

FDZ-Datenreport

08/2011

DE

Dokumentation zu Arbeitsmarktdaten

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

Band I: Datenreport Welle 4

Aktualisierte Version \$3.0- .2012

Marco Berg,
Ralph Cramer,
Christian Dickmann,
Reiner Gilberg,
Birgit Jesske,
Martin Kleudgen,
Arne Bethmann,
Benjamin Fuchs,
Daniel Gebhardt



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

Datenreport Welle 4

Marco Berg, Ralph Cramer, Christian Dickmann, Reiner Gilberg, Birgit Jesske, Martin Kleudgen, infas Institut für angewandte Sozialwissenschaft GmbH - Arne Bethmann, Benjamin Fuchs, Daniel Gebhardt, 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 4. Welle von PASS. Das Dokument basiert auf dem Datenreport der 3. Welle: Marco Berg, Ralph Cramer, Christian Dickmann, Reiner Gilberg, Birgit Jesske, Karen Marwinski, (alle infas Institut für angewandte Sozialwissenschaft GmbH), Daniel Gebhardt, Claudia Wenzig, Martin Wetzel (alle Institut für Arbeitsmarkt- und Berufsforschung (IAB)): Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS) Welle 3 (2008/2009), FDZ Datenreport, 06/2010 (de), Nürnberg.

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 fourth PASS wave and is based upon the third wave's data report: Marco Berg, Ralph Cramer, Christian Dickmann, Reiner Gilberg, Birgit Jesske, Karen Marwinski, (alle infas Institut für angewandte Sozialwissenschaft GmbH), Daniel Gebhardt, Claudia Wenzig, Martin Wetzel (alle Institut für Arbeitsmarkt- und Berufsforschung (IAB)): Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS) Welle 3 (2008/2009), FDZ Datenreport, 06/2010 (de), Nürnberg.

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 4. Welle	10
2	Kennwerte	15
2.1	Stichprobenumfang	15
2.2	Response-Raten	19
2.3	Panelbereitschaft, Zuspielbereitschaft, Verknüpfung mit Prozessdaten	21
2.4	Abgespaltene Haushalte	23
3	Datensatzstruktur	23
4	Generierte Variablen	25
4.1	Vercodung offener Angaben	25
4.2	Harmonisierung	28
4.3	Dependent Interviewing	29
4.4	Einfache Generierungen	32
4.5	Konstruktvariablen	57
5	Datenaufbereitung	79
5.1	Strukturprüfungen und entfernte Interviews	81
5.2	Filterbereinigung	87
5.3	Plausibilitätsprüfungen	88
5.4	Rückwirkende Veränderungen der Wellen 1 bis 3	90
5.5	Anonymisierung	114
5.6	Arbeitslosengeld-II-Bezüge	118
5.7	Erwerbsbiografien	123
5.8	Ein-Euro-Job-Spelldatensatz (ee_spells)	127
6	Gewichtung Welle 4	128
6.1	Designgewichte für die Bestandshaushalte der 4. Welle	128
6.2	Designgewichte für die Zugangsstichprobe der 4. Welle	129
6.3	Wiederteilnahme-Propensity Haushalte	129
6.4	Teilnahme-Propensity erstbefragte Split-Haushalte	133
6.5	Ausfallgewichtung für Haushalte aus Auffrischungsstichprobe der 4. Welle	135
6.6	Wiederteilnahme-Propensity Personen	136
6.7	Integration der Gewichte zum Gesamtgewicht vor Kalibrierung	141
6.8	Integration temporärer Ausfälle (Haushalte)	141
6.9	Kalibrierung zum Haushaltsgewicht 4. Welle Querschnitt	145
6.10	Kalibrierung zum Personengewicht 4. Welle Querschnitt	162
6.11	Schätzung der BA-Querschnittgewichte für Haushalte und Personen ohne Arbeitslosengeld-II-Bezug	181
7	Anhang: Kurzbeschreibung des Datensatzes	182

Tabellen- und Abbildungsverzeichnis

Tabelle 1:	Panelstichprobenumfang auf Haushaltsebene nach Wellen und Teilstichproben	17
Tabelle 2:	Panelstichprobenumfang auf Personenebene nach Wellen und Teilstichproben	17
Tabelle 3:	Panelstichprobenumfang fremdsprachige Interviews nach Wellen.....	18
Abbildung 1:	Realisierte Panelstichprobe Haushalte und Personen nach Erhebungswellen	18
Tabelle 4:	Response-Rate der 4. Welle auf Haushaltsebene nach Teilstichproben	19
Tabelle 5:	Durchschnittliche Realisierungsrate innerhalb der befragten Haushalte nach Wellen und Teilstichproben	20
Tabelle 6:	Anteil Personeninterviews in Wellen 2, 3 und 4 mit panelbereiten Befragten der Vorwelle nach Teilstichproben.....	20
Tabelle 7:	Panelbereitschaft bei erstbefragten Haushalten** nach Wellen.....	21
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.....	22
Abbildung 2:	Datensatzstruktur des PASS in Welle 4	25
Tabelle 9:	Vercodung offener Angabe auf Haushaltsebene in Welle 4	26
Tabelle 10:	Vercodung offener Angaben auf Personenebene in Welle 4	26
Tabelle 11:	Harmonisierte Variablen im Personendatensatz (<i>PENDDAT</i>)	28
Tabelle 12:	Wellenübergreifend generierte, aber nicht vollständig harmonisierte Variablen im Personendatensatz (<i>PENDDAT</i>)	29
Tabelle 13a:	Fortgeschriebene Vorwelleninformationen in Welle 4, Haushaltsfragebogen.....	31
Tabelle 13b	Fortgeschriebene Vorwelleninformationen in Welle 4, Personenfragebogen	32
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	33
Tabelle 15:	Einfache Generierungen für Welle 4 im Haushaltsdatensatz (<i>HHENDDAT</i>) (alphabetische Reihenfolge)	35
Tabelle 16:	Einfache Generierungen für Welle 4 im Personendatensatz (<i>PENDDAT</i>) (alphabetische Reihenfolge)	37
Tabelle 17:	Einfache Generierungen für Welle 4 im Arbeitslosengeld-II-Spellldatensatz (<i>alg2_spells</i>) (Reihenfolge analog zur Reihenfolge im Datensatz)	47
Tabelle 18:	Einfache Generierungen für Welle 4 im BIO-Spellldatensatz (<i>bio_spells</i>) (Reihenfolge analog zur Reihenfolge im Datensatz)	50
Tabelle 19:	Einfache Generierungen für Welle 4 im Ein-Euro-Spellldatensatz (<i>ee_spells</i>) (Reihenfolge analog zur Reihenfolge im Datensatz)	53
Tabelle 20:	Einfache Generierungen für Welle 4 im Personenregisterdatensatz (<i>p_register</i>) (alphabetische Reihenfolge)	54
Abbildung 3:	Übersicht über generierte Variablen auf Personenebene in Welle 4	56

Tabelle 21:	Übersicht über die Schritte der Datenaufbereitung in der 4. Welle des PASS.....	81
Tabelle 22:	Übersicht über die vergebenen Missing Codes.....	88
Tabelle 23:	Generierungsgrundlagen von akt1euro.....	92
Tabelle 24:	Generierungslogik <i>erwerb</i> , <i>erwerb2</i> , <i>nichterw</i> , <i>nichtew2</i>	93
Tabelle 25:	Überarbeitung <i>erwerb</i> , <i>erwerb2</i> , <i>nichterw</i> , <i>nichtew2</i>	97
Tabelle 26:	Grobzuweisung –Spell höherer Priorität schlägt Spell niedrigerer Priorität.....	100
Tabelle 27:	Feinzuweisung für Sonderfälle.....	100
Tabelle 28:	Überarbeitung Einkommensvariablen.....	101
Tabelle 29:	Überarbeitung Arbeitszeitvariablen.....	103
Tabelle 30:	Bildung von <i>spelltyp</i>	109
Tabelle 31:	Bildung von <i>spintegr</i>	109
Tabelle 32:	Übersicht über rückwirkende Veränderungen im Haushaltsdatensatz (<i>HHENDDAT</i>).....	110
Tabelle 33:	Übersicht über rückwirkende Veränderungen im Personendatensatz (<i>PENDDAT</i>).....	112
Tabelle 34:	Übersicht über rückwirkende Korrekturen in Spelldatensätzen als Basis für den integrierten Bio-Spelldatensatz (<i>bio_spells</i>).....	113
Tabelle 35:	Übersicht über rückwirkende Veränderungen in Registerdatensätzen (<i>hh_register</i> ; <i>p_register</i>).....	113
Tabelle 36:	Übersicht über rückwirkende Veränderungen in den Gewichtungsdatsätzen (<i>hweights</i> ; <i>pweights</i>).....	114
Tabelle 37:	Übersicht über die in Welle 4 anonymisierten Variablen im Personendatensatz (<i>PENDDAT</i>).....	115
Tabelle 38:	Übersicht über die in Welle 4 anonymisierten Variablen im BIO-Spelldatensatz (<i>bio_spells</i>).....	118
Tabelle 39:	Querschnittvariablen im ALG2-Spelldatensatz (<i>alg2_spells</i>).....	119
Tabelle 40:	ET-spezifische Querschnittvariablen im BIO-Spelldatensatz (<i>bio_spells</i>).....	124
Tabelle 41:	Querschnittvariablen im AL-Spelldatensatz (<i>al_spells</i>).....	125
Tabelle 42:	Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der Wiederholerhaushalte.....	130
Tabelle 43:	Logit-Modelle zur Wiederteilnahme für Panelbereitschaft, Erreichbarkeit und Teilnahme.....	131
Tabelle 44:	Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der erstbefragten Split-Haushalte (W3 und W4).....	133
Tabelle 45:	Logit-Modelle zur Erstteilnahme von Split-W3-Haushalten für Erreichbarkeit und Teilnahme.....	134
Tabelle 46:	Logit-Modelle zur Erstteilnahme von Split-W4-Haushalten für Erreichbarkeit und Teilnahme.....	134
Tabelle 47:	Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der Auffrischungsstichprobe Welle 4.....	135
Tabelle 48:	Logit-Modelle zur Erstteilnahme für Erreichbarkeit und Teilnahme der Auffrischungsstichprobe Welle 4.....	136
Tabelle 49:	Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der Wiederholer.....	137

Tabelle 50:	Logit-Modelle zur Wiederteilnahme für Erreichbarkeit und Teilnahme.....	139
Tabelle 51:	Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der temporären Ausfälle	142
Tabelle 52:	Logit-Modelle temporärer Ausfälle	144
Tabelle 53:	Soll-Verteilungen und Verteilungen nach Kalibrierung (BA- Stichprobe, Haushalte)	148
Tabelle 54:	Kenngößen der Verteilung der Gewichte	150
Tabelle 55:	Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm- Stichprobe, Haushalte)	151
Tabelle 56:	Kenngößen der Verteilung der Gewichte	156
Tabelle 57:	Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Haushalte)	157
Tabelle 58:	Kenngößen der Verteilung der Gewichte	162
Tabelle 59:	Soll-Verteilungen und Verteilungen nach Kalibrierung (BA- Stichprobe, Personen)	164
Tabelle 60:	Kenngößen der Verteilung der Gewichte	165
Tabelle 61:	Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm- Stichprobe, Personen)	167
Tabelle 62:	Kenngößen der Verteilung der Gewichte	172
Tabelle 63:	Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen)	175
Tabelle 64:	Kenngößen der Verteilung der Gewichte	180

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 Armutsforschung 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 vierte Befragungswelle, für die zwischen Februar 2010 und September 2010 11.768 Personen in 7.848 Haushalten² befragt wurden. Darunter befanden sich 10.049 Personen und 6.945 Haushalte mit denen bereits zum wiederholten Mal ein Interview im Rahmen von PASS geführt³.

Der vorliegende wellenspezifische Datenreport⁴ der 4. Welle dokumentiert die wellenbezogenen Aspekte der Studie. Nach einem Kurzüberblick zu den Neuerungen und Besonderhei-

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 Nach der 3. Welle erfolgte ein Wechsel des Erhebungsinstituts. Die Durchführung der Befragungen sowie die Aufbereitung der erhobenen Daten wurde vom infas Institut für angewandte Sozialwissenschaften, Bonn, vorgenommen.

⁴ Mit der Dokumentation der 3. Welle wurde der Report erstmalig in zwei Teile aufgeteilt: einem wellenspezifischen Datenreport (inkl. Codebuch) und einem wellenübergreifenden User Guide. Der wellenübergreifende User Guide wird in Verantwortung des PASS-Projektteams beim IAB erstellt. Die Dokumentation des wellenspezifischen Datenreports ab der 3. Welle 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

ten der 4. Welle (Kapitel 1.3) werden die zentralen Kennzahlen zu Stichprobe und Ausschöpfungsquoten der 4. Welle 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 4 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 4. Welle basieren auf der 3. Welle des PASS, wobei jedoch Einzelfragen und -module überarbeitet oder neu entwickelt wurden (siehe User Guide PASS Wave 3, FDZ-Datenreport 04/2011) sowie für einen Überblick Kapitel 1.3)⁶.

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 ermöglichen⁷. Um bei der erneuten Befragung Seam-Effekte⁸ zu vermeiden und die Datenqualität zu erhöhen, wurden bereits seit der 3. Welle bei bestimmten Fragen per dependent interviewing Informationen, die der Befragte

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).

⁶ Durch den Wechsel des Erhebungsinstituts war eine Neuprogrammierung der Erhebungsinstrumente erforderlich. Dabei wurden in Einzelfällen auch inhaltlich unveränderte Fragen bzw. Module im Hinblick auf den Fragenablauf optimiert. Dies betraf insbesondere in der Kontaktierungsphase das Intro, die Haushalts- und Beziehungsmatrix zur Erfassung sämtlicher Mitglieder im Haushalt, sowie im Personenfragebogen die Abfrageschleifen im Erwerbsmodul.

⁷ Zu den erstmals befragten Haushalten gehören: (1) Haushalte der Zugangsstichprobe aktuellen Welle sowie (2) Haushalte, die sich von bereits in den Vorwellen befragten Haushalten abgespalten haben (Splithaushalte) (weitere Erläuterungen vgl. Methodenbericht der Welle 4 (Jesske & Quandt, 2011)).

⁸ 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).

beim letzten Interview gegeben hatte, fortgeschrieben. Außerdem wurde weiterhin auf die erneute Erhebung unveränderlicher Merkmale weitgehend verzichtet. Auf Haushaltsebene existiert ab der 4. Welle abweichend zum bisherigen Vorgehen ein integrierter Fragebogen für wiederbefragte Haushalte (HHalt) und für erstmals befragte Haushalte (HHneu)⁹.

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 4. Welle.

1.3 Besonderheiten und Neuerungen der 4. Welle

Nutzern, die bereits mit den Daten der Panelwellen gearbeitet haben, möchten wir an dieser Stelle einen schnellen Überblick über Besonderheiten der 4. PASS-Welle geben.

Die Besonderheiten und Neuerungen in Welle 4 betreffen das Frageprogramm¹⁰ (Kontaktierungsphase, Fortschreibung der in der Vorwelle erhobenen erwerbsbiografischen Informationen, chronologische Erhebung von Aktivitäten bzw. Episoden im Erwerbsmodul, Einsatz eines neuen Schwerpunktmoduls im Bereich „Kinderbetreuung“, sowie Kürzungen in bestehenden Modulen), die Stichprobe, die Aufbereitung der Daten und die Dokumentation.

1.3.1 Personenfragebogen

Im Personenfragebogen werden die in Welle 2 bzw. Welle 3 erhobenen erwerbsbiografischen Informationen fortgeschrieben¹¹. Dabei wurde jedoch das bisherige Konzept der modularen Abfrage getrennt nach Erwerbstätigkeitsepisoden, Arbeitslosigkeiten und anderen Aktivitäten (z.B. berufliche Ausbildung, Hausfrau/Hausmann, Rentner) durch eine chronologische Restrospektiverhebung ersetzt. Die Befragten sollten alle Aktivitäten bzw. Episoden chronologisch vorwärts berichten. Für jede Episode wurde sowohl das Anfangs- als auch das Enddatum erhoben. Aufsatzpunkt für den Start der Chronologie bildete dabei für wiederholt befragte Personen das Datum des letzten Interviews; für Neubefragte der Januar 2008. Ein Episodenwechsel wurde geknüpft an:

- a) einen Wechsel des Erwerbsstatus
- b) einen Tätigkeitswechsel
- c) eine Tätigkeitsunterbrechung
- d) einen Arbeitgeberwechsel.

⁹ Dabei werden Split-Haushalte wie neue Haushalte für die Abfrage behandelt.

¹⁰ Kleinere Änderungen im Frageprogramm (Neuaufnahme, Umformulierung bzw. Löschen von Einzelfragen) werden in diesem Rahmen nicht aufgeführt, sowie alle Änderungen und Neuerungen der Kontaktierungsphase (vgl. dazu Methodenbericht der Welle 4 (Jesske & Quandt, 2011)).

¹¹ 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.

Die Befragten wurden wiederholt auf die Abgrenzungsdefinition bei ihren Aktivitäten hingewiesen. Die wiederholte Nachfrage weiterer Episoden erfolgte bis hin zu jener Aktivität, die zum aktuellen Interviewzeitpunkt andauerte.

Mit dem neuen Konzept wurden explizit parallele Aktivitäten bzw. Episoden zugelassen. Im Fragenprogramm hinterlegt wurde eine Prüfung, die zeitliche Lücken zwischen den berichteten Episoden von mehr als einem Monat als Fehlermeldung anzeigte. Im Falle einer Fehlermeldung war der Interviewer aufgefordert, diese Unplausibilität unmittelbar im Interviewgespräch zu klären und zu korrigieren. Darüber hinaus wurde am Ende des Erwerbsmoduls für den Interviewer eine chronologische Übersicht aller aufgenommenen Aktivitäten eingeblendet, die vom Befragten noch einmal bestätigt werden sollte.

Innerhalb der Abfrage zu den Episoden wurden die einzelnen Fragen beibehalten und nach der Logik der Fortschreibung wie in Welle 3 erhoben. D.h. für fortgeschriebene Erwerbstätigkeiten werden die Informationen:

- (1) berufliche Stellung¹² und
- (2) Arbeitszeit¹³ sowie,
- (3) die Information ob vormals befristete Tätigkeiten in unbefristete umgewandelt wurden und
- (4) die Information darüber, wie das Arbeitsverhältnis beendet wurde

erneut erhoben. Neu in Welle 4 war die Erhebung des monatlichen Arbeitseinkommens. Bis Welle 3 erfolgte lediglich eine summarische Erhebung des Brutto- und des Nettoeinkommens über sämtliche aktuell zum Interviewzeitpunkt ausgeübten Tätigkeiten. Ab Welle 4 erfolgte zusätzlich zur summarischen Erhebung für aktuelle Erwerbstätigkeiten die episodenspezifische Erhebung des Bruttoeinkommens für jede bereits abgeschlossene Tätigkeit.

Im Falle von Arbeitslosigkeitsepisoden wurde nachgefragt, ob diese Phase noch andauerte sowie ggf. die Beendigungsgründe. Für den Zeitraum seit der letzten Befragung wurde weiterhin erhoben, ob und wenn ja, wie lange Arbeitslosengeld I bezogen wurde. Bei Befragten, die zum Interviewzeitpunkt der vierten Welle Arbeitslosengeld I beziehen, wird zusätzlich die Höhe des Betrags erfragt.

Das neue integrierte chronologische Abfragekonzept hatte auch Konsequenzen für den Scientific Use File. Ab der Welle 4 werden die Spelldatensätze zu Erwerbstätigkeit, Arbeitslosigkeit und Lücke in einem Spelldatensatz (*bio_spell*) zusammengefasst. Dieser Gesamt-Spelldatensatz enthält sowohl die in den Wellen 2 und 3 als auch die in Welle 4 erfassten Erwerbstätigkeits-, Arbeitslosigkeits- und Nichterwerbstätigkeitsepisoden. Die Integration aller Episoden in die jeweiligen Spelldatensätze erfolgt nach bestimmten Regeln (siehe dazu Kapitel 5.6, 5.7, 5.8). Werden Episoden über mehrere Wellen hinweg fortgeschrieben, kann

¹² Bei abgeschlossenen fortgeschriebenen Tätigkeiten wird die berufliche Stellung am Ende der Tätigkeit erfragt und bei andauernden fortgeschriebenen Tätigkeiten wird die derzeitige berufliche Stellung erhoben.

¹³ Bei abgeschlossenen fortgeschriebenen Tätigkeiten wird die Arbeitszeit am Ende der Tätigkeit und bei andauernden fortgeschriebenen Tätigkeiten die derzeitige Arbeitszeit erhoben.

ein Spell auch mehrere wellenspezifische Informationen enthalten (z.B. Arbeitsstunden zum Interviewzeitpunkt der Wellen 2 und 3). Sie werden in wellenspezifischen Variablen abgelegt. Wellenspezifische Variablen bezogen auf die Welle 2 enden auf der Ziffer „0“, bezogen auf die Welle 3 enden diese auf der Ziffer „1“ usw. (siehe dazu wellenübergreifender User Guide).

Die Teilnahme an arbeitsmarktpolitischen Maßnahmen wird in der 4. Welle mit einem komplett inhaltlich überarbeiteten Fragenprogramm erhoben. Dieses bezieht sich ausschließlich auf Ein-Euro-Jobs (*ee_spells*). Der Anknüpfungszeitpunkt für diese Maßnahmen war für alle Befragten der Januar 2009. Eine Fortschreibung des bisherigen Maßnahmenspelldatensatz (*mn_spells*) findet ab Welle 4 nicht mehr statt.

Des Weiteren wurden im Personenfragebogen der 4. Welle die vertiefenden Fragen aus Welle 3 in den Themenbereichen „Netzwerke“ und „Gesundheit“ nicht erhoben. Dabei handelt es sich nur um ein kurzfristiges Aussetzen. Gemäß eines Rotationsplans werden die Fragen in späteren Wellen wieder aufgenommen.

Darüber hinaus wurden im Personenfragebogen:

- im Modul „Bildung“ einzelne Fragen für Schüler¹⁴ sowie Azubis und Studierende¹⁵ gestrichen;
- im Modul „Trägerkontakte“ die Frage nach den persönlichen Erfahrungen und Meinungen zu den Trägern der Grundsicherung für Arbeitssuchende¹⁶ wieder aufgenommen, und die relevante Befragtengruppe für dieses Modul auf die aktuellen Arbeitslosengeld II Bezieher eingegrenzt;
- im Fragenbereich „elterliche Bildungsaspiration“ die Fragen PEO0500, PEO0600 und PEO0700 gestrichen;
- Fragen zur Religionszugehörigkeit und Religiosität wieder aufgenommen (PD0200-PD0400);
- sowie bei den Migrationsfragen zwei neue Fragen¹⁷ zur Einschätzung der Deutschkenntnisse für Befragte mit Migrationshintergrund eingefügt.

Das Schwerpunktmodul zur „Altersvorsorge“, das speziell für den Einsatz in Welle 3 konzipiert wurde, entfiel ebenfalls im Fragenprogramm des Personenfragebogens.

1.3.2 Haushaltsfragebogen

Zum Thema Kinderbetreuung wurde ein Schwerpunktmodul eingefügt, das bereits in Welle 3 zum Einsatz kam. Für den Einsatz in Welle 4 wurden die Fragestellungen zur Kinderbetreuung überarbeitet. Dabei ging es um die Ganztagsbetreuung von Kindern unter 7 Jahren, die

¹⁴ Fragen: PB0500, PB0600, PB0700, PB0800, PB0900.

¹⁵ Frage PB1900

¹⁶ Frage PTK1600

¹⁷ Fragen PMI1200 und PMI1300

noch nicht zur Schule gehen¹⁸. Die Fragen zur Betreuung wurden um Items „Betreuung durch Privatpersonen“¹⁹ ergänzt.

Wesentliche weitere Änderungen im Haushaltsfragebogen der Welle 4 betrafen das Arbeitslosengeld II Modul. Um die Komplexität zu verringern, wurde das bisherige Fortschreibungskonzept über die Zielperson des Ursprungshaushalts aufgegeben. Der Anknüpfungzeitpunkt für die Fortschreibung wurde in Welle 4 auf den Zeitpunkt seit dem letzten Interviewdatum gesetzt. Innerhalb des Arbeitslosengeld-II-Moduls wurden die Fragen zu den möglichen Gründen des Bezugs neu entwickelt. Die Fragen zu den Gründen (AL20550) erhalten in Welle 4 alle Neubefragten, die zum Interviewzeitpunkt Arbeitslosengeld II beziehen, sowie Wiederholungsbefragte, deren Bezug in Welle 4 erstmalig aufgenommen wurde, d.h. das Beginndatum für den Bezug liegt nach dem Interviewdatum der Welle 3, oder deren Bezug noch andauert und damit fortgeschrieben wird. Die Fragen zu den Gründen werden nicht gestellt, wenn bei Wiederholungsbefragten aktuell kein Arbeitslosengeld-II-Bezug mehr vorliegt, d.h. der Bezug vor dem Interviewdatum der Welle 4 endet.

Darüber hinaus wurden im Arbeitslosengeld II Modul:

- die Frage zu den möglichen Gründen bei Bezugsende modifiziert (AL22200),
- Fragen zur möglichen Verzögerung der Bewilligung von Arbeitslosengeld II neu aufgenommen (AL20560 und AL20570) und
- die Nachfrage, welchem Haushaltsmitglied das Arbeitslosengeld II gekürzt wurde (AL22150a-e), gestrichen.

Im Haushaltsfragebogen insgesamt gab es die folgenden Streichungen bei:

- den Nachfragen zu eigenen Wohnräumen in Wohnheimen (HW0910 und HW0920)
- den Fragen zu den Sprachen im Haushalt (HD0600 – HD1000)
- der Schlussfrage zu den gegenwärtigen Lebensbedingungen des Haushalts

1.3.3 Stichprobe und Datenaufbereitung

Wie in den Vorwellen wurde auch in Welle 4 eine sog. Zugangsstichprobe für die BA-Teilstichprobe²⁰ gezogen. Ziel dabei 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 2009 Arbeitslosengeld II bezogen haben, nicht jedoch zum Zeitpunkt der Stichprobenziehung der 1., 2. oder 3. Welle (vgl. Kapitel 2.1 und zum Auffrischungskonzept

¹⁸ Fragen: HKI0250b-o

¹⁹ Fragen: HKI 0205 und HKI0405

²⁰ Die 1. Welle 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. Bevölkerungsstichprobe.

Trappmann et al 2009: 11 ff.). Diese in der 4. Welle 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.
- Zugangsstichprobe 3 (BA-Stichprobe) benennt die Stichprobe aus den SGB II-Zugängen zwischen Welle 3 und Welle 4.

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 1. Welle, aber auch die erstbefragten Haushalte aus den Zugangsstichproben in den Wellen 2, 3 und 4. Dabei muss berücksichtigt werden, dass nur solche erstbefragten Haushalte auch bei den Wiederholungsbefragungen zur Verfügung stehen, für welche die Panelbereitschaft²¹ 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 In-

²¹ Die Panelbereitschaft wird nur im ersten Interview mit dem Haushaltsbevollmächtigten erhoben und zählt damit für alle Haushaltsmitglieder. Panelbereite Haushalte haben der Speicherung ihrer Adresse zum Zwecke von Wiederholungsbefragungen im Rahmen der Studie zugestimmt.

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 4 beschrieben²².

PASS ist in der 1. Welle mit 12.794 realisierten Haushaltsinterviews gestartet, von denen 12.000 Haushalte ihre Panelbereitschaft erklärt haben. Die panelbereiten Haushalte der 1. Welle 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 in Folge kein Haushaltsinterview realisiert werden konnte. Dieser Fall konnte erstmalig am Ende der 3. Welle eintreten und wirkt sich auf das Panelbrutto in der 4. Welle aus.

Zusätzlich auf das Panelbrutto in der 4. Welle wirkt 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). Die Bruttoeinsatzstichprobe für die Welle 4 vom PASS umfasste insgesamt 9.497 Panelhaushalte.

Die Fallzahlen für den Bruttostichprobenumfang der jeweiligen Erhebungswellen und Teilstichproben wird in der nachfolgenden Tabelle ausgewiesen. In der 4. Welle konnte in 7.848 Haushalten der 9.465 Panelhaushalte mindestens ein Interview geführt werden. Dazu kommen 748 befragte Haushalte aus der BA-Zugangsstichprobe, von denen 723 Haushalte ihre Panelbereitschaft erklärt haben.

²² Vgl. Hartmann et al. (2008); Büngeler et al. (2009); Büngeler et al. (2010), Jesske & Quandt (2011).

Tabelle 1: Panelstichprobenumfang auf Haushaltsebene nach Wellen und Teilstichproben²³

n		Stichprobe				Gesamt	
		BA	Microm	BA-Zugang 1	BA-Zugang 2		BA-Zugang 3
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	3.917	3.618	861	1.069	9.465	
	HH-Interview realisiert	2.815	2.977	563	745	748	7.848
	davon: HH panelbereit	2.754	2.933	554	727	723	7.691

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

* Reduktion der Bruttostichprobe durch Widerrufsverfahren

Den 7.848 realisierten Haushaltsinterviews in der 4. Welle entsprechen 11.768 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		Stichprobe				Gesamt	
		BA	Microm	BA-Zugang 1	BA-Zugang 2		BA-Zugang 3
Welle 1	abs.	9.386	9.568			18.954	
Welle 2	abs.	4.753	6.392	1.342		12.487	
Welle 3	abs.	4.913	6.207	898	1.421	13.439	
Welle 4*	abs.	3.958	5.016	786	983	1.025	11.768

Quelle: P_Register; Scientific Use File IAB

* Reduktion der Bruttostichprobe durch Widerrufsverfahren

Für Befragte ohne ausreichende Deutschkenntnisse gab es die Möglichkeit, sich in einer anderen Sprache als Deutsch befragen zu lassen. Als alternative Interviewsprachen wurden dabei Türkisch und Russisch angeboten. Tabelle 3 gibt an, wie viele Haushalte bzw. Personen in den beiden Erhebungssprachen befragt wurden.

²³ 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.

Tabelle 3: Panelstichprobenumfang fremdsprachige Interviews nach Wellen

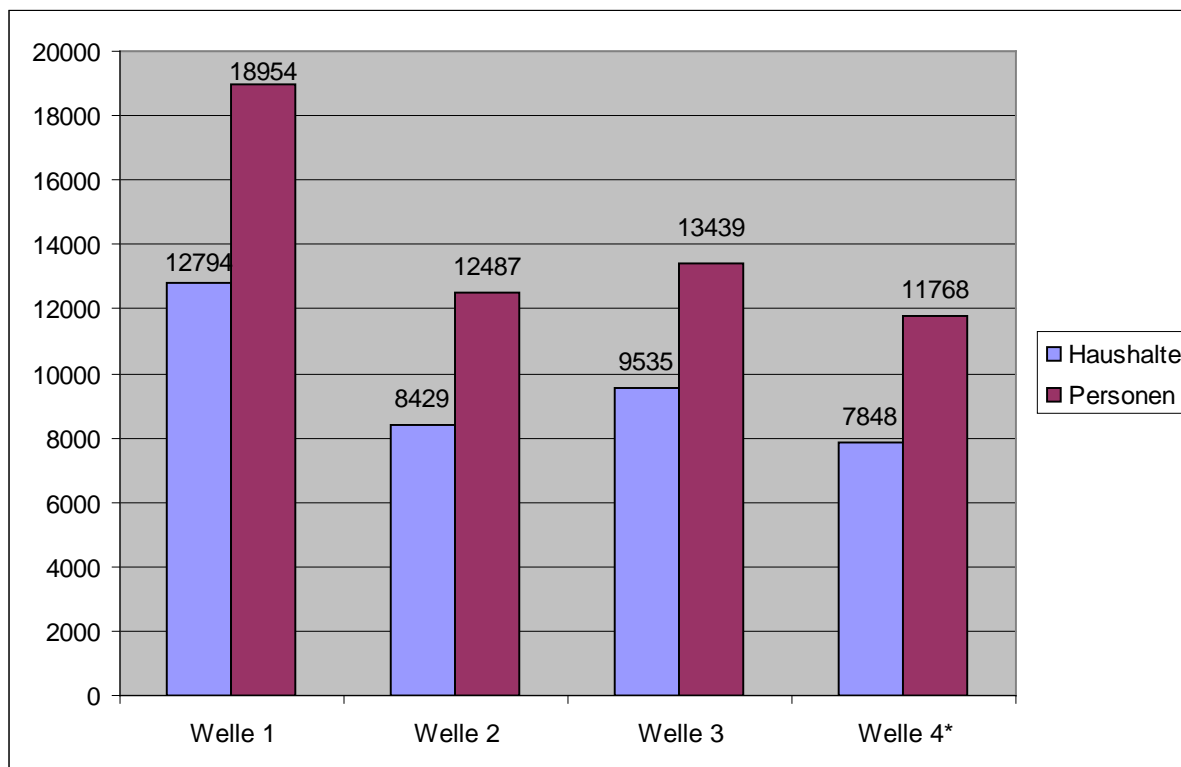
		Russisch abs.	Türkisch abs.
Welle 1	Haushalte	275	163
	Personen	432	305
Welle 2	Haushalte	156	39
	Personen	219	31
Welle 3	Haushalte	210	69
	Personen	330	109
Welle 4*	Haushalte	179	42
	Personen	285	78

Quelle: PENDDAT; Scientific Use File IAB

* Reduktion der Bruttostichprobe durch Widerrufsverfahren

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

Abbildung 1: Realisierte Panelstichprobe Haushalte und Personen nach Erhebungswellen



* Reduktion der Bruttostichprobe durch Widerrufsverfahren

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 annimmt²⁴. Als Ausschöpfungsquote auf der Haushaltsebene wird der Anteil der auswertbaren Haushaltsinterviews an der Summe aus auswertbaren Haushaltsinterviews und nicht neutralen Ausfällen 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.

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

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

Welle 4		Stichprobe					Gesamt
		BA	Microm	BA-Zugang 1	BA-Zugang 2	BA-Zugang 3	
HH brutto	abs.	3.917	3.618	861	1.096	2.440	11.932
	%	100,0	100,0	100,0	100,0	100,0	100,0
neutrale Ausfälle	abs.	12	7	5	3	19	46
	%	0,3	0,2	0,6	0,3	0,8	0,4
HH brutto bereinigt*	abs.	3.905	3.611	856	1.093	2.421	11.886
	%	100,0	100,0	100,0	100,0	100,0	100,0
HH-Interview realisiert	abs.	2.815	2.977	563	745	748	7.100
	%	72,1	82,4	65,8	68,2	30,9	59,7
davon: HH panelbereit	abs.					723	
	%					29,9	

* HH brutto - neutrale Ausfälle

Quelle: HH-Register; Scientific Use File IAB - für BA-Zugang 3: 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.

²⁴ 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.

