FDZ·Datenreport

Documentation of labour market data

Codebook and Documentation of the Panel Study
'Labour Market and Social Security' (PASS)

Datenreport Wave 5

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

Bundesagentur für Arbeit

Codebook and Documentation of the Panel Study

'Labour Market and Social Security' (PASS)

Datenreport Wave 5

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

Table of Contents

List of Tables

List of Figures

Data availability

The dataset described in this document is available for use by professional researchers. For further information, please refer to http://fdz.iab.de/.

1 Introduction

1.1 Objectives and research questions of the panel study 'Labour Market and Social Security'

The panel study 'Labour Market and Social Security' (PASS), established by the Institute for Employment Research (IAB), is a new dataset for labour market, welfare state and poverty research in Germany, creating a new empirical basis for the scientific community and for policy counselling.

The study is carried out as part of the IAB's research into the German Social Code Book II (SGB II)[1]. The IAB has the statutory mandate to study the effects of benefits and services under SGB II aimed at integration into the labour market and subsistence benefits. However, due to its complex sample design, the study also enables researchers to answer questions far beyond these issues. Five core questions influenced the development of the new study, which are explained in detail in Achatz, Hirseland and Proberger (2007):

1. What options are there for regaining independence from Unemployment Benefit II (Arbeitslosengeld II)?

2. How does the social situation of a household change when it receives benefits?

3. How do the individuals concerned cope with their situation? Does their attitude towards action necessary to improve their situation change over time?

4. In what form does contact between benefit recipients and institutions providing basic social security take place? What are the actual institutional procedures applied in practice?

5. What employment history patterns or household dynamics lead to receipt of Unemployment Benefit II?

This Datenreport provides an overview of the fifth survey wave, for which 15,607 individuals were interviewed in 10,235 households[2] between February 2011 and September 2011. This included 9,693 individuals and 6,547 households that had already been interviewed in the context of PASS.[3]

---

[1] Social Code Book II – basic security for job-seekers (Sozialgesetzbuch (SGB) Zweites Buch (II) - Grundsicherung für Arbeitsuchende).

[2] The figures comprise evaluable interviews only. For repeatedly interviewed households also those were considered for which only a household interview without a personal or senior citizens' interview could be conducted.

[3] The panel household sample was supplemented for both recipients of Unemployment Benefit II and the general population sample from new postcode regions in wave 4.

The present wave-specific Datenreport[4] of wave 5 documents the wave-related aspects of the study. Following a short overview of the innovations and characteristics of wave 5 (Chapter 1.3), the Datenreport reports the key figures on samples and response rates of wave 5 (Chapter 2). Moreover, the steps of data preparation and the decisions made as part of this process are described (Chapter 5) and an overview of the variables generated is presented (Chapter 4). Additionally, the weighting procedure is presented (Chapter 6). The separate table reports list the frequencies of all variables included in the scientific use file that were recorded in wave 5, divided into their respective datasets (Volume II to Volume V).

1.2 Instruments and interview programme

Information in PASS is collected by means of separate questionnaires at the household and the individual level. First, a household interview is conducted with each household. This interview gathers information referring to the entire household. The target person for this household interview[5] is already selected during the contact phase which precedes the actual interviews. Personal interviews with the individual household members follow the household interview. The aim is to conduct a personal interview with all of the individuals living in the household who are aged 15 or over – household members who are 65 or over receive a short version of the questionnaire (senior citizens' questionnaire) which does not include questions that are irrelevant for this age group.

The survey instruments and interview programme of wave 5 are based on those used in wave 4 of PASS. However, individual questions and modules have been revised or redeveloped (see Chapter 1.3. for an overview).

---

[4] The report was divided into two components for the first time starting with the wave 3 documentation: a wave-specific Datenreport (including codebook) and a cross-wave Uof the PASS user guideser Guide. The PASS project team at the IAB is responsible for creating the cross-wave User Guide. As of wave 3, infas has been creating the documentation of the wave-specific Datenreport. It is based on the Datenreport of wave 2. The cross-wave User Guide aims to document the study as a whole. It describes in detail the objectives and the design of PASS and presents the contents and instruments of the survey. Moreover, it describes the structure of the scientific use file and the concept of the variable types and their names.

[5] The target person for the household interview should know as much as possible about general issues regarding the household. The selection was based on certain rules and is documented in detail in the methods report (Jesske & Quandt, 2011).

The PASS survey instruments are designed in such a way that they allow repeat interviews of individuals and households that already participated in a previous wave but also first-time interviews[6].

In order to avoid seam effects[7] in the repeat interviews and to increase data quality, dependent interviewing has been used for certain questions since wave 3 to update information that the respondent had provided in the previous interview. Furthermore, to a great extent, information about constant characteristics was not gathered again. Unlike in waves 1 to 3, there has been an integrated questionnaire at the household level for repeatedly interviewed households (HHalt) and for first-time interviewed households (HHneu) as of wave 4[8].

The cross-wave PASS User Guide describes the individual instruments and the interview programme in detail. The following section provides an overview of the characteristics and innovations of wave 5.

1.3 Characteristics and innovations of wave 5

At this point we would like to provide a brief outline of the characteristics of wave 5 of PASS for users who have already worked with the data from the panel waves.

The characteristics and innovations in wave 5 affect the set of questions for the household and personal questionnaire (change of reference periods, modification of individual questions and new question modules)[9], the sample and data preparation.

1.3.1 Personal questionnaire

The personal questionnaire updates the employment history information surveyed since wave 2[10]. Wave 5 maintains the logic of chronological retrospective surveying which was introduced in wave 4 (see section 1.3.1 in Berg et al., FDZ Datenreport 08/2011).

---

[6] First-time interviewed households include: (1) Households from the refreshment and replenishment samples of the current wave and (2) households which split off from households interviewed in previous waves (split-off households) (for further explanations, please the see the wave 4 methods report (Jesske & Quandt, 2011)).

[7] In a panel dataset the number of changes observed at the interface (seam) between one interview and the interview conducted in the subsequent panel wave is often considerably higher than the number of changes observed within one interview (see Jäckle 2008).

[8] Split-off households are treated like new households in the survey.

[9] Minor changes in the set of questions (adding, modifying or deleting individual questions) are not listed completely.

[10] Among others, this is made using the so-called "dependent interviewing" method. Dependent interviewing includes information which repeatedly interviewed individuals provided in the previous wave interview in the interview text of the current interview to check whether this information must be updated.

A structural change was made in the employment biography module in wave 5. The current gross and net income was no longer surveyed as summary value across all continuing employments but relating to the respective employment (ET2800-ET3900). This leads to the generation of new variables which will be explained in detail in Chapter 4. Consequently, the former variables on gross and net income (PEK0100b-PEK1200) are omitted. A survey of summary values exists in wave 5 only for special payments from the past year (PEK1360b) and for government payments for employed persons (PEK2100).

Moreover, the employment module now again includes the question from wave 3 regarding the time of cancellation of limitation of an initially limited employment (ET1753) and a variable which enquires from which sources the respondents with a new employment had first heard about this employment (ET2400).

Further additions in the personal questionnaire in wave 5 concern:

- A special focus module on networks which was already used in wave 3 (in addition to the questions posed in each wave (PSK0100-PSK0400), there are questions regarding network partners outside the household (PSK0205-PSK0270) and social resources (PSK0280a-j and PSK0285a-f).

- The module "job-seeking", in which respondents not seeking employment were asked why this is the case using the item list from wave 1 (PAS0850a-k).

- The module "attitudes (role models)", in which questions regarding gender role allocation (PEO0400a-d) from wave 2 and how money is handled in partnerships (PEO0415, PEO0420, PEO0430, PEO0440, PEO0450) were reintroduced.

- 21 items to enquire about personal characteristics according to "big five" (big five inventory (BFI-K) according to Rammstedt & John (2005)) (PEO1400a-s).

- The module "attitudes", which was supplemented with the subsection "family and employment", for which a set of new questions was developed (PEO0800a-b, PEO0900a-b, PEO1000a-b, PEO1100a-b), and the subsection "working hours" which was extended by the question regarding desired own working hours (PEO1200) and those of the partner (PEO1300).

- Questions regarding affinity for the place of residence (PSK0070a-c and PSK0080).

- The question regarding updating of one-euro jobs (PEE0600).

Furthermore, the personal questionnaire was extended in the face-to-face field by a module regarding "readiness to accept a job". This module surveyed under which conditions respondents were ready to accept a new job offered to them. The question was posed in a factorial survey design using vignettes[11].

---

[11] Vignettes include descriptions of situations or case examples made up of different characteristics which are presented to the respondent instead of individual items. The particularly interesting characteristics in terms of influence are varied in their degree between the case examples.

Five fictitious job offers (vignettes) were varied. They differed regarding income, workload, technical requirements, in-company advancement opportunities, type of contract (limited contract) or distance to the current place of residence. The respondents assessed this regarding attractiveness of the job offers, the probability with which they would accept the job offer and the readiness to move to a new location alone or together with the partner.

Aside from modifications and supplements, the personal questionnaire was reduced as follows:

- The question regarding generalised perceived self-efficacy (PEO0100a-e) in the "life attitudes" module was removed and will be reintroduced in wave 6.

- The question regarding the language spoken in the respondent's circle of friends (PMI1110, PMI1120 und PMI1130) in the "migration" module.

- The questions regarding religious affiliation (PD0200 und PD0300) and religiousness (PD0400) in the "religion" module are only posed to new respondents.

- In the "leisure time" module, the standardised items PA0950a-r are replaced with an open-ended question on leisure time activities (PA1100 and PA1200) and reasons for leisure time activities not pursued (PA1300).

1.3.2 Household questionnaire

There were minor changes in the household questionnaire of wave 5.

- A new feature is a standardised item list of reasons why the own child is not (predominantly) taken care of in a daycare facility or by a childminder.

- In the questions regarding the housing situation the questions regarding the condition of the apartment (HW2000) and the year of moving into this apartment are omitted (HW0900).

- Selected items were omitted in the "deprivation" module: Availability of a heating system (HLS0500a and HLS0500b), availability of a freezer (HLS1300a and HLS1300b) and usage of over-the-counter drugs (HLS2400a and HLS2400b).

1.3.3 Sample and data preparation

In wave 5, like in the previous waves, a so-called refreshment sample was drawn for the BA subsample[12]. The aim is to guarantee the representativeness of the BA sample in the cross-section, and to be able to observe sufficient new transitions into receipt of Unemployment Benefit II over time. For the refreshment sample, benefit unit are drawn which were in receipt of Unemployment Benefit II in July 2010 but not on the sampling date of the first, second, third or fourth wave (see Chapter 2.1 and, on the concept of the refreshment sample, Trappmann et al., 2009). Additionally, there was a panel replenishment of the existing sample in wave 5 by selecting 100 new postcode regions. The panel replenishment includes both the BA and the population sample. However, unlike in wave 1, the population sample was drawn from the registration offices' registers. A detailed description of the procedure can be found in Chapter 6.3. All households which were surveyed for the first time in wave 5 can be identified via the sample indicator (*sample*).

The data preparation was again performed in close cooperation with the IAB. Basic procedures, e. g. for updating datasets and correcting problems in the household structures, were discussed during the preparation process and decided on by the IAB.

The concept for the integration of the spell datasets in the employment module and the necessary preparation steps were discussed and agreed upon with the IAB. The procedure is documented in Chapter 5.7.

2 Key figures

This chapter provides a brief overview of important key figures of the study, such as sample sizes (gross and net) and response rates. For the panel sample, they are represented over the course of the previous four waves and reported both separately for the two original subsamples and the replenishment sample, and for the study as a whole.

- Subsample 1 (BA sample) hereafter refers to the sample of benefit recipients from the process data of the Federal Employment Agency.

- Subsample 2 (MICROM sample) refers to the stratified population sample.

- Refreshment sample 1 (BA sample) is the name of the sample drawn from the SGB II inflow between wave 1 and wave 2.

- Refreshment sample 2 (BA sample) is the name of the sample drawn from the SGB II inflow between wave 2 and wave 3.

- Refreshment sample 3 (BA sample) is the name of the sample drawn from the SGB II inflow between wave 3 and wave 4.

---

[12] Wave 1 of PASS consists of two subsamples: (1) a sample of households in receipt of Unemployment Benefit II drawn from the process data of the Federal Employment Agency (Bundesagentur für Arbeit – BA), and (2) a general population sample, stratified by status, drawn from a database provided by the commercial provider MICROM.

- Refreshment sample 4 (BA sample) is the name of the sample drawn from the SGB II inflow between wave 4 and wave 5.

- Panel replenishment/supplement 1 (municipal register sample) is the name of the sample drawn from the registration office inflows in ten new postcode regions in wave 5.

- Panel replenishment/supplement 2 (BA sample) is the name of the sample drawn from the SGB II inflows in 100 new postcode regions in wave 5.

2.1 Sample size

The sample size in a panel starts with the interviewed households from the first survey wave. In PASS, the gross panel sample contains the interviewed households from wave 1 but also the first-time interviewed households from the refreshment samples of waves 2, 3, 4 and 5. It must be taken into account that only those households interviewed for the first time are available for repeat interviews that are willing to participate in the panel[13]. Agreement to participate in the panel is only recorded in the first interview. A new confirmation of willingness for these households in the subsequent waves is not required. Besides the confirmation of willingness, access to the panel is already induced during the first interview by the general willingness to participate, that is, by realising an interview. Measures to ensure a best possible selection-free access to the panel as part of PASS are described in detail in the method and field report of waves 1 to 5[14].

PASS started with 12,794 conducted household interviews in wave 1; 12,000 of these households agreed to participate in the panel. These households from wave 1 constitute the sample size for the start of the first tracking survey.

The panel concept in PASS assumes that new households or split-off households emerge due to move-outs of individuals from panel households, which are counted as separate households as soon as a household interview was conducted.

This results in an increasing number of households compared to the original sample. Detailed information on the procedures of the panel concept in PASS can be found under "split-off households". Besides the expansion of the panel, there may also be a loss of households due to panel mortality. Households in which all respondents passed away or moved abroad will be removed from the panel gross in the subsequent waves. Moreover, panel losses may occur if no household interview could be conducted for one household for a period of two consecutive waves. This situation could arise for the first time at the end of wave 3 and affects the panel gross in waves 4[15] and 5. The gross sample used for wave 5 comprised a to-

---

[13] The willingness to participate in the panel is granted by the household reference person and is thus valid for all household members. Households willing to participate in the panel have agreed that their address was stored for the purpose of repeat interviews as part of the study.

[14] See Hartmann et al. (2008); Büngeler et al. (2009); Büngeler et al. (2010), Jesske & Quandt (2011), Jesske & Schulz (2012).

[15] The change of the survey institute is another factor influencing the panel gross in wave 4. Transferring the addresses of the panel participants from the IAB to infas required the target person's permission for circulation. For detailed explanations on this procedure and the results, please refer to the methods report of wave 4 (Jesske & Quandt, 2011).

tal of 9,155 panel households. Additionally, each wave includes first-time interviewed households from the refreshment sample and the split-off households, and in wave 5 from the replenishment samples.

The case numbers for the gross sample size of the respective survey waves and subsamples are reported in the following table. In wave 5, at least one interview could be conducted in 6,547 households of the panel sample. In addition, there are 753 first-time interviewed households from the refreshment sample, of which 702 were willing to participate in the panel, and 2,831 from the replenishment samples, of which 2,672 were willing to participate in the panel. The first-time interviewed households of wave 5 covere 104 split-off households which originate from six subsamples of the previous waves.

Table 1: Panel sample on the household level by waves and subsamples[16]

	n	BA	Microm	BA-refreshment 1	BA-refreshment 2	BA-refreshment 3	BA-refreshment 4	EWO supplement	BA supplement	Total
Wave 1	HH-interview realised	6,804	5,990							12,794
	davon: HH willing to participate in panel	6,452	5,548							12,000
Wave 2	Panel-HH gross	6,520	5,611							12,131
	HH-interview realised	3,491	3,897	1,041						8,429
	davon: HH willing to participate in panel	3,360	3,766	1,003						8,129
Wave 3	Panel-HH gross	5,851	5,150	1,010						12,011
	HH-interview realised	3,754	3,901	694	1,186					9,535
	davon: HH willing to participate in panel	3,576	3,777	669	1,145					9,167
Wave 4*	Panel-HH gross	3,926	3,628	863	1,069					9,486
	HH-interview realised	2,815	2,977	563	745	748				7,848
	davon: HH willing to participate in panelt	2,754	2,933	554	727	723				7,691
Wave 5**	Panel-HH gross	3,376	3,319	674	956	726				9,051
	HH-interview realised	2,382	2,680	464	608	517	753	1,510	1,321	10,235
	of this: HH willing to participate in panel	2,347	2,633	456	598	512	702	1,415	1,257	9,920

Source: HH-Register and PENDDAT; Scientific Use File IAB
* Reduction of the gross sample due to objection procedures
** Expansion of the gross sample by supplementation

[16] The scientific use file's register files always comprise the net sample of realised interviews of the respective waves. In the case of split-off households it is possible that there is a subsequent expansion of the panel household gross of the previous wave if the split-off household was identified in the previous wave but could not be realised yet.

The 10,235 household interviews conducted in wave 5 correspond to 15,607 personal interviews. The following table lists the distribution of the respondents across the subsamples and the respective survey waves.

Table 2: **Panel sample size on the individual level by waves and subsamples**

	Personal interview realised	Wave 1	Wave 2	Wave 3	Wave 4*	Wave 5**
		abs.	abs.	abs.	abs.	abs.
Sample	BA	9,386	4,753	4,913	3,958	3,394
	Microm	9,568	6,392	6,207	5,016	4,511
	BA-refreshment 1		1,342	898	786	653
	BA-refreshment 2			1,421	983	822
	BA-refreshment 3				1,025	760
	BA-refreshment 4					1,019
	EWO supplement					2,589
	BA supplement					1,859
	Total	18,954	12,487	13,439	11,768	15,607

Source: P_Register; Scientific Use File IAB

* Reduction of the gross sample due to objection procedures

** Expansion of the gross sample by supplementation

For people without sufficient knowledge of the German language, the interviews were offered in Turkish and Russian. Table 3 indicates how many households or persons were interviewed in the two additional survey languages.

Table 3:   **Panel sample size of foreign-language interviews by waves**

		Russian	Turkish
		abs.	abs.
Wave 1	Households	275	163
	Individuals	432	305
Wave 2	Households	156	39
	Individuals	219	31
Wave 3	Households	210	69
	Individuals	330	109
Wave 4	Households	179	42
	Individuals	285	78
Wave 5	Households	159	36
	Individuals	259	58

Source: PENDDAT; Scientific Use File IAB

For the overall data pool of the realised panel sample the following outline can be drawn regarding households and individuals over the five survey waves.
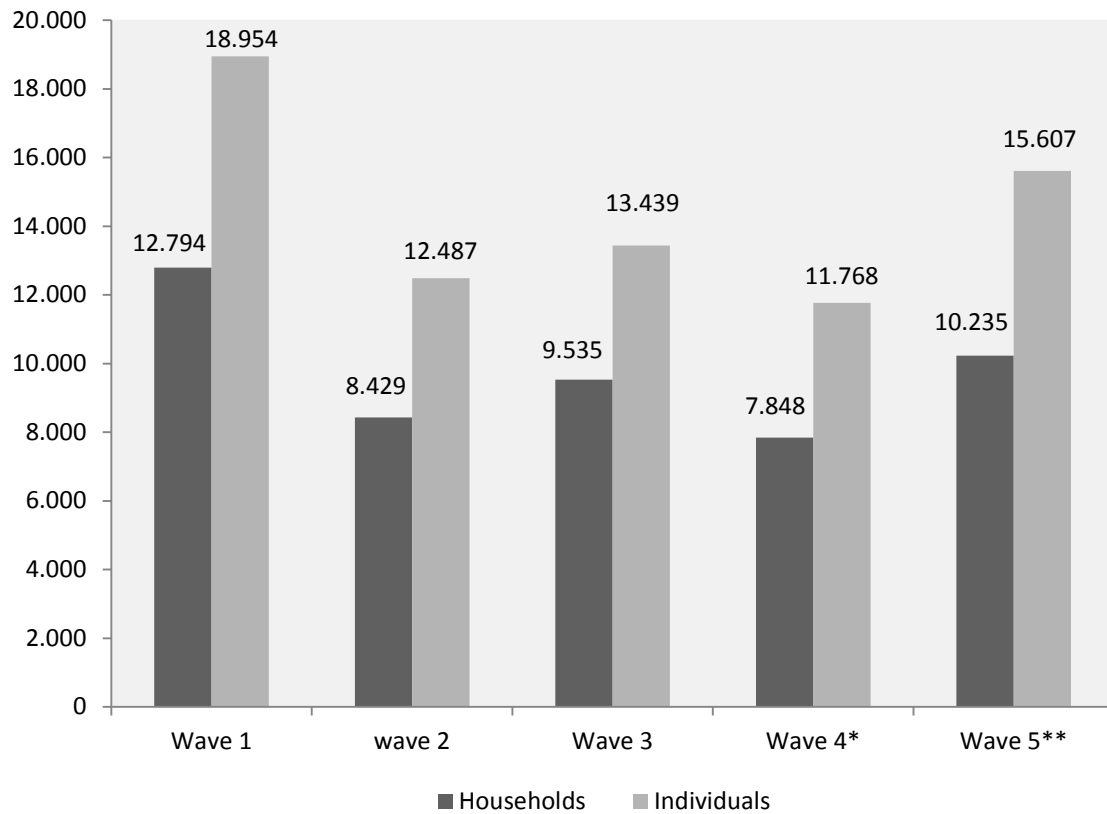
Figure 1:   Realised panel sample from households and individuals by survey waves

* Reduction of the gross sample due to objection procedures
** Expansion of the gross sample by supplementation

2.2 Response rates

The response rate is calculated in accordance with AAPOR standards (AAPOR, 2006). The response rate RR1 is reported, which also includes all cases of unknown eligibility in the denominator and which therefore assumes the lowest value of all response rates[17]. The response rate on the household level is calculated from the share of usable household interviews as a proportion of the total of all usable household interviews and non-neutral non-responses. Only households in which all members passed away and households in which all members moved abroad permanently are regarded as cases of neutral non-response. Households are considered usable if at least one complete household interview is available. New households are only considered usable if not only the household interview but also at least one complete personal interview is available.

---

[17] This is dealt with in very different ways in Germany. Frequently, a large number of individuals or households that were not interviewed are counted as "ineligible" and are removed from the denominator when the response rate is calculated. When a sample is drawn from registers, however, neither a household that is not living at the expected address nor a household that claims not to belong to the target group may be counted as a case of neutral non-response. Moreover, the population of PASS is not restricted to German-speaking respondents or to individuals who are able to be interviewed, so the non-response reasons "does not speak German" or "respondent is sick / unable to be interviewed" cannot be regarded as cases of neutral non-response either.

The following response rates were obtained at the household level for wave 5:

Table 4: **Response rate of wave 5 at the household level by subsamples**

Sample

Wave 5		BA	Microm	BA-refreshment 1	BA-refreshment 2	BA-refreshment 3	BA-refreshment 4	EWO supplement	BA supplement	Total
HH gross	abs.	3,376	3,319	674	956	726	2,792	6,237	5,428	23,508
	%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
neutral nonresponses	abs.	27	21	3	2	3	18	80	27	181
	%	0.8	0.6	0.4	0.2	0.4	0.6	1.3	0.5	0.8
HH gross corrected*		3,349	3,298	671	954	723	2,774	6,157	5,401	23,327
	%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
HH-interview realised	abs.	2,382	2,680	464	608	517	753	1,510	1,321	10,235
	%	71.1	81.3	69.2	63.7	71.5	27.1	24.5	24.5	43.9
of this: HH willing to participate in panel	abs.						702	1,415	1,257	
	%						25.3	23.0	23.3	

* HH gross - neutral non-responses

Source: HH-Register; Scientific Use File IAB - for BA refreshment 4 and supplementary samples: methodological research dataset by infas

In a household survey, one can distinguish between the response rate at the household level and the response rate within households.

The response rate within households is used to denote the average proportion of all household members aged 15 or over within evaluable households for whom a complete personal interview is available.

