

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

**Band I: Einführung und Überblick**

**Welle 1 (2006/2007)**

*Bernhard Christoph, Gerrit Müller, Daniel Gebhardt, Claudia Wenzig,  
Mark Trappmann, Juliane Achatz, Anita Tisch und Christine Gayer*

## 1. Das 'Panel Arbeitsmarkt und soziale Sicherung': Einleitung und Überblick

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. Dieser Datenreport gibt einen Überblick über die erste Befragungswelle, für die zwischen Dezember 2006 und Juli 2007 18.954 Personen in 12.794 Haushalten befragt wurden.

Die Studie ist im Rahmen der SGB II - Forschung 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?

Im Folgenden sollen die Motivation für die Durchführung der Befragung, die Befragungsinhalte und das Studiendesign in einem kurzen Überblick dargestellt werden.

---

### a) Ergänzung bisher verfügbarer Datengrundlagen

Die deutsche Arbeitsmarkt-, Armuts- und Sozialstaatsforschung kann bereits auf verschiedene Mikrodatensätze zurückgreifen. Insbesondere gibt es eine Reihe von Längsschnittdaten, die bereits für relativ lange Erhebungszeiträume zur Verfügung stehen. Im Bereich der Erhebungsdaten ist dabei insbesondere das Sozio-ökonomische Panel (SOEP) (Wagner et al. 2007) zu nennen, das bereits seit 1984 jährlich Daten auf Personen- und Haushaltsebene liefert. Daneben stehen die im IAB aufbereiteten und vom Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit (BA) im IAB angebotenen prozessproduzierten Daten der BA, wie beispielsweise die integrierten Erwerbsbiographien (IEBS), die IAB-Beschäftigtenstichprobe (IABS) oder der Linked-Employer-Employee-Datensatz (LIAB) zur Verfügung.

Das Fragespektrum und das Design des PASS sind darauf ausgerichtet, Lücken der vorhandenen Datenbestände zu schließen. Gegenüber den Prozessdaten der Bundesagentur sind es vor allem drei Eigenschaften des PASS, die das Analysepotenzial erweitern:

1. die Berücksichtigung des Haushaltskontexts – auch vor und nach dem Arbeitslosengeld-II-Bezug.
2. die Vollständigkeit im Hinblick auf Personengruppen und Erwerbsverläufe, da nicht nur abhängig Beschäftigte, Arbeitslose und Hilfebedürftige enthalten sind. Der Datensatz gibt damit insbesondere auch Aufschluss über den Status in Phasen der Nichterwerbstätigkeit, Selbständigkeit oder Beschäftigung im Beamtenverhältnis.
3. die zusätzliche bzw. wesentlich differenziertere Erfassung relevanter Merkmale wie z.B. Einstellungen, Erwerbspotenzial oder Arbeitssuchverhalten.

Gegenüber den bereits verfügbaren Personen- oder Haushaltsbefragungen zielt PASS insbesondere auf die Verbesserung der Datenlage in Bezug auf die folgenden Punkte:

1. Die hohe Fallzahl im Arbeitslosengeld-II-Leistungsbezug (in Welle 1: 10.197 befragte Personen in 7.350 Haushalten, die eine solche Leistung erhalten) ermöglicht differenziertere Analysen – zum Beispiel zur Wirkung des SGB II auf bestimmte Zielgruppen – und genauere Schätzungen von Kennwerten und Modellkoeffizienten als Datensätze, in denen Leistungsempfänger nur proportional zu ihrem Bevölkerungsanteil enthalten sind.
-

2. Die Erhebung zusätzlicher Merkmale wie der Intensität und Art von Kontakten zu Trägern der Grundsicherung oder die Teilnahme an Arbeitsmarktprogrammen ermöglicht Analysen zur Bedeutung institutioneller Hilfeleistungen für die Armutspopulation.
3. Die Verknüpfung der Befragungsdaten mit Prozessdaten der Bundesagentur für Arbeit, kann zum einen zur Validierung der erhobenen Merkmale genutzt werden, zum anderen ermöglicht sie Analysen, in denen die höhere Messgenauigkeit der Prozessdaten mit weiteren Variablen und dem Haushaltskontext aus dem Survey kombiniert werden kann.

#### b) Instrumente und Befragungsschwerpunkte

In jedem Haushalt wurde zunächst ein Haushaltsinterview mit dem Haushaltsvorstand durchgeführt. Anschließend wurde angestrebt, mit jeder Person im Haushalt ab 15 Jahren ein Personeninterview zu führen. Für Personen ab 65 Jahren wurde nur eine Kurzfassung des Personenfragebogens (der sogenannte Seniorenbogen) eingesetzt.

Abbildung 1 gibt einen Überblick über das Themenspektrum der Befragung. Die soziodemographischen Merkmale werden überwiegend mit Standarditems erhoben. Besonders zu erwähnen sind hier die detaillierte Erfassung des Migrationshintergrunds bis in die 3. Generation (Fragebogen p191-p202, Variablenkürzel PMI) und die spezifische Erfassung des Erwerbsstatus, welche die im Leistungsbezug besonders relevante Erfassung paralleler Status besser zulässt als die Standarddemographie (siehe Items p24-p55 im Fragebogen, Variablenkürzel PET).

Der Frageblock zur materiellen Lage umfasst neben einer detaillierten Erhebung von Einkommens- und Vermögensbestandteilen (in der 1. Welle allerdings noch nicht alle Einkommenskomponenten) auch einen Deprivationsindex (Fragebogen: hh7, hh8, p141, Variablenkürzel: HLS), mit dem die Ausstattung der Haushalte mit (Konsum-) Gütern erfasst wird.

Mit dem Fragekomplex zur sozialen Lage werden nicht-materielle Aspekte der Lebenssituation erfasst, die in den Prozessdaten der BA überwiegend nicht abgebildet sind: Insbesondere sind hier die Module zu Gesundheit (p152-p164, PG) und zur sozialen

---

Einbindung (p143-p146, PSK) zu nennen, zu denen die dritte Welle des PASS einen Schwerpunkt anbieten wird.

Daneben werden Einstellungen und Orientierungen (p3, p23, p69-p70, p129, p190, PEO) sowie Aktivitäten (z.B. Arbeitssuche p116-p138, PAS) erhoben, die ebenfalls in den Prozessdaten nicht vorhanden sind.

Schließlich erfasst das PASS den Bezug von Versicherungs- bzw. Sozialleistungen (hh 42-hh55, p71-p75), den Kontakt zu den hierfür zuständigen Institutionen (p76-p90, PTK), sowie die Teilnahme an Maßnahmen (p92-p115, ALM) in jeweils eigenen Modulen.

**Abbildung 1: Befragungsschwerpunkte**

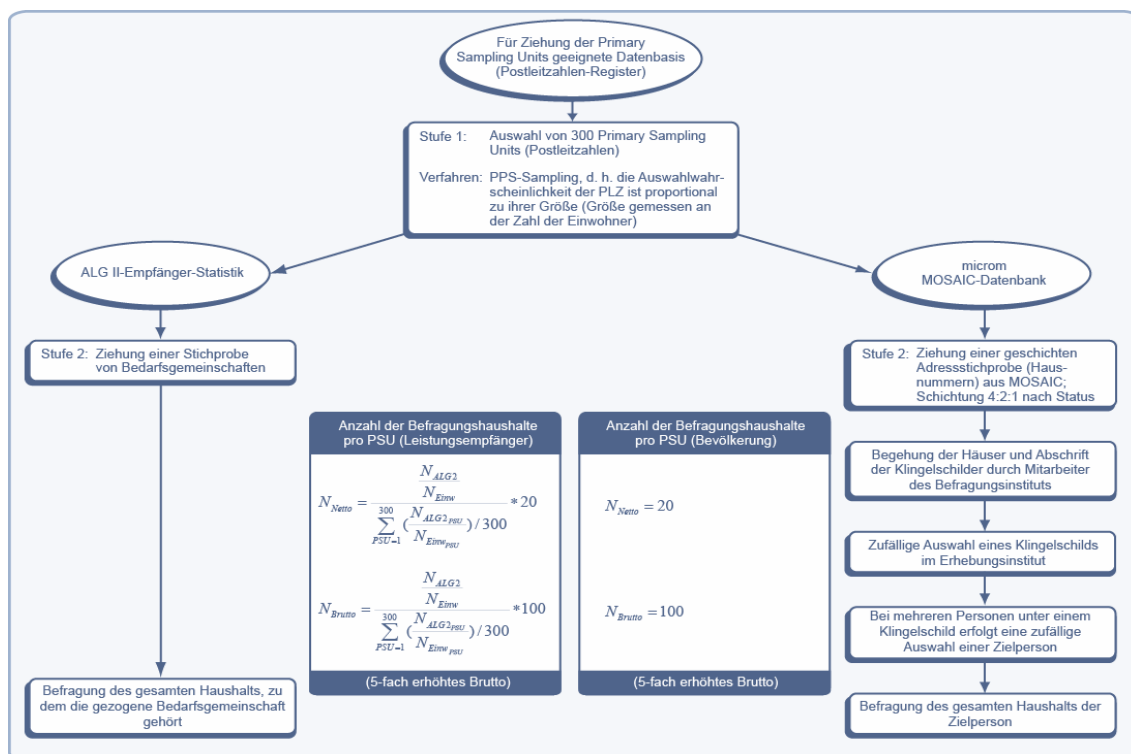
Bereich Ebene	Demographie	Materielle Lage	Soziale Lage	Orientierungen & Aktivitäten	Leistungsbezug
<b>Haus-</b> <b>halt</b>	Liste der Haushaltsmitglieder (incl. Alter und Geschlecht)	Lebensstandard des HH (Deprivationsindex)	Wohnung / Wohnumfeld		ALGII-Leistungsbezug (Spelldaten)
	Verwandtschaftsverhältnisse	Haushaltsnettoeinkommen			
	Kinder, die nicht im HH leben	Vermögen			
		Schulden und Kredite			
<b>Person</b>	Schul- und Berufsbildung	Erwerbseinkommen, brutto und netto	Gesundheit (SP Welle 3)	Geschlechterrollenorientierungen	ALG I -Leistungsbezug (Spelldaten)
	Erwerbsstatus	Sonstige Einkommenskomponenten (z.B. Sonderzahlg.)	Soziale Einbindung / Netzwerke, (SP Welle 3)	Erwerbsorientierungen	Institutionelle Einbindung (z.B. Trägerkontakte, Angebote)
	Beruf		Subjektive Bewertungen (z.B. soziale Selbstpositionierung)	Arbeitssuche (Suchaktivitäten, Reservationslohn; SP W1)	Maßnahmeteilnahme (z.B. Typ, Dauer; Spelldaten)
	Soziale Herkunft (Bildung (W1) & Beruf (W2) der Eltern)		ISCO-basierte Maße	Lebenszufriedenheit	Andere Sozialleistungen (z.B. Renten)
	Staatsbürgerschaft / Migrationshintergrund			Ängste und Sorgen	

c) Erhebungsdesign

Die vielfältigen Anforderungen lassen sich nur durch ein komplexes Studiendesign realisieren. Wesentliche Überlegungen hierzu beinhaltet ein Gutachten von Schnell (2007), auf dessen Basis im IAB das endgültige Design der Studie entwickelt wurde. Insbesondere die folgenden Punkte sind dabei hervorzuheben (vgl. Rudolph und Trappmann 2007):

1. Um trotz eines möglichst kosteneffizienten Vorgehens Erreichbarkeitsprobleme bei telefonischen Befragungen von Personen im Niedrigeinkommensbereich bzw. mit Transferleistungsbezug zu vermeiden, wurde ein CATI/CAPI-Mix gewählt<sup>1</sup>.
2. Da davon auszugehen ist, dass ein beträchtlicher Anteil der Zielpopulation nicht über ausreichende Deutschkenntnisse verfügt, wurden die Erhebungsinstrumente in die häufigsten Muttersprachen von Migranten in Deutschland, nämlich ins Türkische und ins Russische und zudem ins Englische übersetzt.
3. Für alle Personen, die ihre Zustimmung erteilen, wurde außerdem eine Verknüpfung der Befragungsdaten mit den Prozessdaten der Bundesagentur für Arbeit hergestellt. Diese Zuspiegung, die für 72,0 % der befragten Personen zwischen 15 und 64 Jahren gelang, ist allerdings nicht Teil des Scientific Use Files, sondern kann aus Datenschutzgründen nur vor Ort im Forschungsdatenzentrum der BA im IAB genutzt werden.

Abbildung 2: Das Stichprobendesign



<sup>1</sup> Ein kürzlich durchgeführter Mixed Mode Survey (CATI und CAPI) des IAB (Meßmann et al. 2008) ergab, dass mehr als 20 % der befragten Leistungsempfänger (Arbeitslosengeld-II-Empfänger und Sozialhilfeempfänger) keinen Festnetzanschluss besitzen. Diese Personen erwiesen sich im Vergleich mit den übrigen Leistungsempfängern als deprivierter.

4. Da auch Zugangsprozesse in den Leistungsbezug untersucht werden sollen, und zudem für viele Fragestellungen Vergleiche der Empfänger von Leistungen nach dem SGB II mit der allgemeinen Bevölkerung notwendig bzw. sinnvoll sind, wurde neben einer Stichprobe von Haushalten in denen mindestens eine Person im Juli 2006 SGB-II-Leistungen bezogen hat auch eine Bevölkerungsstichprobe befragt. Diese auf einem Verzeichnis von Gebäudeadressen der Firma Microm basierende Bevölkerungsstichprobe wurde zudem nach dem Statusindex dieses Anbieters disproportional geschichtet, so dass Personen mit geringem sozialem Status und somit erhöhtem Zugangsrisiko zum Leistungsbezug höhere Inklusionswahrscheinlichkeiten besitzen (siehe Abbildung 2)<sup>2</sup>. Von einer Verwendung der Daten ohne die resultierenden Gewichte, deren Generierung und Verwendung in Kapitel 5 beschreiben wird, wird aufgrund des komplexen Designs dringend abgeraten.

Abbildung 2 stellt die wesentlichen Elemente des Stichprobendesigns grafisch dar. Grundgesamtheit der Bevölkerungsstichprobe sind alle Privathaushalte in der Bundesrepublik Deutschland. Grundgesamtheit der Prozessdatenstichprobe sind alle Haushalte, in denen mindestens eine Bedarfsgemeinschaft (und damit mindestens eine Person) Leistungen nach dem SGB II bezieht.

---

<sup>2</sup> Ausführlich ist das Verfahren in Rudolph und Trappmann (2007) dargestellt. Die Ergebnisse der disproportionalen Schichtung werden in Trappmann et al. (2007) zusammengefasst.

## 2. Zentrale Kennwerte der Studie

Im folgenden Kapitel soll ein kurzer Überblick über zentrale Kennziffern der Studie, wie z.B. Stichprobengröße oder Ausschöpfungsquoten gegeben werden. Diese werden i.d.R. sowohl für die beiden oben dargestellten Teilstichproben (vgl. Abbildung 2) 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.

### a) Stichprobenumfang

Insgesamt umfasst die Stichprobe des PASS 12.794 Haushalte. Tabelle 1 stellt die Verteilung dieser Haushalte über die beiden Teilstichproben und die Erhebungsmodi dar.

**Tabelle 1: Stichprobenumfang auf Haushaltsebene**

n	CATI	CAPI	Gesamt
Teilstichprobe 1 (BA)	5.139	1.665	6.804
Teilstichprobe 2 (Microm)	3.316	2.674	5.990
<b>Gesamt</b>	<b>8.455</b>	<b>4.339</b>	<b>12.794</b>

Diesen 12.794 Haushalten entsprechen auf Personenebene 18954 Befragte. Tabelle 2 stellt die entsprechende Verteilung dieser Personen über Stichproben und Erhebungsmodi dar.

**Tabelle 2: Stichprobenumfang auf Personenebene**

n	CATI	CAPI	Gesamt
Teilstichprobe 1 (BA)	7.079	2.307	9.386
Teilstichprobe 2 (Microm)	5.335	4.233	9.568
<b>Gesamt</b>	<b>12.414</b>	<b>6.540</b>	<b>18.954</b>

Wie im ersten Kapitel dargestellt, gab es für Befragte ohne ausreichende Deutschkenntnisse die Möglichkeit, sich in einer anderen Sprache als Deutsch befragen zu lassen. Als alternative Interviewsprachen wurden dabei Türkisch, Russisch sowie Eng-



lisch angeboten. Tabelle 3 gibt an, wie viele Haushalte bzw. Personen in den einzelnen Erhebungssprachen befragt wurden.