Im Durchschnitt ergeben sich innerhalb der befragten Haushalte folgende Realisierungsraten:

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

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

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 Wellen 2, 3 und 4 mit panelbereiten Befragten der Vorwelle nach Teilstichproben

		Stichprobe				Gesamt
		BA	Microm	BA-Zugang 1	BA-Zugang 2	
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>	%	47,9	65,2		56,6
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>	%	71,8	78,8	63,2	74,5
Welle 4*	panelbereite Befragungspersonen W3	abs.	4.844	6.100	894	13.218
	in W4 wiederholt befragte Personen	abs.	3.287	4.347	626	854
	<i>Anteil</i>	%	67,9	71,3	70,0	61,9

Quelle: PENDDAT; Scientific Use File IAB

* Reduktion der Bruttostichprobe durch Widerrufsverfahren

2.3 Panelbereitschaft, Zuspielbereitschaft, Verknüpfung mit Prozessdaten

Zur Speicherung der Adressen zum Zwecke 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 Haushalte²⁵ 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 erstmalig 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

* Reduktion der Bruttostichprobe durch Widerrufsverfahren

**erstmalig befragte HH aus Auffrischung und Split

Quelle: PENDDAT und HH_Register; Scientific Use File IAB

Die Panelbereitschaft wurde bei erstmals befragten Haushalten in Welle 3 im Anschluss an das erste Personeninterview erfasst. Die Angabe dieser Person, wurde dann auf den Haushalt übertragen. War die Person panelbereit, wurde der Haushalt als panelbereit gewertet. War die Person nicht panelbereit, wurde der Haushalt als nicht panelbereit gewertet²⁶.

²⁵ Bei allen Haushalten der Welle 1 handelt es sich um erstmals befragte Haushalte. Ab Welle 2 handelt es sich nur bei Haushalten der jeweiligen Auffrischungsstichproben 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

²⁶ 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.

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.

Im Gegensatz zur Panelbereitschaft wurde die Erlaubnis zur Zuspielung 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 Zuspielbereitschaft wird dabei nicht in jeder Welle neu erhoben²⁷.

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

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

	Realisierte Personen-Interviews der Welle, in der die Zuspielfrage gestellt wurde abs.	Realisierte Personen-Interviews der Welle, in der die Zuspielbereitschaft erteilt wurde abs.	Anteil mit Zuspielbereitschaft %
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

* Reduktion der Bruttostichprobe durch Widerrufsverfahren

Basis: Personen 15 bis 64 Jahre

Quelle: PENDDAT; Scientific Use File IAB

²⁷ Auf Grund von Umstellungen bei der Filterung konnte es vorkommen, dass in Welle 2 und 3 die Frage nach der Zuspielbereitschaft erneut gestellt wurde, sofern die befragte Person ihre Zuspielbereitschaft in einer der Vorwellen noch nicht erklärt hatte.

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. Aber 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 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.

Vom Befragungszeitpunkt der 1. bis zur 4. Welle haben sich bereits insgesamt 343 Haushalte abgespalten, die in den Folgewellen weiter befragt werden konnten. In Welle 4 wurden 459 neue Split-Haushalte identifiziert, von den 126 befragt werden konnten. Die nicht befragten Splithaushalte werden in der 5. Welle erneut versucht zu kontaktieren, sofern sie die Teilnahme nicht verweigert haben.

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 so erstellte, wellenübergreifende Datensatz dann in das so genannte 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.

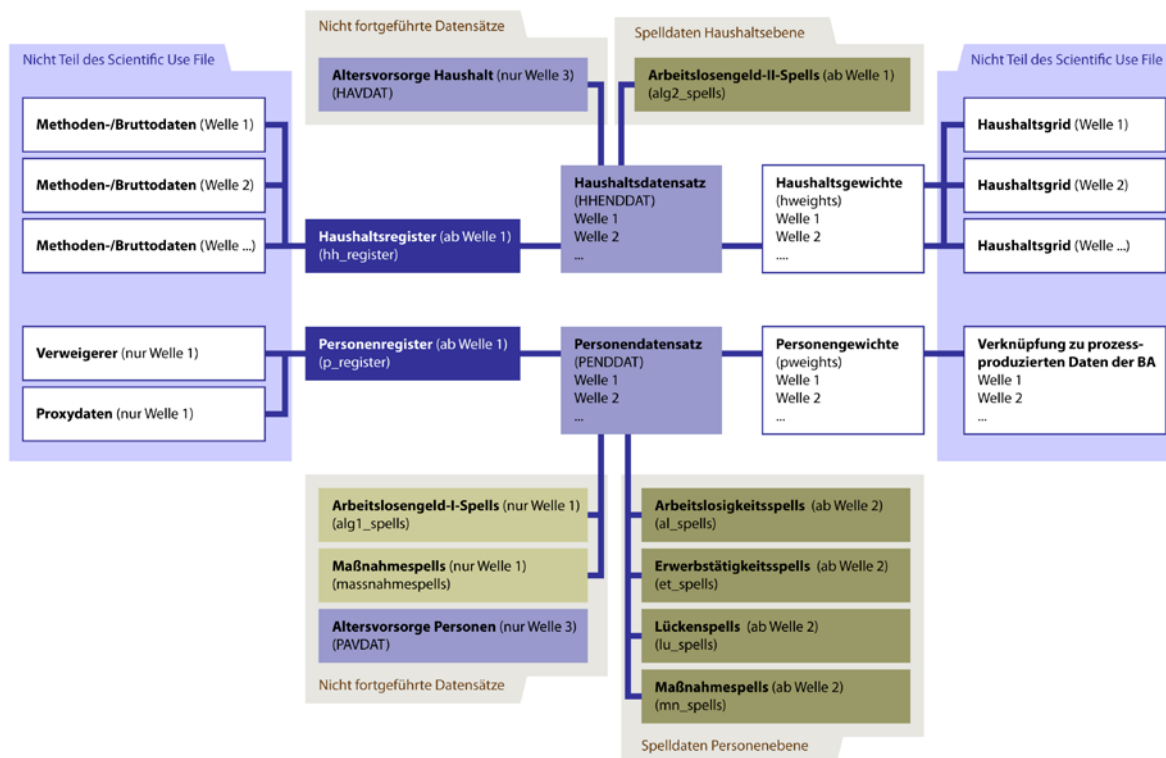
Auch die Aufbereitung im Long-Format ist mit Vor- und Nachteilen verbunden. Der entscheidende Vorteil dieser Variante ist, dass die Daten bereits in der für viele Längsschnittanalysen (wie z.B. Verweildaueranalysen) benötigten Struktur vorliegen. Der zusätzliche Aufwand für die Erstellung eines wellenübergreifenden 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 biographischer 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.

Abbildung 2: Datensatzstruktur des PASS in Welle 4



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 generiert²⁸, 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 sprechende Variablennamen²⁹. Die Ta-

²⁸ Andere offene Angaben wurden dagegen nicht vercodet, bspw. der Name des Trägers der Grundsicherung (PTK0100).

²⁹ ogebland (Geburtsland); ostaatn (Staatsangehörigkeit); ozulanda bis ozulandf (Zuzugsländer Eltern/Großeltern)

belle 9 und 10 geben einen Überblick über die offenen Angaben, die in der 4. Welle vercodet wurden³⁰.

Tabelle 9: Vercodung offener Angabe auf Haushaltsebene in Welle 4

fester Variablenname	Zugecodet zu Variable	Datensatz	Bezeichnung
HD1100a-o	HD1101a-o	HHENDDAT	Erwerbsstatus der HH-Mitglieder, ggf. Proxy-Angabe
HW0880	HW0881a-j	HHENDDAT	anderer, nicht in Liste enthaltener Auszugsgrund
AL21300a-h	AL21301a-h	alg2_spells	anderer, nicht in Liste enthaltener Grund für Kürzung
–	AL21401a-h		
AL22100a-h	AL21501a-h		
	AL21601a-h		
	AL21701a-h		
	AL21801a-h		
	AL21851a-h		
	AL21901a-h		
	AL22001a-h		
	AL22101a-h		
	AL22102a-h		
	AL22103a-h		
AL22200a – AL22200h	AL22201a-h	alg2_spells	anderer, nicht in Liste enthaltener Grund, warum kein ALG2 mehr bezogen wurde
AL20550a-h	AL20551a-h	alg2_spells	anderer, nicht in Liste enthaltener Grund, warum ALG2-Bezug begann

Tabelle 10: Vercodung offener Angaben auf Personenebene in Welle 4

fester Variablenname	Zugecodet zu Variable	Datensatz	Bezeichnung
PB0230 (Code 6)	PB0231	PENDDAT	anderer, nicht in Liste enthaltener deutscher Schulabschluss (Fortschreibung)
PB0230 (Code 7)	PB0231	PENDDAT	anderer, nicht in Liste enthaltener ausländischer Schulabschluss (Fortschreibung)
PB0400 (Code 9)	PB0401	PENDDAT	anderer, nicht in Liste enthaltener deutscher Schulabschluss (Ersterhebung o. keine Angabe in VW)
PB0400 (Code 10)	PB0401	PENDDAT	anderer, nicht in Liste enthaltener ausländischer Schulabschluss (Ersterhebung o. keine Angabe in VW)
PB1000	PB1001	PENDDAT	anderer, nicht in Liste enthaltener ausländischer Schulabschluss (Ersterhebung o. keine Angabe in VW)
PB1300a-j (Code 9)	PB1301a-j	PENDDAT	anderer, nicht in Liste enthaltener deutscher Berufschluss (Fortschreibung o. Ersterhebung)
PB1300a-j (Code 10)	PB1301a-j	PENDDAT	anderer, nicht in Liste enthaltener ausländischer Berufsabschluss (Fortschreibung o. Ersterhebung)

³⁰ Variablen, für die zwar in Vorwellen offene Angaben erhoben und vercodet wurden, jedoch nicht mehr in der aktuellen Welle, sind (mit Ausnahme des ALG2-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 4 (Fortsetzung)

<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-i</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 Zuzugsland des Elternteils/ Großelternteils
<i>PMI1110</i>	<i>PMI1111</i>	<i>PENDDAT</i>	Sprache im Freundeskreis: andere, nicht in Liste enthaltene Sprache
<i>PMI1120</i>	<i>PMI1121</i>	<i>PENDDAT</i>	Sprache im Freundeskreis, gleiche Nutzung zweier Sprachen: erste Sprache ist andere, nicht in Liste enthaltene Sprache
<i>PMI1130</i>	<i>PMI1131</i>	<i>PENDDAT</i>	Sprache im Freundeskreis, gleiche Nutzung zweier Sprachen: zweite Sprache ist andere, nicht in Liste enthaltene Sprache
<i>PSH0200</i>	<i>PSH0201</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener deutscher Schulabschluss der Mutter
<i>PSH0200</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</i>	<i>PSH0501</i>	<i>PENDDAT</i>	anderer, nicht in Liste enthaltener deutscher Schulabschluss des Vaters
<i>PSH0500</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 Haupterwerbstätigkeit, dem Status der Nichterwerbstätigkeit und dem Bezug von Arbeitslosengeld I anders erhoben wurden als noch in der 1. Welle. 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 damitversucht, diese soweit möglich über die Wellen hinweg zu vereinheitlichen, bevor auf ihrer Basis die Generierung durchgeführt wird.

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 Welle 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 die-

se 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 ist³¹.

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-Vercodung), 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 Fragetextes im aktuellen Interview eingebundet.

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 vorlagen³². 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, um den Berichtszeitraum deutlicher abzugrenzen³³. An anderen Stellen, insbesondere bei der Fortschreibung von Episodeninformationen³⁴, 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 in Folge 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 sind³⁵.

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 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.

³¹ 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.

³² 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.

³³ 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?“

³⁴ 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.

³⁵ 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

Tabelle 13a: Fortgeschriebene Vorwelleninformationen in Welle 4, 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
Jahr des Einzugs in aktuelle Wohnung	HW0900	in generierter Variable fortgeschrieben	HHENDDAT: <i>einzugj</i>
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	p_register: <i>bgbez4; bgbezb4</i>

höchsten Schulabschluss handelt. Die Information in der Folgewelle ist in diesem Sinne in den berichteten Variablen unvollständig.

Tabelle 13b Fortgeschriebene Vorwelleninformationen in Welle 4, Personenfragebogen

Personenfragebogen			
Konstrukt	Fragenr.	Anmerkung	Fortschreibung in Var.
Höchster allge- meinbildender Schulabschluss	PB0220- PB1100	in generierter Variable fortge- schrieben	PENDDAT: <i>schul1</i> (ohne offene Ang.) <i>schul2</i> (mit offenen Ang.)
Jahr des höchsten Schulabschlusses	PB0410	in generierter Variable fortge- schrieben	PENDDAT: <i>schulabj</i>
Berufsabschluss	PB1200- PB1600	höchster Berufsabschluss in ge- nerierter Variable fortgeschrieben	PENDDAT: <i>beruf1</i> (ohne offene Ang.) <i>beruf2</i> (mit offenen Ang.)
Jahr des Beruf- abschluss	PB1310	in generierter Variable fortge- schrieben	<i>berabj</i>
Zeiten von fortge- schriebenen Aktivi- täten im BIO- Spelldatensatz	BIO0200, BIO0800, BIO0300	in BIO-Spelldatensatz für Aufsatzspells fortgeschrieben	<i>bio_spells</i> <i>BIO0400</i> , <i>BIO0500</i> , <i>BIO0600</i>
		in BIO-Spelldatensatz für Aufsatzspells fortgeschrieben Informationen zur aktuellen Er- werbstätigkeit in generierten Vari- ablen fortgeschrieben	<i>bio_spells</i> : ET2300 PENDDAT: <i>isco88</i> ; <i>kldb</i> ; <i>stib</i> ; <i>stibkz</i> ; <i>arbeitszeit</i> ; <i>befrist</i> ; <i>mps</i> ; <i>siops</i> ; <i>isei</i> ; <i>egp</i> ; <i>esec</i>
		Informationen zum aktuellen (Nicht-)Erwerbstätigkeitsstatus in generierten Variablen fortge- schrieben	PENDDAT: <i>etakt</i> ; <i>alakt</i> ; <i>statakt</i>
Zeiten von Arbeits- losengeld-I-Bezug bei fortgeschrie- benen Arbeitslo- sigkeits-Spells		Informationen zum aktuellen Ar- beitslosengeld-I-Bezug	<i>bio_spells</i> : <i>AL0700</i> , <i>AL0800</i> , <i>AL0900</i> , <i>AL1000</i> , <i>AL1100</i> , <i>AL1200</i>
		in BIO-Spelldatensatz für Aufsatzspells fortgeschrieben	<i>bio_spells</i> : <i>AL0600</i> , <i>AL0601</i> PENDDAT: <i>alg1abez</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 von Quelldaten aus		Beschreibung
	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 Split-Haushalt umziehen (siehe auch Kapitel 2.4), nehmen ihre Preloadinformationen mit. Somit können sie auch im Split-Haushalt korrekt als erst- oder wiederzubefragende Person gesteuert werden. Wenn eine Person hingegen aus einem Split-Haushalt in einen in einer Vorwelle bewohnten Haushalt des Panels zurückzieht, findet keine Übernahme des Preloads für diese Person aus dem Split-Haushalt in den Ursprungs-Haushalt 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.³⁶ Dieser Sachverhalt tritt seit Welle 4 auf, da hier erstmals ein früheres Personeninterview mehr als zwei Wellen zurück liegen kann.

Für diese beiden Personengruppen werden die den „unveränderlichen“ Generierungen zu Grunde 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 werden. 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.

³⁶ 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 zurück liegt. 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.

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 4 notwendigen Quellvariablen aufgeführt³⁷. 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 4 im Haushaltsdatensatz (*HHENDDAT*) (alphabetische Reihenfolge)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<i>alg2abez</i>	<i>Aktueller ALG2-Bezug des HH, generiert</i> Indikator für aktuellen Arbeitslosengeld-II-Bezug des Haushalts (neu)	<i>zensiert; AL20300; AL20400; 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 generiert und gelieferte Information zum Bundesland, in dem Haushalt zum Befragungszeitpunkt lebt

³⁷ 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 4 im Haushaltsdatensatz (HHENDDAT) (alphabetische Reihenfolge) (Fortsetzung)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<i>einzugj</i>	<i>Einzugsjahr in akt. Wohnung, generiert</i> Information, in welchem Jahr der Haushalt in die aktuelle Wohnung eingezogen ist. Bei wiederbefragten Haushalten wurde das Einzugsjahr ab der 2. Welle nur erhoben, wenn es sich um einen in einem Wohnheim lebenden Haushalt handelte oder wenn seit der Vorwelle ein Umzug stattgefunden hatte (fs)	<u>Bei Ersterfassung:</u> <i>HW0900 (HHENDDAT)</i> <u>Bei wiederholter Erfassung:</u> <i>einzugj</i> aus Vorwelle; <i>HW0900, umzug (HHENDAT)</i>
<i>hinckat</i>	<i>kategorisiertes Haushaltseinkommen pro Monat (in Euro), generiert</i> Zusammenfassung der mehrstufig erhobenen kategorialen Angaben zum Haushaltseinkommen zu einer Variablen (neu)	<i>HEK0700; HEK0800; HEK0900; HEK1000; HEK1100 (HHENDDAT)</i>
<i>hincome</i>	<i>Haushaltseinkommen pro Monat (in Euro) inkl. kategorisierter Angaben, generiert</i> Generierung einer integrierten Variable aus kategorisierter und offener Angabe zum Haushaltsnettoeinkommen (neu)	<i>HEK0600; HEK0700; HEK0800; HEK0900; HEK1000; HEK1100 (HHENDDAT)</i>
<i>hintdat</i>	<i>Interviewdatum Haushaltsinterview</i> Generierte Variable mit dem Datum, an dem das Haushaltsinterview geführt wurde im Format JMMTT (neu)	<i>hintjahr, hintmon, hinttag (HHENDDAT)</i>
<i>kindu4</i>	<i>Steuervariable: Kind unter 4 Jahre im HH</i> Die Variable zeigt an, dass in der Welle mindestens 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)	<i>HD0200a - HD0200o (HHENDDAT)</i>
<i>kindu13</i>	<i>Steuervariable: Kind unter 13 Jahre im HH</i> Die Variable zeigt an, dass in der Welle mindestens 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)	<i>HD0200a - HD0200o (HHENDDAT)</i>
<i>kindu15</i>	<i>Steuervariable: Kind unter 15 Jahre im HH</i> Die Variable zeigt an, dass in der Welle mindestens 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)	<i>HD0200a - HD0200o; kategoriale Nachfrage zu Altersgruppe (bei KA in HD0200) (HHENDDAT)</i>

Tabelle 15: Einfache Generierungen für Welle 4 im Haushaltsdatensatz (HHENDDAT) (alphabetische Reihenfolge) (Fortsetzung)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<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 2. Welle nur erhoben, wenn seit der Vorwelle ein Umzug oder eine Veränderung der Wohnung stattgefunden hatte (fs)	<u>Bei Ersterfassung:</u> <i>HW1000 (HHENDDAT)</i> <u>Bei wiederholter Erfassung:</u> <i>wohnfl</i> aus Vorwelle; <i>HW1000; (HHENDDAT)</i>

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

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<i>akt1euro</i>	<i>Akt. Teiln. an einem 1-Euro-Job, generiert</i> Indikator: Befragungsperson nimmt zum Interviewzeitpunkt an einem Ein-Euro-Job teil (neu)	<i>zensiert (ee_spells);</i>
<i>alakt</i>	<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) <u>Anmerkung:</u> Die Variable wurde in W4 neu und auch rückwirkend für die Wellen 2 und 3 gebildet (siehe dazu Kapitel 5.4).	zensiert, spinteg, BIO0101 (<i>bio_spells</i>)
<i>alg1abez</i>	<i>aktueller ALG1-Bezug, generiert</i> Indikator: Befragungsperson bezieht zum Interviewzeitpunkt Arbeitslosengeld I. In der 4. Welle wurden die Zeiten seit Januar 2008 erhoben, in denen der Befragte arbeitslos gemeldet war. Für jede Episode wurde zudem erfragt, ob und in welchem Zeitraum Arbeitslosengeld I bezogen wurde. (neu)	<i>AL0700, AL 1000, AL 1100, AL 1200 (bio_spells)</i>
<i>apartner</i>	<i>Steuervariable: Unverh. Partner im HH</i> Indikator: Befragungsperson hat einen unspezifischen oder unverheirateten Partner im HH (neu)	Informationen zu Beziehungen im Haushalt (Haushaltsgrid); <i>PD0500 - PD0900 (PENDDAT)</i>
<i>azhpt1</i>	<i>Akt. vertragl. Arb.zeit Haupterwerb. (ohne Mini-jobs), gen.</i> Wöchentliche vertraglich vereinbarte Arbeitszeit der Haupterwerbstätigkeit, die der Befragte zum Interviewzeitpunkt ausübt, generiert aus offenen Angaben zur Arbeitszeit (neu) <u>Anmerkung:</u> Die Variable wurde in W4 neu und auch rückwirkend für die Wellen 2 und 3 gebildet (siehe dazu Kapitel 5.4).	ET2002 (<i>bio_spells</i>)

Tabelle 16: Einfache Generierungen für Welle 4 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 1)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<i>azhpt2</i>	<i>Akt. tatsaechl. Arb.zeit Haupterwerbst. (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)	ET2102, ET2202 (bio_spells)
<i>azges1</i>	<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)	ET2002 (bio_spells)
<i>azges2</i>	<i>Akt. tatsaechl. 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)	ET2102, ET2202 (bio_spells)
<i>befrist</i>	<i>jetzige Tätigkeit: Befristung? Generiert (alle Wellen)</i> Indikator: Die Erwerbstätigkeit, die der Befragte zum Interviewzeitpunkt ausübt, ist befristet (neu)	PET2510a; PET2510b (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 2008 genannt hat (uv)	<u>Bei Ersterfassung:</u> <i>bjahr</i> (bio_spells); PET3200b (PENDDAT) <u>Nach Ersterfassung:</u> <i>begjeewt</i> aus Vorwelle (PENDDAT)

Tabelle 16: Einfache Generierungen für Welle 4 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 2)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<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 (et_spells); PET3200a (PENDDAT)</i> <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 1. Welle berichteten Berufsabschlüsse wurden in der 2. Welle nacherhoben.	<u>Bei Ersterfassung:</u> <i>PB1310aj-kj (PENDDAT)</i> <u>Bei wiederholter Erfassung:</u> <i>berabj</i> aus Vorwelle; <i>PB1310aj-kj (PENDDAT)</i>
<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; (PENDDAT)</i> <u>Bei wiederholter Erfassung:</u> <i>beruf1</i> aus Vorwelle; <i>PB0100; PB0200; PB1200a; PB1300a-j (PENDDAT)</i>
<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 (PENDDAT)</i> <u>Bei wiederholter Erfassung:</u> <i>beruf2</i> aus Vorwelle; <i>PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601 (PENDDAT)</i>
<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) <u>Anmerkung:</u> Die Variable wurde in W4 neu und auch rückwirkend für die Wellen 2 und 3 gebildet (siehe dazu Kapitel 5.4).	<i>PEK0100b; PEK0200; PEK0300; PEK0400; PEK0500; PEK0600</i>

Tabelle 16: Einfache Generierungen für Welle 4 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 3)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<i>brutto</i>	<i>Bruttoeinkommen inkl. kategorisierter Angaben, generiert</i> Generierung einer integrierten Variable aus kategorisierter und offener Angabe zum Bruttoeinkommen (neu)	<i>PEK0100b; PEK0200; PEK0300; PEK0400; PEK0500; PEK0600 (PENDDAT)</i>
	<u>Anmerkung:</u> Die Variable wurde auch rückwirkend für die Vorwellen grundsätzlich überarbeitet (siehe dazu Kapitel 5.4).	
<i>bruttokat</i>	<i>kategorisiertes Bruttoeinkommen, generiert</i> Zusammenfassung der mehrstufig erhobenen kategorialen Bruttoeinkommensangaben zu einer Variable (neu)	<i>PEK0200; PEK0300; PEK0400; PEK0500; PEK0600 (PENDDAT)</i>
	<u>Anmerkung:</u> Die Variable wurde auch rückwirkend für die Vorwellen grundsätzlich überarbeitet (siehe dazu Kapitel 5.4)	
<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 2008 nicht erwerbstätig war (fs)	<u>Bei Ersterfassung:</u> <i>PET1200b (PENDDAT); ejahr; emonat (bio_spells)</i> <u>Bei wiederholter Erfassung:</u> <i>ejhrlewt</i> aus Vorwelle (PENDDAT); <i>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 Jahren 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)	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 Jahren 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 Jahren 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 Jahren im Haushalt hat (neu)	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)

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

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<i>emonlewt</i>	<i>Endzeitpunkt letzte Erwerbstätigkeit (Monat)</i> Monat, in dem der Befragte zuletzt erwerbstätig war (Generierung siehe <i>ejhrlewt</i>) (fs)	<u>Bei Ersterfassung:</u> <i>PET1200a</i> (<i>PENDDAT</i>); <i>emonat2</i> (<i>bio_spells</i>) <u>Bei wiederholter Erfassung</u> <i>emonlewt</i> aus Vorwelle (<i>PENDDAT</i>); <i>emonat</i> (<i>et_spells</i>)
<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 erwerbstaetig (>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)	zensiert, spintegr, BIO0101 (<i>bio_spells</i>)
<i>famstand</i>	<u>Anmerkung:</u> Die Variable wurde in W4 neu und auch rückwirkend für die Wellen 2 und 3 gebildet (siehe dazu Kapitel 5.4). <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</i> ; <i>PD0500</i> ; <i>PD0700</i> (<i>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: Alg2-Bezug aktuell</i> Indikator, für den Arbeitslosengeld-II-Bezug des Haushalts zum Zeitpunkt des Haushaltsinterviews (neu)	<i>HA0250b</i> (<i>HHENDDAT</i>) <i>AL20400</i> ; <i>AL20500</i> (<i>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</i> ; <i>PD1000</i> ; <i>PD1100</i> (<i>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 Elternteil handelt) (neu)	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)

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

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
	<i>Anmerkung:</i> 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>mberuf1</i>	<i>höchster ber. Abschluss der Mutter, inkl. Mut. im HH, ohne of. Ang., gen.</i> Der Berufsabschluss der Mutter wurde in der 1. Welle nur dann erhoben, wenn diese nicht im Haushalt wohnt. Wohnte sie im Haushalt, wurden die Angabe zum Berufsabschluss aus ihrem Personeninterview ergänzt. Ab der 2. Welle wurde die Information zum Berufsabschluss der Mutter bei allen Neubefragten Personen erhoben, unabhängig davon, ob die Mutter im Haushalt lebte oder nicht. Ab der 2. Welle wiederholt Befragte wurden die Werte aus der generierten Variable <i>mberuf1</i> aus der Vorwelle übernommen (uv)	<u>Bei Ersterfassung:</u> <i>PSH0300a-i</i> (PENDDAT) <u>Nach Ersterfassung:</u> <i>mberuf1</i> aus Vorwelle (PENDDAT)
<i>mberuf2</i>	<i>höchster ber. Abschluss der Mutter, inkl. Mut. im HH, mit of. Ang., gen.</i> Wie <i>mberuf1</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>PSH0301a-i</i> (PENDDAT) <u>Nach Ersterfassung:</u> <i>mberuf2</i> aus Vorwelle (PENDDAT)
<i>mhh</i>	<i>Steuervariable: Mutter lebt im HH</i> Indikator, der anzeigt, dass die unspezifische, leibliche, Stief-/Adoptivmutter der Befragungsperson im Haushalt lebt (neu)	Informationen zu Beziehungen im Haushalt (Haushaltsgrid)
<i>migration</i>	<i>Migrationshintergrund d. Befragten, generiert</i> 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)	<u>Bei Ersterfassung:</u> <i>PMI0100; PMI0700; PMI0800a-f; PMI0900a-f</i> (PENDDAT) <u>Nach Ersterfassung:</u> <i>migration</i> aus Vorwelle (PENDDAT)

Tabelle 16: Einfache Generierungen für Welle 4 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 6)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<i>mschul1</i>	<p><u>Anmerkung:</u> Das Generierungskonzept der Variable wurde ab der 2. Welle 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 1. Welle ebenfalls neu generiert.</p> <p><i>höchster allg. Schulabschluss d. Mutter, inkl. Mutter im HH, o. offene Ang., gen.</i></p> <p>Der höchste Schulabschluss der Mutter wurde in der 1. Welle nur dann erhoben, wenn diese nicht im Haushalt wohnt. Wohnte sie im Haushalt, wurden die Angabe zum Schulabschluss aus ihrem Personeninterview ergänzt (uv)</p> <p>Ab der 2. Welle wurde die Information zum höchsten Schulabschluss der Mutter bei allen neubefragten Personen erhoben, unabhängig davon ob die Mutter im Haushalt lebte oder nicht.</p>	<p><u>Bei Ersterfassung:</u> PSH0200 (PENDDAT)</p> <p><u>Nach Ersterfassung:</u> <i>mschul1</i> aus Vorwelle (PENDDAT)</p>
<i>mschul2</i>	<p><i>höchster allg. Schulabschluss d. Mutter, inkl. Mut. im HH, inkl. of. Ang., gen.</i></p> <p>Wie <i>mschul1</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> PSH0201 (PENDDAT)</p> <p><u>Nach Ersterfassung:</u> <i>mschul2</i> aus Vorwelle (PENDDAT)</p>
<i>mstib</i>	<p><i>Stellung im Beruf der Mutter, Kennziffer, generiert</i></p> <p>Detaillierte Stellung der Mutter im Beruf, generiert aus den Einzelvariablen (uv)</p>	<p><u>Bei Ersterfassung:</u> PSH0320; PSH0330; PSH0340; PSH0360; PSH0370; PSH0380 (PENDDAT)</p> <p><u>Nach Ersterfassung:</u> <i>mstib</i> (PENDDAT)</p>
<i>netges</i>	<p><i>Akt. Nettoeinkommen insgesamt (ohne Minijobs, inkl. kat. Ang.), gen.</i></p> <p>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)</p> <p><u>Anmerkung:</u> Die Variable wurde in W4 neu und auch rückwirkend ab Welle 1 gebildet (siehe dazu Kapitel 5.4).</p>	<p>PEK0700b; PEK0800; PEK0900; PEK1000; PEK1100; PEK1200</p>

Tabelle 16: Einfache Generierungen für Welle 4 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 7)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<i>netto</i>	<i>Nettoeinkommen inkl. kategorisierter Angaben, generiert</i> Generierung einer integrierten Variable aus kategorisierter und offener Angabe zum Nettoeinkommen (neu)	<i>PEK0700b; PEK0800; PEK0900; PEK1000; PEK1100; PEK1200 (PENDDAT)</i>
	<u>Anmerkung:</u> Die Variable wurde auch rückwirkend für die Vorwellen grundsätzlich überarbeitet (siehe dazu Kapitel 5.4)	
<i>nettokat</i>	<i>kategorisiertes Nettoeinkommen, generiert</i> Zusammenfassung der mehrstufig erhobenen kategorialen Nettoeinkommensangaben zu einer Variable (neu)	<i>PEK0800; PEK0900; PEK1000; PEK1100; PEK1200 (PENDDAT)</i>
	<u>Anmerkung:</u> Die Variable wurde auch rückwirkend für die Vorwellen grundsätzlich überarbeitet (siehe dazu Kapitel 5.4)	
<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
<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 (PENDDAT)</i>
<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 (PENDDAT)</i> <u>Bei wiederholter Erfassung:</u> <i>schul1</i> aus Vorwelle; <i>PB0200; PB0220; PB0230; PB0300; PB0400 (PENDDAT)</i>
<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 (PENDDAT)</i> <u>Bei wiederholter Erfassung:</u> <i>schul2</i> aus Vorwelle; <i>PB0200; PB0220; PB0231; PB0300; PB0401 (PENDDAT)</i>

Tabelle 16: Einfache Generierungen für Welle 4 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 8)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<i>schulabj</i>	<p><i>Jahr des höchsten Schulabschlusses</i> Jahr, in dem die Befragungsperson ihren höchsten Schulabschluss erworben hat (fs)</p> <p><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 4. Welle vor Mai 2010 geführt, wurde davon ausgegangen, dass der Abschluss in 2009 erworben wurde, bei später geführten Interviews wurde 2010 als Abschlussjahr angenommen.</p>	<p><u>Bei Ersterfassung:</u> <i>PB0220; PB0230; PB0410; pintjahr; pintmon (PENDDAT)</i></p> <p><u>Bei wiederholter Erfassung:</u> <i>schulabj</i> aus Vorwelle; <i>PB0220; PB0230; PB0410; pintjahr; pintmon (PENDDAT)</i></p>
<i>statakt</i>	<p><i>Aktueller Hauptstatus, generiert (ab Welle 2)</i> Zeigt an, welchen Hauptstatus die ZP zum Zeitpunkt des Personeninterviews der jeweiligen Welle hatte (neu)</p> <p><u>Anmerkung:</u> Die Variable wurde in W4 neu und auch rückwirkend für die Wellen 2 und 3 gebildet (siehe dazu Kapitel 5.4).</p>	<p>zensiert, spintegr, BIO0101, az2ges (bio_spells)</p>
<i>stib</i>	<p><i>Stellung im Beruf, Kennziffer, generiert</i> Generierung der detaillierten Kennziffer der Stellung im Beruf aus den Einzelvariablen.</p> <p>Generierung der Variable mit Informationen aus dem Erwerbstätigkeitsmodul (<i>ET0601-ET1201</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 (neu)</p>	<p><i>ET0500; ET0601 ET0701; ET0801; ET0901; ET1001; ET1101; ET1201 (et_spells)</i></p>
<i>stibeewt</i>	<p><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 2008 genannt zu haben (uv)</p>	<p><u>Bei Ersterhebung:</u> <i>PET3300b; PET3000; PET3400; PET3500; PET3600; PET3700; PET3800; PET3900 (PENDDAT) ET0601; ET0701; ET0801; ET0901; ET1001; ET1101; ET1201 (et_spells)</i></p> <p><u>Nach Ersterhebung:</u> <i>stibeewt</i> aus Vorwelle (<i>PENDDAT</i>)</p>

Tabelle 16: Einfache Generierungen für Welle 4 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 9)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<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)	<i>PET1510 (PENDDAT)</i>
<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 2008 nicht erwerbstätig war (fs)	<u>Bei Ersterfassung:</u> <i>PET1210b; PET1210; PET1220; PET1230; PET1240; PET1250; PET1260; PET1270 (PENDDAT)</i> <i>ET0601; ET0701; ET0801; ET0901; ET1001; ET1101; ET1200 (et_spells)</i> <u>Bei wiederholter Erfassung:</u> <i>stiblewt</i> aus Vorwelle <i>(PENDDAT); ET0601; ET0701; ET0801; ET0901; ET1001; ET1101; ET1200 (et_spells)</i>
<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> <i>PSH0600a-i (PENDDAT)</i> <u>Nach Ersterfassung:</u> <i>vberuf1</i> aus Vorwelle <i>(PENDDAT)</i>
<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> <i>PSH0601a-i (PENDDAT)</i> <u>Nach Ersterfassung:</u> <i>vberuf2</i> aus Vorwelle <i>(PENDDAT)</i>
<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> <i>PSH0500 (PENDDAT)</i> <u>Nach Ersterfassung:</u> <i>vschul1</i> aus Vorwelle <i>(PENDDAT)</i>
<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 (PENDDAT)</i> <u>Nach Ersterfassung:</u> <i>vschu2</i> aus Vorwelle <i>(PENDDAT)</i>

Tabelle 16: Einfache Generierungen für Welle 4 im Personendatensatz (PENDDAT) (alphabetische Reihenfolge) (Fortsetzung 10)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<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 (PENDDAT)</i> <u>Nach Ersterfassung:</u> <i>vstib</i> aus Vorwelle (<i>PENDDAT</i>)

