.65523	0.024	.8945813	0.169
schulbil_2	-.3336857	0.492	.4011616	0.199
schulbil_4	-.6393656	0.281	.4509772	0.231
anzkon_1	-1.277773	0.007		
anzkon_3	-.8179033	0.135		
anzkon_4	1.002637	0.362		
stichprobe1	-.1322999	0.790	-.2529847	0.415
stichprobe3	-.0239096	0.978	-.4737019	0.443
stichprobe4	.4028368	0.651	-.3263525	0.551
stichprobe5	-1.185011	0.097	-.2990074	0.637
cons	3.313871	0.000	-.849377	0.008
n	338		307	
Log likelihood	-92.172315		-175.50534	
Pseudo R²	0.1103		0.0205	

6.7 Ausfallgewichtung für Haushalte aus der BA-Auffrischungs- sowie der BA-Bestandsaufstockungsstichprobe der Welle 5

Für die Haushalte aus der Auffrischungsstichprobe der BA-Neuzugänge zum ALG II Bezug (sample = 8) sowie erstmalig auch für Haushalte einer Aufstockungsstichprobe des BA-Bestandes (sample = 7) wurde erneut eine zweistufige Ausfallmodellierung (Erreichbarkeit und Teilnahme) analog zur Auffrischungsstichprobe der Welle 4 vorgenommen. Die daraus abgeleitete Teilnahmewahrscheinlichkeit befindet sich in der Variable *prop_t0*. Da der Variablensatz und die Herkunft beider Teilstichproben identisch ist, wurde nur ein gemeinsames Ausfallmodell mit der Teilstichprobenkennung als Kontrollvariable gerechnet.

Tabelle 40: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der BA-Auffrischungs- bzw. BA-Aufstockungsstichprobe Welle 5

Variablenkürzel und Referenzkategorie	Erläuterung
alter_1 alter_2 alter_4 Referenzkategorie	Haushaltsauskunftsperson (HBV) jünger als 30 Jahre HBV 30-39 Jahre alt HBV 50-64 Jahre alt HBV 40-49 Jahre alt
sex_1 Referenzkategorie	HBV Männlich HBV Weiblich
nichtdeutsch Referenzkategorie	HBV hat nicht-deutsche Staatsangehörigkeit HBV hat deutsche Staatsangehörigkeit oder fehlende Angabe
stichprobe2 Referenzkategorie	Auffrischungsstichprobe (BA-Neuzugänge) Welle 5 Aufstockungsstichprobe (BA-Bestand) Welle 5
schulbil_1 schulbil_2 schulbil_4 schulbil_5 Referenzkategorie	Schulabschluss HBV: ohne Abschluss Schulabschluss HBV: Volks-/Hauptschulabschluss Schulabschluss HBV: Fach-/Hochschulreife Schulabschluss HBV: Keine Angabe Schulabschluss HBV: Realschule/mittlerer Abschluss/noch Schüler
anz_persBG_2 anz_persBG_3 Referenzkategorie	Anzahl Personen in der Bedarfsgemeinschaft: 2 Personen Anzahl Personen in der Bedarfsgemeinschaft: 3 und mehr Personen Anzahl Personen in der Bedarfsgemeinschaft: 1 Person
anz_verwfBG_1 anz_verwfBG_3 anz_verwfBG_4 Referenzkategorie	Anzahl erwerbsfähige Personen in der Bedarfsgemeinschaft: keine Person Anzahl erwerbsfähige Personen in der Bedarfsgemeinschaft: 2 Personen Anzahl erwerbsfähige Personen in der Bedarfsgemeinschaft: 3 und mehr Personen Anzahl erwerbsfähige Personen in der Bedarfsgemeinschaft: 1 Person
BG_typ_2 BG_typ_3 BG_typ_4 BG_typ_5 Referenzkategorie	Bedarfsgemeinschafts-Typ: Alleinerziehend Bedarfsgemeinschafts-Typ: Paar ohne Kinder Bedarfsgemeinschafts-Typ: Paar mit Kindern unter 18 Bedarfsgemeinschafts-Typ: sonstige Bedarfsgemeinschaft Bedarfsgemeinschafts-Typ: Alleinstehend
famstand_2 famstand_3 famstand_4 famstand_5 famstand_6 Referenzkategorie	Familienstand: verheiratet Familienstand: verwitwet Familienstand: geschieden Familienstand: getrennt lebend Familienstand: eheähnliche oder eingetragene Partnerschaft Familienstand: ledig

Tabelle 40: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der BA-Auffrischungs- bzw. BA-Aufstockungsstichprobe Welle 5 (Fortsetzung)

Variablenkürzel und Referenzkategorie	Erläuterung
bundesld_1 bundesld_2 bundesld_3 bundesld_4 bundesld_6 bundesld_7 bundesld_8 bundesld_9 bundesld_10 bundesld_11 bundesld_12 bundesld_13 bundesld_14 bundesld_15 bundesld_16 Referenzkategorie	Bundesland: Schleswig-Holstein Bundesland: Hamburg Bundesland: Niedersachsen Bundesland: Bremen Bundesland: Hessen Bundesland: Rheinland-Pfalz Bundesland: Baden-Württemberg Bundesland: Bayern Bundesland: Saarland Bundesland: Berlin Bundesland: Brandenburg Bundesland: Mecklenburg-Vorpommern Bundesland: Sachsen Bundesland: Sachsen-Anhalt Bundesland: Thüringen Bundesland: Nordrhein-Westfalen
bik_1 bik_2 bik_3 bik_4 bik_5 bik_6 bik_7 bik_8 bik_9 Referenzkategorie	BIK-Größenklasse der Gemeinde: unter 2.000 Einwohner BIK-Größenklasse der Gemeinde: 2.000 bis u. 5.000 Einwohner BIK-Größenklasse der Gemeinde: 5.000 bis u. 20.000 Einwohner BIK-Größenklasse der Gemeinde: 20.000 bis u. 50.000 Einwohner BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner STYP 2/ 3/ 4 BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner STYP 1 BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner STYP 2/ 3/ 4 BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner STYP 1 BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner STYP 2/ 3/ 4 BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner STYP 1
anzkon_1 anzkon_3 anzkon_4 Referenzkategorie	Anzahl Kontaktversuche CATI/CAPI: 1 Kontaktversuch Anzahl Kontaktversuche CATI/CAPI: 4-9 Kontaktversuche Anzahl Kontaktversuche CATI/CAPI: 10 und mehr Kontaktversuche Anzahl Kontaktversuche CATI/CAPI: 2-3 Kontaktversuche

Tabelle 41: Logit-Modelle zur Erstteilnahme für Erreichbarkeit und Teilnahme der BA-Auffrischungs- bzw. BA-Aufstockungstichprobe Welle 5

	Kontakt		Teilnahme	
	Coef.	p	Coef.	p
alter_1	-.0015206	0.992	-.0692774	0.402
alter_2	.2276912	0.139	-.0437775	0.582
alter_4	.246158	0.126	.1627027	0.037
sex_1	-.171073	0.142	-.010268	0.863
nichtdeutsch	-.2467698	0.079	-.3465132	0.000
stichprobe2	.3263159	0.005	.1669613	0.004
schulbil_1	-.4839556	0.005	-.283361	0.002
schulbil_2	-.3921248	0.009	-.1049563	0.141
schulbil_4	-.1269403	0.525	.1992542	0.031
schulbil_5	-.404894	0.040	-.1888125	0.055
anz_persBG_2	1.137636	0.021	.1648316	0.474
anz_persBG_3	.9215946	0.090	.3766071	0.145
anz_verwfBG_1	.1733314	0.833	-.3751382	0.510
anz_verwfBG_3	.2088258	0.494	-.0132664	0.916
anz_verwfBG_4	.6764131	0.198	.0427438	0.824
BG_typ_2	-.5693997	0.265	.0542831	0.819
BG_typ_3	-.7152423	0.168	-.0431914	0.852
BG_typ_4	-.2865151	0.603	-.0150196	0.951
BG_typ_5	-.4500966	0.230	-.0384718	0.861
famstand_2	.3633272	0.113	-.1309437	0.233
famstand_3	.7126889	0.244	-.0205194	0.927
famstand_4	.3735799	0.032	.1744604	0.039
famstand_5	.1017989	0.585	-.2219197	0.031
famstand_6	-.2185996	0.425	-.0498226	0.726
bundesld_1	-.441315	0.102	.0290337	0.858
bundesld_2	-.2615607	0.294	.0738912	0.623
bundesld_3	-.046195	0.823	-.0430577	0.675
bundesld_4	.8261968	0.088	-.1091292	0.636
bundesld_6	-.8061113	0.000	-.2272539	0.084
bundesld_7	-.9465943	0.000	.20312	0.172
bundesld_8	.115628	0.605	-.0744395	0.491
bundesld_9	.0074188	0.970	-.0037252	0.971
bundesld_10	.523671	0.479	.8561815	0.001
bundesld_11	.2789074	0.188	-.3266551	0.007
bundesld_12	.6622326	0.053	-.0233711	0.847
bundesld_13	-.5655013	0.079	-.0466962	0.807
bundesld_14	.5755524	0.098	.2690728	0.029
bundesld_15	.0071509	0.981	.0044224	0.975
bundesld_16	-.472969	0.101	.0559909	0.703
bik10_1	2.328615	0.024	-.2967277	0.295
bik10_2	.2433743	0.509	.0324813	0.877
bik10_3	.3823661	0.129	.1538863	0.197
bik10_4	1.059255	0.000	.0331483	0.741
bik10_5	.2207301	0.361	.4259753	0.001
bik10_6	.7946102	0.032	.6038444	0.000
bik10_7	.0115889	0.949	.1962604	0.060
bik10_8	.4518842	0.005	.0162977	0.849
bik10_9	.7751251	0.008	.2964757	0.011
anzkon_1	-1.814473	0.000		
anzkon_3	-.7893974	0.001		
anzkon_4	-2.369228	0.000		
cons	4.057722	0.000	-1.13626	0.000
n	8220		7763	
Log likelihood	-1500.1312		-4406.527	
Pseudo R²	0.1499		0.0220	

6.8 Ausfallgewichtung für Haushalte aus der EWO-Aufstockungsstichprobe der Welle 5

Für die Haushalte der erstmalig gezogenen EWO-Aufstockungsstichprobe für die allgemeine Bevölkerung (sample = 6) wurde eine zweistufige Ausfallmodellierung (Erreichbarkeit und Teilnahme) vorgenommen. Die daraus abgeleitete Teilnahmewahrscheinlichkeit befindet sich in der Variable *prop_t0*.

Tabelle 42: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der EWO-Aufstockungsstichprobe Welle 5

Variablenkürzel und Referenzkategorie	Erläuterung
alter_1 alter_2 alter_4 alter_5 Referenzkategorie	15-29 Jahre alt 30-39 Jahre alt 50-64 Jahre alt 65+ Jahre alt 40-49 Jahre alt
sex_1 Referenzkategorie	HBV Männlich HBV Weiblich
nichtdeutsch Referenzkategorie	HBV hat nicht-deutsche Staatsangehörigkeit HBV hat deutsche Staatsangehörigkeit oder fehlende Angabe
blneualt_2 Referenzkategorie	Neue Bundesländer Alte Bundesländer
bundesld_1 bundesld_2 bundesld_3 bundesld_4 bundesld_6 bundesld_7 bundesld_8 bundesld_9 bundesld_10 bundesld_11 bundesld_12 bundesld_13 bundesld_14 bundesld_15 bundesld_16 Referenzkategorie	Bundesland: Schleswig-Holstein Bundesland: Hamburg Bundesland: Niedersachsen Bundesland: Bremen Bundesland: Hessen Bundesland: Rheinland-Pfalz Bundesland: Baden-Württemberg Bundesland: Bayern Bundesland: Saarland Bundesland: Berlin Bundesland: Brandenburg Bundesland: Mecklenburg-Vorpommern Bundesland: Sachsen Bundesland: Sachsen-Anhalt Bundesland: Thüringen Bundesland: Nordrhein-Westfalen
anzkon_1 anzkon_3 anzkon_4 Referenzkategorie	Anzahl Kontaktversuche CATI/CAPI: 1 Kontaktversuch Anzahl Kontaktversuche CATI/CAPI: 4-9 Kontaktversuche Anzahl Kontaktversuche CATI/CAPI: 10 und mehr Kontaktversuche Anzahl Kontaktversuche CATI/CAPI: 2-3 Kontaktversuche

Tabelle 43: Logit-Modelle zur Erstteilnahme für Erreichbarkeit und Teilnahme der EWO-Aufstockungsstichprobe Welle 5

	Kontakt		Teilnahme	
	Coef.	p	Coef.	p
alter_1	-.1734983	0.252	-.2048812	0.053
alter_2	-.2268192	0.147	-.4239397	0.000
alter_4	.2374297	0.104	.2044171	0.021
alter_5	.7053023	0.000	-.0647211	0.477
sex_1	-.050245	0.609	.1124317	0.065
nichtdeutsch	-.4406449	0.002	-.5264382	0.000
blneualt_2	-.3393163	0.003		
bundesld_1			-.5270543	0.039
bundesld_2			.1204216	0.629
bundesld_3			.1418745	0.204
bundesld_4			-.4679943	0.169
bundesld_6			-.2911917	0.038
bundesld_7			.2796866	0.048
bundesld_8			-.2024863	0.060
bundesld_9			.2596503	0.009
bundesld_10			.5289703	0.051
bundesld_11			-.2185617	0.217
bundesld_12			-.1186459	0.468
bundesld_13			-.90868	0.004
bundesld_14			-.0081744	0.960
bundesld_15			.4191263	0.016
bundesld_16			.133729	0.351
anzkon_1	-1.337765	0.000		
anzkon_3	-.7611378	0.000		
anzkon_4	-.1804869	0.401		
cons	3.201539	0.000	-1.030231	0.000
n	6237		5769	
Log likelihood	-1569.4437		-3249.2651	
Pseudo R²	0.0557		0.0203	

6.9 Wiederteilnahme-Propensity Personen

Das entscheidende Längsschnittgewicht ist nicht dasjenige auf Haushalts-, sondern das auf Personenebene, da es sich hierbei um die über die Zeit stabilen Einheiten handelt. Die Teilnahmepropensities für Personen in Welle 5 wurden unter Einschluss von zusätzlichen Personenmerkmalen über Logit-Modelle für Panelbereitschaft, Erreichbarkeit und Teilnahme geschätzt. Die über den Haushaltszusammenhang vermittelte Abhängigkeit der Personenstichprobe und die dadurch notwendige Korrektur der Schätzung der Standardfehler wurde bei diesen Modellen berücksichtigt, indem die Fehlerterme auf Haushaltsebene geclustert wurden. Die vorhergesagten Wahrscheinlichkeiten der Modelle wurden wiederum multipliziert. Der Kehrwert dieses Produkts ist in der Variable *ppbleib* zu finden. Das Längsschnittgewicht einer Person für den Zeitraum [t₁; t₂; t₃; t₄; t₅] über alle fünf Wellen lässt sich dann als Produkt aus Querschnittgewicht zu t₁, *ppbleib* (Welle 1 zu Welle 2) und *ppbleib* (Welle 2 zu Welle 3 usw.) berechnen.

Tabelle 44: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der Wiederholer

Variablenkürzel und Referenzkategorie	Erläuterung
alter_1 alter_2 alter_4 alter_5 Referenzkategorie	Person jünger als 30 Jahre Person 30-39 Jahre alt Person 50-64 Jahre alt Person 65 Jahre und älter Person 40-49 Jahre alt
sex_1 Referenzkategorie	Person Männlich Person Weiblich
nichtdeutsch Referenzkategorie	Person hat nicht-deutsche Staatsangehörigkeit Person hat deutsche Staatsangehörigkeit oder fehlende Angabe
schulbil_1 schulbil_2 schulbil_4 Referenzkategorie	Schulabschluss Person: ohne Abschluss Schulabschluss Person: Volks-/Hauptschule Schulabschluss Person: Fach-/Hochschulreife Schulabschluss Person: Realschule/mittlerer Abschluss/noch Schüler
gesundheit_3 gesundheit_4 gesundheit_5 Referenzkategorie	Subjektive Beurteilung des Gesundheitszustands der Person: Zufriedenstellend Subjektive Beurteilung des Gesundheitszustands der Person: Weniger gut Subjektive Beurteilung des Gesundheitszustands der Person: Schlecht Subjektive Beurteilung des Gesundheitszustands der Person: Sehr gut bis Gut
zufrieden_1 zufrieden_2 zufrieden_3 Referenzkategorie	Generelle Lebenszufriedenheit der Person: Skalenwert 0-2 Generelle Lebenszufriedenheit der Person: Skalenwert 3-5 Generelle Lebenszufriedenheit der Person: Skalenwert 6-8 Generelle Lebenszufriedenheit der Person: Skalenwert 9-10
anz_0_3 anz_4_6 anz_7_14 anz_65 Referenzkategorie	Anzahl Personen im Haushalt im Alter 0-3 Jahre Anzahl Personen im Haushalt im Alter 4-6 Jahre Anzahl Personen im Haushalt im Alter 7-14 Jahre Anzahl Personen im Haushalt im Alter 65 Jahre und älter Anzahl Personen im Haushalt im Alter 15-64 Jahre
eigentum Referenzkategorie	Art des Wohneigentums: Eigentümer Art des Wohneigentums: Mieter, fehlende Angabe
wnka_1 wnka_3 Referenzkategorie	Anzahl w.n.- und k.A.-Angaben in Haushalts- und Personeninterview der Person: keine Anzahl w.n.- und k.A.-Angaben in Haushalts- und Personeninterview der Person: 11 und mehr Anzahl w.n.- und k.A.-Angaben in Haushalts- und Personeninterview der Person: 1-10
hhincome_1 hhincome_2 hhincome_4 Referenzkategorie	Haushaltseinkommen: bis 870 Euro Haushaltseinkommen: 871-1.400 Euro Haushaltseinkommen: mehr als 2.200 Euro Haushaltseinkommen: 1.401-2.200 Euro
alg2_1 Referenzkategorie	ALG II-Bezug des Haushalts: bezieht aktuell ALG II ALG II-Bezug des Haushalts: bezieht aktuell kein ALG II
stichprobe1 stichprobe3 stichprobe4 stichprobe5 Referenzkategorie	BA-Stichprobe Auffrischungsstichprobe (BA) Welle 2 Auffrischungsstichprobe (BA) Welle 3 Auffrischungsstichprobe (BA) Welle 4 Microm-Stichprobe
anzkon_1 anzkon_3 anzkon_4 Referenzkategorie	Anzahl Kontaktversuche CATI/CAPI: 1 Kontaktversuch Anzahl Kontaktversuche CATI/CAPI: 4-9 Kontaktversuche Anzahl Kontaktversuche CATI/CAPI: 10 und mehr Kontaktversuche Anzahl Kontaktversuche CATI/CAPI: 2-3 Kontaktversuche

Tabelle 44: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der Wiederholer (Fortsetzung)

Variablenkürzel und Referenzkategorie	Erläuterung
blinealt_2 Referenzkategorie	Neue Bundesländer Alte Bundesländer
bundesld_1 bundesld_2 bundesld_3 bundesld_4 bundesld_6 bundesld_7 bundesld_8 bundesld_9 bundesld_10 bundesld_11 bundesld_12 bundesld_13 bundesld_14 bundesld_15 bundesld_16 Referenzkategorie	Bundesland: Schleswig-Holstein Bundesland: Hamburg Bundesland: Niedersachsen Bundesland: Bremen Bundesland: Hessen Bundesland: Rheinland-Pfalz Bundesland: Baden-Württemberg Bundesland: Bayern Bundesland: Saarland Bundesland: Berlin Bundesland: Brandenburg Bundesland: Mecklenburg-Vorpommern Bundesland: Sachsen Bundesland: Sachsen-Anhalt Bundesland: Thüringen Bundesland: Nordrhein-Westfalen
bik_1 bik_2 bik_3 bik_4 bik_5 bik_6 bik_7 bik_8 bik_9 Referenzkategorie	BIK-Größenklasse der Gemeinde: unter 2.000 Einwohner BIK-Größenklasse der Gemeinde: 2.000 bis u. 5.000 Einwohner BIK-Größenklasse der Gemeinde: 5.000 bis u. 20.000 Einwohner BIK-Größenklasse der Gemeinde: 20.000 bis u. 50.000 Einwohner BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner STYP 2/ 3/ 4 BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner STYP 1 BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner STYP 2/ 3/ 4 BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner STYP 1 BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner STYP 2/ 3/ 4 BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner STYP 1