**Tabelle 3: Stichprobenumfang fremdsprachige Interviews**

n	Haushalte	Personen
Russisch	276	426
Türkisch	163	285
Englisch	9	12
<b>Gesamt</b>	<b>448</b>	<b>723</b>

### b) Response Raten

In einer Haushaltserhebung kann zwischen der Response-Rate auf Haushaltsebene und der Response-Rate innerhalb von Haushalten unterschieden werden. Mit „Ausschöpfung auf Haushaltsebene“ ist der Anteil der verwertbaren Haushalte (hier: vollständiges Haushaltsinterview und mindestens ein vollständiges Personeninterview) an der Summe aus verwertbaren Haushalten und nicht-neutralen Ausfällen gemeint. Mit Response-Rate innerhalb von Haushalten bezeichnen wir den mittleren Anteil der Personen ab 15 Jahren innerhalb von Haushalten, die teilgenommen haben, wenn ihr Haushalt teilnimmt.

#### *Response Rate auf Haushaltsebene*

TNS Infratest Sozialforschung, das mit der Durchführung der Feldarbeit beauftragte Institut, berichtet die folgenden Ausschöpfungsquoten auf Haushaltsebene<sup>3</sup>

Response-Rate Teilstichprobe 1 (BA):	35,1 %
Response-Rate Teilstichprobe 2 (Microm):	26,6 %
Response-Rate insgesamt:	30,5 %

<sup>3</sup> Da sich in der Deutschen Sozialforschung leider keine Standards für die Berechnung von Ausschöpfungsquoten etabliert haben, geben wir diese Werte hier so wieder. Im Vergleich zu anderen kürzlich durchgeführten Mixed-Mode-Erhebungen vergleichbarer Populationen, nämlich der LSS 2005 (Meßmann et al. 2008, Infas 2006) und der Kundenbefragung im Rahmen der Evaluation der Experimentierklausel (ZEW et al. 2007), liegt hier eine strengere Definition von „neutral“ und von „Ausfall“ zugrunde. Der Anteil neutraler Ausfälle ist bei PASS trotz vergleichbarer Ziehungsgrundlagen der Stichproben 7 bzw. 18 Prozentpunkte geringer. Bei Verwendung vergleichbarer Maßstäbe erreichen alle drei Studien eine fast gleich Ausschöpfung in der Population der Leistungsempfänger. Legt man die strengeren Maßstäbe der RR1 der American Association for Public Opinion Research (AAPOR 2006) zugrunde, so liegen die Ausschöpfungen der Leistungsempfängerstichproben aller drei Studien um 30 %. Eine Aufschlüsselung der finalen Dispositionscores ist im Methoden- und Feldbericht von TNS Infratest (Hartmann et al. 2008, Tabelle A2.1 und A2.2) zu finden

### *Response Rate innerhalb der Haushalte*

Innerhalb der Haushalte werden folgende Ausschöpfungsquoten berichtet:

Response-Rate Teilstichprobe 1 (BA):	85,6%
Response-Rate Teilstichprobe 2 (Microm):	84,3%
Response Rate Gesamt:	85,0%

Die auf Haushaltsebene relativ geringe Ausschöpfungsquote in der Teilstichprobe für die allgemeine Bevölkerung ist vor dem Hintergrund der subjektiven Bedeutsamkeit des Themas zu bewerten. Schnell (1997) argumentiert, dass es sich bei dem in vielen Studien anzutreffenden „Mittelschichtsbias“ (d.h. Personen aus mittleren sozialen Schichten sind in Bevölkerungsumfragen häufig überrepräsentiert) in Wirklichkeit um einen Bias handelt, der auf der einen Seite durch die schlechte Erreichbarkeit und die hohen Opportunitätskosten der oberen sozialen Schichten, sowie auf der anderen Seite durch die häufig geringe Bedeutsamkeit des Themas für die unteren sozialen Schichten hervorgerufen wird (Schnell 1997, 201 ff.). Unsere Daten scheinen genau das zu bestätigen. In einer Erhebung, in der es um das Thema soziale Sicherung geht, verschieben sich allerdings die Personengruppen, für die das Thema von hoher Bedeutsamkeit ist. Dies sind vor allem die Betroffenen. Dagegen ist es nicht gelungen, die Allgemeinbevölkerung, die ja im Design als Kontrollgruppe fungiert, im gewünschten Maße von der Bedeutsamkeit der Studie zu überzeugen.

Daneben wurde eine bessere Ausschöpfung auch deshalb nicht erreicht, weil am Ende der Feldarbeit eine sehr große Zahl nicht erreichter Haushalte stand, die in der BA-Stichprobe mit über 35 % besonders hoch ist. Die Gründe hierfür sind vielfältig. Probleme mit dem Adressmaterial – die vor allem im Bereich der optierenden Kommunen vorliegen, für die die BA keine regelmäßigen Adress-Updates erhält<sup>4</sup> – belegen auch die ähnlich hohen Zahlen in den beiden Referenzstudien (vgl. Fußnote 3). Das Design von PASS sah vor, diesen bekannten Schwierigkeiten mit dem zusätzlichen CAPI-Feld und intensiven Recherchen zu begegnen. Die Kombination aus einer kurzen Feldzeit, extrem kurzer Vorbereitungszeit für das Erhebungsinstitut und Fehlern in der Feldsteuerung führte aber letztlich dazu, dass die Fälle mit anfangs fehlerhaften Adressen bzw.

---

<sup>4</sup> Erstaunlicherweise zeigen die Selektivitätsmodelle jedoch einen signifikant positiven Effekt der Datenherkunft aus optierenden Kommunen auf die Erreichbarkeit.

Telefonnummern nur noch teilweise und mit sehr kurzen Feldzeiten abgearbeitet werden konnten.

PASS kann den relativ hohen Nonresponse jedoch durch eine ausgesprochen günstige Datenlage zumindest teilweise kompensieren, da nicht nur die üblicherweise in Surveys zur Korrektur von Nonresponse verwendeten regionalen und demographischen Variablen vorliegen: Darüber hinaus können im Falle der Prozessdatenstichprobe individuelle Informationen zu Arbeitslosigkeit, höchstem allgemeinbildendem Schulabschluss, Bedarfsgemeinschaftsgröße und -typ genutzt werden. In beiden Stichproben konnten zudem microgeographische Informationen – das sind kleinräumig aggregierte Informationen z.B. zu Haustypen, Altersstruktur, Status, Ausländeranteil, Umzugsvolumen, etc. – der Firma Microm zugespielt und für Ausfallanalysen verwendet werden. Mit diesen Modellen befasst sich das Gewichtungskapitel (Kapitel 5)<sup>5</sup>. Nur wenn der Ausfallprozess konditioniert auf all diese Variablen nicht „zufällig“ ist, kommt es zu Nonresponse-Bias. Für einige der erhobenen Variablen liegen zudem Eckwerte für die beiden Grundgesamtheiten vor. Diese Eckwerte werden für die Kalibrierung des Panels verwendet (vgl. ebenfalls Kapitel 5). Damit werden Verzerrungen in Bezug auf diese Variablen mit bekannter Verteilung in der jeweiligen Grundgesamtheit beseitigt.

### c) Panel- und Zuspielbereitschaft

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. Hier konnten in beiden Fällen hohe Zustimmungsraten erzielt werden:

Panelbereitschaft:	93,8%
Zustimmung zur Zuspielung von Prozessdaten:	79,8%

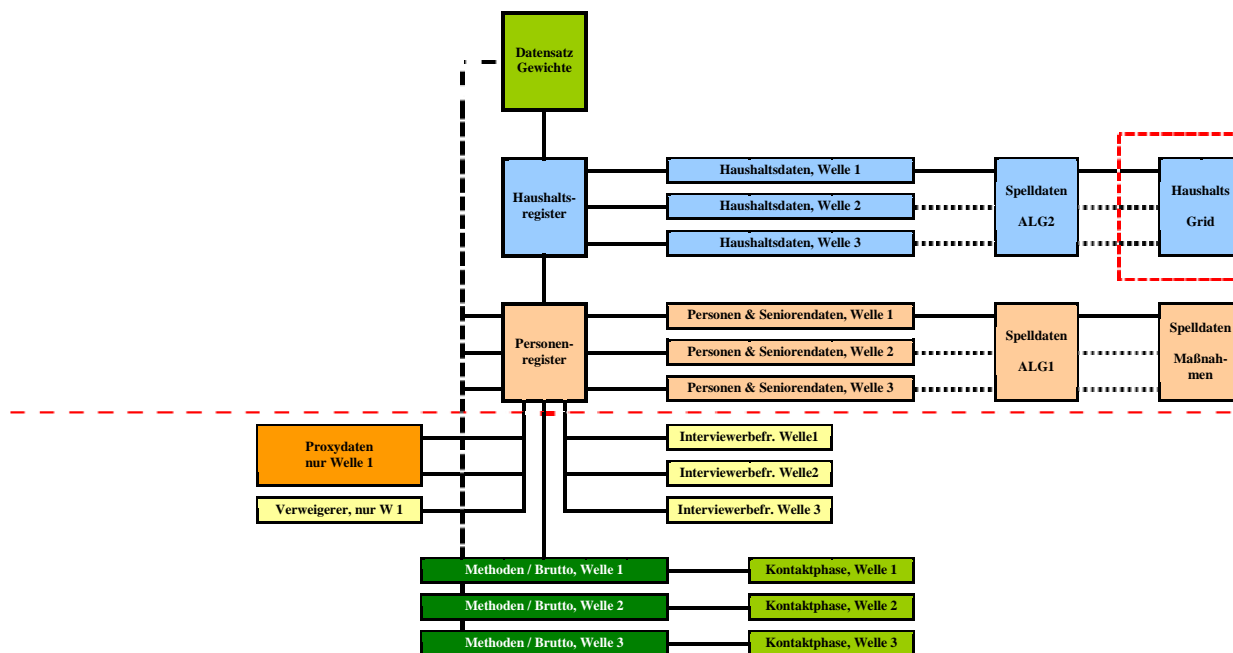
---

<sup>5</sup> Die dort erreichte Modellgüte, die ohnehin zum Teil schon sehr hoch ist – mit einem maximalen McFadden  $R^2=0,1432$  für die Modellierung der Erreichbarkeit in der Bevölkerungsstichprobe – lässt sich durch Berücksichtigung der Variable „Tranche“ noch einmal deutlich steigern (auf 0,29 für die Erreichbarkeit in der Bevölkerungsstichprobe und 0,069 für die Erreichbarkeit in der Prozessdatenstichprobe). Das weist darauf hin, dass gerade bei den späten Einsatztranchen die kurze Feldzeit starken Einfluss auf die Erreichbarkeit hatte. Da die Tranche mit keiner inhaltlich relevanten Variable zusammenhängen sollte, weil sie eine zusätzliche Partition der Stichprobe darstellt, wurde diese Variable jedoch in den Ausfallanalysen nicht verwendet. Nichtsdestotrotz belegt dieses Ergebnis, dass ein großer Teil des Ausfallprozesses mit den vorhandenen Variablen erklärt werden kann.

### 3. Die Datensatzstruktur im Überblick

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. Abbildung 3 stellt dar, wie eine solche Datenstruktur für das PASS aussehen könnte.

Abbildung 3: Übliche Struktur für die Aufbereitung von Paneldaten

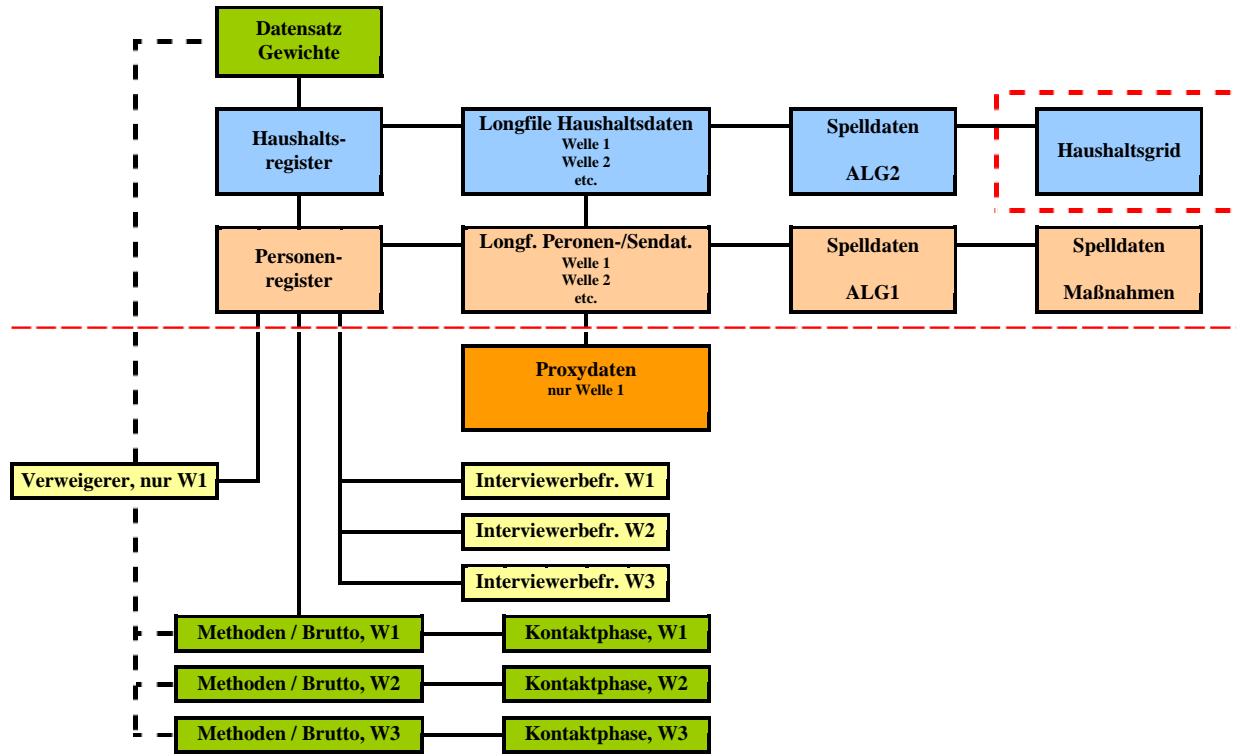


Nach einem längeren Diskussionsprozess und der Konsultation verschiedener Experten<sup>6</sup> wurde jedoch beschlossen, die Daten in einer alternativen Struktur aufzubereiten. Dabei werden die zentralen Datensätze (Haushalts- und Personendaten) nicht als jährliche Rechteckfiles aufbereitet, sondern als wellenübergreifende Long-Files, in denen alle einer Person/einem Haushalt zugeordneten Erhebungswellen untereinander geschrieben und über eine Variable für die Jahreszahl identifiziert werden (vgl. Abbildung 4). Vorteil dieser Variante ist, dass die Daten bereits in der Form vorliegen, wie sie für typische Längsschnitthanwendungen wie z.B. Verweildaueranalysen benötigt werden

<sup>6</sup> Unser Dank gilt hier vor allem der SOEP-Gruppe des DIW, insbes. Herrn Prof. Dr. Schupp, Herrn Dr. Krause und Herrn Dr. Goebel sowie Herrn Bender vom FDZ der BA im IAB.

und in die sie ansonsten erst mühsam gebracht werden müssen. Gebräuchliche Umformatierungen sind außerdem vergleichsweise einfach zu bewerkstelligen.<sup>7</sup>

Abbildung 4: Datensatzstruktur des Panels Arbeitsmarkt und Soziale Sicherung (PASS)



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 Longfiles 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. Noch vor wenigen Jahren hätte sich die von uns gewählte Datenstruktur aufgrund dieses Arguments schlicht verboten. Angesichts der Entwicklungen im IT-Bereich und der heutzutage selbst auf einfachen Desktop-PCs verfügbaren Prozessoren und Speichergrößen ist dieses Argument u.E. inzwischen jedoch zu vernachlässigen.

<sup>7</sup> Umformatierungen ins ebenfalls gebräuchliche Wide-Format, in dem die einer Einheit zuzuordnenden Informationen in unterschiedliche Spalten des Datensatzes geschrieben werden, sind mit Statistikprogrammen wie z.B. Stata automatisch möglich (mit Hilfe des ‚reshape‘-Befehls). Die Erstellung von Rechteckfiles für einzelne Jahre ist ebenfalls jederzeit über eine einfache Filteranweisung machbar.

Damit bleibt als einziger weiterer Nachteil, dass das Zuschlüsseln von Informationen aus Haushalts- und Personendatensatz etwas komplizierter wird, da hierzu zusätzlich zur eigentlichen Schlüsselnummer ein Wellenidentifikator benötigt wird, z.B. in Form einer Jahresvariable. Im Vergleich zu den mit dem Long-Format verbundenen Vorteilen erscheint dieser Nachteil jedoch als gering.