Tabelle 17: Einfache Generierungen für Welle 4 im Arbeitslosengeld-II-Spelldatensatz (*alg2_spells*) (Reihenfolge analog zur Reihenfolge im Datensatz)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<i>bmonat</i>	<i>ALG2-Episode: Anfangsmonat, generiert</i> Beginnmonat der ALG2-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 2. Welle 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 (alg2_spells)</i>
<i>bjahr</i>	<i>ALG2-Episode: Anfangsjahr, generiert</i> Endjahr der Arbeitslosengeld-II-Episode.	<i>AL20200 (alg2_spells)</i>
<i>emonat</i>	<u>Anmerkung:</u> siehe <i>bmonat</i> <i>ALG2-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 (alg2_spells)</i> <i>hintmon (HHENDDAT)</i>
<i>ejahr</i>	<u>Anmerkung:</u> siehe <i>bmonat</i> <i>ALG2-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 (alg2_spells)</i> <i>hintjahr (HHENDDAT)</i>

Tabelle 17: Einfache Generierungen für Welle 4 im Arbeitslosengeld-II-Spelldatensatz (alg2_spells) (Reihenfolge analog zur Reihenfolge im Datensatz) (Fortsetzung 1)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
alg2kbma - alg2kbmh	<p>ALG2: 1. Kürzung: Anfangsmonat, generiert Beginnmonat der Arbeitslosengeld-II-Kürzung, in der für die Generierung die Jahreszeitangaben in konkrete Monatsangaben umgerechnet wurden.</p> <p><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 2. Welle in den Quellvariablen enthalten.</p>	<p>1. Kürzung: AL21000a (alg2_spells) bis 8. Kürzung: AL21000h (alg2_spells)</p>
alg2kbja – alg2kbjh	<p>ALG2: 1. Kürzung: Anfangsjahr, generiert Beginnjahr der Arbeitslosengeld-II-Kürzung. <u>Anmerkung:</u> siehe alg2kma - alg2kbmf</p>	<p>1. Kürzung: AL21100a (alg2_spells) bis 8. Kürzung: AL21100h (alg2_spells)</p>
alg2kema - alg2kemh	<p>ALG2: 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 - alg2kbmf</p>	<p>1. Kürzung: alg2kbma; alg2kbja; AL21200a; AL21201a; AL21202a (alg2_spells) bis 8. Kürzung: alg2kbmh; alg2kbjh; AL21200h; AL21201h; AL21202h (alg2_spells)</p>
alg2keja - alg2kejf	<p>ALG2: 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 - alg2kbmf</p>	<p>1. Kürzung: alg2kbma; alg2kbja; AL21200a; AL21201a; AL21202a (alg2_spells) bis 8. Kürzung: alg2kbmh; alg2kbjh; AL21200f; AL21201f; AL21202f (alg2_spells)</p>
AL22150a bis AL22150h	<p>ALG2: Kürzung: Welchem HH-Mitglied wurde gekürzt, generiert 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 erst ab der 2. Welle erhoben wurden, ist die Frage in allen in der ersten Welle berichteten Kürzungen an</p>	<p>Information, für welches Haushaltsmitglied in der jeweiligen Kürzungsepisode gekürzt wurde (HH102 im Haushaltsfragebogen für wiederbefragte Haushalte; HH53 im Haushaltsfragebogen für abgespaltene und neubefragte Haushalte).</p>

Tabelle 17: Einfache Generierungen für Welle 4 im Arbeitslosengeld-II-Spelldatensatz (alg2_spells) (Reihenfolge analog zur Reihenfolge im Datensatz) (Fortsetzung 3)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
	<p>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> 1 - dem HH-Mitglied wurde das ALG2 gekürzt 2 - dem HH-Mitglied wurde das ALG2 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>	
zensiert	<p><i>ALG2-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: (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. (b) Ein in Welle 3 befragter Haushalt gibt an, dass ein ALG2-Spell auch zum Interviewdatum der Welle 4 noch andauert. Oder ein Enddatum wird genannt, das identisch mit dem Interviewdatum der Welle 4 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 4 nicht mehr im Haushalt lebt und in Welle 4 nicht befragt wurde.</p>	<p>AL20300; AL20400, AL20500 (alg2_spells)</p>

Tabelle 18: Einfache Generierungen für Welle 4 im BIO-Spell Datensatz (bio_spells) (Reihenfolge analog zur Reihenfolge im Datensatz)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<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</i> (bio_spells)
<i>bjahr</i>	<p><i>Erwerbstätigkeit: Anfangsjahr, generiert</i> Beginnjahr der Erwerbstätigkeitsepisode.</p>	<i>BIO0300</i> (bio_spells)
<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</i> (Bio_spells) <i>pintmon</i> (PENDDAT)
<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</i> (et_spells) <i>pintjahr</i> (PENDDAT)
<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</i> (et_spells)

Tabelle 18: Einfache Generierungen für Welle 4 im BIO-Spell Datensatz (bio_spells) (Reihenfolge analog zur Reihenfolge im Datensatz) (Fortsetzung 1)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<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 4</u> <i>ET0602; ET0702; ET0802;</i> <i>ET0902; ET1002; ET1102;</i> <i>ET1202 (bio_spells)</i>
<i>az1</i>	<i>wöchentliche vertragliche Arbeitszeit.</i>	Ansonsten bleibt der Wert der Vorwelle bestehen <u>Erfassung von</u> <u>Spellinformationen in Welle 4</u> <i>ET2002 (bio_spells)</i> Ansonsten bleibt der Wert der Vorwelle bestehen
<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.	<u>Erfassung von</u> <u>Spellinformationen in Welle 4</u> <i>ET2102; ET2202 (bio_spells)</i> 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. Umrechnung der Monatsangaben siehe <i>bmonat</i> .	<i>AL0800 (bio_spells)</i>
<i>alg1bj</i>	<i>ALG1-Bezug: Anfangsjahr, generiert</i> Beginnjahr des Arbeitslosengeld-I-Bezugs. <u>Anmerkung:</u> siehe <i>alg1bm</i>	<i>AL0900 (bio_spells)</i>

Tabelle 18: Einfache Generierungen für Welle 4 im BIO-Spelldatensatz (bio_spells) (Reihenfolge analog zur Reihenfolge im Datensatz) (Fortsetzung 2)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<i>alg1em</i>	<p><i>ALG1-Bezug: Endmonat, generiert</i> 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>Anmerkung: siehe alg2kma - alg2kbme</i></p>	<p><i>AL1000; AL1200 (bio_spells)</i> <i>pintmon (PENDDAT)</i></p>
<i>alg1ej</i>	<p><i>ALG1-Bezug: Endjahr, generiert</i> 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>Anmerkung: siehe alg2kma - alg2kbme</i></p>	<p><i>AL1100; AL1200 (bio_spells)</i> <i>pintjahr (PENDDAT)</i></p>
<i>alg1akt</i>	<p><i>ALG1-Bezug: ggw. noch andauernder Spell (Rechtszensierung)</i> 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><i>Anmerkung: 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. 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.</i></p>	<p><i>emonat, ejahr, AL1000; AL1100;</i> <i>AL1200 (bio_spells)</i></p>

Tabelle 19: Einfache Generierungen für Welle 4 im Ein-Euro-Spell Datensatz (ee_spells) (Reihenfolge analog zur Reihenfolge im Datensatz)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<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 (mn_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: ggw. 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 4 im Personenregisterdatensatz (*p_register*) (alphabetische Reihenfolge)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
<i>alter4</i>	<p><i>Alter der Person in Welle 4 (2010)</i> Variable enthält die „beste“ verfügbare Altersinformation einer Person. Dabei handelt es sich entweder (a) um das aus dem Geburtsdatum, das in Welle 4 berichte wurde, berechnete Alter oder (b), wenn kein Geburtsdatum aus der 4. Welle vorliegt, um die Altersangabe aus dem Haushaltsinterview. Die Information aus <i>alter4</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 4 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)</i> <i>HD0200a</i> bis <i>HD0200o (HHENDDAT)</i></p>
<i>erwprox4</i>	<p><i>Erwerbsstatus laut HH-Interview in Welle 4 (2010)</i> 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> 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</i> bis <i>HD0100o</i> aller Wellen (<i>HHENDDAT</i>)</p>
<i>lastint</i>	<p><i>Befragungswelle des letzten Interviews auf Personenebene</i> 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>neuj4</i>	<p><i>Jahr, seit dem P. neu im aktuellen HH ist, berichtet in Welle 4 (2010)</i> Variable zeigt an, seit welchem Jahr die Person neu zum dem Haushalt gehört, dessen Haushaltsmitglied sie in der 4. Welle ist. <u>Anmerkung:</u> Information zum Datum stammt aus dem Welle 4-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 4 im Personenregisterdatensatz (p_register) (alphabetische Reihenfolge) (Fortsetzung 1)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
neum4	<p>Monat, seit dem P. neu im aktuellen HH ist, berichtet in Welle 4 (2010)</p> <p>Variable zeigt an, seit welchem Monat die Person neu zum dem Haushalt gehört, dessen Haushaltsmitglied sie in der 4. Welle ist.</p>	Informationen zum Datum, seit dem eine Person neu im Haushalt ist. Erhoben im Haushaltsgrid
wegj4	<p><u>Anmerkung:</u> siehe <i>neuj4</i></p> <p>Jahr, seit dem P. nicht mehr im vorherigen HH ist, berichtet in Welle 4 (2010)</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
wegm4	<p><u>Anmerkung:</u> Information zum Datum stammt aus dem Welle 4-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 4 (2010)</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
zdub4	<p><u>Anmerkung:</u> siehe <i>wegj4</i></p> <p>Zeiger: Personennr. der durch die ZP gedoppelten Person in Welle 4 (2010)</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
zmhh4	<p><u>Anmerkung:</u> Eine detaillierte Erläuterung zu den Gründen der Einführung dieser Variable findet sich in Kapitel 5.4.</p> <p>Zeiger: Personennummer der Mutter der Zielperson im HH in Welle 4 (2010)</p> <p>Identifiziert die Personennummer 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 der 4. Welle (Haushaltsgrid)
zparth4	<p>Zeiger: Personennummer des Partners der Zielperson im HH in Welle 4 (2010)</p> <p>Identifiziert die Personennummer 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 der Welle 4 (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 4 im Personenregisterdatensatz (p_register) (alphabetische Reihenfolge) (Fortsetzung 2)

Variable	Label und Beschreibung	Quellvar. für Gen. in Welle 4
zvhh4	<p><i>Zeiger: Personnummer des Vaters der Zielperson im HH in Welle 4 (2010)</i></p> <p>Identifiziert die Personnummer 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 der Welle 4 (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 4

	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		
	iscsd97			miscsd97	viscd97		
	bilzeit			mbilzeit	vbilzeit		
Informationen zum aktuellen Status	akt1euro						
	alakt						
	etakt						
	statakt					speltyp	
Sozio-ökonomische Position	egp	egplewt	egpeewt	megp	vegp	egp	
	esec	eseclewt	eseceewt	mesec	vesec	esec	
	isei	iseilewt	iseileewt	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
Datierung der Arbeitslosigkeit		ejhrlewt				ejahr	ejahr
						alg1bm	
Informationen zur Erwerbstätigkeit						alg1bj	
						alg1em	
						alg1ej	
Informationen zur Erwerbstätigkeit	befrist						
	azhpt1					az1	
	azhpt2					az2	
	azges1						
Berufliche Tätigkeit	azges2						
	isco88	iscolewt	iscoeewt	misco	visco	isco88	
Tätig in Branche	kldb	kldblewt	kldbeewt	mkldb	vkldb	kldb	
	branche					branche	

Abbildung 3: Übersicht über generierte Variablen auf Personenebene in Welle 4 (Fortsetzung)

	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						
Leistungsbezug	alg1abez				alg1akt		
	hhalg2						
Haushaltskontext und Familienstand	hhgr						
	famstand						
	vhh						
	mhh						
	apartner						
	epartner						
	ekind						
	ekin614						
	ekinu15						
	ekinu18						
	ekin1517						
	kindzges						
	kindzhh						
	Migrationshintergrund	ogebland					
ostaatan							
ozulanda							
ozulandb							
ozulandc							
ozulandd							
ozulande							
ozulandf							
migration							
Informationen zur Person	gebhalbj						
	palter						
	zpalthh						
	zpsex						
Allgemein	altbefr						
	fb_vers						
	panel						
	pintdat						
	RegP0100						
	sample						

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>	<i>bilzeit</i>																										
<u>Variablenlabel</u>	Schul- und Berufsbildung in Jahren, generiert																										
<u>Ausgangsvariablen</u>	<i>schul2; beruf2</i>																										
<u>Typ / Datensatz</u>	Bildung / Personendaten																										
<u>Bearbeitung</u>	Bernhard Christoph																										
<u>Erläuterung</u>	<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>	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																										
<u>Literatur</u>	Helberger (1988)																										

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.

schul beruf	nicht erhob.	Schüler	n. 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	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
Berufs- fachsch.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Schul. d. Gesewes.	-	-	-	-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/ 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.	Plnt 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
Plnt 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.	Plint 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
Plint 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)

ISCED 97

Variablenname

Variablenlabel

Ausgangsvariablen

Typ / Datensatz

Bearbeitung

Erläuterung

iscsed97

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. Zeilen 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
Univ/ 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

misced97

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.

Schulberuf	nicht erhob.	PInt fehlt	Eltern. unbek.	nicht gest.	TNZ	KA	WN	ohne Abschl.	Sonder-Schule	HS	RS	FHR	Abi	And. dt. Abschl.	And. au. Abschl.
nicht erhob.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
umplaus. Wert	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
PInt fehlt	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Eltern. 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)

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.	PInt 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	-	-	-	-	-	-	-	-	-	-	-	-	-	-
umplaus. Wert	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
PInt 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-Vercodung

<u>Gebildet</u>	<u>Erwerbstätigkeit</u>	<u>Variablenname</u>	<u>Ausgangsvariablen</u>
	aktuelle	<i>isco88</i>	ET2500
	Spelldaten (bio_spells)	<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-Vercodung), generiert		
	Spelldaten (<i>et_spells</i>): ISCO 88 (ZUMA-Vercodung), generiert		
	erste Ewt.: ISCO 88 (ZUMA-Vercodung), erste Erwerbstätigkeit, generiert		
	letzte Ewt.: ISCO 88 (ZUMA-Vercodung), letzte Erwerbstätigkeit, generiert		
	Vater: ISCO 88 (ZUMA-Vercodung) des Vaters, generiert		
	Mutter: ISCO 88 (ZUMA-Vercodung) 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 Berufscode 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>et_spells</i>): Klassifizierung der Berufe 1992 , generiert erste Ewt.: Klassifiz. d. Berufe 1992, erste Erwerbstaetigkeit, 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 (bio_ <i>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>	<p>aktuelle Ewt.: Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), akt. Beruf, gen.</p> <p>Spelldaten (bio_ <i>spells</i>): Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), gen.</p> <p>erste Ewt.: Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), erste Erwerbst., gen.</p> <p>letzte Ewt.: Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), letzte Erwerbst., gen.</p> <p>Vater: Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), Beruf d. Vaters, gen.</p> <p>Mutter: Klassenschema nach Erikson, Goldthorpe & Portocarrero (EGP), Beruf d. Mutter, gen.</p>		
<u>Typ / Datensatz</u>	sozioökonomische Position / Personendaten		
<u>Bearbeitung</u>	Bernhard Christoph		
<u>Erläuterung</u>	<p>Das Klassenschema von Erikson, Goldthorpe und Portocarero (Erikson et al. 1979, 1982; Erikson & Goldthorpe 1992) ist eines der gebräuchlichsten Instrumente zur Operationalisierung der Klassenposition.</p> <p>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.</p> <p>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 angaben, „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.</p>		
<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 (bio_ <i>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>vmps</i>	<i>visco</i>
	der Mutter	<i>mmmps</i>	<i>misco</i>
<u>Variablenlabel</u>	aktuelle Ewt.:Magnitude-Prestigeskala, akt. Beruf, gen. Spelldaten (bio_ <i>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 das Zentrum für Umfragen, Methoden und Analysen (GESIS-ZUMA).		
<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 (bio_ <i>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>vslops</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 (bio_ <i>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 das Zentrum für Umfragen, Methoden und Analysen (GESIS-ZUMA).		
<u>Literatur</u>	Ganzeboom & Treiman (1996, 2003); Treiman (1977)		

International Socio-Economic Index (ISEI)

<u>Gebildet</u>	<u>Erwerbstätigkeit</u>	<u>Variablenname</u>	Ausgangsvariablen
	aktuelle	<i>isei</i>	<i>isco88</i>
	Spelldaten (bio_ <i>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 (bio_ <i>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 das Zentrum für Umfragen, Methoden und Analysen (GESIS-ZUMA).		
<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>	Ausgangsvariablen
	aktuelle	<i>branche</i>	ET2600
	Spelldaten (bio_ <i>spells</i>)	<i>branche</i>	ET2600
<u>Variablenlabel</u>	aktuelle Ewt.: Jetzige Tätigkeit: Wirtschaftszweig/Branche (WZ2003) Spelldaten (bio_ <i>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)		

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. Hagenaaars et al. (1994).</p>
<u>Literatur</u>	Hagenaaars et al. (1994)

Deprivationsindex, ungewichtet

<u>Variablenname</u>	<i>depindug</i>
<u>Variablenlabel</u>	Deprivationsindex, ungew. (n. aus fin. Gründ. fehl. Items; ungew. Itemssumme: 26)
<u>Ausgangsvariablen</u>	<i>HLS0100a-HLS2600a; HLS0100b-HLS2600b</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 klassifizieren. 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 26 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 26 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>
<u>Literatur</u>	Andreß & Lipsmeier (2001); Halleröd (1995); Nolan & Whelan (1996); Ringen (1988)

Deprivationsindex, gewichtet

<u>Variablenname</u>	<i>depindg</i>
<u>Variablenlabel</u>	Deprivationsindex, gew. (n. aus fin. Gründ. fehl. Items; gew. Itemssumme: 12,8)
<u>Ausgangsvariablen</u>	<i>HLS0100a-HLS2600a; HLS0100b-HLS2600b; PLS0100-PLS2600</i>
<u>Typ / Datensatz</u>	materielle Situation / Haushaltsdaten (Gewichtung auf Personenebene)
<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>
<u>Literatur</u>	Andreß & Lipsmeier (1995, 2001); Andreß et al. (1996); Halleröd (1995); Lipsmeier (1999); Nolan & Whelan (1996)

Haushaltstypologie

Variablenname

Variablenlabel

Ausgangsvariablen

Typ / Datensatz

Bearbeitung

Erläuterung

hhtyp

Haushaltstyp, generiert

Haushaltsinformationen zu Alter und Verwandtschaftsbeziehungen

Haushaltsstruktur / Haushaltsdaten

Daniel Gebhardt

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 Personen 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 4

<u>Variablenname</u>	<i>bgnr4</i>
<u>Variablenlabel</u>	Nummer der Bedarfsgemeinschaft in Welle 4
<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>bgnr4</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>bgnr4</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 Arbeitssuchende (SGB II)

Bedarfsgemeinschaftstypologie, Welle 4

<u>Variablenname</u>	<i>bgtyp4</i>
<u>Variablenlabel</u>	Typ der Bedarfsgemeinschaft in Welle 4
<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>bgnr4</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>bgnr4</i>). Code -5, Generierung nicht möglich (fehlende Werte), wurde für Haushalte mit fehlenden Beziehungs- bzw. Altersangaben zu einzelnen Personen (vgl. Code 99 bei <i>bgnr4</i>) vergeben.</p>
<u>Literatur</u>	–

Arbeitslosengeld-II-Bezug der Bedarfsgemeinschaft zum Zeitpunkt der Stichprobenziehung, Welle 4

<u>Variablenname</u>	<i>bgbez4</i>
<u>Variablenlabel</u>	ALG2-Bezug der Bedarfsgemeinschaft bei Stichprobenziehung in Welle 4 (2010)
<u>Ausgangsvariablen</u>	<i>HA0250*</i> , <i>HA0300</i> , <i>AL20100</i> , <i>AL20200</i> , <i>AL20300</i> , <i>AL20400</i> , <i>AL20603</i> , <i>AL20703*</i> , <i>HA0400</i> , <i>sample</i> , <i>hnr</i> , <i>bgnr4</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>bgnr4</i> beschriebenen Vorgehen gebildet wurde, an, ob diese Bedarfsgemeinschaft zum Zeitpunkt der Stichprobenziehung für Welle 4 Arbeitslosengeld II bezogen hat oder nicht.
<u>Literatur</u>	–

Arbeitslosengeld-II-Bezug der Bedarfsgemeinschaft zum Befragungszeitpunkt, Welle 4

<u>Variablenname</u>	<i>bgbezb4</i>
<u>Variablenlabel</u>	ALG2-Bezug der Bedarfsgemeinschaft zum Befragungszeitpunkt in Welle 4 (2010)
<u>Ausgangsvariablen</u>	<i>AL20603</i> , <i>AL20703</i> , <i>zensiert (alg2_spells)</i> , <i>sample</i> , <i>hhgr</i> , <i>bgnr4</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>bgnr4</i> beschriebenen Vorgehen gebildet wurde, an, ob diese Bedarfsgemeinschaft zum Befragungszeitpunkt in Welle 4 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>bgnr4, 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>bgnr4</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>bgbez4, bgnr4, 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>bgbez4</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 dennoch 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 4. Welle erhobenen Informationen liegen bei infas zunächst als ASCII-Daten vor. In einer ersten Stufe erstellte infas aus diesen Rohdaten zunächst die folgenden Datensätze³⁸:

- 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
- Personendatensatz für im Längsschnitt erhobene Daten I (Modul „Erwerbsbiographie [Spells]“)
- Personendatensatz für im Längsschnitt erhobene Daten II (Modul „Maßnahmen“)
- Datensatz für offene Texte (übergreifend über alle Haushalts-, Personen- und Seniorenbefragung)

In einer zweiten Stufe werden die Daten einer ausführlichen formalen und inhaltlichen Prüfung unterzogen und zum Scientific Use File aufbereitet. Darüber hinaus lieferte 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.

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

³⁸ Im Rahmen der Datenaufbereitung zum Einsatz kamen die Softwarepakete Stata Version 11 und PASW Version 18.

Tabelle 21: Übersicht über die Schritte der Datenaufbereitung in der 4. Welle 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 4. Welle 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 Prüfungen 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 wer-

den 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 sein muss, die 15 Jahre oder älter ist.

- 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 vorlag³⁹.
- 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 Nicht-Realisierung 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 werden 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 Preload für alle durch den Panelhaushalt benannten Splithaushalte erzeugt. Wenn jedoch tatsächlich Personen, die laut Panelhaushalt in verschiedenen Splithaushalten leben, tatsächlich 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.

³⁹ Bei neubefragten Haushalten, für die zwar ein Haushalts-, aber kein gültiges Personeninterview vorlag, wurden die Haushaltsinterviews analog zum Vorgehen in der 1. Welle entfernt. Bei wiederbefragten und abgespaltenen Haushalten, wurden die Haushaltsinterviews dagegen beibehalten.

- 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 Panelhaushaltes (und ggf. aller Personen in weiteren Splithaushalten dieses Panelhaushaltes) 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 neubefragt 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 ALG2-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 ist. 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 Splithaushaltes 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 Splithaushaltes 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. Ein solcher Fall hatte die Einführung der Variable *zdub4* im Personenregister zur Folge. Weitere Erläuterungen hierzu sind in Kapitel 4.4 und Kapitel 5.4.1.4 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 Netzwerkes 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 im geringen Umfang geprüft werden. Bei erstbefragten Haushalten wird anhand der Angaben zu Vornamen, Alter und Geschlecht betrachtet, ob einzelne Haushaltsmitglieder mehrfach aufgelistet werden. In diesem Fall werden von den doppelt berichteten Personen nur die zuerst berichtete Haushaltsposition behalten, die weiteren Haushaltspositionen 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 wie auch ein Verschieben der personenbezogenen Angaben im Haushaltsinterview.
- Durch die Rückmeldung eines Feldinterviewers wurde ein Haushalt entdeckt, der doppelt in der Panelstichprobe enthalten war. Der Haushalt 10015439 ist als identischer Haushalt 15044862 bereits seit W1 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" HH zugeteilt worden waren, ist die Dopplung aufgefallen. Die HH-Zusammensetzung blieb in beiden HH 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 Gewich-

tung die Folge wären. Im *hh_register* ist der Dublettenhaushalt in Variable *hnettd4* mit dem Code 26 gekennzeichnet, der den Grund für die Nicht-Realisierung transparent macht. Im *p_register* sind die Haushaltsmitglieder des Dublettenhaushaltes in der Variable *pnettd4* 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 4. Welle unklar war, wurden alle Interviews der 4. Welle entfernt. Sie werden in der 5. Welle wie Haushalte behandelt, die in der 4. Welle 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 3. Welle 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 Auffrischungsstichprobe unterschieden. Die als nicht realisiert geltenden Haushalte aus der Auffrischungsstichprobe 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* (*hnettok4*, *hnettd4*, *pnettok4*, *pnettd4*) in den Haushalts- und Personenregisterdatensätzen. Über die entsprechenden Variablen im Haushaltsregister ist nachvollziehbar für welche wiederbefragten Haushalte die Haushaltsinterviews nachträglich entfernt wurden. Mittels der 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 4. Welle, 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.⁴⁰ 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 wird 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 hingegen 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 Kategorie für Einkommen oberhalb von € 99.999,- bei der offenen Einkommensangabe) handeln. Diese Codes wurden nur im Bedarfsfall vergeben.

⁴⁰ Die Bereinigung erfolgte dabei, wie in solchen Fällen üblich, immer von den zuerst erhobenen zu den später erfragten Items.

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“ ⁴¹

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 2. Welle 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 2. Welle in keiner Fragebogenversion mehr erfragt wurden, für die Beobachtungen dieser Welle den Wert „-9“. Ebenso verhält es sich mit Beobachtungen der 1. Welle. Erstmals nach der 1. Welle 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 Vorwelle durchgeführten Prüfungen für die aktuelle Welle angepasst und sinnvoll erweitert. Daneben wurde auch die Haushaltsstruktur sowie die Spelldaten plausibilitätsgeprüft – insbesondere auf unzulässige Überschneidungen innerhalb der einzelnen Spelldaten. Dabei wurden grundsätzlich nur die im Querschnitt der 4. Welle erhobenen Daten für die Prüfung herangezogen. Eine Prüfung im Längsschnitt, also zwischen den in der aktuellen und der Vorwelle gemachten Angaben, wurde nicht durchgeführt.

⁴¹ 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.

Im Einzelnen wurden die folgenden Schritte ausgeführt:

1. Widerspruchsprüfung: Generell wurden Widersprüche nur dann korrigiert, wenn entweder die Unplausibilität als besonders gravierend oder bzw. und wenn 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 Vornamen wurde eine Plausibilitätsprüfung der Haushaltsstruktur durchgeführt. Im Vorfeld dieser Prüfung wurden die Beziehungsinformationen im Haushalt durch die im Personeninterview berichteten Partnerschaften ergänzt.
 - Um un plausible 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. Un plausible Beziehungskonstellationen wurden dabei als solche markiert („-8“) bzw. auf Basis hinzugezogener Informationen über den Haushaltskontext korrigiert, so

fern 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 ähnliche unplausible Kombination von Beziehungen identifiziert, so wurde auch hier versucht, die Beziehung im Rahmen des Haushaltskontextes 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 3

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:

- W1: Panelkonzept, d.h. nur Erhebung am aktuellen Rand
- W2/3: Modulare Erhebung von ET-/AL-Episoden⁴² + Auffüllen von Lücken von > 3 Monaten und am aktuellen Rand

⁴² Hier und im folgenden: ET = Erwerbstätigkeiten; AL = Arbeitslosigkeiten; LU = Lücken (d.h. Aktivitäten, die nicht ET oder AL sind)

- W4: 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.

5.4.1.1 Variablen zur aktuellen Maßnahmenteilnahme im PENDDAT

akt1euro (*Akt. Teiln. an einem 1-Euro-Job, generiert*)

In W1⁴³ bis W3 wurden auf Personenebene jeweils Zeiten erfasst, in denen die ZP an arbeitsmarktpolitischen Maßnahmen teilgenommen hat. Trotz der Überarbeitung des Moduls nach W1 war das Erkenntnisinteresse vergleichbar: Berichtet werden sollten Maßnahmen unterschiedlichen Typs, von eintägigen Trainingsmaßnahmen bis hin zu länger dauernden Maßnahmen, wie bspw. ABMs oder Ein-Euro-Jobs. Ziel war es dabei, ein möglichst vollständiges Bild der Maßnahmengeschichte der ZP zu erhalten, weshalb ein sehr umfangreiches Spektrum an Maßnahmen erhoben wurde.

Auf Basis der so erhobenen Episodendaten (und den daraus generierten Spelldatensätzen *massnahmespells* und *mn_spells*) wurde die Variable *aktmassn* für den Personendatensatz generiert. Diese Variable zeigt an, ob die ZP zum Interviewzeitpunkt der jeweiligen Welle in einer andauernden Maßnahme war.

Ab W4 beschränkt sich das Erkenntnisinteresse auf die Ein-Euro-Jobs, an denen die ZP in einem bestimmten Zeitraum teilgenommen hat. Die bisher für den Personendatensatz generierte Variable *aktmassn* kann wegen des reduzierten Umfangs der erhobenen Maßnahmen nicht fortgeführt werden (ab W4: Code -9).

Stattdessen wird eine neue Variable erstellt, die anzeigt, ob die ZP zum Befragungszeitpunkt aktuell an einem Ein-Euro-Job teilgenommen hat. Diese Variable wird für alle Wellen generiert - jeweils auf Basis des für die Welle relevanten Spelldatensatzes, d.h. auf Basis von *massnahmespells* für W1, *mn_spells* für W2/W3 und *ee_spells* für W4.

Ausprägungen der zu generierenden Variable:

- | | |
|-----|--|
| -10 | Item fuer Fragebogenversion nicht relevant |
| -5 | Generierung nicht mgl. (fehlende Werte) |
| -3 | Trifft nicht zu (Filter) |
| 1 | Nimmt aktuell an 1-Euro-Job teil |
| 2 | Nimmt aktuell nicht an 1-Euro-Job teil |

⁴³ Hier und im folgenden: W1 = Welle 1, W2 = Welle 2 etc.

Tabelle 23: Generierungsgrundlagen von akt1euro

Welle	Datensatz	Bestimmung, dass Maßnahme Ein-Euro-Job ist	Bestimmung, dass Ein-Euro-Job andauert
1	massnahmespells	massntyp=1	zensiert=1
2	mn_spells	MN0100	spwelle=2 & zensiert=1
3	mn_spells	MN0100	spwelle=3 & zensiert=1
4	ee_spells	-	zensiert=1

5.4.1.2 Variablen zum (Nicht-)Erwerbsstatus im PENDDAT

Das Konzept zur Erhebung der Erwerbstätigkeiten wurde über die Wellen hinweg mehrfach überarbeitet:

- W1: Panelkonzept, d.h. nur Erhebung am aktuellen Rand
- W2/3: Modulare Erhebung von ET-/AL-Episoden + Auffüllen von Lücken von > 3 Monaten und am aktuellen Rand
- W4: Integrierte Erhebung von ET-/AL-/LU-Episoden

Aufgrund der Änderungen des Erhebungskonzeptes variieren die für die einzelnen Wellen vorliegenden Informationen hinsichtlich:

- Form der vorliegenden Informationen (Panel vs. Episoden)
- Detailgrad der vorliegenden Informationen (Hauptstatus vs. parallele Zustände)
- Konsistenz der vorliegenden Parallelitäten (Lückenfüllung vs. vollständige Erhebung paralleler Zustände)

Das bisherige Konzept der generierten Variablen zum (Nicht-)Erwerbsstatus orientiert sich sehr stark an der Erhebungslogik der 1. Welle. Diese Logik sah – vereinfacht dargestellt – wie folgt aus:

- Liegt eine ET von min. 1/h pro Woche vor?
- Wenn ET: Eine oder mehrere ET?
- Wenn ET (Angabe für HauptET): Sukzessive Bestimmung, ob ET ein Minijob, Ein-Euro-Job o.ä., oder Teil einer Ausbildung ist
- Wenn keine ET (oder HauptET=Minijob): Bestimmung des Status der Nichterwerbstätigkeit (AL oder anderer Status)

Das Konzept der generierten Variablen (*erwerb*, *erwerb2*, *nichterw*, *nichtew2*) folgt im weitesten Sinne dieser Erhebungslogik aus W1. Während in W1 durch die Abfragelogik keine konkurrierenden Zustände möglich waren (ZP mit ET, die kein Minijob ist, wurden nicht nach weiteren Aktivitäten gefragt) ergab sich ab W2 die Notwendigkeit, Entscheidungen zu treffen, falls mehr als eine andauernde Episode vorlag. Bei der Generierung der Variablen zum (Nicht-)Erwerbsstatus wurde dabei in W1 und W2 die folgende Logik angelegt:

Tabelle 24: Generierungslogik *erwerb*, *erwerb2*, *nichterw*, *nichtew2*

Variable	Generierungslogik Welle 1	Generierungslogik Welle 2
erwerb	<p><i>(1) Ausdifferenzierung Status HauptET</i></p> <ul style="list-style-type: none"> - Keine HauptET - HauptET: nicht Lehre/ABM/Minijob - HauptET: Teil einer Lehre - HauptET: ABM etc. - HauptET: Minijob <p><i>(2) Ausdifferenzierung Status HauptET ist Basis für weitere Generierung</i></p> <ul style="list-style-type: none"> - HauptET: nicht Lehre/ABM/Minijob → ET als Erwerbsstatus <p>(Ausnahmen:</p> <ul style="list-style-type: none"> Auszubildende (aus PB0100) mit arbeit < 21 → Auszubildende; Schüler (aus PB0100) mit arbeit >0 & arbeit < 24 → Schüler; Studenten (aus PB0100) mit arbeit >0 & arbeit < 21 → Studenten; Erwerbstätige mit arbeit > 0 & arbeit < 16 → Sonstige) <ul style="list-style-type: none"> - keine HauptET oder HauptET: Minijob → Erwerbsstatus übernehmen aus PET0801 (d.h. Einsetzen des Status der Nichterwerbtätigkeit) - keine HauptET + laut PB0100 Schüler/Studenten → Erwerbsstatus übernehmen aus PB0100 - HauptET: ABM etc. → Übernahme als Erwerbsstatus (ABM, 1-Euro-Job, etc.) <p><i>(3) Widersprüchliche Angaben entscheiden</i></p> <ul style="list-style-type: none"> - erwerb: ABM etc. + PB0100: Schüler/Student/Azubi → -8 - erwerb: Schüler + PB0100: Student → -8 - erwerb: Rentner + PB0100: Azubi → -8 - erwerb: Schüler + PB0100: Azubi → Übernahme Status aus PB0100 - erwerb: Sonstiges + PB0100: Schüler/Student/Azubi → Erwerbsstatus aus PB0100 	nicht generiert (-9)

Tabelle 24: Generierungslogik erwerb, erwerb2, nichterw, nichtew2 (Fortsetzung)

Variable	Generierungslogik Welle 1	Generierungslogik Welle 2
erwerb2	<p>(1) Recode von erwerb</p> <ul style="list-style-type: none"> - Zusammenlegung von Kategorien: - Arbeitslos + ABM/1-Euro-Job/u.ä → Arbeitslos - Lehre/Ausbildung/Weiterbildung/Umschulung + Student → Lehre/Ausbildung/Studium 	<p>(1) Recode von nichtew2</p> <p>(2) ETs integrieren</p> <ul style="list-style-type: none"> - Ersetzen der Werte, wenn aktuelle ET (>400 Euro aus ET-Spells) vorliegt <p>(3) Anpassungen vornehmen</p> <ul style="list-style-type: none"> - erwerb2: ET + PB0100: Student + Arbeitszeit ≤ 20h → Student - erwerb2: AL + PB0100: Student → Student - erwerb2: Schüler + PB0100: Student → Unklarer Status
nichterw	<p>(1) Recode von PET0800</p>	<p>(1) Recode von LU0100 (Lückenstatus ohne offene Angaben) + aktuelle AL aus AL-Spells</p> <ul style="list-style-type: none"> - Zusammenfassung von Kategorien: - Arbeitslos gemeldet + nicht gemeldet → Arbeitslos - Ausbildung/Lehre/Studium + Sonstiges → Sonstiges - Bestimmung MV aus PET0151/PET0911 + Indikator für fehlerhaft nicht ins Lückenmodul gefilterte Fälle
nichtew2	<p>(1) Recode von PET0801</p> <ul style="list-style-type: none"> - Zusammenlegung von Kategorien: - Arbeitslos + ABM/1-Euro-Job/u.ä → Arbeitslos - Lehre/Ausbildung/Weiterbildung/Umschulung + Student → Lehre/Ausbildung/Studium 	<p>(1) Recode von LU0101 (Lückenstatus mit offenen Angaben)</p> <ul style="list-style-type: none"> - Zusammenfassung von Kategorien: - Arbeitslos gemeldet + nicht gemeldet → Arbeitslos - Etwas anderes/Hauptstatus unklar → Sonstiges/Hauptstatus unklar <p>(2) Schüler/Studenten/Azubis aus PB0100 berücksichtigen</p> <ul style="list-style-type: none"> - Wenn bisher kein gültiger Status vorhanden → Übernahme der Information aus PB0100

Die generierten Variablen führen also die Logik des Erhebungskonzepts von W1 fort, die sich auch als Grundlogik in der generierten Variable findet:

- ET schlägt grundsätzlich alle anderen Zustände (bis auf Ausnahmen); AL schlägt alle Zustände außer ET (bis auf Ausnahmen)

In W1 wäre eine andere Logik (bspw. AL schlägt ET) nicht umsetzbar gewesen, da die Erhebungslogik die ET der ZP priorisiert hat und andere Zustände nur ersatzweise erhoben wurden. Das Vorgehen bei der Variablengenerierung folgt also dem Vorgehen bei der Erhebung.