Tabelle 45: Logit-Modelle zur Wiederteilnahme für Panelbereitschaft, Erreichbarkeit und Teilnahme

	Panelbereitschaft		Kontakt		Teilnahme	
	Coef.	p	Coef.	p	Coef.	p
alter_1	-.7675201	0.101	-.4065086	0.003	-.6125301	0.000
alter_2	-1.034716	0.051	-.0873256	0.602	-.2002677	0.017
alter_4	-.4785485	0.409	.160481	0.325	.2760516	0.001
alter_5	1.041629	0.240	-.2175728	0.642	.1415955	0.446
sex_1	-.1488998	0.623	-.1431344	0.091	-.0685735	0.126
nichtdeutsch	-.7950164	0.131	-.1364546	0.495	-.2731369	0.006
schulbil_1	-.450964	0.471	-.1547037	0.503	-.2547237	0.038
schulbil_2	-.1375017	0.737	-.0750139	0.525	-.1896411	0.004
schulbil_4	-.5940769	0.236	.1736765	0.218	.0645598	0.370
gesundheit_3	-.5303662	0.191	-.2860039	0.010	.1289823	0.033
gesundheit_4	-.237548	0.625	-.1048915	0.495	.2510621	0.002
gesundheit_5	-.0466248	0.940	-.166069	0.452	-.02595	0.824
zufrieden_1	-.9588241	0.405	.0755368	0.821	-.2488858	0.112
zufrieden_2	-.716456	0.391	-.2780909	0.133	-.0905377	0.322
zufrieden_3	-.6356465	0.417	-.2344541	0.129	-.0029418	0.967

Tabelle 45: Logit-Modelle zur Wiederteilnahme für Panelbereitschaft, Erreichbarkeit und Teilnahme (Fortsetzung)

	Panelbereitschaft		Kontakt		Teilnahme	
	Coef.	p	Coef.	p	Coef.	p
anz_0_3	-.125206	0.793	-.1034997	0.527	-.0185916	0.829
anz_4_6	-.3209503	0.569	.0826405	0.678	.0741661	0.417
anz_7_14	-.3420701	0.410	-.0797108	0.503	.0125192	0.823
anz_65	-1.270068	0.007	.4417246	0.171	.140469	0.241
eigentum	-.6838766	0.242	.7462691	0.001	.0186042	0.831
wnka_1	-.5217555	0.255	.0493292	0.660	.2954034	0.000
wnka_3	-1.547987	0.000	.0660804	0.677	-.1146345	0.163
hhincome_1	-.958492	0.181	-.2721554	0.127	.1612614	0.106
hhincome_2	-.5737491	0.344	-.0398392	0.815	.0505144	0.562
hhincome_4	.9467236	0.246	.34564	0.124	.2493001	0.008
alg2_1	-.6571527	0.299	.2022105	0.161	.0431921	0.584
stichprobe1			-.6368302	0.001	-.1063695	0.252
stichprobe3			-.8485139	0.000	-.1634342	0.231
stichprobe4			-.7266907	0.002	-.1135386	0.373
stichprobe5			-.5145691	0.036	-.3092601	0.011
blneualt_2	1.326077	0.010				
anzkon_1			-.5881563	0.014		
anzkon_3			-.3647335	0.048		
anzkon_4			-1.366091	0.000		
bundesld_1			-.8153592	0.002	-.6512667	0.000
bundesld_2			-.2045841	0.626	.0130608	0.956
bundesld_3			.0135992	0.961	-.3224352	0.005
bundesld_4			-.0781003	0.891	-.10384	0.738
bundesld_6			-.5307905	0.041	-.2260133	0.094
bundesld_7			.1321791	0.731	-.5868427	0.000
bundesld_8			-.4056052	0.089	-.4287236	0.000
bundesld_9			-.1506979	0.523	-.2695108	0.015
bundesld_10			.2442566	0.653	-.3633051	0.181
bundesld_11			-.8108077	0.003	-.131471	0.384
bundesld_12			.0565758	0.881	-.1744911	0.291
bundesld_13			-.0330389	0.939	.2577315	0.250
bundesld_14			-.4681859	0.112	.4859923	0.006
bundesld_15			-.2825624	0.351	-.3608326	0.018
bundesld_16			.3102048	0.446	-.0823932	0.656
bik_1			.5100487	0.630	.5687608	0.219
bik_2			.5193127	0.297	.221327	0.252
bik_3			-.0797132	0.786	.1251445	0.347
bik_4			-.5064409	0.034	.3166536	0.013
bik_5			-.3559797	0.180	.3606542	0.006
bik_6			-.2955571	0.412	.174933	0.358
bik_7			-.2154511	0.362	.3458393	0.003
bik_8			-.2644422	0.187	.1943944	0.053
bik_9			-.0031591	0.993	.2181262	0.107
cons	8.899969	0.000	4.885233	0.000	1.442663	0.000
n	11768		11730		11252	
Log likelihood	-224.70204		-1772.3739		-5215.4557	
Pseudo R²	0.1219		0.1129		0.0465	

Anmerkung: Die Korrektur der Standardfehler erfolgt mittels einer über Haushalte geclusterten Schätzung.

6.10 Integration der Gewichte zum Gesamtgewicht vor Kalibrierung

In diesem Schritt ging es erneut darum, die durch die Ausfallmodellierung modifizierten Haushaltsgewichte der neuen Auffrischungs- und der Bestandsstichprobe (einschließlich der Auffrischer aus Welle 2 bis 4) zu kombinieren. Die Integration der Bestandsaufstockungen in die jeweiligen Bestände erfolgte durch Konvexkombination der jeweiligen Stichprobe. Die mehrfache Auswahlwahrscheinlichkeit, die ein nun gezogener Leistungsempfänger, der in den vorausgegangenen Jahren in einem gemeinsamen Haushalt mit Leistungsempfängern gelebt hat, ohne Mitglied der Bedarfsgemeinschaft gewesen zu sein, besitzt, wurde wiederum ignoriert. Die neuen Designgewichte der Leistungsempfängerstichprobe rechnen im Querschnitt hoch auf alle Personen, die in 7/2006, in 7/2007, in 7/2008, in 7/2009 oder in 7/2010 in einem Haushalt, der mindestens eine BG enthält, gewohnt haben. Erst bei der Berechnung neuer Gewichte für die Gesamtstichprobe besteht Bedarf an einer Anpassung der Gewichte für alle Leistungsempfängerhaushalte mit Bezug in 7/2010. Dazu wurde die Inklusionswahrscheinlichkeit in der jeweils anderen Stichprobe für Fälle aus der Microm-Stichprobe (Welle 1) und der neuen Auffrischungstichprobe (Welle 5) geschätzt. Für Fälle aus der Auffrischungstichprobe wurde die mittlere Welle 1-Auswahlwahrscheinlichkeit in der Microm-Stichprobe in der jeweiligen Postleitzahl und jeweils die durchschnittliche Teilnahmewahrscheinlichkeit (für Welle 1, Welle 2, Welle 3, Welle 4 und Welle 5) in dieser Stichprobe unterstellt. Für Fälle aus der Microm-Stichprobe wurde, falls es sich (laut Befragungsdaten) um Neuzugänge zum Arbeitslosengeld-II-Leistungsbezug zwischen den vier letzten Stichprobenziehungszeitpunkten (Welle 2, Welle 3, Welle 4, Welle 5) handelt, die durchschnittliche Auswahlwahrscheinlichkeit eines Auffrischungshaushalts in der jeweiligen PLZ und die durchschnittliche Teilnahmewahrscheinlichkeit in dieser Stichprobe unterstellt. Danach wurden die beiden Gewichte zu einem neuen Gesamtgewicht integriert.

6.11 Integration temporärer Ausfälle (Haushalte)

In Welle 5 konnten, wie auch schon in Welle 4, Haushalte wieder teilnehmen, die eine Welle ausgesetzt, d.h. nicht teilgenommen hatten (temporäre Ausfälle). Längsschnittgewichte für diese Haushalte werden nicht berechnet, d.h. (gewichtete) Längsschnittauswertungen können nur mit Teilnehmern über alle in Frage kommenden Wellen durchgeführt werden. Die Nichtteilnahme eines Haushalts kann nur über eine Welle stattfinden, fällt ein Haushalt über zwei aufeinander folgende Wellen aus, wird er nicht mehr kontaktiert. Für die Berechnung von gemeinsamen Querschnittsgewichten einschließlich der temporären Ausfälle erfolgte eine Konvexkombination der modifizierten Haushaltsgewichte der temporären Ausfälle und der modifizierten Haushaltsgewichte der Bestandsstichprobe (nicht der Auffrischerstichprobe) vor Kalibrierung. Die Konvexkombination der Haushaltsgewichte erfolgte also vor der Kalibrierung, diese wurde dann mit den neuen kombinierten Haushaltsgewichten durchgeführt.

Während mit den durch Ausfallmodellierung modifizierten Haushaltsgewichten bereits Hochrechnungsfaktoren für die Bestands- und Auffrischerstichprobe existierten, war es für die temporären Ausfälle notwendig, solche modifizierten Haushaltsgewichte als Schätzer

für die jeweilige Population neu zu berechnen. Ausgangspunkt waren dabei die kalibrierten Haushaltsgewichte der Welle 3 (die Welle 4 ist der temporäre Ausfall).

Für die temporären Ausfälle wurde dann die Wahrscheinlichkeit der Nichtteilnahme in Welle 4 gegeben die Teilnahme in Welle 3 (Nichtteilnahme-Propensities Welle 4) und die Wahrscheinlichkeit der Teilnahme in Welle 5 gegeben die Nichtteilnahme in Welle 4 (Teilnahme-Propensities Welle 5) bestimmt. Die Wahrscheinlichkeit der Nichtteilnahme in Welle 4 ergibt sich über 1-Teilnahmewahrscheinlichkeit in Welle 4.

Die beschriebenen Teilnahme- bzw. Nichtteilnahme-Propensities wurden wiederum über Logit-Modelle geschätzt. Die vorhergesagten Wahrscheinlichkeiten der jeweiligen Modelle wurden multipliziert. Das modifizierte Haushaltsgewicht der temporären Ausfälle ergibt sich dann über die Multiplikation der kalibrierten Haushaltsgewichte der Welle 2 mit dem Kehrwert dieses Produkts.

Tabelle 46: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der temporären Ausfälle

Variablenkürzel und Referenzkategorie	Erläuterung
panelber	Panelbereitschaft (Widerspruchsverfahren von W3 nach W4 berücksichtigt)
alter_1	Haushaltsauskunftsperson (HBV) jünger als 30 Jahre
alter_2	HBV 30-39 Jahre alt
alter_4	HBV 50-64 Jahre alt
alter_5	HBV 65 Jahre und älter
Referenzkategorie	HBV 40-49 Jahre alt
sex_1	HBV Männlich
Referenzkategorie	HBV Weiblich
nichtdeutsch	HBV hat nicht-deutsche Staatsangehörigkeit
Referenzkategorie	HBV hat deutsche Staatsangehörigkeit oder fehlende Angabe
sprache_nichtdeutsch	Sprache HBV: überwiegend im Haushalt gesprochene Sprache nicht-deutsch
Referenzkategorie	Sprache HBV: überwiegend im Haushalt gesprochene Sprache deutsch
schulbil_1	Schulabschluss HBV: ohne Abschluss
schulbil_2	Schulabschluss HBV: Volks-/Hauptschule
schulbil_4	Schulabschluss HBV: Fach-/Hochschulreife
Referenzkategorie	Schulabschluss HBV: Realschule/mittlerer Abschluss/noch Schüler
gesundheit_3	Subjektive Beurteilung des Gesundheitszustands HBV: Zufriedenstellend
gesundheit_4	Subjektive Beurteilung des Gesundheitszustands HBV: Weniger gut
gesundheit_5	Subjektive Beurteilung des Gesundheitszustands HBV: Schlecht
Referenzkategorie	Subjektive Beurteilung des Gesundheitszustands HBV: Sehr gut bis Gut
zufrieden_1	Generelle Lebenszufriedenheit HBV: Skalenwert 0-2
zufrieden_2	Generelle Lebenszufriedenheit HBV: Skalenwert 3-5
zufrieden_3	Generelle Lebenszufriedenheit HBV: Skalenwert 6-8
Referenzkategorie	Generelle Lebenszufriedenheit HBV: Skalenwert 9-10
anz_0_3	Anzahl Personen im Haushalt im Alter 0-3 Jahre
anz_4_6	Anzahl Personen im Haushalt im Alter 4-6 Jahre
anz_7_14	Anzahl Personen im Haushalt im Alter 7-14 Jahre
anz_65	Anzahl Personen im Haushalt im Alter 65 Jahre und älter
DinvalidAge	Nicht auswertbare Altersangaben
Referenzkategorie	Anzahl Personen im Haushalt im Alter 15-64 Jahre
eigentum	Art des Wohneigentums: Eigentümer
Referenzkategorie	Art des Wohneigentums: Mieter, fehlende Angabe

Tabelle 46: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der temporären Ausfälle (Fortsetzung)

Variablenkürzel und Referenzkategorie	Erläuterung
wnka_1 wnka_3 Referenzkategorie	Anzahl w.n.- und k.A.-Angaben in Haushalts- und Personeninterview des HBV: keine Anzahl w.n.- und k.A.-Angaben in Haushalts- und Personeninterview des HBV: 11 und mehr Anzahl w.n.- und k.A.-Angaben in Haushalts- und Personeninterview des HBV: 1-10
hhincome_1 hhincome_2 hhincome_4 Referenzkategorie	Haushaltseinkommen: bis 870 Euro Haushaltseinkommen: 871-1.400 Euro Haushaltseinkommen: mehr als 2.200 Euro Haushaltseinkommen: 1.401-2.200 Euro
alg2_1 Referenzkategorie	ALG II-Bezug des Haushalts: bezieht aktuell ALG II ALG II-Bezug des Haushalts: bezieht aktuell kein ALG II
bundesld_1 bundesld_2 bundesld_3 bundesld_4 bundesld_6 bundesld_7 bundesld_8 bundesld_9 bundesld_10 bundesld_11 bundesld_12 bundesld_13 bundesld_14 bundesld_15 bundesld_16 Referenzkategorie	Bundesland: Schleswig-Holstein Bundesland: Hamburg Bundesland: Niedersachsen Bundesland: Bremen Bundesland: Hessen Bundesland: Rheinland-Pfalz Bundesland: Baden-Württemberg Bundesland: Bayern Bundesland: Saarland Bundesland: Berlin Bundesland: Brandenburg Bundesland: Mecklenburg-Vorpommern Bundesland: Sachsen Bundesland: Sachsen-Anhalt Bundesland: Thüringen Bundesland: Nordrhein-Westfalen
bik_1 bik_2 bik_3 bik_4 bik_5 bik_6 bik_7 bik_8 bik_9 Referenzkategorie	BIK-Größenklasse der Gemeinde: unter 2.000 Einwohner BIK-Größenklasse der Gemeinde: 2.000 bis u. 5.000 Einwohner BIK-Größenklasse der Gemeinde: 5.000 bis u. 20.000 Einwohner BIK-Größenklasse der Gemeinde: 20.000 bis u. 50.000 Einwohner BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner STYP 2/ 3/ 4 BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner STYP 1 BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner STYP 2/ 3/ 4 BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner STYP 1 BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner STYP 2/ 3/ 4 BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner STYP 1

Tabelle 47: Logit-Modelle temporärer Ausfälle

	Wiederteilnahme in Welle 4 zur Ermittlung der Nicht-Teilnahmewahrscheinlichkeit W4 (1-Teilnahmewahrscheinlichkeit W4)		Wiederteilnahme in Welle 5 gegeben Nichtteilnahme in Welle 4	
	Coef.	p	Coef.	p
panelber	6.444816	0.000		
alter_1	-.604372	0.000	.2578263	0.186
alter_2	-.3010882	0.000	.5127758	0.006
alter_4	.3945154	0.000	-.1195183	0.554
alter_5	.3410748	0.113	.1252457	0.818
sex_1	-.2162814	0.000	.1542812	0.246
nichtdeutsch	-.1043709	0.358	-.4117408	0.150
sprache_nichtdeutsch	-.2617484	0.008	-.1490322	0.507
schulbil_1	-.4240614	0.000	.1739415	0.540
schulbil_2	-.108591	0.110	.176298	0.268
schulbil_4	.153694	0.048	.2044274	0.257
gesundheit_3	.0780563	0.231	-.1306583	0.384
gesundheit_4	.1476658	0.081	-.1636986	0.415
gesundheit_5	-.0967201	0.393	-.2391248	0.399
zufrieden_1	-.2558134	0.090	-.1264077	0.731
zufrieden_2	-.0972692	0.348	.1449795	0.550
zufrieden_3	.0310376	0.737	.0347136	0.874
anz_0_3	-.129093	0.078	-.042039	0.800
anz_4_6	.080956	0.351	.1281763	0.472
anz_7_14	.0284811	0.569	-.1055763	0.365
anz_65	.1458783	0.252	-.3949094	0.231
DinvalidAge	-.3512535	0.124	-1.022052	0.303
eigentum	.2236381	0.008	-.4783	0.021
wnka_1	.123889	0.052	.0003759	0.998
wnka_3	-.206712	0.006	-.168878	0.327
hhincome_1	-.1278953	0.146	-.0152265	0.940
hhincome_2	.0574915	0.469	.0921635	0.617
hhincome_4	.1562073	0.086	.0148841	0.946
alg2_1	-.0250443	0.715	.381393	0.017
bundesld_1	-.0864298	0.551	.0067697	0.983
bundesld_2	.6313826	0.029	.8994607	0.097
bundesld_3	-.068764	0.518	-.3142115	0.194
bundesld_4	-.1460367	0.589	-1.361038	0.189
bundesld_6	-.2853048	0.016	-.0240956	0.925
bundesld_7	-.303385	0.043	-.5268514	0.168
bundesld_8	.0348457	0.765	-.3758776	0.195
bundesld_9	-.1258062	0.208	-.3918875	0.099
bundesld_10	-.2249574	0.335	-.0158734	0.975
bundesld_11	.0444861	0.741	-.7900351	0.022
bundesld_12	-.0316614	0.833	-.3226686	0.378
bundesld_13	-.0207816	0.908	-.8455606	0.086
bundesld_14	.0632895	0.644	-.0539551	0.855
bundesld_15	-.0155162	0.909	-.1350229	0.645
bundesld_16	-.0044367	0.979	-.1735856	0.644

Tabelle 47: Logit-Modelle temporärer Ausfälle (Fortsetzung)

	Wiederteilnahme in Welle 4 zur Ermittlung der Nicht-Teilnahmewahrscheinlichkeit W4 (1-Teilnahmewahrscheinlichkeit W4)		Wiederteilnahme in Welle 5 gegeben Nichtteilnahme in Welle 4	
	Coef.	p	Coef.	p
bik_1	.1435061	0.610	-.3076576	0.632
bik_2	.0894075	0.616	-.3209982	0.490
bik_3	.0932077	0.451	.2167444	0.422
bik_4	.0531367	0.649	-.2123762	0.453
bik_5	.0524389	0.672	.2087075	0.459
bik_6	.2081499	0.246	-.3048071	0.495
bik_7	-.1223254	0.232	.337995	0.127
bik_8	-.0280089	0.754	-.1502369	0.467
bik_9	.1887004	0.154	-.3209531	0.349
cons	-4.947284	0.000	-2.283381	0.000
n	9535		2892	
Log likelihood	-4168.0111		-890.03657	
Pseudo R²	0.2875		0.0461	

Die Konvexkombination der Gewichte der Teilnehmer über alle Wellen (Bestandsstichprobe) und der temporären Ausfälle erfolgte für die Gewichte aller drei Teilstichproben i (Microm, BA und Gesamt) über die Multiplikation der jeweiligen modifizierten Haushaltsgewichte mit dem Anteil der Bestandsstichprobe bzw. der temporären Ausfälle am Gesamtstichprobenumfang, d.h. der Summe aus Bestandsstichprobe und temporären Ausfällen:

$dw_{i, hh_{temp.Ausfall}} * (n_{temp.Ausfall i} / (n_{temp.Ausfall i} + n_{Bestand i}))$ für die temporären Ausfälle und

$dw_{i, hh_{Bestand}} * (n_{Bestand i} / (n_{temp.Ausfall i} + n_{Bestand i}))$ für die Bestandsstichprobe.