Neben den Hauptdatensätzen enthält der Scientific-Use-File die in Spellform aufbereitete Daten zum Bezug von Arbeitslosengeld I und II sowie zur Teilnahme an arbeitsmarktpolitischen Maßnahmen, Registerdatensätze auf Haushalts- und Personenebene sowie Gewichtungsdatensätze. Daneben existieren noch Datensätze mit detaillierten Informationen zu den Verwandtschaftsverhältnissen im Haushalt (Haushaltsgrid) sowie mehrere nicht inhaltliche Datensätze (Interviewernachbefragung, Bruttodatensätze etc.), die aufgrund ihrer rein technischen Natur oder aus Datenschutzgründen jedoch nicht im Scientific-Use-File enthalten sind (in der Darstellung sind diese Datensätze durch eine rote gestrichelte Linie von den übrigen Datensätzen abgetrennt).<sup>8</sup>

---

<sup>8</sup> Falls für bestimmte Forschungsvorhaben der Zugriff auf die entsprechenden Teildatensätze unabdinglich sein sollte, möchten wir Sie bitten, sich an das Forschungsdatenzentrum zu wenden, um eine geeignete Datenzugangsmöglichkeit zu finden. Wie diese aussehen kann, wird vom jeweiligen Forschungsvorhaben und den dafür benötigten Daten abhängen.

---

## 4. Das Konzept der Variablenbenennung

### a) Generelles

Bei der Variablenbenennung gibt es u.E. zwei Hauptalternativen, zwischen denen es sich zu entscheiden gilt. Die erste ist eine am Fragebogen orientierte Benennung der Variablen, wie sie z.B. im SOEP praktiziert wird. Vorteil dieser Art der Benennung ist, dass sich die den Variablen zuzuordnenden Items im Fragebogen schnell lokalisieren lassen und so den Wert des Fragebogens als Dokumentationsinstrument deutlich erhöhen. Zentraler Nachteil dieses Vorgehens ist, dass identische Items aufgrund von Veränderungen im Fragebogen unterschiedlich benannt werden, so dass mit zunehmender Verfügbarkeit von Panelwellen selbst für einfache Trendanalysen umfangreiche Vorarbeiten für die Zusammenstellung und ggf. Umbenennung der gewünschten Variablen notwendig werden.

Die zweite Hauptalternative ist die Vergabe unabhängiger Variablennamen, die (bis auf einen evtl. benötigten Wellenindikator) über die Wellen hinweg konstant gehalten werden. Vor- und Nachteile dieser Strategie verhalten sich spiegelbildlich zur ersten Alternative: die Identifizierung der einem Item zuzuordnenden Variablen über die Wellen hinweg ist unproblematisch, die Nutzung des Fragebogens als Dokumentationsinstrument wird jedoch schwieriger, da es nicht mehr möglich ist, aus dem Variablennamen auf die Position des zugehörigen Items im Fragebogen zu schließen.

Gerade bei einer längerfristig angelegten Panelstudie überwiegen u.E. die Vorteile fixer Variablennamen deren Nachteile deutlich. Zudem erfordert die bereits dargestellte Entscheidung für die Organisation der Daten im Long-Format die Verwendung einheitlicher Variablennamen.

### b) Variablentypen

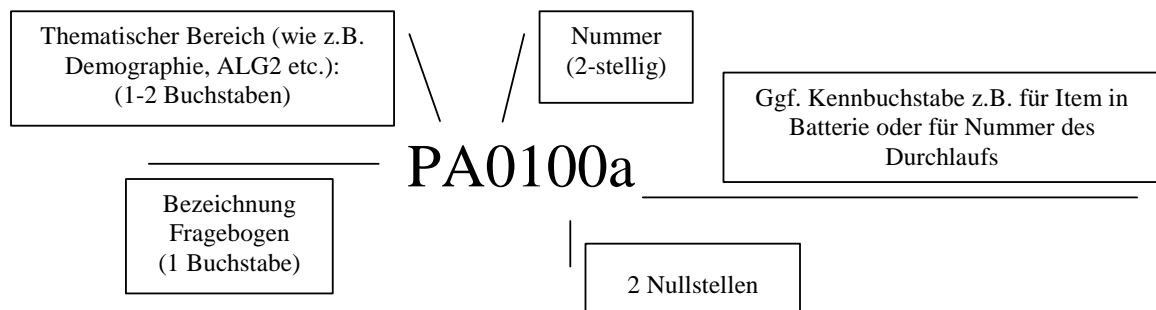
Im Codebuch werden drei verschiedene Variablentypen unterschieden:

1. Systemvariablen: Systemvariablen sind Variablen die im Rahmen des Erhebungsprozesses erzeugt wurden. Sie können zum einen dazu genutzt werden, die im Fragebogen dokumentierten Filterführungen nachzuvollziehen. Zumindest einige Systemvariablen können zum anderen auch von inhaltlichem bzw. methodischem Interesse sein, so z.B. der Interviewmodus oder die Zahl der im Haushalt lebenden

Kinder einer bestimmten Altersgruppe. Systemvariablen erhalten individuelle Bezeichnungen, für die Kleinbuchstaben und Ziffern kombiniert werden können. Zu den Systemvariablen zählen auch die Gewichte.

2. Erhobene Variablen: Erhobene Variablen sind Variablen, die in dieser Form direkt im Fragebogen erhoben wurden. Diese Variablen erhielten einen komplett neuen, abstrakten Variablennamen. Das dieser Benennung zu Grunde liegende Konzept ist in Abbildung 5 anhand eines Beispiels dargestellt.

**Abbildung 5: Darstellung des Schemas zur Variablenbenennung**



- Die 1. Stelle des Variablennamens bezeichnet die Fragebogenebene, also Haushalts- bzw. Personendatensatz mittels des Buchstabens H bzw. P (Großbuchstabe).
- Danach folgen ein oder zwei Großbuchstaben, die das inhaltliche Themengebiet der Variable angeben (für eine vollständige Liste vgl. Tabelle 4).
- In den in Spellform aufbereiteten Datensätzen wird auf das einführende P bzw. H verzichtet. Stattdessen erhalten die Variablen dieser Datensätze eine einheitliche, aus zwei oder drei Buchstaben bzw. zwei Buchstaben und einer Ziffer bestehende, thematische Benennung.
- Auf die einführende Buchstabenkombination folgen dann zwei fortlaufend vergebene Ziffern, welche die Nummer der Frage innerhalb des Themenbereichs bezeichnen.
- An diese beiden Ziffern schließen sich zwei Nullstellen an, die in späteren Wellen das Einfügen weiterer Variablen ermöglichen sollen. Bisher wurde von dieser Möglichkeit nur dort Gebrauch gemacht, wo zusätzlich zur ursprünglichen Versi-



on der Variable eine zweite Variante inklusive zugedeter offener Angaben zur Verfügung gestellt wurde. Für diese Variablen wurde die letzte Nullstelle auf '1' gesetzt (also z.B. PA0101a anstatt PA0100a).

**Tabelle 4: Liste der in den Variablennamen verwendeten thematischen Bezeichnungen**

Personenebene		Haushaltsebene	
Kürzel	Thematischer Bereich	Kürzel	Thematischer Bereich
PA	Allgemein	HA	Allgemein
PAS	Arbeitssuche	HD	Demographie
PB	Bildung	HEK	Einkommen
PD	Demographie	HKI	Kinderbetreuung
PEO	Einstellungen und Orientierungen	HLS	Lebensstandard
PEK	Einkommen	HW	Wohnung
PET	Erwerbstätigkeit		
PG	Gesundheit		
PLS	Lebensstandard		
PMI	Migration		
PP	Pflege		
PSH	Soziale Herkunft		
PSK	Soziale Kontakte		
PTK	Trägerkontakte		
AL	Bezug ALG 1 (Spelldaten, Personenebene)	AL2	Bezug ALG 2 (Spelldaten, Haushaltsebene)
ALM	Arbeitsmarktpolitische Maßnahmen (Spelldaten, Personenebene)		

- Bei Variablen für Items aus Multi-Item-Batterien oder innerhalb einer verschleiften Erhebung kann sich ggf. noch ein Kleinbuchstabe für das Item bzw. den Durchlauf innerhalb der Schleife anschließen.
3. Generierte Variablen: Die Gruppe der generierten Variablen zerfällt nochmals in zwei Untergruppen. Bei den generierten Variablen im engeren Sinne handelt es sich um umfangreichere Zusammenfassungen von Variablen, z.B. aus der offenen und kategorialen Erhebung zusammengefasste Einkommensmaße, oder um komplexere Konstruktvariablen, wie z.B. das bedarfsgewichtete Haushaltseinkommen oder Klassifikationen für Bildung (wie ISCED oder Casmin) bzw. Status (z.B. EGP,

ESEC). Generierte Variablen in diesem engeren Sinne werden mit möglichst prägnanten bzw. 'klingenden' Klarnamen gekennzeichnet. Die Schreibung dieser Klarnamen erfolgt grundsätzlich in Kleinbuchstaben. Für einen Überblick über die generierten Variablen vgl. Kapitel 6.

Die zweite Gruppe generierter Variablen umfasst solche, bei denen lediglich offene Angaben zu einer direkt erhobenen Variable hinzugecodet wurden. Obwohl es sich hierbei streng genommen ebenfalls um generierte Variablen handelt und diese im Auszählungsteil des Codebuchs auch als solche klassifiziert sind, wurden keine Klarnamen verwendet, sondern ihre Bezeichnung wurde an die ursprünglichen Variablen angeglichen und lediglich die '0' in der letzten Ziffer wurde durch eine '1' ersetzt.

---

## 5. Dokumentation der Datenbereinigung

Die Datenbereinigung erfolgte in zwei Schritten: Zunächst erfolgte eine detaillierte Überprüfung und ggf. Bereinigung der Filterführung und ein Setzen spezifischer Codes für die Missing-Values. Anschließend wurde bei ausgewählten Items eine Überprüfung der Plausibilität der von den Befragten gemachten Angaben durchgeführt. 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. Zusätzlich war es durch Probleme mit einer fremdsprachigen Version des Befragungsinstruments (russisch) notwendig geworden, Teile der Befragung für diese Population nachzuerheben. Die Integration dieser nach-erhobenen Daten erfolgte ebenfalls im Rahmen der Datenbereinigung.

### a) Filterprüfung

In der Filterbereinigung wurde die im Instrument vorgesehene Filterführung in den beiden vom Feldinstitut ausgelieferten Datensätzen (je einer für die Daten des Personen- und des Haushaltsfragebogens) mithilfe eines Statistikprogramms nachvollzogen. Waren bestimmte Items durchlaufen worden, obwohl der Wert der relevanten Filtervariable etwas Anderes verlangt hätte (wurden also z.B. detaillierte Angaben zur Berufsausbildung erhoben, obwohl der/die 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.<sup>9</sup> Des Weiteren kam es vor, dass Items nicht erhoben wurden, obwohl dies laut der entsprechenden Filtervariable notwendig gewesen wäre (z.B. wenn keine weiteren Angaben zur Berufsausbildung erhoben wurden, obwohl der/die Befragte angegeben hatte, eine solche Ausbildung gemacht zu haben). In diesen Fällen wurde der spezifische Missing Code '-4' (Frage irrtümlich nicht gestellt) vergeben.

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 5 im Überblick dargestellt. '-1' und '-2' sind die einheitliche Umcodierung der in der Befragung erhobenen 'WN' und 'KA' Wer-

---

te. '-3' ist der generelle 'TNZ'-Code für aufgrund der Filterführung nicht gestellte Fragen. '-4' wurde bereits oben erläutert. 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.

**Tabelle 5: 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“

Der Wert '-8' ist ein spezifischer Missing Code, der im Rahmen der Plausibilitätsprüfung vergeben wurde – für eine detaillierte Darstellung vgl. den folgenden Abschnitt. '-9' ist ein Missing Code, der erst ab der zweiten Welle benötigt wird. Er wird immer dann vergeben, wenn ein bestimmtes Item in einer bestimmten Welle nicht erhoben wurde. Aufgrund des weiter oben dargestellten Long-Formats der Daten würden also z.B. in den zur zweiten Welle ausgelieferten Daten nicht nur die dort nicht mehr erhobenen Variablen in Welle 2 den Wert '-9' erhalten. Ebenso würden die dann neu erhobenen Variablen nachträglich für Welle 1 mit '-9' vercodet. Mit dem Code '-10' können Unterschiede zwischen den Fragebogenversionen berücksichtigt werden. I.d.R. wird dieser Code bei den mit dem Seniorenfragebogen befragten Personen immer dann vergeben, wenn eine Frage zwar im Personenfragebogen, nicht aber im kürzeren Seniorenfragebogen enthalten war.

<sup>9</sup> Die Bereinigung erfolgte dabei, wie in solchen Fällen üblich, immer von der zuerst erhobenen zu den später erfragten Items.

## b) Plausibilitätsprüfung

Im Rahmen der Plausibilitätsprüfung wurde eine umfangreiche Liste theoretisch möglicher Widersprüche in den Angaben der Befragten kontrolliert. Zudem wurde die Haushaltsstruktur auf Plausibilität geprüft und die Spelldaten wurden bereinigt. Im Einzelnen wurden die folgenden Schritte durchgeführt:

1. Widerspruchsprüfung: Generell wurden Widersprüche nur dann korrigiert, wenn entweder die Implausibilität 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 implausibler Angaben wurden zwei Strategien verwendet: eine direkte Korrektur der implausiblen Angaben oder die Vergabe eines spezifischen Missing Codes.
    - Eine Korrektur implausibler Angaben wurde ausschließlich dann vorgenommen, wenn mit an Sicherheit grenzender Wahrscheinlichkeit eine Fehleingabe des Interviewers vorlag. Beispiele sind die Angabe eines monatlichen Bruttoeinkommens von € 22.000,- bei einem gleichzeitig angegebenen (aufgrund des Berufes plausiblen) Nettoeinkommen von € 1.750,- im Monat. In diesem Fall wurde z.B. die letzte '0' des Bruttoeinkommens gelöscht. Einen ähnlichen Fall stellte u.E. eine Einkommensangabe von € 99.999,- oder € 99.998,- dar. Hier wurde im Rahmen der Datenbereinigung davon ausgegangen, dass es sich um Fehleingaben der sechsstellig erhobenen Missing-Codes '999998' (WN) bzw. 999999 (KA) handelte und die Angaben wurden ebenfalls entsprechend bereinigt.
    - Allerdings war es bei den durch die Prüfung gefundenen Implausibilitäten 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 Implausibilität 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, wurde im Einzelfall entschieden.
-

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 im Personeninterview berichtete Partnerschaften mit im Haushalt lebenden Personen ergänzt. Im ersten Teil der Prüfungen wurden die Beziehungsinformationen und die demographischen Angaben zu den einzelnen Haushaltsmitgliedern kombiniert, um unplausible Haushaltskonstellationen zu identifizieren. Für die in diesen Prüfungen als unplausibel identifizierten Haushalte, wurden Einzelfallentscheidungen unter Einbeziehung der gesamten Haushaltsstruktur und anderer Interviewinformationen (z.B. zum Familienstand im Personeninterview) getroffen. Unplausible Beziehungskonstellationen wurden dabei als solche markiert (,-8') bzw. auf Basis hinzugezogener Informationen über den Haushaltskontext korrigiert, sofern es sich plausibel 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. Darüber hinaus wurden Prüfungen durchgeführt, bei denen jeweils drei Verwandtschaftsbeziehungen 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 B ist leibliches Kind von Person C. Person A ist leibliches Kind von Person C. Wurde im Rahmen dieser Prüfung eine solche oder ähnliche unplausible Kombination von Beziehungen identifiziert, so wurden diese als unplausible Angaben (,-8') markiert.
  3. Bei der Überprüfung der Spelldaten wurde einerseits die korrekte zeitliche Reihenfolge der Spells überprüft. Waren diese im Rahmen der Erhebung fälschlicherweise nicht in chronologischer Reihenfolge aufgenommen worden, wurde die korrekte zeitliche Abfolge wiederhergestellt. Darüber hinaus kam es in einigen Fällen auch vor, dass Spells versehentlich doppelt aufgenommen worden waren. Lagen für eine
-

Person bzw. einen Haushalt zwei komplett identische Spells vor, wurde einer der beiden Spells gelöscht.

#### c) Nacherhebung russisches CATI-Feld

Die russische Version der CATI-Programmierung war aufgrund der kyrillischen Schriftzeichen, die mit der für das Deutsche, Türkische und Englische Instrument verwendeten Software nicht darstellbar waren, in einer eigenen Software neu programmiert worden. Es war dabei, im Gegensatz zu den oben genannten anderen fremdsprachigen Instrumenten, nicht möglich, die deutsche Version der CATI-Programmierung als Grundlage beizubehalten und lediglich die Fragetexte zu übersetzen und auszutauschen.

Im Rahmen dieser Neuprogrammierung ist es bei insgesamt 294 der 432 mit dem russischen Instrument interviewten Personen zu Fehlern in der Filterführung bzw. bei der Generierung von Steuerungsvariablen gekommen. Dies hat eine Nacherhebung einzelner Module des Personenbogens notwendig gemacht und betraf dabei insbesondere die Fragen zu Kontakten mit den Trägern (P76-P90), Maßnahmeteilnahme (P93-P115) und Arbeitssuche (P116-P138).