On average, the following response rates are obtained within the interviewed households:

Table 5: **Average response rate within the interviewed households by waves and sub-samples**

		Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
		%	%	%	%	%
Sample	**BA**	85.6	85.5	83.1	88.4	88.7
	Microm	84.2	85.1	83.6	88.0	88.3
	BA-refreshment		86.2	84.3	90.2	89.5
	BA-refreshment			84.2	88.3	89.3
	BA-refreshment				89.6	91.2
	BA-refreshment					88.9
	EWO supplement					84.4
	BA supplement					90.0
	Gesamt	84.9	85.4	83.5	88.5	88.3

Source: P_Register; Scientific Use File IAB

In addition to the response rates at the household level and within the households, the following table shows the repeat interview rate at the individual level. This is the proportion of individuals willing to participate in the panel with whom an interview could be conducted in the subsequent wave.

Table 6: **Proportion of personal interviews in waves 2 to 5 with respondents from the previous wave willing to participate in the panel by subsamples**

Sample

			BA	Microm	BA-refreshment 1	BA-refreshment 2	BA-refreshment 3	Total
Wave 2	individuals willing to participate in the panel W1	abs.	8,925	8,938				17,863
	re-interviewed individuals in W2	abs.	4,274	5,829				10,103
	Share	%	*47.9*	*65.2*				*56.6*
Wave 3	individuals willing to participate in the panel W2	abs.	4,686	6,292	1,298			12,276
	re-interviewed individuals in W3	abs.	3,365	4,956	820			9,141
	Share	%	*71.8*	*78.8*	*63.2*			*74.5*
Wave 4*	individuals willing to participate in the panel W3	abs.	4,844	6,100	894	1,380		13,218
	re-interviewed individuals in W4	abs.	3,287	4,347	626	854		9,114
	Share	%	*67.9*	*71.3*	*70.0*	*61.9*		*69.0*
Wave 5	individuals willing to participate in the panel W4	abs.	3,946	5,004	785	979	993	11,707
	re-interviewed individuals in W5	abs.	2,972	4,151	570	714	702	9,109
	Share	%	*75.3*	*83.0*	*72.6*	*72.9*	*70.7*	*77.8*

Source: PENDDAT; Scientific Use File IAB

* Reduction of the gross sample due to objection procedures between Wave 3 and 4

2.3 Agreement to panel participation and merging of data, linking with process data

The respondents' consent is always required for storing addresses for the purpose of repeat interviews in the next wave and for merging the survey data with the process data of the Federal Employment Agency.

Agreement to panel participation was explained in detail in Chapter 2.1 within the scope of the sample size. The agreement to participate in the panel for first-time interviewed households[18] in a wave in PASS can be illustrated as follows:

Table 7: Agreement to panel participation of first-time interviewed households by waves

	Realised HH interviews with first-time interviewed HH***	Realised HH interviews with first-time interviewed HH willing to participate in the panel	Share willing to participate in the panel
	abs.	abs.	%
Wave 1	12,794	12,000	93.8
Wave 2	1,086	1,048	96.5
Wave 3	1,327	1,285	96.8
Wave 4*	903	866	95.9
Wave 5**	3,688	3,476	94.3

* Reduction of the gross sample due to objection procedures

** Expansion of the gross sample by supplementation

***first-time interviewed HH from refreshment, supplement and split

Source: PENDDAT and HH_Register; Scientific Use File IAB

The agreement to participate in the panel of first-time interviewed households in each wave is recorded following the first personal interview. The information given by this individual is then assumed for the household. If the individual agrees to participate in the panel, the household is considered willing to participate in the panel. If the individual does not agree to participate in the panel, the household is considered unwilling to participate in the panel (see also Chapter 2.1)[19].

---

[18] All households in wave 1 are first-time interviewed households. From wave 2 onwards, only the households from the refreshment samples and split-off households participating for the first time are counted as first-time interviewed households. Therefore, households interviewed for the first time have been the minority from wave 2 onwards – the majority of the household interviews conducted in these waves are interviews with households that were already interviewed at an earlier point in time.

[19] Hence, one individual provides the information regarding willingness to participate in the panel for the whole household. The information available on the household level was integrated in the individual dataset (*PENDDAT*) during data preparation. The individual respondents in the household assumed the corresponding information available for the household. The same procedure was applied in wave 2. In wave 1, however, the participation agreement was recorded after each individual and senior citizen's interview specifically for each individual –

.

In contrast to the agreement to participation, the permission to merge process data of the Federal Employment Agency with the survey data was obtained for each respondent who was interviewed using the personal questionnaire. This question does not apply to individuals aged 65 and over, because it is not included in the senior citizens' questionnaire. Agreement to merging of data is not obtained again in each new wave[20].

Table 8 provides an overview of the agreement to merging of data in the individual waves. Only those interviews are listed in which agreement to merging of data was requested in the respective wave as part of the personal questionnaire.

Table 8: Agreement to merging of data in personal interviews (15- to under 65-year-olds), in which the merging question was raised in the respective wave, by waves

	Realised personal interviews from the wave in which the merging question was posed	Realised personal interviews from the wave in which consent to merging was granted	Share with granted consent to merging
	abs.	abs.	%
Wave 1	17,249	13,766	79.8
Wave 2	3,358	2,560	76.2
Wave 3	2,656	2,128	80.1
Wave 4*	2,032	1,774	87.3
Wave 5**	5,145	4,414	85.8

* Reduction of the gross sample due to objection procedures

** Expansion of the gross sample by supplementation

Basis: individuals 15 to 64 years of age

Source: PENDDAT; Scientific Use File IAB

2.4 Split-off households

PASS is designed as a dynamic panel. Individuals who move into or are born into sample households are also interviewed as long as they are aged 15 or over. Individuals who move out of sample households or do not live in the household for one year or more

---

therefore varying data within a household are possible. Households with at least one individual willing to participate in the panel were considered willing to participate in the panel.

As part of updating address information after the first personal interview in re-interviewed households, it was explained that an interview would be conducted again in the following year. If the respondent did not explicitly object to this notification, the household was considered as still agreeing to participate in the panel, and the panel variable in the individual dataset (*PENDDAT*) was updated accordingly.

[20] Due to filtering modifications, there were cases in which the question regarding consent to merging of data was raised again in wave 2 and 3 if the respondent had not granted his/her agreement to this in the previous waves.

should continue to be interviewed, however. These individuals' new households are considered as split-offs from the original sample households. These split-off parts of the households (or split-off households) become sample households of PASS themselves. All of the individuals aged 15 or over living in these households become target persons for personal interviews. Should it occur in one of the subsequent waves that part of this split-off household in turn splits off, then this new split-off household, too, becomes a PASS sample household, irrespective of whether there is still anyone from one of the original samples living there ("infinite degree contagion model", Rendtel & Harms 2009, 267). Individuals who moved abroad, on the other hand, cease to be included in the survey as they no longer belong to the population and because the research questions specific to SGB II no longer apply. Individuals who do not live in the household for less than one year continue to be counted as household members and do not constitute a new PASS household.

There are a total of 477 split-off households from the interviews from waves 1 to 5, 283 of which could be interviewed in wave 5. Among them were 83 new split-off households from wave 5 and 21 first-time interviewed households which could already be identified in wave 4. Please refer to the methods report of wave 5 for further information on split-off households (Jesske & Schulz, 2012).

The interviewed split-off households can be identified in the datasets by comparing the current household number (*hnr*) with the original household number (*uhnr*), which differs in these cases. The original household number (*uhnr*) contains the household number of the panel household from which the new household has separated. Split-off households assume the sample indicator (*sample*), the information as to the sampling year (*jahrsamp*), the primary sampling unit (*psu*) and its stratification (*strpsu*) from their original household.

3   Dataset structure

The usual structure for editing a panel dataset, as used for example in surveys such as the German Socio-Economic Panel (GSOEP) or the British Household Panel Survey (BHPS), is to store information on individuals and households in annual individual datasets. If required, these can be supplemented with specific datasets, which might have a cross-wave data structure, for example for register or spell data.

This data structure makes it possible to store the information using relatively little storage space. Which variables were surveyed in which year can be identified immediately when looking into the datasets. The merging with additional information – via key variables, such as household or personal identification numbers – is also quite simple. However, this structure, which is usual for panel data, also has disadvantages which make it quite difficult to work with these datasets. If analyses are to be conducted not only in the cross-section but also in the longitudinal section, then first all of the relevant variables from the individual datasets of the respective waves have to be integrated into a common dataset, whereby care must be taken to ensure that the constructs selected really are the same with regard to contents. For typical longitudinal analyses the cross-wave dataset created in this way then has to be reshaped into so-called long format. In contrast to wide format, in which the data matrix contains precisely one row for each observation unit (e. g. a household or an individual), and then several datasets exist for each survey wave, in long format all of the waves allocated to one observation unit are arranged below one another. Instead of arranging the information in wave-specific variables in the same row, in long format the information is assigned to the same variable in each case in wave-specific rows of the observation units.

Preparing the data in long format has both advantages and disadvantages. The decisive advantage of this variant is that the data are already available in the structure required for many longitudinal analyses (such as event history analyses). It is no longer necessary to invest additional time and effort for creating a cross-wave file. The switch from long format to wide format is also quite easy to perform. STATA, for example, provides an option to switch between the two formats with little effort using the "reshape" command. Until a few years ago, the central argument against using this type of dataset structure was the significantly larger storage space required, which mainly results from the fact that even variables recorded in only one or a small number of survey waves always require a complete column across all waves in the dataset. In addition, the long files become quite large with increasing duration of the panel, simply as a result of all annual waves being appended to one another, which significantly increases the storage space required and the time to perform individual operations using the data. The wide availability of fast processors and large storage capacities even on simple desktop PCs makes this objection seem irrelevant in the meantime. Another disadvantage is the merging with additional information. Unlike the datasets prepared in wide format, an additional key variable is now required in order to be able to identify an observation clearly. This may be a wave identifier in the household or individual datasets, or alternatively the spell number in the spell datasets, which are also available in long format. Furthermore, it is not apparent at first sight which variables were surveyed for which waves, as all of the variables ever surveyed are pre-

sent in the dataset. These variables are given a special code (-9) for waves in which they were not surveyed.

When the advantages and disadvantages of long format for the user are weighed up, the advantages clearly outweigh the disadvantages in our opinion. Accordingly, the household and individual datasets of PASS (*HHENDDAT; PENDDAT*) and the corresponding weighting data (*hweights; pweights*) were prepared in long format.

At the household level, the scientific use file contains the data on the household's receipt of Unemployment Benefit II processed in spell form (*alg2_spells*). From wave 4 onwards, the individual level contains an integrated biographic spell dataset (bio_spells) which integrates and replaces the spell datasets et_spells, al_spells and lu_spells existing until wave 3. Furthermore, a one-euro spell dataset (ee_spells) was introduced in wave 4. The household and person registers (*hh_register; p_register*) are available in wide format. In wave 5, the scientific use file was extended at the individual level by one dataset for the vignette module (*VIGDAT*).

Figure 2:   Dataset structure of PASS in wave 5
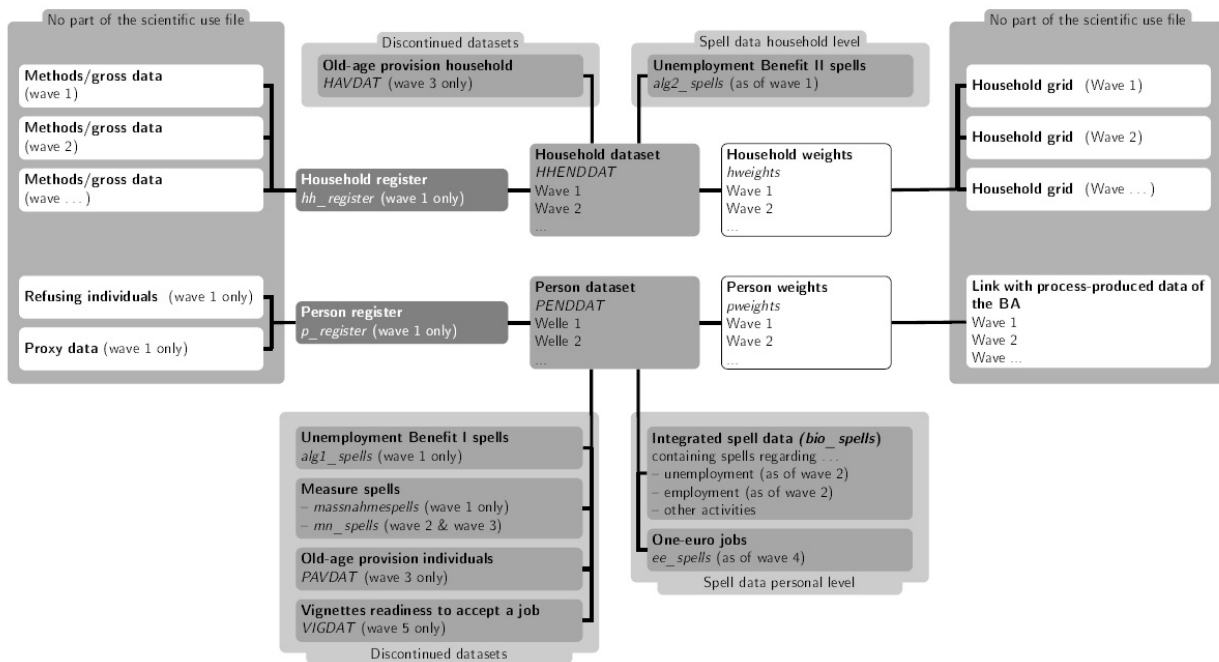
4 Generated variables

4.1 Coding of responses to open-ended survey questions

Some items of the survey were gathered as closed items with an open residual category or as open-ended items. In such cases, additional variables were usually generated[21] which differed from the original variable only insofar as the information from the open-ended responses was coded to the corresponding categories where possible. Moreover, in some cases new categories were created based on the information from open-ended questions. The name of these additional variables frequently differs from that of the original variable in the last digit only, where "0" was replaced by "1". The items on country of birth, nationality and the parents'/grandparents' country of residence before migration were also anonymised and given corresponding variable names[22]. Table 9 and table 10 give an overview of the open-ended survey questions which were coded in wave 5[23].

Table 9: Coding of responses to open-ended survey questions at the household level in wave 5

Regular variable name	Coded to variable	Dataset	Name
HD1100a-o	HD1101a-o	HHENDDAT	Employment status of HH members, proxy information, if necessary
HW0880a-i	HW0881a-j	HHENDDAT	Other reason for moving out, not listed
AL21300a-h – AL22100a-h	AL21301a-h AL21401a-h AL21501a-h AL21601a-h AL21701a-h AL21801a-h AL21851a-h AL21901a-h AL22001a-h AL22101a-h AL22102a-h AL22103a-h	alg2_spells	Other reason for benefit cut, not listed
AL22200a – AL22200h	AL22201a-h	alg2_spells	Other reason for discontinuation of receipt of UB II, not listed
AL20550a-h	AL20551a-h	alg2_spells	Other reason for why receipt of UB II started, not listed

---

[21] Other information from open-ended survey questions was not coded, for example the name of the institution providing basic social security (PTK0100).

[22] ogebland (country of birth); ostaatan (nationality); ozulanda to ozulandf (parents'/grandparents' country of residence before migration).

[23] Variables for which information was surveyed via open-ended questions and coded in the previous waves but not in the current wave are not listed (with the exception of the spell dataset for Unemployment Benefit II). For the observations in waves without obtaining information on these variables, these variables are allocated the code -9 (item not surveyed in wave) and are documented in the Datenreport of the survey wave.

Table 10: **Coding of responses to open-ended survey questions at the individual level in wave 5**

Regular variable name	Coded to variable	Dataset	Name
PB0230 (code 6)	PB0231	PENDDAT	Other German school qualification, not listed (update)
PB0230 (code 7)	PB0231	PENDDAT	Other foreign school qualification, not listed (update)
PB0400 (code 9)	PB0401	PENDDAT	Other German school qualification, not listed (first survey or not reported in previous wave)
PB0400 (code 10)	PB0401	PENDDAT	Other foreign school qualification, not listed (first survey or not reported in previous wave)
PB1000	PB1001	PENNDAT	Other foreign school qualification, not listed (first survey or not reported in previous wave)
PB1300a-j (code 9)	PB1301a-j	PENDDAT	Other German vocational qualification, not listed (update or first survey)
PB1300a-j (code 10)	PB1301a-j	PENDDAT	Other foreign vocational qualification, not listed (update or first survey)
PB1600	PB1601	PENDDAT	Other qualification to which the foreign qualification corresponds, not listed
AL0600	AL0601	bio_spells	Other reason for no longer being registered as unemployed, not listed
BIO0100	BIO0101	bio_spells	Other type of activity, not listed
EE0300a-h	EE0301a-h	ee_spells	Other reason for not participating in a one-euro job
EE1000a-e	EE1001a-e	ee_spells	Other reason why one-euro job was terminated prematurely
PTK0320a-g	PTK0321a-g	PENDDAT	Other reason for not having to seek employment, not listed
PEE0200a-d	PEE0201a-e	PENDDAT	Other source of information of one-euro jobs
PAS0900a-g	PAS0901a-g PAS0901i	PENDDAT	Other places where target pers. obtained information about job vacancies, not listed
PG0900a-f	PG0901a-g	PENDDAT	Other health problems, not listed
PG1300	PG1301	PENDDAT	Other health insurance, not listed
PP1300a-e	PP1301a-e	PENDDAT	Other private caretaking activities
PMI0200	ogebland	PENDDAT	Other country of birth, not listed
PMI0500	ostaatan	PENDDAT	Other nationality, not listed
PMI1000a-f	ozulanda-f	PENDDAT	Other country of birth, not listed Country from which parent/grandparent migrated

Table 10: Coding of responses to open-ended survey questions at the individual level in wave 5 (continued)

Regular variable name	Coded to variable	Dataset	Name
PA1100[24]	freiz1-3	PENDDAT	First to third leisure time activity
PA1200	frwunsch	PENDDAT	Desired leisure time activity
PA1300a-f	PA1301a-g	PENDDAT	Other reason for not pursuing the leisure time activity, not listed
PSH0200 (code 9)	PSH0201	PENDDAT	Other German school qualification of mother, not listed
PSH0200 (code 10)	PSH0201	PENDDAT	Other foreign school qualification of mother, not listed
PSH0300a-i (code 7)	PSH0301a-i	PENDDAT	Other German vocational qualification of mother, not listed
PSH0300a-i (code 8)	PSH0301a-i	PENDDAT	Other foreign vocational qualification of mother, not listed
PSH0500 (code 9)	PSH0501	PENDDAT	Other German school qualification of father, not listed
PSH0500 (code 10)	PSH0501	PENDDAT	Other foreign school qualification of father, not listed
PSH0600a-i (code 7)	PSH0601a-i	PENDDAT	Other German vocational qualification of father, not listed
PSH0600a-i (code 8)	PSH0601a-i	PENDDAT	Other foreign vocational qualification of father, not listed

4.2 Harmonisation

The survey instruments of some variables changed across the waves. In particular the integration of the employment biography module in wave 2 resulted in the fact that critical information on employment status, current main employment, the status of economic inactivity and the receipt of Unemployment Benefit I was surveyed in a different way than in wave 1. Since then, information has been collected not only with regard to the date of the interview but also in spell form for certain periods of time.

In order to facilitate cross-wave analyses in such cases, variables are generated for important indicators which are harmonised across the waves. Therefore, harmonisations are a special group within the generated variables (see Section 4.4) that are used to standardise differently collected indicators in retrospect.

Changes between the waves can affect the entire survey concept, categories and the interviewed groups. Harmonised variables thus consider different source variables that result from changed survey concepts, changes in categories and interviewed groups. This was an effort to standardise them as far as possible across the waves before generation was performed based on the variables.

---

[24] The variable PA1100 is not included in PENDDAT itself, since it does not include any additional information aside from the fact whether a target person has provided an open response or replied to the question with "don't know" or "details refused". Responses of "don't know" or "details refused" in PA1100 were included in the variables freiz1-3.

So far, the simple classification of occupational status (stibkz) has been harmonised. However, the number of necessary harmonisations can be expected to increase with the duration of the panel.

Table 11:   **Harmonised variables in the individual dataset (*PENDDAT*)**

Variable	Subject area	Name
stibkz	Employment	Current occupational status, simple classification, harmonised (anonymised)

While explicitly harmonised variables also consider changes in categories and interviewed group across the waves – besides changes in the survey concept – a second type of variables does not explicitly consider changes in the interviewed groups. These variables are generated for all waves, but they may contain information for different groups of respondents, depending on the wave. These differences result from revisions of the filtering process which were performed between the waves and affect the respective source variables of a generated variable.

Therefore, cross-wave variables of this type apply in addition to the actual harmonisations and standardise individual aspects between the waves. In contrast to the harmonised variables they are generated in each wave for all groups respectively, for which in that wave the corresponding source variables were collected. Hence, they can easily be used for evaluations in the cross-section of a specific wave. However, in the longitudinal section these differences must be considered before statements about changes between the waves can be made.

Therefore, it should be checked before working with the cross-wave but not harmonised variables, whether differences in the interviewed groups could cause problems for the respective evaluations and whether standardisation might be necessary[25].

---

[25]   For example, in wave 1 other groups of respondents were questioned on their employment than in the following waves. Accordingly, also the respective groups which provided information on occupational status, occupational activities, working hours, fixed-term employment, etc. varied.

Especially the subsequent cross-wave variables show differences regarding the groups for which they are generated:

Table 12: **Variables in the individual dataset (*PENDDAT*) which are generated across waves, but not completely harmonised**

Variable	Subject area	Name
isco88	Employment	ISCO 88 (ZUMA coding), current employment, gen.
kldb	Employment	Classification of occupations 1992, current employment
azhpt2	Employment	Current actual working hrs. main employment (without marginal employment, incl. cat. info.), gen.
azges2	Employment	Current total actual working hrs. (without marginal employment, incl. cat. info.), gen.
befrist	Employment	Current activity: limited contract? Generated (all waves)
mps	Employment	Magnitude Prestige Scale, current employment, gen.
siops	Employment	Standard International Occupational Prestige Scale, current employment, gen.
isei	Employment	International Socio-Economic Index, current employment, gen.
egp	Employment	Class scheme acc. to Erikson, Goldthorpe and Portocarrero (EGP), current occupation, gen.
esec	Employment	European Socio-economic Classification (ESeC), current occupation, gen.
stib	Employment	Occupational status, code number, current employment, gen.
netges	Employment	Current total net income (without marginal employment, incl. cat. info.), gen.
alg1abez	Benefit receipt	Current receipt of UB I, gen.
aktmassn	Participation in measures	Current participation in a programme funded/promoted by the employment agency, gen.

4.3 Dependent interviewing

In various places in both the household interviews and the personal interviews, information was gathered via dependent interviewing, i.e. depending on responses given in the previous wave. In this approach, data from the last interview was used for controlling the filter questions or it was integrated directly as part of the question text in the current interview.

There were mainly two goals that were pursued by utilising information from previous waves. Firstly, in some places only changes since the previous wave were to be recorded, partly depending on information on a certain set of questions already being available in the previous wave[26]. At these points, information from previous waves was used for controlling the filter. Secondly, the respondent should receive content information. In the places where changes since the previous wave were to be collected, the interview date of the previous wave was included in the question text to define the reporting period more clear-

---

[26] For example, individuals were only asked about their highest school qualification once. If they answered this question once, only new school qualifications obtained since the last interview are reported in the subsequent waves.

ly[27]. In other places, in particular where spell information was updated[28], also replies the respondent gave in the previous wave were integrated in the question texts. This was used to remind the respondent of his/her replies in the previous wave. This was to prevent that changes in status were reported which did not take place in reality but are an artefact of the open-ended survey arising from wrong memories or imprecise information.

If information from a single wave in the dataset is reviewed, only incomplete information is available for some respondents due to dependent interviewing, which only represents the changes between two survey dates. For respondents who are interviewed for the first time about a certain topic there might be information available which is complete regarding this wave[29].

In the course of data preparation, the recorded changes are combined with information from the previous wave to create variables and datasets with complete information as well. The spells in the existing spell datasets are updated with the newly recorded spell information. In the cross-section datasets (*HHENDDAT, PENDDAT*), however, generated variables are created in which the information from the previous wave is combined with the surveyed changes.

In the following, Table 13a and Table 13b provide a brief overview of all of the relevant places in the questionnaires and show in which variable the updated information can be found. The cases where generated variables were updated or continued are additionally listed in Chapter 4.4 of this Datenreport.

---

[27]   If, for example, only new school qualifications since the last interview were to be reported, the question was: "Have you obtained a general school qualification since our last interview on [display of interview date in previous wave]?"