6.12 Kalibrierung zum Haushaltsgewicht Welle 5 Querschnitt

Anschließend fand wieder eine Kalibrierung der modifizierten Designgewichte inklusive der Ausfallgewichtung auf Haushaltsebene mittels GREG-Verfahren an die Eckwerte des Statistischen Bundesamts für 2010 und für Haushalte mit Leistungsbezug an die Eckwerte der Statistik der Bundesagentur für Arbeit für Juli 2010 statt. Ebenso wie im Vorjahr wurde dabei als Eckwert auch die Zahl der Neuzugänge zum Arbeitslosengeld-II-Leistungsbezug seit dem Vorjahr auf Ebene von Bedarfsgemeinschaften (367.870) verwendet. Dabei werden diejenigen Fälle der bisherigen Stichproben aus den Wellen 1 bis 5 auf die Eckwerte zum Arbeitslosengeld-II-Bezug der Bundesagentur für Arbeit hochgerechnet, die laut Welle 5 des Surveys im Juli 2010 Arbeitslosengeld II bezogen haben.

Das Hauptziel der Gewichtung ist der Ausgleich von Verzerrungen, die durch das Stichprobendesign (mit ungleichen Auswahlwahrscheinlichkeiten) und durch selektive Teilnahme bzw. Nichtteilnahme entstehen. Unter Verwendung der Gewichte lassen sich Populationswerte aus der Stichprobe erwartungstreu schätzen. Weisen die Gewichte al-

lerdings eine starke Streuung auf, so kann dies zu einer großen Varianz der Schätzfunktionen führen. Hier tritt der in der Statistik immer wieder anzutreffende Trade-off zwischen Bias und Varianz auf. Die Gewichtung reduziert den Bias; eine allzu große Steigerung der Varianz durch die Gewichtung sollte allerdings auch vermieden werden. Aus diesem Grunde wird versucht, besonders große Gewichtungsfaktoren (und nachrangig auch besonders kleine Faktoren) möglichst zu umgehen und ggf. geeignete Korrekturen an den Gewichten vorzunehmen. Im Rahmen der vorliegenden Kalibrierung erfolgte dies an zwei Punkten:

- Die Eingangsgewichte für die Kalibrierung (die modifizierten Designgewichte nach Berücksichtigung der Ausfallanalysen) wurden vor der Kalibrierung getrimmt, d.h. sie wurden durch neue Eingangsgewichte ersetzt. Unter- bzw. Obergrenze der getrimmten Designgewichte wurden durch bestimmte Perzentile der Verteilung, in Abhängigkeit von der Verteilung der Designgewichte, ermittelt.
- Bei der Kalibrierung wurde zudem das Intervall der Gewichte begrenzt, d.h. es wurde eine obere und eine untere Grenze für die Gewichte festgelegt. Festgelegt wurde dabei die Gesamtbreite der Gewichte; die Spannbreite der reinen Kalibrierungsgewichte ergibt sich aus dem Verhältnis von ursprünglichen Gewichten zum getrimmten Eingangsgewicht. Zu beachten war dabei, dass engere Grenzen für die Gewichte in geringerer Streuung und damit geringerer Varianz der Schätzungen resultieren, bei zu eng gesetzten Grenzen kann allerdings die Kalibrierung aller Eckwerte unmöglich werden.

Zur Beurteilung der Gewichte wird im Folgenden neben dem Mittelwert und der Standardabweichung auch das Effektivitätsmaß angegeben (E). Das Effektivitätsmaß E basiert auf der Varianz des Gewichtungsfaktors. Das Effektivitätsmaß gibt in Prozent der realisierten Fallzahl an, wie groß bei Verwendung des Gewichts die effektive Fallzahl bei einem passiven Merkmal ist, das mit den aktiven Merkmalen nicht korreliert. Die effektive Fallzahl entspricht der Anzahl Befragter, die bei einer uneingeschränkten Zufallsauswahl, gegeben die Varianz des Merkmals in der Stichprobe, den gleichen Stichprobenfehler produziert hätte. Das Effektivitätsmaß drückt das Verhältnis von n zu n' als Prozentwert aus.

6.12.1 Kalibrierung der BA-Stichprobe

Die Grundgesamtheit der kumulierten BA-Stichprobe aus allen fünf Wellen bilden alle Haushalte in Deutschland mit mindestens einer Bedarfsgemeinschaft mit Leistungen nach dem SGB II zu einem der bisher fünf Ziehungszeitpunkte (im Juli 2006, Juli 2007, Juli 2008, Juli 2009 oder Juli 2010). Eine Kalibrierung findet in Welle 5 dagegen nur an die Eckwerte der BA-Statistik vom Juli 2010 statt. Damit beeinflusst die Kalibrierung nur die Gewichte der Haushalte aus der BA-Stichprobe, in denen auch im Juli 2010 noch mindestens eine Bedarfsgemeinschaft mit Leistungen nach dem SGB II lebt. Ausgangspunkt für die Kalibrierung waren die modifizierten Designgewichte inklusive der Ausfallgewichtung. Die modifizierten Designgewichte wurden am 5%-Perzentil und am 95%-Perzentil ihrer Verteilung getrimmt und anschließend so reskaliert, dass ihre Summe wieder die Summe der ungetrimmten Designgewichte ergab. Die Hochrechnungsfaktoren der getrimmten Designgewichte reichen von 242,61 bis 1976,16. Das Verhältnis der Gesamthochrech-

nungsfaktoren nach Kalibrierung zu getrimmten Designgewichten wurde nach unten auf minimal 0,3 und nach oben auf maximal 2,0 begrenzt. Danach bewegen sich die Gesamthochrechnungsfaktoren nach Kalibrierung im Bereich von minimal 72,78 bis maximal 3014,12.

Für folgende Merkmale wurde kalibriert:

Bedarfsgemeinschaften Grundlage BA-Statistik:

- Zahl der BG Neuzugänge zum Arbeitslosengeld-II-Leistungsbezug
- Anzahl BG im Leistungsbezug nach SGB II nach Bundesländern
- Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Personen unter 65 Jahre in der Bedarfsgemeinschaft, nach West/Ost
- Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Kinder unter 15 Jahre in der Bedarfsgemeinschaft, nach West/Ost
- Anzahl BG im Leistungsbezug nach SGB II, die aus einer/einem Alleinerziehenden mit Kind(ern) besteht, nach West/Ost

Daneben wurde wie im Vorjahr ein zusätzlicher Eckwert mit aufgenommen, und zwar die Zahl der Neuzugänge zum Arbeitslosengeld-II-Leistungsbezug seit dem Vorjahr auf Ebene von Bedarfsgemeinschaften (367.870).

Um die Kalibrierung durchzuführen, muss jede Eckwertvariable für jeden Haushalt einen gültigen Wert besitzen. Deshalb wurde der sehr geringe Item-Nonresponse vor der Kalibrierung imputiert. Die Imputation erfolgte über den Mittelwert bzw. den Modalwert der jeweiligen Variablen.

Da die Imputation einzig der Durchführbarkeit der Kalibrierung dient, wurden die imputierten Werte nach der Kalibrierung wieder auf fehlende Werte zurückgesetzt. Eine Hochrechnung mit den kalibrierten Gewichten ohne Berücksichtigung des Item-Nonresponse führt daher zu geringfügigen Abweichungen von den im Folgenden dargestellten Werten.

Tabelle 48: Soll-Verteilungen und Verteilungen nach Kalibrierung (BA-Stichprobe, Haushalte)

Eckwert	Ausprägung Eckwert aus der BA-Statistik	ungewichtete Verteilung	Soll-Werte aus der BA-Statistik	Verteilung mit kalibrierten Gewichten
Anzahl BG im Leistungsbezug nach SGB II nach Bundesländern (16 Kategorien)	Zahl BGs Schleswig-Holstein	161	122.816	122.816
	Zahl BGs Hamburg	138	108.473	108.473
	Zahl BGs Niedersachsen	436	334.073	334.073
	Zahl BGs Bremen	51	51.949	51.949
	Zahl BGs Nordrhein-Westfalen	1.080	849.532	849.532
	Zahl BGs Hessen	217	219.564	219.564
	Zahl BGs Rheinland-Pfalz	151	123.522	123.522
	Zahl BGs Baden-Württemberg	314	256.618	256.618
	Zahl BGs Bayern	387	267.375	267.375
	Zahl BGs Saarland	71	44.787	44.787
	Zahl BGs Berlin	303	334.106	334.106
	Zahl BGs Brandenburg	272	167.160	167.160
	Zahl BGs Mecklenburg-Vorpommern	129	127.174	127.174
	Zahl BGs Sachsen	339	280.039	280.039
	Zahl BGs Sachsen-Anhalt	261	186.400	186.400
	Zahl BGs Thüringen	174	130.035	130.035
Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Personen unter 65 Jahre in der Bedarfsgemeinschaft (1, 2, 3, 4 und „5 oder mehr“) und nach West/Ost (10 Kategorien)	Zahl BGs mit 1 Person unter 65 (West)	1.225	1.287.479	1.287.479
	Zahl BGs mit 2 Personen unter 65 (West)	792	480.035	480.035
	Zahl BGs mit 3 Personen unter 65 (West)	506	297.590	297.590
	Zahl BGs mit 4 Personen unter 65 (West)	285	185.355	185.355
	Zahl BGs mit 5 oder mehr Personen unter 65 (West)	198	128.250	128.250
	Zahl BGs mit 1 Person unter 65 (Ost)	682	714.572	714.572
	Zahl BGs mit 2 Personen unter 65 (Ost)	422	264.338	264.338
	Zahl BGs mit 3 Personen unter 65 (Ost)	221	135.917	135.917
	Zahl BGs mit 4 Personen unter 65 (Ost)	90	71.425	71.425
	Zahl BGs mit 5 oder mehr Personen unter 65 (Ost)	63	38.662	38.662

Tabelle 48: Soll-Verteilungen und Verteilungen nach Kalibrierung (BA-Stichprobe, Haushalte) (Fortsetzung)

Eckwert	Ausprägung Eckwertaus der BA-Statistik	ungewichtete Verteilung	Soll-Werte aus der BA-Statistik	Verteilung mit kalibrierten Gewichten
Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Kinder unter 15 Jahre in der Bedarfsgemeinschaft (0, 1, 2, 3, „4 oder mehr“) und nach West/Ost (10 Kategorien)	Zahl BGs ohne Kinder unter 15 Jahre (West)	1.997	1.629.124	1.629.124
	Zahl BGs mit 1 Kind unter 15 Jahre (West)	550	406.480	406.480
	Zahl BGs mit 2 Kindern unter 15 Jahre (West)	325	230.977	230.977
	Zahl BGs mit 3 Kindern unter 15 Jahre (West)	93	80.359	80.359
	Zahl BGs mit 4 oder mehr Kindern unter 15 Jahre (West)	41	31.769	31.769
	Zahl BGs ohne Kinder unter 15 Jahre (Ost)	1.095	909.771	909.771
	Zahl BGs mit 1 Kind unter 15 Jahre (Ost)	247	188.910	188.910
	Zahl BGs mit 2 Kindern unter 15 Jahre (Ost)	100	90.185	90.185
	Zahl BGs mit 3 Kindern unter 15 Jahre (Ost)	23	25.972	25.972
	Zahl BGs mit 4 oder mehr Kindern unter 15 Jahre (Ost)	13	10.076	10.076
Anzahl BG im Leistungsbezug nach SGB II, die aus einer/einem Alleinerziehenden mit Kindern besteht, nach West/Ost (4 Kategorien)	Zahl BGs mit einem/einer Alleinerziehenden (West)	715	450.759	450.759
	Rest BGs ohne Alleinerziehenden (West)	2.291	1.927.950	1.927.950
	Zahl BGs mit einem/einer Alleinerziehenden (Ost)	258	189.640	189.640
	Rest BGs ohne Alleinerziehenden (Ost)	1.220	1.035.274	1.035.274

Tabelle 49: Kenngrößen der Verteilung der Gewichte

1%-Perzentil	105,1712
5%-Perzentil	173,7548
10%-Perzentil	239,6466
25%-Perzentil	372,7158
50%-Perzentil	682,7329
75%-Perzentil	1085,275
90%-Perzentil	1588,228
95%-Perzentil	2006,884
99%-Perzentil	2652,644
Mittelwert	813,7423
Standardabweichung	567,7689
Minimum	72,78169
Maximum	3014,117
Fallzahl	4482
Effektivitätsmaß	66,8%

6.12.2 Bevölkerungsstichprobe

Die Grundgesamtheit bilden alle Privathaushalte in Deutschland. Ausgangspunkt für die Kalibrierung waren die modifizierten Designgewichte inklusive der Ausfallgewichtung. Die modifizierten Designgewichte wurden am 5%-Perzentil und am 95%-Perzentil ihrer Verteilung getrimmt und anschließend so reskaliert, dass ihre Summe wieder die Summe der ungetrimmten Designgewichte ergab. Die Hochrechnungsfaktoren der getrimmten Designgewichte reichen von 2775,61 bis 27305,53. Das Verhältnis der Gesamthochrechnungsfaktoren nach Kalibrierung zu getrimmten Designgewichten wurde nach unten auf minimal 0,1 und nach oben auf maximal 4,0 begrenzt. Danach bewegen sich die Gesamthochrechnungsfaktoren nach Kalibrierung im Bereich von minimal 277,56 bis maximal 79154,33.

Für folgende Merkmale wurde kalibriert:

Bedarfsgemeinschaften: Grundlage BA-Statistik:

- Anzahl BG im Leistungsbezug nach SGB II nach Bundesländern
- Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Personen unter 65 Jahre in der Bedarfsgemeinschaft, nach West/Ost
- Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Kinder unter 15 Jahre in der Bedarfsgemeinschaft, nach West/Ost
- Anzahl BG im Leistungsbezug nach SGB II, die aus einer/einem Alleinerziehenden mit Kind(ern) besteht, nach West/Ost

Haushalte: Grundlage Mikrozensus 2010:

- Anzahl der Haushalte nach Bundesland und BIK-Typ
- Anzahl der Haushalte nach Haushaltsgröße und West/Ost
- Anzahl der Haushalte nach „Kinder unter 15 Jahre im Haushalt ja/nein“ und West/Ost

Um die Kalibrierung durchzuführen, muss jede Eckwertvariable für jeden Haushalt einen gültigen Wert besitzen. Deshalb wurde der sehr geringe Item-Nonresponse vor der Kalibrierung imputiert. Die Imputation erfolgte über den Mittelwert bzw. den Modalwert der jeweiligen Variablen.

Da die Imputation einzig der Durchführbarkeit der Kalibrierung dient, wurden die imputierten Werte nach der Kalibrierung wieder auf fehlende Werte zurückgesetzt. Eine Hochrechnung mit den kalibrierten Gewichten ohne Berücksichtigung des Item-Nonresponse führt daher zu geringfügigen Abweichungen von den im Folgenden dargestellten Werten.

Tabelle 50: Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungstichprobe, Haushalte)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl BG im Leistungsbezug nach SGB II nach Bundesländern (16 Kategorien)	Zahl BGs Schleswig-Holstein	9	122.816	123.812
	Zahl BGs Hamburg	6	108.473	108.473
	Zahl BGs Niedersachsen	32	334.073	334.073
	Zahl BGs Bremen	6	51.949	51.949
	Zahl BGs Nordrhein-Westfalen	93	849.532	848.046
	Zahl BGs Hessen	7	219.564	219.564
	Zahl BGs Rheinland-Pfalz	11	123.522	124.333
	Zahl BGs Baden-Württemberg	10	256.618	256.618
	Zahl BGs Bayern	26	267.375	267.375
	Zahl BGs Saarland	7	44.787	44.787
	Zahl BGs Berlin	17	334.106	333.949
	Zahl BGs Brandenburg	24	167.160	167.160
	Zahl BGs Mecklenburg-Vorpommern	4	127.174	127.174
	Zahl BGs Sachsen	18	280.039	280.039
	Zahl BGs Sachsen-Anhalt	12	186.400	186.400
	Zahl BGs Thüringen	15	130.035	129.872

Tabelle 50: Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungsstichprobe, Haushalte) (Fortsetzung 1)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Personen unter 65 Jahre in der Bedarfsgemeinschaft (1, 2, 3, 4 und „5 oder mehr“) und nach West/Ost (10 Kategorien)	Zahl BGs mit 1 Person unter 65 (West)	72	1.287.479	1.287.479
	Zahl BGs mit 2 Personen unter 65 (West)	50	480.035	479.859
	Zahl BGs mit 3 Personen unter 65 (West)	35	297.590	297.090
	Zahl BGs mit 4 Personen unter 65 (West)	29	185.355	185.355
	Zahl BGs mit 5 oder mehr Personen unter 65 (West)	21	128.250	129.246
		28	714.572	714.572
	Zahl BGs mit 1 Person unter 65 (Ost)	27	264.338	264.338
	Zahl BGs mit 2 Personen unter 65 (Ost)	22	135.917	135.597
	Zahl BGs mit 3 Personen unter 65 (Ost)	7	71.425	71.425
	Zahl BGs mit 4 Personen unter 65 (Ost)	6	38.662	38.662
	Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Kinder unter 15 Jahre in der Bedarfsgemeinschaft (0, 1, 2, 3, „4 oder mehr“) und nach West/Ost (10 Kategorien)	Zahl BGs ohne Kinder unter 15 Jahre (West)	139	1.629.124
Zahl BGs mit 1 Kind unter 15 Jahre (West)		27	406.480	406.480
Zahl BGs mit 2 Kindern unter 15 Jahre (West)		26	230.977	231.973
Zahl BGs mit 3 Kindern unter 15 Jahre (West)		8	80.359	80.359
Zahl BGs mit 4 oder mehr Kindern unter 15 Jahre (West)		7	31.769	31.769
Zahl BGs ohne Kinder unter 15 Jahre (Ost)		65	909.771	909.614
Zahl BGs mit 1 Kind unter 15 Jahre (Ost)		11	188.910	188.747
Zahl BGs mit 2 Kindern unter 15 Jahre (Ost)		11	90.185	90.185
Zahl BGs mit 3 Kindern unter 15 Jahre (Ost)		2	25.972	25.972
Zahl BGs mit 4 oder mehr Kindern unter 15 Jahre (Ost)		1	10.076	10.076

Tabelle 50: Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungsstichprobe, Haushalte) (Fortsetzung 2)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl BG im Leistungsbezug nach SGB II, die aus einer/einem Alleinerziehenden mit Kindern besteht, nach West/Ost (4 Kategorien)	Zahl BGs mit einem/einer Alleinerziehenden (West)	44	450.759	451.755
	Rest BGs ohne Alleinerziehenden (West)	163	1.927.950	1.927.274
	Zahl BGs mit einem/einer Alleinerziehenden (Ost)	11	189.640	189.640
	Rest BGs ohne Alleinerziehenden (Ost)	79	1.035.274	1.034.954
Anzahl der Haushalte nach Bundesland und BIK-Typ (Schreibweise: "Bundesland.BIK-Typ")	1.1 bis 1.4	12	308.000	308.000
	1.5 bis 1.6	21	139.000	139.000
	1.7 bis 1.8	37	518.000	518.000
	1.9	15	179.000	179.000
	1.10	32	221.000	221.000
	2.10	55	972.000	972.000
	3.2 bis 3.3	59	500.000	500.000
	3.4	47	442.000	442.000
	3.5	47	458.000	458.000
	3.7	131	852.000	852.000
	3.8	62	553.000	553.000
	3.9	84	623.000	623.000
	3.10	39	393.000	393.000
	4.8 bis 4.10	31	358.000	358.000
	5.2 bis 5.3	55	356.000	356.000
	5.4	97	1.038.000	1.038.000
	5.5	91	640.000	640.000
	5.6	39	355.000	355.000
	5.7	91	741.000	741.000
	5.8	199	2.143.000	2.143.000
	5.9	79	457.000	457.000
	5.10	289	2.821.000	2.821.000
	6.2	11	65.000	65.000
	6.3	50	319.000	319.000
	6.4 bis 6.5	19	322.000	322.000
	6.7	71	565.000	565.000
	6.8	24	463.000	463.000
	6.9	62	360.000	360.000
	6.10	56	794.000	794.000
	7.1 bis 7.2	37	305.000	305.000
	7.3	15	187.000	187.000
	7.4	33	178.000	178.000
7.5 bis 7.6	19	222.000	222.000	
7.7	67	394.000	394.000	
7.8	32	321.000	321.000	
7.9 bis 7.10	11	258.000	258.000	
8.2 bis 8.3	60	640.000	640.000	
8.4	71	517.000	517.000	
8.5 bis 8.6	28	475.000	475.000	