Von diesen 294 Fällen konnten durch die Nacherhebung insgesamt 202 Personen zu den oben genannten Modulen erneut befragt werden. Dabei ist jede Person jeweils nur zu dem Modul bzw. den Modulen befragt worden, die im Rahmen der ersten, regulären, Befragung fälschlicherweise nicht gestellt wurden. Diese Fälle sind im Personendatensatz mit der Variable „nachbef“ gekennzeichnet, die zudem angibt, welche (Kombination der) Module für den jeweiligen Fall nacherhoben und in den Datensatz integriert wurden. Darüber hinaus sind diejenigen Variablen, bei denen die Fragen P93 bis P115 nacherhoben wurden, in den Spelldaten für die Maßnahmen durch die Variable "nachbmas" identifizierbar.

---

## 6. Generierte Variablen

### a) Variablen incl. offener Angaben

Einige Items der Befragung wurden als geschlossene Items mit offener Restkategorie erhoben. In diesen Fällen wurden i.d.R. zusätzliche Variablen generiert, 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 zumeist lediglich dadurch, dass die '0' in der letzten Ziffer durch eine '1' ersetzt wurde. Die folgenden Variablen gehören zu diesem Typ:

*Personenebene:* PB0401; PB1001; PB1001a – PB1301j; PB1601; PG0901a – PG0901g; PSH0201; PSH0301a – PSH0301i; PSH0501; PSH0601a – PSH0601i; ozulanda – ozulandf

*Haushaltsebene:* AL21301a-AL22203a, AL21301b-AL22203b, AL21301c-AL22203c, AL21301d-AL22203d, AL21301e-AL22203e

### b) Einfache Umcodierungen oder Zusammenfassungen

Dieser Typ umfasst z.B. Variablen, bei denen mehrere, aus erhebungstechnischen Gründen getrennte Items eines Konstrukts zusammengefasst wurden (wie z.B. beim höchsten Bildungsabschluss) oder bei denen wichtige Informationen aus anderen Teildatensätzen zu den Personen- oder Haushaltsdaten hinzugespielt wurden (z.B. Indikatoren für den aktuellen ALG-I- oder ALG-II-Bezug). Die entsprechenden Variablen sind in Tabelle 6 und Tabelle 7 zusammengefasst.



**Tabelle 6: Einfache Umcodierungen oder Zusammenfassungen auf Personenebene**

Variable	Label und Beschreibung	Quellvariablen
aktmassn	<i>Akt. Teiln. an einem v. Arbeitsamt finanz./gefoerd. Programm, generiert</i> Indikator für aktuelle Teilnahme an einer arbeitsmarktpolitischen Maßnahme.	ALM0400/zensiert (massnahmespells) (p96)
alg1abez	<i>Aktueller ALG1-Bezug, generiert</i> Indikator für aktuellen Bezug von Arbeitslosengeld I.	zensiert (ALG1_Spells) (p72; p73)
arbzeit	<i>Woechentliche Arbeitszeit incl. Angaben bei unregelmässiger Arbeitszeit, gen.</i> Generierung einer integrierten Variablen aus offenen und kategorialen Angaben zur Arbeitszeit.	PET1300/PET1400 (p36/p37)
beruf1	<i>Hoehster beruflicher Abschluss, ohne ausl. Abschl. und off. Angaben, generiert</i> Identifikation des höchsten Berufsabschlusses durch Hierarchisierung der von der Befragungsperson gemachten beruflichen Abschlüsse; ohne offene Angaben.	PB1200; PB13000a-PB1300I (p15; p16_01-p16ka)
beruf2	<i>Hoehster beruflicher Abschluss, mit ausl. Abschl. und off. Angaben, generiert</i> Wie beruf1 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).	PB1200; PB13000a-PB1300I; P1400; P1500a-P1500e (p15; p16_01- p16ka; p17; p18_01- p18ka)
brutto	<i>Bruttoeinkommen incl. kategorisierter Angaben, generiert</i> Generierung einer integrierten Variable aus kategorisierter und offener Angabe zum Bruttoeinkommen.	PEK0200-PEK0600 (p56-p61)
bruttokat	<i>Kategorisiertes Bruttoeinkommen, generiert</i> Zusammenfassung der mehrstufig erhobenen kategorialen Brutto-Einkommensangaben zu einer Variable.	PEK0100-PEK0600 (p57-p61)
erwerb	<i>Erwerbsstatus, generiert</i> Generierung einer integrierten Erwerbsstatusvariablen aus den Fragen zur Erwerbstätigkeit und der Frage zum Status der Nichterwerbstätigkeit.	PET0100; PET0300; PET0400; PET0800 (p24, p26, p27, p28, p31)
famstand	<i>Familienstand, generiert</i> Generierung einer integrierten Familienstandsvariable aus den Angaben im Personen- und Haushaltsbogen.	PD0500; epartner (=Übertragsvariable aus dem HH-Bogen; 1=Ehepartner lebt im HH) (p177; epartner)
gebhalbj	<i>Geburtshalbjahr, generiert</i> Indikator: Geburtstag liegt im 1. oder 2. Halbjahr des Geburtsjahres	PD0100 (p1)

Variable	Label und Beschreibung	Quellvariablen
kindzges	<p><i>Anzahl eigene Kinder gesamt (inner- und außerhalb HH), generiert</i></p> <p>Variable für die Gesamtzahl der Kinder einer Person als Summe der Kinder die im Haushalt leben und der Kinder außerhalb des Haushalts.</p>	<p>HD0100x (HH-Befragter); G0600xy / G0800xy (andere Personen im HH); PD0900-PD1100</p> <p>(hh28, hh33, hh35, p183-p185)</p>
kindzihh	<p><i>Anzahl Kinder im HH, generiert</i></p> <p>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).</p> <p><u>Anmerkung:</u> 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.</p>	<p>HD0100x (HH-Befragter); G0600xy / G0800xy (andere Personen im HH) (hh28, hh33, hh35)</p>
massntyp	<p><i>Massnahme: Art der Massnahme, incl. 1-Euro-Jobs, generiert</i></p> <p>Diese Variable wurde als Kombination aus der Frage, ob es sich bei der Maßnahme um einen 1-Euro-Job handle (p102) und der Vercodung der offenen Angabe zur Art der Maßnahme (sofern es sich nicht um einen 1-Euro-Job handelte) gebildet.</p>	<p>ALM1000, ALM 1500 (p102, p107)</p>
mberuf1	<p><i>Höchster ber. Abschluss der Mutter, incl. Mutter im HH, ohne offene Ang., gen.</i></p> <p>Der Berufsabschluss der Mutter wurde nur dann erhoben, wenn diese nicht im Haushalt wohnt. Wohnt sie im Haushalt, musste die Angabe zum Berufsabschluss aus ihrem Personeninterview ergänzt werden.</p>	<p>PSH0300a-l; beruf1 d. Mutter (p205_01-p205_99; beruf1/Mutter)</p>
mberuf2	<p><i>Höchster ber. Abschluss der Mutter, incl. Mutter im HH, mit offenen Ang., gen.</i></p> <p>Der Berufsabschluss der Mutter wurde nur dann erhoben, wenn diese nicht im Haushalt wohnt. Wohnt sie im Haushalt, musste die Angabe zum Berufsabschluss aus ihrem Personeninterview ergänzt werden.</p>	<p>PSH0300a-l; beruf2 d. Mutter (p205_01-p205_99; beruf2/Mutter)</p>
migration	<p><i>Migrationshintergrund d. Befragten, generiert</i></p> <p>Generierte Variable für eine vierstufige Unterscheidung des Migrationshintergrunds: Ohne Migrationshintergrund; eigene Migration (1. Generation); Migration mind. eines Elternteils, aber nicht selbst migriert (2. Generation); Migration mind. eines Großelternteils, aber nicht selbst migriert und kein Elternteil migriert (3. Generation).</p>	<p>PMI0100, PMI0700-PMI0900f (p191; p198-p200f)</p>

Variable	Label und Beschreibung	Quellvariablen
mschul1	<i>Höchster allg. Schulabschluss d. Mutter, incl. Mutter im HH, o. offene Ang., gen</i> Der Schulabschluss der Mutter wurde nur dann erhoben, wenn diese nicht im Haushalt wohnt. Wohnt sie im Haushalt, musste die Angabe zum Schulabschluss aus ihrem Personeninterview ergänzt werden.	PSH0200; schul1 d. Mutter (p204; schul1/Mutter)
mschul2	<i>Höchster allg. Schulabschluss d. Mutter, incl. Mutter im HH, incl. offene Ang., gen</i> Der Schulabschluss der Mutter wurde nur dann erhoben, wenn diese nicht im Haushalt wohnt. Wohnt sie im Haushalt, musste die Angabe zum Schulabschluss aus ihrem Personeninterview ergänzt werden.	PSH0200; schul2 d. Mutter (p204; schul2/Mutter)
netto	<i>Nettoeinkommen incl. kategorisierter Angaben, generiert</i> Generierung einer integrierten Variable aus kategorisierter und offener Angabe zum Nettoeinkommen.	PEK0700-PEK1200 p62-p67
nettokat	<i>Kategorisiertes Nettoeinkommen, generiert</i> Zusammenfassung der mehrstufig erhobenen kategorialen Netto-Einkommensangaben zu einer Variable.	PEK0800-PEK1200 p63-p67
ogebland	<i>Geburtsland, incl. offene Angaben, Kategorien (anonymisiert)</i> Aus Geburtsland = Deutschland und der offenen Angabe zum (anderen) Geburtsland zusammengefasste Variable.	PMI0100; PMI0200 (p191; p192)
ostaatan	<i>Staatsangehörigkeit, incl. offene Angaben, Kategorien (anonymisiert)</i> Aus Staatsangehörigkeit = Deutsch und der offenen Angabe zur (anderen) Staatsangehörigkeit zusammengefasste Variable.	PMI0400; PMI0500 (p195; p196)
palter	<i>Alter (aus p1), generiert</i> Aus dem Geburtsdatum generierte Altersvariable.	PD0100 (p1)
pintjahr	<i>Interviewdatum Personeninterview – Jahr, generiert</i> Jahr in dem das Interview stattfand.	pintdat
pintmon	<i>Interviewdatum Personeninterview – Monat, generiert</i> Monat in dem das Interview stattfand.	pintdat
pinttag	<i>Interviewdatum Personeninterview – Tag, generiert</i> Tag an dem das Interview stattfand.	pintdat
schul1	<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.	PB0300 - PB0500 (p5-p7)
schul2	<i>Höchster Schulabschluss, mit ausl. Abschl. und offenen Angaben</i> Entspricht schul1, bezieht zusätzlich die offenen Angaben zum Schulabschluss mit ein.	PB0300 - PB0500; PB1000; PB1100 (p5-p7; p13, p14)

Variable	Label und Beschreibung	Quellvariablen
stib	<i>Stellung im Beruf, Kennziffer, generiert</i> Generierung der detaillierten Kennziffer der Stellung im Beruf aus den Einzelvariablen.	PET1600-PET2200 (p39-p45)
vberuf1	<i>Hoehster ber. Abschluss des Vaters, incl. Vater im HH, ohne offene Ang., gen.</i> Der Berufsabschluss des Vaters wurde nur dann erhoben, wenn dieser nicht im Haushalt wohnt. Wohnt er im Haushalt, musste die Angabe zum Berufsabschluss aus seinem Personeninterview ergänzt werden.	PSH0600a-l; beruf1 d. Vaters (p208_01-p208_99; beruf1/ Vater)
vberuf2	<i>Hoehster ber. Abschluss des Vaters, incl. Vater im HH, mit offenen Ang., gen.</i> Der Berufsabschluss des Vaters wurde nur dann erhoben, wenn dieser nicht im Haushalt wohnt. Wohnt er im Haushalt, musste die Angabe zum Berufsabschluss aus seinem Personeninterview ergänzt werden.	PSH0600a-l; beruf2 d. Vaters (p208_01-p208_99; beruf2/ Vater)
vschul1	<i>Hoehster allg. Schulabschluss d. Vaters, incl. Vater im HH, o. offene Ang., gen</i> Der Schulabschluss des Vaters wurde nur dann erhoben, wenn dieser nicht im Haushalt wohnt. Wohnt er im Haushalt, musste die Angabe zum Schulabschluss aus seinem Personeninterview ergänzt werden.	PSH0500; schul1 d. Vaters (p207; schul1/ Vater)
vschul2	<i>Hoehster allg. Schulabschluss d. Vaters, incl. Vater im HH, incl. offene Ang., gen</i> Der Schulabschluss des Vaters wurde nur dann erhoben, wenn dieser nicht im Haushalt wohnt. Wohnt er im Haushalt, musste die Angabe zum Schulabschluss aus seinem Personeninterview ergänzt werden.	PSH0500; schul2 d. Vaters (p207; schul2/ Vater)
zmhh	<i>Zeiger: Personenummer der Mutter der Zielperson</i> Identifiziert die Personenummer der Mutter, sofern sie im Haushalt lebt. Als Mutter gelten: leibliche Mütter, Stiefmütter, Adoptivmütter oder Pflegemütter sowie Mütter mit nicht weiter spezifiziertem Status.	HD0200a-HD0200o; Haushaltsgrid (HH26, HH28-HH32; HH33-HH37)
zparthh	<i>Zeiger: Personenummer des Partners der Zielperson im HH in Welle 1 (2006/2007)</i> Identifiziert die Personenummer eines im Haushalt lebenden Partners. Als Partner gelten: Ehepartner, eingetragene Lebenspartner, unverheiratete Partner und Partner mit nicht weiter spezifiziertem Status.	HD0200a-HD0200o; Haushaltsgrid (HH26, HH28-HH32; HH33-HH37)
zvhh	<i>Zeiger: Personenummer des Vaters der Zielperson</i> Identifiziert die Personenummer des Vaters, sofern er im Haushalt lebt. Als Vater gelten: leibliche Väter, Stiefväter, Adoptivväter oder Pflegeväter sowie Väter mit nicht weiter spezifiziertem Status.	Haushaltsgrid (HH28-HH32; HH33-HH37)

**Tabelle 7: Einfache Umcodierungen oder Zusammenfassungen auf Haushaltsebene**

Variable	Label und Beschreibung	Quellvariablen
alg2abez	<i>Aktueller ALG2-Bezug des HH, generiert</i> Indikator für aktuellen Bezug von Arbeitslosengeld 2.	zensiert (ALG2_Spells) (hh43; hh44)
blneualt	<i>Alte oder Neue Bundeslaender, generiert</i> Zusammenfassung der Bundesländer zu alten Bundesländern (ohne Berlin) und neuen Bundesländern (mit Berlin).	bundesld
hhinckat	<i>Kategorisiertes Haushaltseinkommen, generiert</i> Zusammenfassung der mehrstufig erhobenen kategorialen Angaben zum Haushaltseinkommen zu einer Variable.	HEK0700-HEK1100 (hh61-hh65)
hhincome	<i>Haushaltseinkommen incl. kategorisierter Angaben, generiert</i> Generierung einer integrierten Variable aus kategorisierter und offener Angabe zum Haushaltsnettoeinkommen.	HEK0600-HEK1100 (hh60-hh65)
hintjahr	<i>Interviewjahr Haushaltsinterview – Jahr, generiert</i> Jahr in dem das Interview stattfand.	hintdat
hintmon	<i>Interviewmonat Haushaltsinterview – Monat, generiert</i> Monat in dem das Interview stattfand.	hintdat
hinttag	<i>Interviewtag Haushaltsinterview – Tag, generiert</i> Tag an dem das Interview stattfand.	hintdat

**Tabelle 8: Einfache Umcodierungen oder Zusammenfassungen in sonstigen Datensätzen**

Variable	Datensatz	Label und Beschreibung	Quellvariablen
bmonat	alg1_spells	Umgerechnete Anfangs- und Enddaten der Spelldatensätze mit umgerechneten Jahreszeitangaben bei den Monatswerten und konkreten Monats- bzw. Jahresangaben beim gegenwärtigen Bezug. Bei den Maßnahmespells Berechnung des Enddatums aus Anfangsdatum und Dauer.	AL10100 - AL10400
bjahr	alg2_spells		AL20100 - AL20400
emonat	Massnahme-spells		ALM0200, ALM0300, ALM0400, ALM0500, ALM0800
ejahr	Massnahme-spells		
zensiert	alg1_spells alg2_spells Massnahme-spells	Die Variable 'zensiert' gibt an, ob ein Spell zum Erhebungszeitpunkt der letzten Befragungswelle noch andauert, d.h. der Spell ist rechtszensiert. In den Datensätzen zum Bezug des Arbeitslosengeld I und II stellt die Variable eine Kombination aus der bei der Erhebung des Bezugszeitraums gemachten Angabe, dass gegenwärtig noch Leistungen bezogen werden und der im Fall eines Spellendes im Befragungsmonat gestellten Nachfrage, ob ggw. noch Leistungen bezogen werden. Bei den Maßnahmespells handelt es sich lediglich um eine Umcodierung der Frage nach der ggw. Programmteilnahme, wobei fehlende Werte auf dieser Variablen als Nicht-Teilnahme gewertet wurden.	AL10300-AL10500 AL20300-AL20500 ALM0400

### c) 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, incl. eines kurzen Überblicks über den theoretischen Hintergrund des Konstrukts und der wichtigsten Literaturverweise.