[28]   Examples are updates of Unemployment Benefit II receipt from the previous wave in the household interview of the respective current wave or updates of employments or unemployments in the individual interview.

[29]   Individuals who were asked about their school qualification for the first time reported their respective highest school qualification. Therefore, complete information on the highest school qualification is available for this wave in the recorded variables. In the subsequent wave only newly obtained school qualifications are recorded. For example, if a school qualification was newly recorded, this information is available from the recorded variables, but it is not clear if this qualification is actually the highest school qualification. In this sense, the information of the subsequent wave is incomplete in the reported variables.

Table 13a: **Updated information from the previous wave in wave 5, household questionnaire**

Household questionnaire for re-interviewed households (HHalt)			
Construct	**Q. no.**	**Note**	**Update in variable**
Housing situation		Form of accommodation, type of tenancy and type of hostel/home/hall of residence updated during the interview	HHENDDAT: *HW0200 to HW0400*
Household structure		Household size updated during the interview	HHENDDAT: *HA0100*
		Sex of the individuals in the household corrected during the interview, if necessary	HHENDDAT: *HD0100a to HD0100o*
		Age of the individuals in the household updated during the interview	HHENDDAT: *HD0200a to HD0200o*
		Family relationships updated during the interview	not provided in the SUF
Size of dwelling in sqm	HW1000	Updated in generated variable	HHENDDAT: *wohnfl*
Receipt of Unemployment Benefit II	Module "Unemployment Benefit II"	Updated in Unemployment Benefit II spell dataset	alg2_spells: Variables of the Unemployment Benefit II spell dataset
		Information on the HH's current receipt of Unemployment Benefit II	HHENDDAT: alg2abez
		Information on the benefit units's Unemployment Benefit II receipt	p_register: bgbezs5; bgbezb5

Table 13b: Updated information from the previous wave in wave 5, personal question-naire

Personal questionnaire Construct	Q. no.	Note	Update in variable
Highest general school qualification	PB0220-PB1100	Updated in generated variable	PENDDAT: *schul1* (without responses to open-ended questions) *schul2* (with responses to open-ended questions)
Year in which highest school qual. was gained	PB0410	Updated in generated variable	PENDDAT: *schulabj*
Vocational qualification	PB1200-PB1600	Highest vocational qualification, updated in generated variable	PENDDAT: *beruf1* (without responses to open-ended questions) *beruf2* (with responses to open-ended questions)
Year of vocational qualification	PB1310	Updated in generated variable	*berabj*
Periods of updated activities in the BIO spell dataset	BIO0200, BIO0800, BIO0300	Updated in the BIO spell dataset for attached spells	bio_spells *BIO0400, BIO0500, BIO0600*
		Updated in the BIO spell dataset for attached spells Information on current employment, updated in generated variables	bio_spells: ET2300, ET2700 PENDDAT: *isco88; kldb; stib; stibkz; arbzeit; befrist; mps; siops; isei; egp; esec*
		Information on current economic inactivity/employment status, updated in generated variables	PENDDAT: *etakt; alakt; statakt*
Periods of receipt of Unemployment Benefit I in updated unemployment spells		Information on current receipt of Unemployment Benefit I	bio_spells: AL0700, AL0800, AL0900, *AL1000, AL1100, AL1200*
		Updated in the BIO spell dataset for attached spells	bio_spells: AL0600, AL0601 PENDDAT: *alg1abez*
Periods of updated activities in the EE spell dataset			*ee_spells: EE0800a, EE0800b*
Information regarding premature end in the EE spell dataset			*ee_spells: EE0900, EE1000a-EE1000e, EE1001a-EE1001e*

A distinction has to be drawn between these characteristics, where information collected in the past is updated with information on changes between the survey dates, and the so-called "constant characteristics". They are expected not to change over time. Therefore, these characteristics are recorded only once in PASS, although later corrections may be

possible in some cases. Since information on these characteristics is usually only available in the surveyed variables at the date of the first interview, they are afterwards provided in the form of generated variables (see Chapter 4.4, Bethmann & Gebhardt, 2011).

4.4 Simple generated variables

Simple generated variables cover, for example, variables for which different items of one construct that were surveyed separately for technical reasons were aggregated or for which information from the current wave was combined with information from the previous wave (see Chapter 4.3) (such as the highest educational qualification) or for which important information was merged from other partial datasets (e. g. indicators for current receipt of Unemployment Benefit I or Unemployment Benefit II).

The simple generated variables for households and individuals that are interviewed on a topic for the first time can always be generated on the basis of information surveyed in the current wave. For households and individuals that provided information on a topic in a previous wave, they can be differentiated in the cross-section datasets (*HHENDDAT; PENDDAT*) regarding the origin of the respective variables necessary for their generation. The three different types of simple generated variables are provided in Table 14.

Table 14: **Types of simple generated variables in the cross-section datasets (*HHENDDAT; PENDDAT*) for households and individuals that already provided information on the respective topic in a previous wave**

Type	Generation based on source data from		Description
	wave of the first survey of the topic for HH/individual	current wave	
unverän-derlich (uv)	yes	no	Information gathered in the first survey is generally adopted in the subsequent wave – unless input errors were corrected in the current wave. Example: *zpsex* (sex)
fort-geschrieben (fs)	yes	yes	Information that was current in the previous wave is combined with information of the current wave and updated, if necessary. Example: *schul1* (highest school qualification)
unabhängig neu (neu)	no	yes	The variable is newly generated from the data of the current wave in each wave, regardless of the information from the previous wave. Example: *hhincome* (net income of household)

More detailed explanations must be provided on the type "unveränderlich (uv)" regarding simple generation for *PENDDAT*. A first-time survey of a topic with an individual does not

always have to take place only in the first wave in which the individual gives a personal/senior citizens' interview. Two groups of individuals are again treated as first-time interviewed respondents even if they give a repeat personal/senior citizens' interview.

On the one hand, theses are individuals moving back into a household. Individuals moving from their previous household to a split-off household (see also Chapter 2.4) take their preload information with them. Thus, they can be treated correctly as first-time interviewed individuals or repeatedly interviewed individuals also in the split-off household. If an individual, however, moves back from a split-off household to a panel household he/she lived in at the time of a previous wave, the preload of this individual is not transferred from the split-off household to the original household. Individuals moving back in are thus treated like first-time interviewed individuals. This situation has been existing since wave 3, as in wave 2 the first move-outs of repeatedly interviewed households may occur and thus since wave 3 returns of individuals previously moved out may occur.

On the other hand, only an individual-related preload for dependent interviewing is created for an individual (see Chapter 4.3) if he/she gave a personal/senior citizens' interview in one of the two directly preceding waves. The background is that there shall be a distinction up to which point in time an individual should remember the results surveyed in spell form. The reference date for individuals who gave their personal/senior citizens' interview for the last time in the third preceding wave or earlier were before the relevant date for first-time interviewed respondents. In order to limit stress on the target person and assuming that the validity of the surveyed information is too severely threatened beyond this limit, individuals whose reference date for information on spell results is before the date relevant for first-time interviewed individuals are treated like first-time interviewed respondents.[30] This situation has been occurring since wave 4 as this is the first time that a previous personal interview may be more than two waves past.

For these two groups of individuals the information on which the "constant" generations are based are collected again (e. g. in the module "social origin") since these individuals are again treated as first-time interviewed individuals. Data preparation treats this surveyed information just like the information from individuals who are actually interviewed for the first time within the framework of PASS. These generations, e. g. for the status information of the mother and father, are thus based on the current wave. No transfer of information from the previous wave takes place and no data is made plausible with previous information. It can basically be assumed that the information of the target persons, which are processed to become "constant" generations, is consistent with previous information in case of a repeated survey. Inconsistencies and thus deviations as compared to information from the previous waves cannot, however, be generally excluded. Individuals included in one of the two groups described can be indentified in *PENDDAT* by them being

---

[30]   This excludes the information whether an individual has already been asked about his/her consent to merging of data in an earlier interview. This preload information is generated irrespective of the fact of how long a previous personal interview dates back. This is to avoid that individuals who gave their consent in a previous wave negate this question *RegP0100* in a subsequent wave and thus de facto withdraw their consent. The option for the target person to withdraw his/her consent to merging of data remains unaffected by this decision.

flagged in more than one wave with the code variable *altbefr* as first-time respondent (code "0" or code "-9" for wave 1).

The simple generated variables are shown in the dataset-specific Table 15 to 20. They include short descriptions of the individual variables. Furthermore, the source variables necessary for the generation of the variable in wave 5 are indicated[31]. For the cross-section datasets (*HHENDDAT; PENDDAT*) there is additional information on which type of simple generated variables shown in Table 16 they are (uv; fs; neu). This division does not make sense for spell datasets since there are no wave-specific observations. Instead, the generated variables are newly generated at spell level if the spell was newly included in the current wave or was updated with information surveyed in the current wave. Also register datasets follow a different logic so that no further differentiation was made here.

Table 15: **Simple generated variables for wave 5 in the household dataset (*HHENDDAT*) (in alphabetical order)**

Variable	Variable label and description	Source var. for generated var. in wave 5
alg2abez	*Current receipt of UB II of the HH, generated* Indicator for the household's current receipt of Unemployment Benefit II (neu)	*zensiert; AL20300; AL20400; AL20500* (*alg2_spells*); information on further receipts of Unemployment Benefit II (*AL22700*); *hintjahr* (*HHENDDAT*)
bik	*BIK region size classes (GKBIK10), generated* The information on region size class was generated by infas by converting the postcode available in the address data to *GKBIK10* (neu).	Supplied by survey institute
blneualt	*Western German States or Eastern German States, generated* Aggregation of German federal states into the Western German States of the former FRG (without Berlin) and the Eastern German States of the former GDR (with Berlin). Infas determined the federal states based on the postcodes available from the address data (neu).	Information generated and supplied by the survey institute on the federal state in which the household is resident at the survey date.

---

[31]   The respective Datenreport documents how the variables in the cross-section datasets (*HHENDDAT; PENDDAT*) were generated for observations in the previous waves. The documentation of the respective waves also describes the generation of wave-specific variables in the register datasets. The generated variables in the spell datasets were always generated in the already updated datasets. If a spell was not updated, the respective generated variables remained unchanged (if necessary with the exception that a special code was set in the censoring indicator if the spell could not be continued for technical reasons). If a spell was updated, then always the most current information was used, i.e. the variables filled with information from the current wave or the cross-section variables in the spells relevant for the current wave.

Table 16: Simple generated variables for wave 5 in the household dataset (*HHENDDAT*) (in alphabetical order)

Variable	Variable label and description	Source var. for generated var. in wave 5
hhinckat	*Categorised household income per month (in EUR), gen.* Categorised information on the household's income aggregated from several survey items into one variable (neu)	*HEK0700; HEK0800; HEK0900; HEK1000; HEK1100 (HHENDDAT)*
hhincome	*Household income per month (in EUR) incl. categorised information, gen.* Generation of a variable integrating information from categorised and open-ended survey questions on net household income (neu).	*HEK0600; HEK0700; HEK0800; HEK0900; HEK1000; HEK1100 (HHENDDAT)*
hintdat	*Date of household interview* Generated variable indicating the date on which the household interview was conducted in the format YYMMDD (neu)	*hintjahr; hintmon; hinttag (HHENDDAT)*
kindu4	*Control variable: child under the age of 4 in the HH* The variable indicates that at least one individual in the household is under the age of four in the wave. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this individual aged four is actually the child of another individual living in the household (neu).	*HD0200a - HD0200o (HHENDDAT)*
kindu13	*Control variablechild under the age of 13 in theHH* The variable indicates that at least one individual in the household is under the age of 13 in the wave. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this individual aged 13 is actually the child of another individual living in the household (neu).	*HD0200a - HD0200o (HHENDDAT)*
kindu15	*Control variable: child under the age of 15 in the HH* The variable indicates that at least one individual in the household is under the age of 15 in the wave. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this individual aged 15 is actually the child of another individual living in the household. If the response to the open-ended question on age was missing, the categorical follow-up question about the age groups was also used to generate the variable (neu).	*HD0200a - HD0200o*; categorical follow-up question about age group (in cases of no response in *HD0200*) (*HHENDDAT*)
wohnfl	*Living space in sqm, gen.* Information on the size of the living space in the household's current dwelling. In the case of re-interviewed households, the size of the living space was only asked as of the second wave if the household had moved house or if the house/apartment had changed since the previous wave (fs).	*For first survey:* *HW1000 (HHENDDAT)* For repeated survey: *wohnfl from previous wave; HW1000; (HHENDDAT)*

Table 17: Simple generated variables for wave 5 in the individual dataset (*PENDDAT*) (in alphabetical order)

Variable	Variable label and description	Source var. for generated var. in wave 5
akt1euro	*Current part. in one-euro job, generated* Indicator: respondent is participating in a one-euro job measure at the time of the interview (neu).	*zensiert (ee_spells)*
alakt	*Currently reported as unemployed, generated (as of wave 2)* Indicates that the TP was reported unemployed at the date of the personal interview of the respective wave (neu).	*zensiert; spintegr; BIO0101 (bio_spells)*
alg1abez	*Current receipt of UB I, generated* Indicator: respondent is in receipt of Unemployment Benefit I at the interview date. In wave 5, the periods since January 2009 during which the respondent was registered as unemployed were surveyed. For each spell additional questions were asked as to whether the respondent received UB I and if so, during which period (neu).	*AL0700; AL1000; AL1100; AL1200 (bio_spells)*
apartner	*Control variable: unmarried partner living in HH* Indicator: respondent has a cohabitee or a partner whose status is not specified in the household (neu).	Information on relationships between household members (household grid); *PD0500 - PD0900 (PENDDAT)*
azhpt1	*Current contractual working hrs. main employment (without marginal employment), gen.* Weekly contractual working hours in the main employment the respondent holds at the time of the interview, generated from open-ended questions on working hours (neu).	*ET2003 (bio_spells)*
azhpt2	*Current actual working hrs. main employment (without marginal employment, incl. cat. info.), gen.* Weekly actual working hours in the main employment held by the respondent at the interview date, generated from responses to open-ended questions on working hours and the categorical follow-up question in the case of irregular working hours (neu).	*ET2103; ET2203 (bio_spells)*
azges1	*Current total contractual working hrs. (without marginal employment), gen.* Weekly contractual working hours in all employments the respondent holds at the time of the interview, generated from open-ended questions on working hours (neu).	*ET2003 (bio_spells)*
azges2	*Current total actual working hrs. (without marginal employment, incl. cat. info.), gen.* Weekly actual working hours in all employments held by the respondent at the interview date, generated from responses to open-ended questions on working hours and the categorical follow-up question in the case of irregular working hours (neu).	*ET2103; ET2203 (bio_spells)*

Table 16: Simple generated variables for wave 5 in the individual dataset (PENDDAT) (in alphabetical order) (continued 1)

Variable	Variable label and description	Source var. for generated var. in wave 5
befrist	*Current employment: limited contract?* *Generated (all waves)* Indicator: the employment held by the respondent at the interview date is on a limited contract (neu).	*PET2510a; PET2510b* *(PENDDAT)*
begjeewt	*Start year of first employment, generated* Year in which the respondent first worked in a regular employment. To generate the variable, information about the first regular employment was combined with information from the employment spells if the respondent had already reported his/her first regular employment during the questions on employment spells since January 2009 (uv).	*For first survey:* *bjahr (bio_spells);* *PET3200b (PENDDAT)* *After first survey:* *begjeewt* from previous wave *(PENDDAT)*
begmeewt	*Start month of first employment, generated* Month in which the respondent first had a regular employment (generation: see *begjeewt*) (uv).	*For first survey:* *bmonat* (*bio_spells*); *PET3200a* (*PENDDAT*) *After first survey:* *begmeewt* from previous wave *(PENDDAT)*
berabj	*Year of the highest vocational qualification* Year in which the respondent gained his/her highest vocational qualification at the interview date (fs). Note: The years in which the vocational qualifications reported in wave 1 were achieved were surveyed in wave 2.	*For first survey:* *PB1310aj-kj* (*PENDDAT*) *For repeated survey:* *berabj* from previous wave; *PB1310aj-kj* (*PENDDAT*)
beruf1	*Highest vocational qual., excl. foreign qual and open info., generated* Identification of the highest vocational qualification at the interview date by hierarchising the vocational qualifications cited by the respondents, excl. information from open-ended questions (fs).	*For first survey:* *PB0100; PB0200; PB0300; PB1200b; PB1200c; PB1300a-j;* (*PENDDAT*) *For repeated survey:* *beruf1* from previous wave; *PB0100; PB0200; PB1200a; PB1300a-j* (*PENDDAT*)
beruf2	*Highest vocational qual., incl. foreign qual and open info., generated* As *beruf1* with the following differences: 1. Inclusion of responses to open-ended questions; 2. inclusion of information on foreign qualifications; 3. degrees are not distinguished by type of institution (e. g. university or other institution of higher education) but by the qualification level (Bachelor's degree; Master's degree; Ph.D.) (fs).	*For first survey:* *PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601* (*PENDDAT*) *For repeated survey:* *beruf2* from previous wave; *PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601* (*PENDDAT*)

brges	*Current total gross income (without marginal employment, incl. cat. info.), gen.*	ET2800; ET2900; ET3000; ET3100;
	Contains the cumulated information on gross income from all employments (>EUR 400). Generated from answers to open-ended questions on gross income and categorical follow-up question in case of "don't know" or "details refused" answers to open-ended questions (neu)	ET3200; ET3300 (bio_spells)

Table 16:   Simple generated variables for wave 5 in the individual dataset (PENDDAT) (in alphabetical order) (continued 2)

Variable	Variable label and description	Source var. for generated var. in wave 5
brutto	*Gross income from the current main employment incl. categorised information, generated* Generation of a variable integrating information from categorised and open-ended survey questions on gross income (neu).	ET2800; ET2900; ET3000; ET3100; ET3200; ET3300 (bio_spells)
bruttokat	*Categorised gross income from the current main employment, generated* Aggregation of the categorised information on gross income for a specific variable, combined from several items on income categories (neu).	ET2800; ET2900; ET3000; ET3100; ET3200; ET3300 (bio_spells)
ejhrlewt	*Time when last employment ended (year)* Last year in which the respondent was in employment. To generate this variable, information from the employment spells was combined with information on the last employment if the respondent had been out of work since January 2009 (fs).	<u>For first survey:</u> *PET1200b* (*PENDDAT*); *ejahr; emonat* (*bio_spells*) <u>For repeated survey:</u> *ejhrlewt from previous wave* (*PENDDAT*); *ejahr; emonat* (*bio_spells*)
ekin1517	*Control variable: own child aged between 15 and 17 in the household* This variable indicates that the respondent has a natural child, a stepchild/adopted child or a child of non-specified status aged between 15 and 17 in the household (neu).	Information on relationships between household members (household grid)
ekind	*Control variable: own child in HH* This variable indicates that the respondent has a natural child, a stepchild/adopted child or a child of non-specified status of any age in the household (neu). It can occur in rare household constellations that according to *ekind*, an individual has children living in the household, but their *pnr* does not appear in the pointers *zmhh* and *zvhh* of *p_register*. This can occur in case of same-sex relationships with children or if both the current and the former partner live in the household.	Information on relationships between household members (household grid)
ekin614	*Control variable: own child aged between 6 and 14 in the household*	Information on relationships between household members

	This variable indicates that the respondent has a natural child, a stepchild/adopted child or a child of non-specified status aged between 6 and 14 in the household (neu).	(household grid)
ekinu15	*Control variable: own child under the age of 15 in HH* This variable indicates that the respondent has a natural child, a stepchild/adopted child or a child of non-specified status under the age of 15 in the household (neu).	Information on relationships between household members (household grid)
ekinu18	*Control variable: own child under the age of 18 in HH* This variable indicates that the respondent has a natural child, a stepchild/adopted child or a child of non-specified status under the age of 18 in the household (neu).	Information on relationships between household members (household grid)

Table 16:   Simple generated variables for wave 5 in the individual dataset (PENDDAT) (in alphabetical order) (continued 3)

Variable	Variable label and description	Source var. for generated var. in wave 5
epartner	*Control variable: spouse or registered partner in HH* This variable indicates that the respondent has a spouse or a same-sex registered partner in the household (neu).	Information on relationships between household members (household grid)
etakt	*Currently employed (>EUR 400 per month), gen. (as of wave 2)* This variable indicates that the TP had an ongoing spell of employment at the time of the personal interview of the respective wave (i.e. employment earning > EUR 400) (neu).	*zensiert, spintegr, BIO0101 (bio_spells)*
famstand	*Marital status, gen.* Generation of a marital status variable integrating information from the personal questionnaire and the control variable *epartner* generated from the household dataset (neu).	*epartner; PD0500; PD0700 (PENDDAT)*
gebhalbj	*Half-year of birth, gen.* This variable indicates whether the date of birth is in the first or second half of the year of birth (neu).	Information on month of birth
kindzges	*Total number of own children (living in and outside the household), gen.* Total number of the respondent's children including the children living in his/her household and the children living outside the household (neu).	Information on relationships between household members (household grid); *PD0900; PD1000; PD1100 (PENDDAT)*
kindzihh	*Number of own children in the household, gen.* Variable generated on the basis of the responses in the household questionnaire concerning the number of children that an individual in the household has (total number of individuals in the household (half) matrix who count as children of the respondent plus the number of individuals in the household (half) matrix for whom the respondent is classified as being a parent) (neu). Note: *When using this variable it should be borne in mind that it relates to each individual person. This means that a child who lives in a household together with his/her parents is counted as a "child in the household" for both the father and the mother.* *Aggregating this variable across the household members will therefore not produce any meaningful results.*	Information on relationships between household members (household grid)

Table 16: Simple generated variables for wave 5 in the individual dataset (PENDDAT) (in alphabetical order) (continued 4)

Variable	Variable label and description	Source var. for generated var. in wave 5
mberuf1	*Highest vocational qualification attained by the mother, incl. mother in the HH, excl. information from open-ended survey questions, gen.* In wave 1, the question regarding the mother's vocational qualification was only asked if the mother was not living in the survey household. If she was living in the household, the information regarding her vocational qualification was taken from her personal interview. As of wave 2, the question regarding the mother's vocational qualification has been posed to all newly interviewed individuals, irrespective of whether the mother was living in the household or not. For people taking part in a repeat interview as of wave 2, the values were transferred from the generated variable *mberuf1* from the previous wave (uv).	For first survey: PSH0300a-i (*PENDDAT*) After first survey: *mberuf1* from previous wave (*PENDDAT*)
mberuf2	*Highest vocational qualification attained by the mother, incl. mother in the household, incl. information from open-ended survey questions, gen.* Same as *mberuf1*, apart from the fact that responses to open-ended questions were also taken into account for the generation of *mberuf2* (uv).	For first survey: PSH0301a-i (*PENDDAT*) After first survey: *mberuf2* from previous wave (*PENDDAT*)
mhh	*Control variable: mother living in HH* Variable indicating that the respondent's natural mother, stepmother, adoptive mother or mother of non-specified status is living in the household (neu).	Information on relationships between household members (household grid)
migration	*Respondent's migration background, generated* Generated variable for four categories of migration backgrounds: no migration background; personal migration (first generation); migration of at least one parent but no personal migration of the respondent (second generation); migration of at least one grandparent but no personal migration of respondent or of either parent (third generation) (uv). *Note: The concept for generating this variable has been revised as of wave 2. To generate the variable in earlier waves, only the information on whether the respondent was born in Germany and on which ancestor moved to Germany was used; now the information on whether an ancestor was born outside Germany and, if applicable, which ancestor, is also included. In order to guarantee a consistent logic across the waves, the variable for wave 1 was also regenerated.*	*For first survey:* PMI0100; PMI0700; PMI0800a-f; PMI0900a-f (*PENDDAT*) After first survey: *migration* from previous wave (*PENDDAT*)

Table 16: Simple generated variables for wave 5 in the individual dataset (PENDDAT)

Variable	Variable label and description	Source var. for generated var. in wave 5
mschul2	*Highest general school qualification attained by the mother, incl. mother in HH, incl. information from open-ended questions, gen.* Same as *mschul1*, apart from the fact that responses to open-ended questions were also taken into account for the generation of *mberuf2*. (uv).	For first survey: PSH0201 (*PENDDAT*) After first survey: *mschul2* from previous wave (*PENDDAT*)
mschul1	*Highest general school qualification attained by the mother, incl.* *mother in HH, excl. information from open-ended questions, gen.* In wave 1, the question on the mother's school qualification was only asked if the mother was not living in the survey household. If she was living in the household, the information on her highest school qualification was taken from her personal interview (uv). As of wave 2, the question on the mother's highest school qualification has been posed to all newly interviewed individuals, regardless of whether the mother was living in the survey household or not.	For first survey: PSH0200 (PENDDAT) After first survey: mschul1 from previous wave (PENDDAT)
mstib	*Mother's occupational status, code number, gen.* Detailed occupational status of mother, generated from the individual variables (uv).	For first survey: *PSH0320; PSH0330; PSH0340; PSH0360; PSH0370; PSH0380* (*PENDDAT*) After first survey: *mstib (PENDDAT)*
netges	*Current total net income (without marginal employment, incl. cat. info.), gen.* Contains the cumulated information on net income from all employments (>EUR 400). Generated from answers to open-ended questions on net income and the categorical follow-up question as of wave 2 in case of "don't know" or "details refused" answers to open-ended questions (neu).	ET3400; ET3500; ET3600; ET3700; ET3800; ET3900 (*bio_spells*)
netto	*Net income of the current main employment incl. categorised information, gen.* Generation of a variable integrating information from categorised and open-ended survey questions on net income (neu).	ET3400; ET3500; ET3600; ET3700; ET3800; ET3900 (*bio_spells*)
nettokat	*Categorised net income from the current main employment, gen.* Aggregation of the categorised information on net income for a specific variable, combined from several items on income categories (neu).	ET3400; ET3500; ET3600; ET3700; ET3800; ET3900 (*bio_spells*)
palter	*Age (from PD010), gen.* Respondent's age, generated based on the date of birth and the date of the personal interview in	*PD0100; pintjahr, pintmon, pinttag* (*PENDDAT*)

the current wave (neu).

panel *Willingness to participate in the panel* (neu)

Information supplied by the survey institute regarding the households' willingness to participate in the panel.