Tabelle 50: Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungsstichprobe, Haushalte) (Fortsetzung 3)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Haushalte nach Bundesland und BIK-Typ (Schreibweise: "Bundesland.BIK-Typ")	8.7	77	940.000	940.000
	8.8	75	691.000	691.000
	8.9	52	491.000	491.000
	8.10	106	1.195.000	1.195.000
	9.1 bis 9.2	27	358.000	358.000
	9.3	85	492.000	492.000
	9.4	84	692.000	692.000
	9.5	77	370.000	370.000
	9.6 bis 9.7	136	1.096.000	1.096.000
	9.8	52	643.000	643.000
	9.9	99	723.000	723.000
	9.10	127	1.499.000	1.499.000
	10.3 bis 10.4	14	155.000	155.000
	10.7 bis 10.8	39	330.000	330.000
	11.10	127	1.974.000	1.974.000
	12.1 bis 12.3	23	270.000	270.000
	12.4	28	239.000	239.000
	12.5 bis 12.6	33	174.000	174.000
	12.7	13	75.000	75.000
	12.8	16	149.000	149.000
	12.9 bis 12.10	38	332.000	332.000
	13.1 bis 13.4	23	360.000	360.000
	13.5 bis 13.6	11	161.000	161.000
	13.7	4	100.000	100.000
	13.8	21	223.000	223.000
	14.1	13	19.000	19.000
	14.2	32	132.000	132.000
	14.3 bis 14.4	24	470.000	470.000
	14.5	11	164.000	164.000
	14.6	15	118.000	118.000
	14.7 bis 14.8	26	376.000	376.000
	14.9 bis 14.10	76	912.000	912.000
	15.1 bis 15.2	16	111.000	111.000
	15.3 bis 15.4	27	292.000	292.000
15.5 bis 15.6	42	275.000	275.000	
15.7	23	234.000	234.000	
15.8	42	282.000	282.000	
16.1 bis 16.2	36	146.000	146.000	
16.3 bis 16.4	44	361.000	361.000	
16.5	34	242.000	242.000	
16.6	31	53.000	53.000	
16.7 bis 16.8	33	305.000	305.000	

Tabelle 50: Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungsstichprobe, Haushalte) (Fortsetzung 4)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Haushalte nach Haushaltsgröße (1,2,3,4,“5 und mehr Personen“) und West/Ost (10 Kategorien)	Anzahl Haushalte mit 1 Person (West)	797	12.066.000	12.066.000
	Anzahl Haushalte mit 2 Personen (West)	1.272	10.591.000	10.591.000
	Anzahl Haushalte mit 3 Personen (West)	514	3.971.000	3.971.000
	Anzahl Haushalte mit 4 Personen (West)	532	3.273.000	3.273.000
	Anzahl Haushalte mit 5 oder mehr Pers. (West)	213	1.226.000	1.226.000
	Anzahl Haushalte mit 1 Person (Ost)	254	3.651.000	3.651.000
	Anzahl Haushalte mit 2 Personen (Ost)	359	3.087.000	3.087.000
	Anzahl Haushalte mit 3 Personen (Ost)	147	1.099.000	1.099.000
	Anzahl Haushalte mit 4 Personen (Ost)	71	563.000	563.000
	Anzahl Haushalte mit 5 oder mehr Pers. (Ost)	31	149000	149000
Anzahl der Haushalte nach „Kinder unter 15 Jahre im Haushalt ja/nein“ und West/Ost	Anzahl Haushalte mit Kindern unter 15 (West)	759	5.640.000	5.640.000
	Anzahl Haushalte ohne Kinder unter 15 (West)	2.569	25.487.000	25.487.000
	Anzahl Haushalte mit Kindern unter 15 (Ost)	134	1.232.000	1.232.000
	Anzahl Haushalte ohne Kinder unter 15 (Ost)	728	7.317.000	7.317.000

Tabelle 51: Kenngrößen der Verteilung der Gewichte

1%-Perzentil	705,2742
5%-Perzentil	2162,21
10%-Perzentil	2903,944
25%-Perzentil	4492,12
50%-Perzentil	7049,734
75%-Perzentil	11456,63
90%-Perzentil	20196,32
95%-Perzentil	26008,26
99%-Perzentil	37335,97
Mittelwert	9469,212
Standardabweichung	7867,704
Minimum	277,5614
Maximum	79154,33
Fallzahl	4190
Effektivitätsmaß	59,2%

6.12.3 Gesamtstichprobe

Die Grundgesamtheit bilden alle Privathaushalte in Deutschland. Ausgangspunkt für die Kalibrierung waren die modifizierten Designgewichte inklusive der Ausfallgewichtung. Die modifizierten Designgewichte wurden am 5%-Perzentil und am 95%-Perzentil ihrer Verteilung getrimmt und anschließend so reskaliert, dass ihre Summe wieder die Summe der ungetrimmten Designgewichte ergab. Die Hochrechnungsfaktoren der getrimmten Designgewichte reichen von 209,1 bis 18628,5. Das Verhältnis der Gesamthochrechnungsfaktoren nach Kalibrierung zu getrimmten Designgewichten wurde nach unten auf minimal 0,1 und nach oben auf maximal 5,0 begrenzt. Danach bewegen sich die Gesamthochrechnungsfaktoren nach Kalibrierung im Bereich von minimal 20,9 bis maximal 46186.

Für folgende Merkmale wurde kalibriert:

Bedarfsgemeinschaften Grundlage BA-Statistik:

- Anzahl BG im Leistungsbezug nach SGB II nach Bundesländern
- Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Personen unter 65 Jahre in der Bedarfsgemeinschaft, nach West/Ost
- Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Kinder unter 15 Jahre in der Bedarfsgemeinschaft, nach West/Ost
- Anzahl BG im Leistungsbezug nach SGB II, die aus einer/einem Alleinerziehenden mit Kind(ern) besteht, nach West/Ost

Haushalte Grundlage Mikrozensus 2010:

- Anzahl der Haushalte nach Bundesland und BIK-Typ
- Anzahl der Haushalte nach Haushaltsgröße und West/Ost
- Anzahl der Haushalte nach „Kinder unter 15 Jahre im Haushalt ja/nein“ und West/Ost

Daneben wurde auch bei der Gesamtstichprobe als zusätzlicher Eckwert die Zahl der Neuzugänge zum Arbeitslosengeld-II-Leistungsbezug seit dem Vorjahr auf Ebene von Bedarfsgemeinschaften (367.870) mit aufgenommen.

Um die Kalibrierung durchzuführen, muss jede Eckwertvariable für jeden Haushalt einen gültigen Wert besitzen. Deshalb wurde der sehr geringe Item-Nonresponse vor der Kalibrierung imputiert. Die Imputation erfolgte über den Mittelwert bzw. den Modalwert der jeweiligen Variablen.

Da die Imputation einzig der Durchführbarkeit der Kalibrierung dient, wurden die imputierten Werte nach der Kalibrierung wieder auf fehlende Werte zurückgesetzt. Eine Hochrechnung mit den kalibrierten Gewichten ohne Berücksichtigung des Item-Nonresponse führt daher zu geringfügigen Abweichungen von den im Folgenden dargestellten Werten.

Tabelle 52: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Haushalte)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl BG im Leistungsbezug nach SGB II nach Bundesländern (16 Kategorien)	Zahl BGs Schleswig-Holstein	170	122.816	122.795
	Zahl BGs Hamburg	144	108.473	108.494
	Zahl BGs Niedersachsen	468	334.073	333.999
	Zahl BGs Bremen	57	51.949	51.948
	Zahl BGs Nordrhein-Westfalen	1.173	849.532	849.304
	Zahl BGs Hessen	224	219.564	219.541
	Zahl BGs Rheinland-Pfalz	162	123.522	123.634
	Zahl BGs Baden-Württemberg	324	256.618	256.578
	Zahl BGs Bayern	413	267.375	268.778
	Zahl BGs Saarland	78	44.787	44.749
	Zahl BGs Berlin	320	334.106	333.546
	Zahl BGs Brandenburg	296	167.160	166.990
	Zahl BGs Mecklenburg-Vorpommern	133	127.174	127.127
	Zahl BGs Sachsen	357	280.039	279.896
	Zahl BGs Sachsen-Anhalt	273	186.400	186.234
	Zahl BGs Thüringen	189	130.035	129.989

Tabelle 52: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Haushalte) (Fortsetzung 1)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten	
Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Personen unter 65 Jahre in der Bedarfsgemeinschaft (1, 2, 3, 4 und „5 oder mehr“) und nach West/Ost (10 Kategorien)	Zahl BGs mit 1 Person unter 65 (West)	1.297	1.287.479	1.287.478	
	Zahl BGs mit 2 Personen unter 65 (West)	842	480.035	481.373	
	Zahl BGs mit 3 Personen unter 65 (West)	541	297.590	297.435	
	Zahl BGs mit 4 Personen unter 65 (West)	314	185.355	185.402	
	Zahl BGs mit 5 oder mehr Personen unter 65 (West)	219	128.250	128.132	
		710	714.572	714.572	
	Zahl BGs mit 1 Person unter 65 (Ost)	449	264.338	263.868	
	Zahl BGs mit 2 Personen unter 65 (Ost)	243	135.917	135.359	
	Zahl BGs mit 3 Personen unter 65 (Ost)	97	71.425	71.465	
	Zahl BGs mit 4 Personen unter 65 (Ost)	69	38.662	38.520	
	Zahl BGs mit 5 oder mehr Personen unter 65 (Ost)				
	Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Kinder unter 15 Jahre in der Bedarfsgemeinschaft (0, 1, 2, 3, „4 oder mehr“) und nach West/Ost (10 Kategorien)	Zahl BGs ohne Kinder unter 15 Jahre (West)	2.136	1.629.124	1.630.364
		Zahl BGs mit 1 Kind unter 15 Jahre (West)	577	406.480	406.504
Zahl BGs mit 2 Kindern unter 15 Jahre (West)		351	230.977	230.817	
Zahl BGs mit 3 Kindern unter 15 Jahre (West)		101	80.359	80.376	
Zahl BGs mit 4 oder mehr Kindern unter 15 Jahre (West)		48	31.769	31.759	
Zahl BGs ohne Kinder unter 15 Jahre (Ost)		1.160	909.771	908.804	
Zahl BGs mit 1 Kind unter 15 Jahre (Ost)		258	188.910	188.887	
Zahl BGs mit 2 Kindern unter 15 Jahre (Ost)		111	90.185	90.044	
Zahl BGs mit 3 Kindern unter 15 Jahre (Ost)		25	25.972	25.972	
Zahl BGs mit 4 oder mehr Kindern unter 15 Jahre (Ost)		14	10.076	10.076	

Tabelle 52: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Haushalte) (Fortsetzung 2)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl BG im Leistungsbezug nach SGB II, die aus einer/einem Alleinerziehenden mit Kindern besteht, nach West/Ost (4 Kategorien)	Zahl BGs mit einem/einer Alleinerziehenden (West)	759	450.759	450.656
	Rest BGs ohne Alleinerziehenden (West)	2.454	1.927.950	1.929.164
	Zahl BGs mit einem/einer Alleinerziehenden (Ost)	269	189.640	189.580
	Rest BGs ohne Alleinerziehenden (Ost)	1.299	1.035.274	1.034.203
Anzahl der Haushalte nach Bundesland und BIK-Typ (Schreibweise: "Bundesland.BIK-Typ")	1.1 bis 1.4	48	308.000	308.000
	1.5 bis 1.6	45	139.000	139.000
	1.7 bis 1.8	134	518.000	518.000
	1.9	35	179.000	179.000
	1.10	78	221.000	221.000
	2.10	226	972.000	972.000
	3.2 bis 3.3	129	500.000	500.000
	3.4	102	442.000	442.000
	3.5	67	458.000	458.000
	3.7	261	852.000	852.000
	3.8	193	553.000	553.000
	3.9	155	623.000	623.000
	3.10	135	393.000	393.000
	4.8 bis 4.10	96	358.000	358.000
	5.2 bis 5.3	126	356.000	356.000
	5.4	228	1.038.000	1.038.000
	5.5	231	640.000	640.000
	5.6	103	355.000	355.000
	5.7	204	741.000	741.000
	5.8	535	2.143.000	2.143.000
	5.9	152	457.000	457.000
	5.10	770	2.821.000	2.821.000
	6.2	18	65.000	65.000
	6.3	93	319.000	319.000
	6.4 bis 6.5	46	322.000	322.000
	6.7	122	565.000	565.000
	6.8	83	463.000	463.000
	6.9	105	360.000	360.000
	6.10	131	794.000	794.000
	7.1 bis 7.2	59	305.000	305.000
	7.3	23	187.000	187.000
	7.4	53	178.000	178.000
7.5 bis 7.6	61	222.000	222.000	
7.7	86	394.000	394.000	
7.8	87	321.000	321.000	
7.9 bis 7.10	39	258.000	258.000	
8.2 bis 8.3	96	640.000	640.000	
8.4	114	517.000	517.000	
8.5 bis 8.6	42	475.000	475.000	

Tabelle 52: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Haushalte) (Fortsetzung 3)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Haushalte nach Bundesland und BIK-Typ (Schreibweise: "Bundesland.BIK-Typ")	8.7	154	940.000	940.000
	8.8	154	691.000	691.000
	8.9	92	491.000	491.000
	8.10	240	1.195.000	1.195.000
	9.1 bis 9.2	40	358.000	358.000
	9.3	122	492.000	492.000
	9.4	127	692.000	692.000
	9.5	128	370.000	370.000
	9.6 bis 9.7	218	1.096.000	1.096.000
	9.8	151	643.000	643.000
	9.9	140	723.000	723.000
	9.10	305	1.499.000	1.499.000
	10.3 bis 10.4	39	155.000	155.000
	10.7 bis 10.8	103	330.000	330.000
	11.10	556	1.974.000	1.974.000
	12.1 bis 12.3	108	270.000	270.000
	12.4	86	239.000	239.000
	12.5 bis 12.6	126	174.000	174.000
	12.7	44	75.000	75.000
	12.8	33	149.000	149.000
	12.9 bis 12.10	107	332.000	332.000
	13.1 bis 13.4	102	360.000	360.000
	13.5 bis 13.6	44	161.000	161.000
	13.7	39	100.000	100.000
	13.8	63	223.000	223.000
	14.1	19	19.000	19.000
	14.2	104	132.000	132.000
	14.3 bis 14.4	82	470.000	470.000
	14.5	40	164.000	164.000
	14.6	96	118.000	118.000
	14.7 bis 14.8	68	376.000	376.000
	14.9 bis 14.10	253	912.000	912.000
	15.1 bis 15.2	62	111.000	111.000
	15.3 bis 15.4	69	292.000	292.000
15.5 bis 15.6	105	275.000	275.000	
15.7	101	234.000	234.000	
15.8	158	282.000	282.000	
16.1 bis 16.2	73	146.000	146.000	
16.3 bis 16.4	115	361.000	361.000	
16.5	89	242.000	242.000	
16.6	59	53.000	53.000	
16.7 bis 16.8	110	305.000	305.000	

Tabelle 52: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Haushalte) (Fortsetzung 4)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Haushalte nach Haushaltsgröße (1,2,3,4,“5 und mehr Personen“) und West/Ost (10 Kategorien)	Anzahl Haushalte mit 1 Person (West)	2.310	12.066.000	12.066.000
	Anzahl Haushalte mit 2 Personen (West)	2.392	10.591.000	10.591.000
	Anzahl Haushalte mit 3 Personen (West)	1.202	3.971.000	3.971.000
	Anzahl Haushalte mit 4 Personen (West)	946	3.273.000	3.273.000
	Anzahl Haushalte mit 5 oder mehr Pers. (West)	474	1.226.000	1.226.000
	Anzahl Haushalte mit 1 Person (Ost)	1.144	3.651.000	3.651.000
	Anzahl Haushalte mit 2 Personen (Ost)	958	3.087.000	3.087.000
	Anzahl Haushalte mit 3 Personen (Ost)	484	1.099.000	1.099.000
	Anzahl Haushalte mit 4 Personen (Ost)	213	563.000	563.000
	Anzahl Haushalte mit 5 oder mehr Pers. (Ost)	112	149000	149000
Anzahl der Haushalte nach „Kinder unter 15 Jahre im Haushalt ja/nein“ und West/Ost	Anzahl Haushalte mit Kindern unter 15 (West)	2.086	5.640.000	5.640.000
	Anzahl Haushalte ohne Kinder unter 15 (West)	5.238	25.487.000	25.487.000
	Anzahl Haushalte mit Kindern unter 15 (Ost)	647	1.232.000	1.232.000
	Anzahl Haushalte ohne Kinder unter 15 (Ost)	2.264	7.317.000	7.317.000

Tabelle 53: Kenngrößen der Verteilung der Gewichte

1%-Perzentil	42,12403
5%-Perzentil	108,0065
10%-Perzentil	174,8335
25%-Perzentil	402,8683
50%-Perzentil	1049,692
75%-Perzentil	5171,642
90%-Perzentil	11155,47
95%-Perzentil	17561,72
99%-Perzentil	25649,94
Mittelwert	3876,502
Standardabweichung	5832,188
Minimum	20,90576
Maximum	46186
Fallzahl	10235
Effektivitätsmaß	30,6%

6.13 Kalibrierung zum Personengewicht Welle 5 Querschnitt

Die Kalibrierung der Personengewichte wurde – wie in den Vorgängerwellen – unter der Restriktion möglichst geringer Abweichung von den kalibrierten Haushaltsgewichten vorgenommen. Die kalibrierten Haushaltsgewichte wurden an die einzelnen Haushaltsmitglieder gewissermaßen vererbt. Anschließend wurden diese Eingangsgewichte auf Personenebene kalibriert.

Ebenso wie im Vorjahr wurde dabei als Eckwert auch die Zahl der Neuzugänge zum Arbeitslosengeld-II-Leistungsbezug seit dem Vorjahr auf Ebene von Personen zwischen 15 und 64 Jahre (499.235) verwendet. Es werden erneut diejenigen Fälle der Stichproben aus Welle 1, Welle 2, Welle 3 und Welle 4 auf die Eckwerte zum Arbeitslosengeld-II-Bezug der Bundesagentur für Arbeit hochgerechnet, die laut Welle 5 des Surveys im Juli 2010 Arbeitslosengeld II bezogen haben.

Vor der Kalibrierung wurden die kalibrierten Haushaltsgewichte, die das Eingangsgewicht bilden, ebenfalls getrimmt. Auch bei der Kalibrierung der Personengewichte wurde zudem der Bereich der Gewichte auf ein bestimmtes Intervall festgelegt.

6.13.1 BA-Stichprobe

Die Grundgesamtheit der kumulierten BA-Stichprobe aller fünf Wellen bilden alle Personen im Alter ab 15 Jahre, die in einem Haushalt leben, in dem es zu einem der bisher fünf Ziehungszeitpunkte (im Juli 2006, Juli 2007, Juli 2008, Juli 2009 oder Juli 2010) mindestens eine Bedarfsgemeinschaft mit Bezug von Leistungen nach dem SGB II gab. Für die Kalibrierung wurden lediglich all diejenigen Personen im Alter ab 15 Jahre, die in einer Bedarfsgemeinschaft mit Leistungen nach dem SGB II im Juli 2010 leben, berücksichtigt.

Personen, die in Haushalten leben, die den Bezug inzwischen komplett verlassen haben und Personen, die zwar noch in einem Haushalt mit mindestens einer Bedarfsgemeinschaft nach SGB II leben, selbst aber nicht Teil einer Bedarfsgemeinschaft sind, wurden für die Kalibrierung aus dem Datensatz entfernt. Die Gewichte für diese Personen wurden über einen anderen Weg berechnet (s.u.).