#### **Personenebene**

##### Bildung in Jahren

##### Variablenname

bilzeit

##### Variablenlabel

Schul- und Berufsbildung in Jahren, generiert

##### Ausgangsvariablen

schul2 beruf2

##### Typ / Datensatz

Bildung / Personendaten

##### Bearbeitung

Bernhard Christoph

##### Erläuterung

Für viele Analyseverfahren ist die Verwendung einer linearen Bildungsvariablen 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 dreizehn Jahren erreicht wurde. Hier wurden die Schulabschlüsse wie folgt veranschlagt:

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

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).

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.

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

##### Literatur

Helberger (1988)

Bildung in Jahren, Mutter

<u>Variablenname</u>	mbilzeit																
<u>Variablenlabel</u>	Schul- und Berufsbildung in Jahren, generiert																
<u>Ausgangsvariablen</u>	mschul2 mberuf2																
<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 wurden (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>	vbilzeit																
<u>Variablenlabel</u>	Schul- und Berufsbildung in Jahren, generiert																
<u>Typ / Datensatz</u>	Bildung / Personendaten																
<u>Typ</u>	Bildung																
<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 wurden (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)																

CASMINVariablenname

casmin

Variablenlabel

Bildung klassifiziert nach CASMIN, aktual. Version, generiert

Ausgangsvariablen

schul2 beruf2

Typ / Datensatz

Bildung / Personendaten

Bearbeitung

Bernhard Christoph

Erläuterung

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.

schulberuf	nicht erhob.	Schüler	n. gest.	TNZ	KA	WN	ohne Abschl.	Sonderschule	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
Anlernausbild.	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Lehre	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Berufsfachsch.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Schul. d. Ges.wes.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Meister	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
BA	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
FH/Bachelor	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
Unif/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

mcasmin

Variablenlabel

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

Ausgangsvariablen

mschul2 mberuf2

Typ / Datensatz

Bildung / Personendaten

Bearbeitung

Bernhard Christoph

Erläuterung

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	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
Plint 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	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)

VCASMINVariablenname

vcasmin

Variablenlabel

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

Ausgangsvariablen

vschul2 vberuf2

Typ / Datensatz

Bildung / Personendaten

Bearbeitung

Bernhard Christoph

Erläuterung

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

iscsed

Variablenlabel

Bildung klassifiziert nach isced97, aktual. Version, generiert

Ausgangsvariablen

mschul2 mberuf2

Typ / Datensatz

Bildung / Personendaten

Bearbeitung

Bernhard Christoph

Erläuterung

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 bis 6.

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

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

Literatur

BMBF (2003); OECD (1999)

MISCED 97Variablenname

misced

Variablenlabel

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

AusgangsvARIABLEN

mschul2 mberuf2

Typ / Datensatz

Bildung / Personendaten

Bearbeitung

Bernhard Christoph

Erläuterung

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 bis 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
Aulernausbild.	-	-	-	-	-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

visced

Variablenlabel

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

Ausgangsvariablen

vschul2 vberuf2

Typ / Datensatz

Bildung / Personendaten

Bearbeitung

Bernhard Christoph

Erläuterung

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 bis 5. Zu Details der Vercodung vgl. die folgende Abbildung.

Schulberuf	nicht erhob.	Pfint 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
Pfint 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
Aulern-ausbild.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
Lehre	-	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Meister	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
BA	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
FH	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Uni	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
And. dt. Abschl.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
And. aus Abschl.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2

Literatur

BMBF (2003); OECD (1999)

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

<u>Variablenname</u>	isco88
<u>Variablenlabel</u>	ISCO 88 (ZUMA-Vercodung), generiert
<u>Ausgangsvariablen</u>	P46
<u>Typ / Datensatz</u>	Beruf / Personendaten
<u>Ansprechpartner</u>	Bernhard Christoph
<u>Erläuterung</u>	<p>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 Bundesamtes, 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.</p> <p>Die Vercodung erfolgte durch das Zentrum für Umfragen, Methoden und Analysen (Gesis-ZUMA). Die ZUMA-Vercodung von ISCO-88 stellt im Gegensatz zur Infratest-Variante eine originäre Vercodung von ISCO-88 dar. Sie bildet die Grundlage für die Generierung der ISCO-basierten Status- und Prestigemaße.</p> <p>Eine detaillierte Tabelle der ISCO 88 sowie der zugehörigen Status- und Prestigewerte findet sich in Tabelle A 3 auf S. 623 ff.</p>
<u>Literatur</u>	ILO (1990)

International Standard Classification of Occupations 1988 (ISCO88); Infratest Ver-  
codung

<u>Variablenname</u>	isco88it
<u>Variablenlabel</u>	ISCO 88 (Infratest-Vercodung), generiert
<u>Ausgangsvariablen</u>	P46
<u>Typ / Datensatz</u>	Beruf / Personendaten
<u>Ansprechpartner</u>	Bernhard Christoph
<u>Erläuterung</u>	<p>Die ISCO ist eine von der ILO entwickelte, international vergleichbare Berufsklassifikation. Besonderheit der ISCO-88 ist, insbesondere im Vergleich zur hier ebenfalls zur Verfügung gestellten Klassifizierung der Berufe des statistischen Bundesamtes, 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.</p> <p>Die Vercodung erfolgte durch Infratest, das Feldinstitut des PASS, als Umschlüsselung aus der Klassifizierung der Berufe des Statistischen Bundesamtes.</p> <p>Eine detaillierte Tabelle der ISCO-88 sowie der zugehörigen Status- und Prestigewerte findet sich in Tabelle A 3 auf S. 623 ff.</p>
<u>Literatur</u>	ILO (1990)

Klassifizierung der Berufe 1992 (KldB92); Infratest Vercodung

<u>Variablenname</u>	kldb_it
<u>Variablenlabel</u>	Klassifizierung der Berufe 1992 (Infratest-Vercodung), generiert
<u>Ausgangsvariablen</u>	P46
<u>Typ / Datensatz</u>	Beruf / Personendaten
<u>Ansprechpartner</u>	Bernhard Christoph
<u>Erläuterung</u>	Die KldB92 ist die aktuelle Version der Klassifizierung der Berufe in der Variante des Statistischen Bundesamtes. Sie ist eine spezifisch auf Deutschland bezogene Berufsklassifikation. Die Klassifikation basiert ausschließlich auf der beruflichen Tätigkeit. Die Vercodung erfolgte durch Infratest, das Feldinstitut des PASS.
<u>Literatur</u>	StBA (1992).

Berufsvercodung, Problemindikator (Infratest Vercodung)

<u>Variablenname</u>	berpr_it
<u>Variablenlabel</u>	Berufsvercodung, Problemindikator (Infratest-Vercodung), generiert
<u>Ausgangsvariablen</u>	P46
<u>Typ / Datensatz</u>	Beruf / Personendaten
<u>Bearbeitung</u>	-
<u>Erläuterung</u>	Von Infratest erstellter Indikator für die Beurteilung der Vercodung der Berufe nach KldB 1992.
<u>Literatur</u>	

---

Klassenschema nach Erikson, Goldthorpe und Portocarrero (EGP)

<u>Variablenname</u>	egp
<u>Variablenlabel</u>	Klassenschema nach Erikson, Goldthorpe & Portocarero (EGP), akt. Beruf, gen.
<u>Ausgangsvariablen</u>	isco88, stib
<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 &amp; 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 schienen (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>Variablenname</u>	esec
<u>Variablenlabel</u>	European Socio-economic Classification (ESeC) akt. Beruf, gen.
<u>Ausgangsvariablen</u>	isco88, stib, PET2000, PET2700
<u>Typ / Datensatz</u>	Sozioökonomische Position / Personendaten
<u>Bearbeitung</u>	Bernhard Christoph
<u>Erläuterung</u>	<p>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 &amp; Harrison 2007; für Deutschland Müller et al. 2006, 2007).</p> <p>Der Stata do-file für die Generierung von ESeC wurde uns dankenswerterweise von Heike Wirth von Gesis-ZUMA zur Verfügung gestellt (Fischer &amp; Wirth 2007) und von uns lediglich für die vorliegende Studie angepasst. Dieser do-file stellt eine Stata-Umsetzung der von Harrison &amp; Rose (2006) geschriebenen Standard-SPSS-Syntax zur Generierung von ESeC dar.</p>
<u>Literatur</u>	Fischer & Wirth (2007); Harrison & Rose (2006); Müller et al. (2006, 2007); Rose & Harrison (2007)



Magnitude-Prestigeskala (MPS)

<u>Variablenname</u>	mps
<u>Variablenlabel</u>	Magnitude-Prestigeskala, aktueller Beruf, gen.
<u>Ausgangsvariablen</u>	isco88
<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 KldB 1992 des Statistischen Bundesamtes 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>Variablenname</u>	siops
<u>Variablenlabel</u>	Standard International Occupational Prestige Scale, aktueller Beruf, gen.
<u>Ausgangsvariablen</u>	isco88
<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>Variablenname</u>	isei
<u>Variablenlabel</u>	International Socio-Economic Index, aktueller Beruf, gen.
<u>Ausgangsvariablen</u>	isco88
<u>Typ / Datensatz</u>	Sozioökonomische Position / Personendaten
<u>Ansprechpartner</u>	Bernhard Christoph
<u>Erläuterung</u>	<p>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.</p> <p>Der ISEI wurde zunächst für ISCO-68 entwickelt (Ganzeboom et al. 1992), später jedoch auch für ISCO-88 umgesetzt (Ganzeboom &amp; Treiman 1996, 2003).</p> <p>Die Zuspiegelung erfolgte im Rahmen der Berufsvercodung durch das Zentrum für Umfragen, Methoden und Analysen (Gesis-ZUMA).</p>
<u>Literatur</u>	Ganzeboom et al. (1992); Ganzeboom & Treiman (1996, 2003)

---

## Haushalts- oder Bedarfsgemeinschaftsebene

### Bedarfsgewichtetes Haushaltseinkommen, alte OECD-Gewichtung

<u>Variablenname</u>	oecdinca
<u>Variablenlabel</u>	Bedarfsgewichtetes Haushaltseinkommen, alte OECD-Gewichtung (gerundet)
<u>Ausgangsvariablen</u>	HD0200a-HD0200o; HA0100; hhincome (hh27_01-hh27_15; hh1; hhincome)
<u>Typ / Datensatz</u>	Sozioökonomische Position / Haushaltsdaten
<u>Bearbeitung</u>	Bernhard Christoph
<u>Erläuterung</u>	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). Zur alten OECD-Skala vgl. OECD (1982) für einen Überblick zum Thema vgl. Hauser (1996). 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.
<u>Literatur</u>	Hauser (1996); OECD (1982)

### Bedarfsgewichtetes Haushaltseinkommen, neue OECD-Gewichtung

<u>Variablenname</u>	oecdincn
<u>Variablenlabel</u>	Bedarfsgewichtetes Haushaltseinkommen, neue OECD-Gewichtung (gerundet)
<u>Ausgangsvariablen</u>	HD0200a-HD0200o; HA0100; hhincome (hh27_01-hh27_15; hh1; hhincome)
<u>Typ / Datensatz</u>	Sozioökonomische Position / Haushaltsdaten
<u>Bearbeitung</u>	Bernhard Christoph
<u>Erläuterung</u>	<u>Generelle Darstellung:</u> vgl. Bedarfsgewichtetes Haushaltseinkommen, alte OECD-Gewichtung. Zur neuen OECD-Skala vgl. Hagenaars et al. (1994). 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.
<u>Literatur</u>	Hagenaars et al. (1994)

Deprivationsindex, ungewichtet

<u>Variablenname</u>	depindug
<u>Variablenlabel</u>	Deprivationsindex, ungew. (n. aus fin. Gründ. fehl. Items; ungew. Itemsumme: 26)
<u>Ausgangsvariablen</u>	HLS0100a-HLS2600a; HLS0100b-HLS2600b (HH7a_01-HH7b_13; HH8a_01-HH8b_13)
<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 Armutspopulation im engeren Sinne zu betrachten (vgl. Halleröd 1995; Nolan &amp; 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 ('listwise 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>	depindg
<u>Variablenlabel</u>	Deprivationsindex, gew. (n. aus fin. Gründ. fehl. Items; gew. Itemssumme: 12,8)
<u>Ausgangsvariablen</u>	HLS0100a-HLS2600a; HLS0100b-HLS2600b; PLS0100-PLS2600 (HH7a_01-HH7b_13; HH8a_01-HH8b_13; P141a-P141z)
<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 Index auf die von mind. einem bestimmten Anteil der Befragten als notwendig erachteten Items (z.B. Andreß &amp; Lipsmeier 1995, Andreß et al. 1996) oder eine theoretische Beschränkung auf wenige, als grundlegend angesehene Items (z.B. Nolan &amp; 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 &amp; Lipsmeier (2001, insbes. S. 28ff.)</p>
<u>Literatur</u>	Andreß & Lipsmeier (1995, 2001); Andreß et al. (1996); Halleröd (1995); Lipsmeier (1999); Nolan & Whelan (1996)

HaushaltstypologieVariablenname

hhtyp

Variablenlabel

Haushaltstyp, generiert

Ausgangsvariablen

HH27-HH37

Typ / Datensatz

Haushaltsstruktur / Haushaltsdaten

Bearbeitung

Daniel Gebhardt

Erläuterung

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

**Definition der Beziehungen bei der Generierung des Haushaltstyps:**

• **Paare:** Ehepaare; eingetragene Lebenspartnerschaften; unverheiratete Partnerschaften sowie Partnerschaften mit nicht weiter spezifiziertem Status (fehlender Wert in Nachfrage nach Art der Partnerschaft).

• **Kind einer Person:** Leibliches Kind; Stief-, Adoptiv- oder Pflegekind; Kind mit nicht weiter spezifiziertem Status (fehlender Wert in Nachfrage nach Art der Beziehung zum Kind).

• **Elternteil einer Person:** Leibliches Elternteil; Stief-, Adoptiv- und Pflegeelternteil; Eltern mit nicht weiter spezifiziertem Status (fehlender Wert in Nachfrage nach Art der Elternschaft).

**Definition der Haushaltstypen:**

• **1-Personen Haushalt:** Haushalt mit nur einer Person im Haushalt.

• **Paar ohne Kinder:** Haushalt besteht ausschließlich aus zwei in Partnerschaft lebenden Personen.

• **1-Eltern-Haushalt:** Haushalt besteht ausschließlich aus einem Elternteil und dessen Kindern. Hinsichtlich des Alters der Kinder werden keine Einschränkungen vorgenommen.

• **Paar mit Kindern < 16 Jahren:** Haushalt besteht ausschließlich aus zwei in Partnerschaft lebenden Personen und deren jeweiligen bzw. gemeinsamen Kindern. Alle Kinder sind unter 16 Jahren alt.

• **Paar mit Kindern >= 16 Jahren:** 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 Jahren und >= 16 Jahren:** Haushalt besteht ausschließlich aus zwei in Partnerschaft lebenden Personen und deren jeweiligen bzw. gemeinsamen Kindern. Es leben sowohl Kinder unter 16 Jahren als auch im Alter von 16 Jahren oder älter im Haushalt.

• **Mehrgenerationenhaushalt:** Haushalt besteht aus Angehörigen mindestens dreier linear aufeinanderfolgender Generationen. Der Haushalt besitzt einen Mehrgenerationenkern, d.h. mindestens eine Person im Haushalt ist sowohl Kind als auch Elternteil einer anderen Person im Haushalt. Ausgehend von den zu diesem Mehrgenerationenkern gehörenden Personen leben im Haushalt ausschließlich Eltern, Kinder, Geschwister, Partner sowie Geschwister der Partner.

• **Anderer Haushaltstyp:** Haushalt, der keinem der anderen definierten Haushaltstypen zugeordnet werden konnte.

• **Generierung nicht möglich (fehlende Werte):** Grundsätzlich alle Haushalte mit mindestens einem fehlenden (-1,-2,-4) oder unplausiblen Wert (-8) in der Oberkategorie in einer Beziehungsvariable oder bei der Altersangabe. (Ausnahme: In Haushalten mit drei und weniger Mitgliedern wurde in eindeutigen Konstellationen auch bei fehlenden Altersangaben der Haushaltstyp gebildet.)