Allerdings ist dieses Vorgehen für die Bestimmung des Hauptstatus der Person nicht wirklich sinnvoll und übergeht darüber hinaus Grundgedanken, die sich bspw. in der Definition von Arbeitslosigkeit (§16, 119 SGB III; gilt auch für SGB II nach §53a SGB II) finden.

Arbeitslosigkeit hat (laut der Definition des SGB III) bestimmte Voraussetzungen:

- Beschäftigungslosigkeit (d.h. keine bezahlte Beschäftigung bzw. Beschäftigung nur bis zur Grenze von 15h/Woche; Schwankungen sind möglich) (§119 SGB III)
- Verfügbarkeit (d.h. steht Vermittlungsbemühungen der BA zur Verfügung; Suche nach und Bereitschaft zu Beschäftigung \geq 15h/Woche; Kann Eingliederungsvorschlägen zeitnah Folge leisten; Bereitschaft zu Teilnahme an beruflichen Eingliederungsmaßnahmen) (§119 SGB III)
- Eigenbemühungen (d.h. Bemühungen, die Beschäftigungslosigkeit zu beenden) (§119 SGB III)
- Meldung (d.h. persönliche Arbeitslosmeldung bei BA) (§16 SGB III)
- Kein aktueller Maßnahmeteilnehmer (§16 SGB III)

Die bisherige Logik, in der AL von ET unabhängig vom Stundenumfang geschlagen wird, ist demnach eher durch die Erhebungslogik der W1 getrieben als durch eine inhaltliche Abwägung, was tatsächlich als Hauptstatus anzusehen ist.

Weitere Kritik an den (Nicht-)Erwerbsstatusvariablen betrifft das grundlegende Ziel dieser Variablen. Was sollen sie anzeigen? Den Hauptstatus der Person? Den Erwerbsstatus (wenn ja, was soll das genau sein)? Das Ziel erscheint bei genauerer Betrachtung nicht konsistent, da zwei Konzepte vermischt werden:

- Die Darstellung des Hauptstatus der ZP (d.h. bei konkurrierenden Zuständen wird eine Entscheidung getroffen, welcher Status unter welchen Bedingungen einen anderen „schlägt“)
- Die Darstellung, ob die ZP aktuell in einem bestimmten Zustand ist (auch wenn dieser mglw. im Hauptstatus von einem anderen „geschlagen“ wird)

Für die Generierung der (Nicht-)Erwerbsstatusvariablen ab W4 gibt es grundsätzlich zwei Möglichkeiten:

- Weiterführen der bisherigen Generierungslogik auf neuer Datenbasis
- Überarbeitung der Generierungslogik, mit dem Ziel:
 - Konzeptionelle Straffung (Was genau bilden die Variablen ab?)
 - Verbesserung von Entscheidungen, die in der Vergangenheit vor dem Hintergrund der vorhandenen Daten getroffen wurden aber inhaltlich suboptimal sind (d.h. nicht nur Weiterführung auf neuer Datenbasis, sondern inhaltliche Nutzung der detaillierteren Datenbasis)
 - Entschlackung (d.h. Entfernung von Variablen mit äußerst begrenztem Zusatznutzen)

Es wurde entschieden, die Generierungslogik grundlegend zu überarbeiten. Mit den bisherigen Variablen wird wie folgt verfahren:

Tabelle 25: Überarbeitung *erwerb*, *erwerb2*, *nichterw*, *nichtew2*

Variable	Entscheidung	Begründung
erwerb	beibehalten (W1: inhaltlich generiert) (W2ff.: -9)	<p>Variable bildet Erhebungskonzept W1 optimal ab. Der Schwerpunkt liegt auf ET (diese schlagen vereinfacht gesprochen Arbeitslosigkeit, die wiederum alles andere schlägt). Gegen eine Fortführung sprechen inhaltliche Bedenken, denen dank der detaillierteren Datenbasis durch eine Neukonzeption begegnet werden kann.</p> <p>Für W1 bleibt die Variable erhalten, da sie auf das Erhebungskonzept zugeschnitten ist. Die Besonderheiten (keine Parallelitäten; Konzentration auf ET; keine Differenzierung der Arbeitslosigkeit in gemeldete und nicht gemeldete) bleiben so auf W1 begrenzt.</p>
erwerb2	entfällt aus SUF	<p>Mit dieser Variable wird die Logik des Erhebungskonzepts aus W1 in harmonisierter Form fortgeführt.</p> <p>Dabei treten aber verschiedene Probleme auf:</p> <p>(1) Welche ETs erhoben werden, ändert sich (W1: 1h/Woche vs. W2ff.: >400 Euro)</p> <p>(2) Änderung der Schwerpunktsetzung (W1: Wenn ET [nicht Minijob] vorhanden → keine Erhebung paralleler ALs/LU-Zustände; W2ff.: ET/AL/(teilweise auch LU) parallel möglich)</p> <p>(3) Durch Festhalten an Logik der W1 können die Möglichkeiten der neuen Datenbasis nicht sinnvoll genutzt werden (um bspw. inhaltlich sinnvollere Entscheidungen zu treffen)</p> <p>Fazit: Harmonisierte Variable mit Schwerpunkt auf ET (wie bisher in <i>erwerb2</i>) ist einzige Möglichkeit für wellenübergreifend harmonisierte Variable. Deren Erstellung wäre zwar möglich, aber nur auf Basis inhaltlicher/konzeptioneller Entscheidungen die nicht sinnvoll sind. Da das Konzept der W1 als problematisch angesehen wird, wird auf die Auslieferung einer harmonisierten Variable verzichtet.</p>
nichterw	entfällt aus SUF	<p>Bisherige Aufteilung in Erwerbs- und Nichterwerbsstatus wird aufgegeben und durch Hauptstatus + Indikator für aktuelle (sozialversicherungspflichtige) Erwerbstätigkeit + Indikator für aktuelle Arbeitslosmeldung ersetzt</p> <p>W1: Variable bietet im Vergleich zu neuer Hauptstatusvariable keine Zusatzinformation</p> <p>W2ff.: Zusatzinformation im Vergleich zu neuer Hauptstatusvariable nur gering</p> <p>Fazit: Generell eher zusätzliche Komplexität mit sehr begrenztem Nutzen (bspw. Studenten > 20h/Arbeitszeit pro Woche). Für Auswertungen wahrscheinlich eine eigene Bestimmung der Nebenzustände eher sinnvoll als bisher ausgelieferte Variablen.</p>
nichtew2	entfällt aus SUF	(siehe <i>nichterw</i>)

Ab Welle 4 werden die folgenden Variablen generiert:

etakt: *Aktuell erwerbstaetig (>400 Euro pro Monat), generiert (ab Welle 2)*
alakt: *Aktuell arbeitslos gemeldet, generiert (ab Welle 2)*
statakt: *Aktueller Hauptstatus, generiert (ab Welle 2)*

Mit der Überarbeitung werden die folgenden Ziele verfolgt:

- Trennung der Information zum Hauptstatus (*statakt*) von der Information zu aktuell andauernden Episodentypen (*etakt*, *alakt*)
- Klarere Dokumentation der Regeln bei der Erstellung des Hauptstatus
- Differenzierung zwischen gemeldeter und nicht gemeldeter AL (wo möglich)

etakt (Aktuell erwerbstaetig (>400 Euro pro Monat), generiert (ab Welle 2))

Die Variable zeigt an, dass die ZP zum Zeitpunkt des Personeninterviews der jeweiligen Welle einen andauernden ET-Spell (d.h. eine ET > 400 Euro) hatte. Für W1 kann die Variable entsprechend nicht generiert werden, da sich das Erhebungskonzept zwischen W1 und den Folgewellen unterscheidet (W1: min 1h/Woche; W2ff. > 400 Euro/Monat).

Eine aktuelle ET liegt vor, wenn im Spelldatensatz der jeweiligen Welle ein zensierter ET-Spell vorhanden ist.

Ausprägungen der zu generierenden Variable:

-10	Item fuer Fragebogenversion nicht relevant
-5	Generierung nicht mgl. (fehlende Werte)
-3	Trifft nicht zu (Filter)
1	ZP akt. erwerbstaetig (> 400 Euro)
2	ZP akt. nicht erwerbstaetig (> 400 Euro)

alakt (Aktuell arbeitslos gemeldet, generiert (ab Welle 2))

Die Variable zeigt an, dass die ZP zum Zeitpunkt des Personeninterviews der jeweiligen Welle arbeitslos gemeldet war. Für Welle 1 kann die Variable entsprechend nicht generiert werden, da sich das Erhebungskonzept zwischen W1 und den Folgewellen unterscheidet (W1: AL nur erhoben, wenn keine ET vorliegt; W1: arbeitslos; W2ff.: arbeitslos gemeldet).

Eine aktuelle, gemeldete AL liegt vor, wenn im Spelldatensatz der jeweiligen Welle ein zensierter AL-Spell (gemeldete AL) vorhanden ist.

Ausprägungen der zu generierenden Variable:

-10	Item fuer Fragebogenversion nicht relevant
-5	Generierung nicht mgl. (fehlende Werte)
-3	Trifft nicht zu (Filter)
1	ZP akt. arbeitslos gemeldet
2	ZP akt. nicht arbeitslos gemeldet

statakt (Aktueller Hauptstatus, generiert (ab Welle 2))

Die Variable zeigt an, welchen Hauptstatus die ZP zum Zeitpunkt des Personeninterviews der jeweiligen Welle hatte.

Basis für die Generierung sind die Spelldatensätze (W2/3: *ET-/AL-/LU-Spells*; W4f.: *BIO-Spells*) sowie der Status als Schüler/Student/Azubi in PB0100.

Wenn in der jeweiligen Welle ein Spell eines bestimmten Typs aktuell andauert, dann liegt für die Person der entsprechende Zustand vor. Der Spelltyp wird in W2/W3 über den jeweiligen Spelldatensatz (*ET-/AL-Spells*) bzw. den Lückenzustand (*LU0101* in *LU-Spells*) bestimmt. Ab W4 kann die Variable *spelltyp* genutzt werden. In allen Wellen werden nur zum Interviewdatum andauernde Spells (d.h. *zensiert=1* im SUF der jeweiligen Welle) berücksichtigt. Der aktuelle Status als Schüler bzw. als Student/Azubi aus PB0100 wird berücksichtigt, als ob in der jeweiligen Welle ein aktuell andauernder Spell vorliegen würde.

Ausprägungen der zu generierenden Variable:

- 10 Item fuer Fragebogenversion nicht relevant
- 5 Generierung nicht mgl. (fehlende Werte)
- 3 Trifft nicht zu (Filter)
- 1 Erwerbstaetig m. Eink.> 400 Euro/Monat
- 2 Arbeitslos gemeldet
- 3 Schueler (bis W3 in Spells n. kat. erh.)
- 4 Ausbildung/Lehre/Studium
- 5 Wehr-/Zivildienst/etc.
- 6 Hausfrau/Hausmann
- 7 Mutterschutz/Erzieh.urlaub/Elternzeit
- 8 Rentner/Pensionaer/Vorruhestand
- 9 Sonstiges/Hauptstatus unklar
- 10 Arb.los, n. gemeldet (ab W4 aus off. A.)
- 11 Krank/beruf.unf./erw.unf./beh.(off. A.)
- 12 Selbst./mithelf. Fam.angeh. (off. A.)

Die Zuweisung der Codes sollte dabei schrittweise erfolgen:

Tabelle 186: Grobzuweisung –Spell höherer Priorität schlägt Spell niedrigerer Priorität

Priorität eines aktuellen Spells (bzw. analoger Zustand aus PB0100)	Code in statakt (analog zu Variable spelltyp)	Bedeutung
1	2	Arbeitslos gemeldet/Massnahmenteiln.
2	1	Erwerbstaetig m. Eink.> 400 Euro/Monat
3	8	Rentner/Pensionaer/Vorruhestand
4	7	Mutterschutz/Erzieh.urlaub/Elternzeit
5	5	Wehr-/Zivildienst/etc.
6	4	Ausbildung/Lehre/Studium
7	3	Schueler (bis W3 aus off. A.)
8	12	Selbst./mithelf. Fam.angeh. (off. A.)
9	11	Krank/beruf.unf./erw.unf./beh.(off. A.)
10	10	Arb.los, n. gemeldet (ab W4 aus off. A.)
11	6	Hausfrau/Hausmann
12	9	Sonstiges/Hauptstatus unklar

Liegen für die Zusatzinformation keine gültigen Werte vor, bleibt die Grobzuweisung (siehe Tabelle 27) erhalten.

Tabelle 27: Feinzuweisung für Sonderfälle

Grobzuweisung	Zusatzinformation	Entscheidung
Arbeitslos gemeldet	Erwerbstaetig m. Eink.> 400 Euro/Monat + Arbeitszeit (az2ges; tatsächliche Arbeitszeit, Summe über zensierte ETs) >= 15h	Erwerbstaetig m. Eink.> 400 Euro/Monat
Erwerbstaetig m. Eink.> 400 Euro/Monat	Ausbildung/Lehre/Studium + Arbeitszeit (az2ges; tatsächliche Arbeitszeit, Summe über zensierte ETs) <= 20h	Ausbildung/Lehre/Studium

Eine aktuelle, gemeldete AL liegt vor, wenn im Spelldatensatz der jeweiligen Welle (W2/3: AL-Spells; W4f.: BIO-Spells) ein zensierter AL-Spell (gemeldete AL) vorhanden ist.

5.4.1.3 Einkommensvariablen und Arbeitszeit im PENDDAT und in BIO-Spells

In W1-W4 beziehen sich die Variablen zur aktuellen Erwerbstätigkeit auf die Haupterwerbstätigkeit⁴⁴. Ausgenommen sind hier die Angaben zum Brutto-/Nettoeinkommen in W2-W4 – diese beziehen sich auf alle aktuell andauernden Erwerbstätigkeiten > 400 Euro (Unschärfe hinsichtlich Minijoblöhnen). Spellspezifische Angaben liegen nicht vor und werden erst ab W5 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 W2 auf verschiedene Erwerbstätigkeiten (Haupt-ET bzw. alle ET). Werden auf dieser Basis Stundenlöhne berechnet, kommt es zu Fehlern bei ZP mit mehreren ETs.
- II. Die unterschiedlichen Bezüge sind aus den Variablenlabels nicht ersichtlich.

Die generierten Variablen zu Einkommen und Arbeitszeit werden deshalb in W4 entsprechend überarbeitet.

Einkommensvariablen

Das Erhebungskonzept der Einkommensvariablen hat sich zwischen Welle 1 und 2 stark verändert, ohne dass dies zur Bildung neuer Variable geführt hat: *brutto* (*bruttokat*) und *netto* (*nettokat*) geben in W1 das Einkommen der Haupttätigkeit wieder, ab W 2 das Einkommen aller nicht geringfügigen Tätigkeiten. Dies ist inkonsistent und führt potenziell zu Fehlern in der Auswertung. Dieses Problem soll mit der Überarbeitung korrigiert werden:

Tabelle 28: Überarbeitung Einkommensvariablen⁴⁵

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

⁴⁴ W2-W3; Es handelt sich um die zensierte ET im ET-Spelldatensatz. 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.

⁴⁵ In W1 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 W1-W3):

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 W1 auf die jeweilige HauptET bezogen. Ab W2 wurden sie aber mit den kumulierten Angaben für alle ETs (>400 Euro) gefüllt, da nur diese erhoben wurden. Die Variablenlabel werden ab W4 entsprechend angepasst. Für W2-W4 werden die Variablen mit -9 gefüllt, da eine zu W1 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 ETs (>400 Euro) kumulierte Angabe zum Bruttoeinkommen. Für W1 ist die Variable in dieser Form nicht generierbar, da nur für die HauptET das Bruttoeinkommen erhoben wurde. Für W2 und W3 ist die Variable inhaltlich identisch zur Variable *brutto*, die im SUF W3 ausgeliefert wurde (d.h. vor der oben beschriebenen Überarbeitung). In W2-W4 wurde nur das kumulierte Bruttoeinkommen erhoben – die in W2/W3 genutzten Quellvariablen enthalten also schon die entsprechenden Angaben zum Gesamteinkommen aus ETs > 400 Euro. Für W4 soll die Variable analog zu W2/W3 erstellt werden. Ab W5 erfolgt die Generierung auf Basis von spellspezifischen Einkommensangaben.

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

Diese Variable enthält die über alle ETs (>400 Euro) kumulierte Angabe zum Nettoeinkommen. Für W1 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 W2 und W3 ist die Variable identisch zur Variable *netto*, die im SUF W3 ausgeliefert wurde. In W2-W4 wurde nur das kumulierte Nettoeinkommen erhoben – die in W2/W3 genutzten Quellvariablen enthalten also schon die entsprechenden Angaben zum Gesamteinkommen aus ETs > 400 Euro. Für W4 wurde die Variable analog zu W2/W3 erstellt. Ab W5 erfolgt die Generierung auf Basis von spellspezifischen Einkommensangaben.

Arbeitszeit

Durch die Korrektur der Variablen zum (Brutto-/Netto)Einkommen (s.o im gleichen Abschnitt) können im Personendatensatz keine Stundenlöhne mehr generiert werden, da als Arbeitszeit nur die tatsächliche Arbeitszeit der HauptET vorliegt (Variable *arbzeit* im *PENDDAT* des SUF W3).

Analog zur Überarbeitung der Einkommensvariablen ist demnach auch eine Überarbeitung der Arbeitszeitvariablen im *PENDDAT* als auch in den *BIO-Spells* nötig.

Tabelle 29: Überarbeitung Arbeitszeitvariablen

Variable	Inhalt	Datensatz	Generiert für			Basis		Anmerkung
			W1	W2	W3	OFFA	KatA	
az1	SpellET, vertrag.	Bio-Spells	0	1	1	1	0	Kat W2f.
azhpt1	HauptET, vertrag.	PENDDAT	0	1	1	1	0	Kat W2f.
azges1	Gesamt, vertrag.	PENDDAT	0	1	1	1	0	Kat W2f.
az2	SpellET, vertrag.	Bio-Spells	0	1	1	1	1	entspricht bisheriger Variable arbzeit (BIO-Spells); Kat W2f.; ET mit max(az2) = HauptET (wenn zwei identische: ET mit frühestem Beginn
azhpt2	HauptET, tatsaechl.	PENDDAT	1	1	1	1	1	entspricht bisher Variable arbzeit (PENDDAT); Kat. W1 != Kat W2f.
azges2	Gesamt, tatsaechl.	PENDDAT	1	1	1	1	1*	Kat W1!= Kat W2f.; in W1 keine Kat für NebenET

Überarbeitete Variablen (bereits im Datensatz in W1-W3):

arbzeit (Wöchentliche Arbeitszeit inkl. Angaben bei unregelmässiger Arbeitszeit, gen.)

Variable entfällt aus *PENDDAT* und *BIO-Spells*. Sie wird inhaltlich durch *azhpt2 (PENDDAT)* und *az2 (BIO-Spells)* ersetzt.

Neue Variablen in W4:

az1 (Vertragl. Arbeitszeit, gen.)

Die Variable wird für alle Episoden in den *BIO-Spells* generiert. Sie enthält die jeweils für den Spell aktuellste Angabe zur vertraglichen Arbeitszeit (ET > 400 Euro). Basis für die Generierung sind dabei jeweils die Querschnittsvariablen, für die im jeweiligen Spell zuletzt Angaben erhoben wurden.

Bsp.:

- Spell angelegt in W2, beendet in W2: Querschnittsvariablen W2
- Spell angelegt in W2, fortgeschrieben in W3 und W4: Querschnittsvariablen W4

azhpt1 (*Vertragl. akt. Arb.zeit Haupterwerb. (ohne Minijobs), gen.*)

Die Variable wird für den *PENDDAT* generiert. Sie enthält die vertragliche Arbeitszeit der in der jeweiligen Welle aktuell andauernden HauptET aus den Spelldaten (ET > 400 Euro). Für W1 ist die Variable nicht generierbar (-9), da die entsprechenden Angaben erst ab W2 erhoben werden. Ab W2 wird aus den jeweiligen Spelldaten die generierte Variable zur vertraglichen Arbeitszeit der HauptET (*az1*) in den *PENDDAT* übernommen. Bei welchem aktuell andauernden Spell es sich um die HauptET handelt, wird auch hier auf Basis der tatsächlichen Arbeitszeit bestimmt (generierte Variable *az2* in den Spelldaten; analog zum Vorgehen in W2/W3 wo die Variable *arbzeit* zur Bestimmung verwendet wurde).

azges1 (*Vertragl. akt. Arb.zeit insgesamt (ohne Minijobs), gen.*)

Die Variable wird für den *PENDDAT* generiert. Sie enthält die kumulierte vertragliche Arbeitszeit aller in der jeweiligen Welle aktuell andauernden Erwerbstätigkeiten aus den Spelldaten (ET > 400 Euro). Für W1 ist die Variable nicht generierbar (-9), da die entsprechenden Angaben erst ab W2 erhoben werden.

Ab W2 wird die Variable auf Basis der generierten Variable zur vertraglichen Arbeitszeit (*az1*) aus den Spelldaten generiert. Bei der Generierung werden die Angaben in der generierten Variable zur vertraglichen Arbeitszeit (*az1*) über alle Spells hinweg kumuliert, die zum Befragungszeitpunkt aktuell andauern. Diese Angabe wird in den *PENDDAT* übernommen.

az2 (*Tatsaechliche. Arbeitszeit inkl. Angaben bei unregelmässiger Arbeitszeit, gen.*)

Die Variable wird für alle Episoden in den *BIO-Spells* generiert. Sie enthält die jeweils für den Spell aktuellste Angabe zur tatsächlichen Arbeitszeit und integriert dabei auch die kategorialen Angaben bei unregelmässiger Arbeitszeit. Basis für die Generierung sind dabei jeweils die Querschnittsvariablen, für die im jeweiligen Spell zuletzt Angaben erhoben wurden.

Bsp.:

- Spell angelegt in W2, beendet in W2: Querschnittsvariablen W2
- Spell angelegt in W2, fortgeschrieben in W3 und W4: Querschnittsvariablen W4

Die Variable ersetzt die bisher in den ET-Spells generierte Variable *arbeitszeit* (die entsprechend entfällt). Die Generierung erfolgt analog zur Generierung von *arbeitszeit* in der Datenaufbereitung W2/W3.

Die Variable dient der Bestimmung der Haupterwerbstätigkeit (HauptET) in einer Welle, für die verschiedene Angaben in den *PENDDAT* übernommen werden. Die HauptET ist die in der jeweiligen Welle andauernde ET mit dem größten Stundenumfang. Existieren mehrere Erwerbstätigkeiten mit gleichem Stundenumfang, so wird die zuerst beginnende ausgewählt.

azhpt2 (*Tatsaechl. akt. Arb.zeit Haupterwerb. (ohne Minijobs, inkl. kat. Ang.), gen.*)

Die Variable wird für den *PENDDAT* generiert. Sie enthält die tatsächliche Arbeitszeit der aktuell andauernden Haupterwerbstätigkeit und integriert dabei auch die kategoriale Nachfrage bei unregelmässiger Arbeitszeit. Die Variable ersetzt inhaltlich die im *PENDDAT* entfallene Variable *arbeitszeit*. Die Generierung erfolgt für W1-W4 analog zur Generierung der entfallenen Variable.

In W1 wird die Variable auf Basis der Querschnittsdaten generiert. Sie kombiniert also sowohl die offenen Angaben zur tatsächlichen Arbeitszeit als auch die kategoriale Nachfrage. 1-Euro-Jobs, ABM, Minijobs und Tätigkeiten, die Teil einer Lehre sind, werden dabei nicht berücksichtigt – für diese Fälle kann die Variable nicht generiert werden (-3), da in den W2-W4 analoge Informationen nicht erhoben werden.

Ab W2 wird aus den jeweiligen Spelldaten die generierte Variable zur tatsächlichen Arbeitszeit der HauptET (*az2*) in den *PENDDAT* übernommen. Bei welchem aktuell andauernden Spell es sich um die HauptET handelt, wird auch hier auf Basis der tatsächlichen Arbeitszeit bestimmt (generierte Variable *az2* in den Spelldaten; analog zum Vorgehen in W2/W3 wo die Variable *arbzeit* zur Bestimmung verwendet wurde).

Die kategoriale Nachfrage bei unregelmäßiger Arbeitszeit unterscheidet sich zwischen W1 und W2ff. Trotzdem findet eine Integration über die Wellen hinweg statt.

azges2 (*Tatsaechl. akt. Arb.zeit insgesamt (ohne Minijobs, inkl. kat. Ang.), gen.*)

Die Variable wird für den *PENDDAT* generiert. Sie enthält die kumulierte tatsächliche Arbeitszeit aller in der jeweiligen Welle aktuell andauernden Erwerbstätigkeiten.

In W1 werden dazu die Stunden der HauptET (nach Integration der kategorialen Angaben bei unregelmäßiger Arbeitszeit) mit den Angaben zur tatsächlichen Arbeitszeit der weiteren ETs kombiniert. 1-Euro-Jobs, ABM, Minijobs und Tätigkeiten, die Teil einer Lehre sind, werden dabei nicht berücksichtigt – für diese Fälle kann die Variable nicht generiert werden (-3), da in den W2-W4 analoge Informationen nicht erhoben werden.

Ab W2 wird die Variable auf Basis der generierten Variable zur tatsächlichen Arbeitszeit (*az2*) aus den Spelldaten generiert. Bei der Generierung werden die Angaben in der generierten Variable zur tatsächlichen Arbeitszeit (*az2*) über alle Spells hinweg kumuliert, die zum Befragungszeitpunkt aktuell andauern. Diese Angabe wird in den *PENDDAT* übernommen.

5.4.1.4 Dublettenzeiger im p_register

zdub4 (*Zeiger: Personennr. der durch die ZP gedoppelten Person in Welle 4 (2009/2010)*)

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 (*zdub4*) 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 werden⁴⁶. Die beteiligten Haushalte wurden befragt und die Person taucht in den jeweiligen Haushaltsstrukturen auf.

⁴⁶ Ob es sich um dieselbe Person handelt, wird in der Prüfung der Haushaltsstrukturen geprüft. Basis sind demografische Angaben (Name, Alter, Geschlecht, Geburtsdatum).

Da beide Personen im SUF nicht die gleiche Personennummer haben dürfen, erhält die Person, die in der aktuellen Welle neu angelegt wurde (=Dublette in DublettenHH), eine neue feste Personennummer nach den bekannten Regeln. Die Person, die bereits in einer Vorwelle Panel war (d.h. die Person ist bereits mit einer festen Nummer im Personenregister der Vorwelle bekannt), behält die bisherige Personennummer (=Original in UrHH).

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 *pnetto**4. Die Dublette wird in der Folgewelle als Neubefragter behandelt, für den keine Vorinformationen vorliegen. In der Gewichtung wird also 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 Dubletten-Interview keine konkurrierende Information aus dem Interview mit Original gibt, bleibt das Dubletten-Interview im SUF. Die Informationen aus dem Personeninterview werden hier auch für die Preload-Generierung der Folgewelle genutzt.

Im Personenregister wird ab Welle 4 eine wellenspezifische Zeigervariable 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 Befragungshaushaltes ist (analog zur Vergabe des -6 Codes in den übrigen Variablen des Personenregisters).

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

2. potenzielle Dubletten

Eine Person ist als Mitglied eines Haushaltes bekannt, der bereits früher im PASS befragt wurde (=UrHH). Obwohl dieser Haushalt in der aktuellen Welle nicht befragt wurde, taucht die Person nun in einem anderen Haushalt auf (=DublettenHH). Da der UrHH dieser Person seit dem Auftauchen der Person in dem DublettenHH, 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 potenzielle Dublette des Originals im UrHH dar.

Bei potenziellen Dubletten wird davon ausgegangen, dass es sich um einen noch nicht berichteten Auszug aus dem UrHH handelt. Entsprechend wird der potenziellen Dublette im SUF die feste Personennummer des Originals zugewiesen, d.h. die Person wird behandelt,

als ob sie vom UrHH in den DublettenHH umgezogen wäre. In der aktuellen Welle auf Personenebene geführte Interviews bleiben erhalten.

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

Für den Preload der Folgewelle wird wie folgt vorgegangen:

- UrHH der Person

Die potenzielle Dublette wird nicht aus der Haushaltsstruktur entfernt, damit der Auszug berichtet werden kann. Die Personinterviews, die in DublettenHH geführt wurden, werden für die Preload-Generierung der potenziellen Dublette im UrHH nicht genutzt. Person und Haushalt werden hier also wie reguläre Aussetzer behandelt.

- DublettenHH der Person

Die Preload-Generierung erfolgt regulär auf Basis der jeweils letzten Information, d.h. die Person ist in der Haushaltsstruktur des Dubletten-HH enthalten und die im Dubletten-HH geführten Personeninterviews werden bei der Generierung berücksichtigt.

In der Folgewelle kann es sich herausstellen, dass es sich um eine echte Dublette (siehe 1) handelt. In diesem Fall wird die Personennummer der Person im DublettenHH auch rückwirkend korrigiert, d.h. auch rückwirkend erhält die Person eine neue Personennummer wie unter (1) beschrieben (in allen relevanten Datensätzen). Auch der Dublettenzeiger wird rückwirkend gefüllt.

Ausprägungen der zu generierenden Variable:

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

5.4.1.5 Spelldaten zur Erwerbsbiografie in BIO-Spells

spellnr (*Alle Aktivitaeten: Spellnummer*)

Es handelt sich um den spelltypenübergreifenden, chronologischen Spellzähler über alle Spelltypen. (Generierung auf Basis der gleichen Ausgangsvariablen wie bei chronologischer Sortierung in W2/W3)

spellnret (*Erwerbstaetigkeiten: Spellnummer*)

Es handelt sich um den spelltypeninternen, chronologischen Spellzähler für die ET-Spells (ET > 400 Euro). (Generierung auf Basis der gleichen Ausgangsvariablen wie bei chronologischer Sortierung in W2/W3)

spellnral (*Arbeitslosigkeiten: Spellnummer*)

Es handelt sich um den spelltypeninternen, chronologischen Spellzähler für die AL-Spells (gemeldete AL). (Generierung auf Basis der gleichen Ausgangsvariablen wie bei chronologischer Sortierung in W2/W3)

spellnrlu (*Sonstige Aktivitaeten: Spellnummer*)

Es handelt sich um den spelltypeninternen, chronologischen Spellzähler für die LU-Spells (d.h. alle anderen Aktivitäten, bei denen es sich nicht um ET oder gemeldete AL handelt). (Generierung auf Basis der gleichen Ausgangsvariablen wie bei chronologischer Sortierung in W2/W3)

spelltyp (*Alle Aktivitaeten: Aktivitaetstyp, generiert*)

In W2/W3 wurden nacheinander zuerst alle relevanten ET-Episoden und im Anschluss alle AL-Episoden der ZP erfragt. Danach wurden Lücken von > 3 Monaten bzw. Lücken am aktuellen Rand aufgenommen (LU). Für den SUF wurden auf Basis dieser Informationen drei Spelldatensätze erstellt, die jeweils nur einen Spelltyp enthielten (ET, AL oder LU).

Ab W4 werden die ET/AL-Spells nicht mehr modular nacheinander erfragt und danach durch eine Lückenerhebung vervollständigt, sondern die Erhebung erfolgt integriert wobei auch LU-Episoden (d.h. Episoden, bei denen es sich nicht um ET oder gemeldete AL handelt) direkt erhoben werden. Daraus ergibt sich ein integrierter Biografiespelldatensatz. Die in W2/W3 typenspezifisch erhobenen Spells werden in den integrierten Datensatz von W4 integriert.

Um über die einzelnen Wellen hinweg die einzelnen Spelltypen einfach unterscheiden zu können, wird mit *spelltyp* eine generierte Variable gebildet.