Ausgangspunkt für die Kalibrierung sind die kalibrierten Haushaltsgewichte der BA-Stichprobe. Diese wurden am 5%-Perzentil und am 95%-Perzentil ihrer Verteilung getrimmt und anschließend so reskaliert, dass ihre Summe wieder die Summe der ungetrimmten kalibrierten Haushaltsgewichte ergab. Die getrimmten Hochrechnungsfaktoren reichen von 284,1 bis 3407,7. Das Verhältnis der Gesamthochrechnungsfaktoren nach Kalibrierung zu getrimmten Designgewichten wurde nach unten auf minimal 0,2 und nach oben auf maximal 3,5 begrenzt. Danach bewegen sich die Gesamthochrechnungsfaktoren nach Kalibrierung im Bereich von minimal 78,9 bis maximal 4699,1.

Für folgende Merkmale wurde kalibriert:

Leistungsempfänger Grundlage BA-Statistik:

- Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Bundesländern
- Anzahl der Personen in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Alter (15-24 und 25-64)
- Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Geschlecht, nach West/Ost
- Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach „Alleinerziehend ja/nein“, nach West/Ost
- Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Staatsangehörigkeit (deutsch/nicht deutsch)

Daneben wurde wie im Vorjahr als zusätzlicher Eckwert die Zahl der Neuzugänge zum Arbeitslosengeld-II-Leistungsbezug seit dem Vorjahr auf Ebene von Personen zwischen 15 und 64 Jahre (499.235) mit aufgenommen.

Um die Kalibrierung durchzuführen, muss jede Eckwertvariable für jede Person einen gültigen Wert besitzen. Deshalb wurde der sehr geringe Item-Nonresponse vor der Kalibrierung imputiert. Die Imputation erfolgte über den Mittelwert bzw. den Modalwert der jeweiligen Variablen.

Da die Imputation einzig der Durchführbarkeit der Kalibrierung dient, wurden die imputierten Werte nach der Kalibrierung wieder auf fehlende Werte zurückgesetzt. Eine Hochrechnung mit den kalibrierten Gewichten ohne Berücksichtigung des Item-Nonresponse führt daher zu geringfügigen Abweichungen von den im Folgenden dargestellten Werten.

Tabelle 54: Soll-Verteilungen und Verteilungen nach Kalibrierung (BA-Stichprobe, Personen)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Bundesländern (16 Kategorien)	Zahl Personen in BGs Schleswig-Holstein	203	170.430	170.430
	Zahl Personen in BGs Hamburg	173	146.202	146.202
	Zahl Personen in BGs Niedersachsen	562	472.434	472.434
	Zahl Personen in BGs Bremen	63	70.746	70.746
	Zahl Personen in BGs Nordrhein-Westfalen	1.470	1.217.381	1.217.381
	Zahl Personen in BGs Hessen	284	315.063	315.063
	Zahl Personen in BGs Rheinland-Pfalz	198	175.315	175.315
	Zahl Personen in BGs Baden-Württemberg	403	355.805	355.805
	Zahl Personen in BGs Bayern	493	358.390	358.390
	Zahl Personen in BGs Saarland	91	61.906	61.906
	Zahl Personen in BGs Berlin	357	450.129	450.129
	Zahl Personen in BGs Brandenburg	358	225.513	225.513
	Zahl Personen in BGs Mecklenburg-Vorpommern	156	171.727	171.727
	Zahl Personen in BGs Sachsen	441	379.470	379.470
	Zahl Personen in BGs Sachsen-Anhalt	373	256.044	256.044
Zahl Personen in BGs Thüringen	234	175.557	175.557	
Anzahl der Personen in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Alter (15-24 und 25-64; 2 Kategorien)	Zahl Personen in BGs im Alter 15-24	906	939.152	939.152
	Zahl Personen in BGs im Alter 25-64	4.953	4.062.960	4.062.960
Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Geschlecht und nach West/Ost (4 Kategorien)	Zahl Männer in BGs (West)	1.792	1.637.643	1.637.643
	Zahl Frauen in BGs (West)	2.148	1.706.029	1.706.029
	Zahl Männer in BGs (Ost)	923	854.217	854.217
	Zahl Frauen in BGs (Ost)	996	804.223	804.223
Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach „Alleinerziehend ja/nein“, nach Geschlecht und nach West/Ost (8 Kategorien)	Zahl nicht alleinerziehende in BGs (West)	3.210	2.892.913	2.892.913
	Zahl alleinerziehende in BGs (West)	730	450.759	450.759
	Zahl nicht alleinerziehende in BGs (Ost)	1.646	1.468.800	1.468.800
	Zahl alleinerziehende in BGs (Ost)	273	189.640	189.640
Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Staatsangehörigkeit (deutsch/nicht deutsch)	Zahl nicht deutscher Personen in BGs	678	990.860	990.860
	Zahl deutscher Personen in BGs	5.181	4.011.252	4.011.252

Tabelle 55: Kenngrößen der Verteilung der Gewichte

1%-Perzentil	118,4711
5%-Perzentil	168,7721
10%-Perzentil	225,2746
25%-Perzentil	357,9115
50%-Perzentil	669,5241
75%-Perzentil	1152,045
90%-Perzentil	1825,846
95%-Perzentil	2114,982
99%-Perzentil	3300,052
Mittelwert	854,1858
Standardabweichung	655,8097
Minimum	78,91148
Maximum	4699,068
Fallzahl	5856
Effektivitätsmaß	62,9%

6.13.2 Bevölkerungsstichprobe

Die Grundgesamtheit bilden alle über 14-jährigen Personen in Privathaushalten in Deutschland. Ausgangspunkt für die Kalibrierung waren die kalibrierten Haushaltsgewichte der Bevölkerungsstichprobe. Diese wurden am 5%-Perzentil und am 95%-Perzentil ihrer Verteilung getrimmt und anschließend so reskaliert, dass ihre Summe wieder die Summe der ungetrimmten kalibrierten Haushaltsgewichte ergab. Die getrimmten Hochrechnungsfaktoren reichen von 2358,6 bis 28801. Das Verhältnis der Gesamthochrechnungsfaktoren nach Kalibrierung zu getrimmten Designgewichten wurde nach unten auf minimal 0,2 und nach oben auf maximal 3,5 begrenzt. Danach bewegen sich die Gesamthochrechnungsfaktoren nach Kalibrierung im Bereich von minimal 471,7 bis maximal 79063,4.

Für folgende Merkmale wurde kalibriert:

Leistungsempfänger Grundlage BA-Statistik:

- Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Bundesländern
- Anzahl der Personen in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Alter (15-24 und 25-64)
- Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Geschlecht, nach West/Ost
- Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach „Alleinerziehend ja/nein“, nach West/Ost
- Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Staatsangehörigkeit (deutsch/nicht deutsch)

Bevölkerung Grundlage Mikrozensus 2010:

- Anzahl der Personen ab 15 Jahre in Privathaushalten nach Bundesland
- Anzahl der Personen ab 15 Jahre in Privathaushalten nach Alter, Geschlecht und West/Ost
- Anzahl der Personen ab 15 Jahre in Privathaushalten nach Haushaltsgröße, nach West/Ost
- Anzahl der Personen ab 15 Jahre in Privathaushalten nach höchstem Schulabschluss, nach West/Ost
- Anzahl der Personen ab 15 Jahre in Privathaushalten nach Familienstand und West/Ost
- Anzahl der Personen ab 15 Jahre in Privathaushalten nach Staatsangehörigkeiten

Bevölkerung Grundlage BA Statistik:

- Anzahl der Arbeitslosen inklusive Maßnahmeteilnehmer nach West/Ost
- Anzahl der sozialversicherungspflichtig Beschäftigten nach West/Ost

Als Quelle für den Eckwert des Erwerbsstatus wurde die BA-Statistik verwendet, da die Definition der Arbeitslosigkeit und der sozialversicherungspflichtigen Beschäftigung im PASS nicht dem ILO-Konzept des statistischen Bundesamtes entspricht, den Statistiken der BA aber entnommen werden kann.

Um die Kalibrierung durchzuführen, muss jede Eckwertvariable für jede Person einen gültigen Wert besitzen. Deshalb wurde der sehr geringe Item-Nonresponse vor der Kalibrierung imputiert. Die Imputation erfolgte über den Mittelwert bzw. den Modalwert der jeweiligen Variablen.

Da die Imputation einzig der Durchführbarkeit der Kalibrierung dient, wurden die imputierten Werte nach der Kalibrierung wieder auf fehlende Werte zurückgesetzt. Eine Hochrechnung mit den kalibrierten Gewichten ohne Berücksichtigung des Item-Nonresponse führt daher zu geringfügigen Abweichungen von den im Folgenden dargestellten Werten.

Tabelle 56: Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungstichprobe, Personen)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Bundesländern (16 Kategorien)	Zahl Personen in BGs Schleswig-Holstein	10	170.430	170.430
	Zahl Personen in BGs Hamburg	8	146.202	146.202
	Zahl Personen in BGs Niedersachsen	45	472.434	472.434
	Zahl Personen in BGs Bremen	9	70.746	70.746
	Zahl Personen in BGs Nordrhein-Westfalen	143	1.217.381	1.217.381
	Zahl Personen in BGs Hessen	15	315.063	315.063
	Zahl Personen in BGs Rheinland-Pfalz	18	175.315	175.315
	Zahl Personen in BGs Baden-Württemberg	19	355.805	355.805
	Zahl Personen in BGs Bayern	35	358.390	358.390
	Zahl Personen in BGs Saarland	10	61.906	61.906
	Zahl Personen in BGs Berlin	19	450.129	450.129
	Zahl Personen in BGs Brandenburg	34	225.513	225.513
	Zahl Personen in BGs Mecklenburg-Vorpommern	8	171.727	171.727
	Zahl Personen in BGs Sachsen	27	379.470	379.470
	Zahl Personen in BGs Sachsen-Anhalt	19	256.044	256.044
	Zahl Personen in BGs Thüringen	21	175.557	175.557
Anzahl der Personen in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Alter (15-24 und 25-64; 2 Kategorien)	Zahl Personen in BGs im Alter 15-24	76	939.152	939.152
	Zahl Personen in BGs im Alter 25-64	364	4.062.960	4.062.960
Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Geschlecht und nach West/Ost (4 Kategorien)	Zahl Männer in BGs (West)	148	1.637.643	1.637.643
	Zahl Frauen in BGs (West)	164	1.706.029	1.706.029
	Zahl Männer in BGs (Ost)	65	854.217	854.217
	Zahl Frauen in BGs (Ost)	63	804.223	804.223
Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach „Alleinerziehend ja/nein“, nach Geschlecht und nach West/Ost (8 Kategorien)	Zahl nicht alleinerziehende in BGs (West)	267	2.892.913	2.892.913
	Zahl alleinerziehende in BGs (West)	45	450.759	450.759
	Zahl nicht alleinerziehende in BGs (Ost)	115	1.468.800	1.468.800
	Zahl alleinerziehende in BGs (Ost)	13	189.640	189.640
Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Staatsangehörigkeit (deutsch/nicht deutsch)	Zahl nicht deutscher Personen in BGs	59	990.860	990.860
	Zahl deutscher Personen in BGs	381	4.011.252	4.011.252

Tabelle 56: Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungsstichprobe, Personen) (Fortsetzung 1)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Bundesland (16 Kategorien)	Zahl Personen in Privathaushalten Schleswig-Holstein	188	2.398.000	2.398.000
	Zahl Personen in Privathaushalten Hamburg	82	1.550.000	1.550.000
	Zahl Personen in Privathaushalten Niedersachsen	789	6.736.000	6.736.000
	Zahl Personen in Privathaushalten Bremen	43	577.000	577.000
	Zahl Personen in Privathaushalten Nordrhein-Westfalen	1.560	15.266.000	15.266.000
	Zahl Personen in Privathaushalten Hessen	532	5.201.000	5.201.000
	Zahl Personen in Privathaushalten Rheinland-Pfalz	360	3.437.000	3.437.000
	Zahl Personen in Privathaushalten Baden-Württemberg	818	9.138.000	9.138.000
	Zahl Personen in Privathaushalten Bayern	1.214	10.646.000	10.646.000
	Zahl Personen in Privathaushalten Saarland	96	885.000	885.000
	Zahl Personen in Privathaushalten Berlin	177	3.017.000	3.017.000
	Zahl Personen in Privathaushalten Brandenburg	253	2.207.000	2.207.000
	Zahl Personen in Privathaushalten Mecklenburg-Vorpommern	102	1.450.000	1.450.000
	Zahl Personen in Privathaushalten Sachsen	319	3.673.000	3.673.000
	Zahl Personen in Privathaushalten Sachsen-Anhalt	262	2.061.000	2.061.000
	Zahl Personen in Privathaushalten Thüringen	305	1.961.000	1.961.000
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Alter (in 5-Jahres-Klassen), Geschlecht und West/Ost (56 Kategorien)	Zahl Männer in PH (West), 15-19 Jahre	195	1.884.000	1.884.000
	Zahl Männer in PH (West), 20-24 Jahre	191	1.939.000	1.939.000
	Zahl Männer in PH (West), 25-29 Jahre	110	1.948.000	1.948.000
	Zahl Männer in PH (West), 30-34 Jahre	123	1.908.000	1.908.000
	Zahl Männer in PH (West), 35-39 Jahre	147	2.062.000	2.062.000
	Zahl Männer in PH (West), 40-44 Jahre	252	2.826.000	2.826.000
	Zahl Männer in PH (West), 45-49 Jahre	284	2.736.000	2.736.000
	Zahl Männer in PH (West), 50-54 Jahre	299	2.383.000	2.383.000
Zahl Männer in PH (West), 55-59 Jahre	237	2.068.000	2.068.000	

Tabelle 56: Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungsstichprobe, Personen) (Fortsetzung 2)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Alter (in 5-Jahres-Klassen), Geschlecht und West/Ost (56 Kategorien)	Zahl Männer in PH (West), 60-64 Jahre	232	1.781.000	1.781.000
	Zahl Männer in PH (West), 65-69 Jahre	208	1.749.000	1.749.000
	Zahl Männer in PH (West), 70-74 Jahre	241	1.787.000	1.787.000
	Zahl Männer in PH (West), 75-79 Jahre	100	1.109.000	1.109.000
	Zahl Männer in PH (West), 80+ Jahre	83	1.048.000	1.048.000
	Zahl Frauen in PH (West), 15-19 Jahre	215	1.768.000	1.768.000
	Zahl Frauen in PH (West), 20-24 Jahre	148	1.896.000	1.896.000
	Zahl Frauen in PH (West), 25-29 Jahre	122	1.944.000	1.944.000
	Zahl Frauen in PH (West), 30-34 Jahre	144	1.916.000	1.916.000
	Zahl Frauen in PH (West), 35-39 Jahre	200	2.052.000	2.052.000
	Zahl Frauen in PH (West), 40-44 Jahre	292	2.722.000	2.722.000
	Zahl Frauen in PH (West), 45-49 Jahre	359	2.678.000	2.678.000
	Zahl Frauen in PH (West), 50-54 Jahre	324	2.426.000	2.426.000
	Zahl Frauen in PH (West), 55-59 Jahre	300	2.120.000	2.120.000
	Zahl Frauen in PH (West), 60-64 Jahre	260	1.841.000	1.841.000
	Zahl Frauen in PH (West), 65-69 Jahre	203	1.894.000	1.894.000
	Zahl Frauen in PH (West), 70-74 Jahre	222	2.037.000	2.037.000
	Zahl Frauen in PH (West), 75-79 Jahre	118	1.375.000	1.375.000
	Zahl Frauen in PH (West), 80+ Jahre	73	1.937.000	1.937.000
	Zahl Männer in PH (Ost), 15-19 Jahre	34	317.000	317.000
	Zahl Männer in PH (Ost), 20-24 Jahre	51	556.000	556.000
	Zahl Männer in PH (Ost), 25-29 Jahre	43	550.000	550.000
	Zahl Männer in PH (Ost), 30-34 Jahre	41	509.000	509.000
	Zahl Männer in PH (Ost), 35-39 Jahre	33	507.000	507.000
	Zahl Männer in PH (Ost), 40-44 Jahre	53	655.000	655.000
	Zahl Männer in PH (Ost), 45-49 Jahre	84	718.000	718.000
	Zahl Männer in PH (Ost), 50-54 Jahre	77	635.000	635.000
	Zahl Männer in PH (Ost), 55-59 Jahre	62	621.000	621.000
	Zahl Männer in PH (Ost), 60-64 Jahre	58	438.000	438.000
	Zahl Männer in PH (Ost), 65-69 Jahre	54	504.000	504.000
	Zahl Männer in PH (Ost), 70-74 Jahre	45	502.000	502.000
	Zahl Männer in PH (Ost), 75-79 Jahre	37	301.000	301.000
	Zahl Männer in PH (Ost), 80+ Jahre	16	237.000	237.000

Tabelle 56: Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungsstichprobe, Personen) (Fortsetzung 3)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Alter (in 5-Jahres-Klassen), Geschlecht und West/Ost (56 Kategorien)	Zahl Frauen in PH (Ost), 15-19 Jahre	28	279.000	279.000
	Zahl Frauen in PH (Ost), 20-24 Jahre	38	497.000	497.000
	Zahl Frauen in PH (Ost), 25-29 Jahre	37	491.000	491.000
	Zahl Frauen in PH (Ost), 30-34 Jahre	29	461.000	461.000
	Zahl Frauen in PH (Ost), 35-39 Jahre	31	475.000	475.000
	Zahl Frauen in PH (Ost), 40-44 Jahre	56	624.000	624.000
	Zahl Frauen in PH (Ost), 45-49 Jahre	88	681.000	681.000
	Zahl Frauen in PH (Ost), 50-54 Jahre	78	642.000	642.000
	Zahl Frauen in PH (Ost), 55-59 Jahre	88	636.000	636.000
	Zahl Frauen in PH (Ost), 60-64 Jahre	71	454.000	454.000
	Zahl Frauen in PH (Ost), 65-69 Jahre	54	571.000	571.000
	Zahl Frauen in PH (Ost), 70-74 Jahre	68	604.000	604.000
	Zahl Frauen in PH (Ost), 75-79 Jahre	38	404.000	404.000
	Zahl Frauen in PH (Ost), 80+ Jahre	26	500.000	500.000
	Anzahl der Personen ab 15 Jahre in Privathaushalten nach Haushaltsgröße (1, 2, 3, 4, „5 oder mehr Personen“) und West/Ost (10 Kategorien)	Zahl Personen in PH mit 1 Person (West)	792	12.066.000
Zahl Personen in PH mit 2 Personen (West)		2.144	20.714.000	20.714.000
Zahl Personen in PH mit 3 Personen (West)		1.005	9.811.000	9.811.000
Zahl Personen in PH mit 4 Personen (West)		1.177	9.128.000	9.128.000
Zahl Personen in PH mit 5 oder mehr Personen (West)		564	4.115.000	4.115.000
Zahl Personen in PH mit 1 Person (Ost)		253	3.651.000	3.651.000
Zahl Personen in PH mit 2 Personen (Ost)		633	5.998.000	5.998.000
Zahl Personen in PH mit 3 Personen (Ost)		304	2.721.000	2.721.000
Zahl Personen in PH mit 4 Personen (Ost)		147	1.527.000	1.527.000
Zahl Personen in PH mit 5 oder mehr Personen (Ost)		81	472.000	472.000