Literatur

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

Bedarfsgemeinschaftsnummer, Welle 1

<u>Variablenname</u>	bgnr1
<u>Variablenlabel</u>	Nummer der Bedarfsgemeinschaft in Welle 1
<u>Ausgangsvariablen</u>	HH27-HH37
<u>Typ / Datensatz</u>	Bedarfsgemeinschaft / Personenregister
<u>Bearbeitung</u>	Gerrit Müller
<u>Erläuterung</u>	<p>Die Variable bgnr1 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 bgnr1 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 zum tatsächlichen 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 selber 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 Jahren 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>	SGB II

Bedarfsgemeinschaftstypologie, Welle 1

<u>Variablenname</u>	bgtyp1
<u>Variablenlabel</u>	Typ der Bedarfsgemeinschaft in Welle 1
<u>Ausgangsvariablen</u>	HH27-HH37
<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 „bgnr1“ 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 Jahren 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 SGBII fallen (vgl. Code 0 bei bgnr1). Code -5, Generierung nicht möglich (fehlende Werte), wurde für Haushalte mit fehlenden Beziehungs- bzw. Altersangaben zu einzelnen Personen (vgl. Code 99 bei bgnr1) vergeben.</p>
<u>Literatur</u>	–

ALG2-Bezug der Bedarfsgemeinschaft zum Zeitpunkt der Stichprobenziehung, Welle 1

<u>Variablenname</u>	bgbezs1
<u>Variablenlabel</u>	ALG2-Bezug der Bedarfsgemeinschaft bei Stichprobenziehung in Welle 1 (2006/2007)
<u>Ausgangsvariablen</u>	HH43, HH45, HH46, HH54, sample, bgnr1
<u>Typ / Datensatz</u>	Bedarfsgemeinschaft / Personenregister
<u>Bearbeitung</u>	Mark Trappmann
<u>Erläuterung</u>	<p>Diese Variable gibt für jede Bedarfsgemeinschaft, die entsprechend dem bei Variable bgnr1 beschriebenen Vorgehen gebildet wurde, an, ob diese Bedarfsgemeinschaft zum Zeitpunkt der Stichprobenziehung für Welle 1 Arbeitslosengeld II bezogen hat, oder nicht.</p>
<u>Literatur</u>	–



ALG2-Bezug der Bedarfsgemeinschaft zum Befragungszeitpunkt, Welle 1

<u>Variablenname</u>	bgbezb1
<u>Variablenlabel</u>	ALG2-Bezug der Bedarfsgemeinschaft zum Befragungszeitpunkt in Welle 1 (2006/2007)
<u>Ausgangsvariablen</u>	HH43-HH46, sample, bgnr1
<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 bgnr1 beschriebenen Vorgehen gebildet wurde, an, ob diese Bedarfsgemeinschaft zum Befragungszeitpunkt in Welle 1 Arbeitslosengeld II bezogen hat, oder nicht.
<u>Literatur</u>	–

Anzahl der Bedarfsgemeinschaften im Haushalt

<u>Variablenname</u>	anzbg
<u>Variablenlabel</u>	Anzahl der synthetischen Bedarfsgemeinschaften im HH, generiert
<u>Ausgangsvariablen</u>	bgnr1, hnr
<u>Typ / Datensatz</u>	Bedarfsgemeinschaft / Haushaltsdatensatz
<u>Bearbeitung</u>	Mark Trappmann
<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 bgnr1 beschriebenen Vorgehensweise.
<u>Literatur</u>	–

Anzahl der zum Stichprobenzeitpunkt Leistungen beziehenden Bedarfsgemeinschaften im Haushalt

<u>Variablenname</u>	nbgbezug
<u>Variablenlabel</u>	Anzahl zum Stichprobenzeitpunkt beziehender BGs im HH
<u>Ausgangsvariablen</u>	bgbezs1, bgnr1, hnr
<u>Typ / Datensatz</u>	Bedarfsgemeinschaft / Haushaltsdatensatz
<u>Bearbeitung</u>	Mark Trappmann
<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 bgbezs1 aus dem Personenregister beziehenden Bedarfsgemeinschaften über die Haushaltsnummer gebildet.
<u>Literatur</u>	–

## 7. Gewichtung und Verwendung der Gewichte

### a) Dokumentation der Gewichtung

Die Gewichtung beruht auf einem dreistufigen Gewichtungskonzept:

- In der ersten Stufe erfolgt die Designgewichtung der Bruttoeinsatzstichprobe.
- Daran anschließend wird in der zweiten Stufe die Modellierung von Ausfällen durchgeführt.
- Zuletzt schließt sich in der dritten Stufe eine Kalibrierung der Gewichte an.

#### *1. Stufe: Designgewichtung:*

Die Designgewichte sind reziproke Auswahlwahrscheinlichkeiten für die Bruttoeinsatzstichprobe. Das Verfahren zu ihrer Generierung ist ausführlich in Rudolph und Trappmann (2007) beschrieben. Die Designgewichte befinden sich im Datensatz *hweights*. Die bereitgestellten Designgewichte sind im Einzelnen:

dw_ba_1	Designgewicht eines Haushalts in der BA-Stichprobe (Grundgesamtheit: Haushalte, in denen es im Juli 2006 mindestens eine Bedarfsgemeinschaft mit Bezug von Leistungen nach dem SGB II gab)
dw_mi_1	Designgewicht eines Haushalts in der Microm-Stichprobe (Grundgesamtheit: Haushalte in der Bundesrepublik Deutschland)
dw_1	Designgewicht eines Haushalts in der Gesamtstichprobe (Grundgesamtheit: Haushalte in der Bundesrepublik Deutschland)

#### *2. Stufe: Modellierung von Ausfällen*

Mit Hilfe zweier Logit-Modelle wurde für alle Haushalte der Bruttoeinsatzstichprobe die Teilnahmewahrscheinlichkeit geschätzt. Das erste Logit-Modell erklärt die Wahrscheinlichkeit eines Kontakts. Das zweite Logit-Modell erklärt die Teilnahme (mindestens Haushaltsinterview und ein Personeninterview vollständig) im Falle eines erfolgreichen Kontakts. Diese Logit-Modelle wurden für die beiden Teilstichproben getrennt berechnet, da der Kontaktierung jeweils unterschiedliche Prozesse zugrunde liegen. Bei den Modellen für die Microm-Stichprobe wurden jeweils nur die von Microm gelieferten micro-geographischen Variablen verwendet. Bei den Modellen für die BA-Stichprobe konnten zusätzlich Merkmale aus der Ziehungsgrundlage (A2LL bzw. XSozial) verwen-

det werden. Die verwendeten Modelle enthalten lediglich Variablen mit signifikanten Effekten (LR-Test, zweiseitig, 10%-Niveau). Die verwendeten Variablen und die Koeffizienten der Modelle sind in den nachfolgenden Tabellen ausgewiesen. Eine detaillierte Darstellung der Modellierung der Ausfälle findet sich auch in Kapitel 5.2 des Methoden- und Feldberichts von TNS Infratest (Hartmann et al. 2008).

Der Datensatz *hweights* enthält die Variablen *pr\_ba\_1* und *pr\_mi\_1*. Dies sind – separat für BA- und Microm-Stichprobe – die Produkte der vorhergesagten Wahrscheinlichkeiten beider Modelle.

*pr\_ba\_1* geschätzte Teilnahmewahrscheinlichkeit für Fälle aus BA-Stichprobe

*pr\_mi\_1* geschätzte Teilnahmewahrscheinlichkeit für Fälle aus Microm-Stichprobe

Dividiert man die Designgewichte durch die geschätzten Teilnahmewahrscheinlichkeiten, so erhält man die modifizierten Designgewichte, die den Ausgangspunkt für den dritten Schritt – die Kalibrierung – darstellten.

### 3. Stufe: Kalibrierung

Eine ausführliche Dokumentation der Kalibrierung ist in Kiesel (2008) zu finden. Daher wird an dieser Stelle nur das grundlegende Verfahren skizziert.

#### HAUSHALTSEBENE

In einem ersten Schritt wurden beide Teilstichproben und die Gesamtstichprobe auf Haushaltsebene an Eckwerten der amtlichen Statistik kalibriert.

Die Gesamt- und BA-Gewichte für Leistungsbezieher beider Teilstichproben wurden an Eckwerte der Statistik der Bundesagentur für Arbeit (Berichtsmonat Juli 2006) angepasst. Die Gesamt- und Micromgewichte wurden zusätzlich an Eckwerte des Statistischen Bundesamts zu Privathaushalten in Deutschland für 2007 angepasst: Die verwendeten Eckwerte sind Kiesel (2008) zu entnehmen.

Alle Gewichte sind Haushaltsgewichte. Die BA-Statistik weist jedoch Werte auf Bedarfsgemeinschaftsebene aus. Die Verbindung wird über die synthetischen Bedarfsgemeinschaften (BGs) hergestellt, deren Generierung in Kapitel 7 dokumentiert ist (Variable *bgnr1* im Datensatz *p\_register*). Haushalte werden zunächst in syntheti-

sche Bedarfsgemeinschaften zerlegt. Die Merkmale, die für die Kalibrierung verwendet werden, werden sodann auf BG-Ebene generiert. Dazu gehört auch das Merkmal, ob die BG zum Stichprobenziehungszeitpunkt ALG II-Leistungen erhalten hat. Nach der Kalibrierung ergeben die Merkmale aller BGs, die zum Stichprobenziehungszeitpunkt Leistungen bezogen haben, multipliziert mit den Hochrechnungsfaktoren für Haushalte die Eckwerte. Beziehende BGs innerhalb derselben Haushalte erhalten folglich immer gleiche Hochrechnungsfaktoren.

Der Leistungsbezug eines Haushalts oder gar einer BG lässt sich nicht in allen Fällen sauber bestimmen. Eine möglichst breite Datenlieferung soll dem Nutzer die Möglichkeit geben, Entscheidungen eigenständig zu treffen. So wird auf Haushaltsebene die Variable *alg2samp* im Datensatz *hh\_register* ausgeliefert. Sie enthält den Leistungsbezug zum Stichprobenziehungszeitpunkt für alle Haushalte in den Kategorien: *0 kein Bezug*, *1 Bezug*, *2 laut Survey kein Bezug (aber BA-Stichprobe und damit Bezug laut Prozessdaten)*, *3 laut Survey Bezug unklar (aber BA-Stichprobe und damit Bezug laut Prozessdaten)*, *4 laut Survey Bezug unklar (Microm-SP)*. Darüber hinaus kann jeder Nutzer mit Hilfe der Arbeitslosengeldspells (Datensatz *alg2\_spells*) diese Variable selbst bilden. Zusätzlich hilfreich sind dabei die Variablen *AL20600* und *AL20700a-o* (für wen bezieht der Haushalt Leistungen?) und die Variable *HA0400* aus *HHENDDAT*, in der für Haushalte, die erst nach Juli 2006 gegründet wurden, erhoben wird, ob mindestens eines ihrer Mitglieder im Juli 2006 Leistungen bezogen hat. Die Variable *sample* in *hweights* gibt für jeden Haushalt an, aus welcher Stichprobe er stammt.

Für die Durchführung der Gewichtung ist jedoch eine eindeutige Entscheidung erforderlich, welche Bedarfsgemeinschaften man als zum Stichprobenziehungszeitpunkt Arbeitslosengeld II beziehende Bedarfsgemeinschaften betrachtet. Die folgenden Ausführungen legen offen, welche Entscheidungen der Gewichtung zugrunde liegen.

Auf der Ebene der Haushalte wurde entschieden:

1. Alle Haushalte aus der BA-Stichprobe (*sample=1*) haben zum Stichprobenziehungszeitpunkt Leistungen bezogen, auch wenn sie dem widersprechen, insofern mindestens eine Person zwischen 15 und 64 in diesem Haushalt lebt.
2. Haushalte aus der Microm-Stichprobe, für die der Bezug nicht eindeutig aus den Surveydaten ableitbar ist, werden für die Gewichtung als Arbeitslosengeld-II-Empfängerhaushalte betrachtet, wenn sie angeben jemals ALG II bezogen zu ha-

ben ( $HA0300=1$ ) und wenn das Anfangs- oder Enddatum mindestens eines Spells in 2006 liegt (bei unbestimmtem Ende bzw. Anfang).

Die Übertragung von der Haushalts- auf die BG-Ebene ist mit noch größerer Unsicherheit behaftet. Die retrospektive Angabe, welche Teile des Haushalts im Juli 2006 Leistungen bezogen haben, ist nicht zuverlässig erfragbar. In den meisten Fällen besteht der gesamte Haushalt aus nur einer BG. Hier erübrigt sich die Frage ohnehin. Die BG bezieht hier genau dann, wenn der Haushalt bezieht. In Fällen, in denen der Haushalt aus mehreren BGs besteht, wurde die folgende Näherung gewählt:

Es wurde die Information genutzt, für wen der Haushalt aktuell Leistungen bezieht (AL20600 und AL20700a-o). Eine BG wurde als beziehend betrachtet, wenn mindestens eine der zu dieser gehörenden Personen als Bezieher berichtet wurde. In einem Haushalt mit mehreren BGs und ohne Information darüber, für wen der Haushalt Leistungen bezieht (z.B. weil laut Befragung kein aktueller Bezug vorliegt), wurden alle synthetischen BGs als beziehende BGs betrachtet. Das Ergebnis dieser Generierung ist in der Variable *bgbez1* im Datensatz *p\_register* nachzuvollziehen.

Die Gewichte nach abgeschlossener Kalibrierung auf Haushaltsebene befinden sich ebenfalls im Datensatz *hweights*.

wqbahh1	kalibriertes Haushaltsgewicht der BA-Stichprobe
wqmihh1	kalibriertes Haushaltsgewicht der Microm-Stichprobe
wqhh1	kalibriertes Haushaltsgewicht der Gesamt-Stichprobe

## PERSONENEBENE

Im Anschluss an die Kalibrierung auf Haushaltsebene werden die Personen, die ein Personen- oder Senioreninterview gegeben haben, an Personeneckwerte angepasst. Ausgangspunkt für diesen Schritt sind die kalibrierten Haushaltsgewichte.

Die Gesamt- und BA-Gewichte für Leistungsbezieher beider Teilstichproben wurden an Eckwerte der Statistik der Bundesagentur für Arbeit (Berichtsmonat Juli 2006) angepasst. Die Gesamt- und Micromgewichte wurden zusätzlich an Eckwerte des Statistischen Bundesamts für Privathaushalte in Deutschland für 2007 angepasst. Die verwendeten Eckwerte sind wiederum Kiesel (2008) zu entnehmen.

Senioreninterviews wurden bei der Kalibrierung an Bevölkerungseckwerte ebenso behandelt wie Personeninterviews. Die BA-Statistik weist jedoch keine Eckwerte zur

---

Anzahl der Senioren in Haushalten von Leistungsempfängern aus. Ebenso sind in Haushalten von Leistungsempfängern lebende Personen, die nicht zu einer Bedarfsgemeinschaft gehören, dort nicht ausgewiesen. Die BA-Personengewichte für diese Personen konnten daher nicht durch Kalibrierung gewonnen werden. Die Teilnahme-wahrscheinlichkeit dieser Personen, gegeben ihr Haushalt nimmt an der Befragung teil, wurde durch ein Logit-Modell mit den Kovariaten Anzahl der Personen ab 15 Jahren im Haushalt, Interviewmodus, Alter und Geschlecht geschätzt. Anschließend wurde das modifizierte Designgewicht durch diesen Wert dividiert.

Die kalibrierten Personengewichte befinden sich im Datensatz *pweights*.

wqbpap1	kalibriertes Personengewicht der BA-Stichprobe
wqmip1	kalibriertes Personengewicht der Microm-Stichprobe
wqgesp1	kalibriertes Personengewicht der Gesamt-Stichprobe

#### b) Empfehlungen zum Surveyset in Stata

Bei allen hier dargestellten Gewichten handelt es sich um sogenannte *probability weights*: Das Gewicht eines Haushalts bzw. einer Person entspricht dem Kehrwert seiner bzw. ihrer (durch Ausfallmodellierung und Kalibrierung korrigierten) Inklusionswahrscheinlichkeit. In Stata sind *probability weights* ab Version 9 im Rahmen des *Surveyset* zu setzen. Das *Surveyset* hat aber nicht nur den Zweck, die zu verwendenden Gewichte zu definieren, sondern auch diejenigen Aspekte des Surveydesigns zu definieren, die Einfluss auf die Standardfehler haben.

Dazu gibt es in Stata zwei grundsätzlich verschiedene Möglichkeiten: Die Spezifikation des Designs oder die Verwendung von Replikationsgewichten. Bei der ersten Möglichkeit werden diejenigen Aspekte des Surveydesigns angegeben, die Einfluss auf den Standardfehler haben: Das sind – neben den Gewichten – Cluster, Stratifizierungsmerkmale und Endlichkeitskorrekturen bei der Auswahl ohne Zurücklegen. Nicht berücksichtigt werden können der Effekt der Kalibrierung auf den Standardfehler und Besonderheiten wie pps-Sampling. Bei der zweiten Möglichkeit wird dagegen auf einen Satz von Replikationsgewichten zugegriffen, die mit Verfahren wie Jackknife, BRR oder Bootstrap für alle Untersuchungseinheiten errechnet werden. Dieses Verfahren ermöglicht potentiell auch die Berücksichtigung der Kalibrierung.