Ausprägungen der zu generierenden Variable:

- 1 Erwerbstaetig m. Eink.> 400 Euro/Monat
- 2 Arbeitslos gemeldet/Massnahmenteiln.
- 3 Arb.los, n. gemeldet (ab W4 aus off. A.)
- 4 Ausbildung/Lehre/Studium
- 5 Wehr-/Zivildienst/etc.
- 6 Hausfrau/Hausmann
- 7 Mutterschutz/Erzieh.urlaub/Elternzeit
- 8 Rentner/Pensionaer/Vorruhestand
- 9 Etwas anderes gemacht
- 10 Schueler (bis W3 aus off. A.)
- 11 Hauptstatus unklar (aus off. Angaben)
- 12 Krank/beruf.unf./erw.unf./beh.(off. A.)
- 13 Selbst./mithelf. Fam.angeh. (off. A.)
- 1 Weiss nicht
- 2 Keine Angabe

Tabelle 30: Bildung von spelltyp

Inhaltliche Ausprägungen von spelltyp	Ausprägung trifft bei Spell aus W2/W3 zu, wenn ...	Ausprägung trifft bei Spell aus W4 zu, wenn ...
Erwerbstätig m. Einkommen > 400 Euro	alle ET-Spells	BIO0101=1
Arbeitslos gemeldet/Massnahmenteiln.	alle AL-Spells	BIO0101=2
Arb.los, n. gemeldet (ab W4 aus off. A.)	LU-Spells + LU0101=3	(entspr. Code in BIO0101)
Ausbildung/Lehre/Studium	LU-Spells + LU0101=4	BIO0101=4
Wehr-/Zivildienst/FSJ	LU-Spells + LU0101=5	BIO0101=5
Hausfrau/Hausmann	LU-Spells + LU0101=6	BIO0101=6
Mutterschutz/Erzieh.urlaub/Elternzeit	LU-Spells + LU0101=7	BIO0101=7
Rentner/Pensionaer/Vorruhestand	LU-Spells + LU0101=8	BIO0101=8
Etwas anderes gemacht	LU-Spells + LU0101=9	BIO0101=9
Schueler (bis W3 aus off. A.)	LU-Spells + LU0101=10	BIO0101=3
Hauptstatus unklar (aus off. Angaben)	LU-Spells + LU0101=11	(entspr. Code in BIO0101)
Krank/erwerbsunf./behindert/berufsunf. (aus off. Angaben)	LU-Spells + LU0101=12	(entspr. Code in BIO0101)
Selbstaendig/mithelf. Familienangeh. (aus off. Angaben)	LU-Spells + LU0101=13	(entspr. Code in BIO0101)
Weiss nicht	LU-Spells + LU0101=-1	BIO0101=-1
Keine Angabe	LU-Spells + LU0101=-2	BIO0101=-2

spinteg (Alle Aktivitaeten: Betroffen von Spellintegration W2/W3 und W4ff.?)

Um für den Nutzer deutlich zu machen, mit welcher Erhebungslogik ein Spell im integrierten Biografiespell Datensatz erhoben wurde, wird eine generierte Variable erstellt, mit der drei Typen von Spells unterschieden werden können: (1) Spells, die nur Informationen enthalten, die in W2/W3 berichtet wurden, (2) Spells, die in W2/W3 angelegt wurden aber mit dem neuen Biografiemodul in W4 fortgeschrieben wurden und (3) Spells, die nur Informationen enthalten, die in W4ff. berichtet wurden.

Tabelle 31: Bildung von spinteg

Inhaltliche Ausprägung von spinteg	Ausprägung trifft bei Spell aus W2/W3 zu, wenn ...	Ausprägung trifft bei Spell aus W4 zu, wenn ...
Angaben vollstaendig aus W2/W3	ET/AL: angelegt + beendet LU: angelegt + beendet	-
Angaben aus W2/W3 und W4(ff.)	ET/AL: angelegt + zensiert LU: angelegt + zensiert	ET/AL: Fortschreibung VW LU: Fortschreibung VW
Angaben vollstaendig aus W4(ff.)	-	ET/AL/LU: Angelegt

Variablen zur Arbeitszeit (az1, az2)

Die Variablen werden im Rahmen der Überarbeitung der Arbeitszeitvariablen beschrieben (siehe weiter oben).

5.4.2 Fehlerkorrekturen

Im Rahmen der Datenaufbereitung für den Scientific Use File der 4. Welle 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 File der 3. Welle festgestellt wurden. Tabelle 32 bis 36 geben einen Überblick über die rückwirkenden Veränderungen an bereits ausgelieferten Wellen des PASS⁴⁷.

Tabelle 32: Übersicht über rückwirkende Veränderungen im Haushaltsdatensatz (HHENDDAT)

Geänderte Variable	Betroffener Datensatz	Geänderte Welle	Art der Änderung	Beschreibung der Änderung
HEK1630	HHENDDAT	3	Korrektur	HEK1630 wird seit Welle 3 erhoben. Die Filterbedingung für HEK1630 lautet korrekt „HEK1600>0 & HEK1600<99 & HHALG2_an=0“. In der Erhebung und Aufbereitung der W3-Daten wurde der Term „HHALG2_an=0“ nicht berücksichtigt, weil er in der Fragebogenvorlage für Welle 3 nicht dokumentiert war. Inhaltlich ist es jedoch notwendig, die Steuerungsvariable HHALG2_an zu berücksichtigen, da ausschließlich Haushalte, die kein ALG2 beziehen, den Kinderzuschlag erhalten können. Haushalte mit in Welle 3 aktuell andauerndem ALG2-Bezug wurden nachträglich auf Code „-3“ vercodet. Hierdurch wird Datenkonsistenz mit den Folgewellen hergestellt. Zudem wurde das Variablenlabel korrigiert. Das Label muss „Kinderzuschlag“ und nicht „Kindergeldzuschlag“ heißen.
HD0300	HHENDDAT	2,3	Korrektur	Die Wertelabels der Codes 1 bis 4 bezogen sich auf den Referenzzeitpunkt 01/2005. Dies war nur im Welle1-SUF korrekt. In jeder Erhebungswelle wird der Referenzzeitpunkt um Jahr nach hinten verschoben. Die Labels wurden angepasst: Statt „01.2005“ wird nun auf „Ref.dat“ verwiesen.

⁴⁷ Anpassungen an Werte- oder Variablenlabels sind hier nur berücksichtigt, wenn sie dadurch die Interpretation von Variablen oder Werten verändert.

Tabelle 32: Übersicht über rückwirkende Veränderungen im Haushaltsdatensatz (HHENDDAT) (Fortsetzung)

Geänderte Variable	Betroffener Datensatz	Geänderte Welle	Art der Änderung	Beschreibung der Änderung
<i>nextstra</i>	<i>HHENDDAT</i>	1-3	Entfällt	Die Variable "nextstra" ist in der PASS-Auslieferung der Welle 4 nicht mehr enthalten. Die Variable beschreibt die nächste Schicht bei der Ziehung, mit der eine Schicht fusioniert werden kann, falls in einer Schicht nur noch eine einzige PSU enthalten ist. Diese Variable muss aber über alle vorigen und nachfolgenden Wellen fortgeschrieben werden, da ansonsten strpsu (die Schichtkennung) innerhalb einer PSU nicht mehr konstant ist. Angesicht einer großen Zahl zu fusionierender Schichten in der Auffrischungsstichprobe ALG II 2007 in Welle 4 und der Tatsache, dass jährlich weitere kleine Substichproben hinzukommen, ist davon auszugehen, dass relativ schnell eine sehr grobe Zusammenfassung von Schichten erfolgen muss, die mit den ursprünglichen Schichten nicht mehr korrespondiert. Deshalb wurde entschieden, die Variable zu entfernen.
<i>bik</i>	<i>HHENDDAT</i>	1-3	Korrektur	Die Variable bik wurde in einer unüblichen Codierung ausgeliefert, bei der Code 0 für die hochverdichteten Ballungsräume und Code 9 für die kleinsten und am schwächsten verflochtenen Regionen stand. Dies wurde nun auf die gebräuchliche Codierung recodiert, bei der Code 10 für die hochverdichteten Ballungsräume und Code 1 für die kleinsten und am schwächsten verflochtenen Regionen steht.

Tabelle 33: Übersicht über rückwirkende Veränderungen im Personendatensatz (PENDDAT)

Geänderte Variable	Betroffener Datensatz	Geänderte Welle	Art der Änderung	Beschreibung der Änderung
<i>*isco_it berpr_it</i>	<i>PENDDAT</i>	1-3	Entfällt	Die Infratestvercodung liegt ab Welle 4 nicht mehr vor. Die ZUMA-Vercodung ist Basis für die themenbezogenen Generierungen.
<i>*kldb*</i>	<i>PENDDAT</i>	1-3	Korrektur	Ab W4 keine Vercodung mehr durch Infratest. Integration in bestehende Variablen möglich - dafür mussten Namen und Variablenlabels angepasst werden.
<i>alg1s05</i>	<i>PENDDAT</i>	1-3	Entfällt	Die Variable bezieht sich dem Label nach darauf, ob der Befragte seit Januar 2005 ALG1 bezogen hat. Dieser Sachverhalt kann in den neueren Teilstichproben nicht mehr bestimmt werden, weil der retrospektive Befragungszeitraum nicht so weit zurückreicht. Das Fortführen dieser Variable ist insofern nicht sinnvoll.
<i>arbeits</i>	<i>PENDDAT</i>	1-3	Entfällt	Siehe Kapitel 5.4.1.3
<i>brutto* netto*</i>	<i>PENDDAT</i>	2-3	Korrektur	Siehe Kapitel 5.4.1.3
<i>nichterw nichtew2 erwerb2</i>	<i>PENDDAT</i>	1-3	Entfällt	Siehe Kapitel 5.4.1.2
<i>alakt etakt statakt</i>	<i>PENDDAT</i>	2-3	Eingefügt	Siehe Kapitel 5.4.1.2
<i>brges</i>	<i>PENDDAT</i>	2-3	Eingefügt	Siehe Kapitel 5.4.1.3
<i>netges</i>	<i>PENDDAT</i>	1-3	Eingefügt	Siehe Kapitel 5.4.1.3
<i>azhpt1 azges1</i>	<i>PENDDAT</i>	2-3	Eingefügt	Siehe Kapitel 5.4.1.3
<i>azhpt2 azges2</i>	<i>PENDDAT</i>	1-3	Eingefügt	Siehe Kapitel 5.4.1.3
<i>mhh vhh</i>	<i>PENDDAT</i>	3	Korrektur	Für Senioreninterviews (fb_vers=3) sind die beiden Variablen mit -10 zu füllen. In Welle 3 waren die beiden Variablen statt dessen inhaltlich gefüllt.
<i>beruf1 beruf2</i>	<i>PENDDAT</i>	3	Korrektur	Die beiden Variablen werden fortgeschrieben. Es werden also zunächst die Angaben der aktuellen Welle berücksichtigt. Wenn es keine neuen Angaben aus der aktuellen Welle gibt, wird der Wert aus der Vorwelle übernommen. Beim Fortschreiben ist es zu kleineren Fehlern gekommen. Bei der <i>beruf1</i> mussten 19 Fälle, bei der <i>beruf2</i> 34 Fälle korrigiert werden.

Tabelle 34: Übersicht über rückwirkende Korrekturen in Spelldatensätzen als Basis für den integrierten Bio-Spelldatensatz (bio_spells)

Geänderte Variable	Betroffener Datensatz	Geänderte Welle	Art der Änderung	Beschreibung der Änderung
<i>egp</i> <i>esec</i>	<i>et_spells</i>	2-3	Korrektur	Ab W2 wird StiB (einfache Einteilung) mit veränderter Skala abgefragt im Vgl. zu W1 (W2 ff: PET1510 bzw. ET060*; W1: PET1500). Dies wurde in den <i>et_spells</i> bei der Bildung von <i>egp</i> und <i>esec</i> nicht beachtet. Da StiB eine Quelle bei der Generierung dieser beiden Indikatoren ist, kam es zu Fehlzusammenordnungen.
<i>zensiert</i>	<i>lu_spells</i>	2-3	Korrektur	Spells von Zielpersonen mit einem zensierten LU-Spell aus W2, aber ohne überlappenden Spell gleichen Typs in W3, die in W3 befragt wurden, wurden in Zensierungsvariable mit Sondercode -5 geschlossen.
<i>zensiert</i>	<i>lu_spells</i> , <i>et_spells</i> , <i>al_spells</i>	2-3	Korrektur	Andauernde Spells aus W2, bei denen die ZP in W3 mit dem Senioirefragebogen befragt wurde, wurden in der Zensierungsvariable mit Sondercode -5 geschlossen.

Tabelle 35: Übersicht über rückwirkende Veränderungen in Registerdatensätzen (*hh_register*; *p_register*)

Geänderte Variable	Betroffener Datensatz	Geänderte Welle	Art der Änderung	Beschreibung der Änderung
<i>bgbezb3</i>	<i>p_register</i>	3	Korrektur	Es gab deutliche Unterschiede in der Aufbereitungslogik zwischen Welle 3 einerseits und den Vorwellen andererseits. So wurde z.B. kein Abgleich mehr mit <i>alg2abez</i> durchgeführt. Zudem wurden die Codes -1, -2 und -4 aus den Variablen <i>AL20602</i> und <i>AL20702*</i> uneinheitlich weiterverarbeitet. Die Aufbereitungslogik wurde nachträglich dem Vorgehen aus Welle 2 angeglichen.

Tabelle 36: Übersicht über rückwirkende Veränderungen in den Gewichtungsdatensätzen (hweights; pweights)

Geänderte Variable	Betroffener Datensatz	Geänderte Welle	Art der Änderung	Beschreibung der Änderung
<i>ppleib</i>	pweights	1	Korrektur	Bei der Erstellung der <i>ppleib</i> -Variable wurde bei den Bleibewahrscheinlichkeiten von Personen von Welle 1 auf Welle 2 die Ausfälle durch Verweigerung der Panelbereitschaft nicht berücksichtigt. Dies ist in der Auslieferung des SUF Welle 4 korrigiert. Für die Korrektur der Variable <i>ppleib</i> wurde wegen der geringen Ausfallrate anstelle eines Propensity-Modells als konstanter Faktor die mittlere Ausfallrate (Propensity-Modell nur mit der Konstanten) auf Grund von Verweigerungen der Panelbereitschaft von Welle 1 zu Welle 2 verwendet. Die Variable <i>ppleib</i> wurde also konstant mit dem Faktor 1,059 (18954 Teilnehmer der ersten Welle geteilt durch 17900 panelbereite Teilnehmer der ersten Welle) multipliziert.

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*⁴⁸, *bgtyp*⁴⁹), Indikatorvariablen zu Partnern im Haushalt (*apartner*, *epartner*⁵⁰), Zeigervariablen auf Eltern und Partner im Haushalt (*zmhh*; *zvhh*; *zparth*⁵¹) sowie verschiedene Indikatorvariablen, die anzeigen ob Eltern (*mhh*; *vhh*⁵²) oder Kinder der Zielperson (z. B. *ekind*⁵³) im Haushalt leben.

⁴⁸ Enthalten im Haushaltsdatensatz (*HHENDDAT*), siehe Kapitel 4.5.2.

⁴⁹ Wellenspezifische Variablen enthalten im Personenregister (*p_register*), siehe Kapitel 4.4.

⁵⁰ Enthalten im Personendatensatz (*PENDDAT*), siehe Kapitel 4.4.

⁵¹ Wellenspezifische Variablen enthalten im Personenregister (*p_register*), siehe Kapitel 4.4.

⁵² Enthalten im Personendatensatz (*PENDDAT*), siehe Kapitel 4.4.

⁵³ Enthalten im Personendatensatz (*PENDDAT*), siehe Kapitel 4.4.

Tabelle 37 gibt einen Überblick über die betroffenen Variablen und die vorgenommenen Anonymisierungsschritte⁵⁴ im Personendatensatz. In Tabelle 38 sind die anonymisierten Variablen des Erwerbstätigkeitsspelldatensatzes dargestellt.

Tabelle 37: Übersicht über die in Welle 4 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 PET1240 wurden der entsprechenden Beamtenkategorie zugeschlagen. Die Variable für Berufs- und Zeitsoldaten PET1240 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 PET1240 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 W2 (anon.)	Vorgehen analog zu <i>PET1210</i> .

⁵⁴ Falls für Ihr Forschungsvorhaben die Verwendung der nicht anonymisierten Versionen einer oder mehrere 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 37: Übersicht über die in Welle 4 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> .

Tabelle 37: Übersicht über die in Welle 4 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> .

Tabelle 37: Übersicht über die in Welle 4 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 38: Übersicht über die in Welle 4 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 der 1. bis 3. Welle wurden die Arbeitslosengeld-II-Bezüge auf Haushaltsebene in Episodenform erfasst. Dieses Konzept wurde in Welle 4 mit einem leicht überarbeiteten Fragenprogramm fortgeführt. Neben Formulierungsänderungen wurde die Frage nach den Gründen für den Beginn des Arbeitslosengeld-II-Bezugs neu aufgenommen (*AL20550a* bis *AL2055h* bzw. mit zugecodeten offenen Angaben *AL20551a* bis *AL20551h*).

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 4 erstmals befragten Haushalte der Auffrischungsstichprobe wurden nach dem Bezug von Arbeitslosengeld II in der Zeit seit der letzten Veränderung des Haushalts gefragt. Lag diese Veränderung vor Januar 2008 oder wurde dazu keine Angabe gemacht, wurden die Arbeitslosengeld-II-Bezüge des Haushalts seit Januar 2008 erhoben.

5.6.2 Struktur des Arbeitslosengeld-II-Spelldatensatzes

Durch die Integration der in Welle 4 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 3 nicht mehr erhobenen Variablen.

1. Auch in Welle 4 wurden mit *AL20603*, *AL20703a bis AL20703o*, *AL20803* und *AL20903* neue, wellenspezifische Querschnittvariablen in den Arbeitslosengeld-II-Spelldatensatz eingefügt. Diese Variablen sind auf den Interviewzeitpunkt der 4. Welle bezogen. Für die Befragungszeitpunkte der Vorwellen existieren ebenfalls Querschnittvariablen, in denen die auf die jeweilige Welle bezogenen, analogen Informationen abgelegt sind. Tabelle 39 gibt einen Überblick über die im Arbeitslosengeld-II-Spelldatensatz enthaltenen Querschnittinformationen.

Tabelle 39: Querschnittvariablen im ALG2-Spelldatensatz (*alg2_spells*)

	Querschnittvariable mit Information bezogen auf			
	Welle 1	Welle 2	Welle 3	Welle 4
Erhält der HH ALG2 für alle HH-Mitglieder?	<i>AL20600</i>	<i>AL20601</i>	<i>AL20602</i>	<i>AL20603</i>
Erhält der HH ALG2 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>
Höhe des monatlichen ALG2-Bezugs?	<i>AL20800</i>	<i>AL20801</i>	<i>AL20802</i>	<i>AL20803</i>
Hat eine Kürzung des ALG2 begonnen?	<i>AL20900</i>	<i>AL20901</i>	<i>AL20902</i>	<i>AL20903</i>

2. Eingebettet in den Arbeitslosengeld-II-Bezugsepisoden finden sich Informationen zu Zeiten, in denen Kürzungen des Leistungsbezugs stattgefunden haben. Bis Welle 3 kamen innerhalb einer Bezugsepisode bis zu sechs Kürzungen vor. Im Rahmen der Fortschrei-

bung der Arbeitslosengeld-II-Bezüge, die in der Vorwelle zensiert waren, werden auch Informationen zu seit dem letzten Interview neu begonnenen Kürzungen erhoben. Diese neuen Kürzungen werden in die bestehenden, fortzuschreibenden Arbeitslosengeld-II-Episoden übertragen. Da die bestehende maximale Anzahl von sechs Kürzungen pro Bezugsepisode durch die erneute Fortschreibung nicht mehr ausreichte, wurden zwei zusätzliche Kürzung eingefügt, die die Kürzel „g“ und „h“ tragen⁵⁵. Des Weiteren sind bei Frage 22200 zwei neue Items hinzugekommen (AL22200g und AL22200h; inklusive offener Vercodung AL22201g und AL22201h). Darüber hinaus entspricht die Datenstruktur der Datenstruktur aus Welle 3.

3. Der Kürzungsgrund *AL21900a bis AL21900e* wurde auch in Welle 4 nicht erhoben. Den Variablen *AL21901a bis AL21901h* wurden entsprechend auch keine offenen Angaben mehr zugecodet.

5.6.3 Plausibilitätsprüfung und –bereinigung des Arbeitslosengeld-II-Spelldatensatzes

Wie in Welle 1 bis 3 wurden die Informationen zu Arbeitslosengeld-II-Bezügen auch in der 3. Welle 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 Datumvariablen (*bmonat; bjahr; emonat; ejahr*) des Arbeitslosengeld-II-Spells, der Kürzungsepisoden (*alg2kbn; alg2kbn; alg2kbn; alg2kbn*) sowie im Zensierungsindikator des Arbeitslosengeld-II-Spells (zensiert) vorgenommen. Konnten Unplausibilitäten nicht durch Korrektur der Datumvariablen 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 4 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 4 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 4 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 Fortschrei-

⁵⁵ Die zu einer Kürzung gehörenden Variablen sind über einen Buchstaben am Ende der Variable zu erkennen. Kürzungsvariablen die sich auf die erste Kürzung beziehen, enden mit einem „a“, die auf die zweite Kürzung bezogenen mit einem „b“ usw.

bung von in der Vorwelle zensierten Spells werden dem Spelldatensatz neue, in Welle 4 berichtete Spells zugespielt (3). Diese drei Varianten sollen im Folgenden kurz dargestellt werden:

1. *Fälle, in denen der Haushalt in Welle 4 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.⁵⁶ Das generierte 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 4 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 4. Welle berichteten Kürzungsepisoden sowie die auf die Welle 4 bezogenen Querschnittinformationen (*AL20603*; *AL20703a bis AL20703o*, *AL20803*, *AL20903*) übernommen.

3. *In Welle 4 neu berichtete Arbeitslosengeld-II-Spells, die keine in der Vorwelle zensierte Episode fortschreiben.*

In Welle 4 neu berichtete Spells wurden dem Arbeitslosengeld-II-Spelldatensatz angehängt. Danach wurde der Spellzähler neu generiert, um eine lückenlose Variable *spellnr* zu schaffen.

⁵⁶ Hier gilt dasselbe. Verändert wird nur der Zensierungsindikator. Das berichtete Enddatum, die Nachfrage für andauernde Spells und das generierte Enddatum bleiben unverändert.

5.6.5 Datenedition in Folge fehlerhafter Preloadeinspielung

In Welle 4 kam es zu einem Fehler bei der Übertragung des Preloads in die zur Fragenbogensteuerung genutzte Datenbank. Die ursprünglich in die Datenbank übertragene Preload-Version musste noch einmal aktualisiert werden. Dabei kam es zu einem Fehler, in dessen Folge die Variable *alg2bez_h_vw* (aktueller ALG2-Bezug des HH in VW) nicht oder falsch aktualisiert wurde.

In der Folge wurde eine bestimmte Gruppe (372 Haushalte) in den Welle 4-Interviews mit einem falschen Preload konfrontiert. Diese Fälle hatten in der Vorwelle keinen aktuellen ALG2-Bezug. Trotzdem wurde in Welle 4 fälschlich die ALG2-Fortschreibungsfrage gestellt.

Betroffen sind Fälle, die in Welle 3 einen Aufsatzspell der Vorwelle fortgeschrieben und beendet haben und bei denen kein zum Befragungszeitpunkt Welle 3 andauernder ALG2-Spell vorlag.

In diesem Zusammenhang wurden für Spells Konventionen der Datenaufbereitung wie folgt getroffen, welche über die Variable *fallgruppe* im Datensatz ermittelt werden können:

a.) Aufsatzspell widersprochen (Fallgruppe 1)

Diese Spells wurden nicht in den ALG2-Spell Datensatz übernommen, da ein Widerspruch vorliegt.

b.) Aufsatzspell fortgeschrieben und bis Interviewdatum Welle 4 durchgängig andauernd (Fallgruppe 2)

Hier wurde die fälschliche Angabe aus dem Preload fortgeschrieben, da kein Widerspruch stattgefunden hat. Den Angaben am aktuellen Rand der Befragung wurde hier erst einmal vertraut. Daraus ergibt sich, dass das Spellende in diesem Fall als unproblematisch angesehen wird. Fraglich ist allerdings der Spellbeginn. Bei der Korrektur gab es folgende Varianten:

- a. Fortschreibung des falschen Aufsatzspells akzeptieren (d.h. Spell wird Teil des ALG2-Spell Datensatzes)
- b. Setzung für das generierte Beginndatum (*bmonat*, *bjahr*): Interviewdatum der Vorwelle
- c. Setzung für das berichtete Beginndatum (*AL20100*, *AL20200*): -4

c.) Aufsatzspell fortgeschrieben und vor Interviewdatum Welle 4 beendet mit weiteren ALG2-Spells (Fallgruppe 3) bzw. ohne weitere ALG2-Spells (Fallgruppe 4)

In diesen Fällen muss befürchtet werden, dass es sich um künstliche Spells handelt, die tatsächlich nicht vorgekommen sind. Im Gegensatz dazu wird in Fallgruppe 2 nur ein künstliches Beginndatum vermutet. In Fällen, in denen der fälschlich fortgeschriebene Aufsatzspell direkt im Interviewdatum der Vorwelle oder früher wieder endete, wurde davon ausgegan-

gen, dass es sich um einen nicht korrekt aufgenommenen Widerspruch handelt. Diese Spells wurden entsprechend entfernt.

In Fällen, in denen der fälschlich fortgeschriebene Aufsatzspell nach dem Interviewdatum der Vorwelle endete, ist unklar, ob es sich um einen künstlichen Spell handelt. Diese Spells bleiben deshalb im Datensatz erhalten:

- Setzung für das generierte Beginndatum (bmonat, bjahr): Interviewdatum der Vorwelle
- Setzung für das berichtete Beginndatum (AL20100, AL20200): -4

d.) Aufsatzspell nicht fortgeschrieben, da ALG2-Block nicht angesteuert

In diesen 17 Spells wurde die Zensierung in der Variable *zensiert* mit Sondercode -6 geschlossen.

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 Erwerbsbiographie. Für Personen, die in Welle 4 erstmals nach ihrer Erwerbsbiografie gefragt wurden, wurde das Referenzdatum für den Beginn des Retrospektivintervalls angepasst. In Welle 4 sollten hier alle Erwerbstätigkeits- und Arbeitslosigkeitsepisoden seit Januar 2008 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 4. Welle wurden neben andauernden ET- und AL-Spells aus der Vorwelle auch erstmals Lückenspells fortgeschrieben. Für die Fortschreibung der in der Vorwelle andauernden und damit im Spelldatensatz rechtszensierten Spells sind in den Personenfragebögen Dependent-Interviewing-Fragen enthalten. Es konnten bis zu zwei ET-Spells und ein AL-Spell aus der Vorwelle fortgeschrieben werden. Bei Befragten mit mehr als zwei zum Interviewzeitpunkt andauernden ET-Spells wurden jeweils die Erwerbstätigkeiten mit dem höchsten Stundenumfang fortgeschrieben.

5.7.2 Struktur des BIO-Spelldatensatzes

Der BIO-Spelldatensatz orientiert sich hinsichtlich seiner Struktur an den modularen Spelldatensätzen ET, AL und LU. Die in Welle 2 und 3 modular erhobenen Informationen wurden mit den integriert erhobenen Informationen der Welle 4 zusammengeführt. ET-spezifische Variablen haben dabei im BIO-Spelldatensatz ihren Variablennamen im Ver-

gleich zum ET-SUF der Welle 3 behalten, analog auch AL- bzw. LU-spezifische Variablen. Zwischen ET, AL und LU gleiche Variablen wurden 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 wurde eine Variable für den Aktivitätstyp (spelltyp) und für die Spellintegration (spintegr) aufgenommen sowie eine übergreifende Spellnummer generiert (spellnr) und die vorhandenen Spellnummern umbenannt (spellnret, spellnral, spellnrlu) .

Durch die Integration der in Welle 4 berichteten Erwerbstätigkeits- und Arbeitslosigkeitsphasen in den BIO-Spelldatensatz werden neue ET- bzw. der 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-Spelldatensatz werden die ET-spezifischen Variablen *ET0600 bis ET2200* als wellenspezifische, auf Welle 2 bezogene Querschnittsinformationen verstanden, die Variablen *ET0601 bis ET2201* auf Welle 3 bezogene und die neuen Variablen *ET0552 bis ET2202* auf Welle 4 bezogene Querschnittsinformationen. Tabelle 40 gibt einen Überblick über die im BIO-Spelldatensatz enthaltenen ET-spezifischen Querschnittsinformationen.

Tabelle 40: ET-spezifische Querschnittvariablen im BIO-Spelldatensatz (bio_spells)

	Querschnittvariable mit Information bezogen auf ...		
	Welle 2	Welle 3	Welle 4
Stellung im Beruf (einfache und detaillierte Einteilung)	<i>ET0600</i>	<i>ET0601</i>	<i>ET0502</i>
	<i>ET0700</i>	<i>ET0701</i>	<i>ET0602</i>
	<i>ET0800</i>	<i>ET0801</i>	<i>ET0702</i>
	<i>ET1000</i>	<i>ET1001</i>	<i>ET0802</i>
	<i>ET1100</i>	<i>ET1101</i>	<i>ET1002</i>
	<i>ET1200</i>	<i>ET1201</i>	<i>ET1102</i> <i>ET1202</i>
Aufsichtungsfunktion; Zahl der beaufsichtigten Arbeitskräfte	<i>ET1300</i>	<i>ET1301</i>	<i>ET1302</i>
	<i>ET1400</i>	<i>ET1401</i>	<i>ET1402</i>
Entfristung einer anfangs un- fristeten Tätigkeit	<i>ET1700</i>	<i>ET1701</i>	<i>ET1702</i>
Arbeitszeit (vertraglich; tatsächlich; durch- schnittlich bei unregelmäßiger Arbeitszeit)	<i>ET2000</i>	<i>ET2001</i>	<i>ET2002</i>
	<i>ET2100</i>	<i>ET2101</i>	<i>ET2102</i>
	<i>ET2200</i>	<i>ET2201</i>	<i>ET2202</i>

Im BIO-Spelldatensatz gibt es darüber hinaus eine AL-spezifische Variable, die als wellenspezifische Querschnittsinformation zu verstehen ist (AL1300 für Welle 2; AL1301 für

Welle 3 und AL1302 für Welle 4). Tabelle 41 gibt einen Überblick über die im Spelldatensatz enthaltenen Querschnittinformationen.

Tabelle 41: Querschnittvariablen im AL-Spelldatensatz (*al_spells*)

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

- Die nicht wellenspezifische ET-Variable *ET2700* (Bruttoeinkommen bei beendeter Tätigkeit) ist in Welle 4 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 und Welle 3. Die Prüfungen fanden analog zu Welle 3 nur innerhalb eines Spelldatentyps statt. Spelldatentypübergreifende Prüfungen wurden nicht durchgeführt. Bereinigungen und Umcodierungen wurden, wie bei den Arbeitslosengeld-II-Spelldaten, 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 4 berichteten Episoden ins Spelldatensatzformat 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 4 erhobenen Informationen fortgeschrieben.

Dabei sind drei Varianten zu unterscheiden, bei denen in der ersten (1) lediglich der Zensurindikator *zensiert* verändert wird. Die zweite Variante (2) stellt eine Fortschreibung des in der Vorwelle zensierten Spells mit Welle 3-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 4 berichtete Spells zugespielt (3). Diese drei Varianten sollen im Folgenden kurz dargestellt werden:

1. *Fälle, in denen die Person in Welle 4 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 unverändert⁵⁷. 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 4 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 4 erhobenen Informationen überschrieben. Weiterhin wurden die auf die Welle 4 bezogenen Querschnittinformationen (*ET0552* bis *ET2202*) übernommen.

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 4 erhobenen Informationen überschrieben. Weiterhin wurden die auf die Welle 4 bezogene Querschnittinformation (*AL1302*) über-

⁵⁷ 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.

nommen. Bei den AL-Spellldaten 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 4 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 4 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 4 neu berichtete Spells, die keine in der Vorwelle zensierte Episode fortschreiben.

In Welle 4 neu berichtete Spells wurden dem BIO-Spellldatensatz angehängt. Danach wurde der Spellzähler neu generiert, um eine lückenlose Variable *spellnr* zu schaffen.

Das Fortschreiben der Spellldatensätze hat keine Auswirkung auf die Modul-spezifischen Spellnummern der bereits im SUF zu Welle 3 enthaltenen Spells (*spellnret*, *spellnral*, *spellnrlu*). Diese Spells behalten ihre Spellnummer. Die neuen Spells aus Welle 4 werden an den entsprechenden Datensatz angehängt und die Spellnummer fortgeschrieben.

5.8 Ein-Euro-Job-Spellldatensatz (*ee_spells*)

In der 4. Welle wurde das Konzept zur Erhebung der Maßnahmeteilnahme grundlegend überarbeitet. An Stelle des MN-Spellldatensatzes tritt ab Welle 4 der Ein-Euro-Spellldatensatz (*ee_spells*). Als Referenzzeitpunkt, ab dem Ein-Euro-Jobs erfragt wurden, diente Januar 2009.

5.8.1 Struktur des EEJ-Spell Datensatzes

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

5.8.2 Plausibilitätsprüfung und –bereinigung des EEJ-Spell Datensatzes

Der EEJ-Spell Datensatz zur Teilnahme an Ein-Euro-Jobs wurde plausibilitätsgeprüft und bereinigt. Die Plausibilitätsprüfungen beinhalteten Prüfungen der Zeitangaben, des Referenzzeitpunktes (1/2009) sowie Prüfungen auf logische Inkonsistenzen bei Befragten mit mehreren Ein-EuroJobSpells.

Bereinigungen und Umcodierungen wurden nur in den generierten Datumsvariablen vorgenommen. Jahreszeitangaben wurden in konkrete Monatsangaben umkodiert, „-8“ Werte für unplausible Angaben gesetzt oder Datumangaben ersetzt oder plausibilisiert. Danach wurde ein Spellzähler *spellnr* generiert. Die Generierung wurde analog zu den chronologischen Zählern in den BIO-Spell Datensätzen vorgenommen, wobei nicht teilgenommene Spells bei der Sortierung nicht miteinbezogen wurden und somit ihre ursprüngliche *spellnr* beibehalten haben.

6 Gewichtung Welle 4

Die Konstruktion der Gewichte für die vierte Welle erfolgte im Wesentlichen analog zur dritten Welle (vgl. Berg et.al, FDZ Datenreport 06/2010). Im Gegensatz zu Welle 3 wurde bei der Kalibrierung der Welle 4 wieder das GREG-Verfahren verwendet, was für die Gewichte allerdings keine wesentliche Bedeutung hat⁵⁸ (Welle 3 wurde mit dem Raking-Verfahren (IPF) kalibriert).

Ausgangspunkt für die Gewichtung der vierten Welle und des Längsschnitts von Welle 3 auf Welle 4 waren nunmehr die Querschnittgewichte aus Welle 3 für Haushalte bzw. Personen. Die zwei Gewichte jedes Haushalts und die zwei Gewichte jeder Person wurden erneut fortgeschrieben.

6.1 Designgewichte für die Bestandshaushalte der 4. Welle

Unter Berücksichtigung von Zuzügen aus dem Inland wurden für die vierte Welle nunmehr aus den Haushaltsquerschnittgewichten der dritten Welle neue „Haushaltsdesigngewichte“ für die 4. Welle generiert. Dies erfolgte erneut über das für Welle 2 beschriebene „Weight-Share“-Verfahren. Geburten, Todesfälle oder Auszüge haben keinen Einfluss auf das Gewicht, Zuzüge aus dem Inland erhöhen jedoch die Inklusionswahrscheinlichkeit eines Haus-

⁵⁸ Die Gewichte bei beiden Verfahren sind in der Regel sehr ähnlich. Beim Raking lassen sich aber im Gegensatz zu GREG keine korrigierten Standardfehler schätzen.

halts, 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 Querschnittgewicht w_{i,hh_3} :

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

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

6.2 Designgewichte für die Zugangsstichprobe der 4. Welle

In der vierten Welle wurde das Panel erneut durch Nachziehung von Neuzugängen zum Leistungsbezug aufgefrischt. Eine Ziehungschance hatten alle Haushalte, die zwar im Juli 2009 Leistungen bezogen haben, im selben Monat 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 beiden Jahre Leistungen bezogen hatte. Nachgezogen wurde wiederum in den 300 Points der ersten Welle. 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 $sample=5$ in der Variable dw_ba .