Tabelle 56: Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungsstichprobe, Personen) (Fortsetzung 4)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahre in Privathaushalten nach höchstem Schulabschluss und West/Ost (12 Kategorien)	Zahl Personen in PH mit höchstem Schulabschluss: noch Schüler (West)	245	2.273.000	2.273.000
	Zahl Personen in PH mit höchstem Schulabschluss: ohne Abschluss (West)	161	2.336.000	2.336.000
	Zahl Personen in PH mit höchstem Schulabschluss: Hauptschule (West)	1.903	22.759.000	22.759.000
	Zahl Personen in PH mit höchstem Schulabschluss: Realschule/POS (West)	1.626	13.732.000	13.732.000
	Zahl Personen in PH mit höchstem Schulabschluss:(Fach-)Hochschulreife (West)	1.747	14.734.000	14.734.000
	Zahl Personen in PH mit höchstem Schulabschluss: noch Schüler (Ost)	27	343.000	343.000
	Zahl Personen in PH mit höchstem Schulabschluss: ohne Abschluss (Ost)	26	361.000	361.000
	Zahl Personen in PH mit höchstem Schulabschluss: Hauptschule (Ost)	282	3.270.000	3.270.000
	Zahl Personen in PH mit höchstem Schulabschluss: Realschule/POS (Ost)	697	6.752.000	6.752.000
	Zahl Personen in PH mit höchstem Schulabschluss: (Fach-)Hochschulreife (Ost)	386	3.643.000	3.643.000
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Familienstand und West/Ost (10 Kategorien)	Zahl Personen in PH mit Familienstand: ledig (West)	1.311	10.125.000	10.125.000
	Zahl Personen in PH mit Familienstand: verheiratet, eingetragene Lebensgemeinschaft. (West)	3.654	36.063.000	36.063.000
	Zahl Personen in PH mit Familienstand: geschieden (West)	406	4.883.000	4.883.000
	Zahl Personen in PH mit Familienstand: verwitwet (West)	311	4.763.000	4.763.000
	Zahl Personen in PH mit Familienstand: ledig (Ost)	339	3.351.000	3.351.000
	Zahl Personen in PH mit Familienstand: verheiratet, eingetragene Lebensgemeinschaft. (Ost)	856	8.264.000	8.264.000
	Zahl Personen in PH mit Familienstand: geschieden (Ost)	117	1.466.000	1.466.000
	Zahl Personen in PH mit Familienstand: verwitwet (Ost)	106	1.288.000	1.288.000
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Staatsangehörigkeit	Zahl Personen in PH Nicht-Deutsch	290	6.340.000	6.340.000
	Zahl Personen in PH Deutsch	6.810	63.863.000	63.863.000

Tabelle 56: Soll-Verteilungen und Verteilungen nach Kalibrierung (Bevölkerungsstichprobe, Personen) (Fortsetzung 5)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Arbeitslose inkl. Maßnahmeteilnehmer West/Ost	Nicht arbeitslos West	5.409	53.032.681	53.032.681
	Arbeitslose inkl. Maßnahmeteilnehmer West	273	2.801.319	2.801.319
	Nicht arbeitslos Ost	1.309	13.084.328	13.084.328
	Arbeitslose inkl. Maßnahmeteilnehmer Ost	109	1.284.672	1.284.672
Sozialversicherungspflichtig beschäftigt West/Ost	Nicht-Sozialversicherungspflichtig beschäftigt West	3.431	32.861.296	32.861.296
	Sozialversicherungspflichtig beschäftigt West	2.251	22.972.704	22.972.704
	Nicht-Sozialversicherungspflichtig beschäftigt Ost	790	9.000.117	9.000.117
	Sozialversicherungspflichtig beschäftigt Ost	628	5.368.883	5.368.883

Tabelle 57: Kenngrößen der Verteilung der Gewichte

1%-Perzentil	962,9659
5%-Perzentil	1799,035
10%-Perzentil	2360,424
25%-Perzentil	3893,852
50%-Perzentil	6806,903
75%-Perzentil	12181,16
90%-Perzentil	21342,62
95%-Perzentil	29527,5
99%-Perzentil	47323,72
Mittelwert	9887,747
Standardabweichung	9479,726
Minimum	471,7283
Maximum	79063,38
Fallzahl	7100
Effektivitätsmaß	52,1%

6.13.3 Gesamtstichprobe

Die Grundgesamtheit bilden wie bei der Bevölkerungsstichprobe alle 15-jährigen und älteren Personen in Privathaushalten in Deutschland. Ausgangspunkt für die Kalibrierung waren die kalibrierten Haushaltsgewichte der Gesamtstichprobe. Diese wurden am 5%-Perzentil und am 95%-Perzentil ihrer Verteilung getrimmt und anschließend so reskaliert, dass ihre Summe wieder die Summe der ungetrimmten kalibrierten Haushaltsgewichte ergab. Die getrimmten Hochrechnungsfaktoren reichen von 121,5 bis 20778,5. Das Verhältnis der Gesamthochrechnungsfaktoren nach Kalibrierung zu getrimmten Designgewichten wurde nach unten auf minimal 0,2 und nach oben auf maximal 3,5 begrenzt. Danach bewegen sich die Gesamthochrechnungsfaktoren nach Kalibrierung im Bereich von minimal 24,3 bis maximal 65767,4.

Für folgende Merkmale wurde kalibriert:

Leistungsempfänger Grundlage BA-Statistik:

- Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Bundesländern
- Anzahl der Personen in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Alter (15-24 und 25-64)
- Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Geschlecht, nach West/Ost
- Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach „Alleinerziehend ja/nein“, nach West/Ost
- Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Staatsangehörigkeit (deutsch/nicht deutsch)

Bevölkerung Grundlage Mikrozensus 2010:

- Anzahl der Personen ab 15 Jahre in Privathaushalten nach Bundesland
- Anzahl der Personen ab 15 Jahre in Privathaushalten nach Alter, Geschlecht und West/Ost
- Anzahl der Personen ab 15 Jahre in Privathaushalten nach Haushaltsgröße, nach West/Ost
- Anzahl der Personen ab 15 Jahre in Privathaushalten nach höchstem Schulabschluss, nach West/Ost
- Anzahl der Personen ab 15 Jahre in Privathaushalten nach Familienstand und West/Ost
- Anzahl der Personen ab 15 Jahre in Privathaushalten nach Staatsangehörigkeiten

Bevölkerung Grundlage BA Statistik:

- Anzahl der Arbeitslosen inklusive Maßnahmeteilnehmer nach West/Ost
- Anzahl der sozialversicherungspflichtig Beschäftigten nach West/Ost

Als Quelle für den Eckwert des Erwerbsstatus wurde die BA-Statistik verwendet, da die Definition der Arbeitslosigkeit und der sozialversicherungspflichtigen Beschäftigung im PASS nicht dem ILO-Konzept des statistischen Bundesamtes entspricht, den Statistiken der BA aber entnommen werden kann.

Daneben wurde auch bei der Gesamtstichprobe als zusätzlicher Eckwert die Zahl der Neuzugänge zum Arbeitslosengeld-II-Leistungsbezug seit dem Vorjahr auf Ebene von Personen zwischen 15 und 64 Jahren (499.235) mit aufgenommen.

Um die Kalibrierung durchzuführen, muss jede Eckwertvariable für jede Person einen gültigen Wert besitzen. Deshalb wurde der sehr geringe Item-Nonresponse vor der Kalibrierung imputiert. Die Imputation erfolgte über den Mittelwert bzw. den Modalwert der jeweiligen Variablen.

Da die Imputation einzig der Durchführbarkeit der Kalibrierung dient, wurden die imputierten Werte nach der Kalibrierung wieder auf fehlende Werte zurückgesetzt. Eine Hochrechnung mit den kalibrierten Gewichten ohne Berücksichtigung des Item-Nonresponse führt daher zu geringfügigen Abweichungen von den im Folgenden dargestellten Werten.

Tabelle 58: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Bundesländern (16 Kategorien)	Zahl Personen in BGs Schleswig-Holstein	213	170.430	170.430
	Zahl Personen in BGs Hamburg	181	146.202	146.202
	Zahl Personen in BGs Niedersachsen	607	472.434	472.434
	Zahl Personen in BGs Bremen	72	70.746	70.746
	Zahl Personen in BGs Nordrhein-Westfalen	1.613	1.217.381	1.217.381
	Zahl Personen in BGs Hessen	299	315.063	315.063
	Zahl Personen in BGs Rheinland-Pfalz	216	175.315	175.315
	Zahl Personen in BGs Baden-Württemberg	422	355.805	355.805
	Zahl Personen in BGs Bayern	528	358.390	358.390
	Zahl Personen in BGs Saarland	101	61.906	61.906
	Zahl Personen in BGs Berlin	376	450.129	450.129
	Zahl Personen in BGs Brandenburg	392	225.513	225.513
	Zahl Personen in BGs Mecklenburg-Vorpommern	164	171.727	171.727
	Zahl Personen in BGs Sachsen	468	379.470	379.470
	Zahl Personen in BGs Sachsen-Anhalt	392	256.044	256.044
Zahl Personen in BGs Thüringen	255	175.557	175.557	
Anzahl der Personen in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Alter (15-24 und 25-64; 2 Kategorien)	Zahl Personen in BGs im Alter 15-24	982	939.152	939.152
	Zahl Personen in BGs im Alter 25-64	5.317	4.062.960	4.062.960
Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Geschlecht und nach West/Ost (4 Kategorien)	Zahl Männer in BGs (West)	1.940	1.637.643	1.637.643
	Zahl Frauen in BGs (West)	2.312	1.706.029	1.706.029
	Zahl Männer in BGs (Ost)	988	854.217	854.217
	Zahl Frauen in BGs (Ost)	1.059	804.223	804.223
Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach „Alleinerziehend ja/nein“, nach Geschlecht und nach West/Ost (8 Kategorien)	Zahl nicht alleinerziehende in BGs (West)	3.477	2.892.913	2.892.913
	Zahl alleinerziehende in BGs (West)	775	450.759	450.759
	Zahl nicht alleinerziehende in BGs (Ost)	1.761	1.468.800	1.468.800
	Zahl alleinerziehende in BGs (Ost)	286	189.640	189.640
Anzahl der Personen ab 15 Jahre in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Staatsangehörigkeit (deutsch/nicht deutsch)	Zahl nicht deutscher Personen in BGs	737	990.860	990.860
	Zahl deutscher Personen in BGs	5.562	4.011.252	4.011.252

Tabelle 58: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen) (Fortsetzung 1)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Bundesland (16 Kategorien)	Zahl Personen in Privathaushalten Schleswig-Holstein	484	2.398.000	2.398.000
	Zahl Personen in Privathaushalten Hamburg	310	1.550.000	1.550.000
	Zahl Personen in Privathaushalten Niedersachsen	1.587	6.736.000	6.736.000
	Zahl Personen in Privathaushalten Bremen	129	577.000	577.000
	Zahl Personen in Privathaushalten Nordrhein-Westfalen	3.643	15.266.000	15.266.000
	Zahl Personen in Privathaushalten Hessen	949	5.201.000	5.201.000
	Zahl Personen in Privathaushalten Rheinland-Pfalz	628	3.437.000	3.437.000
	Zahl Personen in Privathaushalten Baden-Württemberg	1.397	9.138.000	9.138.000
	Zahl Personen in Privathaushalten Bayern	1.973	10.646.000	10.646.000
	Zahl Personen in Privathaushalten Saarland	227	885.000	885.000
	Zahl Personen in Privathaushalten Berlin	710	3.017.000	3.017.000
	Zahl Personen in Privathaushalten Brandenburg	750	2.207.000	2.207.000
	Zahl Personen in Privathaushalten Mecklenburg-Vorpommern	356	1.450.000	1.450.000
	Zahl Personen in Privathaushalten Sachsen	985	3.673.000	3.673.000
	Zahl Personen in Privathaushalten Sachsen-Anhalt	790	2.061.000	2.061.000
	Zahl Personen in Privathaushalten Thüringen	689	1.961.000	1.961.000
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Alter (in 5-Jahres-Klassen), Geschlecht und West/Ost (56 Kategorien)	Zahl Männer in PH (West), 15-19 Jahre	412	1.884.000	1.884.000
	Zahl Männer in PH (West), 20-24 Jahre	387	1.939.000	1.939.000
	Zahl Männer in PH (West), 25-29 Jahre	373	1.948.000	1.948.000
	Zahl Männer in PH (West), 30-34 Jahre	384	1.908.000	1.908.000
	Zahl Männer in PH (West), 35-39 Jahre	381	2.062.000	2.062.000
	Zahl Männer in PH (West), 40-44 Jahre	539	2.826.000	2.826.000
	Zahl Männer in PH (West), 45-49 Jahre	590	2.736.000	2.736.000

Tabelle 58: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen) (Fortsetzung 2)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Alter (in 5-Jahres-Klassen), Geschlecht und West/Ost (56 Kategorien)	Zahl Männer in PH (West), 50-54 Jahre	553	2.383.000	2.383.000
	Zahl Männer in PH (West), 55-59 Jahre	513	2.068.000	2.068.000
	Zahl Männer in PH (West), 60-64 Jahre	428	1.781.000	1.781.000
	Zahl Männer in PH (West), 65-69 Jahre	278	1.749.000	1.749.000
	Zahl Männer in PH (West), 70-74 Jahre	257	1.787.000	1.787.000
	Zahl Männer in PH (West), 75-79 Jahre	107	1.109.000	1.109.000
	Zahl Männer in PH (West), 80+ Jahre	84	1.048.000	1.048.000
	Zahl Frauen in PH (West), 15-19 Jahre	451	1.768.000	1.768.000
	Zahl Frauen in PH (West), 20-24 Jahre	415	1.896.000	1.896.000
	Zahl Frauen in PH (West), 25-29 Jahre	425	1.944.000	1.944.000
	Zahl Frauen in PH (West), 30-34 Jahre	503	1.916.000	1.916.000
	Zahl Frauen in PH (West), 35-39 Jahre	522	2.052.000	2.052.000
	Zahl Frauen in PH (West), 40-44 Jahre	633	2.722.000	2.722.000
	Zahl Frauen in PH (West), 45-49 Jahre	726	2.678.000	2.678.000
	Zahl Frauen in PH (West), 50-54 Jahre	638	2.426.000	2.426.000
	Zahl Frauen in PH (West), 55-59 Jahre	580	2.120.000	2.120.000
	Zahl Frauen in PH (West), 60-64 Jahre	447	1.841.000	1.841.000
	Zahl Frauen in PH (West), 65-69 Jahre	255	1.894.000	1.894.000
	Zahl Frauen in PH (West), 70-74 Jahre	237	2.037.000	2.037.000
	Zahl Frauen in PH (West), 75-79 Jahre	126	1.375.000	1.375.000
	Zahl Frauen in PH (West), 80+ Jahre	83	1.937.000	1.937.000
	Zahl Männer in PH (Ost), 15-19 Jahre	116	317.000	317.000
	Zahl Männer in PH (Ost), 20-24 Jahre	161	556.000	556.000
	Zahl Männer in PH (Ost), 25-29 Jahre	207	550.000	550.000
	Zahl Männer in PH (Ost), 30-34 Jahre	173	509.000	509.000
	Zahl Männer in PH (Ost), 35-39 Jahre	141	507.000	507.000
	Zahl Männer in PH (Ost), 40-44 Jahre	175	655.000	655.000

Tabelle 58: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen) (Fortsetzung 3)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Alter (in 5-Jahres-Klassen), Geschlecht und West/Ost (56 Kategorien)	Zahl Männer in PH (Ost), 45-49 Jahre	248	718.000	718.000
	Zahl Männer in PH (Ost), 50-54 Jahre	228	635.000	635.000
	Zahl Männer in PH (Ost), 55-59 Jahre	232	621.000	621.000
	Zahl Männer in PH (Ost), 60-64 Jahre	194	438.000	438.000
	Zahl Männer in PH (Ost), 65-69 Jahre	76	504.000	504.000
	Zahl Männer in PH (Ost), 70-74 Jahre	53	502.000	502.000
	Zahl Männer in PH (Ost), 75-79 Jahre	38	301.000	301.000
	Zahl Männer in PH (Ost), 80+ Jahre	17	237.000	237.000
	Zahl Frauen in PH (Ost), 15-19 Jahre	108	279.000	279.000
	Zahl Frauen in PH (Ost), 20-24 Jahre	164	497.000	497.000
	Zahl Frauen in PH (Ost), 25-29 Jahre	188	491.000	491.000
	Zahl Frauen in PH (Ost), 30-34 Jahre	187	461.000	461.000
	Zahl Frauen in PH (Ost), 35-39 Jahre	143	475.000	475.000
	Zahl Frauen in PH (Ost), 40-44 Jahre	205	624.000	624.000
	Zahl Frauen in PH (Ost), 45-49 Jahre	309	681.000	681.000
	Zahl Frauen in PH (Ost), 50-54 Jahre	255	642.000	642.000
	Zahl Frauen in PH (Ost), 55-59 Jahre	253	636.000	636.000
	Zahl Frauen in PH (Ost), 60-64 Jahre	181	454.000	454.000
	Zahl Frauen in PH (Ost), 65-69 Jahre	79	571.000	571.000
	Zahl Frauen in PH (Ost), 70-74 Jahre	77	604.000	604.000
Zahl Frauen in PH (Ost), 75-79 Jahre	41	404.000	404.000	
Zahl Frauen in PH (Ost), 80+ Jahre	31	500.000	500.000	
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Haushaltsgröße (1, 2, 3, 4, „5 oder mehr Personen“) und West/Ost (10 Kategorien)	Zahl Personen in PH mit 1 Person (West)	2.300	12.066.000	12.066.000
	Zahl Personen in PH mit 2 Personen (West)	3.741	20.714.000	20.714.000
	Zahl Personen in PH mit 3 Personen (West)	2.158	9.811.000	9.811.000
	Zahl Personen in PH mit 4 Personen (West)	1.951	9.128.000	9.128.000
	Zahl Personen in PH mit 5 oder mehr Personen (West)	1.177	4.115.000	4.115.000
	Zahl Personen in PH mit 1 Person (Ost)	1.139	3.651.000	3.651.000
	Zahl Personen in PH mit 2 Personen (Ost)	1.533	5.998.000	5.998.000
	Zahl Personen in PH mit 3 Personen (Ost)	903	2.721.000	2.721.000
	Zahl Personen in PH mit 4 Personen (Ost)	426	1.527.000	1.527.000
	Zahl Personen in PH mit 5 oder mehr Personen (Ost)	279	472.000	472.000

Tabelle 58: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen) (Fortsetzung 4)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahre in Privathaushalten nach höchstem Schulabschluss und West/Ost (12 Kategorien)	Zahl Personen in PH mit höchstem Schulabschluss: noch Schüler (West)	537	2.273.000	2.273.000
	Zahl Personen in PH mit höchstem Schulabschluss: ohne Abschluss (West)	628	2.336.000	2.336.000
	Zahl Personen in PH mit höchstem Schulabschluss: Hauptschule (West)	4.130	22.759.000	22.759.000
	Zahl Personen in PH mit höchstem Schulabschluss: Realschule/POS (West)	3.151	13.732.000	13.732.000
	Zahl Personen in PH mit höchstem Schulabschluss:(Fach-)Hochschulreife (West)	2.881	14.734.000	14.734.000
	Zahl Personen in PH mit höchstem Schulabschluss: noch Schüler (Ost)	111	343.000	343.000
	Zahl Personen in PH mit höchstem Schulabschluss: ohne Abschluss (Ost)	148	361.000	361.000
	Zahl Personen in PH mit höchstem Schulabschluss: Hauptschule (Ost)	912	3.270.000	3.270.000
	Zahl Personen in PH mit höchstem Schulabschluss: Realschule/POS (Ost)	2.249	6.752.000	6.752.000
	Zahl Personen in PH mit höchstem Schulabschluss: (Fach-) Hochschulreife (Ost)	860	3.643.000	3.643.000
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Familienstand und West/Ost (10 Kategorien)	Zahl Personen in PH mit Familienstand: ledig (West)	3.537	10.125.000	10.125.000
	Zahl Personen in PH mit Familienstand: verheiratet, eingetragene Lebensgemeinschaft. (West)	5.739	36.063.000	36.063.000
	Zahl Personen in PH mit Familienstand: geschieden (West)	1.592	4.883.000	4.883.000
	Zahl Personen in PH mit Familienstand: verwitwet (West)	459	4.763.000	4.763.000
	Zahl Personen in PH mit Familienstand: ledig (Ost)	1.585	3.351.000	3.351.000
	Zahl Personen in PH mit Familienstand: verheiratet, eingetragene Lebensgemeinschaft. (Ost)	1.854	8.264.000	8.264.000
	Zahl Personen in PH mit Familienstand: geschieden (Ost)	663	1.466.000	1.466.000
	Zahl Personen in PH mit Familienstand: verwitwet (Ost)	178	1.288.000	1.288.000

Tabelle 58: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen) (Fortsetzung 5)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2010	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahre in Privathaushalten nach Staatsangehörigkeit	Zahl Personen in PH Nicht-Deutsch	1.166	6.340.000	6.340.000
	Zahl Personen in PH Deutsch	14.441	63.863.000	63.863.000
Arbeitslose inkl. Maßnahmeteilnehmer West/Ost	Nicht arbeitslos West	8.918	53.032.681	53.032.681
	Arbeitslose inkl. Maßnahmeteilnehmer West	2.409	2.801.319	2.801.319
	Nicht arbeitslos Ost	2.996	13.084.328	13.084.328
	Arbeitslose inkl. Maßnahmeteilnehmer Ost	1.284	1.284.672	1.284.672
Sozialversicherungspflichtig beschäftigt West/Ost	Nicht-Sozialversicherungspflichtig beschäftigt West	7.325	32.861.296	32.861.296
	Sozialversicherungspflichtig beschäftigt West	4.002	22.972.704	22.972.704
	Nicht-Sozialversicherungspflichtig beschäftigt Ost	2.655	9.000.117	9.000.117
	Sozialversicherungspflichtig beschäftigt Ost	1.625	5.368.883	5.368.883

Tabelle 59: Kenngrößen der Verteilung der Gewichte

1%-Perzentil	39,67916
5%-Perzentil	94,32583
10%-Perzentil	141,2273
25%-Perzentil	353,4838
50%-Perzentil	1387,97
75%-Perzentil	5820,786
90%-Perzentil	13003,01
95%-Perzentil	19010,12
99%-Perzentil	31854,19
Mittelwert	4498,174
Standardabweichung	6954,562
Minimum	24,30088
Maximum	65767,37
Fallzahl	15607
Effektivitätsmaß	29,5%

6.14 Schätzung der BA-Querschnittsgewichte für Haushalte und Personen ohne Arbeitslosengeld-II-Bezug

Schließlich verblieben auch in der Welle 5 noch Haushalte und Personen, denen durch Kalibrierung kein BA-Haushaltsquerschnittsgewicht oder BA-Personenquerschnittsgewicht zugewiesen werden konnte. Die Anzahl dieser Haushalte ist nun in Welle 5 größer als in Welle 4, da ein größerer Teil der BA-Stichprobe der Wellen 1 bis 4 mittlerweile aus dem Leistungsbezug ausgeschieden ist. Es handelt sich um die folgenden drei Gruppen ohne Leistungsbezug im Juli 2010, die aber zur Grundgesamtheit der BA-Stichprobe (Haushalte mit Arbeitslosengeld-II-Bezug in 7/2006 oder 7/2007 oder 7/2008 oder 7/2009 oder 7/2010 bzw. Personen in Haushalten mit Arbeitslosengeld-II-Bezug in 7/2006 oder 7/2007 oder 7/2008 oder 7/2009 oder 7/2010) gehören.