Derzeit stehen noch keine Replikationsgewichte für PASS zur Verfügung. Daher ist beim *Surveyset* für PASS die erste Variante zu verwenden. Das komplexe Stichprobendesign des PASS lässt sich jedoch nicht in allen Details mit dem *Surveyset*-Befehl für die Varianzschätzung nutzbar machen. Wir empfehlen die folgende Näherung:

```
svyset psu [pw=wqX], strata(strpsu)
```

Dabei steht *wqX* für das für die intendierten Analysen adäquate Gewicht. Ein Indikator für die (in beiden Teilstichproben gleichen) *primary sampling units* ist die Variable *psu* im Haushaltsdatensatz HHENDDAT.dta. Die Schichten bei der Auswahl der *primary sampling units* sind durch die Variable *strpsu* im selben Datensatz repräsentiert. Schichten mit weniger als zwei Einheiten in der Stichprobe wurden zusammengefasst. Beim Ziehen mit Zurücklegen spielen Schichten und Cluster ab der 2. Ebene für die Varianzschätzung keine Rolle (Särndal et al. 1992, 144ff.). Ist die Sampling-Rate sehr gering, so kann die Varianzschätzung beim Ziehen ohne Zurücklegen durch die Formeln für das Ziehen mit Zurücklegen sehr gut angenähert werden. Dies ist hier der Fall (Nur etwa 3,6% der Postleitzahlen in Deutschland wurden für die Befragung ausgewählt). Auf die Angabe von Endlichkeitskorrekturen und weiteren Clustern (hier: Haushalten) kann daher verzichtet werden. Das empfohlene *Surveyset* berücksichtigt dann allerdings weder Kalibrierung noch pps-Sampling, noch die ohnehin geringe Endlichkeitskorrektur für das Ziehen ohne Zurücklegen. Die resultierenden Standardfehler fallen zu groß und damit konservativ aus.

### c) Verwendung der Gewichte

Alle Gewichte sind Hochrechnungsfaktoren. Teilt man diese Gewichte durch Ihren Mittelwert, so erhält man Gewichte, die sich zur Stichprobengröße summieren. Die Designgewichte (*dw\_mi\_1*, *dw\_ba\_1*, *dw\_1*) und die geschätzten Teilnahmewahrscheinlichkeiten (*pr\_ba\_1*, *pr\_mi\_1*) werden mit ausgeliefert, empfohlen wird jedoch die Verwendung der kalibrierten Gewichte. Wer dennoch auf die Kalibrierung verzichten möchte, sei darauf hingewiesen, dass man zwar durch Division der Haushaltsgewichte durch die für die jeweilige Teilstichprobe adäquaten geschätzten Teilnahmewahrscheinlichkeiten modifizierte Haushaltsdesigngewichte gewinnen kann, dass diese aber nicht berücksichtigen, dass es auch innerhalb teilnehmender Haushalte zu Nonrespon-

se gekommen ist und die Anwendung auf Personen daher zunächst die Schätzung einer Teilnahmewahrscheinlichkeit der Person gegeben die Teilnahme ihres Haushalts erfordert.

Die folgenden Abschnitte zeigen an Beispielen, wie man die Gewichte für unterschiedliche Fragestellungen einsetzt.

### *Analysen zu Leistungsbeziehern im Juli 2006*

Aussagen über die Grundgesamtheit der BA-Stichprobe (Haushalte, in denen es im Juli 2006 mindestens eine Bedarfsgemeinschaft mit Bezug von Leistungen nach dem SGB II gab; im Folgenden bezeichnet als: Haushalte mit Leistungsbezug im Juli 2006) gewinnt man am besten, indem man nur die BA-Stichprobe und die zugehörigen Gewichte verwendet. Dieses Vorgehen ist dann effizienter als die Verwendung der Gesamtstichprobe, da die Gewichte in der BA-Stichprobe eine kleinere Varianz aufweisen.

#### ANALYSEN AUF HAUSHALTSEBENE

Will man Aussagen über Haushalte mit Leistungsbezug im Juli 2006 treffen, so verwendet man folglich `wqbahh1`. Dies sei im Folgenden an einem Beispiel für Stata (ab Version 9) demonstriert. Berechnet werden soll die Anzahl bzw. der Anteil der Leistungsbezieherhaushalte, die über ein Auto verfügen (Variable `HLS0800a`). Zunächst müssen dem Haushaltsdatensatz die Haushaltsgewichte zugespielt werden, dann muss das *Surveyset* durchgeführt werden, danach kann der hochgerechnete Wert berechnet werden:

```
use HHENDDAT.dta, clear
merge hnr using hweights.dta
svyset psu [pw=wqbahh1], strata(strpsu)
svy: tab HLS0800a, count cell format(%9.0g)
```

Etwa 37,9% der Haushalte mit mindestens einem Leistungsbezieher verfügen über ein Auto, 62,1% verfügen über kein Auto, der Anteil ohne gültige Angabe ist verschwindend gering.

#### ANALYSEN AUF BEDARFSGEMEINSCHAFTSEBENE

---



Bei Fragestellungen zum SGB II - Leistungsbezug interessieren häufig nicht Haushalte, sondern Bedarfsgemeinschaften. Soll die obige Fragestellung nach dem Anteil der Haushalte mit Leistungsbezug im Juli 2006, die über ein Auto verfügen, auf Bedarfsgemeinschaften übertragen werden, so kann man mit den Daten des PASS die Fragestellung beantworten, wie viele Bedarfsgemeinschaften in einem Haushalt leben, der über ein Auto verfügt (da die Bedarfsgemeinschaften nachträglich identifiziert wurden, gibt es keine Fragen im Fragebogen, die sich direkt auf Bedarfsgemeinschaften beziehen – man kann also in Haushalten mit mehreren Bedarfsgemeinschaften nicht zuordnen, welcher Bedarfsgemeinschaft das Auto gehört). Diese Frage ist relativ einfach zu beantworten, wenn die Variable *nbgbezug* zu Hilfe genommen wird, die angibt wie viele zum Stichprobenziehungszeitpunkt Arbeitslosengeld II beziehende Bedarfsgemeinschaften ein Haushalt enthält<sup>10</sup>. Der schnellste Weg ist die Multiplikation der Haushaltsgewichte mit diesem Wert.

```
use HHENDDAT.dta, clear
merge hnr using hweights.dta
gen bgweight=wqbahh1*nbgbezug
svyset psu [pw=bgweight], strata(strpsu)
svy: tab HLS0800a, count cell format(%9.0g)
```

Abweichend zu der oben präsentierten Analyse verschieben sich die Anteile leicht (37,9% der beziehenden Haushalte, aber 38,2% der beziehenden Bedarfsgemeinschaften verfügen über ein Auto im Haushalt). Vor allem aber haben sich die Absolutzahlen verändert: Die Summe aller Haushalte mit Bezug war 3.882.013, die Summe aller beziehenden BGs ist dagegen nun 4.011.889 und stimmt aufgrund der Kalibrierung mit dem Eckwert der BA-Statistik überein.

---

<sup>10</sup> In dieser Variable sind die zu treffenden Entscheidungen, wenn aus den Angaben nicht eindeutig hervorgeht, wie viele Bedarfsgemeinschaften im Haushalt Arbeitslosengeld II beziehen, genau so getroffen worden wie bei der Kalibrierung. Jedem Nutzer steht es natürlich frei, eigene Entscheidungen auf Basis der Arbeitslosengeld-II-Spells zu treffen.

---

## ANALYSEN AUF PERSONENEbene

Analysen auf Personenebene sind ähnlich einfach. Zu verwenden ist hier das Gewicht `wqbap1`. Ein Zwischenschritt wird notwendig, da die Variablen `psu` und `strpsu` nur im Haushaltsdatensatz enthalten sind. Im folgenden Beispiel wird die Anzahl der Personen ab 15 Jahren<sup>11</sup> in Haushalten mit Leistungsbezug berechnet, die einen Migrationshintergrund (Variable *migration*) haben.

```
use HHENDDAT.dta, clear
keep hnr psu strpsu
sort hnr
save psui nfo, replace

use PENDDAT.dta, clear
merge pnr using pweights.dta
drop _m
sort hnr
merge hnr using psui nfo

svyset psu [pw=wqbap1], strata(strpsu)
svy: tab migration, count cell format(%9.0g)
```

Knapp 61,3% haben demnach keinen Migrationshintergrund, 24,4% sind selbst zugezogen, bei weiteren 7,6% ist mindestens ein Elternteil zugezogen und bei weiteren 1,9% mindestens ein Großelternteil. Für 3,6% erhält man den Code „Item in Fragebogenversion nicht erhoben“<sup>12</sup>. Das liegt daran, dass die Daten des Kurzfragebogens für Personen ab 65 Jahren im selben Datensatz gehalten werden wie Daten des Personenbogens. Personen ab 65 Jahren erhalten dann bei Fragen, die im Seniorenfragebogen nicht gestellt wurden, diesen Code. Will man bei den Auswertungen auf diese Personen verzichten, so kann man die Auszählung auf Daten aus Personenfragebögen (`fb_vers=1`) einschränken

```
svy: tab migration if fb_vers==1, count cell format(%9.0g)
```

<sup>11</sup> Da jüngere Personen nicht selbst befragt werden, kann man über sie mit den Daten des PASS nur Aussagen zu den Merkmalen machen, die in den Haushaltsbögen erhoben werden (z.B. Alter, Geschlecht). Dafür sind dann wiederum die Haushaltsgewichte zu verwenden.

<sup>12</sup> Für weitere 1,2% kann das Merkmal aufgrund fehlender Informationen nicht gebildet werden

Die Personengewichte der BA-Stichprobe rechnen auf alle Personen in Haushalten mit Leistungsbezug hoch. Manche Haushalte bestehen jedoch aus mehreren synthetischen Bedarfsgemeinschaften, von denen nicht alle Leistungen beziehen. Will man lediglich auf Personen hochrechnen, die selbst zu SGB II-Bedarfsgemeinschaften gehören, so muss man wiederum Personen mit *bgbez1=0* ausschließen (dafür muss man dem Personendatensatz die Variable *bgbez1* aus dem Personenregister *p\_register* zuspielen):

```
drop _m
sort pnr
merge pnr using p_register.dta
tab _m pnetto1
drop if _merge==2
svy, subpop(if bgbez1==1): tab migration if fb_vers==1, count ///
cell format(%9.0g)
```

Der Anteil der selbst zugezogenen ist folglich unter den zu einer Bedarfsgemeinschaft gehörenden Personen mit 25,5% geringfügig größer als unter den Personen, die in einem Haushalt mit Leistungsbezug leben (25,3%).

### *Analysen zur Wohnbevölkerung der Bundesrepublik*

Analysen zur Wohnbevölkerung der Bundesrepublik können jeweils sowohl mit den Gesamtgewichten als auch mit den Microm-Gewichten durchgeführt werden. Die Ergebnisse werden sich in den meisten Fällen nur geringfügig unterscheiden. Der Anteil der Haushalte mit Auto in der Gesamtbevölkerung berechnet sich demnach entweder mit den folgenden Befehlszeilen mit Hilfe der Gesamtgewichte:

```
use HHENDDAT.dta, clear
merge hnr using hweights.dta
svyset psu [pw=wqhh1], strata(strpsu)
svy: tab HLS0800a, cell ci format(%10.0g)
```

oder alternativ mit den Microm-Gewichten

```
svyset psu [pw=wqmi hh1], strata(strpsu)
svy: tab HLS0800a, cell ci format(%10.0g)
```

---

Der Anteil der Haushalte mit Auto ist im ersten Fall 76,0% (95%-Konfidenzintervall von 74,1% bis 77,9%), im anderen 75,8% (95%-Konfidenzintervall von 73,7% bis 77,8%). Etwas schmaler ist das Konfidenzintervall bei Verwendung der Gesamtgewichte, weil der Teil der Bevölkerung, der Leistungen nach dem SGB II bezieht, hier viel genauer abgebildet wird, weshalb wir deren Verwendung bevorzugen. Für die Personengewichte gilt dasselbe.

### *Analysen zu Leistungsbeziehern am aktuellen Rand*

Arbeitet man mit der BA-Stichprobe und den zugehörigen Gewichten, so beziehen sich Ergebnisse jeweils auf die Bezieher im Juli 2006. Für Aussagen über diese Population erzielt man mit diesem Vorgehen die größte statistische Power, da die BA-Gewichte eine relativ geringe Varianz haben. Viele Analysen – vor allem zu schnell veränderlichen Merkmalen – wird man dagegen am aktuellen Rand der Befragung durchführen wollen, auf den sich viele Merkmale – wie etwa Erwerbstatus, Einkommen oder Beschäftigungsumfang - beziehen. Der Befragungszeitpunkt liegt zwischen 6 und 13 Monaten nach dem Zeitpunkt der Stichprobenziehung. Arbeitet man am aktuellen Rand ausschließlich mit der BA-Stichprobe, so kann man lediglich Aussagen über so genannte *Stayer* treffen, die vom Stichprobenziehungszeitpunkt bis zum Befragungszeitpunkt im Bezug verblieben sind. Angesichts eines recht hohen Turnovers (37% der Personen, die im Januar 2005 Leistungen nach dem SGB II bezogen haben, haben den Bezug bis Dezember 2006 verlassen (Graf 2007)), kann diese Gruppe sich in ihrer Zusammensetzung deutlich von den aktuellen Beziehern unterscheiden.

### ANALYSEN ZU BEZIEHERN AM AKTUELLEN RAND AUF HAUSHALTSEBENE

Für die aktuellen Leistungsbezieher repräsentative Ergebnisse können daher nur mit den Gesamtgewichten erzielt werden. Das Merkmal, ob ein aktueller Bezug des Haushalts vorliegt (*alg2abez*), ist im Haushaltsdatensatz (HHENDDAT.dta) enthalten. Für Analysen auf Haushaltsebene ist die Berechnung daher relativ einfach. Das folgende Beispiel zeigt das wiederum für die Frage des Autobesitzes.

---

```
use HHENDDAT.dta, clear
merge hnr using hweights.dta
svyset psu [pw=wqhh1], strata(strpsu)
svy, subpop(if alg2abez==1): tab HLS0800a, cell ci format(%9.0g)
```

Von den aktuellen Leistungsbezieherhaushalten verfügen 36,2% über ein Auto. Der Wert ist etwas geringer als für die Bezieherhaushalte im Juli 2006.

Hätte man dies mit den BA-Gewichten und der BA-Stichprobe geschätzt,

```
svyset psu [pw=wqba1], strata(strpsu)
svy, subpop(if alg2abez==1): tab HLS0800a, cell ci format(%9.0g)
```

so wäre ein Wert von 33,7% errechnet worden. Da es sich hierbei nur um *Stayer* handelt, also um Haushalte, die sowohl im Juli 2006 als auch zum Befragungszeitpunkt Leistungen bezogen haben, ist es plausibel, dass diese Haushalte seltener über ein Auto verfügen als Haushalte, die in diesem Zeitraum aus dem Bezug abgegangen oder in den Bezug gelangt sind.

Eine Konsequenz aus der Verwendung der Gesamtgewichte gegenüber den BA-Gewichten ist die deutliche Vergrößerung der Konfidenzintervalle. Die Varianz der Gesamtgewichte ist aufgrund der stark unterschiedlichen Sampling-Raten in beiden Teilstichproben wesentlich größer. Die Analysen zum Autobesitz der Haushalte mit Arbeitslosengeld-II-2-Bezug zum Ziehungszeitpunkt (s.o.) ergaben ein 95%-Konfidenzintervall von 36,0% bis 39,7%, für den Befragungszeitpunkt erhalten wir ein beinahe doppelt so großes 95%-Konfidenzintervall von 33,5% bis 39,1%.

#### ANALYSEN ZU BEZIEHERN AM AKTUELLEN RAND AUF BEDARFGEMEINSCHAFTSEBENE

Im Vergleich zu den Analysen zum Stichprobenziehungszeitpunkt (in 5.3.1.2) ist ein weiterer Schritt zu vollziehen, da keine zu *nbgbezug* äquivalente Variable für den Leistungsbezug zum Befragungszeitpunkt existiert. Diese muss erst gebildet werden. Dazu kann die Variable *bgbezb1* in *p\_register* verwendet werden, die für jede Bedarfsgemeinschaft kennzeichnet, ob ein ALG-II-Bezug zum Befragungszeitpunkt vorliegt<sup>13</sup>.