6.3 Wiederteilnahme-Propensity Haushalte

In diesem Schritt wurde wiederum analog zum Vorgehen für Welle 3 für jeden Haushalt, der in der dritten Welle teilgenommen hatte, die Wahrscheinlichkeit zur Wiederteilnahme an Welle 4 über Logit-Modelle für Panelbereitschaft, Erreichbarkeit und Teilnahme geschätzt. Auch Haushalte mit letzter Teilnahme in Welle 2, ohne Teilnahme in Welle 3 (temporäre Ausfälle), wurden in der Modellierung für Welle 4 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. 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 ersten Welle für den gesamten möglichen Zeitraum $[t_1; t_2; t_3, t_4]$ zwischen allen vier 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.

Tabelle 42: 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
sprache_nichtdeutsch Referenzkategorie	Sprache HBV: überwiegend im Haushalt gesprochene Sprache nicht-deutsch Sprache HBV: überwiegend im Haushalt gesprochene Sprache deutsch
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 DinvalidAge 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 Nicht auswertbare Altersangaben 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	ALG2-Bezug des Haushalts: bezieht aktuell ALG2 ALG2-Bezug des Haushalts: bezieht aktuell kein ALG2
stichprobe1 stichprobe3 stichprobe4 Referenzkategorie	BA-Stichprobe Auffrischungsstichprobe (BA) Welle 2 Auffrischungsstichprobe (BA) Welle 3 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 42: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der Wiederholerhaushalte (Fortsetzung)

Variablenkürzel und Referenzkategorie	Erläuterung
blneualt_2 Referenzkategorie	Neue Bundesländer Alte Bundesländer
bundesld_1	Bundesland: Schleswig-Holstein
bundesld_2	Bundesland: Hamburg
bundesld_3	Bundesland: Niedersachsen
bundesld_4	Bundesland: Bremen
bundesld_6	Bundesland: Hessen
bundesld_7	Bundesland: Rheinland-Pfalz
bundesld_8	Bundesland: Baden-Württemberg
bundesld_9	Bundesland: Bayern
bundesld_10	Bundesland: Saarland
bundesld_11	Bundesland: Berlin
bundesld_12	Bundesland: Brandenburg
bundesld_13	Bundesland: Mecklenburg-Vorpommern
bundesld_14	Bundesland: Sachsen
bundesld_15	Bundesland: Sachsen-Anhalt
bundesld_16	Bundesland: Thüringen
Referenzkategorie	Bundesland: Nordrhein-Westfalen
bik_1	BIK-Größenklasse der Gemeinde: unter 2.000 Einwohner
bik_2	BIK-Größenklasse der Gemeinde: 2.000 bis u. 5.000 Einwohner
bik_3	BIK-Größenklasse der Gemeinde: 5.000 bis u. 20.000 Einwohner
bik_4	BIK-Größenklasse der Gemeinde: 20.000 bis u. 50.000 Einwohner
bik_5	BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner STYP 2/ 3/ 4
bik_6	BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner STYP 1
bik_7	BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner STYP 2/ 3/ 4
bik_8	BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner STYP 1
bik_9	BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner STYP 2/ 3/ 4
Referenzkategorie	BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner STYP 1

Tabelle 43: Logit-Modelle zur Wiederteilnahme für Panelbereitschaft, Erreichbarkeit und Teilnahme

	Panelbereitschaft		Kontakt		Teilnahme	
	Coef.	p	Coef.	p	Coef.	p
alter_1	.2449415	0.048	-.563471	0.000	-.5253855	0.000
alter_2	.2266293	0.038	-.3007197	0.053	-.2684374	0.002
alter_4	-.1421211	0.119	.4448706	0.009	.3424315	0.000
alter_5	-.0611345	0.760	.6635783	0.196	.1243719	0.605
sex_1	.1283306	0.059	-.2378574	0.032	-.1962515	0.002
nichtdeutsch	-.0153975	0.926	.0860101	0.714	-.1611078	0.194
sprache_nichtdeutsch	.1221601	0.421	.02652	0.902	-.1991273	0.078
schulbil_1	.0868956	0.621	-.3618542	0.100	-.440921	0.001
schulbil_2	.006146	0.940	-.1451147	0.272	-.077809	0.297
schulbil_4	-.1066143	0.220	-.0391485	0.797	.2277794	0.008
gesundheit_3	.051909	0.498	-.0757783	0.547	.0519512	0.467
gesundheit_4	-.0000444	1.000	.1316258	0.433	.1421355	0.129
gesundheit_5	.221958	0.160	-.2521271	0.214	-.0532613	0.674

Tabelle 43: Logit-Modelle zur Wiederteilnahme für Panelbereitschaft, Erreichbarkeit und Teilnahme /Fortsetzung)

	Panelbereitschaft		Kontakt		Teilnahme	
	Coef.	p	Coef.	p	Coef.	p
zufrieden_1	.0539952	0.791	-.3393382	0.209	-.1988758	0.241
zufrieden_2	-.0645234	0.596	.0201944	0.921	-.1162082	0.309
zufrieden_3	.0124931	0.903	.0522408	0.778	-.0122526	0.904
anz_0_3	.050641	0.645	-.0276337	0.851	-.1646145	0.033
anz_4_6	.0102969	0.931	.1046487	0.565	.1133878	0.231
anz_7_14	-.0768915	0.202	-.0787451	0.433	.0555317	0.307
anz_65	-.2021364	0.073	.1684241	0.586	.2093217	0.137
DinvalidAge	-.2654724	0.327	-.125904	0.804	-.348494	0.133
eigentum	-.506994	0.000	.7619415	0.000	.1077725	0.242
wnka_1	.1641973	0.029	.0567571	0.648	.1499191	0.033
wnka_3	-.2358067	0.011	-.1162425	0.412	-.169655	0.040
hhincome_1	.0008659	0.994	-.741978	0.000	.1242989	0.203
hhincome_2	-.0202731	0.833	-.2303618	0.161	.1538367	0.075
hhincome_4	.14235	0.140	.1749497	0.392	.079187	0.424
alg2_1	.2756459	0.003	.1676447	0.197	.0346835	0.653
stichprobe1	.2533853	0.009	.1790433	0.246	-.3167451	0.000
stichprobe3	.577445	0.001	.0320593	0.881	-.2583328	0.044
stichprobe4	.012362	0.923	-.1104379	0.544	-.4345083	0.000
blneualt_2	-.1545256	0.043				
bundesld_1			-.3111224	0.259	.0005076	0.997
bundesld_2			1.942047	0.057	.4931614	0.114
bundesld_3			-.6117696	0.001	.1431391	0.238
bundesld_4			.0929839	0.863	-.2019436	0.491
bundesld_6			.0939481	0.712	-.3272168	0.009
bundesld_7			-.4709838	0.089	-.2198406	0.183
bundesld_8			.1280441	0.599	.0518479	0.683
bundesld_9			-.2806843	0.140	-.0561062	0.612
bundesld_10			.4919014	0.418	-.2852845	0.243
bundesld_11			.5878931	0.048	-.0696639	0.633
bundesld_12			-.3199865	0.256	.1060142	0.527
bundesld_13			.0702071	0.845	.0033396	0.986
bundesld_14			-.3191026	0.189	.2199778	0.161
bundesld_15			-.1700517	0.508	.0787719	0.602
bundesld_16			.114062	0.741	.0351705	0.846
bik_1			.1152843	0.831	.0188483	0.952
bik_2			.0849133	0.794	.0916075	0.651
bik_3			.3005287	0.212	-.0438891	0.747
bik_4			.0311147	0.883	-.0084268	0.949
bik_5			.0469815	0.836	.0105957	0.939
bik_6			1.005097	0.016	.0036167	0.985
bik_7			.1692014	0.392	-.2447706	0.031
bik_8			.123732	0.452	-.1187737	0.239
bik_9			.5941939	0.036	.0476738	0.743
anzkon_1			-.6760014	0.000		
anzkon_3			-.0374	0.798		
anzkon_4			-.5533189	0.000		
cons	1.966568	0.000	3.549948	0.000	1.863116	0.000
n	9526		8436		8044	
Log likelihood	-3272.2333		-1451.5222		-3543.0662	
PseudoR2	0.0342		0.0847		0.0471	

6.4 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 3 entstandener Split-Haushalt (Split-W3 Haushalte) und in Welle 4 entstandener Split-Haushalt (Split-W4 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 44: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der erstbefragten Split-Haushalte (W3 und W4)

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 Referenzkategorie	BA-Stichprobe Auffrischungsstichprobe (BA) Welle 2 Auffrischungsstichprobe (BA) Welle 3 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 45: Logit-Modelle zur Erstteilnahme von Split-W3-Haushalten für Erreichbarkeit und Teilnahme

	Kontakt		Teilnahme	
	Coef.	p	Coef.	p
alter_1	-.2418126	0.716	.5429442	0.340
alter_2	-.4146809	0.570	-.1851265	0.791
alter_4	-.7330313	0.198	-.0424319	0.937
alter_5	-1.770686	0.241	2.556612	0.055
sex_1	-.6319621	0.188	.3655047	0.410
nichtdeutsch	-.1078378	0.931	1.012818	0.281
schulbil_1	-.2794609	0.724	-.3160496	0.716
schulbil_2	.47696	0.384	-1.09537	0.060
schulbil_4	.5901841	0.384	.7000542	0.189
anzkon_1	-1.314818	0.011		
anzkon_3	1.649364	0.145		
anzkon_4	.1804791	0.838		
stichprobe1	1.034975	0.035	-.1910267	0.702
stichprobe3	-.2036549	0.782	-.6937177	0.457
cons	1.963171	0.004	-1.323982	0.012
n	187		158	
Log likelihood	-67.599448		-75.077429	
Pseudo R²	0.1621		0.0875	

Tabelle 46: Logit-Modelle zur Erstteilnahme von Split-W4-Haushalten für Erreichbarkeit und Teilnahme

	Kontakt		Teilnahme	
	Coef.	p	Coef.	p
alter_1	-.3103986	0.495	.5612502	0.121
alter_2	-.0615905	0.911	.2041636	0.602
alter_4	-.4011371	0.236	.1618036	0.556
alter_5	1.05171	0.220	1.097297	0.075
sex_1	-.3221642	0.275	-.1990163	0.429
nichtdeutsch	-.1466203	0.837	.0084587	0.988
schulbil_1	-.5262951	0.463	.9209869	0.072
schulbil_2	-.0384243	0.915	.165165	0.548
schulbil_4	.2198667	0.553	.3418422	0.283
anzkon_1	-.3701073	0.235		
anzkon_3	2.284895	0.000		
anzkon_4	2.754416	0.000		
stichprobe1	3.516057	0.000	-.9831306	0.000
stichprobe3	.289362	0.500	-.2668892	0.548
stichprobe4	.0596691	0.902	-.3008521	0.563
cons	.4192136	0.235	-.347527	0.211
n	459		368	
Log likelihood	-153.64726		-226.55592	
Pseudo R²	0.3278		0.0590	

6.5 Ausfallgewichtung für Haushalte aus Auffrischungsstichprobe der 4. Welle

Für die Haushalte aus der Auffrischungsstichprobe wurde erneut eine zweistufige Ausfallmodellierung (Erreichbarkeit und Teilnahme) analog zur Auffrischungsstichprobe der dritten Welle vorgenommen. Die daraus abgeleitete Teilnahmewahrscheinlichkeit befindet sich in der Variable *prop_t0*.

Tabelle 47: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der Auffrischungsstichprobe Welle 4

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
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
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
blneualt_2 Referenzkategorie	Neue Bundesländer Alte Bundesländer

Tabelle 48: Logit-Modelle zur Erstteilnahme für Erreichbarkeit und Teilnahme der Auffrischungsstichprobe Welle 4

	Kontakt		Teilnahme	
	Coef.	p	Coef.	p
alter_1	-.2709841	0.450	-.0748669	0.585
alter_2	.5520724	0.191	.1400829	0.311
alter_4	.0117875	0.975	.2277945	0.118
sex_1	-.1020647	0.701	.0284883	0.776
nichtdeutsch	-.0272298	0.936	-.2228597	0.076
anz_persBG_2	-1.702769	0.063	.3772027	0.264
anz_persBG_3	-1.524367	0.197	.5694042	0.153
anz_verwfBG_1	-.9406263	0.436	.2104376	0.678
anz_verwfBG_3	.1560236	0.821	-.1769682	0.496
anz_verwfBG_4	-.0581056	0.958	-.7962362	0.060
BG_typ_2	1.785789	0.074	-.3529742	0.328
BG_typ_3	1.017784	0.247	-.6112298	0.095
BG_typ_4	.8798209	0.436	-.5367045	0.193
BG_typ_5	1.407635	0.105	-.0245079	0.931
famstand_2	.8228556	0.202	.3322433	0.124
famstand_3	-1.126793	0.186	.2749957	0.497
famstand_4	-.5357952	0.169	-.0970361	0.538
famstand_5	.2135627	0.676	-.0288342	0.863
famstand_6	-.3597392	0.578	.5774257	0.045
blneuallt_2	.4622306	0.120	-.1443263	0.172
anzkon_1	-1.265884	0.008		
anzkon_3	-.2950892	0.584		
anzkon_4	-2.737432	0.000		
cons	4.895443	0.000	-.7310508	0.000
n	2440		2359	
Log likelihood	-297.79633		-1487.1327	
Pseudo R²	0.1622		0.0085	

6.6 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 4 wurden unter Einschluss von zusätzlichen Personenmerkmalen über Logit-Modelle für Panelbereitschaft, Erreichbarkeit und Teilnahme geschätzt. Die über den Haushaltszusammenhang vermittelte Abhängigkeit der Personestichprobe 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 $[t_1; t_2; t_3; t_4]$ über alle vier Wellen lässt sich dann als Produkt aus Querschnittgewicht zu t_1 , *ppbleib* (Welle 1 zu Welle 2) und *ppbleib* (Welle 2 zu Welle 3 usw.) berechnen.

Tabelle 49: 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
sprache_nichtdeutsch Referenzkategorie	Sprache Person: überwiegend im Haushalt gesprochene Sprache nicht-deutsch Sprache Person: überwiegend im Haushalt gesprochene Sprache deutsch
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 DinvalidAge 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 Nicht auswertbare Altersangaben 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	ALG2-Bezug des Haushalts: bezieht aktuell ALG2 ALG2-Bezug des Haushalts: bezieht aktuell kein ALG2
stichprobe1 stichprobe3 stichprobe4 Referenzkategorie	BA-Stichprobe Auffrischungsstichprobe (BA) Welle 2 Auffrischungsstichprobe (BA) Welle 3 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 49: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der Wiederholer (Fortsetzung)

Variablenkürzel und Referenzkategorie	Erläuterung
bneualt_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 50: Logit-Modelle zur Wiederteilnahme für Panelbereitschaft, Erreichbarkeit und Teilnahme

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

	Panelbereitschaft		Kontakt		Teilnahme	
	Coef.	p	Coef.	p	Coef.	p
alter_1	.2435204	0.006	-.3796629	0.003	-.6380499	0.000
alter_2	.1779134	0.085	-.3793937	0.012	-.3230357	0.000
alter_4	-.1375531	0.121	.5048395	0.003	.2456517	0.002
alter_5	.027996	0.882	.5818921	0.164	.1317982	0.438
sex_1	.1429213	0.001	-.1086084	0.206	-.142638	0.001
nichtdeutsch	.0696697	0.677	-.1179752	0.610	-.2664062	0.015
sprache_nichtdeutsch	-.0068279	0.970	.0628587	0.788	-.2947304	0.007
schulbil_1	.1374612	0.420	-.3917476	0.065	-.4822471	0.000
schulbil_2	-.0062376	0.931	-.1263243	0.310	-.1339413	0.030
schulbil_4	-.109026	0.171	-.0526074	0.717	.1356339	0.063
gesundheit_3	.0371439	0.578	-.0974953	0.393	.1225001	0.035
gesundheit_4	.0501367	0.576	.0322809	0.826	.100103	0.189
gesundheit_5	.1338005	0.357	-.2967936	0.124	-.012446	0.910
zufrieden_1	.0677153	0.723	-.3747764	0.174	-.0332241	0.826
zufrieden_2	-.0215315	0.841	.0307697	0.874	-.1119732	0.221
zufrieden_3	.0069942	0.936	.0905009	0.600	.0564085	0.461

Tabelle 50: Logit-Modelle zur Wiederteilnahme für Panelbereitschaft, Erreichbarkeit und Teilnahme (Fortsetzung)

	Panelbereitschaft		Kontakt		Teilnahme	
	Coef.	p	Coef.	p	Coef.	p
anz_0_3	.0798781	0.530	.0133438	0.930	-.1653959	0.026
anz_4_6	-.0104995	0.939	.3446568	0.120	.0843735	0.362
anz_7_14	-.0078977	0.908	-.0451142	0.655	-.0053103	0.916
anz_65	-.246353	0.043	.18392	0.531	-.0071902	0.947
DinvalidAge	-.1436989	0.557	-.2284297	0.619	-.2672323	0.323
eigentum	-.6452986	0.000	.8104465	0.000	.131467	0.134
wnka_1	.1024658	0.152	.0668878	0.585	.1685842	0.005
wnka_3	-.1999418	0.021	-.105062	0.420	-.2232673	0.001
hhincome_1	-.0332209	0.807	-.6980556	0.000	.1603949	0.088
hhincome_2	-.0335545	0.758	-.1987762	0.298	.1456451	0.076
hhincome_4	.1431539	0.175	.2071721	0.353	.0296048	0.746
alg2_1	.1801189	0.107	.1525633	0.283	-.0604414	0.426
stichprobe1	.2394516	0.030	.1340638	0.414	-.2718389	0.002
stichprobe3	.7714802	0.000	-.0910405	0.701	-.1922442	0.145
stichprobe4	-.0358222	0.798	-.1917936	0.316	-.3642542	0.001
blneualt_2	-.1490614	0.078				
anzkon_1			-.7274809	0.000		
anzkon_3			.0244946	0.877		
anzkon_4			-.580114	0.000		
bundesld_1			-.4421672	0.142	.0398692	0.802
bundesld_2			1.253176	0.226	.2258841	0.418
bundesld_3			-.4988578	0.013	-.0247771	0.827
bundesld_4			.0021874	0.997	-.313746	0.233
bundesld_6			.2739805	0.301	-.4146181	0.001
bundesld_7			-.4197543	0.177	-.238903	0.123
bundesld_8			.0759532	0.775	.0657664	0.609
bundesld_9			-.2790896	0.185	-.1254794	0.230
bundesld_10			.4008025	0.534	-.5064959	0.022
bundesld_11			.4743736	0.141	.0279104	0.849
bundesld_12			-.3197083	0.300	.1375164	0.405
bundesld_13			.0319421	0.934	.0353293	0.864
bundesld_14			-.1628645	0.532	-.0117899	0.938
bundesld_15			-.0687428	0.807	.05052	0.732
bundesld_16			.3727655	0.316	.0765401	0.671
bik_1			.0821929	0.889	.045135	0.879
bik_2			-.1168019	0.741	.0455952	0.812
bik_3			.3013398	0.265	-.018835	0.883
bik_4			-.0769723	0.752	-.0342773	0.787
bik_5			-.0814823	0.744	-.0441525	0.739
bik_6			1.052581	0.014	-.2114996	0.237
bik_7			.1470519	0.484	-.2208845	0.045
bik_8			.036178	0.841	-.0282934	0.776
bik_9			.6274816	0.046	.0138276	0.917
cons	2.081648	0.000	3.541147	0.000	1.829531	0.000
n	13439		11912		11439	
Log likelihood	-4573.7302		-1824.7061		-5479.8072	
Pseudo R²	0.0387		0.0828		0.0512	

6.7 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 und 3) zu kombinieren. 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 oder in 7/2009 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 7/2009. Dazu wurde die Inklusionswahrscheinlichkeit in der jeweils anderen Stichprobe für Fälle aus der Microm-Stichprobe (Welle 1) und der neuen Auffrischungsstichprobe (Welle 4) geschätzt. Für Fälle aus der Auffrischungsstichprobe wurde die mittlere Welle 1-Auswahlwahrscheinlichkeit in der Microm-Stichprobe in der jeweiligen Postleitzahl und jeweils die durchschnittliche Teilnahmewahrscheinlichkeit (für W1, W2, W3 und W4) in dieser Stichprobe unterstellt. Für Fälle aus der Microm-Stichprobe wurde, falls es sich (laut Befragungsdaten) um Neuzugänge zum Arbeitslosengeld-II-Leistungsbezug zwischen den drei letzten Stichprobenziehungszeitpunkten (W2, W3, W4) 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.

6.8 Integration temporärer Ausfälle (Haushalte)

In Welle 4 konnten, wie auch schon in Welle 3, 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 zweiten Welle (die dritte Welle ist der temporäre Ausfall).

Für die temporären Ausfälle wurde dann die Wahrscheinlichkeit der Nichtteilnahme in Welle 3 gegeben die Teilnahme in Welle 2 (Nichtteilnahme-Propensities W3) und die Wahrscheinlichkeit der Teilnahme in Welle 4 gegeben die Nichtteilnahme in Welle 3 (Teilnahme-Propensities W4) bestimmt. Die Wahrscheinlichkeit der Nichtteilnahme in Welle 3 ergibt sich über 1-Teilnahmewahrscheinlichkeit in Welle 3.

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 zweiten Welle mit dem Kehrwert dieses Produkts.

Tabelle 51: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der temporären Ausfälle

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
sprache_nichtdeutsch Referenzkategorie	Sprache HBV: überwiegend im Haushalt gesprochene Sprache nicht-deutsch Sprache HBV: überwiegend im Haushalt gesprochene Sprache deutsch
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 DinvalidAge 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 Nicht auswertbare Altersangaben Anzahl Personen im Haushalt im Alter 15 - 64 Jahre
eigentum Referenzkategorie	Art des Wohneigentums: Eigentümer Art des Wohneigentums: Mieter, fehlende Angabe

Tabelle 51: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle der temporären Ausfälle (Fortsetzung)

Variablenkürzel und Referenzkategorie	Erläuterung
wnka_1	Anzahl w.n.- und k.A.-Angaben in Haushalts- und Personeninterview des HBV: keine
wnka_3	Anzahl w.n.- und k.A.-Angaben in Haushalts- und Personeninterview des HBV: 11 und mehr
Referenzkategorie	Anzahl w.n.- und k.A.-Angaben in Haushalts- und Personeninterview des HBV: 1-10
hhincome_1	Haushaltseinkommen: bis 870 Euro
hhincome_2	Haushaltseinkommen: 871-1.400 Euro
hhincome_4	Haushaltseinkommen: mehr als 2.200 Euro
Referenzkategorie	Haushaltseinkommen: 1.401-2.200 Euro
alg2_1	ALG2-Bezug des Haushalts: bezieht aktuell ALG2
Referenzkategorie	ALG2-Bezug des Haushalts: bezieht aktuell kein ALG2
bundesld_1	Bundesland: Schleswig-Holstein
bundesld_2	Bundesland: Hamburg
bundesld_3	Bundesland: Niedersachsen
bundesld_4	Bundesland: Bremen
bundesld_6	Bundesland: Hessen
bundesld_7	Bundesland: Rheinland-Pfalz
bundesld_8	Bundesland: Baden-Württemberg
bundesld_9	Bundesland: Bayern
bundesld_10	Bundesland: Saarland
bundesld_11	Bundesland: Berlin
bundesld_12	Bundesland: Brandenburg
bundesld_13	Bundesland: Mecklenburg-Vorpommern
bundesld_14	Bundesland: Sachsen
bundesld_15	Bundesland: Sachsen-Anhalt
bundesld_16	Bundesland: Thüringen
Referenzkategorie	Bundesland: Nordrhein-Westfalen
bik_1	BIK-Größenklasse der Gemeinde: unter 2.000 Einwohner
bik_2	BIK-Größenklasse der Gemeinde: 2.000 bis u. 5.000 Einwohner
bik_3	BIK-Größenklasse der Gemeinde: 5.000 bis u. 20.000 Einwohner
bik_4	BIK-Größenklasse der Gemeinde: 20.000 bis u. 50.000 Einwohner
bik_5	BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner STYP 2/ 3/ 4
bik_6	BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner STYP 1
bik_7	BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner STYP 2/ 3/ 4
bik_8	BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner STYP 1
bik_9	BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner STYP 2/ 3/ 4
Referenzkategorie	BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner STYP 1

Tabelle 52: Logit-Modelle temporärer Ausfälle

	Wiederteilnahme in Welle 3 zur Ermittlung der Nicht-Teilnahmewahrscheinlichkeit W3 (1-Teilnahmewahrscheinlichkeit W3)		Wiederteilnahme in Welle 4 gegeben Nichtteilnahme in Welle 3	
	Coef.	p	Coef.	p
alter_1	-.3587909	0.000	.1904256	0.329
alter_2	-.0786584	0.271	.0982976	0.598
alter_4	.2331384	0.000	.1199264	0.511
alter_5	1.295449	0.000	.2944528	0.551
sex_1	.0457909	0.350	.0354397	0.785
nichtdeutsch	-.1592276	0.143	.1229105	0.616
sprache_nichtdeutsch	-.4898952	0.000	-.2114027	0.347
schulbil_1	-.0481656	0.689	.0929211	0.736
schulbil_2	.0265179	0.648	-.1115634	0.469
schulbil_4	.1430196	0.027	-.187033	0.294
gesundheit_3	.0007108	0.990	-.0279472	0.853
gesundheit_4	-.0573202	0.426	-.0128668	0.946
gesundheit_5	-.0715361	0.512	.4141463	0.094
zufrieden_1	-.2541332	0.052	.5892071	0.069
zufrieden_2	-.1537822	0.086	.4255118	0.090
zufrieden_3	.0186331	0.811	.0713923	0.759
anz_0_3	.11404	0.109	.3899672	0.017
anz_4_6	.1981789	0.010	.1487108	0.428
anz_7_14	.1236042	0.004	.0697032	0.530
anz_65	-.6790136	0.000	-.4909581	0.073
DinvalidAge	-.2154192	0.536	.6949877	0.273
eigentum	-.2323834	0.000	-.4183435	0.024
wnka_1	.1519274	0.004	.1201478	0.396
wnka_3	-.2447711	0.000	-.3784057	0.037
hhincome_1	.6989807	0.000	.503108	0.010
hhincome_2	.2752072	0.000	.1477986	0.405
hhincome_4	.1809726	0.011	.0016841	0.993
alg2_1	-.0208097	0.738	-.1910293	0.206
bundesld_1	.2074965	0.098	-.0419949	0.896
bundesld_2	-.3541349	0.097	.2277456	0.599
bundesld_3	.1478861	0.102	-.3021676	0.225
bundesld_4	-.0433936	0.875	-1.528863	0.142
bundesld_6	.3530065	0.001	-.3394542	0.293
bundesld_7	-.0222644	0.864	-.0498326	0.874
bundesld_8	-.022292	0.814	-.1916105	0.451
bundesld_9	.345654	0.000	.0054215	0.981
bundesld_10	-.2635286	0.239	-.2208627	0.696
bundesld_11	-.2802475	0.015	-.8326922	0.004
bundesld_12	.1859026	0.142	-.196226	0.594
bundesld_13	.0407544	0.797	-.1760452	0.683
bundesld_14	.1063765	0.340	-.4022964	0.174
bundesld_15	.1239814	0.305	-.0528568	0.859
bundesld_16	.1276419	0.363	-.2424729	0.526

Tabelle 52: Logit-Modelle temporärer Ausfälle (Fortsetzung)

	Wiederteilnahme in Welle 3 zur Ermittlung der Nicht-Teilnahmewahrscheinlichkeit W3 (1-Teilnahmewahrscheinlichkeit W3)		Wiederteilnahme in Welle 4 gegeben Nichtteilnahme in Welle 3	
	Coef.	p	Coef.	p
bik_1	-.1718978	0.389	-.314913	0.547
bik_2	-.1070126	0.456	.2166599	0.529
bik_3	-.1862369	0.077	-.6639829	0.029
bik_4	-.2260286	0.019	-.9719687	0.001
bik_5	-.1548135	0.141	-.3110231	0.242
bik_6	.0426827	0.785	-.4441523	0.281
bik_7	.0438941	0.624	-.3476504	0.138
bik_8	.0090798	0.911	-.1945911	0.316
bik_9	-.0567797	0.601	-.7660066	0.024
cons	.3782731	0.003	-1.92955	0.000
n	8429		2933	
Log likelihood	-5253.1301		-924.46904	
Pseudo R²	0.0355		0.0556	

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:

$$dw_i \cdot h_{temp.Ausfall} * (n_{temp.Ausfall\ i} / (n_{temp.Ausfall\ i} + n_{Bestand\ i})) \text{ für die temporären Ausfälle und}$$

$$dw_i \cdot h_{Bestand} * (n_{Bestand\ i} / (n_{temp.Ausfall\ i} + n_{Bestand\ i})) \text{ für die Bestandsstichprobe.}$$

6.9 Kalibrierung zum Haushaltsgewicht 4. Welle Querschnitt

Anschließend fand wieder eine Kalibrierung der modifizierten Designgewichte inklusive der Ausfallgewichtung auf Haushaltsebene mittels GREG-Verfahren an die Eckwerte des Statistischen Bundesamts für 2009 und für Haushalte mit Leistungsbezug an die Eckwerte der Statistik der Bundesagentur für Arbeit für Juli 2009 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 (380.663) verwendet. Dabei werden diejenigen Fälle der bisherigen Stichproben aus Welle 1, Welle 2 und Welle 3 auf die Eckwerte zum Arbeitslosengeld-II-Bezug der Bundesagentur für Arbeit hochgerechnet, die laut Welle 4 des Surveys im Juli 2009 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 Populations-

werte aus der Stichprobe erwartungstreu schätzen. Weisen die Gewichte allerdings 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.

6.9.1 Kalibrierung der BA-Stichprobe

Die Grundgesamtheit der kumulierten BA-Stichprobe der ersten vier Wellen bilden alle Haushalte in Deutschland mit mindestens einer Bedarfsgemeinschaft mit Leistungen nach dem SGB II zu einem der bisher vier Ziehungszeitpunkte (im Juli 2006, Juli 2007, Juli 2008 oder Juli 2009). Eine Kalibrierung findet in Welle 4 dagegen nur an die Eckwerte der BA-Statistik vom Juli 2009 statt. Damit beeinflusst die Kalibrierung nur die Gewichte der Haushalte aus der BA-Stichprobe, in denen auch im Juli 2009 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 313,86 bis 2630,18. Das Verhältnis der Gesamthochrechnungsfaktoren nach Kalibrie-

zung 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 94,15834 bis maximal 4165,227.

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 Kindern 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 (380.663).

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

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 Bundesländern (16 Kategorien)	Zahl BGs Schleswig-Holstein	135	121.765	121.765
	Zahl BGs Hamburg	79	107.774	107.774
	Zahl BGs Niedersachsen	325	331.822	331.822
	Zahl BGs Bremen	36	50.549	50.549
	Zahl BGs Nordrhein-Westfalen	837	823.814	823.814
	Zahl BGs Hessen	203	218.820	218.820
	Zahl BGs Rheinland-Pfalz	117	122.169	122.169
	Zahl BGs Baden-Württemberg	272	246.337	246.337
	Zahl BGs Bayern	293	265.707	265.707
	Zahl BGs Saarland	52	44.384	44.384
	Zahl BGs Berlin	271	330.193	330.193
	Zahl BGs Brandenburg	179	172.677	172.677
	Zahl BGs Mecklenburg-Vorpommern	105	131.213	131.213
	Zahl BGs Sachsen	252	287.711	287.711
	Zahl BGs Sachsen-Anhalt	245	191.791	191.791
	Zahl BGs Thüringen	128	134.303	134.303
Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Personen unter 65 Jahren 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)	981	1.240.872	1.240.872
	Zahl BGs mit 2 Personen unter 65 (West)	595	480.659	480.659
	Zahl BGs mit 3 Personen unter 65 (West)	427	300.506	300.506
	Zahl BGs mit 4 Personen unter 65 (West)	203	184.448	184.448
	Zahl BGs mit 5 oder mehr Personen unter 65 (West)	143	126.656	126.656
	Zahl BGs mit 1 Person unter 65 (Ost)	554	714.495	714.495
	Zahl BGs mit 2 Personen unter 65 (Ost)	309	276.254	276.254
	Zahl BGs mit 3 Personen unter 65 (Ost)	180	143.684	143.684
	Zahl BGs mit 4 Personen unter 65 (Ost)	90	74.190	74.190
	Zahl BGs mit 5 oder mehr Personen unter 65 (Ost)	47	39.265	39.265

Tabelle 53: 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 Jahren in der Bedarfsgemeinschaft (0, 1, 2, 3, „4 oder mehr“) und nach West/Ost (10 Kategorien)	Zahl BGs ohne Kinder unter 15 Jahren (West)	1554	1.585.891	1.585.891
	Zahl BGs mit 1 Kind unter 15 Jahren (West)	466	407.636	407.636
	Zahl BGs mit 2 Kindern unter 15 Jahren (West)	223	229.110	229.110
	Zahl BGs mit 3 Kindern unter 15 Jahren (West)	68	79.014	79.014
	Zahl BGs mit 4 oder mehr Kindern unter 15 Jahren (West)	38	31.490	31.490
	Zahl BGs ohne Kinder unter 15 Jahren (Ost)	886	926.414	926.414
	Zahl BGs mit 1 Kind unter 15 Jahren (Ost)	187	195.946	195.946
	Zahl BGs mit 2 Kindern unter 15 Jahren (Ost)	84	90.169	90.169
	Zahl BGs mit 3 Kindern unter 15 Jahren (Ost)	15	25.468	25.468
	Zahl BGs mit 4 oder mehr Kindern unter 15 Jahren (Ost)	8	9.891	9.891
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)	592	452.688	452.688
	Rest BGs ohne Alleinerziehenden (West)	1757	1.880.453	1.880.453
	Zahl BGs mit einem/einer Alleinerziehenden (Ost)	202	192.016	192.016
	Rest BGs ohne Alleinerziehenden (Ost)	978	1.055.872	1.055.872

Tabelle 54: Kenngrößen der Verteilung der Gewichte

1%-Perzentil	193,7486
5%-Perzentil	267,6218
10%-Perzentil	331,4976
25%-Perzentil	470,8878
50%-Perzentil	734,9039
75%-Perzentil	1433,385
90%-Perzentil	2257,767
95%-Perzentil	2667,289
99%-Perzentil	2921,039
Mittelwert	1030,394
Standardabweichung	736,9397
Minimum	94,15834
Maximum	4165,227
Fallzahl	3404
Effektivitätsmaß	65,65%

6.9.2 Microm-Stichprobe

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 2394,48 bis 35018,87. Das Verhältnis der Gesamthochrechnungsfaktoren nach Kalibrierung zu getrimmten Designgewichten wurde nach unten auf minimal 0,1 und nach oben auf maximal 3,0 begrenzt. Danach bewegen sich die Gesamthochrechnungsfaktoren nach Kalibrierung im Bereich von minimal 239,4479 bis maximal 56098,93.