- Aus der Auffrischungstichprobe: Nicht-BG-Mitglieder im Haushalt: Das Personengewicht wurde hier aus dem BA-Haushaltsgewicht in Welle 5 nach Kalibrierung (*wqbahh*) durch Division durch den Anteil dieser Personen, die – vorausgesetzt ihr Haushalt hat teilgenommen – ein Personen- oder Senioreninterview gegeben haben, gewonnen.
- Bestandshaushalte, in denen im Juli 2010 niemand mehr Arbeitslosengeld II bezogen hat: Der Haushalt behält das BA-Gewicht vor Kalibrierung. Personen in diesen Haushalten mit Interviews in beiden Wellen wurde als neues BA-Personengewicht ihr altes BA-Personengewicht der Vorwelle, multipliziert mit der reziproken Bleibewahrscheinlichkeit *ppbleib*, zugewiesen. Personen in diesen Haushalten ohne Personeninterview in Welle 4 erhalten als neues BA-Personengewicht das BA-Haushaltsgewicht ihres Haushalts für Welle 5 dividiert durch den Anteil solcher Personen, die teilnehmen, vorausgesetzt ihr Haushalt nimmt teil.
- Nicht-BG-Mitglieder in Bestandshaushalten, die im Juli 2010 noch Arbeitslosengeld II beziehen: Personen in diesen Haushalten mit Interviews in beiden Wellen wurde als neues BA-Personengewicht ihr BA-Personengewicht der Vorwelle, multipliziert mit der reziproken Bleibewahrscheinlichkeit *ppbleib*, zugewiesen.

Diese Personen und Haushalte wurden erstmals in Welle 5 zusätzlich an einen Eckwert für die Personen bzw. Bedarfsgemeinschaften, die den Grundsicherungsbezug verlassen haben, angepasst. Die exakte Populationsgröße dieser Gruppe ist dabei unbekannt, sie lässt sich aber aus der Summe aller kumulierten BA-Teilstichproben abzüglich der aktuell im Bezug befindlichen Personen bzw. Bedarfsgemeinschaften näherungsweise bestimmen. Die entsprechende Zahl der aus dem Grundsicherungsbezug abgegangenen Personen bemisst sich auf (3.185.162), die Zahl der aus dem Grundsicherungsbezug abgegangenen Bedarfsgemeinschaften auf (2.368.760).

7 Anhang: Kurzbeschreibung des Datensatzes

Inhaltliche Charakteristika

Kategorien	Erläuterungen
Themen/ Merkmalsgruppen	<p>Soziodemographische Merkmale: Systemfreie Personennummer; Geschlecht; Geburtsjahr; Alter; Familienstand; Kinderzahl in und außerhalb des Haushalts; Staatsangehörigkeit; Herkunftsland und Migrationshintergrund; Deutschkenntnisse und Sprache im Haushalt/Freundeskreis; Schul- und Berufsbildung (inkl. generierter Skalen: CASMIN, ISCED-97, Bildung in Jahren); Schul- und Berufsbildung der Eltern; Gesundheitsindikatoren; Konfession; soziale Kontakte; Freizeitaktivitäten von Jugendlichen; Betreuung und Schulbesuch von Kindern; Pflegetätigkeiten; Haushaltseinkommen (inkl. Einzelkomponenten und bedarfsgewichtetes Haushaltseinkommen); Basisinformationen zu Vermögen und Schulden; Ausstattung des Haushalts (Deprivationsindex); Wohnung und Wohnumfeld; detaillierte Informationen zum Thema Altersvorsorge (nur Welle 3);</p> <p>Beschäftigungsbezogene Merkmale: Erwerbsstatus/Status Nichterwerbstätigkeit; Mini-Job; Arbeitszeit; Stellung im Beruf (detailliert); berufliche Tätigkeit (ISCO-88 und KldB-92); ISCO-basierte Status- und Prestigemaße (ISEI, SIOPS, MPS, EGP, ESeC); Erwerbseinkommen (brutto & netto); Erwerbsbiografien mit Erwerbstätigkeiten/Arbeitslosigkeiten und Zeiten der Nichterwerbstätigkeit seit 01/2005 (ab Welle 2); Be- und Entfristung der Tätigkeit; Aufsichtsfunktion; Arbeitgeber: öfftl. Dienst/Privatwirtschaft; Arbeitgeber: Zahl der Beschäftigten; weitere Erwerbstätigkeit; Zusammenfassende Informationen zur Erwerbs- und Arbeitslosigkeitshistorie; detaillierte Informationen zum Thema Arbeitssuche; Reservationslohn; Vignettenmodul zur Stellenannahmefähigkeit (nur Welle 5);</p> <p>Merkmale zum Leistungsbezug: <u>ALG I:</u> Beginn und Ende der Leistungsbezugsepisode(n) seit 01/2005 (nur Welle 1); Informationen zu ALG-I-Bezugszeiten im Rahmen gemeldeter Arbeitslosigkeiten seit 01/2005 (ab Welle 2); Höhe der Leistung; Beendigungsgrund; <u>ALG II:</u> Beginn und Ende der Leistungsbezugsepisode(n) seit 01/2005; Bezugs- und Beendigungsgrund; Identifizierung beziehende HH-Mitglieder; Höhe der Leistungen; Leistungskürzungen (Anfang, Dauer, Gründe, HH-Mitglieder mit Kürzung); <u>Maßnahmeteilnahme (bis Welle 3):</u> Maßnahmetyp; Beginn und Ende der Maßnahme; Indikator vorzeitige Beendigung; Gründe für vorzeitige Beendigung; Art Maßnahmezugang; Maßnahmebewertung; Arbeitszeit in Maßnahme; Vergleich der Tätigkeit zu regulärer Beschäftigung; Branche; ab Welle 4 nur noch Ein-Euro-Job; <u>Ein-Euro-Job-Teilnahme (ab Welle 4):</u> Kenntnis, Gespräche über Ein-Euro-Jobs; Gründe für Nicht-Teilnahme; Beginn und Ende des Ein-Euro-Jobs; Gründe für vorzeitige Beendigung; Arbeitszeit; Vergleich der Tätigkeit zu regulärer Beschäftigung; Bewertung <u>Kontakte zum ALG-II-Träger:</u> Anzahl und Art der Kontakte; Gesprächsinhalte; Angebote; Eingliederungsvereinbarung; Bewertung des Trägers;</p>

Kategorien	Erläuterungen
Themen/ Merkmalsgruppen (Fortsetzung)	Subjektive Indikatoren: Zufriedenheit; Ängste und Sorgen; Erwerbsorientierungen; Bildungsaspiration; Geschlechterrollenorientierungen; subjektive soziale Position (Oben-Unten-Skala); subjektive Bewertung des Gesundheitszustands; Persönlichkeitsskala „Big Five“
Untersuchungseinheit	<p>Personen und Haushalte mit ALG II – Leistungsbezug in 7/2006 (Sample I)</p> <p>Personen und Haushalte der Wohnbevölkerung der BRD (Sample II)</p> <p>Personen und Haushalte mit ALG II – Leistungsbezug in 7/2007 aber ohne Bezug in 7/2006 (Sample III; Zugangsstichprobe 1)</p> <p>Personen und Haushalte mit ALG II – Leistungsbezug in 7/2008 aber ohne Bezug in 7/2006 oder 7/2007 (Sample IV; Zugangsstichprobe 2)</p> <p>Personen und Haushalte mit ALG II – Leistungsbezug in 7/2009 aber ohne Bezug in 7/2006, 7/2007 oder 7/2008 (Sample V; Zugangsstichprobe 3)</p> <p>Personen und Haushalte mit ALG II – Leistungsbezug in 7/2010 aber ohne Bezug in 7/2006, 7/2007, 7/2008 oder 7/2009 (Sample VIII; Zugangsstichprobe 4)</p> <p>Personen und Haushalte der Wohnbevölkerung der BRD (Sample VI, Bestandsauffrischung/ Aufstockungsstichprobe)</p> <p>Personen und Haushalte mit ALG II – Leistungsbezug in 7/2010 (Sample VII, Bestandsauffrischung/ Aufstockungsstichprobe)</p> <p>Anmerkung: Befragung von Personen ab 65 Jahre nur mittels eines Kurzfragebogens</p>

Kategorien	Erläuterungen
Fallzahlen	<p>Welle 1: Sample I: 9.386 Personen (in 6.804 Haushalten) Sample II: 9.568 Personen (in 5.990 Haushalten)</p> <p>Welle 2: Sample I: 4.753 Personen (in 3.491 Haushalten) Sample II: 6.392 Personen (in 3.897 Haushalten) Sample III: 1.342 Personen (in 1.041 Haushalten)</p> <p>Welle 3: Sample I: 4.913 Personen (in 3.754 Haushalten) Sample II: 6.207 Personen (in 3.901 Haushalten) Sample III: 898 Personen (in 694 Haushalten) Sample IV: 1.421 Personen (in 1.186 Haushalten)</p> <p>Welle 4: Sample I: 3.958 Personen (in 2.815 Haushalten) Sample II: 5.016 Personen (in 2.977 Haushalten) Sample III: 786 Personen (in 563 Haushalten) Sample IV: 983 Personen (in 745 Haushalten) Sample V: 1.025 Personen (in 748 Haushalten)</p> <p>Welle 5: Sample I: 3.394 Personen (in 2.382 Haushalten) Sample II: 4.511 Personen (in 2.680 Haushalten) Sample III: 653 Personen (in 464 Haushalten) Sample IV: 822 Personen (in 608 Haushalten) Sample V: 760 Personen (in 517 Haushalten) Sample VI: 2.589 Personen (in 1.510 Haushalten) Sample VII: 1.859 Personen (in 1.321 Haushalten) Sample VIII: 1.019 Personen (in 753 Haushalten)</p>

Kategorien	Erläuterungen
Erhebungsmodus	<p>CATI und CAPI</p> <p>CAPI Interviews wurden dann durchgeführt, wenn ein Stichprobenhaushalt nicht telefonisch erreichbar war oder persönliche Befragung gewünscht wurde</p> <p>Welle 1: N (CATI): 12.414 Personen (8.445 Haushalte) N (CAPI): 6.540 Personen (4.339 Haushalte)</p> <p>Welle 2: N (CATI): 7.888 Personen (5.378 Haushalte) N (CAPI): 4.599 Personen (3.051 Haushalte)</p> <p>Welle 3: N (CATI): 7.776 Personen (5.664 Haushalte) N (CAPI): 5.663 Personen (3.871 Haushalte)</p> <p>Welle 4: n (CATI): 6.913 Personen (4.669 Haushalte) n (CAPI): 4.855 Personen (3.179 Haushalte)</p> <p>Welle 5: n (CATI): 7.358 Personen (4.987 Haushalte) n (CAPI): 8.249 Personen (5.248 Haushalte)</p>
Erhebungssprachen	<p>Welle 1: Deutsch: 18.205 Personen (12.347 Haushalte) Russisch: 432 Personen (275 Haushalte) Türkisch: 305 Personen (163 Haushalte) Englisch: 12 Personen (9 Haushalte)</p> <p>Welle 2: Deutsch: 12.237 Personen (8.234 Haushalte) Russisch: 219 Personen (156 Haushalte) Türkisch: 31 Personen (39 Haushalte) Englisch: ab Welle 2 nicht mehr angeboten, da zu geringe Fallzahlen in Welle 1</p> <p>Welle 3: Deutsch: 13.000 Personen (9.256 Haushalte) Russisch: 330 Personen (210 Haushalte) Türkisch: 109 Personen (69 Haushalte)</p> <p>Welle 4: Deutsch: 11.405 Personen (7.627 Haushalte) Russisch: 285 Personen (179 Haushalte) Türkisch: 78 Personen (42 Haushalte)</p> <p>Welle 5: Deutsch: 15.290 Personen (10.040 Haushalte) Russisch: 259 Personen (159 Haushalte) Türkisch: 58 Personen (36 Haushalte)</p>

Kategorien	Erläuterungen
Response Rate	<p>Welle 1:</p> <p>Sample I: 35,1 % Sample II: 26,6 % Gesamt: 30,5 %</p> <p>Welle 2:</p> <p>Sample I (nur panelbereite HH): 51,1 % Sample II (nur panelbereite HH): 64,7 % Sample III: 26,3 % Abgespaltene Haushalte (aus Sample I und II): 13,4 % Gesamt: 45,0 %</p> <p>Welle 3:</p> <p>Sample I (nur panelbereite HH): 64,5 % Sample II (nur panelbereite HH): 76,4 % Sample III (nur panelbereite HH): 69,0 % Sample IV: 31,2 % Gesamt: 60,6 %</p> <p>Welle 4:</p> <p>Sample I (nur panelbereite HH): 72,1 % Sample II (nur panelbereite HH): 82,4 % Sample III (nur panelbereite HH): 65,6 % Sample IV (nur panelbereite HH): 68,2 % Sample V: 30,9 % Gesamt: 59,5 %</p> <p>Welle 5:</p> <p>Sample I (nur panelbereite HH): 71,1 % Sample II (nur panelbereite HH): 81,3 % Sample III (nur panelbereite HH): 69,2 % Sample IV (nur panelbereite HH): 63,7 % Sample V: (nur panelbereite HH): 71,5 % Sample VI: 24,5 % Sample VII: 24,5 % Sample VIII: 27,1 % Gesamt: 43,9 %</p>

Kategorien	Erläuterungen
Ausschöpfung innerhalb der Haushalte	<p>Welle 1: Sample I: 85,6 % Sample II: 84,3 % Gesamt: 85,0 %</p> <p>Welle 2: Sample I (nur wiederbefragte HH): 85,5 % Sample II (nur wiederbefragte HH): 85,1 % Sample III: 86,2 % Abgespaltene Haushalte (aus Sample I und II): 88,3 % Gesamt: 85,4 %</p> <p>Welle 3: Sample I (nur wiederbefragte HH): 83,1 % Sample I (nur wiederbefragte HH): 83,6 % Sample III (nur wiederbefragte HH): 84,3 % Sample IV: 84,2 % Abgespaltene Haushalte (aus Sample I-III): 84,2 % Gesamt: 83,5 %</p> <p>Welle 4: Sample I (nur wiederbefragte HH): 88,4 % Sample I (nur wiederbefragte HH): 88,0 % Sample III (nur wiederbefragte HH): 90,2 % Sample IV (nur wiederbefragte HH): 88,3 % Sample V: 89,6 % Abgespaltene Haushalte (aus Sample I-IV): 86,4 % Gesamt: 88,5 %</p> <p>Welle 5: Sample I (nur wiederbefragte HH): 88,7 % Sample I (nur wiederbefragte HH): 88,3 % Sample III (nur wiederbefragte HH): 89,5 % Sample IV (nur wiederbefragte HH): 89,3 % Sample V (nur wiederbefragte HH): 91,2 % Sample VI: 84,4 % Sample VII: 90,0 % Sample VIII: 88,9 % Abgespaltene Haushalte (aus Sample I-V): 89,9 % Gesamt: 88,3 %</p>
Feldzeit	<p>Welle 1: Dezember 2006-Juni 2007 Welle 2: Dezember 2007-Juli 2008 Welle 3: Dezember 2008-August 2009 Welle 4: Februar 2010-September 2010 Welle 5: Februar 2011-September 2011</p>

Kategorien	Erläuterungen
Zeitraum	<p>Welle 1: Feldzeit sowie Spelldaten retrospektiv ab 01/2005</p> <p>Welle 2: Feldzeit sowie Spelldaten retrospektiv ab 01/2005 bzw. jeweiligem Referenzzeitraum des Spelldatentyps</p> <p>Welle 3: Feldzeit sowie Spelldaten retrospektiv ab 01/2006 bzw. jeweiligem Referenzzeitraum des Spelldatentyps</p> <p>Welle 4: Feldzeit sowie Spelldaten retrospektiv ab 01/2008 bzw. jeweiligem Referenzzeitraum des Spelldatentyps</p> <p>Welle 5: Feldzeit sowie Spelldaten retrospektiv ab 01/2009 bzw. jeweiligem Referenzzeitraum des Spelldatentyps</p>
Zeitbezug	Wiederholungsbefragung (Haushaltspanel)
Regionale Gliederung	<p>Bundesland, Ost/West</p> <p>(Weitere regionale Informationen sind vorhanden, aber aus Datenschutzgründen nicht im Scientific Use File enthalten. Detaillierte Informationen hierzu auf Anfrage)</p>
Gebietsstand	Zum Befragungszeitpunkt

Methodische Charakteristika

Kategorien	Erläuterungen
Erhebungsdesign	<p>Ursprungsstichprobe Welle 1: Zweistufige Zufallsstichprobe mit zwei Teilpopulationen</p> <p>1. Stufe: Auswahl von 300 Postleitzahlbezirken als Primary Sampling Unit (PSU) für beide Teilstichproben. Die Ziehungswahrscheinlichkeit der einzelnen Postleitzahlbezirke hing dabei von der jeweiligen Größe (im Sinne der Einwohnerzahl) ab (probability proportional to size/pps).</p> <p>2. Stufe, Sample I: Ziehung von Bedarfsgemeinschaften aus den Registerdaten der Bundesagentur für Arbeit (BA) mit Stichtag Juli 2006. Die Zahl der pro PSU gezogenen Bruttostichprobe war abhängig von der ‚Größe‘ der PSU im Sinne des relativen Anteils der Leistungsempfänger im jeweiligen Postleitzahlbezirk (probability proportional to size/pps). Die durchschnittliche Größe der Bruttostichprobe betrug N=100 pro Postleitzahlbezirk.</p> <p>2. Stufe, Sample II: Bei Sample II wurde zunächst eine Gebäudestichprobe aus einer kommerziellen Datenbank (Micromosaic) gezogen. Diese wurde nach einem in der Datenbank enthaltenen Schichtindex im Verhältnis 4:2:1 für Haushalte mit niedrigem, mittleren bzw. hohem Status geschichtet. Die ausgewählten Gebäude wurden durch einen Interviewer des Erhebungsinstituts begangen. Handelte es sich um Gebäude mit mehreren Parteien, wurden diese notiert und anschließend wurde im Erhebungsinstitut ein zu befragender Haushalt ausgewählt. Die Bruttostichprobe umfasste dabei N=100 Haushalte pro Postleitzahlbezirk.</p> <p>Zugangsstichprobe BA in Welle 2 (Sample III): Zusätzlich zur Fortführung des für Welle 1 gezogenen Samples I wurde in der zweiten Welle eine Auffrischungsstichprobe aus den Registerdaten der Bundesagentur für Arbeit gezogen. Dabei wurden Bedarfsgemeinschaften ausgewählt, die im Juli 2007, aber nicht im Juli 2006 Arbeitslosengeld II bezogen haben, die also Neuzugänge in den Bezug darstellen. Die Ziehung erfolgte in für Welle 1 ausgewählten Postleitzahlbezirken analog zum Vorgehen in der ersten Welle.</p> <p>Zugangsstichprobe BA in Welle 3 (Sample IV): Auch in Welle 3 wurde wieder eine Zugangsstichprobe zum Sample I aus den Registerdaten der Bundesagentur für Arbeit gezogen. Dabei wurden Bedarfsgemeinschaften ausgewählt, die im Juli 2008, aber nicht im Juli 2006 oder Juli 2007 Arbeitslosengeld II bezogen haben, die also Neuzugänge in den Bezug darstellen. Die Ziehung erfolgte in für Welle 1 ausgewählten Postleitzahlbezirken analog zum Vorgehen in der ersten Welle.</p>

Zugangsstichprobe BA in Welle 4 (Sample V):

Auch in Welle 4 wurde wieder eine Zugangsstichprobe zum Sample I aus den Registerdaten der Bundesagentur für Arbeit gezogen. Dabei wurden Bedarfsgemeinschaften ausgewählt, die im Juli 2009, aber nicht im Juli 2006, Juli 2007 oder Juli 2008 Arbeitslosengeld II bezogen haben, die also Neuzugänge in den Bezug darstellen. Die Ziehung erfolgte in für Welle 1 ausgewählten Postleitzahlbezirken analog zum Vorgehen in der ersten Welle.