---

<sup>13</sup> Im Beispielcode werden durch „recode bgbezb1 (-5=0)“ alle Bedarfsgemeinschaften, für die der aktuelle Leistungsbezug aufgrund der Angaben im Survey unklar ist, wie nicht beziehende Bedarfsgemeinschaften behandelt.

```
use p_register.dta, clear
collapse (mean) hnr bgbezb1, by(bgnr)
recode bgbezb1 (-5=0)
by hnr, sort: egen nbgbezak=sum(bgbezb1)
collapse nbgbezak, by(hnr)
sort hnr
save hnr_nbgbezak.dta, replace
use HHENDDAT.dta, clear
merge hnr using hweights.dta
drop _m
sort hnr
merge hnr using hnr_nbgbezak.dta
gen bgw_akt=wqhh1*nbgbezak
svyset psu [pw=bgw_akt], strata(strpsu)
svy, subpop(if alg2abez==1): tab HLS0800a, cell ci format(%9.0g)
```

Der geschätzte Wert von 36,2% unterscheidet sich nicht von der Analyse auf Haushaltsebene. Allerdings bezieht sich der Wert nicht mehr auf eine Subpopulation von knapp 3.489.000 Haushalten wie im vorigen Abschnitt, sondern auf knapp 3.519.000 Bedarfsgemeinschaften mit Bezug zum Befragungszeitpunkt. Im Befragungszeitraum schwankte die Zahl der Bedarfsgemeinschaften laut BA-Statistik zwischen 3.724.000 (Juli 07) und 3.818.000 (März 07). Dieser Eckwert wird also nicht genau getroffen. Die Untererfassung rührt daher, dass anders als zum Stichprobenziehungszeitpunkt der Bezug laut Prozessdaten für den Befragungszeitpunkt nicht für alle Befragten zur Verfügung steht und somit das Unterberichten des Leistungsbezugs<sup>14</sup> am aktuellen Rand nicht korrigiert wird.

---

<sup>14</sup> Da der ALG-II-Leistungsbezug ein sozial unerwünschtes Merkmal sein dürfte, ist ein *underreporting* in einem gewissen Umfang nicht verwunderlich.

## ANALYSEN ZU BEZIEHERN AM AKTUELLEN RAND AUF PERSONENEBENE

Der Übergang auf Personenebene kann analog zum Fall für den Stichprobenziehungszeitpunkt bewerkstelligt werden. Zunächst spielt man dem Personendatensatz wieder die Personengewichte und die Informationen für das *Surveyset* zu. Für Analysen zu Personen aus aktuell beziehenden Haushalten sind die Auszählungen auf Personen mit *alg2abez=1* einzuschränken. Dieses Merkmal ist aus dem Haushaltsdatensatz zuzuspielen.

```
use HHENDDAT.dta, clear
keep hnr psu strpsu alg2abez
sort hnr
save psu_alg2_info, replace

use PENDDAT.dta, clear
merge pnr using pweights.dta
drop _m
sort hnr
merge hnr using psu_alg2_info
drop _m
svyset psu [pw=wqgesp1], strata(strpsu)
svy, subpop(if alg2abez==1): tab migration if fb_vers==1, count ///
cell format(%9.0g)
```

Von den Personen in Haushalten, die aktuell Arbeitslosengeld II beziehen, haben danach 60,6% keinen Migrationshintergrund, 26,1% sind selbst zugezogen, 8,1% haben mindestens ein zugezogenes Elternteil und 2,0% ein zugezogenes Großelternanteil.

In den meisten Fällen wird man aber die Analysen nicht auf Personen in beziehenden Haushalten, sondern auf Personen in beziehenden Bedarfsgemeinschaften einschränken. Dieses Merkmal ist wiederum dem Personenregister zu entnehmen. Die folgende Befehlsfolge produziert die Migrantenanteile an Personen in Bedarfsgemeinschaften zwischen 15 und 64.

```
drop _m  
sort pnr  
merge pnr using p_register.dta  
svy, subpop(if bgbezb1==1): tab migration if fb_vers==1, count cell
```

### *Analysen zu Leistungsbeziehern zu weiteren Zeitpunkten*

Die Arbeitslosengeld-II-Bezugsbiographien auf Haushaltsebene ermöglichen es prinzipiell auch Analysen zu anderen Zeitpunkten zwischen Stichprobenziehung und Erhebung der ersten Welle zu machen. Variablen wie *bgbezs1*, *bgbezb1* oder *nbgbezug* werden aber nur für die beiden oben behandelten Zeitpunkte ausgeliefert. Nutzer die auf andere Zeitpunkte hochrechnen möchten, müssen sich daher analoge Variablen bilden. Dabei sind immer Unschärfen und das Problem der Untererfassung des Leistungsbezugs zu behandeln.

### *Vergleich von Leistungsbeziehern mit der allgemeinen Bevölkerung*

Aus der obigen Vielzahl von Möglichkeiten, Leistungsbezieher und ihre Haushalte bzw. Bedarfsgemeinschaften zu untersuchen, resultiert eine ebensolche Vielzahl von Möglichkeiten, Leistungsbezieher mit der allgemeinen Bevölkerung zu vergleichen. Einen Überblick gibt Tabelle 9. Zu verwenden sind jeweils die Gesamtgewichte.

---



**Tabelle 9: Variablen und ihre Verwendungsmöglichkeiten beim Vergleich von SGB II-Leistungsbeziehern mit der allgemeinen Bevölkerung**

Variable	Datensatz	Ausprägungen	Ist geeignet zum Vergleich von ...
sample	PENDDAT	1 BA-Stichprobe	<p>a) Haushalten mit ALG-II-Leistungsbezug im Juli 2006 (sample=1) mit Haushalten der Wohnbevölkerung (sample=2).</p> <p>b) Personen in Haushalten mit ALG-II-Leistungsbezug im Juli 2006 (sample=1) mit Personen in Haushalten der Wohnbevölkerung (sample=2).</p> <p>ALG-II-Leistungsempfänger-Haushalte werden dabei über die Teilstichprobe definiert.</p>
	HHENDDAT	2 Microm-Stichprobe	
alg2samp	hh_register	0 kein Bezug	<p>a) Haushalten mit ALG-II-Leistungsbezug im Juli 2006 (alg2samp=1) mit Haushalten ohne ALG-II-Leistungsbezug im Juli 2006 (alg2samp=0).</p> <p>b) Personen in Haushalten mit ALG-II-Leistungsbezug im Juli 2006 (alg2samp=1) mit Personen in Haushalten ohne ALG-II-Leistungsbezug im Juli 2006 (alg2samp=0).</p> <p>Dem Nutzer bleibt die Entscheidung überlassen, wie er die Fälle behandelt, die zwar laut Stichprobe ALG-II-Leistungen bezogen haben, laut Survey aber nicht.</p>
		1 Bezug	
		2 Laut Survey kein Bezug (BA-SP)	
		3 Laut Survey Bezug unklar (BA-SP)	
		4 Laut Survey Bezug unklar (Microm-SP)	
bgbezs1	p_register	1 ALG-II-Bezug bei Stichprobenziehung	<p>Personen in Bedarfsgemeinschaften mit ALG-II-Leistungsbezug im Juli 2006 (bgbezs1=1) mit Personen in Bedarfsgemeinschaften ohne ALG-II-Leistungsbezug im Juli 2006 (bgbezs1=0)</p> <p>Da diese Variable bei der Gewichtung verwendet wurde, wurde für jeden unklaren Fall eine Entscheidung getroffen.</p>
		0 Kein ALG-II-Bezug bei Stichprobenziehung	
alg2abez	HHENDDAT	1 HH bezieht aktuell ALG II	<p>a) Haushalten mit ALG-II-Leistungsbezug zum Befragungszeitpunkt (alg2abez=1) mit Haushalten ohne ALG-II-Leistungsbezug zum Befragungszeitpunkt (alg2abez=2).</p> <p>b) Personen in Haushalten mit ALG-II-Leistungsbezug zum Befragungszeitpunkt (alg2abez=1) mit Personen in Haushalten ohne ALG-II-Leistungsbezug zum Befragungszeitpunkt (alg2abez=2).</p>
		2 HH bezieht aktuell kein ALG II	
		-5 Generierung nicht mgl. (fehlende Werte)	
bgbezb1	p_register	1 ALG-II-Bezug in Welle 1	<p>Personen in Bedarfsgemeinschaften mit ALG-II-Leistungsbezug zum Befragungszeitpunkt (bgbezb1=1) mit Personen in Bedarfsgemeinschaften ohne ALG-II-Leistungsbezug zum Befragungszeitpunkt (bgbezb1=0)</p>
		0 Kein ALG-II-Bezug in Welle 1	
		-5 Generierung nicht mgl. (fehlende Werte)	

## 8. Anonymisierung

Vom IAB als Abteilung 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 mehr oder minder stark vergrößerter Form in den Datensatz 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. Tabelle 10 gibt einen Überblick über die betroffenen Variablen und die vorgenommenen Anonymisierungsschritte.<sup>15</sup>

**Tabelle 10: Übersicht über die anonymisierten Variablen**

Frage	Varname	Varlabel	Vorgehen
p1	PD0100	Geburtsjahr	Das genaue Geburtsdatum wurde zu Geburtsjahr und zu einem Indikator für 1. bzw. 2. Halbjahr verkürzt.
	gebhalbj	Geburtshalbjahr	
p38	PET1500	Stellung im Beruf	In der Befragung wurden aus erhebungstechnischen Gründen Berufssoldaten und Richter separat erhoben. Aufgrund der geringen Fallzahlen und da diese ohnehin üblicherweise nicht im Rahmen der Stellung im Beruf erhoben werden, wurden diese beiden Gruppen mit den Beamten zu einer gemeinsamen Gruppe zusammengeführt.
p41	PET1800	Jetzige StiB Berufssoldat	Die Variable wurde nicht ausgeliefert. Die Fälle wurden den entsprechenden Kategorien in PET1900 zugeschlagen.
p42	PET1900	Jetzige StiB Beamter	Diese Variable enthält zusätzliche Fälle. Die Richter aus P38 wurden den Beamten im höheren Dienst zugeschlagen, die Berufssoldaten aus P41 der entsprechenden Beamtenkategorie.
p192	PMI0200	Nicht in D geboren: Geburtsland	Länder mit zu geringen Fallzahlen wurden zu größeren Kategorien zusammengefasst.
p196	PMI0500	Keine D Staatsangehörigkeit: welche Staatsangehörigkeit?	Staatsangehörigkeiten von Ländern mit zu geringen Fallzahlen wurden zu größeren Kategorien zusammengefasst.
p201a – p201f	PMI1000a -PMI1000f	Zuzugsland Eltern und Großeltern	Länder mit zu geringen Fallzahlen wurden zu größeren Kategorien zusammengefasst.
Generiert	ogebland	Geburtsland, incl. offene Angaben	Länder mit zu geringen Fallzahlen wurden zu größeren Kategorien zusammengefasst.
Generiert	ostaatan	Staatsangehörigkeit, incl. offene Angaben	Staatsangehörigkeiten von Ländern mit zu geringen Fallzahlen wurden zu größeren Kategorien zusammengefasst.
Generiert	ozulanda - ozulandf	Zuzugsland Eltern und Großelt., incl. o. Ang.	Länder mit zu geringen Fallzahlen wurden zu größeren Kategorien zusammengefasst.

<sup>15</sup> 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.

## 9. Literatur

- AAPOR (2006), Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 4th Edition. AAPOR, Lanexa.
- Achatz, Juliane, Hirseland, 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.
- 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.
- 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.
- Christoph, Bernhard (2005): Zur Messung des Berufsprestiges: Aktualisierung der Magnitude-Prestigeskala auf die Berufsklassifikation ISCO88. ZUMA-Nachrichten 57: 79-127.
-

- Christoph, Bernhard, Sachweh, Patrick und Müller, Tim (2005): Umfrage zum Forschungsprojekt „Die Akzeptanz des Wohlfahrtsstaats“. Codebuch und Dokumentation. Mannheimer Zentrum für Europäische Sozialforschung (MZES), Mannheim.
- 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.
- Graf, Tobias (2007). Bedarfsgemeinschaften 2005 und 2006: Die Hälfte war zwei Jahre lang durchgehend bedürftig. IAB-Kurzbericht 17/2007.
-

- 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.
- Infas (2006), IAB-Projekt: "Zusammenlegung von Arbeitslosen- und Sozialhilfe mit Einführung des SGB II: Übergangsmuster, Strukturen und Folgen für die Betroffenen". Repräsentative Querschnittsbefragung. Abschlussbericht zur methodischen Umsetzung, Infas, Bonn.
- International Labour Office [ILO] (1990): International Standard Classification of Occupations. ISCO-88. International Labour Office, Genf.
- Kiesl, Hans (2008): Kalibrierte Hochrechnung für das Panel "Arbeitsmarkt und soziale Sicherung" (Welle 1). FDZ-Methodenreport x/2008. Forschungsdatenzentrum der BA im IAB, Nürnberg [im Erscheinen].
-

- 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 Bildungsklassifikation CASMIN für die Volkszählung 1970, die Mikrozensus-Zusatzerhebung 1971 und die Mikrozensus 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.
- Meßmann, Susanne; Bender, Stefan; Rudolph, Helmut; Hirsland, Andreas; Bruckmeier, Kerstin; Wübbecke, Christina; Dundler, Agnes; Städele, Daniela; Schels, Brigitte (2008): Lebenssituation und Soziale Sicherung 2005 (LSS 2005) \* IAB-Querschnittsbefragung SGB II. Handbuch-Version 1.0.0. (FDZ Datenreport, 04/2008 (de)), Nürnberg.
- Müller, Walter, Wirth, Heike, Bauer, Gerrit, Pollak, Reinhard und Weiss, Felix (2007): Entwicklung einer europäischen sozioökonomischen Klassifikation. *Wirtschaft und Statistik*, 5/2007: 527-530.
- 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.
- 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
- 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.
- 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.
- Rudolph, Helmut und Trappmann, Mark (2007), "Design und Stichprobe des Panels „Arbeitsmarkt und Soziale Sicherung“ (PASS)", S. 60-101 in Promberger, Markus. (Hrsg.), Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung "Arbeitsmarkt und Soziale Sicherung", IAB Forschungsbericht 12/2007, Nürnberg.
- Särndal, Carl-Erik, Swensson, Bengt und Wretman, Jan (1992): Model Assisted Survey Sampling. Springer, New York.
- Schnell, Rainer (1997), Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen, Leske+Budrich, Opladen.
- Schnell, Rainer (2007). Alternative Verfahren zur Stichprobengewinnung für ein Haushaltspanelsurvey mit Schwerpunkt im Niedrigeinkommens- und Transferleistungsbezug, S. 33-59, in Promberger, Markus. (Hrsg.), Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung "Arbeitsmarkt und Soziale Sicherung", IAB Forschungsbericht 12/2007, Nürnberg.
- Sozialgesetzbuch Zweites Buch [SGB II]: Grundsicherung für Arbeitssuchende.
- Statistisches Bundesamt [StBA] (1992): Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen. Statistisches Bundesamt, Wiesbaden.
- Trappmann, Mark, Christoph, Bernhard, Achatz, Juliane und Wenzig, Claudia (2007). "Labour Market and Social Security": A New Panel Study for Research on Long
-

Term Unemployment. Paper Presented at the International Conference of the German Association of Political Economy, Nuremberg, October 11th, 2007. [http://doku.iab.de/veranstaltungen/2007/cape\\_2007\\_trappmann.pdf](http://doku.iab.de/veranstaltungen/2007/cape_2007_trappmann.pdf)

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

Wagner, Gert G., Frick, Joachim R. und Schupp, Jürgen (2007), The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements, Schmollers Jahrbuch 127 (1), 139-169.

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

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

Zentrum für europäische Wirtschaftsforschung [ZEW], Institut für Arbeit und Qualifikation [IAQ] und TNS Emnid (2007). Evaluation der Experimentierklausel nach §6c SGB II - Vergleichende Evaluation des arbeitsmarktpolitischen Erfolgs der Modelle der Aufgabenwahrnehmung "Optierende Kommune" und "Arbeitsgemeinschaft". Untersuchungsfeld 3: "Wirkungs- und Effizienzanalyse". Erster Bericht. Mannheim, Gelsenkirchen, Bielefeld.

---





## *Imprint*

### **FDZ** *Datenreport*

No. 5/2008

#### **Publisher**

The Research Data Centre (FDZ)  
of the Federal Employment Service  
in the Institute for Employment Research  
Regensburger Str. 104  
D-90478 Nuremberg

#### **Editorial staff**

Stefan Bender, Dagmar Herrlinger

#### **Technical production**

Dagmar Herrlinger

#### **Copyright**

Reproduction – also in parts – only with permission of  
the FDZ

#### **Download**

[http://doku.iab.de/fdz/reporte/2008/DR\\_05-08\\_I.pdf](http://doku.iab.de/fdz/reporte/2008/DR_05-08_I.pdf)

#### **Internet**

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

#### **Corresponding authors**

Bernhard Christoph, Institute for Employment Re-  
search, Regensburger Str. 104, D-90478 Nuremberg;  
Phone: +49-(0)911/179-3507;  
E-Mail: [bernhard.christoph@iab.de](mailto:bernhard.christoph@iab.de)

Mark Trappmann, Institute for Employment Research,  
Regensburger Str. 104, D-90478 Nuremberg;  
Phone: +49-(0)911/179-3096  
E-Mail: [mark.trappmann@iab.de](mailto:mark.trappmann@iab.de)

---