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 Kindern besteht, nach West/Ost

Haushalte: Grundlage Mikrozensus 2008:

- 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 55: Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm-Stichprobe, Haushalte)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl BG im Leistungsbezug nach SGB II nach Bundesländern (16 Kategorien)	Zahl BGs Schleswig-Holstein	16	121.765	121.765
	Zahl BGs Hamburg	2	107.774	107.774
	Zahl BGs Niedersachsen	43	331.822	331.778
	Zahl BGs Bremen	6	50.549	50.549
	Zahl BGs Nordrhein-Westfalen	75	823.814	823.858
	Zahl BGs Hessen	18	218.820	218.820
	Zahl BGs Rheinland-Pfalz	13	122.169	122.169
	Zahl BGs Baden-Württemberg	14	246.337	246.337
	Zahl BGs Bayern	40	265.707	265.707
	Zahl BGs Saarland	10	44.384	44.384
	Zahl BGs Berlin	18	330.193	330.193
	Zahl BGs Brandenburg	17	172.677	172.677
	Zahl BGs Mecklenburg-Vorpommern	6	131.213	131.213
	Zahl BGs Sachsen	23	287.711	287.711
	Zahl BGs Sachsen-Anhalt	21	191.791	191.791
	Zahl BGs Thüringen	15	134.303	134.303

Tabelle 55: Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm-Stichprobe, Haushalte) (Fortsetzung 1)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Personen unter 65 Jahren 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)	82	1.240.872	1.240.872
	Zahl BGs mit 2 Personen unter 65 (West)	62	480.659	480.658
	Zahl BGs mit 3 Personen unter 65 (West)	32	300.506	300.506
	Zahl BGs mit 4 Personen unter 65 (West)	31	184.448	184.488
	Zahl BGs mit 5 oder mehr Personen unter 65 (West)	30	126.656	126.617
	Zahl BGs mit 1 Person unter 65 (Ost)	31	714.495	714.495
	Zahl BGs mit 2 Personen unter 65 (Ost)	29	276.254	276.254
	Zahl BGs mit 3 Personen unter 65 (Ost)	24	143.684	143.684
	Zahl BGs mit 4 Personen unter 65 (Ost)	10	74.190	74.190
	Zahl BGs mit 5 oder mehr Personen unter 65 (Ost)	6	39.265	39.265
Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Kinder unter 15 Jahren in der Bedarfsgemeinschaft (0, 1, 2, 3, „4 oder mehr“) und nach West/Ost (10 Kategorien)	Zahl BGs ohne Kinder unter 15 Jahren (West)	151	1.585.891	1.585.891
	Zahl BGs mit 1 Kind unter 15 Jahren (West)	29	407.636	407.636
	Zahl BGs mit 2 Kindern unter 15 Jahren (West)	38	229.110	229.110
	Zahl BGs mit 3 Kindern unter 15 Jahren (West)	10	79.014	79.014
	Zahl BGs mit 4 oder mehr Kindern unter 15 Jahren (West)	9	31.490	31.490
	Zahl BGs ohne Kinder unter 15 Jahren (Ost)	74	926.414	926.414
	Zahl BGs mit 1 Kind unter 15 Jahren (Ost)	12	195.946	195.946
	Zahl BGs mit 2 Kindern unter 15 Jahren (Ost)	6	90.169	90.169
	Zahl BGs mit 3 Kindern unter 15 Jahren (Ost)	7	25.468	25.468
	Zahl BGs mit 4 oder mehr Kindern unter 15 Jahren (Ost)	1	9.891	9.891

Tabelle 55: Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm-Stichprobe, Haushalte) (Fortsetzung 2)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	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)	48	452.688	452.688
	Rest BGs ohne Alleinerziehenden (West)	189	1.880.453	1.880.453
	Zahl BGs mit einem/einer Alleinerziehenden (Ost)	16	192.016	192.016
	Rest BGs ohne Alleinerziehenden (Ost)	84	1.055.872	1.055.872
Anzahl der Haushalte nach Bundesland und BIK-Typ (Schreibweise: "Bundesland.BIK-Typ")	1.1 bis 1.4	17	316.000	316.000
	1.5 bis 1.6	24	149.000	149.000
	1.7 bis 1.8	28	517.000	517.000
	1.9	18	172.000	172.000
	1.10	29	211.000	211.000
	2.10	40	969.000	969.000
	3.2 bis 3.3	39	489.000	489.000
	3.4	25	434.000	434.000
	3.5	25	451.000	451.000
	3.7	101	857.000	857.000
	3.8	59	548.000	548.000
	3.9	54	636.000	636.000
	3.10	34	376.000	376.000
	4.8 bis 4.10	22	358.000	358.000
	5.2 bis 5.3	38	349.000	349.000
	5.4	57	1.052.000	1.052.000
	5.5	71	631.000	631.000
	5.6	25	301.000	301.000
	5.7	72	836.000	836.000
	5.8	156	2.333.000	2.333.000
	5.9	37	363.000	363.000
	5.10	183	2.659.000	2.659.000
	6.2	16	64.000	64.000
	6.3	33	330.000	330.000
	6.4 bis 6.5	20	304.000	304.000
	6.7	40	564.000	564.000
	6.8	25	465.000	465.000
	6.9	50	366.000	366.000
	6.10	46	784.000	784.000
	7.1 bis 7.2	23	305.000	305.000
	7.3	20	177.000	177.000
7.4	14	154.000	154.000	
7.5 bis 7.6	24	240.000	240.000	
7.7	34	388.000	388.000	
7.8	13	329.000	329.000	
7.9 bis 7.10	11	269.000	269.000	
8.2 bis 8.3	37	639.000	639.000	
8.4	61	518.000	518.000	
8.5 bis 8.6	18	464.000	464.000	

Tabelle 55: Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm-Stichprobe, Haushalte) (Fortsetzung 3)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Haushalte nach Bundesland und BIK-Typ (Schreibweise: "Bundesland.BIK-Typ")	8.7	51	1.034.000	1.034.000
	8.8	63	825.000	825.000
	8.9	17	369.000	369.000
	8.10	83	1.059.000	1.059.000
	9.1 bis 9.2	12	364.000	364.000
	9.3	62	502.000	502.000
	9.4	62	693.000	693.000
	9.5	34	372.000	372.000
	9.6 bis 9.7	82	1.080.000	1.080.000
	9.8	44	634.000	634.000
	9.9	83	720.000	720.000
	9.10	120	1.485.000	1.485.000
	10.3 bis 10.4	17	155.000	155.000
	10.7 bis 10.8	19	325.000	325.000
	11.10	95	1.971.000	1.971.000
	12.1 bis 12.3	11	277.000	277.000
	12.4	21	223.000	223.000
	12.5 bis 12.6	14	136.000	136.000
	12.7	15	126.000	126.000
	12.8	19	147.000	147.000
	12.9 bis 12.10	21	327.000	327.000
	13.1 bis 13.4	16	363.000	363.000
	13.5 bis 13.6	12	164.000	164.000
	13.7	11	102.000	102.000
	13.8	14	221.000	221.000
	14.1	13	21.000	21.000
	14.2	34	143.000	143.000
	14.3 bis 14.4	17	442.000	442.000
	14.5	12	171.000	171.000
	14.6	15	133.000	133.000
	14.7 bis 14.8	9	378.000	378.000
	14.9 bis 14.10	45	893.000	893.000
	15.1 bis 15.2	16	138.000	138.000
	15.3 bis 15.4	16	231.000	231.000
	15.5 bis 15.6	14	309.000	309.000
	15.7	25	241.000	241.000
	15.8	31	276.000	276.000
	16.1 bis 16.2	9	141.000	141.000
	16.3 bis 16.4	25	371.000	371.000
	16.5	22	238.000	238.000
16.6	14	56.000	56.000	
16.7 bis 16.8	23	312.000	312.000	

Tabelle 55: Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm-Stichprobe, Haushalte) (Fortsetzung 4)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	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)	608	11.877.000	11.877.000
	Anzahl Haushalte mit 2 Personen (West)	868	10.551.000	10.551.000
	Anzahl Haushalte mit 3 Personen (West)	382	3.977.000	3.977.000
	Anzahl Haushalte mit 4 Personen (West)	357	3.305.000	3.305.000
	Anzahl Haushalte mit 5 oder mehr Pers. (West)	173	1.274.000	1.274.000
	Anzahl Haushalte mit 1 Person (Ost)	192	3.627.000	3.627.000
	Anzahl Haushalte mit 2 Personen (Ost)	223	3.064.000	3.064.000
	Anzahl Haushalte mit 3 Personen (Ost)	99	1.140.000	1.140.000
	Anzahl Haushalte mit 4 Personen (Ost)	56	571.000	571.000
	Anzahl Haushalte mit 5 oder mehr Pers. (Ost)	19	149.000	149.000
Anzahl der Haushalte nach „Kinder unter 15 Jahren im Haushalt ja/nein“ und West/Ost	Anzahl Haushalte mit Kindern unter 15 (West)	622	5.704.000	5.704.000
	Anzahl Haushalte ohne Kinder unter 15 (West)	1766	25.280.000	25.280.000
	Anzahl Haushalte mit Kindern unter 15 (Ost)	105	1.217.000	1.217.000
	Anzahl Haushalte ohne Kinder unter 15 (Ost)	484	7.334.000	7.334.000

Tabelle 56: Kenngrößen der Verteilung der Gewichte

1%-Perzentil	1212,23
5%-Perzentil	2331,255
10%-Perzentil	3064,392
25%-Perzentil	5468,877
50%-Perzentil	10128,85
75%-Perzentil	18479,22
90%-Perzentil	28758,1
95%-Perzentil	34701,2
99%-Perzentil	44240,42
Mittelwert	13280,15
Standardabweichung	10250,35
Minimum	239,4479
Maximum	56098,93
Fallzahl	2977
Effektivitätsmaß	62,67%

6.9.3 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 176,35 bis 25479,46. 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 18,79 bis maximal 43378,58.

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 Kindern besteht, nach West/Ost

Haushalte Grundlage Mikrozensus 2008:

- 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 (380.663) 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 57: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Haushalte)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl BG im Leistungsbezug nach SGB II nach Bundesländern (16 Kategorien)	Zahl BGs Schleswig-Holstein	151	121.765	121.767
	Zahl BGs Hamburg	81	107.774	107.781
	Zahl BGs Niedersachsen	368	331.822	331.808
	Zahl BGs Bremen	42	50.549	50.549
	Zahl BGs Nordrhein-Westfalen	912	823.814	823.659
	Zahl BGs Hessen	221	218.820	218.743
	Zahl BGs Rheinland-Pfalz	130	122.169	122.156
	Zahl BGs Baden-Württemberg	286	246.337	246.286
	Zahl BGs Bayern	333	265.707	266.183
	Zahl BGs Saarland	62	44.384	44.384
	Zahl BGs Berlin	289	330.193	330.123
	Zahl BGs Brandenburg	196	172.677	172.671
	Zahl BGs Mecklenburg-Vorpommern	111	131.213	131.231
	Zahl BGs Sachsen	275	287.711	287.601
	Zahl BGs Sachsen-Anhalt	266	191.791	191.768
	Zahl BGs Thüringen	143	134.303	134.320

Tabelle 57: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Haushalte) (Fortsetzung 1)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Personen unter 65 Jahren 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)	1063	1.240.872	1.240.874
	Zahl BGs mit 2 Personen unter 65 (West)	657	480.659	480.666
	Zahl BGs mit 3 Personen unter 65 (West)	459	300.506	300.388
	Zahl BGs mit 4 Personen unter 65 (West)	234	184.448	184.409
	Zahl BGs mit 5 oder mehr Personen unter 65 (West)	173	126.656	126.980
	Zahl BGs mit 1 Person unter 65 (Ost)	585	714.495	714.495
	Zahl BGs mit 2 Personen unter 65 (Ost)	338	276.254	276.239
	Zahl BGs mit 3 Personen unter 65 (Ost)	204	143.684	143.688
	Zahl BGs mit 4 Personen unter 65 (Ost)	100	74.190	74.114
	Zahl BGs mit 5 oder mehr Personen unter 65 (Ost)	53	39.265	39.176
Anzahl BG im Leistungsbezug nach SGB II nach Zahl der Kinder unter 15 Jahren in der Bedarfsgemeinschaft (0, 1, 2, 3, „4 oder mehr“) und nach West/Ost (10 Kategorien)	Zahl BGs ohne Kinder unter 15 Jahren (West)	1705	1.585.891	1.586.151
	Zahl BGs mit 1 Kind unter 15 Jahren (West)	495	407.636	407.605
	Zahl BGs mit 2 Kindern unter 15 Jahren (West)	261	229.110	229.119
	Zahl BGs mit 3 Kindern unter 15 Jahren (West)	78	79.014	79.014
	Zahl BGs mit 4 oder mehr Kindern unter 15 Jahren (West)	47	31.490	31.428
	Zahl BGs ohne Kinder unter 15 Jahren (Ost)	960	926.414	926.318
	Zahl BGs mit 1 Kind unter 15 Jahren (Ost)	199	195.946	195.886
	Zahl BGs mit 2 Kindern unter 15 Jahren (Ost)	90	90.169	90.149
	Zahl BGs mit 3 Kindern unter 15 Jahren (Ost)	22	25.468	25.468
	Zahl BGs mit 4 oder mehr Kindern unter 15 Jahren (Ost)	9	9.891	9.891

Tabelle 57: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Haushalte) (Fortsetzung 2)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	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)	640	452.688	452.612
	Rest BGs ohne Alleinerziehenden (West)	1946	1.880.453	1.880.705
	Zahl BGs mit einem/einer Alleinerziehenden (Ost)	218	192.016	192.033
	Rest BGs ohne Alleinerziehenden (Ost)	1062	1.055.872	1.055.679
Anzahl der Haushalte nach Bundesland und BIK-Typ (Schreibweise: "Bundesland.BIK-Typ")	1.1 bis 1.4	66	316.000	316.000
	1.5 bis 1.6	55	149.000	149.000
	1.7 bis 1.8	108	517.000	517.000
	1.9	41	172.000	172.000
	1.10	63	211.000	211.000
	2.10	147	969.000	969.000
	3.2 bis 3.3	82	489.000	489.000
	3.4	63	434.000	434.000
	3.5	42	451.000	451.000
	3.7	225	857.000	857.000
	3.8	186	548.000	548.000
	3.9	101	636.000	636.000
	3.10	87	376.000	376.000
	4.8 bis 4.10	71	358.000	358.000
	5.2 bis 5.3	89	349.000	349.000
	5.4	140	1.052.000	1.052.000
	5.5	183	631.000	631.000
	5.6	60	301.000	301.000
	5.7	180	836.000	836.000
	5.8	461	2.333.000	2.333.000
	5.9	74	363.000	363.000
	5.10	545	2.659.000	2.659.000
	6.2	26	64.000	64.000
	6.3	69	330.000	330.000
	6.4 bis 6.5	52	304.000	304.000
	6.7	78	564.000	564.000
	6.8	87	465.000	465.000
	6.9	79	366.000	366.000
	6.10	114	784.000	784.000
	7.1 bis 7.2	36	305.000	305.000
	7.3	28	177.000	177.000
	7.4	23	154.000	154.000
7.5 bis 7.6	78	240.000	240.000	
7.7	50	388.000	388.000	
7.8	40	329.000	329.000	
7.9 bis 7.10	41	269.000	269.000	
8.2 bis 8.3	64	639.000	639.000	
8.4	107	518.000	518.000	
8.5 bis 8.6	27	464.000	464.000	

Tabelle 57: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Haushalte) (Fortsetzung 3)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Haushalte nach Bundesland und BIK-Typ (Schreibweise: "Bundesland.BIK-Typ")	8.7	103	1.034.000	1.034.000
	8.8	145	825.000	825.000
	8.9	42	369.000	369.000
	8.10	212	1.059.000	1.059.000
	9.1 bis 9.2	24	364.000	364.000
	9.3	93	502.000	502.000
	9.4	95	693.000	693.000
	9.5	53	372.000	372.000
	9.6 bis 9.7	137	1.080.000	1.080.000
	9.8	136	634.000	634.000
	9.9	113	720.000	720.000
	9.10	271	1.485.000	1.485.000
	10.3 bis 10.4	42	155.000	155.000
	10.7 bis 10.8	61	325.000	325.000
	11.10	477	1.971.000	1.971.000
	12.1 bis 12.3	79	277.000	277.000
	12.4	61	223.000	223.000
	12.5 bis 12.6	62	136.000	136.000
	12.7	49	126.000	126.000
	12.8	39	147.000	147.000
	12.9 bis 12.10	56	327.000	327.000
	13.1 bis 13.4	87	363.000	363.000
	13.5 bis 13.6	48	164.000	164.000
	13.7	32	102.000	102.000
	13.8	42	221.000	221.000
	14.1	18	21.000	21.000
	14.2	104	143.000	143.000
	14.3 bis 14.4	53	442.000	442.000
	14.5	40	171.000	171.000
	14.6	96	133.000	133.000
	14.7 bis 14.8	37	378.000	378.000
	14.9 bis 14.10	158	893.000	893.000
	15.1 bis 15.2	60	138.000	138.000
	15.3 bis 15.4	49	231.000	231.000
	15.5 bis 15.6	59	309.000	309.000
	15.7	126	241.000	241.000
	15.8	127	276.000	276.000
	16.1 bis 16.2	37	141.000	141.000
	16.3 bis 16.4	80	371.000	371.000
	16.5	55	238.000	238.000
16.6	46	56.000	56.000	
16.7 bis 16.8	76	312.000	312.000	

Tabelle 57: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Haushalte) (Fortsetzung 4)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	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)	1853	11.877.000	11.877.000
	Anzahl Haushalte mit 2 Personen (West)	1741	10.551.000	10.551.000
	Anzahl Haushalte mit 3 Personen (West)	969	3.977.000	3.977.000
	Anzahl Haushalte mit 4 Personen (West)	662	3.305.000	3.305.000
	Anzahl Haushalte mit 5 oder mehr Pers. (West)	370	1.274.000	1.274.000
	Anzahl Haushalte mit 1 Person (Ost)	914	3.627.000	3.627.000
	Anzahl Haushalte mit 2 Personen (Ost)	687	3.064.000	3.064.000
	Anzahl Haushalte mit 3 Personen (Ost)	372	1.140.000	1.140.000
	Anzahl Haushalte mit 4 Personen (Ost)	198	571.000	571.000
	Anzahl Haushalte mit 5 oder mehr Pers. (Ost)	82	149.000	149.000
Anzahl der Haushalte nach „Kinder unter 15 Jahren im Haushalt ja/nein“ und West/Ost	Anzahl Haushalte mit Kindern unter 15 (West)	1691	5.704.000	5.704.000
	Anzahl Haushalte ohne Kinder unter 15 (West)	3904	25.280.000	25.280.000
	Anzahl Haushalte mit Kindern unter 15 (Ost)	532	1.217.000	1.217.000
	Anzahl Haushalte ohne Kinder unter 15 (Ost)	1721	7.334.000	7.334.000

Tabelle 58: Kenngrößen der Verteilung der Gewichte

1%-Perzentil	52,711
5%-Perzentil	113,8978
10%-Perzentil	167,971
25%-Perzentil	368,6647
50%-Perzentil	959,6308
75%-Perzentil	6079,649
90%-Perzentil	18154
95%-Perzentil	24881,24
99%-Perzentil	30300,4
Mittelwert	5037,589
Standardabweichung	7894,869
Minimum	18,79417
Maximum	43378,58
Fallzahl	7848
Effektivitätsmaß	28,93%

6.10 Kalibrierung zum Personengewicht 4. Welle 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 (499.434) verwendet. Es werden erneut diejenigen Fälle der Stichproben aus Welle 1 Welle 2 und Welle 3 auf die Eckwerte zum Arbeitslosengeld-II-Bezug der Bundesagentur für Arbeit hochgerechnet, die laut Welle 4 des Surveys im Juli 2009 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.

6.10.1 BA-Stichprobe

Die Grundgesamtheit der kumulierten BA-Stichprobe aller vier Wellen bilden alle Personen im Alter ab 15 Jahre, die in einem Haushalt leben, in dem es zu einem der bisher vier Zeitungszeitpunkte (im Juli 2006, Juli 2007, Juli 2008 oder Juli 2009) 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 2009 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 392,22 bis 4011,3. Das Verhältnis der Gesamthochrechnungsfaktoren nach Kalibrierung zu getrimmten Designgewichten wurde nach unten auf minimal 0,5 und nach oben auf maximal 2,0 begrenzt. Danach bewegen sich die Gesamthochrechnungsfaktoren nach Kalibrierung im Bereich von minimal. 196,11 bis maximal 7392,75.

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 (499.434) 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 59: Soll-Verteilungen und Verteilungen nach Kalibrierung (BA-Stichprobe, Personen)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahren in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Bundesländern (16 Kategorien)	Zahl Personen in BGs Schleswig-Holstein	176	170.449	170.449
	Zahl Personen in BGs Hamburg	93	145.876	145.876
	Zahl Personen in BGs Niedersachsen	417	473.772	473.772
	Zahl Personen in BGs Bremen	51	69.246	69.246
	Zahl Personen in BGs Nordrhein-Westfalen	1.114	1.188.016	1.188.016
	Zahl Personen in BGs Hessen	251	316.524	316.524
	Zahl Personen in BGs Rheinland-Pfalz	145	175.232	175.232
	Zahl Personen in BGs Baden-Württemberg	318	343.728	343.728
	Zahl Personen in BGs Bayern	373	359.156	359.156
	Zahl Personen in BGs Saarland	64	61.900	61.900
	Zahl Personen in BGs Berlin	344	447.970	447.970
	Zahl Personen in BGs Brandenburg	238	237.387	237.387
	Zahl Personen in BGs Mecklenburg-Vorpommern	127	180.515	180.515
	Zahl Personen in BGs Sachsen	331	396.110	396.110
	Zahl Personen in BGs Sachsen-Anhalt	331	268.792	268.792
Zahl Personen in BGs Thüringen	163	184.513	184.513	
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	763	971.951	971.951
	Zahl Personen in BGs im Alter 25-64	3.773	4.047.235	4.047.235
Anzahl der Personen ab 15 Jahren in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Geschlecht und nach West/Ost (4 Kategorien)	Zahl Männer in BGs (West)	1.348	1.608.020	1.608.020
	Zahl Frauen in BGs (West)	1.654	1.695.879	1.695.879
	Zahl Männer in BGs (Ost)	746	879.978	879.978
	Zahl Frauen in BGs (Ost)	788	835.309	835.309
Anzahl der Personen ab 15 Jahren 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)	2.416	2.851.211	2.851.211
	Zahl alleinerziehende in BGs (West)	586	452.688	452.688
	Zahl nicht alleinerziehende in BGs (Ost)	1.327	1.523.271	1.523.271
	Zahl alleinerziehende in BGs (Ost)	207	192.016	192.016
Anzahl der Personen ab 15 Jahren in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Staatsangehörigkeit (deutsch/nicht deutsch)	Zahl nicht deutscher Personen in BGs	516	968.390	968.390
	Zahl deutscher Personen in BGs	4.020	4.050.796	4.050.796

Tabelle 60: Kenngrößen der Verteilung der Gewichte

1%-Perzentil	205,9219
5%-Perzentil	258,8751
10%-Perzentil	307,8417
25%-Perzentil	465,0251
50%-Perzentil	771,7853
75%-Perzentil	1472,96
90%-Perzentil	2390,529
95%-Perzentil	2791,07
99%-Perzentil	4901,122
Mittelwert	1107,988
Standardabweichung	943,3924
Minimum	196,1115
Maximum	7392,752
Fallzahl	4530
Effektivitätsmaß	57,97%

6.10.2 Microm-Stichprobe

Die Grundgesamtheit bilden alle über 14-jährigen Personen in Privathaushalten in Deutschland. Ausgangspunkt für die Kalibrierung waren die kalibrierten Haushaltsgewichte der Microm-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 2676,57 bis 38818,75. 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 535,32 bis maximal 135865,6.

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 2008:

- 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 61: Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm-Stichprobe, Personen)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahren in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Bundesländern (16 Kategorien)	Zahl Personen in BGs Schleswig-Holstein	25	170.449	170.449
	Zahl Personen in BGs Hamburg	2	145.876	145.876
	Zahl Personen in BGs Niedersachsen	56	473.772	473.772
	Zahl Personen in BGs Bremen	8	69.246	69.246
	Zahl Personen in BGs Nordrhein-Westfalen	109	1.188.016	1.188.016
	Zahl Personen in BGs Hessen	30	316.524	316.524
	Zahl Personen in BGs Rheinland-Pfalz	20	175.232	175.232
	Zahl Personen in BGs Baden-Württemberg	20	343.728	343.728
	Zahl Personen in BGs Bayern	46	359.156	359.156
	Zahl Personen in BGs Saarland	11	61.900	61.900
	Zahl Personen in BGs Berlin	25	447.970	447.970
	Zahl Personen in BGs Brandenburg	27	237.387	237.387
	Zahl Personen in BGs Mecklenburg-Vorpommern	11	180.515	180.515
	Zahl Personen in BGs Sachsen	25	396.110	396.110
	Zahl Personen in BGs Sachsen-Anhalt	35	268.792	268.792
Zahl Personen in BGs Thüringen	26	184.513	184.513	
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	102	971.951	971.951
	Zahl Personen in BGs im Alter 25-64	374	4.047.235	4.047.235
Anzahl der Personen ab 15 Jahren in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Geschlecht und nach West/Ost (4 Kategorien)	Zahl Männer in BGs (West)	152	1.608.020	1.608.020
	Zahl Frauen in BGs (West)	175	1.695.879	1.695.879
	Zahl Männer in BGs (Ost)	67	879.978	879.978
	Zahl Frauen in BGs (Ost)	82	835.309	835.309
Anzahl der Personen ab 15 Jahren 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)	279	2.851.211	2.851.211
	Zahl alleinerziehende in BGs (West)	48	452.688	452.688
	Zahl nicht alleinerziehende in BGs (Ost)	132	1.523.271	1.523.271
	Zahl alleinerziehende in BGs (Ost)	17	192.016	192.016
Anzahl der Personen ab 15 Jahren in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Staatsangehörigkeit (deutsch/nicht deutsch)	Zahl nicht deutscher Personen in BGs	62	968.390	968.390
	Zahl deutscher Personen in BGs	414	4.050.796	4.050.796

Tabelle 61: Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm-Stichprobe, Personen) (Fortsetzung 1)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahren in Privathaushalten nach Bundesland (16 Kategorien)	Zahl Personen in Privathaushalten Schleswig-Holstein	204	2.397.000	2.397.000
	Zahl Personen in Privathaushalten Hamburg	58	1.553.000	1.553.000
	Zahl Personen in Privathaushalten Niedersachsen	561	6.735.000	6.735.000
	Zahl Personen in Privathaushalten Bremen	29	575.000	575.000
	Zahl Personen in Privathaushalten Nordrhein-Westfalen	1.050	15.301.000	15.301.000
	Zahl Personen in Privathaushalten Hessen	407	5.192.000	5.192.000
	Zahl Personen in Privathaushalten Rheinland-Pfalz	238	3.438.000	3.438.000
	Zahl Personen in Privathaushalten Baden-Württemberg	590	9.130.000	9.130.000
	Zahl Personen in Privathaushalten Bayern	866	10.627.000	10.627.000
	Zahl Personen in Privathaushalten Saarland	60	878.000	878.000
	Zahl Personen in Privathaushalten Berlin	140	3.011.000	3.011.000
	Zahl Personen in Privathaushalten Brandenburg	170	2.223.000	2.223.000
	Zahl Personen in Privathaushalten Mecklenburg-Vorpommern	90	1.470.000	1.470.000
	Zahl Personen in Privathaushalten Sachsen	223	3.692.000	3.692.000
	Zahl Personen in Privathaushalten Sachsen-Anhalt	169	2.101.000	2.101.000
	Zahl Personen in Privathaushalten Thüringen	161	1.993.000	1.993.000
Anzahl der Personen ab 15 Jahren in Privathaushalten nach Alter (in 5-Jahres-Klassen), Geschlecht und West/Ost (56 Kategorien)	Zahl Männer in PH (West), 15-19 Jahre	171	1.943.000	1.943.000
	Zahl Männer in PH (West), 20-24 Jahre	131	1.938.000	1.938.000
	Zahl Männer in PH (West), 25-29 Jahre	79	1.962.000	1.962.000
	Zahl Männer in PH (West), 30-34 Jahre	88	1.876.000	1.876.000
	Zahl Männer in PH (West), 35-39 Jahre	136	2.178.000	2.178.000
	Zahl Männer in PH (West), 40-44 Jahre	207	2.916.000	2.916.000
	Zahl Männer in PH (West), 45-49 Jahre	204	2.681.000	2.681.000
	Zahl Männer in PH (West), 50-54 Jahre	203	2.304.000	2.304.000
	Zahl Männer in PH (West), 55-59 Jahre	166	2.063.000	2.063.000

Tabelle 61: Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm-Stichprobe, Personen) (Fortsetzung 2)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahren in Privathaushalten nach Alter (in 5-Jahres-Klassen), Geschlecht und West/Ost (56 Kategorien)	Zahl Männer in PH (West), 60-64 Jahre	133	1.724.000	1.724.000
	Zahl Männer in PH (West), 65-69 Jahre	146	1.896.000	1.896.000
	Zahl Männer in PH (West), 70-74 Jahre	127	1.703.000	1.703.000
	Zahl Männer in PH (West), 75-79 Jahre	56	1.037.000	1.037.000
	Zahl Männer in PH (West), 80+ Jahre	44	997.000	997.000
	Zahl Frauen in PH (West), 15-19 Jahre	158	1.810.000	1.810.000
	Zahl Frauen in PH (West), 20-24 Jahre	122	1.878.000	1.878.000
	Zahl Frauen in PH (West), 25-29 Jahre	101	1.959.000	1.959.000
	Zahl Frauen in PH (West), 30-34 Jahre	119	1.889.000	1.889.000
	Zahl Frauen in PH (West), 35-39 Jahre	177	2.186.000	2.186.000
	Zahl Frauen in PH (West), 40-44 Jahre	238	2.797.000	2.797.000
	Zahl Frauen in PH (West), 45-49 Jahre	280	2.656.000	2.656.000
	Zahl Frauen in PH (West), 50-54 Jahre	230	2.338.000	2.338.000
	Zahl Frauen in PH (West), 55-59 Jahre	199	2.113.000	2.113.000
	Zahl Frauen in PH (West), 60-64 Jahre	151	1.766.000	1.766.000
	Zahl Frauen in PH (West), 65-69 Jahre	162	2.058.000	2.058.000
	Zahl Frauen in PH (West), 70-74 Jahre	125	1.941.000	1.941.000
	Zahl Frauen in PH (West), 75-79 Jahre	65	1.335.000	1.335.000
	Zahl Frauen in PH (West), 80+ Jahre	45	1.882.000	1.882.000
	Zahl Männer in PH (Ost), 15-19 Jahre	28	365.000	365.000
	Zahl Männer in PH (Ost), 20-24 Jahre	27	557.000	557.000
	Zahl Männer in PH (Ost), 25-29 Jahre	31	563.000	563.000
	Zahl Männer in PH (Ost), 30-34 Jahre	23	487.000	487.000
	Zahl Männer in PH (Ost), 35-39 Jahre	22	540.000	540.000
	Zahl Männer in PH (Ost), 40-44 Jahre	43	698.000	698.000
	Zahl Männer in PH (Ost), 45-49 Jahre	55	698.000	698.000
	Zahl Männer in PH (Ost), 50-54 Jahre	60	643.000	643.000
	Zahl Männer in PH (Ost), 55-59 Jahre	38	608.000	608.000
	Zahl Männer in PH (Ost), 60-64 Jahre	33	417.000	417.000
	Zahl Männer in PH (Ost), 65-69 Jahre	31	550.000	550.000
	Zahl Männer in PH (Ost), 70-74 Jahre	24	487.000	487.000
	Zahl Männer in PH (Ost), 75-79 Jahre	17	274.000	274.000
	Zahl Männer in PH (Ost), 80+ Jahre	7	221.000	221.000

Tabelle 61: Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm-Stichprobe, Personen) (Fortsetzung 3)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahren in Privathaushalten nach Alter (in 5-Jahres-Klassen), Geschlecht und West/Ost (56 Kategorien)	Zahl Frauen in PH (Ost), 15-19 Jahre	26	335.000	335.000
	Zahl Frauen in PH (Ost), 20-24 Jahre	32	513.000	513.000
	Zahl Frauen in PH (Ost), 25-29 Jahre	20	492.000	492.000
	Zahl Frauen in PH (Ost), 30-34 Jahre	22	448.000	448.000
	Zahl Frauen in PH (Ost), 35-39 Jahre	40	496.000	496.000
	Zahl Frauen in PH (Ost), 40-44 Jahre	40	648.000	648.000
	Zahl Frauen in PH (Ost), 45-49 Jahre	55	676.000	676.000
	Zahl Frauen in PH (Ost), 50-54 Jahre	57	641.000	641.000
	Zahl Frauen in PH (Ost), 55-59 Jahre	65	627.000	627.000
	Zahl Frauen in PH (Ost), 60-64 Jahre	45	439.000	439.000
	Zahl Frauen in PH (Ost), 65-69 Jahre	40	625.000	625.000
	Zahl Frauen in PH (Ost), 70-74 Jahre	35	586.000	586.000
	Zahl Frauen in PH (Ost), 75-79 Jahre	22	378.000	378.000
	Zahl Frauen in PH (Ost), 80+ Jahre	15	478.000	478.000
Anzahl der Personen ab 15 Jahren 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)	598	11.877.000	11.877.000
	Zahl Personen in PH mit 2 Personen (West)	1.445	20.639.000	20.639.000
	Zahl Personen in PH mit 3 Personen (West)	763	9.826.000	9.826.000
	Zahl Personen in PH mit 4 Personen (West)	787	9.229.000	9.229.000
	Zahl Personen in PH mit 5 oder mehr Personen (West)	470	4.255.000	4.255.000
	Zahl Personen in PH mit 1 Person (Ost)	187	3.627.000	3.627.000
	Zahl Personen in PH mit 2 Personen (Ost)	380	5.953.000	5.953.000
	Zahl Personen in PH mit 3 Personen (Ost)	203	2.847.000	2.847.000
	Zahl Personen in PH mit 4 Personen (Ost)	128	1.595.000	1.595.000
	Zahl Personen in PH mit 5 oder mehr Personen (Ost)	55	468.000	468.000

Tabelle 61: Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm-Stichprobe, Personen) (Fortsetzung 4)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahren in Privathaushalten nach höchstem Schulabschluss und West/Ost (12 Kategorien)	Zahl Personen in PH mit höchstem Schulabschluss: noch Schüler (West)	206	2.348.000	2.348.000
	Zahl Personen in PH mit höchstem Schulabschluss: ohne Abschluss (West)	104	1.887.000	1.887.000
	Zahl Personen in PH mit höchstem Schulabschluss: Hauptschule (West)	1.329	23.468.000	23.468.000
	Zahl Personen in PH mit höchstem Schulabschluss: Realschule/POS (West)	1.168	13.583.000	13.583.000
	Zahl Personen in PH mit höchstem Schulabschluss:(Fach-)Hochschulreife (West)	1.256	14.540.000	14.540.000
	Zahl Personen in PH mit höchstem Schulabschluss: noch Schüler (Ost)	35	380.000	380.000
	Zahl Personen in PH mit höchstem Schulabschluss: ohne Abschluss (Ost)	20	275.000	275.000
	Zahl Personen in PH mit höchstem Schulabschluss: Hauptschule (Ost)	226	3.802.000	3.802.000
	Zahl Personen in PH mit höchstem Schulabschluss: Realschule/POS (Ost)	425	6.449.000	6.449.000
	Zahl Personen in PH mit höchstem Schulabschluss: (Fach-)Hochschulreife (Ost)	247	3.584.000	3.584.000
Anzahl der Personen ab 15 Jahren in Privathaushalten nach Familienstand und West/Ost (10 Kategorien)	Zahl Personen in PH mit Familienstand: ledig (West)	992	9.886.000	9.886.000
	Zahl Personen in PH mit Familienstand: verheiratet, eingetragene Lebensgemeinschaft. (West)	2.504	36.324.000	36.324.000
	Zahl Personen in PH mit Familienstand: geschieden (West)	342	4.849.000	4.849.000
	Zahl Personen in PH mit Familienstand: verwitwet (West)	225	4.767.000	4.767.000
	Zahl Personen in PH mit Familienstand: ledig (Ost)	242	3.325.000	3.325.000
	Zahl Personen in PH mit Familienstand: verheiratet, eingetragene Lebensgemeinschaft. (Ost)	540	8.389.000	8.389.000
	Zahl Personen in PH mit Familienstand: geschieden (Ost)	99	1.497.000	1.497.000
	Zahl Personen in PH mit Familienstand: verwitwet (Ost)	72	1.279.000	1.279.000
Anzahl der Personen ab 15 Jahren in Privathaushalten nach Staatsangehörigkeit	Zahl Personen in PH Nicht-Deutsch	178	6.354.000	6.354.000
	Zahl Personen in PH Deutsch	4.838	63.962.000	63.962.000

Tabelle 61: Soll-Verteilungen und Verteilungen nach Kalibrierung (Microm-Stichprobe, Personen) (Fortsetzung 5)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Arbeitslose inkl. Maßnahmeteilnehmer West/Ost	Nicht arbeitslos West	3.828	52.539.412	52.539.412
	Arbeitslose inkl. Maßnahmeteilnehmer West	235	3.286.588	3.286.588
	Nicht arbeitslos Ost	845	12.998.476	12.998.476
	Arbeitslose inkl. Maßnahmeteilnehmer Ost	108	1.491.524	1.491.524
Sozialversicherungspflichtig beschäftigt West/Ost	Nicht-Sozialversicherungspflichtig beschäftigt West	2.400	33.504.848	33.504.848
	Sozialversicherungspflichtig beschäftigt West	1.663	22.321.152	22.321.152
	Nicht-Sozialversicherungspflichtig beschäftigt Ost	551	9.246.200	9.246.200
	Sozialversicherungspflichtig beschäftigt Ost	402	5.243.800	5.243.800

Tabelle 62: Kenngrößen der Verteilung der Gewichte

1%-Perzentil	795,634
5%-Perzentil	1747,698
10%-Perzentil	2419,551
25%-Perzentil	4431,909
50%-Perzentil	9077,082
75%-Perzentil	17655,82
90%-Perzentil	31548,64
95%-Perzentil	43427,17
99%-Perzentil	73782,65
Mittelwert	14018,34
Standardabweichung	15070,33
Minimum	535,3148
Maximum	135865,6
Fallzahl	5016
Effektivitätsmaß	46,39%

6.10.3 Gesamt-Stichprobe

Die Grundgesamtheit bilden wie bei der Microm-Stichprobe 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 141,75 bis 29423,42. Das Verhältnis der Gesamthochrechnungsfaktoren nach Kalibrierung zu getrimmten Designgewichten wurde nach unten auf minimal 0,2 und nach oben auf maximal 5,0 begrenzt. Danach bewegen sich die Gesamthochrechnungsfaktoren nach Kalibrierung im Bereich von minimal 28,35 bis maximal 123298,8.