Table 16: Simple generated variables for wave 5 in the individual dataset (PENDDAT) (in alphabetical order) (continued 6)

Variable	Variable label and description	Source var. for generated var. in wave 5
pintdat	*Date of personal interview* Generated variable indicating the date on which the personal interview was conducted in the format YYMMDD (neu).	*pintjahr, pintmon, pinttag* (*PENDDAT*)
schul1	*Highest school qualification, excl. foreign qualifications and information from open-ended survey questions* Variable for the highest school qualification; equivalent eastern and western German qualifications were combined (e. g. EOS and Abitur); excl. information from open-ended questions (fs).	For first survey: *PB0200; PB0220; PB0230; PB0300; PB0400* (*PENDDAT*) For repeated survey: *schul1 from previous wave*; *PB0200; PB0220; PB0230; PB0300; PB0400* (*PENDDAT*)
schul2	*Highest school qualification, incl. foreign qualifications and information from open-ended survey questions* Like schul1 with the following differences: 1. inclusion of responses to open-ended questions; 2. inclusion of information on foreign qualifications (fs).	For first survey: *PB0200; PB0220; PB0231; PB0300; PB0401* (*PENDDAT*) For repeated survey: *schul2 from previous wave; PB0200; PB0220; PB0231; PB0300; PB0401* (*PENDDAT*)
schulabj	*Year in which highest school qual. was attained* Year in which the respondent attained his/her highest school qualification (fs). Note: Re-interviewed respondents for whom information regarding the highest school qualification was already available from a previous wave were not asked in the current wave about the year when this qualification was attained if they had attained a new qualification since the previous wave. In this case, the year in which the qualification was attained was estimated depending on the month and year of the interview. If the interview in wave 5 was conducted before May 2011, it was assumed that the qualification was gained in 2010, if the interview was conducted later than May, the qualification was assumed to have been gained in 2011.	For first survey: *PB0220; PB0230; PB0410; pintjahr; pintmon* (*PENDDAT*) For repeated survey: *schulabj* from previous wave; *PB0220; PB0230; PB0410; pintjahr; pintmon* (*PENDDAT*)
statakt	*Current main status, generated (as of wave 2)* Indicates which main status the TP had at the date of the personal interview of the respective wave (neu).	*zensiert; spintegr; BIO0101; az2ges* (*bio_spells*)
stib	*Occupational status, code number, generated* Generation of the detailed code number for occupational status from the individual variables.	*ET0603; ET0703; ET0803; ET0903; ET1003; ET1103; ET1203* (*bio_spells*)

Generation of the variable using information from the employment module (*ET0603-ET1203*). If there was more than one ongoing employment spell, the one with the
most hours of work was selected. If there was more than one ongoing spell with exactly the same amounts of hours, the one that started first was selected (neu).

Table 16: **Simple generated variables for wave 5 in the individual dataset (PENDDAT) (in alphabetical order) (continued 7)**

Variable	Variable label and description	Source var. for generated var. in wave 5
stibeewt	*Occupational status, first employment, code number, generated* Detailed code number of the occupational status in the respondent's first regular employment. To generate the variable, information regarding the first regular employment was combined with information from the employment spells if the respondent had already reported his/her first regular employment during the questions on employment spells since January 2009 (uv).	<u>For first survey:</u> *PET3300; PET3400; PET3500; PET3600; PET3700; PET3800; PET3900 (PENDDAT); ET0603; ET0703; ET0803; ET0903; ET1003; ET1103; ET1203 (bio_spells)* <u>After first survey:</u> *stibeewt* from previous wave (*PENDDAT*)
stibkz	*Current occupational status, simple classification, harmonised (anonymised)* Generation of the simple code number for occupational status from the individual variables (neu).	*PET1510* (*PENDDAT*)
stiblewt	*Occupational status, last employment, code number, generated* Detailed code number of the occupational status in the respondent's last employment. Information from the employment spells were combined with information on the last employment for the generation if the respondent has been unemployed since January 2009 (fs).	<u>For first survey:</u> *PET1210; PET1220; PET1230; PET1240; PET1250; PET1260; PET1270 (PENDDAT); ET0603; ET0703; ET0803; ET0903; ET1003; ET1103; ET1203 (bio_spells)* For repeated survey: *stiblewt* from previous wave *(PENDDAT); ET0603; ET0703; ET0803; ET0903; ET1003; ET1103; ET1203 (bio_spells)*
vberuf1	*Highest vocational qualification attained by the father, incl. father in the HH, excl. open info., gen.* Generation of variable for father's highest vocational qualification analogous to *mberuf1* (uv).	<u>For first survey:</u> *PSH0600a-i* (*PENDDAT*) <u>After first survey:</u> *vberuf1 from previous wave* (*PENDDAT*)
vberuf2	*Highest vocational qualification attained by the father, incl. father in the HH, incl. open info., gen.* Generation of variable for father's highest vocational qualification (incl. information from open-ended survey questions) analogous to *mberuf2* (uv).	<u>For first survey:</u> *PSH0601a-i* (*PENDDAT*) <u>After first survey:</u> *vberuf2 from previous wave* (*PENDDAT*)
vhh	*Control variable: father living in HH*	Information on relationships be-

	Variable indicating that the respondent's natural father, stepfather, adoptive father or father of non-specified status is living in the household (neu).	tween household members (household grid)
vschul1	Highest general school qualification attained by the father, incl. father in HH, excl. information from open-ended questions, gen. Generation of variable for father's highest general school qualification analogous to mschul1 (uv).	For first survey: PSH0500 (PENDDAT) After first survey: vschul1 from previous wave (PENDDAT)

Table 16: **Simple generated variables for wave 5 in the individual dataset (PENDDAT) (in alphabetical order) (continued 8)**

Variable	Variable label and description	Source var. for generated var. in wave 5
vschul2	Highest general school qualification attained by the father, incl. father in household, incl. open info., gen. Generation of variable for father's highest general school qualification (incl. information from open-ended survey questions) analogous to mschul2 (uv).	For first survey: PSH0501 (PENDDAT) After first survey: vschul2 from previous wave (PENDDAT)
vstib	Father's occupational status, code number, generated Detailed occupational status of father, generated from the individual variables (uv).	For first survey: PSH0620; PSH0630; PSH0640; PSH0660; PSH0670; PSH0680 (PENDDAT) After first survey: vstib from previous wave (PENDDAT)

Table 18: **Simple generated variables for wave 5 in the spell dataset for Unemployment Benefit II (alg2_spells) (in the same order as in the dataset)**

Variable	Variable label and description	Source var. for generated var. in wave 5
bmonat	Spell of UB II: start month, generated Month in which the spell of Unemployment Benefit II started. To generate the variable, if information was only available on the season when a spell started, it was converted into a definite month. Note: The generated date variables were checked for plausibility and corrected, if necessary. The dates originally reported by the respondent have been included in the source variables as of wave 2. Details regarding the season in which the spell started were recoded into month values as follows: 21 beginning of year/winter → January 24 spring/Easter → April 27 middle of year/summer → July 30 autumn → October	AL20100 (alg2_spells)

	32 end of year → December	
bjahr	*Spell of UB II: start year, generated* Year in which the spell of Unemployment Benefit II ended. Note: see *bmonat*	*AL20200 (alg2_spells)*
emonat	*Spell of UB II: end month, generated* Month in which the spell of UB II receipt ended. To generate the variable information on the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the household was interviewed) the interview month was entered. Note: see *bmonat*	*AL20300 (alg2_spells);* *hintmon (HHENDDAT)*
ejahr	*Spell of UB II: end year, generated* Year in which the spell of Unemployment Benefit II ended. In the case of right-censored spells (i.e. spells that were still ongoing when the household was interviewed) the interview year was entered. Note: see *bmonat*	*AL20400 (alg2_spells);* *hintjahr (HHENDDAT)*

Table 17: **Simple generated variables for wave 5 in the spell dataset for Unemployment Benefit II (alg2_spells) (in the same order as in the dataset) (continued 1)**

Variable	Variable label and description	Source var. for generated var. in wave 5
alg2kbma - alg2kbmh	*UB II: 1st cut: start month, generated* Month in which the reduction of Unemployment Benefit II started. To generate the variable information on the season was converted into a definite month. <u>Note:</u> The UB II cuts are embedded in the spells of UB II receipt. The information on the individual benefit cut spells can be distinguished via the indicator at the end of the respective variable (a - h). The generated date variables were checked for plausibility and corrected, if necessary. The dates originally reported by the respondent have been included in the source variables as of wave 2.	1st benefit cut: *AL21000a* (*alg2_spells*) to 8th benefit cut: *AL21000h* (*alg2_spells*)
alg2kbja- alg2kbjh	*UB II: 1st benefit cut: start year, generated* Year when Unemployment Benefit II cut started. <u>Note:</u> see *alg2kma - alg2kbmf*	1st benefit cut: *AL21100a* (*alg2_spells*)to 8th benefit cut:*AL21100h* (*alg2_spells*)
alg2kema - alg2kemh	*UB II: 1st benefit cut: end month, generated* Month in which the Unemployment Benefit II cut ended. To generate the variable information on the season was converted into a definite month. If the respondent reported a duration for the benefit cut, this was used to calculate the end date of the benefit cut based on the generated start date. <u>Note:</u> see *alg2kma - alg2kbmf*	1st benefit cut: *alg2kbma; alg2kbja; AL21200a; AL21201a; AL21202a (alg2_spells)* to 8th cut:*alg2kbmh; alg2kbjh; AL21200h; AL21201h; AL21202h (alg2_spells)*
alg2keja - alg2kejf	*UB II: 1st benefit cut: end year, generated* Year in which the Unemployment Benefit II cut ended. If the respondent reported a duration for the benefit cut, this was used to calculate the end date of the benefit cut based on the generated start date. <u>Note:</u> see *alg2kma - alg2kbmf*	1st benefit cut: *alg2kbma; alg2kbja; AL21200a; AL21201a; AL21202a (alg2_spells)* to 8th benefit cut: *alg2kbmh; alg2kbjh; AL21200f; AL21201f; AL21202f (alg2_spells)*

Table 17: **Simple generated variables for wave 5 in the spell dataset for Unemployment Benefit II (alg2_spells) (in the same order as in the dataset) (continued 2)**

Variable	Variable label and description	Source var. for generated var. in wave 5
AL22150a to AL22150h	*UB II: benefit cut: which HH member's benefit was cut, gen.* This variable contains coded information about which HH members' Unemployment Benefit II was cut. It is a string variable with 15 positions. Starting from the left, each position of this variable stands for the position of one individual in the household grid. The first position of the variable, for example, indicates whether Unemployment Benefit II was cut for the first individual in the household in the particular benefit cut spell, the second position indicates whether the second individual's benefit was cut and so on. As the source information for the generation was only collected from wave 2 to wave 4, all 15 positions of the question are given the code "I" (item not surveyed in wave) for all benefit cuts reported in the first wave and as of wave 5 (see below). Each of the 15 positions of the variable, which stands for one of a maximum of 15 individuals in the household structure, is given one of the following codes indicating the individual's benefit-cut status. Codes: 1 – the household member's UB II was cut 2 - the household member's UB II was not cut W – don't know K – not specified T – not applicable (filter) F – question mistakenly not asked U – implausible value I – item not recorded in wave.	Information which household member's benefit was cut in the respective benefit cut spell (only surveyed until wave 4).
zensiert	*Spell of UB II: spell ongoing at time of last HH interview (right-censored.), generated* The censoring indicator shows whether a spell was still ongoing at the time of the last household interview. Note: A spell is regarded as censored if one of the following conditions is met: (a) It is a censored spell of a household from one of the previous waves which had not been re-interviewed in the subsequent waves up to the current wave. (b) A household surveyed in wave 4 reports that a spell of UB II is still ongoing on the interview date in wave 5. Or an end date is reported which is identical with the interview date in wave 5 and it is confirmed in the follow-up question that the benefit receipt is still currently ongoing. Code -5 was given if the household reference person of the previous wave was no longer living in	*AL20300; AL20400, AL20500 (alg2_spells)*

the household in wave 5 and was not interviewed
in wave 5.

Table 18: Simple generated variables for wave 5 in the BIO spell dataset (bio_*spells*) (in the same order as in the dataset)

Variable	Variable label and description	Source var. for generated var. in wave 5
bmonat	*Employment: start month, generated* Month in which the employment spell started. To generate the variable information on the season was converted into a definite month. Note: The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent are included in the source variables. Details regarding the season in which the spell started were recoded into months as follows: 21 beginning of year/winter → January 24 spring/Easter → April 27 middle of year/summer → July 30 autumn → October 32 end of year → December	*BIO0200 (bio_spells)*
bjahr	*Employment: start year, generated* Year when the employment spell started Note: see *bmonat*	*BIO0300 (bio_spells)*
emonat	*Employment: end month, generated* Month in which the employment spell ended. To generate the variable information on the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the individual was interviewed) the interview month was entered. Note: see *bmonat*	*BIO0400, BIO0600 (bio_spells);* *pintmon (PENDDAT)*
ejahr	*Employment: end year, generated* Year in which the employment spell ended. For right-censored spells (i.e. spells that were still ongoing when the individual was interviewed) the interview month was entered. Note: see *bmonat*	*BIO0500, BIO0600 (bio_spells);* *pintjahr (PENDDAT)*
zensiert	*Employment: spell still currently ongoing (right censoring)* The censoring indicator shows whether a spell was still ongoing at the time of the personal interview in the previous wave, i.e. whether it is a right-censored spell. Note: A spell is regarded as censored if one of the two following conditions is met: The individual reports with regard to the end date of the BIO spell that the employment is still ongoing on the interview date. Or an end date is reported which is identical with the interview date and it is confirmed in the follow-up question that the activity is still currently ongoing.	*BIO0400; BIO0500; BIO0600 (bio_spells)*

Table 18: **Simple generated variables for wave 5 in the BIO spell dataset (bio_spells) (in the same order as in the dataset) (continued 1)**

Variable	Variable label and description	Source var. for generated var. in wave 5
stib	*Occupational status, code number, generated* Generation of the detailed code number for occupational status on from the individual variables.	*Collection of spell information in wave 5* *ET0603; ET0703; ET0803; ET0903; ET1003; ET1103; ET1203 (bio_spells)* Otherwise, the value from the previous wave remains
az1	*Weekly contractual working hours*	*Collection of spell information in wave 5* *ET2003 (bio_spells)* Otherwise, the value from the previous wave remains
az2	*Weekly working hours incl. details in the case of irregular working hours, gen.* Integrated variable on weekly hours of work in the employment held by the respondent, combining responses to open-ended questions on working hours and the categorical follow-up question. For the closed categories of the follow-up question the mean values for the categories were used, for the open-ended category (40 hours or more) the median of the weekly working hours reported in the open-ended questions was used.	*Collection of spell information in wave 5* *ET2103; ET2203 (bio_spells)* Otherwise, the value from the previous wave remains
alg1bm	*Receipt of UB I: start month, generated* Month in which the spell of Unemployment Benefit I receipt started. To generate the variable information on the season was converted into a definite month. Note: Periods of receipt of Unemployment Benefit I are embedded in the spells of registered unemployment. A maximum of one period of UB I receipt is available per period of registered unemployment. The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent are included in the source variables. Conversion of the month details, see *bmonat*.	*AL0800 (bio_spells)*
alg1bj	*Receipt of UB I: start year, generated* Year in which the spell of Unemployment Benefit I receipt started. Note: see *alg1bm*	*AL0900 (bio_spells)*

Table 18: **Simple generated variables for wave 5 in the BIO spell dataset (bio_spells) (in the same order as in the dataset) (continued 2)**

Variable	Variable label and description	Source var. for generated var. in wave 5
alg1em	*Receipt of UB I: end month, generated* Month in which the spell of Unemployment Benefit I receipt ended. To generate the variable information on the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the individual was interviewed) the interview date was entered. Note: see *alg2kma - alg2kbme*	*AL1000; AL1200 (bio_spells); pintmon (PENDDAT)*
alg1ej	*Receipt of UB I: end year, generated* Year in which the spell of Unemployment Benefit I receipt ended. In the case of right-censored spells (i.e. spells that were still ongoing when the individual was interviewed) the interview date was entered. Note: see *alg2kma - alg2kbme*	*AL1100; AL1200 (bio_spells); pintjahr (PENDDAT)*
alg1akt	*Receipt of UB I: spell still currently ongoing (right censoring)* The censoring indicator shows whether the spell of Unemployment Benefit I receipt was still ongoing at the time of the personal interview in the previous wave, i.e. whether it is a right-censored spell. Note: A spell is regarded as censored if one of the two following conditions is met: The individual reports with regard to the end date of the spell of Unemployment Benefit I receipt that the benefit receipt is still ongoing on the interview date. Or an end date is reported which is identical with the interview date and it is confirmed in the follow-up question that benefit receipt is still currently ongoing. The variable is generated based on the generated date variables, which are checked for plausibility.	*emonat; ejahr; AL1000; AL1100; AL1200 (bio_spells)*
br	*Gross income (incl. categorised info.), gen.*	*ET2800; ET2900; ET3000; ET3100; ET3200; ET3300 (bio_spells)*
net	*Net income (incl. categorised info.), gen.*	*ET3400; ET3500; ET3600; ET3700; ET3800; ET3900 (bio_spells)*

Table 19: **Simple generated variables for wave 5 in the one-euro spell dataset (ee_*spells*) (in the same order as in the dataset)**

Variable	Variable label and description	Source var. for generated var. in wave 5
bmonat	*Measure: start month, generated* Month in which the measure of active labour market policy spell started. To generate the variable information on the season was converted into a definite month. Note: The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent (apart from values identified as implausible when the range of values was checked) are included in the source variables. Details regarding the season in which the spell started were recoded into months values as follows: 21 beginning of year/winter → January 24 spring/Easter → April 27 middle of year/summer → July 30 autumn → October 32 end of year → December	*EE0600a (ee_spells)*
bjahr	*Measure: start year, generated* Year in which the measure of active labour market policy spell started. Note: see *bmonat*	*EE0600b (ee_spells)*
emonat	*Measure: end month, generated* Month in which the measure of active labour market policy ended. To generate the variable information on the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the individual was interviewed) the interview date was entered. Note: see *bmonat*	*EE0600a; EE0600b; EE0700; EE0800a; EE0800b (ee_spells); pintmon, pintjahr (PENDDAT)*
ejahr	*Measure: end year, generated* Year in which the measure of active labour market policy spell ended. For right-censored spells (i.e. spells that were still ongoing when the individual was interviewed) the interview date was entered. Note: see *bmonat*	*EE0600a; EE0600b; EE0700; EE0800a; EE0800b (ee_spells); pintjahr; pintjahr (PENDDAT)*
zensiert	*Measure: spell still currently ongoing (right censoring)* The censoring indicator shows whether a spell was still ongoing at the time of the personal interview in the previous wave, i.e. whether it is a right-censored spell.	*EE0700 (ee_spells)*

Table 20: Simple generated variables for wave 5 in the person register dataset (*p_register*) (in alphabetical order)

Variable	Variable label and description	Source var. for generated var. in wave 5
alter5	*Age of individual in wave 5 (2011)* Variable contains the "best" available information regarding an individual's age. This is either (a) the age calculated from the date of birth reported in wave 5 or (b) if no date of birth is available from wave 5, then the age reported in the household interview. The information from alter5 was also transferred to the household dataset and corresponds to the information in *HD0200a* to *HD0200o*. This procedure is consistent with that followed in the field. Already during the fieldwork, the age variable in the database was populated with the respective "best" information. During fieldwork, a variable in the database is first populated with the age information according to the household interview. If a personal interview is conducted, this variable in the database is overwritten with the age calculated based on the details given in the personal interview (date of birth, date of personal interview). Both the age information provided in the household dataset and the individual dataset are based on this variable of the database. The "best" age information included in the household dataset for wave 5 was considered in the plausibility checks and when generating the benefit unit and household types.	*PD0100; pintjahr; pintmon; pinttag (PENDDAT); HD0200a to HD0200o (HHENDDAT)*
erwprox5	*Employment status according to HH interview in wave 5 (2011)* Variable is an unchanged transfer of HD1101* from the current wave from HHENDDAT.	*HD1101**
korrsex	*Info. on sex was corrected between survey waves* For individuals who belonged to a sample HH in more than one wave this variable indicates whether the sex was corrected in the household interview.	*HD0100a* to *HD0100o* of all waves *(HHENDDAT)*
lastint	*Survey wave of last interview at individual level* This variable indicates the wave in which the last interview at the individual level was conducted with the person (personal interview or senior citizen's interview).	Personal interviews from all waves *(PENDDAT)*
neuj5	*Year in which individual joined current HH, reported in wave 5 (2011)* This variable indicates the year the individual joined the household of which he/she is a member in wave 5. Note: Information on the date comes from the wave 5 interview with the re-interviewed household into which the individual has moved or was born since the previous wave.	Information on the date since which an individual has belonged to a household. Surveyed in the household grid

Table 20: **Simple generated variables for wave 5 in the person register dataset (p_register) (in alphabetical order)**

Variable	Variable label and description	Source var. for generated var. in wave 5
neum5	*Month in which individual joined current HH, reported in wave 5 (2011)* This variable indicates the month the individual joined the household of which he/she is a member in wave 5. Note: see *neuj5*	Information on the date since which an individual has belonged to a household. Surveyed in the household grid
wegj5	*Year since which individual has no longer been living in previous HH, reported in wave 5 (2011)* This variable indicates the year the individual ceased to be a member of the household of the previous wave. Note: Information on the date comes from the wave 5 interview with the household in which the individual was living in the previous wave.	Information on the date since which an individual has ceased to belong to a household. Surveyed in the household grid
wegm5	*Month since which individual has no longer been living in previous HH, reported in wave 5 (2011)* This variable indicates the month the individual ceased to be a member of the household of the previous wave. Note: see *wegj4*	Information on the date since which an individual has ceased to belong to a household. Surveyed in the household grid
zdub5	*Pointer: Personal identification no. of the individual doubled by the TP in wave 5 (2011)* *Indicates that an individual from an original HH currently lives in a split-off HH without the original HH having reported the move of this individual.* Note: Chapter 5.4.1.2 provides a detailed explanation on the reasons for the introduction of this variable.	Information on all household members of an original household and all of its split-off households in the household grid of the current and the previous wave.
zmhh5	*Pointer: personal ID number of target person's mother in HH in wave 5 (2011)* Contains the personal identification number of the mother if she is living in the household. Natural mothers, stepmothers, adoptive or foster mothers, or mothers whose status is not specified are counted as the mother.	Information on relationships between household members (household grid)
zparthh5	*Pointer: personal ID number of target person's partner in HH in wave 5 (2011)* Contains the personal identification number of a partner living in the household. Spouses, registered partners, cohabitees and partners whose status is not specified are counted as a partner.	Information on relationships between household members (household grid)
zupanel	*Survey wave in which individual joined panel* This variable indicates the wave in which the individual was a member of a sample household for the first time.	Information on the individuals living in a household in all waves (household grid)

Table 20: Simple generated variables for wave 5 in the person register dataset (p_register) (in alphabetical order)

Variable	Variable label and description	Source var. for generated var. in wave 5
zvhh5	*Pointer: personal ID number of target person's father in HH in wave 5 (2011)* Contains the personal identification number of the father if he is living in the household. Natural fathers, stepfathers, adoptive or foster fathers, or fathers whose status is not specified are counted as the father.	Information on relationships between household members (household grid)

The datasets at the individual level contain a multitude of generated variables and constructed variables. These also include variables (e. g. for occupational status) that can be found in more than one dataset. Figure 3 provides an overview of the simple and complex generated variables at the individual level.