In Welle 5 wurde der Bestand der Ursprungsstichprobe durch zwei Aufstockungsstichproben aufgefrischt, anhand einer zweistufigen Zufallsstichprobe mit zwei Teilpopulationen

1. Stufe: Auswahl von 100 Postleitzahlbezirken als Primary Sampling Unit (PSU) für beide Teilstichproben. Die Ziehungswahrscheinlichkeit der einzelnen Postleitzahlbezirke hing dabei von der jeweiligen Größe (im Sinne der Einwohnerzahl) ab (probability proportional to size/pps).

2. Stufe, Sample VI: Bei Sample VI erfolgte die Ziehung aus den Registern der Einwohnermeldeämter. Dazu wurden die 100 Postleitzahlbereiche 96 Gemeinden zugeordnet. Die Ziehung der Personendressen aus der Auswahlgesamtheit in den Gemeinden erfolgte über eine systematische Zufallsauswahl (Intervallziehung). Die Ziehung der Adressen aus den Meldeämtern erfolgte aus den Jahrgängen 1992 und früheren Jahrgängen. In jedem Sample Point wurden 144 Adressen aus den Registern der Gemeinden gezogen.

2. Stufe, Sample VII: Ziehung von Bedarfsgemeinschaften aus den Registerdaten der Bundesagentur für Arbeit (Stichtag Juli 2010). Die Zahl der auszuwählenden Leistungsempfänger pro Point wurde als Produkt der festen Stichprobengröße (Personen pro Point) in der Bevölkerungsstichprobe mit dem Quotienten aus Leistungsbeziehendenquote im Point mit der Leistungsbeziehendenquote im Bund gewählt.

Zugangsstichprobe BA in Welle 5 (Sample VIII):

In Welle 5 wurde wieder eine Zugangsstichprobe zum Sample I aus den Registerdaten der Bundesagentur für Arbeit gezogen. Dabei wurden Bedarfsgemeinschaften ausgewählt, die im Juli 2010, aber nicht im Juli 2006, Juli 2007, Juli 2008 oder Juli 2009 Arbeitslosengeld II bezogen haben, die also Neuzugänge in den Bezug darstellen. Die Ziehung erfolgte in für Welle 1 ausgewählten Postleitzahlbezirken analog zum Vorgehen in der ersten Welle.

Kategorien	Erläuterungen
An der Erhebung beteiligte Institutionen	Institut für Arbeitsmarkt- und Berufsforschung; TNS Infratest Sozialforschung (Wellen 1 bis 3), infas Institut für angewandte Sozialwissenschaft GmbH (ab Welle 4)
Frequenz der Datensammlung	Jährlich (Panel)
Dateiformat/-größe	STATA, SPSS (mehrere Files)
Dateiorganisation	<p>Haushaltsdatensatz: HHENDDAT.dta/.sav</p> <p>Personendatensatz: <i>PENDDAT.dta/.sav</i></p> <p>Spelldaten ALG I: alg1_spells.dta/.sav (nur Welle 1)</p> <p>Spelldaten ALG II: alg2_spells.dta/.sav</p> <p>Spelldaten Arbeitslosigkeiten: al_spells.dta/.sav (Wellen 2 und 3)</p> <p>Spelldaten Erwerbstätigkeiten: et_spells.dta/.sav (Wellen 2 und 3)</p> <p>Spelldaten Lücken: lu_spells.dta/.sav (Wellen 2 und 3)</p> <p>ab Welle 4: Spelldaten Erwerbstätigkeiten, Arbeitslosigkeiten und Lücken integriert: bio_spells.dta/.sav</p> <p>Spelldaten Maßnahmen: mn_spells.dta/.sav (ab Welle 2)</p> <p>Spelldaten Maßnahmeteiln.: massnahmespells.dta/.sav (nur Welle 1)</p> <p>Registerdaten Haushaltsebene: hh_register.dta/.sav</p> <p>Registerdaten Personenebene: p_register.dta/.sav</p> <p>Gewichtungsdaten Haushaltsebene: hweights.dta/.sav</p> <p>Gewichtungsdaten Personenebene: pweights.dta/.sav</p> <p>Altersvorsorge Haushaltsebene: HAVDAT.dta/.sav (nur Welle 3)</p> <p>Altersvorsorge Personenebene: PAVDAT.dta/.sav (nur Welle 3)</p> <p>Vignettendaten: VIGDAT.dat/.sav (nur Welle 5)</p>

Datenzugang

Kategorien	Erläuterungen
Zugangswege	Scientific Use File (SUF)
Anonymisierungsgrad	faktisch anonymisiert
Sensible Merkmale	keine

Literatur

AAPOR (2006). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 4th Edition. Lenexa: AAPOR.

Achatz, J., Hirsland, A. & Promberger, M. (2007). Rahmenkonzept für das IAB-Panel „Arbeitsmarkt und Soziale Sicherung“. In M. Promberger (Hrsg.), *Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung „Arbeitsmarkt und Soziale Sicherung“*, IAB-Forschungsbericht 12/2007 (S. 11-32), Nürnberg.

Andreß H.-J., Burkatzki, E., Lipsmeier, G., Salentin, K., Schulte, K. & Strengmann-Kuhn, W. (1996). *Leben in Armut. Analysen der Verhaltensweisen armer Haushalte mit Umfragedaten*. Endbericht des DFG-Projekts "Versorgungsstrategien privater Haushalte im unteren Einkommensbereich (VuE)". Bielefeld.

Andreß, H.-J. & Lipsmeier, G. (1995). Was gehört zum notwendigen Lebensstandard und wer kann ihn sich leisten? Ein neues Konzept zur Armutsmessung. *Aus Politik und Zeitgeschichte* 31-32, S. 35-49.

Andreß, H.-J. & Lipsmeier, G. (2001). *Armut und Lebensstandard. Gutachten im Rahmen des Armuts- und Reichtumsberichts der Bundesregierung*. BMAS. Bonn.

Beckmann, P. & Trometer, R. (1991). Neue Dienstleistungen des ALLBUS: Haushalts- und Familientypologien, Klassenschema nach Goldthorpe. *ZUMA-Nachrichten* 28, S. 7-17.

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Marwinski, K., Gebhardt, D., Wenzig, C. & Wetzel, M. (2010). Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS). Bd. 1: Datenreport Welle 3. *FDZ Datenreport 06/2010*. Nürnberg.

Bethmann, A. & Gebhardt, D. (2011). User Guide "Panel Labour Market and Social Security" (PASS). Wave 3. *FDZ Datenreport 04/2011*. Nuremberg.

Brauns, H. & Steinmann, S. (1999). Educational Reform in France, West-Germany and the United Kingdom. Updating the Casmin Classification. *ZUMA-Nachrichten* 44. pp. 7-45.

Bundesministerium für Bildung und Forschung [BMBF] (2003). *Berufsausbildung sichtbar gemacht*. Schaubildsammlung. 4. Auflage. Bonn: BMBF.

Büngeler, K., Gensicke, M., Hartmann, J., Jäckle, R. & Tschersich, N. (2009). IAB-Haushaltspanel im Niedrigeinkommensbereich Welle 2 (2007/2008). Methoden- und Feldbericht. *FDZ-Methodenreport 08/2009*. Nürnberg.

Büngeler, K., Gensicke, M., Hartmann, J., Jäckle, R. & Tschersich, N. (2010): IAB-Haushaltspanel im Niedrigeinkommensbereich Welle 3 (2008/2009). Methoden- und Feldbericht. *FDZ-Methodenreport 10/2010*. Nürnberg.

Christoph, B. (2005). Zur Messung des Berufsprestiges: Aktualisierung der Magnitude-Prestigeskala auf die Berufsklassifikation ISCO88. *ZUMA-Nachrichten* 57. S. 79-127.

Europäische Gemeinschaften [EG] (2002). Verordnung (EG) Nr. 29/2002 der Kommission vom 19. Dezember 2001 zur Änderung der Verordnung (EWG) Nr. 3037/90 des Rates betreffend die statistische Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft. Amtsblatt der Europäischen Gemeinschaften L6/3-L6-33. Brüssel.

Erikson, R. & Goldthorpe, J. (1992). *The Constant Flux. A Study of Class Mobility in Industrial Society*. Oxford: Clarendon Press.

- Erikson, R., Goldthorpe, J. & Portocarero, L. (1979). Intergenerational Class Mobility in Three Western Societies: England, France and Sweden. *British Journal of Sociology* 30. pp. 415-441.
- Erikson, R., Goldthorpe, J. & Portocarero, L. (1982): Social Fluidity in Industrial Nations: England, France and Sweden. *British Journal of Sociology* 33. pp. 1-34.
- Fischer, A. & Wirth, H. (2007): *Constructing Version 4 of ESEC Classes from 3-digit ISCO (Stata-do file)*. Mannheim: Gesis-ZUMA.
- Frick, J., Göbel, J. & Krause, P. (n.d.). \$HGEN: Generated Household Level Variables. [<http://www.diw.de/documents/dokumentenarchiv/17/60053/hgen.pdf> (8.11.2007)].
- Ganzeboom, H. & Treiman, D. (1996). Internationally Comparable Measures for Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research* 25. pp. 201-239.
- Ganzeboom, H. & Treiman, D. (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. In H. Jürgen, P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables* (pp. 159-193), New York: Kluwer Academic / Plenum Publishers.
- Ganzeboom, H., De Graaf, P. & Treiman, D. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research* 21. pp. 1-56.
- Gebhardt, D., Müller, G., Bethmann, A., Trappmann, M., Christoph, B., Gayer, C., Müller, B., Tisch, A., Siflinger, B., Kiesl, H., Huyer-May, B., Achatz, J., Wenzig, C., Rudolph, H., Graf, T. & Biedermann, A. (2009). Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS). Datenreport Welle 2 (2007/2008). *FDZ Datenreport 06/2009*. Nürnberg.
- Gebhardt, D. & Trappmann, M. (2011): Using the datasets. In A. Bethmann & D. Gebhardt (Hrsg.), *User Guide "Panel Study Labour Market and Social Security" (PASS). Wave 3, FDZ Datenreport, 04/2011* (pp. 62-107), Nürnberg.
- Granato, N. (2000). Mikrodaten-Tools: CASMIN-Bildungsklassifikation. Eine Umsetzung mit dem Mikrozensus 1996. *ZUMA-Technischer Bericht 2000/12*. Mannheim.
- Hagenaars, A., de Vos, K. & Zaidi, M. (1994). *Poverty Statistics in the Late 1980s: Research Based on Micro-data*. Luxembourg: Office for Official Publications of the European Communities.
- Halleröd, B. (1995). The Truly Poor: Direct and Indirect Consensual Measurement of Poverty in Sweden. *Journal of European Social Policy* 5(2). pp. 111-129.
- Harrison, E. & Rose, R. (2006). ESeC User Guide, Appendix 6 (SPSS-Syntax: Esec Full) [<http://www.iser.essex.ac.uk/esec/guide/docs/Appendix6.sps> (31.11.2007)]
- Hartmann, J., Brink, K., Jäckle, R. & Tschersich, N. (2008). IAB-Haushaltspanel im Niedrigeinkommensbereich. Methoden- und Feldbericht. *FDZ Methodenreport 07/2008*. Nürnberg.
- Hauser, R. (1996). Zur Messung individueller Wohlfahrt und Ihrer Verteilung. In Statistisches Bundesamt (Hrsg.), *Wohlfahrtsmessung. Aufgabe der Statistik im gesellschaftlichen Wandel* (S. 13-38), Stuttgart: Metzler-Poeschel.

Helberger, C. (1988). Eine Überprüfung der Linearitätsannahme der Humankapitaltheorie. In H.-J. Bodenhöfer (Hrsg.), *Bildung, Beruf, Arbeitsmarkt* (S. 151-170), Berlin: Duncker & Humblot.

International Labour Office [ILO] (1990). *International Standard Classification of Occupations. ISCO-88*. Geneva: International Labour Office.

Jäckle, A. (2008). The Causes of Seam Effects in Panel Surveys. *ISEP Working Paper Series 2008-14*. Essex.

Jesske, B. & Quandt, S. (2011). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 4. Erhebungswelle – 2010 (Haupterhebung). *FDZ-Methodenreport 08/2011*. Nürnberg.

Jesske, B. & Schulz, S. (im Erscheinen 2012). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 5. Erhebungswelle – 2011 (Haupterhebung), *FDZ Methodenreport (im Erscheinen)*, Nürnberg.

König, W., Lüttinger, P. & Müller, W. (1987). *Eine vergleichende Analyse der Entwicklung und Struktur von Bildungssystemen. Methodologische Grundlagen und Konstruktion einer vergleichbaren Bildungsskala*. CASMIN-Projekt. Arbeitspapier Nr. 12. Mannheim.

Lechert, Y., Schroedter, J. & Lüttinger, P. (2006). Die Umsetzung der Bildungsklassifikation CASMIN für die Volkszählung 1970, die Mikrozensus- Zusatzerhebung 1971 und die Mikrozensus 1976-2004. *ZUMA-Methodenbericht 2006/12*. Mannheim.

Lengerer, A., Bohr, J. & Janßen, A. (2005). Haushalte, Familien und Lebensformen im Mikrozensus – Konzepte und Typisierungen. *ZUMA-Arbeitsbericht 2005/05*. Mannheim.

Lipsmeier, G. (1999). Die Bestimmung des notwendigen Lebensstandards – Einschätzungsunterschiede und Entscheidungsprobleme. *Zeitschrift für Soziologie* 28(4). S. 281-300.

Müller, W., Wirth, H., Bauer, G., Pollak, R. & Weiss, F. (2006). ESeC – Kurzbericht zur Validierung und Operationalisierung einer europäischen sozioökonomischen Klassifikation. *ZUMA-Nachrichten* 59. S. 111–119.

Müller, W., Wirth, H., Bauer, G., Pollak, R. & Weiss, F. (2007): Entwicklung einer europäischen sozioökonomischen Klassifikation. *Wirtschaft und Statistik* 5/2007. S. 527-530.

Nolan, B. & Whelan, C. (1996). Measuring Poverty Using Income and Deprivation Indicators: Alternative Approaches. *Journal of European Social Policy* 6(3). pp. 225-240.

Organisation for Economic Co-Operation and Development [OECD] (Ed.) (1999). *Classifying Educational Programmes. Manual for ISCED-97 Implementation in OECD Countries*. 1999 Edition. Paris: OECD.

Organisation for Economic Co-Operation and Development [OECD] (Ed.) (1982): *The OECD List of Social Indicators*. Paris: OECD.

Porst, R. (1984). Haushalt und Familien 1982. Zur Erfassung und Beschreibung von Haushalts- und Familienstrukturen mit Hilfe repräsentativer Bevölkerungsumfragen. *Zeitschrift für Soziologie* 13(2). S. 164-175.

Rammstedt, B. & John, O. (2005). Kurzversion des Big Five Inventory (BIF-K). *Diagnostica* 51(4). S. 195-206.

Rendtel, U. & Harms, T. (2009). Weighting and calibration for household panels. In P. Lynn (Ed.), *Methodology of Longitudinal Surveys* (pp. 265-286), Chichester: Wiley.

Ringen, S. (1988). Direct and Indirect Measurement of Poverty. *Journal of Social Policy* 17(3). pp. 351-365.

Rose, R. & Harrison, E. (2007). The European Socio-Economic Classification: A New Social Class Schema for Comparative European Research. *European Societies* 9(3). pp. 459-490.

Sozialgesetzbuch Zweites Buch [SGB II]: Grundsicherung für Arbeitssuchende.

Spieß, M. & Rendtel, U. (2000). Combining an ongoing panel with a new cross-sectional sample. *DIW-Discussion Papers* 198. Berlin.

Stata Corp (2007). Survey Data Reference Manual: Release 10. College Station: Stata Press.

Statistisches Bundesamt [StBA] (1992). Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt [StBA] (2002). Klassifikation der Wirtschaftszweige, Edition 2003 (WZ2003). Wiesbaden: Statistisches Bundesamt.

Rudolph, H. & Trappmann, M. (2007). Design und Stichprobe des Panels „Arbeitsmarkt und Soziale Sicherung“ (PASS). In M. Promberger (Hrsg.), Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung „Arbeitsmarkt und Soziale Sicherung“, *IAB-Forschungsbericht* 12/2007 (S. 60-101), Nürnberg.

Trappmann, M., Christoph, B., Achatz, J., Wenzig, C., Müller, G. & Gebhardt, D. (2009). Design and stratification of PASS. A New Panel Study for Research on Long Term Unemployment. *IAB-Discussion Paper* 5/2009. Nuremberg.

Trappmann, M. (2011). Weighting. In A. Bethmann & D. Gebhardt (Hrsg.), User Guide "Panel Study Labour Market and Social Security" (PASS). Wave 3, *FDZ Datenreport*, 04/2011 (pp. 51-61), Nürnberg.

Treiman, D. (1977). *Occupational Prestige in Comparative Perspective*. New York: Academic Press.

Wegener, B. (1985): Gibt es Sozialprestige? *Zeitschrift für Soziologie* 14. S. 209-235.

Wegener, B. (1988): *Kritik des Prestiges*. Opladen: Westdeutscher Verlag.

Impressum

FDZ-Datenreport 6/2012

Herausgeber

Forschungsdatenzentrum (FDZ)
der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt- und Berufsforschung
Regensburger Str. 104
90478 Nürnberg

Redaktion

Stefan Bender, Dagmar Theune

Technische Herstellung

Dagmar Theune

Rechte

Nachdruck - auch auszugsweise - nur mit
Genehmigung des FDZ gestattet

Bezugsmöglichkeit

http://doku.iab.de/fdz/reporte/2012/DR_06-12_I.pdf

Internet

<http://fdz.iab.de/>

Rückfragen zum Inhalt an:

Forschungsdatenzentrum (FDZ) der
Bundesagentur für Arbeit im Institut für
Arbeitsmarkt- und Berufsforschung (IAB),
Regensburger Str. 104, 90478 Nürnberg,
E-Mail: iab.fdz@iab.de

Arne Bethmann, Institut für Arbeitsmarkt- und
Berufsforschung (IAB),
Regensburger Str. 104, 90478 Nürnberg,
Tel.: +49 (0) 911/179-2307
E-Mail: arne.bethmann@iab.de