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 2008:

- 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 (499.434) 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 63: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahren in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Bundesländern (16 Kategorien)	Zahl Personen in BGs Schleswig-Holstein	201	170.449	170449
	Zahl Personen in BGs Hamburg	95	145.876	145876
	Zahl Personen in BGs Niedersachsen	473	473.772	473772
	Zahl Personen in BGs Bremen	59	69.246	69246
	Zahl Personen in BGs Nordrhein-Westfalen	1.223	1.188.016	1188016
	Zahl Personen in BGs Hessen	281	316.524	316524
	Zahl Personen in BGs Rheinland-Pfalz	165	175.232	175232
	Zahl Personen in BGs Baden-Württemberg	338	343.728	343728
	Zahl Personen in BGs Bayern	419	359.156	359156
	Zahl Personen in BGs Saarland	75	61.900	61900
	Zahl Personen in BGs Berlin	369	447.970	447970
	Zahl Personen in BGs Brandenburg	265	237.387	237387
	Zahl Personen in BGs Mecklenburg-Vorpommern	138	180.515	180515
	Zahl Personen in BGs Sachsen	356	396.110	396110
	Zahl Personen in BGs Sachsen-Anhalt	366	268.792	268792
Zahl Personen in BGs Thüringen	189	184.513	184513	
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	865	971.951	971951
	Zahl Personen in BGs im Alter 25-64	4.147	4.047.235	4047235
Anzahl der Personen ab 15 Jahren in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Geschlecht und nach West/Ost (4 Kategorien)	Zahl Männer in BGs (West)	1.500	1.608.020	1608020
	Zahl Frauen in BGs (West)	1.829	1.695.879	1695879
	Zahl Männer in BGs (Ost)	813	879.978	879978
	Zahl Frauen in BGs (Ost)	870	835.309	835309
Anzahl der Personen ab 15 Jahren 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)	2.695	2.851.211	2851211
	Zahl alleinerziehende in BGs (West)	634	452.688	452688
	Zahl nicht alleinerziehende in BGs (Ost)	1.459	1.523.271	1523271
	Zahl alleinerziehende in BGs (Ost)	224	192.016	192016
Anzahl der Personen ab 15 Jahren in Bedarfsgemeinschaften im Leistungsbezug nach SGB II nach Staatsangehörigkeit (deutsch/nicht deutsch)	Zahl nicht deutscher Personen in BGs	578	968.390	968390
	Zahl deutscher Personen in BGs	4.434	4.050.796	4050796

Tabelle 63: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen) (Fortsetzung 1)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahren in Privathaushalten nach Bundesland (16 Kategorien)	Zahl Personen in Privathaushalten Schleswig-Holstein	513	2.397.000	2.397.000
	Zahl Personen in Privathaushalten Hamburg	192	1.553.000	1.553.000
	Zahl Personen in Privathaushalten Niedersachsen	1.178	6.735.000	6.735.000
	Zahl Personen in Privathaushalten Bremen	103	575.000	575.000
	Zahl Personen in Privathaushalten Nordrhein-Westfalen	2.621	15.301.000	15.301.000
	Zahl Personen in Privathaushalten Hessen	768	5.192.000	5.192.000
	Zahl Personen in Privathaushalten Rheinland-Pfalz	447	3.438.000	3.438.000
	Zahl Personen in Privathaushalten Baden-Württemberg	1.080	9.130.000	9.130.000
	Zahl Personen in Privathaushalten Bayern	1.444	10.627.000	10.627.000
	Zahl Personen in Privathaushalten Saarland	148	878.000	878.000
	Zahl Personen in Privathaushalten Berlin	641	3.011.000	3.011.000
	Zahl Personen in Privathaushalten Brandenburg	517	2.223.000	2.223.000
	Zahl Personen in Privathaushalten Mecklenburg-Vorpommern	302	1.470.000	1.470.000
	Zahl Personen in Privathaushalten Sachsen	728	3.692.000	3.692.000
	Zahl Personen in Privathaushalten Sachsen-Anhalt	634	2.101.000	2.101.000
	Zahl Personen in Privathaushalten Thüringen	452	1.993.000	1.993.000
Anzahl der Personen ab 15 Jahren in Privathaushalten nach Alter (in 5-Jahres-Klassen), Geschlecht und West/Ost (56 Kategorien)	Zahl Männer in PH (West), 15-19 Jahre	345	1.943.000	1.943.000
	Zahl Männer in PH (West), 20-24 Jahre	312	1.938.000	1.938.000
	Zahl Männer in PH (West), 25-29 Jahre	289	1.962.000	1.962.000
	Zahl Männer in PH (West), 30-34 Jahre	268	1.876.000	1.876.000
	Zahl Männer in PH (West), 35-39 Jahre	333	2.178.000	2.178.000
	Zahl Männer in PH (West), 40-44 Jahre	424	2.916.000	2.916.000
	Zahl Männer in PH (West), 45-49 Jahre	438	2.681.000	2.681.000

Tabelle 63: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen) (Fortsetzung 2)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahren in Privathaushalten nach Alter (in 5-Jahres-Klassen), Geschlecht und West/Ost (56 Kategorien)	Zahl Männer in PH (West), 50-54 Jahre	404	2.304.000	2.304.000
	Zahl Männer in PH (West), 55-59 Jahre	363	2.063.000	2.063.000
	Zahl Männer in PH (West), 60-64 Jahre	275	1.724.000	1.724.000
	Zahl Männer in PH (West), 65-69 Jahre	204	1.896.000	1.896.000
	Zahl Männer in PH (West), 70-74 Jahre	143	1.703.000	1.703.000
	Zahl Männer in PH (West), 75-79 Jahre	61	1.037.000	1.037.000
	Zahl Männer in PH (West), 80+ Jahre	47	997.000	997.000
	Zahl Frauen in PH (West), 15-19 Jahre	322	1.810.000	1.810.000
	Zahl Frauen in PH (West), 20-24 Jahre	339	1.878.000	1.878.000
	Zahl Frauen in PH (West), 25-29 Jahre	355	1.959.000	1.959.000
	Zahl Frauen in PH (West), 30-34 Jahre	396	1.889.000	1.889.000
	Zahl Frauen in PH (West), 35-39 Jahre	432	2.186.000	2.186.000
	Zahl Frauen in PH (West), 40-44 Jahre	542	2.797.000	2.797.000
	Zahl Frauen in PH (West), 45-49 Jahre	600	2.656.000	2.656.000
	Zahl Frauen in PH (West), 50-54 Jahre	461	2.338.000	2.338.000
	Zahl Frauen in PH (West), 55-59 Jahre	405	2.113.000	2.113.000
	Zahl Frauen in PH (West), 60-64 Jahre	278	1.766.000	1.766.000
	Zahl Frauen in PH (West), 65-69 Jahre	203	2.058.000	2.058.000
	Zahl Frauen in PH (West), 70-74 Jahre	135	1.941.000	1.941.000
	Zahl Frauen in PH (West), 75-79 Jahre	71	1.335.000	1.335.000
	Zahl Frauen in PH (West), 80+ Jahre	49	1.882.000	1.882.000
	Zahl Männer in PH (Ost), 15-19 Jahre	105	365.000	365.000
	Zahl Männer in PH (Ost), 20-24 Jahre	127	557.000	557.000
	Zahl Männer in PH (Ost), 25-29 Jahre	173	563.000	563.000
	Zahl Männer in PH (Ost), 30-34 Jahre	137	487.000	487.000
	Zahl Männer in PH (Ost), 35-39 Jahre	118	540.000	540.000
	Zahl Männer in PH (Ost), 40-44 Jahre	142	698.000	698.000

Tabelle 63: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen) (Fortsetzung 3)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahren in Privathaushalten nach Alter (in 5-Jahres-Klassen), Geschlecht und West/Ost (56 Kategorien)	Zahl Männer in PH (Ost), 45-49 Jahre	190	698.000	698.000
	Zahl Männer in PH (Ost), 50-54 Jahre	185	643.000	643.000
	Zahl Männer in PH (Ost), 55-59 Jahre	166	608.000	608.000
	Zahl Männer in PH (Ost), 60-64 Jahre	116	417.000	417.000
	Zahl Männer in PH (Ost), 65-69 Jahre	45	550.000	550.000
	Zahl Männer in PH (Ost), 70-74 Jahre	30	487.000	487.000
	Zahl Männer in PH (Ost), 75-79 Jahre	17	274.000	274.000
	Zahl Männer in PH (Ost), 80+ Jahre	7	221.000	221.000
	Zahl Frauen in PH (Ost), 15-19 Jahre	110	335.000	335.000
	Zahl Frauen in PH (Ost), 20-24 Jahre	133	513.000	513.000
	Zahl Frauen in PH (Ost), 25-29 Jahre	156	492.000	492.000
	Zahl Frauen in PH (Ost), 30-34 Jahre	152	448.000	448.000
	Zahl Frauen in PH (Ost), 35-39 Jahre	133	496.000	496.000
	Zahl Frauen in PH (Ost), 40-44 Jahre	155	648.000	648.000
	Zahl Frauen in PH (Ost), 45-49 Jahre	229	676.000	676.000
	Zahl Frauen in PH (Ost), 50-54 Jahre	221	641.000	641.000
	Zahl Frauen in PH (Ost), 55-59 Jahre	175	627.000	627.000
	Zahl Frauen in PH (Ost), 60-64 Jahre	113	439.000	439.000
	Zahl Frauen in PH (Ost), 65-69 Jahre	57	625.000	625.000
	Zahl Frauen in PH (Ost), 70-74 Jahre	41	586.000	586.000
	Zahl Frauen in PH (Ost), 75-79 Jahre	23	378.000	378.000
Zahl Frauen in PH (Ost), 80+ Jahre	18	478.000	478.000	
Anzahl der Personen ab 15 Jahren 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)	1.824	11.877.000	11.877.000
	Zahl Personen in PH mit 2 Personen (West)	2.667	20.639.000	20.639.000
	Zahl Personen in PH mit 3 Personen (West)	1.758	9.826.000	9.826.000
	Zahl Personen in PH mit 4 Personen (West)	1.355	9.229.000	9.229.000
	Zahl Personen in PH mit 5 oder mehr Personen (West)	890	4.255.000	4.255.000
	Zahl Personen in PH mit 1 Person (Ost)	904	3.627.000	3.627.000
	Zahl Personen in PH mit 2 Personen (Ost)	1.086	5.953.000	5.953.000
	Zahl Personen in PH mit 3 Personen (Ost)	678	2.847.000	2.847.000
	Zahl Personen in PH mit 4 Personen (Ost)	403	1.595.000	1.595.000
	Zahl Personen in PH mit 5 oder mehr Personen (Ost)	203	468.000	468.000

Tabelle 63: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen) (Fortsetzung 4)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahren in Privathaushalten nach höchstem Schulabschluss und West/Ost (12 Kategorien)	Zahl Personen in PH mit höchstem Schulabschluss: noch Schüler (West)	408	2.348.000	2.348.000
	Zahl Personen in PH mit höchstem Schulabschluss: ohne Abschluss (West)	426	1.887.000	1.887.000
	Zahl Personen in PH mit höchstem Schulabschluss: Hauptschule (West)	3.065	23.468.000	23.468.000
	Zahl Personen in PH mit höchstem Schulabschluss: Realschule/POS (West)	2.432	13.583.000	13.583.000
	Zahl Personen in PH mit höchstem Schulabschluss:(Fach-)Hochschulreife (West)	2.163	14.540.000	14.540.000
	Zahl Personen in PH mit höchstem Schulabschluss: noch Schüler (Ost)	118	380.000	380.000
	Zahl Personen in PH mit höchstem Schulabschluss: ohne Abschluss (Ost)	107	275.000	275.000
	Zahl Personen in PH mit höchstem Schulabschluss: Hauptschule (Ost)	772	3.802.000	3.802.000
	Zahl Personen in PH mit höchstem Schulabschluss: Realschule/POS (Ost)	1.634	6.449.000	6.449.000
	Zahl Personen in PH mit höchstem Schulabschluss: (Fach-) Hochschulreife (Ost)	643	3.584.000	3.584.000
Anzahl der Personen ab 15 Jahren in Privathaushalten nach Familienstand und West/Ost (10 Kategorien)	Zahl Personen in PH mit Familienstand: ledig (West)	2.759	9.886.000	9.886.000
	Zahl Personen in PH mit Familienstand: verheiratet, eingetragene Lebensgemeinschaft. (West)	4.120	36.324.000	36.324.000
	Zahl Personen in PH mit Familienstand: geschieden (West)	1.282	4.849.000	4.849.000
	Zahl Personen in PH mit Familienstand: verwitwet (West)	333	4.767.000	4.767.000
	Zahl Personen in PH mit Familienstand: ledig (Ost)	1.304	3.325.000	3.325.000
	Zahl Personen in PH mit Familienstand: verheiratet, eingetragene Lebensgemeinschaft. (Ost)	1.329	8.389.000	8.389.000
	Zahl Personen in PH mit Familienstand: geschieden (Ost)	514	1.497.000	1.497.000
	Zahl Personen in PH mit Familienstand: verwitwet (Ost)	127	1.279.000	1.279.000

Tabelle 63: Soll-Verteilungen und Verteilungen nach Kalibrierung (Gesamtstichprobe, Personen) (Fortsetzung 5)

Eckwert	Ausprägung Eckwert aus der BA-Statistik bzw. Mikrozensus 2009	ungewichtete Verteilung	Soll-Werte	Verteilung mit kalibrierten Gewichten
Anzahl der Personen ab 15 Jahren in Privathaushalten nach Staatsangehörigkeit	Zahl Personen in PH Nicht-Deutsch	873	6.354.000	6.354.000
	Zahl Personen in PH Deutsch	10.895	63.962.000	63.962.000
Arbeitslose inkl. Maßnahmeteilnehmer West/Ost	Nicht arbeitslos West	6.499	52.539.412	52.539.412
	Arbeitslose inkl. Maßnahmeteilnehmer West	1.995	3.286.588	3.286.588
	Nicht arbeitslos Ost	2.170	12.998.476	12.998.476
	Arbeitslose inkl. Maßnahmeteilnehmer Ost	1.104	1.491.524	1.491.524
Sozialversicherungspflichtig beschäftigt West/Ost	Nicht-Sozialversicherungspflichtig beschäftigt West	5.466	33.504.848	33.504.848
	Sozialversicherungspflichtig beschäftigt West	3.028	22.321.152	22.321.152
	Nicht-Sozialversicherungspflichtig beschäftigt Ost	2.044	9.246.200	9.246.200
	Sozialversicherungspflichtig beschäftigt Ost	1.230	5.243.800	5.243.800

Tabelle 64: Kenngrößen der Verteilung der Gewichte

1%-Perzentil	31,43041
5%-Perzentil	79,49212
10%-Perzentil	134,8719
25%-Perzentil	321,3353
50%-Perzentil	1247,177
75%-Perzentil	6821,727
90%-Perzentil	18490,31
95%-Perzentil	27502,24
99%-Perzentil	48845,01
Mittelwert	5975,187
Standardabweichung	10608,77
Minimum	28,3505
Maximum	123298,8
Fallzahl	11768
Effektivitätsmaß	24,08%

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

Schließlich verblieben auch in der Welle 4 noch Haushalte und Personen, denen durch Kalibrierung kein BA-Haushaltsquerschnittsgewicht oder BA-Personenquerschnittsgewicht zugewiesen werden konnte. Die Anzahl dieser Haushalte ist nun in Welle 4 größer als in Welle 3, da ein größerer Teil der BA-Stichprobe der Welle 1 und 2 mittlerweile aus dem Leistungsbezug ausgeschieden ist. Es handelt sich um die folgenden drei Gruppen ohne Leistungsbezug im Juli 2009, die aber zur Grundgesamtheit der BA-Stichprobe (Haushalte mit Arbeitslosengeld-II-Bezug in 7/2006 oder 7/2007 oder 7/2008 oder 7/2009 bzw. Personen in Haushalten mit Arbeitslosengeld-II-Bezug in 7/2006 oder 7/2007 oder 7/2008 oder 7/2009) gehören.

- Aus der Auffrischungsstichprobe: Nicht-BG-Mitglieder im Haushalt: Das Personengewicht wurde hier aus dem BA-Haushaltsgewicht in Welle 4 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 2009 niemand mehr Arbeitslosengeld II bezogen hat: Der Haushalt behält das BA-Gewicht vor Kalibrierung. Personen in diesen Haushalten mit Interviews in beiden Wellen wurden 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 3 erhalten als neues BA-Personengewicht das BA-Haushaltsgewicht ihres Haushalts für Welle 4 dividiert durch den Anteil solcher Personen, die teilnehmen, vorausgesetzt ihr Haushalt nimmt teil.
- Nicht-BG-Mitglieder in Bestandshaushalten, die im Juli 2009 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.

7 Anhang: Kurzbeschreibung des Datensatzes

Inhaltliche Charakteristika

Kategorien	Erläuterungen
Themen/ Merkmalsgruppen	<p>Soziodemographische Merkmale: systemfreie Personenummer; Geschlecht; Geburtsjahr; Alter; Familienstand; Kinderzahl im und außerhalb des Haushalts; Staatsangehörigkeit; Herkunftsland und Migrationshintergrund; Schul- und Berufsbildung (incl. generierter Skalen: CASMIN, ISCED-97, Bildung in Jahren); Schul- und Berufsbildung der Eltern; Gesundheitsindikatoren; Konfession; soziale Kontakte; Betreuung und Schulbesuch von Kindern; Haushaltseinkommen (incl. 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); Befristung 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;</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; Beendigungsgrund; Identifizierung beziehende HH-Mitglieder; Höhe der Leistungen; Leistungskürzungen (Anfang, Dauer, Gründe, HH-Mitglieder mit Kürzung); <u>Maßnahmeteilnahme:</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>Kontakte zum ALGII-Träger:</u> Anzahl und Art der Kontakte; Gesprächsinhalte; Angebote; Eingliederungsvereinbarung; Bewertung des Trägers;</p> <p>Subjektive Indikatoren: Zufriedenheit; Ängste und Sorgen; Erwerbsorientierungen; Bildungsaspiration; Geschlechterrollenorientierungen; subjektive soziale Position (Oben-Unten-Skala); subjektive Bewertung des Gesundheitszustands</p>

Kategorien	Erläuterungen
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>Anmerkung: Befragung von Personen ab 65 Jahren nur mittels eines Kurzfragebogens</p>
Fallzahlen	<p>Welle 1:</p> <p>Sample I: 9.386 Personen (in 6.804 Haushalten)</p> <p>Sample II: 9.568 Personen (in 5.990 Haushalten)</p> <p>Welle 2:</p> <p>Sample I: 4.753 Personen (in 3.491 Haushalten)</p> <p>Sample II: 6.392 Personen (in 3.897 Haushalten)</p> <p>Sample III: 1.342 Personen (in 1.041 Haushalten)</p> <p>Welle 3:</p> <p>Sample I: 4.913 Personen (in 3.754 Haushalten)</p> <p>Sample II: 6.207 Personen (in 3.901 Haushalten)</p> <p>Sample III: 898 Personen (in 694 Haushalten)</p> <p>Sample IV: 1.421 Personen (in 1.186 Haushalten)</p> <p>Welle 4:</p> <p>Sample I: 3.958 Personen (in 2.815 Haushalten)</p> <p>Sample II: 5.016 Personen (in 2.977 Haushalten)</p> <p>Sample III: 786 Personen (in 563 Haushalten)</p> <p>Sample IV: 983 Personen (in 745 Haushalten)</p> <p>Sample V: 1.025 Personen (in 748 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>
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>

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>

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>
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</p>
Zeitraum	<p>Welle 1: Feldzeit sowie Spelldaten retrospektiv ab 01/2005 Welle 2: Feldzeit sowie Spelldaten retrospektiv ab 01/2005 bzw. jeweiligem Referenzzeitraum des Spelldatentyps Welle 3: Feldzeit sowie Spelldaten retrospektiv ab 01/2006 bzw. jeweiligem Referenzzeitraum des Spelldatentyps Welle 4: Feldzeit sowie Spelldaten retrospektiv ab 01/2008 bzw. jeweiligem Referenzzeitraum des Spelldatentyps</p>
Zeitbezug	Wiederholungsbefragung (Haushaltspanel)
Regionale Gliederung	Bundesland, Ost/West (Weitere regionale Informationen sind vorhanden, aber aus Datenschutzgründen nicht im Scientific Use File enthalten. Detaillierte Informationen hierzu auf Anfrage)
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. 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 Gebäuestichprobe 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 1 für Sample I in Welle 2: Zusätzlich zur Fortführung der für Welle 1 gezogenen Samples I wurde in der 2. 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 1. Welle.</p> <p>Zugangsstichprobe 2 für Sample I in Welle 3: 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 und 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 1. Welle.</p> <p>Zugangsstichprobe 3 für Sample I in Welle 4: 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 1. Welle.</p>

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>

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. AAPOR, Lanexa.

Achatz, Juliane, Hirsland, Andreas und Promberger, Markus (2007), „IAB-Panelbefragung von Haushalten im Niedrigeinkommensbereich. Entwurf für ein Rahmenkonzept“, S. 11-32 in Promberger, Markus (Hrsg.), Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung 'Arbeitsmarkt und Soziale Sicherung', IAB Forschungsbericht 12/2007, Nürnberg.

Andersen, Hanfried H.; Mühlbacher, Axel; Nübling, Matthias (2007a): Die SOEP-Version des SF-12 als Instrument gesundheitsökonomischer Analysen. SOEPpapers 6/2007, Deutsches Institut für Wirtschaftsforschung, Berlin, S.46.

Andersen, Hanfried H.; Mühlbacher, Axel; Nübling, Matthias; Schupp, Jürgen, Wagner, Gert G. (2007b): Computation of Standard Values for Physical and Mental Health Scale Scores Using the SOEP-Version of SF-12v2. Schmollers Jahrbuch 127, S. 171-182.

Andreß Hans-Jürgen, Burkatzki, Eckhard, Lipsmeier, Gero, Salentin, Kurt, Schulte, Katja und Strengmann-Kuhn, Wolfgang (1996): Leben in Armut. Analysen der Verhaltensweisen armer Haushalte mit Umfragedaten. Endbericht des DFG-Projekts „Versorgungsstrategien privater Haushalte im unteren Einkommensbereich (VuE)“. Bielefeld.

Andreß, Hans-Jürgen und Lipsmeier, Gero (1995): Was gehört zum notwendigen Lebensstandard und wer kann ihn sich leisten? Ein neues Konzept zur Armutsmessung. Aus Politik und Zeitgeschichte, B 31-32/1995: 35-49.

Andreß, Hans-Jürgen und Lipsmeier, Gero (2001): Armut und Lebensstandard. Gutachten im Rahmen des Armuts- und Reichtumsberichts der Bundesregierung. BMAS. Bonn.

Beckmann, Petra und Trometer, Reiner (1991): Neue Dienstleistungen des ALLBUS: Haushalts- und Familientypologien, Klassenschema nach Goldthorpe. ZUMA-Nachrichten 28: 7-17.

Bethmann, Arne und Gebhardt, Daniel (2011): User Guide „Panel Labour Market and Social Security“ (PASS), Wave 3, FDZ Datenreport 04/2011 (en), Institut für Arbeitsmarkt und Berufsforschung, Nürnberg

Marco Berg, Ralph Cramer, Christian Dickmann, Reiner Gilberg, Birgit Jesske, Karen Marwinski, (alle infas Institut für angewandte Sozialwissenschaft GmbH), Daniel Gebhardt, Claudia Wenzig, Martin Wetzel (alle Institut für Arbeitsmarkt- und Berufsforschung (IAB): Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS) Welle 3 (2008/2009), FDZ Datenreport, 06/2010 (de), Nürnberg.

Brauns, Hildegard und Steinmann, Susanne (1999): Educational Reform in France, West-Germany and the United Kingdom: Updating the Casmin Classification. ZUMA-Nachrichten 44: 7-45

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

Büngeler, Kathrin, Gensicke, Miriam, Hartmann, Josef, Jäckle, Robert, Tschersich, Nikolai (2009): IAB-Haushaltspanel m Niedrigeinkommensbereich Welle 2 (2007/2008). Methoden- und Feldbericht. FDZ-Methodenreport, 08/2009, Nürnberg.

Büngeler, Kathrin, Gensicke, Miriam, Hartmann, Josef, Jäckle, Robert, Tschersich, Nikolai (2010): IAB-Haushaltspanel im Niedrigeinkommensbereich Welle 3 (2008/2009). Methoden- und Feldbericht. FDZ-Methodenreport, 10/2010 Nürnberg.

Christoph, Bernhard (2005): Zur Messung des Berufsprestiges: Aktualisierung der Magnitude-Prestigeskala auf die Berufsklassifikation ISCO88. ZUMA-Nachrichten 57: 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, Robert und Goldthorpe, John H. (1992): The Constant Flux. A Study of Class Mobility in Industrial Society. Clarendon Press, Oxford.

Erikson, Robert, Goldthorpe, John H. und Portocarero, Lucienne (1979): Intergenerational Class Mobility in Three Western Societies: England, France and Sweden. British Journal of Sociology 30: 415-441.

Erikson, Robert, Goldthorpe, John H. und Portocarero, Lucienne (1982): Social Fluidity in Industrial Nations: England, France and Sweden. British Journal of Sociology 33: 1-34.

Fischer, Andreas und Wirth, Heike (2007): Constructing Version 4 of ESEC Classes from 3-digit ISCO (Stata-do file). Gesis-ZUMA, Mannheim

Frick, Joachim R.; Göbel, Jan und Krause, Peter (o. J.): \$HGEN: Generated Household-Level Variables. Download: [<http://www.diw.de/documents/dokumentenarchiv/17/60053/hgen.pdf> (8.11.2007)].

Ganzeboom, Harry B. G. und Treiman, Donald (1996): Internationally Comparable Measures for Occupational Status for the 1988 International Standard Classification of Occupations. Social Science Research, 25: 201-239.

Ganzeboom, Harry B. G. und Treiman, Donald (2003): Three Internationally Standardised Measures for Comparative Research on Occupational Status. In Jürgen H. P. Hoffmeyer-Zlotnik und Wolf, Christof (Hg.): Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables (S. 159-193). Kluwer Academic/Plenum Publishers, New York et al.

Ganzeboom, Harry B. G.; De Graaf, Paul M. und Treiman, Donald J. (1992). A Standard International Socio-Economic Index of Occupational Status. Social Science Research, 21: 1-56.

Gebhardt, Daniel; Müller, Gerrit; Bethmann, Arne; Trappmann, Mark; Christoph, Bernhard; Gayer, Christine; Müller, Bettina; Tisch, Anita; Siflinger, Bettina; Kiesl, Hans; Huyer-May, Bernadette; Achatz, Juliane; Wenzig, Claudia; Rudolph, Helmut; Graf, Tobias; Biedermann, Anika (2009): Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS) Welle 2 (2007/2008), (FDZ Datenreport, 06/2009 (de), Nürnberg

Granato, Nadia (2000): CASMIN-Bildungsklassifikation. Eine Umsetzung mit dem Mikrozensus 1996. ZUMA-Technischer Bericht 2000/12.

Hagenaars, Aldi, J. M. de Vos, Klaas und Zaidi, M. Ashgar (1994): Poverty Statistics in the Late 1980s: Research Based on Micro-data. Office for Official Publications of the European Communities, Luxembourg.

Halleröd, Björn (1995): The Truly Poor: Direct and Indirect Consensual Measurement of Poverty in Sweden. Journal of European Social Policy. Vol. 5, Nr. 2: 111-129.

Harrison, Eric und Rose, Richard (2006): ESeC User Guide, Appendix 6 (SPSS-Syntax: Esec Full) [<http://www.iser.essex.ac.uk/esec/guide/docs/Appendix6.sps>]

Hartmann, Josef, Brink, Kathrin, Jäckle, Robert und Tschersich, Niko (2008). IAB-Haushaltspanel im Niedrigeinkommensbereich: Methoden- und Feldbericht. München.

Hauser, Richard (1996): Zur Messung individueller Wohlfahrt und Ihrer Verteilung. S. 13-38 in Statistisches Bundesamt (Hg.): Wohlfahrtsmessung. Aufgabe der Statistik im gesellschaftlichen Wandel. Metzler-Poeschel, Stuttgart.

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

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

Jäckle, Annette (2008): The Causes of Seam Effects in Panel Surveys. ISEP Working Paper Series, No. 2008-14, University of Essex.

Jeske, Birgit und Quandt, Sylvia (2011): Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 4. Erhebungswelle – 2010 (Haupterhebung), FDZ Methodenreport, 08/2011 Nürnberg.

König, Wolfgang, Lüttinger, Paul und Müller, Walter (1987): Eine vergleichende Analyse der Entwicklung und Struktur von Bildungssystemen. Methodologische Grundlagen und Konstruktion einer vergleichbaren Bildungsskala. CASMIN-Projekt. Arbeitspapier Nr. 12.

Lechert, Yvonne; Schroedter, Julia; Lüttinger, Paul (2006): Die Umsetzung der Bildungs-klassifikation CASMIN für die Volkszählung 1970, die Mikrozensus- Zusatzerhebung 1971 und die Mikrozensusen 1976-2004. ZUMA-Methodenbericht 2006/12.

Lengerer, Andrea; Bohr, Jeanette und Janßen, Andrea (2005): Haushalte, Familien und Lebensformen im Mikrozensus – Konzepte und Typisierungen. ZUMA-Arbeitsbericht Nr. 2005/05; ZUMA; Mannheim.

Lipsmeier, Gero (1999): Die Bestimmung des notwendigen Lebensstandards – Einschätzungsunterschiede und Entscheidungsprobleme. Zeitschrift für Soziologie, Jg. 28, Heft 4: 281-300.

Müller, Walter, Wirth, Heike, Bauer, Gerrit, Pollak, Reinhard und Weiss, Felix (2006): ESeC – Kurzbericht zur Validierung und Operationalisierung einer europäischen sozioökonomischen Klassifikation. ZUMA-Nachrichten 59: 111–119.

Müller, Walter, Wirth, Heike, Bauer, Gerrit, Pollak, Reinhard und Weiss, Felix (2007): Entwicklung einer europäischen sozioökonomischen Klassifikation. Wirtschaft und Statistik, 5/2007: 527-530.

Nolan, Brian und Whelan, Christopher T. (1996). Measuring Poverty Using Income and Deprivation Indicators: Alternative Approaches. Journal of European Social Policy, Vol. 6, No. 3: 225-240

Nübling, M., Andersen, H. H., Mühlbacher, A., (2006): Entwicklung eines Verfahrens zur Berechnung der körperlichen und psychischen Summenskalen auf Basis der SOEP – Version des SF 12 (Algorithmus). Data Documentation 16, 14 S., Berlin: DIW.

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

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

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

Rendtel, Ulrich, Harms, Torsten (2009): Weighting and calibration for household panels, S. 265-286 in Lynn, Peter (Hrsg.), Methodology of Longitudinal Surveys. Chichester, Wiley.

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

Rose, Richard und Harrison, Eric (2007): The European Socio-Economic Classification: A New Social Class Schema for Comparative European Research. European Societies 9, 3: 459-490.

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

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

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

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

Trappmann, Mark, Christoph, Bernhard, Achatz, Juliane, Wenzig, Claudia, Müller, Gerrit, Gebhardt, Daniel (2009): Design and stratification off PASS. A New Panel Study for Research on Long Term Unemployment. IAB-Discussion Paper, 5/2009, Nürnberg.

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

Ware JE, Kosinski M, Turner-Bowker DM, Gandek B (2002): How to score Version 2 of the SF-12® Health Survey. Lincoln, RI: Qualitymetric Incorporated,

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

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

Impressum

FDZ-Datenreport 8/2011

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/2011/DR_08-11_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