Figure 3: Overview of generated variables at the individual level in wave 5

	PENDDAT					BIO-Spells	EE_Spells
	Current status	Employment history		Social origin		Employment and unemployment biography	One-euro job participation
		Last employment	First employment	Mother	Father		
Education	berabj						
	beruf1			mberuf1	vberuf1		
	beruf2			mberuf2	vberuf2		
	schulabj						
	schul1			mschul1	vschul1		
	schul2			mschul2	vschul2		
Education classification	casmin			mcasmin	vcasmin		
	isced97			misced97	visced97		
	bilzeit			mbilzeit	vbilzeit		
Information on current status	akt1euro						
	alakt						
	etakt						
	statakt					spelltyp	
Socio-economic position	egp	egplewt	egpeewt	megp	vegp	egp	
	esec	eseclewt	eseceewt	mesec	vesec	esec	
	isei	iseilewt	iseieewt	misei	visei	isei	
	mps	mpslewt	mpseewt	mmps	vmps	mps	
	siops	siopslewt	siopseetw	msiops	vsiops	siops	
Occupational status	stip	stiblewt	stibeewt	mstib	vstib	stib	
	stibkz						
Date of employment			begmeewt			bmonat	bmonat
			begjeewt			bjahr	bjahr
		emonlewt				emonat	emonat
		ejhrlewt				ejahr	ejahr
Date of unemployment						alg1bm	
						alg1bj	
						alg1em	
						alg1ej	
Information on employment	befrist						
	azhpt1					az1	
	azhpt2					az2	
	azges1						
	azges2						
Occupation	isco88	iscolewt	iscoeewt	misco	visco	isco88	
	kldb	kldblewt	kldbeewt	mkldb	vkldb	kldb	
Employed in which industry	branche					branche	

Figure 3: Overview of generated variables at the individual level in wave 5 (continued)

		PENDDAT				BIO-Spells	EE_Spells
	Current status	Employment history		Social origin		Employment and unem-ploy-ment bi-ography	One-euro job participation
		Last employ-ment	First employ-ment	Mother	Father		
Income	netges						
	brges						
	netto						
	nettokat						
	brutto						
	bruttokat						
Benefit receipt	alg1abez					alg1akt	
	hhalg2						
Houshold context and civil status	hhgr						
	famstand						
	vhh						
	mhh						
	apartner						
	epartner						
	ekind						
	ekin614						
	ekinu15						
	ekinu18						
	ekin1517						
	kindzges						
	kindzihh						
Migration backround	ogebland						
	ostaatan						
	ozulanda						
	ozulandb						
	ozulandc						
	ozulandd						
	ozulande						
	ozulandf						
	migration						
Information on individual	gebhalbj						
	palter						
	zplathh						
	zpsex						
General	altbefr						
	fb_vers						
	panel						
	pintdat						
	RegP0100						
	sample						
Leisure time be-haviour	freiz1						
	freiz2						
	freiz3						
	frwunsch						

4.5 Constructed variables

Constructed variables are variables the generation of which requires more extensive re-coding and/or coding. In most cases, these variables have been empirically tested elsewhere and have a foundation in theoretical concepts. Moreover, at least some of them are standardised instruments used in social sciences or economics. Examples of such standardised instruments are the European Socio-economic Classification (ESeC), the International Standard Classification of Education (ISCED) or the equivalised household income. This chapter provides a detailed description of the constructed variables made available in the PASS data as well as a short overview of their theoretical background and the most important references.

Individual level

Education in years

<u>Variable name</u>	*bilzeit*
<u>Variable label</u>	Duration of school education and vocational training in years, generated
<u>Source variables</u>	*schul2; beruf2*
<u>Type / dataset</u>	Education / individual-level data
<u>Prepared by</u>	Bernhard Christoph
<u>Explanation</u>	For many statistical models, using a linear variable for education and training is more appropriate than using a categorical one. For school qualifications, it is fairly easy to convert the categorical information into linear information. The linear value simply corresponds to the time spent at school until attainment of the final school-leaving qualification. Care must be taken here, however, to ensure that equivalent qualifications are always allocated identical durations. An upper secondary school leaving certificate, for example, should always be labelled with the same duration, irrespective of whether it was attained after twelve or thirteen years of education. School-leaving qualifications were allocated the following education durations for this variable:

Lower secondary school leaving certificate;
lower secondary school leaving certificate from the former
GDR (POS) after completion of grade 8;
other Degree: 9 years

Intermediate secondary school leaving certificate;
intermediate secondary school leaving certificate from the former
GDR (POS) after completion of grade 10:   10 years

Entrance qualification for university of applied sciences: 12 years
General qualification for university entrance or subject-specific
higher education entrance qualification (incl. EOS –
similar qualification in the former GDR)   13 years

The situation is different for vocational qualifications. Due to the numerous different ways to gain a vocational qualification and the related potentially large differences in income even for qualifications with similar training durations, the training duration may not be subjected to a simple one-to-one conversion process. This problem can be avoided by attempting to operationalise the growth in human capital related to a certain vocational qualification (see e. g. Helberger, 1988).

This study uses a similar approach. For the conversion process, only the respondent's highest vocational qualification was considered and the years estimated to represent the human capital growth resulting from this qualification were added to the years of school education.

Training as a semi-skilled worker:	+1 year
Apprenticeship, vocational school, school for health care occupations:	+1.5 years
Master craftsman's certificate:	+3 years
Vocational academy:	+3 years
University of applied sciences/Bachelor's degree:	+3 years
University/Master's degree:	+5 years
PhD.:	+8 years
Other German qualification:	+1.5 years
Other foreign qualification:	+1.5 years

<u>Literature:</u>	Helberger (1988)

Education in years, mother

Variable name	*mbilzeit*
Variable label	Duration of school education and vocational training in years, generated
Source variables	*mschul2; mberuf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	General description: see "Education in years"

When generating the variable for the parents' years of education and training, the values added for vocational qualifications differ from those used when constructing the corresponding variable for the respondents, since information on vocational education/training was collected in less detail for the parents (especially as far as tertiary education is concerned). The values corresponding to particular courses of education/training are as follows:

Training as a semi-skilled worker:	+1 year
Apprenticeship, vocational school, school for health care occupations:	+1.5 years
Master craftsman's certificate:	+3 years
Vocational academy:	+3 years
University of applied sciences:	+3 years
University:	+5 years
Other German qualification:	+1.5 years
Other foreign qualification:	+1.5 years

Literature:	Helberger (1988)

Education in years, father

Variable name	*vbilzeit*
Variable label	Duration of school education and vocational training in years, generated
Source variables	*vschul2; vberuf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	General description: see "Education in years"

When generating the variable for the parents' years of education and training, the values added for vocational qualifications differ from those used when constructing the corresponding variable for the respondents, since information on vocational education/training was collected in less detail for the parents (especially as far as tertiary education is concerned). The values corresponding to particular courses of education/training are as follows:

Training as a semi-skilled worker:	+1 year
Apprenticeship, vocational school, school for health care occupations:	+1.5 years
Master craftsman's certificate:	+3 years
Vocational academy:	+3 years
University of applied sciences:	+3 years
University:	+5 years
Other German qualification:	+1.5 years
Other foreign qualification:	+1.5 years

Literature:	Helberger (1988)

CASMIN

Variable name	*casmin*
Variable label	Education classified acc. to CASMIN, updated version, generated
Source variables	*schul2; beruf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	The CASMIN educational classification was developed within the framework of the CASMIN project (Comparative Analysis of Social Mobility in Industrial Nations) in order to compare school and vocational qualifications on an international scale (König, Lüttinger & Müller,. 1987). An updated version is now available (Brauns & Steinmann, 1999).

The procedures for recoding qualifications acc. to CASMIN applied in the panel, especially for problematic cases, follow the procedures described in Lechert, Schroedter and Lüttinger (2006) and Granato (2000). For this, the slightly differing category values of the education variable in this dataset are of course taken into account. Details can be found in the table below. Cells containing valid combinations according to CASMIN are highlighted in light grey, those containing defined missing values are dark grey.

School / Occup.	Not surv.	Pupil	Not asked	NA	No details	Don't know	No qual.	Special needs school	Lower sec. school	Interm. Sec. school	Entrance qual. for uni. of app. Sci.	Upper sec. leaving cert.	Other Ger. qual.	Other foreign qual.
Not surv.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-
Implaus. value	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Pupil	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
Not asked	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
NA	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
No details	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
Don't know	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
No qual.	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Semi-skilled	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Apprenticeship	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Voc. school	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Health care school	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Master craftsman	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Vocational academy	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
UAS/ Bachelor's	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
Uni./Master's	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
PhD	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
Other Ger. qual.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Other foreign qual.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c

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

MCASMIN

Variable name	*mcasmin*
Variable label	Education of mother classified acc. to CASMIN, updated version, generated
Source variables	*mschul2; mberuf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	General description: see CASMIN

Since the education variable has different category values for respondents and their parents, the coding pattern of *mcasmin* and *vcasmin* differs slightly from the pattern used in *casmin*. The following table shows the differences in detail.

School / Occup.	Not surv.	Personal inter-view missing	Parent un-known	Not asked	NA	No de-tails	Don't know	No qual.	Special needs school	Lower sec. school	Interm. Sec. school	En-trance qual. for uni. of app. Sci.	Upper sec. leaving cert.	Other Ger. qual.	Other foreign qual.
Not surv.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Implaus. value	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Personal inter-view missing	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Parent unknown	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
Not asked	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
NA	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
No details	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
Don't know	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
No qual.	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Semi-skilled	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_voc	2c_voc	1b	1b
Apprentice-ship	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Master craftsman	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Vocational acad-emy	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
UAS	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
Uni.	-	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
Other Ger. qual.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Other foreign qual.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c

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

VCASMIN

Variable name		*vcasmin*
Variable label		Education of father classified acc. to CASMIN, updated version, generated
Source variables		*vschul2; vberuf2*
Category / dataset		Education / individual-level data
Prepared by		Bernhard Christoph
Explanation		General description: see CASMIN

Since the education variable has different category values for respondents and their parents, the coding pattern of *mcasmin* and *vcasmin* differs slightly from the pattern used in *casmin*. The following table shows the differences in detail.

Occup. \ School	Not surv.	Personal inter-view missing	Parent un-known	Not asked	NA	No de-tails	Don't know	No qual.	Special needs school	Lower sec. school	Interm. Sec. school	En-trance qual. for uni. of app. Sci.	Upper sec. leaving cert.	Other Ger. qual.	Other foreign qual.
Not surv.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Implaus. value	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Personal inter-view missing	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Parent unknown	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
Not asked	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
NA	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
No details	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
Don't know	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
No qual.	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Semi-skilled	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Apprenticeship	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Master craftsman	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Vocational acad-emy	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
UAS	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
Uni.	-	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
Other Ger. qual.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Other foreign qual.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c

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

ISCED 97

Variable name	*isced97*
Variable label	Education classified acc. to isced97, updated version, generated
Source variables	*schul2; beruf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	ISCED-97 (International Standard Classification of Education) developed by the OECD (OECD 1999, for an outline, see also BMBF, 2003) is an education classification which can be used as an alternative to CASMIN.

What must be taken into account regarding the coding of the ISCED-97 classification is that it includes categories which cannot reasonably be assigned to the present data. The ISCED values '0' (pre-primary education / kindergarten) and '1' (primary education) do not apply, because the respondents are at least 15 years of age. Instead, a separate group was generated for individuals with an education below ISCED level 2 (ISCED 2 = lower or intermediate secondary school leaving certificate). Therefore, only ISCED levels 2 to 6 are covered in the coding applied in this dataset.

Coding details are shown in the table below. Cells containing valid combinations according to ISCED are highlighted in light grey, those containing defined missing values are dark grey.

Occup. \ School	Not surv.	Pupil	Not asked	NA	No details	Don't know	No qual.	Special needs school	Lower sec. school	Interm. Sec. school	Entrance qual. for uni. of app. Sci.	Upper sec. leaving cert.	Other Ger. qual.	Other foreign qual.
Not surv.	-10													
Implaus. value					-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Pupil		-5												
Not asked			-4											
NA				-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
No details				-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
Don't know				-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
No qual.				-3	-2	-1	1	1	2	2	3a	3a	2	2
Semi-skilled				-3	-2	-1	2	2	2	2	3a	3a	2	2
Apprenticeship				-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Voc. school				-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Health care school				5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
Master craftsman				5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
Vocational academy				5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
UAS/Bachelor's				5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Uni./Master's				5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
PhD				6	6	6	6	6	6	6	6	6	6	6
Other Ger. qual.				-3	-2	-1	2	2	2	2	3a	3a	2	2
Other foreign qual.				-3	-2	-1	2	2	2	2	3a	3a	2	2

Literature:	BMBF (2003); OECD (1999)

MISCED 97

Variable name	*misced97*
Variable label	Education of mother classified acc. to isced97, updated version, generated
Source variables	*mschul2; mberuf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	For the theoretical background and generation details, see ISCED-97.

In contrast to the ISCED-97 coding applied to data on the respondents' education, it is not possible to generate ISCED level 6 for data on their parents. This is so, since data on the corresponding qualifications (i.e. PhD or equivalent) were not collected for the parents. Therefore, only ISCED levels 2 to 5 are covered in the coding applied in this dataset. The following table shows the coding details.

Occup. \ School	Not surv.	Personal interview missing	Parent unknown	Not asked	NA	No details	Don't know	No qual.	Special needs school	Lower sec. school	Interm. Sec. school	Entrance qual. for uni. of app. Sci.	Upper sec. leaving cert.	Other Ger. qual.	Other foreign qual.
Not surv.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Implaus. value	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Personal interview missing	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Parent unknown	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
Not asked	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
NA	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
No details	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
Don't know	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
No qual.	-	-	-	-	-3	-2	-1	1	1	2	2	3a	3a	2	2
Semi-skilled	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
Apprenticeship	-	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Master craftsman	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
Vocational academy	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
UAS	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Uni.	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Other Ger. qual.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
Other foreign qual.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2

Literature:	BMBF (2003); OECD (1999)

VISCED 97

<u>Variable name</u>	*visced97*
<u>Variable label</u>	Education of father classified acc. to isced97, updated version, generated
<u>Source variables</u>	*vschul2; vberuf2*
<u>Category / dataset</u>	Education / individual-level data
<u>Prepared by</u>	Bernhard Christoph
<u>Explanation</u>	For the theoretical background and generation details, see ISCED-97. In contrast to the ISCED-97 coding applied to data on the respondents' education, it is not possible to generate ISCED level 6 for data on their parents. This is so, since data on the corresponding qualifications (i.e. PhD or equivalent) were not collected for the parents. Therefore, only ISCED levels 2 to 5 are covered in the coding applied in this dataset. The following table shows the coding details.

School Occup.	Not surv.	Personal interview missing	Parent unknown	Not asked	NA	No details	Don't know	No qual.	Special needs school	Lower sec. school	Interm. Sec. school	Entrance qual. for uni. of app. Sci.	Upper sec. leaving cert.	Other Ger. qual.	Other foreign qual.
Not surv.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Implaus. value	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Personal interview missing	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Parent unknown	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
Not asked	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
NA	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
No details	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
Don't know	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
No qual.	-	-	-	-	-3	-2	-1	1	1	2	2	3a	3a	2	2
Semi-skilled	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
Apprenticeship	-	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Master craftsman	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
Vocational academy	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
UAS	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Uni.	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Other Ger. qual.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
Other foreign qual.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2

<u>Literature:</u>	BMBF (2003); OECD (1999)

International Standard Classification of Occupations 1988 (ISCO-88); ZUMA coding

Generated	Employment	Variable name	Source variables
	current	*isco88*	ET2500
	Spell data (*bio_spells*)	*isco88*	ET2500
	first	*iscoeewt*	ET2500, PET1280, PET3950
	last	*iscolewt*	ET2500, PET1280
	of father	*visco*	PSH0800
	of mother	*misco*	PSH0700
Variable label	Current empl.: ISCO-88 (ZUMA coding), generated		
	Spell data (*bio_spells*): ISCO-88 (ZUMA coding), generated		
	First empl.: ISCO-88 (ZUMA coding), first employment, generated		
	Last empl.: ISCO 88 (ZUMA coding), last employment, generated		
	Father: ISCO-88 (ZUMA coding) of the father, generated		
	Mother: ISCO-88 (ZUMA coding) of the mother, generated		
Category / dataset	Occupation / individual-level data		
Contact person	Bernhard Christoph		
Explanation	The International Standard Classification of Occupations (ISCO) was developed by the International Labour Organization (ILO), as an internationally comparative classification. The special feature of the ISCO-88 is that in addition to the employment performed, the qualification level generally necessary to perform the employment is taken into account when assigning an occupation to a particular occupational code. This constitutes a major difference to the Classification of Occupations provided by the German Federal Statistical Office (KldB), which is also provided in this dataset.		
Literature:	ILO (1990)		

Classification of Occupations 1992 (KldB92)

Generated	Employment	Variable name	Source variables
	current	*kldb_it*	ET2500
	Spell data (*bio_spells*)	*kldb*	ET2500
	first	*kldbeewt*	ET2500, PET1280, PET3950
	last	*kldblewt*	ET2500, PET1280
	of father	*vkldb*	PSH0800
	of mother	*mkldb*	PSH0700

Variable label	Current empl.: Classification of Occupations 1992, current employment
	Spell data (*bio_spells*): Classification of Occupations 1992, generated
	First empl.: Classification of Occup. 1992, first empl., gen.
	Last empl.: Classification of Occupations 1992, last empl., gen.
	Father: Classification of Occupations 1992 of father, generated
	Mother: Classification of Occupations 1992 of mother, generated
Category / dataset	Occupation / individual-level data
Contact person	Bernhard Christoph
Explanation	The KldB92 is the current version of the Classification of Occupations published by the German Federal Statistical Office (Statistisches Bundesamt). It is a classification system that was specifically constructed to match the particularities of the German occupational structure. It is based solely on employment.
Literature:	StBA (1992)

Class scheme according to Erikson, Goldthorpe and Portocarrero (EGP)

Generated	Employment	Variable name	Source variables
	current	*egp*	*isco88, stib*
	Spell data (*bio_spells*)	*egp*	*isco88, stib*
	first	*egpeewt*	*iscoeewt, stibeewt*
	last	*egplewt*	*iscolewt, stiblewt*
	of father	*vegp*	*visco, vstib*
	of mother	*megp*	*misco, mstib*

Variable label	Current empl.: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), current occupation, generated
	Spell data (*bio_spells*): Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), gen.
	First empl.: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), first employment, gen.
	Last empl.: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), last employment, gen.
	Father: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), occupation of father, gen.
	Mother: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), occupation of mother, gen.
Category / dataset	socio-economic position / individual-level data
Prepared by	Bernhard Christoph
Explanation	The class scheme developed by Erikson, Goldthorpe and Portocarrero (Erikson et al.,1979, 1982; Erikson & Goldthorpe, 1992) is one of the most common instruments for operationalising class position.
	For this variable, data are coded exclusively based on the ISCO-88 occupational classification and the occupational status. The coding procedure is based on an earlier approach elaborated by Christoph et al. (2005), where a detailed description of the procedure can be found. In contrast to the procedure described by Christoph et al., here unpaid family workers were not coded as self-employed but as individuals in dependent employment in accordance with the coding applied in the European Socio-Economic Classification (ESeC), which is described in the next section.
	One difference between the EGP codings applied here and the ESeC codings is that in the EGP coding procedure cases were set to "missing" (-7) where the occupational activity seemed to be incompatible with the occupational status (e. g. "directors and chief executives" [ISCO=1210] who reported that they were "employees performing simple duties" [StiB=51]). For reasons of compatibility with the strongly standardised coding procedure that we adopted, we did not apply a comparable revision procedure when using EseC codings.
Literature:	Christoph (2005); Erikson and Goldthorpe (1992); Erikson et al. (1982); Erikson et al. (1979):

European Socio-economic Classification (ESeC)

Generated	Employment	Variable name	Source variables
	current	*esec*	*isco88, stib, PET2000, PET2700*
	Spell data (*bio_spells*)	*esec*	*isco88, stib, ET1100, ET1101, ET1102, ET1300, ET1301, ET1302,*
	first	*eseceewt*	*iscoeewt, stibeewt, PET1261*
	last	*eseclewt*	*iscolewt, stiblewt, PET3801*
	of father	*vesec*	*visco, vstib, PSH0670*
	of mother	*mesec*	*misco, mstib, PSH0370*
Variable label	Current empl.: European Socio-economic Classification (ESeC), current occupation, gen.		
	Spell data (*bio_spells*): European Socio-economic Classification (ESeC), gen.		
	First empl.: European Socio-economic Classification (ESeC), first employment, gen.		
	Last empl.: European Socio-economic Classification (ESeC), last employment, gen.		
	Father: European Socio-economic Classification (ESeC), occupation of father, gen.		
	Mother: European Socio-economic Classification (ESeC), occupation of mother, gen.		
Category / dataset	socio-economic position / individual-level data		
Prepared by	Bernhard Christoph		
Explanation	With regard to its theoretical conception, the European Socio-economic Classification is largely based on the EGP class scheme. In contrast to the latter, however, great importance was attached to international comparability of operationalisation procedures and comprehensive validation of the classification scheme (for a general description, see: Rose & Harrison, 2007, and Müller et al. 2006, 2007 for Germany).		
	The Stata do-file required to generate the ESeC was kindly provided by Heike Wirth from GESIS-ZUMA (Fischer & Wirth 2007). We simply adjusted it to the requirements of this study. This do-file, originally written in standard SPSS syntax by Harrison and Rose (2006) as a standard program for the generation of the ESeC, was converted into Stata.		
Literature:	Fischer and Wirth (2007); Harrison Rose (2006); Müller et al. (2006, 2007); Rose and Harrison (2007)		

Magnitude Prestige Scale (MPS)

Generated	Employment	Variable name	Source variables
	current	*mps*	*isco88*
	Spell data (*bio_spells*)	*mps*	*isco88*
	first	*mpseewt*	*iscoeewt*
	last	*mpslewt*	*iscolewt*
	of father	*vmps*	*visco*
	of mother	*mmps*	*misco*
Variable label	Current empl.: Magnitude Prestige Scale, current occupation, gen.		
	Spell data (*bio_spells*): Magnitude Prestige Scale, generated		
	First empl.: Magnitude Prestige Scale, first employment, gen.		
	Last empl.: Magnitude Prestige Scale, last employment, gen.		
	Father: Magnitude Prestige Scale, occupation of father, gen.		
	Mother: Magnitude Prestige Scale, occupation of mother, gen.		
Category / dataset	socio-economic position / individual-level data		
Contact person	Bernhard Christoph		
Explanation	The Magnitude Prestige Scale [MPS] (Wegener, 1985, 1988) is the only specifically German instrument available so far to operationalise social prestige based on detailed occupation information. It was originally developed for the older 1968 version of the International Standard Classification of Occupations (ISCO-68). Since occupation coding in the study at hand was conducted based on the more recent ISCO-88 classification and the Classification of Occupations (KldB) developed by the Federal Statistical Office, a variant of the scale transferred to ISCO-88 was used (Christoph 2005). Infas merged the data as part of the occupational coding procedure.		
Literature:	Christoph (2005); Wegener (1985, 1988)		

Standard International Occupational Prestige Scale (SIOPS/Treiman Scale)

Generated	Employment	Variable name	Source variables
	current	*siops*	*isco88*
	Spell data (*bio_spells*)	*siops*	*isco88*
	first	*siopseewt*	*iscoeewt*
	last	*siopslewt*	*iscolewt*
	of father	*vsiops*	*visco*
	of mother	*msiops*	*misco*

Variable label	Current empl.: Standard International Occupational Prestige Scale, current occupation, gen.
	Spell data (*bio_spells*): Standard International Occupational Prestige Scale, generated
	First empl.: Standard International Occupational Prestige Scale, first employment, gen.
	Last empl.: Standard International Occupational Prestige Scale, last employment, gen.
	Father: Standard International Occupational Prestige Scale, occupation of father, gen.
	Mother: Standard International Occupational Prestige Scale, occupation of mother, gen.
Category / dataset	socio-economic position / individual-level data
Contact person	Bernhard Christoph
Explanation	The Treiman Prestige Scale, which was originally constructed by Treiman (1977) for ISCO-68, is the first and only prestige scale available so far which can be used for internationally comparative research into occupations. Since its adaptation to the ISCO-88 (Ganzeboom & Treiman, 1996, 2003), the scale has commonly been used under the name "Standard International Occupational Prestige Scale". Infas merged the data as part of the occupational coding procedure.
Literature:	Ganzeboom and Treiman (1996, 2003); Treiman (1977)

International Socio-Economic Index (ISEI)

Generated	Employment	Variable name	Source variables
	current	*isei*	*isco88*
	Spell data (*bio_spells*)	*isei*	*isco88*
	first	*iseieewt*	*iscoeewt*
	last	*iseilewt*	*iscolewt*
	of father	*visei*	*visco*
	of mother	*misei*	*misco*

Variable label	Current empl.: International Socio-Economic Index, current employment, gen.
	Spell data (*bio_spells*): International Socio-Economic Index, generated
	First empl.: International Socio-Economic Index, first employment, gen.
	Last empl.: International Socio-Economic Index, last employment, gen.
	Father: International Socio-Economic Index, occupation of father, gen.
	Mother: International Socio-Economic Index, occupation of mother, gen.
Category / dataset	socio-economic position / individual-level data
Contact person	Bernhard Christoph
Explanation	The International Socio-Economic Index is certainly one of the most common indices of its kind. This is due not least to the fact that, in contrast to most other SEIs, the ISEI is based on an original theoretical concept which sees the occupation and its socio-economic status as an "intervening variable" between education and income.
	Initially, the ISEI was developed for the ISCO-68 (Ganzeboom, De Graaf & Treimann, 1992) and was later adapted to the ISCO-88 (Ganzeboom & Treiman, 1996, 2003).
	Infas merged the data as part of the occupational coding procedure.
Literature:	Ganzeboom et al. (1992); Ganzeboom and Treiman (1996, 2003)

Classification of Economic Activities 2003 (Klassifikation der Wirtschaftszweige 2003 (WZ2003)

Generated	Employment	Variable name	Source variables
	current	*branche*	ET2600
	Spell data (*bio_spells*)	*branche*	ET2600

Variable label	Current empl.: Current activity: economic sector/industry (WZ2003)
	Spell data (*bio_spells*): economic sector/industry (WZ2003), generated
Category / dataset	socio-economic position / individual-level data
Contact person	Bernhard Christoph
Explanation	The information from the open-ended survey question about the sector / industry in which the respondent works was coded based on the 2-digit code in the Classification of Economic Activities of the Federal Statistical Office (WZ2003). At the two-digit level, this classification largely corresponds to the European "Nomenclature générale des Activités économiques dans les Communautés Européennes (NACE)" in revision 1.1.
Literature:	StaBA (2002); EG (2002)

Pursued and desired leisure time activities by young people

Variable name	*freiz1, freiz2, freiz3, frwunsch*

Variable label		freiz1: leisure time activity 1, pursued
		freiz2: leisure time activity 2, pursued
		freiz3: leisure time activity 3, pursued
		frwunsch: leisure time activity, desired
Source variables		*PA1100 (for freiz1-freiz3); PA1200 (for frwunsch)*
Category / dataset		leisure time / individual-level data
Prepared by		Johanna Eckert (DJI), Arne Bethmann, Claudia Wenzig
Explanation		Explanation:

The variables freiz1, freiz2, freiz3 and frwunsch are based on a newly developed scheme of categories regarding young people's leisure time activities. The scheme of categories' origin lies in the open-ended responses regarding the three most popular leisure time activities (PA1100) and the desired leisure time activity (PA1200).

The most popular leisure time activities were converted to a maximum of three individual variables according to the question text. The question regarding the desired leisure time activity considered only one reply according to the question text. Responses beyond that were not included in the coding.

The scheme was developed inductively based on the open, corrected information. In order to achieve comparability between the waves, the new scheme of categories also includes all leisure time activities which were asked in restricted questions in the previous waves. Furthermore, the scheme is designed in such a way that it can possibly be expanded in the next waves with new main and subcategories, if necessary.

The scheme of categories comprises a total of 16 main categories plus the categories "no leisure time activities" and "information cannot be assigned". The sequence of the 14 main categories with regards to content arises from the frequency of their mention. The main categories can be differentiated with the help of 77 subcategories.

	Main category / variable characteristic	Number of subcategories
1000	Sports and exercise	31
2000	Spending time with family and friends	4
3000	Computer, games and communication	5
4000	Making / listening to music	6
5000	Reading	-
6000	Culture, cinema, TV and events	8
7000	Creative hobbies, handicrafts, cooking and baking	11
8000	Going out, partying, nightlife	3
9000	Hanging out, relaxing	-
10000	Shopping	-
11000	Travelling, trips, making tours and being mobile	3
12000	Spending time with pets	-
13000	Voluntary work	4
14000	Learning and education	-
15000	Games and mental exercise	2
16000	Side job	-
99998	No leisure time activity	-
99999	Information cannot be assigned	-

Literature:	Johanna Eckert, Arne Bethmann, Claudia Wenzig (planned): Manual coding "Pursued and desired leisure time activities by young people". PASS wave 5 (2011).

Household or benefit unit level

Equivalised household income, old OECD weighting.

Variable name	*oecdinca*
Variable label	equivalised household income, old OECD weighting (rounded)
Source variables	*HD0200a-HD0200o; HA0100; hhincome*
Category / dataset	socio-economic position / household-level data
Prepared by	Bernhard Christoph
Explanation	With what is called the "equivalised household income", statisticians try to take into account the savings achievable by means of joint housekeeping in multi-individual households as compared to single households. To do this, the per-capita income in multi-individual households is not calculated based on the actual number of individuals living in the household, but by using a divisor which is usually below this figure and is calculated based on the assumed needs of the household members (equivalised household size).
	According to the old OECD scale, only the first household member (aged 15 or over) is assigned a weighting factor of 1.0. Further household members aged 15 or over are assigned a weighting factor of 0.7; children up to the age of 14 are counted with a weighting factor of 0.5 to calculate the equivalised household size.
	For more information on the old OECD scale, see OECD (1982); an overview of the topic is provided by Hauser (1996).
Literature:	Hauser (1996); OECD (1982)

Equivalised household income, modified OECD weighting

Variable name	*oecdincn*
Variable label	equivalised household income, modified OECD weighting (rounded)
Source variables	*HD0200a-HD0200o; HA0100; hhincome*
Category / dataset	socio-economic position / household-level data
Prepared by	Bernhard Christoph
Explanation	General description: see "Equivalised household income, old OECD weighting".
	The modified OECD equivalence scale assumes a weighting factor of 1.0 only for the first household member (aged 15 or over). Further household members aged 15 or over are assigned a weighting factor of 0.5; children up to the age of 14 are counted with a weighting factor of 0.3 to calculate the equivalised household size.
	For more information on the modified OECD scale, see Hagenaars, de Vos, and Zaidi (1994).
Literature:	Hagenaars et al. (1994)

Deprivation index, unweighted

<u>Variable name</u>	*depindug2*
<u>Variable label</u>	All waves: deprivation index, unweighted (item total: 23)
<u>Source variables</u>	*HLS0100a-HLS0400a; HLS0100b-HLS0400b; HLS0600a-HLS1200a; HLS0600b-HLS1200b; HLS1400a-HLS2500a; HLS1400b-HLS2500b;*
<u>Category / dataset</u>	material situation / household-level data
<u>Prepared by</u>	Bernhard Christoph
<u>Explanation</u>	Following a proposal by Ringen (1988), a distinction is usually made in poverty research between a direct and an indirect measurement of poverty. Indirect measurement focuses on the resources available to attain a certain standard of living, in particular the (equivalised household) income. For this reason this is also referred to as the resource-based approach to measuring poverty.
	In contrast, direct measurement attempts to record the households' actual ownership of goods and tries to determine the extent to which the households cannot afford certain goods or activities which are considered to be relevant, for financial reasons. This is also referred to as the deprivation approach (see e. g. Halleröd 1995).
	According to the general tenor of previous scientific research, the population classified as poor by the resource-based approach is not always identical to that defined by the deprivation approach. In order to define exactly who is to be considered poor in the narrow sense, it has therefore often been suggested to combine the measures of income-related poverty and deprivation and to count only those who are classified as poor by both approaches as belonging to the population living in poverty in the narrow sense (see Halleröd 1995; Nolan & Whelan 1996; Andreß & Lipsmeier 2001).
	The index is based on a list of 23 goods or activities. The households surveyed are asked to indicate whether they possessed these goods or participated in the activities mentioned. The unweighted index calculated on this basis simply adds up the number of items which the respondents indicated that they did not possess or did not participate in. However, only items which are missing for financial reasons are counted, in order to avoid certain consumer preferences (e. g. a household deliberately doing without a car or a television) being misinterpreted as a reduction in the standard of living.
	Additionally, an item was only accepted as missing for financial reasons if the answers to both questions explicitly confirmed this. "Don't know" or "details refused" answers were evaluated either as if the particular good was available in the household or as if it was missing for a reason other than financial reasons. This assumption does certainly not apply to all cases. Alternatively, it would have been possible not to calculate an index value for households that failed to answer a question for (at least) one particular good ("listwise deletion"). With respect to the total of 23 goods and activities surveyed, however, this method could quickly have led to a large number of missing index values. For this reason, the first method described was selected. Nevertheless, compared to the listwise deletion procedure, there is a risk of the number of goods missing for financial reasons being underestimated with this method.
	For waves 1 to 4 the variable *depindug* provides a version of the unweighted deprivation index which is based on 26 instead of 23 items, i. e. in addition to the items mentioned above also on HLS0500*, HLS1300* and HLS2600*. These three items have no longer been surveyed since wave 5. Thus, *depindug2* was newly integrated in the dataset and has been generated retroactively since wave 1.
<u>Literature:</u>	Andreß and Lipsmeier (2001); Halleröd (1995); Nolan and Whelan (1996); Ringen (1988)

Deprivation Index, weighted

<u>Variable name</u>	*depindg2*
<u>Variable label</u>	Deprivation index, weighted (items not missing for financial reasons; total of weighted items: 13,14)
<u>Source variables</u>	*HLS0100a-HLS0400a; HLS0100b-HLS0400b; HLS0600a-HLS1200a; HLS0600b-HLS1200b; HLS1400a-HLS2500a; HLS1400b-HLS2500b; PLS0100-PLS0400; PLS0600-PLS1200; PLS1400-PLS2500;*
<u>Category / dataset</u>	All waves: Deprivation Index, weighted (item total: 11.08)
<u>Prepared by</u>	Bernhard Christoph
<u>Explanation</u>	For a general description: see deprivation index, unweighted

<u>Explanation</u>

With respect to unweighted indices, such as the one described above, there is often criticism that all of the items included are given identical weightings. When comparing two items, for example the question as to whether the dwelling has an indoor toilet or the one as to whether there is a VCR / DVD player in the household, it immediately becomes clear that there is a vast difference in the extent to which a household's standard of living would be restrained by the lack of one of these items. It therefore seems reasonable to weight the individual items, even if empirical research has proven that in most cases weighted and unweighted index variants do not deliver significantly different results (see Lipsmeier, 1999).

For the present survey, we decided to weight items according to the proportion of respondents who regarded a particular item as necessary. We chose this procedure not only because it is convincing in conceptual terms and is a commonly used procedure (applied by Halleröd 1995, for example), but also because it could be implemented without unreasonable costs. As the deprivation weightings to be determined for the individual questionnaire items can be assumed highly stable over time, these items need only be administered once or at comparably long intervals. Moreover, thanks to the large population of the PASS sample, we were able to split the population into several randomly selected subsamples, each of which was presented with only some of the items.

Alternative weighting methods, such as restricting the indices to those items which are considered necessary by a certain minimum proportion of the respondents (e. g. Andreß & Lipsmeier 1995, Andreß et al. 1996) or a theoretical restriction to a few fundamental items (e. g. Nolan & Whelan 1996), were not applied in this survey, but can be generated, if necessary, based on the data provided. A discussion summarising the different methods of index weighting can be found in Andreß and Lipsmeier (2001, esp. p. 28 ff..).

For waves 1 to 4 the variable *depindg* provides a version of the weighted deprivation index which is based on 26 instead of 23 items, i. e. in addition to the items mentioned above also on HLS0500*, HLS1300* and HLS2600*, and PLS0500, PLS1300 and PLS2600. These three HLS items have no longer been surveyed since wave 5. Thus, *depindg2* was newly integrated in the dataset and has been generated retroactively since wave 1.

<u>Literature:</u>	Andreß and Lipsmeier (1995, 2001); Andreß et al. (1996); Halleröd (1995); Lipsmeier (1999); Nolan and Whelan (1996)

Household typology

Variable name	*hhtyp*
Variable label	Household type, generated
Source variables	Household information on age and relationships between household members
Category / dataset	Household structure / household data
Prepared by	Daniel Gebhardt
Explanation	A number of variants and suggestions exist regarding the definition of household types (see e. g. Lengerer, Bohr & Jansen, 2005 for the Micro-census household typology, Porst (1984) and Beckmann & Trometer 1991 for the ALLBUS typology and Frick, Göbel & Krause (n.d.) for the SOEP). The household typology used in PASS follows the latter typology. The decisive criteria of differentiation are existing partnerships, the number and age of children and existing generation relationships. Whereas the SOEP typology is merely based on the relationship of the household members to the head of the household, PASS uses information on interrelationships between all household members for the generation. In addition, the PASS typology includes the age of the household members as indicated in the household interview and the household size.

Definition of relationships for generating the household type:

- Couples: married couples; registered partnerships; non-married partnerships and partnerships whose status is not further specified (missing value for the follow-up question about the type of partnership).
- Child of an individual: natural child; stepchild; adopted or foster child; child whose status is not further specified (missing value for the follow-up question about type of relationship to the child).
- Parent of an individual: natural parent: step-parent; adoptive or foster parent: parent whose status is not further specified (missing value in follow-up question about type of parenthood).

Definition of household types:

- One-person household: Household consisting of only one individual.
- Couple without children: Household consists of two individuals living together as a couple.
- One-parent household: Household consists solely of one parent and his/her children. No restrictions are made with respect to the children's ages.
- Couple with children under the age of 16: Household consists solely of two individuals living as a couple and their respective and/or mutual children. All of the children are under the age of 16.
- Couple with children aged 16 or over: Household consists solely of two individuals living as a couple and their respective and/or mutual children. All of the children are aged 16 or over.
- Couple with children under the age of 16 and children aged 16 or over: Household consists solely of two individuals living as a couple and their respective and/or mutual children. There are both children under the age of 16 and children aged 16 or over living in the household.
- Multi-generation household: Household consists of members of at least three generations in linear succession. The core of the household is multi-generational, i.e. at least one individual in the household is both a child and a parent of another member of the household. The other people living in the household are parents, children, siblings, partners of the central member(s) and partners' siblings.
- Other household type: Household which could not be assigned to one of the other defined household types.
- Generation not possible (missing values): Basically, all households with at least one missing value (-1, -2, -4) or implausible value (-8) in the main category of a relationship variable or the age variable (Exception: for households with three or less members in unambiguous relationship constellations, the household type was also generated even if age details were missing).

Literature:	Beckmann and Trometer (1991); Frick et al. (n.d.); Lengerer et al. (2005); Porst (1984)

Benefit unit ID, wave 5

Variable name	*bgnr5*
Variable label	Benefit unit ID in wave 5
Source variables	Household information on age and relationships between household members
Category / dataset	Benefit unit / person register
Prepared by	Gerrit Müller
Explanation	The *bgnr5* variable is created at the individual level. It assigns an identification number to each household member indicating the individual's affiliation to a particular benefit unit. Consequently, household members with the same ID constitute a benefit unit together. The *bgnr5* variable is composed of the known household number and a two-digit indicator to identify the benefit unit within the household.
	The identification of a household member's affiliation to a benefit unit is based solely on the information on the relationships between the different household members from the household grid table as well as on the members' ages according to the household interview. The benefit units identified in this way are, therefore, to be regarded as "synthetic" benefit units. The identification process does not consider information on actual benefit receipt or on the individual members' ability to work and qualification status. It is more a case of identifying groups of individuals in the same household who are or would be regarded as benefit units in joint receipt of benefits according to the provisions of the German Social Code Book II in the event that they required benefits. This artificial allocation procedure is necessary, since information on the existence of a benefit unit and the identification of individuals affiliated to this unit cannot be collected directly in the context of an interview.
	With regard to content, the allocation of an individual to a benefit unit is based on the latest version of the German Social Code Book II, Section 7, Sub-section 3 (last amended on 26 March 2007). According to this, each individual who has reached the age of 25 and has not reached the age of 65 constitutes a separate benefit unit unless this individual is living in a partnership and/or has a child / children aged under 25 who has/have no own partner/children. In the latter case, the benefit unit comprises the individual, his/her partner and the child(ren). If two individuals live in the same household with a mutual child, but do not indicate in the household grid table that they are living in a partnership, a partnership is nevertheless assumed to exist in terms of Section 7, Sub-section (3a), and the corresponding individuals and their child(ren) are assigned to the same benefit unit. Individuals who have reached the age of 15 and who have not reached the age of 25 are generally assigned to their parents unless they are already living together with a partner (or a child of their own) in a joint household. Individuals aged between 15 and 25 who live without their parents (or partner / children) constitute a separate benefit unit.
	Individuals aged 65 and over are not covered by the German Social Code Book II and are therefore not counted as members of a benefit unit (code 0) unless they live together with a partner who is aged under 65 (or a child aged under 25) in the same household. Likewise, children who have not reached the age of 15 who live in a household without their parents are not counted as members of a benefit unit (code 0). They are covered by the provisions of the German Social Code Book XII. Allocations to benefit units were not made for households with missing information on relationships and/or the age of certain household members; instead, all members of these households were assigned code 99. By approximation, such households may be interpreted as households consisting of one benefit unit only.
Literature:	German Social Code Book II – basic security for job-seekers (Sozialgesetzbuch, Zweites Buch - Grundsicherung für Arbeitssuchende (SGB II))

Benefit unit typology, wave 5

Variable name	*bgtyp5*
Variable label	Type of benefit unit in wave 5
Source variables	Household information on age and relationships between household members
Category / dataset	Benefit unit / person register
Prepared by	Gerrit Müller
Explanation	The benefit unit typology is based on the same concept of the synthetic benefit unit as was used for variable *bgnr5*. Until reaching the age of 25, children are counted as members of the benefit unit of their parents unless they themselves have a partner or child of their own. This is handled differently from the BA statistics, where typologies are often still established based on majority (18th birthday). As an example: households in which the youngest child is aged between 18 and 24 and which are classified as one-parent benefit units according to our typology are counted as single households in the BA statistics. This difference must be borne in mind when comparing PASS data with figures from the official statistics.
	Code 0, no benefit unit, was assigned to households in which one or more member(s) were not covered by the Social Code Book II (see also code 0 for variable *bgnr5*). Code -5, generation impossible (missing values), was allocated to households with missing information on relationships and/or the age of individual household members (see code 99 for *bgnr5*).
Literature:	–

Benefit unit in receipt of Unemployment Benefit II on the sampling date, wave 5

Variable name	*bgbezs5*
Variable label	Benefit unit in receipt of UB II on the sampling date in wave 5
Source variables	*HA0250*, HA0300, AL20100, AL20200, AL20300, AL20400, AL20604, AL20704*, HA0400, sample, hnr, bgnr5, hhgr*
Category / dataset	Benefit unit / person register
Prepared by	Mark Trappmann
Explanation	For each benefit unit that was identified in accordance with the procedure described for variable *bgnr5* this variable indicates whether the benefit unit was in fact receiving Unemployment Benefit II on the sampling date of wave 5 or not.
Literature:	–

Benefit unit in receipt of Unemployment Benefit II on the survey date, wave 5

Variable name	*bgbezb5*
Variable label	Benefit unit in receipt of UB II on the survey date in wave 5 (2010)
Source variables	*AL20604, AL20704, zensiert (alg2_spells), sample, hhgr, bgnr5*
Category / dataset	Benefit unit / person register
Prepared by	Daniel Gebhardt
Explanation	For each benefit unit that was identified in accordance with the procedure described for variable *bgnr5* this variable indicates whether the benefit unit was in fact receiving Unemployment Benefit II on the survey date of wave 5 or not.
Literature:	–

Number of benefit units within the household

Variable name	*anzbg*
Variable label	Number of synthetic benefit units in the HH, generated
Source variables	*bgnr5, hnr*
Category / dataset	Benefit unit / household dataset
Prepared by	Daniel Gebhardt
Explanation	This variable indicates the number of benefit units existing in the household. The benefit units were identified in accordance with the procedure described for the generation of variable *bgnr5*.
Literature:	–

Number of benefit units in the household actually receiving benefits on the sampling date

Variable name	*nbgbezug*
Variable label	Number of benefit units in the HH receiving benefits on the sampling date
Source variables	*bgbezs5, bgnr5, hnr*
Category / dataset	Benefit unit / household dataset
Prepared by	Daniel Gebhardt
Explanation	This variable indicates the number of benefit units within the household which were in receipt of benefits in accordance with the Social Code Book II on the sampling date. The value was calculated via the household number by aggregating the benefit units within each household which were actually receiving benefits according to the variable *bgbezs5* from the person register.
Literature:	–

5   Data preparation

Since wave 3, not the IAB but infas has been responsible for preparing the data. In order to guarantee the consistency of data preparation in the longitudinal section, infas was provided with the relevant syntax files of the data preparation in wave 2 together with the necessary source and intermediary datasets and a documentation of the individual operations. Important decisions, such as on the correction of structural problems in the participating households or on the development of the *bio_spells* dataset, which was first developed in wave 4, were made together with the IAB. The IAB was also available for questions beyond that during the period of data preparation.

The information gathered in the interviews of wave 5 is initially available at infas in the form of ASCII data. In a first step, infas prepared the following datasets from these raw data[32]:

- Household dataset for questions surveyed in the cross-section

- Household dataset for data surveyed in the longitudinal section (module "Unemployment Benefit II")

- Dataset on the update of the household composition (matrix)

- Dataset on the update of the family relationships in the household (relationship matrix)

- Individual/senior citizens' dataset for questions surveyed in the cross-section in wave 5 including the questions from the vignette module which is later converted into spell format

- Individual dataset for data surveyed in the longitudinal section I (module "employment biography [spells]")

- Individual dataset for data surveyed in the longitudinal section II (module "measures")

- Dataset for open texts (across all household, personal and senior citizens' interviews)

A second step included more detailed, formal and content-related checks of the data, which were then prepared as the scientific use file. Furthermore, infas provides a gross dataset as well as other special datasets which do not derive directly from the actual survey instruments.

The data checks subsequently conducted at infas can be divided into three steps, which are described in more detail in the following sections. First, the household structure of the re-interviewed households was checked and corrected if necessary. If serious problems were found in the structure, the corresponding interviews were removed (see Chapter 5.1 on this issue). This was followed by a detailed check of the filter questions (applying corrections if necessary). On the one hand, filter errors were marked and on the other hand, specific codes were set for missing values (see Chapter 5.2 on this issue). After this, selected items were checked regarding plausibility of content. Clearly implausible or contradictory responses were marked as such by a specific missing code. Such corrections of the data were, however, carried out in a very restrictive way.

---

[32]   The software packages Stata version 11 and PASW version 18 were used for data preparation.

The following table provides an overview of all of the steps conducted in the context of the data preparation and their sequence:

Table 21:   Overview of the steps involved in preparing the data of wave 5 of PASS

No.	Step of the procedure
1	Import of the surveyed raw data in working datasets
2	Check of the household structure (see Chapter 5.1)
3	Removal of problematic interviews (household and/or individual level) (see Chapter 5.1)
4	Integration of individual dataset and senior citizens' dataset
5	Correction of the household structure of re-interviewed households (see Chapter 5.1)
6	Filter checks at the household level (see Chapter 5.2)
7	Construction of a household grid dataset and plausibility checks on this (see Chapter 5.3)
8	Generation of the synthetic benefit units (see description of variables, Chapter 4.5)
9	Generation of new control variables based on the household data after filter checks and the household grid dataset after plausibility checks
10	Filter checks at the individual level (see Chapter 5.2)
11	Coding of information from open-ended survey questions (see Chapter 4.1)
12	Plausibility checks of the household and individual-level data (excluding spell data) (see Chapter 5.3)
13	Preparation, plausibility checks and construction of the spell datasets (see Chapters 5.6 to 5.8 and Chapter 5.3)
14	Simple generations (see Chapter 4.4)
15	Complex generations (see Chapter 4.5)
16	Generation of the data structure for the scientific use file (household datasets, individual datasets, register datasets)
17	Anonymisation (see Chapter 5.5)

5.1   Structure checks and interviews removed from the dataset

A structure check was conducted before the filter checks were carried out. Here interviews which are regarded as not successfully surveyed in the sense of PASS were to be identified and were, if necessary, removed from the datasets for this reason. In addition, the structure of the re-interviewed households was compared with the structure reported in the previous wave in order to identify and, if necessary, correct implausible or problematic changes in the household composition and errors in the allocation of the personal interviews to their respective position in the household. For observing the households in the longitudinal section it is essential that the individuals are assigned consistently to their position in the household and that the respondents can be identified clearly across the waves.

 A definite personal identification number must not be allocated to different individuals in different waves. If the correct household composition was unclear, all of the interviews conducted with this household in wave 5 were removed from the dataset. If one of the personal interviews was conducted with the wrong individual but without any further problems emerging in the household composition, then just the personal interview was removed.

Different checks were carried out to identify problematic cases. The cases concerned were discussed in a formalised procedure between infas and the IAB. The final decision on how to proceed with these cases was made by the IAB. It should be considered that the following specifies the extent of the checks conducted. Not every check in every wave leads to the identification of problems. The result of a check is usually that a checked issue occurs in a low case number or not at all. Furthermore, known error sources are absorbed already during the interview. The survey instrument thus, for example, intends that not all known target persons can move out of a panel household at the same time and that among the individuals remaining after the moves at least one must be 15 years of age or older.

- By comparing the first names reported in the current and the previous wave, cases were identified in which changes in the household composition had not been recorded correctly. Instead of including moves into and out of the household in the relevant places in the household interview, it sometimes happened that interviewers renamed household members or changed their age or sex. All cases where a first name had been changed and this could not be put down to a correction of spelling and where the year of birth reported in the previous wave differed by more than one year from that reported in the current wave were subjected to individual case reviews. Here a decision was made as to whether the change in the data was simply a matter of correcting the first name, age or sex, or whether the interviewer had made an inadmissible change to the household structure.

- Furthermore, it was checked whether more than one individual with the same date of birth was living in the household. In the household context of the two waves, it was decided whether these were plausible or implausible cases. The remaining cases then underwent another check. For this, households were identified in which a date of birth was reported in the current and previous wave by individuals in different positions in the household structure. Here it seemed reasonable to suspect that a different individual from that in the previous wave conducted the particular personal interview in the current wave. In the context of the household and individual-level data of the current and previous wave, individual case decisions were made regarding the respective household and personal interviews.

- In general, the date of birth from the personal/senior citizens' interview of the current wave displaces all other age information on this individual, e. g. from the household grid, and is the basis for all generations which are among others based on age. In a special constellation, the date of birth is, however, corrected in *PD0100*. If the year of birth of an individual changes significantly according to *PD0100*, the day and month, however, stay the same, the hitherto known date of birth has never changed according to *PD0100* and at least two pieces of information on the date of birth from *PD0100* are available from previous waves, then the year of birth is reset to the value known from the previous waves considering the whole household constellation. A theoretical example is an individual whose date of birth is known as 01 February 1972 from at least two previous waves and whose date of birth is now recorded as 01 February 1992, which would make this individual younger than the children living in the household. Without a correction, such a constellation would lead to implausibility in the relation-

ship structure, which would consequently also lead to the fact that, for example, the synthetic benefit units cannot be generated. Hence, the information from the example is being corrected to the value 01 February 1972 in the current wave.

- In order to identify households which are regarded as not successfully surveyed in the sense of PASS, the datasets at the household and the individual level were merged. Personal interviews without a full household interview were marked, as were household interviews for which no interview at the individual level was available[33].

- Also moves into and out of the household are another important factor. Panel households with reported move-outs of the household were generally inspected regarding their household context and correlated with the realised split-off households. Evaluations were made as to whether the remaining household context of the panel household is self-evidently plausible. Interviews from panel households in which all household members leave the household, except for individual children under 15 years of age, were discarded with regard to the panel household as well as with regard to split-off households. If more than one individual moved out, it was checked whether these individuals formed a joint split-off household or several different ones, and whether this is plausible. Such cases were considered implausible, for instance, where one partner left the panel household together with young children, but the individuals moving out formed several different split-off households according to field information, i.e. the young children allegedly forming individual households. In case of the non-realisation of the split-off household, the move-outs were considered as plausible, but all individuals that moved out were retroactively merged into one joint split-off household.

- Individual cases occurred in which, according to the interview in the panel household, individual persons formed a split-off household, however, all members of the panel household could be found in the split-off household. In an alternative situation not all members of the panel household live in the split-off household, but at least one member of the panel household who, in the interview there, was not reported as having moved out or having moved to another split-off household than the one observed. Here, too, differentiated decisions were made as to which reported move-outs were considered valid and which were discarded as implausible. If a reported move-out was retroactively discarded as implausible, the individual that had allegedly moved out was retroactively re-integrated into the household context of the panel household.

- In split-off households it is verified whether individuals who are not known from the panel household but join PASS through the split-off household might still originate from the panel household. Two constellations promote these cases. On the one hand, it occurs that a panel household reports in case of several individuals moving out that the split-off individuals formed more than one split-off household. In this case, a dynamical preload is created for the current filed for all the split-off households known through the panel household. If, however, individuals who, according to the panel household, live in various split-off households are actually found in a shared split-off

---

[33]   In the case of new sample households for which a household interview but no valid personal interview was available, the household interviews were removed from the dataset following the procedure used in wave 1. In contrast, the household interviews of re-interviewed households and split-off households were retained.

household, those individuals who were not assigned to this split-off household by the panel household but to another split-off household do not have a preload in this split-off household and are included as new individuals.

- On the other hand, it is possible that individuals from a panel household move out of or into a household which was formed as split-off household in a previous wave and was already successfully surveyed back then. Thus, there is another move from the original panel household into this split-off household after the separation of the split-off household. Regardless of whether the panel household from which the respective split-off household emerged was successfully surveyed in the wave of the new move from the panel household to the split-off household, such cases cannot be controlled in the field. To do so, the split-off household would have to be provided with the personal information of all individuals from the panel household (and possibly all individuals in other split-off households of this panel household) as preload. The few cases in which such a constellation might occur do not justify efforts like that in the field. Instead, cases like this must be found in the structure checks. Please note in this context that regarding structure checks split-off households must be considered as split-off households also in the waves following their first successful survey even if they are considered panel households in field control after the first successful survey.

  In both cases the personal identification number of the respective individuals in the split-off household is corrected retrospectively. It must also be considered that these individuals are treated as new respondents in the personal/senior citizens' interview although they might have already participated. This deviation is generally not corrected (see also Chapter 4.4).

- In panel households that reported a move-out as of wave 2, there can also be moves back in of members formerly belonging to the household as of the wave 3. The requirement of recognising these individuals as moving back in and assigning them their former household position instead of assigning them a new household position is a component of the household grid. It was evaluated subsequently whether these requirements were met in the field in all cases. For individuals who were subsequently identified in the current wave as moving back in based on a comparison of first name, age and sex with the members who previously moved out of the households, the household structure had to be changed. This led to retroactive changes of the personal identification number of the individual to be positioned and also an adjustment in the individual-related information in the household interview, e. g. on childcare or the reasons for a cut in Unemployment Benefit II to the position defined as correct within the framework of the structural check. Conversely, it is also checked whether an individual who is marked in the field as moving back in really is the same individual who moved out in a previous wave. If not, this is a move-in of an individual who is new to PASS. The described changes in the household structure are also made in this case.

- In case of moves back in it is checked whether the split-off household in which the individual lived before he/she moved back into the panel household was successfully surveyed in the current wave and whether the split-off household considers the individual moving back in as having moved out. Also individuals who moved back into their panel household in a previous wave must continue to be checked regarding their

status in the split-off household as long as the split-off household is part of the current panel sample. If an individual who moves back in is still considered a current household member in his/her split-off household, a decision was made for these cases during data preparation as to whether this was plausible or whether the household structure of the panel or split-off household had to be corrected.

- Not only moves back can lead to individuals being considered as current household member of several households. It can also occur that an individual is considered a member of a split-off household although he/she was not recorded as having moved out of the panel household. Individual cases of this can be acknowledged as plausible after examination of the household structure of the respective households. Cases like that are documented in the *zdub** variables in the person register. For further explanations, please refer to Chapter 4.4 and Chapter 5.4.1.2.

- There can be other issues regarding the relationship of a panel household and its split-off households. There is a possibility that individuals who joined PASS via a split-off household move to the panel household. Another possibility is that individuals move from one split-off household to another split-off household. Generally, all individuals in a panel household and all split-off households connected to it must be considered a network. The structure checks are designed in such a way that individual moves between the households of such a network are detected regardless of the direction in which an individual moves in the network.

- Household structure checks generally do not evaluate the structure of the household in terms of plausibility but they consider the changes between the waves. Therefore, the household structure of households interviewed for the first time can only be checked to a limited extent. For households interviewed for the first time a check is made based on information concerning first name, age and sex as to whether individual household members are being listed multiple times. In this case, only the initially reported household position is maintained for the individuals reported twice, the other household positions are discarded. This might lead to other changes in the household structure. If, for example, in a household interviewed for the first time there are four individuals and the individuals on position 2 and 3 are identical, not only individual 3 is removed but also individual 4 is retroactively moved to position 3. As a rule, in a household interviewed for the first time with X household members, the positions 1 to X are to be filled without gaps. Just like for someone retroactively recognised as moving back in, a subsequent change in the personal identification number of the individual to be moved also requires moving the individual-related information in the household interview.

- Thanks to feedback by a field interviewer, a household was detected which was included twice in the panel sample in wave 4. Household 10015439 has been in the sample as identical household 15044862 since wave 1. Both households were successfully surveyed in wave 1 and wave 3 and not surveyed in wave 2. In wave 4, household 10015439 was successfully surveyed. This duplicate was detected since "both" households were given to the CAPI interviewer of this point. The household composition in the two households remained the same across all waves. Household 15044862, which was not surveyed in wave 4, will be deleted from the sample for wave 5. There will be no retroactive removal of the duplicate from waves 1 to 3 since this would affect weighting. The duplicate household is marked with code 26 in the hnettod4 variable in hh_register which makes the reason for non-surveying transparent. All household members of the duplicate household are marked with code 56 in the *pnettod4* variable in *p_register*.

- Individual case decisions were also made to deal with the cases which proved to be problematic during the structure checks. What was of significance here was how serious the particular problem was considered to be. In cases where the correct household composition in wave 5 was unclear, all of the interviews from wave 5 were removed. In wave 6 these households will be treated as households that did not participate in wave 5. If in retroactively removed household interviews moves-out were reported, also the split-off households were discarded. This concerned both the interviews conducted in the current wave in these split-off households and also the sample of the subsequent wave. Split-off households that developed from a discarded interview of a panel household are retroactively classified as not having been conducted and do not count to the panel sample of the subsequent wave. If there was merely a problem in assigning individuals to their respective position in the household, i.e. if it was suspected that a personal interview had been conducted with the wrong individual in wave 5, then only the respective personal or senior citizens' interview was removed. If it was a structural problem that had no serious consequences and could be solved, for example, by removing a personal interview, additional corrections of the first name, age and sex were made at the household level. The incorrect information concerned was then set back to the last valid value from the previous wave or in the case of age to the value from the previous wave + the number of years since the last valid interview in this household.

In addition, all interviews with individuals for whose household no complete household interview was available were removed. In the opposite case, i.e. households for which no individual-level interview was available, a distinction was made between re-interviewed households and households from the refreshment sample. The households from the refreshment sample which were regarded as not successfully surveyed were removed following the procedure used in the previous waves. In the case of re-interviewed households without interviews at the individual level, however, the household interview was not deleted.

The *Netto* variables (*hnettok5*, *hnettod5*, *pnettok5*, *pnettod5*) in the household and person register datasets indicate removed interviews. Via the corresponding variables in the household register it is possible to trace the re-interviewed households whose household

interviews were removed later. By means of net variables in the person register it is possible to trace the cases where only single individual-level interviews or all of the interviews of the household were deleted. In the case of households from the refreshment sample of wave 5 without at least one valid household and personal interview it is not possible to trace deleted interviews in the register datasets, as these households were not included in the datasets.

5.2 Filter checks

During the filter checks, the correct operation of the filter questions in the respective instruments was checked using a statistical program. If certain questions were asked although the value of the relevant filter variable would have required something else (for example, if detailed information was requested on vocational training although the respondent had stated that he/she did not have any vocational qualification), these variables were set to the missing code "-3" (not applicable), which they would also have received through correct use of the filters.[34] Moreover, some items were not surveyed in individual cases although this would have been necessary according to the relevant filter variable (e. g. if no further information was recorded on vocational training although the respondent had stated that he/she had undergone such training). In these cases, the specific missing code "-4" (question mistakenly not asked) was assigned. An assignment of the code "-4" can also be based on the household structure evaluation as described in Chapter 5.1. If the move-out of an individual is retroactively discarded as implausible and the individual is retroactively classified as still belonging to the former household, then this also means that individual-related information on these individuals in the household interview must be coded retroactively as mistakenly not surveyed. Thus, the code "-4" does not always refer to a problem in the survey instrument. If the code "-4" is assigned to a question that is relevant for filtering subsequent questions, then the subsequent questions are also coded with "-4" in case these subsequent questions were actually not surveyed. If subsequent questions were, however, surveyed, because, for instance, several filter questions linked to this subsequent question and another filter question triggered the subsequent question correctly, the value surveyed there remains.

In an additional step of the filter checks, the missing codes allocated by the field institute and the system missings were replaced by standard values for all variables. Table 22 provides an overview of the assigned values. "-1" and "-2" are the standard recoding for the values "don't know" and "details refused" recorded during the survey. "-3" is the general "not applicable" code for questions not asked due to filters. As described above, the code "-4" was assigned if a question was not asked as a result of a filter error. Codes "-5" to "-7" are question-specific codes. These can be either specific missing codes (e. g. "Not applicable, not available for the labour market"), or special categories for valid values (e. g. a category for an income above € 99,999 in the open question on income). These codes were only assigned as required.

---

[34] As is usual in such cases, the filter checks were conducted beginning with the items which were asked first and then moving on to those asked later.

Table 22:   Overview of the missing codes used

Code	Explanation
-1	"don't know"
-2	"details refused"
-3	"not applicable (filter)" (question not asked due to filter)
-4	"question mistakenly not asked" (question should, however, have been asked)
-5	question-specific code no. 1, only assigned as required
-6	question-specific code no. 2, only assigned as required
-7	question-specific code no. 3, only assigned as required
-8	"implausible value"
-9	"item not surveyed in wave"
-10	"item not surveyed in questionnaire version"[35]

The value "-8" is a specific missing code assigned during the plausibility checks (see Chapter 5.3 on plausibility checks). The missing code "-9" became necessary for the first time in wave 2. It is assigned if a certain item was not surveyed in a specific wave. Due to the dataset being prepared in long format, as was described above, variables that have no longer been surveyed in any version of the questionnaire as of wave 2 are given the value "-9" for the observations in this wave. Variables that were surveyed for the first time after wave 1 are retroactively coded "-9" for observations of waves in which they were not surveyed. Code "-10" can be used to consider differences between the questionnaire versions, in other words between the personal questionnaire and the senior citizens' questionnaire or between the two versions of the household questionnaire until wave 3.

5.3 Plausibility checks

For the plausibility checks an extensive list of theoretically possible contradictions in the respondents' statements was checked. For this, the list of checks conducted in the previous waves was adapted and extended for the current wave. Furthermore, also the household structure and the spell data were checked for plausibility – in particular with regard to inadmissible overlaps within the individual spell types. Generally, only the data gathered in the cross-section of wave 5 were checked here. No checks were carried out in the longitudinal section, in other words comparing the information provided in the current wave with that given in the previous wave.

In detail, the following steps were carried out:

1. Contradiction check: In general, contradictions were only corrected if either the implausibility could be defined as particularly serious and/or if the alteration was regard-

---

[35]   As of wave 4, code "-10" has only been used to differentiate between personal and senior citizens' questionnaires. Up to and including wave 3 there was an additional differentiation at the household level between first-time interviewed and repeatedly interviewed households. The differentiation at the household level is not continued in wave 4 due to the merger of the formerly separate questionnaire versions to one comprehensive household questionnaire.

ed as comparatively minor. The latter applied, for example, if only a small number of cases were affected or if one missing code (e. g. "-3") was simply replaced by another one (e. g. "-8"). Two strategies were used to filter implausible statements: either the implausible responses were corrected directly or they were allocated a specific missing code.

- Implausible responses were only corrected if it was highly probable that the interviewer had entered information incorrectly. An example of this is a statement of a monthly total rent of EUR 9,998.-. Here it was assumed in the plausibility check that the five-digit missing code "99998" (don't know) was entered incorrectly. This response and other similar responses were recoded to the corresponding missing categories. If the recoded missing categories triggered a filter in subsequent questions, as is the case for the categorical question of income, then the categorical questions were retroactively set to code "-4" (question mistakenly not asked).

- However, it was rarely the case that a value could be recognised as an incorrect entry with sufficient certainty. In most cases, it was only possible to establish a contradiction between two statements but not to identify specific incorrect entries or such that had led to the implausible statement. Therefore, in these cases no corrections were made and the specific missing value code "-8" was allocated instead. It was decided on an individual basis whether the code was allocated to one of the two variables involved in the contradiction or to both of them.

2. Plausibility check of the household structure: This check was carried out based on the information collected in the household interview on the family relationships between the household members, and the information on age, sex and first name. Prior to this check, the information on relationships in the household was supplemented by the information on partnerships reported in the personal interview.

- In order to identify implausible household structures, first the information on relationships was combined with the demographic information about the individual household members. For the households that were identified as implausible during these checks, individual case decisions were made which took into account the overall household structure and other information gathered during the interviews (e. g. on marital status in the personal interview). Implausible relationships were marked as such ("-8") or were corrected based on additional information on the household context if it was highly probable that an error had occurred. An example: In the case of two people of the same sex who were both natural parents of a third member of the household, the sex was corrected based on the first name. If the first names also indicated that the two people were of the same sex, and if there was no other relevant information available, then the relationship was marked as implausible based on the household structure.

- In a second step, checks were carried out comparing sets of three family relationships with one another for plausibility. An example of a relationship structure that would be classified as implausible in this check is: individual A is individual B's spouse. Individual A is the natural parent of individual C. Individual C is a sibling of individual B. If such a combination or another similarly implausible combination of relationships was identified during the plausibility checks, then here, too, an attempt was made to make the relationship plausible based on the household context. In the case described, the relationship data was corrected by individual C being coded as a child of individual B whose status was not further specified. The aim is to correct as many of the implausibilities identified as possible in terms of content, since a plausible and complete constellation of relationships is the necessary requirement for generating the benefit unit.

3. Also the spell datasets were subjected to a number of plausibility checks as described in detail in Chapters 5.6 to 5.8.

5.4 Retroactive changes of waves 1 to 4

5.4.1 Conceptional revisions

Conceptional adjustments were made to several generated variables in the course of the work on the SUF of wave 4. This is due to three different reasons.

On the one hand, changes in the survey logic had to be considered. Firstly, this concerns labour market policy measures in which the target persons participated. While waves 1 to 3 surveyed a comprehensive range of measures, the interest in results as of wave 4 is limited to one-euro jobs. Secondly, this concerns the concept to survey employments. The following shows how this presents over the waves:

- Wave 1: panel concept, i.e. only survey of latest available data

- Wave 2/wave 3: modular survey of ET/AL spells[36] + filling of gaps of > 3 months and of latest available data

- from wave 4 onwards: integrated survey of ET/AL/LU spells

On the other hand, conceptional flaws in the distinction of main and secondary employments for generated variables on income and working hours had to be corrected. Furthermore, decisions had to be made regarding the current survey concept in the person register as well as in *bio_spells*.

These revisions were already described in detail in Chapter 5.4 of the Datenreport for wave 4 of PASS (see Berg et.al, FDZ Datenreport 08/2011). Two subject areas will be covered again. On the one hand, this affects generated income variables. For the first time since wave 1, the variables *brutto(kat)* and *netto(kat)* can be generated again in *PENDDAT* in wave 5; in the *bio_spells* spell dataset, wave 5 provides the variables *br* and *net* for the first time. In order to clarify the function of the variables which are new or available again, we will include the applicable explanations from the wave 4 Datenreport here.

On the other hand, it is explained again how duplicate individuals are handled. Wave 4 included the first constellations where an individual lived in two households at the same time. Wave 5 was the first time that such duplicate individuals gave an interview themselves at the individual level. The conceptional considerations on the handling of duplicate individuals were further developed against this background. However, it was not necessary to adapt the SUF data of the previous waves since the conceptual adjustments concern the handling of interviews at the individual level. The following information regarding duplicate individuals thus replaces the respective chapter from the wave 4 Datenreport.

5.4.1.1 Income variables in PENDDAT and in BIO spells

The variables on current employment refer to the main employment in waves 1 to 4[37]. Excluded from that is information on gross/net income in waves 2 to 4 – this information re-

---

[36]   Here and in the following: ET = employment; AL = unemployment; LU = gaps (i.e. activities which are not ET or AL).

[37]   Wave 2 to wave 3; this is the censored ET in the ET spell dataset. In case of several censored spells, the spell with the highest amount of hours was selected. In case of several spells with

fers to all currently ongoing employments > EUR 400 (imprecision regarding marginal employment wages). Spell-specific information is not available and will only be surveyed as of wave 5. The information is only surveyed as total value across all employments. This leads to two partial problems:

I.   The generated variables on working hours and gross/net wage have referred to different employments (main ET or all ETs) as of wave 2. If hourly wages are calculated on this basis, this leads to errors for TPs with several ET.

II.   The different earnings cannot be recognised from the variable labels.

The generated variables on income and working hours will thus be revised accordingly in wave 4.

The survey concept of income variables changed significantly between wave 1 and 2 without this leading to the formation of new variables: *brutto* (*bruttokat*) and *netto* (*nettokat*) reflect the income from the main employment in wave 1; as of wave 2, the income from all employments which are not marginal. This is inconsistent and potentially leads to errors in the evaluation. The revision is to correct this problem:

Table 23:   Revision of income variables[38]

Variable	Content	Dataset	Generated for					Basis	
			W1	W2	W3	W4	W5	OFFA	KatA
bruttokat	Main ET, gross	PENDDAT	1	0	0	**0**	1	0	1
brutto	Main ET, gross	PENDDAT	1	0	0	**0**	1	1	1
nettokat	Main ET, net	PENDDAT	1	0	0	**0**	1	0	1
netto	Main ET, net	PENDDAT	1	0	0	**0**	1	1	1
brges	Total ET, gross	PENDDAT	0	1	1	**1**	1	1	1
netges	Total ET, net	PENDDAT	1	1	1	**1**	1	1	1
br	Spell ET, gross	BIO spells	0	0	0	**0**	1	1	1
net	Spell ET, net	BIO spells	0	0	0	**0**	1	1	1

Revised variables (in waves 1 to 3 already in the dataset)

bruttokat *(current gross income main empl. (without marginal employment, categorised), gen.*
brutto *(current gross income main empl. (without marginal employment, incl. cat. info.), generated)*
nettokat *(current net income main empl. (without marginal employment, categorised), generated)*
netto *(current net income main empl. (without marginal employment, incl. cat. info.), generated)*

---

        the same amount of hours the longest lasting spell was selected. Only one employment was surveyed for senior citizens.

[38]   In wave 1, there is only a categorical follow-up question for the main employment's net wage but not for other activities. This is accepted when generating *netges*. If the information (MV) on net income from other activities is missing, the variable *netges* cannot be generated.

These variables refer to the respective main ET in wave 1. As of wave 2, they have been, however, filled with the cumulated information for all ETs (>EUR 400) since only this information was surveyed. The variable labels have been adjusted respectively as of wave 4. For waves 2 to 4, the variables were filled with -9 since a generation analogous to wave 1 is not possible.

New variables in W4

brges *(current total gross income (without marginal employment, incl. cat. info.), gen.)*

This variable contains the cumulated information on gross income from all ET (>EUR 400). This variable cannot be generated in this form for wave 1 since only the gross income for the main ET was surveyed. For waves 2 and 3, the variable is identical in terms of content with the *brutto* variable, which was included in the SUF of wave 3 (i. e. prior to the revision as explained above). In waves 2 to 4, only the cumulated gross income was surveyed – the source variables used in wave 2/wave 3 thus already include the respective information on total income from ET > EUR 400. The variable for wave 4 shall be generated analogous to wave 2/wave 3. As of wave 5, it will be generated based on spell-specific income information.

netges *(current total net income (without marginal employment, incl. cat. info.), gen.)*

This variable contains the cumulated information on net income from all ET (>EUR 400). The variable can be generated for wave 1 by combining the open-ended and categorical information on net income from the main employment with the information for other activities (however, the categorical follow-up question is missing here). For waves 2 and 3, the variable is identical with the *netto* variable, which was provided in the SUF of wave 3. In waves 2 to 4, only the cumulated net income was surveyed – the source variables used in wave 2/wave 3 thus already include the respective information on total income from ET > EUR 400. The variable for wave 4 shall be generated analogous to wave 2/wave 3. As of wave 5, it will be generated based on spell-specific income information.

5.4.1.2 Duplicate pointer in p_register

zdub* *(pointer: personal identification no. of the individual doubled by the TP in wave X (20XY)*

The data structure in PASS (e. g. in the person register) is designed in such a way that a personal identification number can only be allocated to one household in each wave.

Thus, individuals who de facto belong to more than one household or for whom a change of households (move) was not reported properly must be treated differently.

A wave-specific pointer variable (*zdub**) marking these cases is created in the person register to achieve this.

Two different types of problems must be differentiated:

1. Real duplicates

Real duplicates are individuals who de facto belong to two households in a wave[39]. The households concerned were interviewed and the individual is included in the respective household structures.

If there were individual-level interviews with duplicate and original in the current wave, then the interview of the duplicate is removed and will not be used for the preload generation in the next wave either. Analogous to the other personal interviews deleted during data preparation, marking occurs in the *pnetto** variables of the respective wave. Weighting only uses one of the two observations of the individual in the current wave. Special treatment of these cases is thus not necessary for weighting.

If there is only one individual-level interview for either duplicate or original in the current wave, then this interview is not removed, i. e. if there is no competing information from the interview with the original, the duplicate interview remains in the SUF.

The information from the personal interview considered as valid is used for both the duplicate and the original for the preload generation of the next wave. This is made in particular regarding the spell information to be updated. It is possible that real duplicates give personal interviews in their original household and in their split-off household over the waves. If the individual-related preload for the next wave were generated depending on whether the individual provided information as original or duplicate, this would lead to multiple surveying of biographic information over the waves. This would then either have to be combined retroactively during data preparation or stored in the data as redundant information. The preparation of individual-related preload information irrespective of the household in which the information was provided avoids problems like this. However, the household-related preload is different for the two households of the duplicate individual.

Irrespective of the fact whether the individual gives a personal interview as original or duplicate, the individual maintains his/her known personal identification number *pnr* from the original household. This procedure is possible in the individual cross-section since the household number *hnr* shows in which household the duplicate gave the personal interview. This procedure is even mandatory in the spell datasets since an individual's biography is updated here and the biography of this individual shall not be divided onto two personal identification numbers.

Original and duplicate are documented in two data rows in the person register. A wave-specific pointer variable *zdub** is integrated which points from a duplicate to the original (irrespective of the interview status of duplicate and original on the individual level). For the observation of the duplicate in the person register, this pointer variable thus contains the permanent personal identification number of the original, i.e. it can only be filled with a personal identification number for individuals who are duplicates. If an observation is no duplicate, the variable is filled with "0" (analogous to the proceedings with other pointer variables) or with "-6" if the individual's household was not surveyed in the current wave or the individual is no longer part of a survey household (analogous to the allocation of code -6 in the other variables of the person register). A duplicate individual is thus included

---

[39] Whether this is the same individual is ensured during the household structure test. This is based on demographic information (name, age, sex, date of birth).

twice in the person register. On the one hand as original: There, the *pnr* is the permanent personal identification number of the original under which the individual is known since entering the panel, *zdub** equals 0. On the other hand as duplicate: There, the *pnr* is newly generated from the *hnr* of the household in which the individual is a duplicate and the position of the duplicate in the household. *zdub** contains the original's permanent personal identification number.

In the household in which the individual is a duplicate, the personal identification number stored in *pnrzp** in the *hh_register* is also changed to the personal identification number of the duplicate stored in *p_register* if the duplicate individual is the HRP of this household in one wave.

In the following waves, skipping one of the two households does not lead to a cancellation of the duplicate.

Thus, analyses based on several SUF datasets can largely be performed as usual despite the occurrence of duplicate individuals. Please observe the following when using *p_register*: For matchings with the *p_register* via the personal identification number, you must first generate a match variable equalling *zdub**, if it exceeds 0, or otherwise equalling *pnr*. Furthermore, not interviewed cases must be distinguished using *pnetto** to avoid, for instance, that information of the original is merged if the duplicate individual gave the personal interview in a wave.

2. Potential duplicates

An individual is known as member of a household which was already interviewed in PASS in the past (=original HH). Although this household was not interviewed in the current wave, the individual appears in another household (=duplicate HH). Since the original HH of this individual has not been interviewed since the appearance of this individual in the duplicate HH, it remains unclear whether the newly integrated individual is a duplicate or a regular move (which just has not been recorded yet). This individual is thus a potential duplicate of the original in the original HH.

In case of potential duplicates it is assumed that this is a move-out from the original household that has not been reported yet. Consequently, the potential duplicate is assigned the permanent personal identification number of the original in the SUF, i.e. the individual is treated as if he/she moved from the original HH to the duplicate HH. Individual-level interviews conducted in the current wave remain in place.

Since it is not certain that this is a duplicate, but instead the personal identification number of the individual concerned is changed, the pointer variable does thus not include a personal identification number.

The procedure for the following wave's preload is as described under (1). The individual-related preload is thus updated across households, regardless of whether it is a real or potential duplicate.

The following wave can determine whether this is a real duplicate (see 1). In this case, a second row is retroactively included in *p_register* for the individual whose *pnr* is newly generated from the *hnr* of the household in which the individual is a duplicate and the position of the duplicate in this household and in which *zdub** is filled with the permanent personal identification number of the original. This is then also made retroactively for all waves in which the individual now recognised as duplicate lived in his/her household which originated from the original household. As of the SUF of wave 5, there have thus been *zdub** variables in the *p_register* for all waves as of wave 2, although the first real duplicate was only observed in wave 4. If necessary, also *pnrzp** is changed retroactively in the *hh_register* in these cases.

Categories of the variable to be generated:

> -6 HH n. interv./TP no memb. of interv. HH
> 0 TP is no duplicate of another indiv.
> (permanent personal identification number of the "original" if TP is a duplicate)

5.4.2 Error corrections

During the data preparation process for the scientific use file of wave 5, some changes were also made to the waves of PASS, which had already been delivered. These changes included corrections of errors that were detected after the completion of the scientific use file of wave 4. Tables 24 to 28 give an overview of the retroactive changes to the already delivered waves of PASS[40].

---

[40]   Adjustments to value labels or variable labels are only taken into account here if this changes the interpretation of variables or values.

Table 24: **Overview of retroactive changes in the household dataset (*HHENDDAT*)**

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
depindug2 depindg2	*HHENDDAT*	1-4	Added	See Chapter 4.5.1
HD1101*	*HHENDDAT*	4	Correction	Code 14 was formerly labelled with "mini job, marginal employment (>= EUR 400)", correctly, this must mean "mini job, marginal employment (<= EUR 400)".
HW0881	*HHENDDAT*	2+4	Correction	Previously, responses such as "don't know" and "details refused" from open texts were handled inconsistently. Partly, the variables of these cases were set to code -1 or -2 including the open response, partly the "other" code triggering the recording of the open response was maintained. This has now been standardised in such a way that in such cases always the "other" code triggering recording of the open response is maintained since the "other" information has the higher informational content. 56 cases were corrected in wave 2 and one case in wave 4.

Table 25: Overview of retrospective alterations in the individual dataset (*PENDDAT*)

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
PB0401 ostaatan(na) PG0901* PG1301 PSH0201 mschul2 PSH0301* mberuf2 PSH0501 vschul2 PTK0321* PEE0201*	*PENDDAT*	1-4	Correction	Previously, open responses like "don't know" and "details refused" from open texts were handled inconsistently. Partly, the variables of these cases were set to code -1 or -2 including the open response, partly the "other" code triggering the recording of the open response was maintained. This has now been standardised in such a way that in such cases always the "other" code triggering recording of the open response was maintained since the "other" information has the higher informational content. In waves 1 to 4, the following numbers of cases occurred for the individual variables: PB0401: W3: 4 ostaatan(na): W3: 1 PG0901*: W2: 53; W4: 3 PG1301: W2: 2; W3: 3 PSH0201: W1: 1; W3: 3; W4: 1 mschul2: W1: 1; W2: 1; W3: 4; W4: 3 PSH0301*: W1: 1; W3: 1 mberuf2: W1: 1; W3: 1; W4: 1 PSH0501: W1: 1; W2: 1; W3: 4; W4: 4 vschul2: W1: 1; W2: 1; W3: 5; W4: 9 PTK0321*: W3: 5; W4: 1 PEE0201*: W4: 1
migration	*PENDDAT*	1-4	Correction	There were several mistakes in the previous generation. Individuals without an own migration background were only set to code "3 - at least 1 parent migrated to Germany" if exactly one parent had migrated to Germany. If, however, both parents had migrated to Germany, these individuals were assigned code "1 - no migration background". Additionally, it was insufficiently checked whether there were only missing values for both parents regarding the questions on migration from abroad. Migration of the mother's mother was only considered in part.
hhalg2	*PENDDAT*	1-4	Omitted	This variable includes the control information regarding the UB II receipt in the household for the personal questionnaire at the field time and is not to be used for analyses.

Table 26: Overview of retroactive corrections in spell datasets (bio_*spells, alg2_spells, ee_spells*)

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
branche	*bio_spells*	2-4	Correction	Undefined values 8 and 42 were recoded to -5.
ET0552	*bio_spells*	4	Correction	If ET0602 = -6, the missing value in ET0552 was altered from -3 to -6.
ET1952	*bio_spells*	4	Correction	If ET0602 = -6, the missing value in ET1952 was altered from -3 to -6.
AL0601	*bio_spells*	2-4	Correction	If AL0600 = -6, the missing value in AL0601 was altered from -3 to -6.
BIO0101 AL0601 LU0101	*bio_spells*	2-4	Correction	Previously, open responses like "don't know" and "details refused" from open texts were handled inconsistently. Partly, the variables of these cases were set to code -1 or -2 including the open response, partly the "other" code triggering the recording of the open response was maintained. This was now standardised in such a way that in such cases always the "other" code triggering recording of the open response was maintained since the "other" information has the higher informational content.
AL21300a-AL22103a AL21300b-AL22103b AL21300c-AL22103c	*alg2_spells*	1-4	Correction	Previously, open responses like "don't know" and "details refused" from open texts were handled inconsistently. Partly, the variables of these cases were set to code -1 or -2 including the open response, partly the "other" code triggering the recording of the open response was maintained. This was now standardised in such a way that in such cases always the "other" code triggering recording of the open response was maintained since the "other" information has the higher informational content.

Table 27: Overview of retrospective alterations in the register datasets (*hh_register; p_register*)

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
*weg*4*	*p_register*	4	Correction	The moving-out date in W4 had to be complemented for four persons. These TP had been reported as having moved out of their original HH in W4 and also the split-off HH was realised in W4. Since the original HH, however, provided an aborted interview, the information was not included in the SUF.
zupanel	*p_register*	4	Correction	The value "zupanel==4" was assigned to the duplicate individual known in W4, although he/she has already belonged to the panel via his/her original household since wave 2.
erwprox	*p_register*	4	Correction	Code 14 was formerly labelled with "mini job, marginal employment (>= EUR 400)", correctly, this must mean "mini job, marginal employment (<= EUR 400)".

Table 28: Overview of retrospective alterations in the weighting datasets (hweights; pweights)

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
-	-	-	-	-

5.5 Anonymisation

All data surveyed by the IAB as a special department of the Federal Employment Agency (BA) are social data, which places high demands on data protection. It was therefore necessary to include some of the variables in the scientific use file in simplified form. These variables are generally labelled with the flag "anonymised" in the variable label. For the same reason it was also necessary to exclude available regional information, with the exception of the German federal states and information on East/West Germany derived from this. For reasons of data protection, neither the data on family relationships in the household nor the first names of the household members are part of the scientific use file. References to the household structure are provided, however, by generated variables, for example on the household and benefit unit type (*hhtyp*[41], *bgtyp*[42]), indicator variables on partners in the household (*apartner; epartner*[43]), pointer variables pointing to parents and partners in the household (*zmhh; zvhh; zparthh*[44]) and various indicator variables which show whether parents (*mhh; vhh*[45]) or children of the target person (e. g. *ekind*[46]) are living in the household.

Table 29 provides an overview of the variables concerned and the process of anonymisation[47] in the individual dataset. Table 30 shows the anonymised variables of the employment spell dataset.

---

[41] Contained in the household dataset (*HHENDDAT*), see Chapter 4.5.2
[42] Wave-specific variables contained in the person register (*p_register*), see Chapter 4.4.
[43] Contained in the individual dataset (*PENDDAT*), see Chapter 4.4.
[44] Wave-specific variables contained in the person register (*p_register*), see Chapter 4.4.
[45] Contained in the individual dataset (*PENDDAT*), see Chapter 4.4.
[46] Contained in the individual dataset (*PENDDAT*), see Chapter 4.4.
[47] If non-anonymised versions of one or several variables are indispensable for your research, please contact the Forschungsdatenzentrum (Research Data Centre) to find a suitable possibility of obtaining access to the data. The form of this access will depend on the research project and the variables necessary for it.

Table 29: Overview of the anonymised variables in the individual dataset (*PENDDAT*) in wave 5

Varname	Variable label	Procedure
PD0100	Year of birth (date of birth, anon.)	The precise date of birth was shortened to the year of birth.
gebhalbj	Half-year of birth, gen.	The precise date of birth was shortened to an indicator for the first or second half of the year.
PET1210	Last occupational status, simple classification (before January 2005) (anon.)	For technical reasons, professional and regular soldiers were recorded separately in the survey. Due to the small amount of case numbers and as this group is not usually asked about occupational status anyway, this group was merged with that of civil servants and judges.
PET1250	Last occup. status civil servant: detailed info., incl. soldiers (before January 2005) (anon.)	This variable contains additional cases. The professional and regular soldiers from PET1240 were added to the corresponding civil servants category. The variable for professional and regular soldiers PET1240 is not supplied.
PET1211	Last occup. status, simple class. (incl. spell info.) (anon.), gen.	Procedure as for *PET1210*.
PET1251	Last occup. status civil servant: detailed info., incl. soldiers (incl. spell info.) (anon.), gen.	Procedure as for *PET1250*. The variable for professional and regular soldiers PET1240 is not supplied.
stiblewt	Occupational status, last employment, code number, gen.	When generating the occupational status variable, professional and regular soldiers are assigned to the corresponding civil servant category.
PET1510	Current occup. status, simple classification, surv. as of wave 2 (anon.)	Procedure as for *PET1210*.

Table 29: **Overview of the anonymised variables in the individual dataset (PENDDAT) in wave 5 (continued 1)**

Varname	Variable label	Procedure
PET1900	Current occup. status civil servant: detailed info., incl. soldiers (anon.)	Procedure as for *PET1250*. The variable for professional and regular soldiers PET1800 surveyed in the senior citizens' interviews is not supplied. For the personal interviews, no generated variable for professional and regular soldiers is incorporated into the individual dataset from the employment spells ET090*.
stibkz	Current occupational status, simple classification, harmonised (anon.)	When generating the occupational status variable, professional and regular soldiers are assigned to the corresponding civil servants category.
stib	Occupational status, code number, gen.	Procedure as for *stiblewt*.
PET3300	First occup. status, simple classification (anon.)	Procedure as for *PET1210*.
PET3700	First occup. status civil servant: detailed info., incl. soldiers	Procedure as for *PET1250*. The variable for professional and regular soldiers PET3600 is not supplied.
PET3301	First occup. status, simple class. (merged, incl. spell info.) (anon.), gen.	Procedure as for *PET1210*.
PET3701	First occup. status civil servant: detailed info., incl. soldiers, (merged, incl. spell info) (anon.), gen.	Procedure as for *PET1250*. The variable for professional and regular soldiers PET3600 is not supplied.
stibeewt	Occupational status, first employment, code number, gen.	Procedure as for *stiblewt*.
PSH0320	Mother's occup. status at that time, simple classification (anon.)	Procedure as for *PET1210*.
PSH0360	Mother's occup. status at that time, civil servant, incl. soldiers: detailed info. (anon.)	Procedure as for *PET1250*. The variable for professional and regular soldiers PSH0350 is not supplied.
mstib	Mother's occupational status, code number, gen.	Procedure as for *stiblewt*.
PSH0620	Father's occup. status at that time, simple classification (anon.)	Procedure as for *PET1210.*

Table 29: **Overview of the anonymised variables in the individual dataset (PENDDAT) in wave 5 (continued 2)**

Varname	Variable label	Procedure
PSH0660	Father's occup. status at that time, civil servant, incl. soldiers: detailed info. (anon.)	Procedure as for *PET1250*. The variable for professional and regular soldiers PSH0650 is not supplied.
vstib	Father's occupational status, code number, gen.	Procedure as for *stiblewt*.
PMI0200	Not born in Germany: country of birth	Countries with very low case numbers were grouped into larger categories.
ogebland	Country of birth, incl. open info., categories (anon.)	Procedure as for *PMI0200*.
PMI0500	No German nationality: which nationality? (anon.)	Nationalities of countries with very low case numbers were grouped into larger categories.
ostaatan	Nationality, incl. open info., categories (anon.)	Procedure as for *PMI0500*.
PMI1000a	Father: country of res. before migration (anon.)	Countries of residence before migration with very low case numbers were grouped into larger categories.
PMI1000b	Mother: country of residence before migration (anon.)	Procedure as for *PMI1000a*.
PMI1000c	Father's father: country of residence before migration (anon.)	Procedure as for *PMI1000a*.
PMI1000d	Father's mother: country of res. before migration (anon.)	Procedure as for *PMI1000a*.
PMI1000e	Mother's father: country of residence before migration (anon.)	Procedure as for *PMI1000a*.
PMI1000f	Mother's mother: country of residence before migration (anon.)	Procedure as for *PMI1000a*.
ozulanda	Father: country of residence before migration, incl. open info., categories (anon.)	Procedure as for *PMI1000a*.
ozulandb	Mother: country of residence before migration, incl. open info., categories (anon.)	Procedure as for *PMI1000a*.

Table 29: Overview of the anonymised variables in the individual dataset (PENDDAT) in wave 5 (continued 3)

Varname	Variable label	Procedure
ozulandc	Father's father: country of residence before migration, incl. open info., categories (anon.)	Procedure as for *PMI1000a*.
ozulandd	Father's mother: country of residence before migration, incl. open info., categories (anon.)	Procedure as for *PMI1000a*.
ozulande	Mother's father: country of residence before migration, incl. open info., categories (anon.)	Procedure as for *PMI1000a*.
ozulandf	Mother's mother: country of residence before migration, incl. open info., categories (anon.)	Procedure as for *PMI1000a*.

Table 30: Overview of the anonymised variables in the BIO spell dataset (bio_*spells*) in wave 5

Varname	Variable label	Procedure
ET0601	Occup. status, simple classification (anon.)	Procedure as for *PET1210*.
ET1001	Occ. status civil servant: detailed info. (anon.)	Procedure as for *PET1250*. The variable for professional and regular soldiers is not supplied.
stib	Occ. status, code number, gen.	Procedure as for *stiblewt*.

5.6 Receipt of Unemployment Benefit II

Receipt of Unemployment Benefit II at the household level was already recorded in spell form in waves 1 to 4. This concept was continued in wave 5 but with a slightly revised set of questions.

5.6.1 Concept for updating the spells of Unemployment Benefit II receipt that were still ongoing in the previous wave

In order to update the spells of Unemployment Benefit II receipt which were still ongoing in the previous wave and were therefore right-censored in the spell dataset, dependent interviewing questions are included. Households with ongoing spells from the previous wave start here again with the interview.

The households from the refreshment sample which were interviewed for the first time in wave 5 were asked about their receipt of Unemployment Benefit II during the period since the last change in the household composition. If this was before January 2009 or if no information was provided about changes in the household, then the household's receipt of Unemployment Benefit II from January 2009 onwards was recorded.

5.6.2   Structure of the Unemployment Benefit II spell dataset

The structure and the contents of the spell dataset on Unemployment Benefit II change due to the integration of the spells of Unemployment Benefit II receipt reported in wave 5. Here it is necessary to distinguish between (1) new variables that refer to a particular wave, (2) new variables that do not refer to a particular wave and (3) variables that are no longer surveyed in wave 5.

1. Also in wave 5, new wave-specific cross-sectional variables were included in the Unemployment Benefit II spell dataset. These are: *AL20604, AL20704a to AL20704o, AL20804 and AL20904*. These variables refer to the interview date of wave 5. Cross-sectional variables also exist for the interview dates of the previous waves which contain the analogous information referring to the respective wave. Table 31 provides an overview of the cross-sectional information contained in the Unemployment Benefit II spell dataset.

Table 31:   **Cross-sectional variables in the UB II spell dataset (*alg2_spells*)**

	Cross-sectional variable with information referring to				
	Wave 1:	**Wave 2:**	**Wave 3:**	**Wave 4:**	**Wave 5:**
Does the HH receive UB II for all HH members?	*AL20600*	*AL20601*	*AL20602*	*AL20603*	*AL20604*
Does the HH receive UB II for individuals 1 to 15?	*AL20700a* to *AL20700o*	*AL20701a* to *AL20701o*	*AL20702a* to *AL20702o*	*AL20703a* to *AL20703o*	*AL20704a* to *AL20704o*
Amount of monthly UB II receipt?	*AL20800*	*AL20801*	*AL20802*	*AL20803*	*AL20804*
Has a cut of UB II begun?	*AL20900*	*AL20901*	*AL20902*	*AL20903*	*AL20904*

2. Not available in wave 5 as compared to wave 4.

3. Not available in wave 5 as compared to wave 4.

5.6.3 Plausibility checks and corrections in the Unemployment Benefit II spell dataset

As was done in waves 1 to 4, the information on receipt of Unemployment Benefit II was also subjected to a number of plausibility checks in wave 5. Inadmissible overlaps and dates of spells of Unemployment Benefit II receipt or of benefit cuts were corrected if necessary. In principle, changes were only made to the generated date variables (*bmonat; bjahr; emonat; ejahr*) of the spell of Unemployment Benefit II receipt, the spells of benefit cuts (*alg2kbm; alg2kbj; alg2kem; alg2kej*) and the censoring indicator of the spell of Unemployment Benefit II receipt (*zensiert*). If it was not possible to remove implausibilities by correcting the date variables, then, in a small number of cases, spells of Unemployment Benefit II receipt were merged, or spells of Unemployment Benefit II receipt or benefit cuts were deleted entirely.

5.6.4 Updating the Unemployment Benefit II spell dataset

After the spells of Unemployment Benefit II receipt that were reported in wave 5 had been converted into spell format and following the plausibility checks and corrections, where inadmissible overlaps and spells with implausible dates were corrected, the spells of Unemployment Benefit II receipt which were still ongoing at the time of the interview in the previous wave were updated using the information gathered in wave 5. Two variants are to be distinguished here. In the first (1), only the censoring indicator *zensiert* is changed. The second variant (2) is an update of the spell which was censored in the previous wave using information gathered in wave 5 in the narrow sense. Here the censoring indicator is integrated into the spell of Unemployment Benefit II receipt which was still ongoing in the previous wave, as are the generated and surveyed end dates, the wave-specific cross-sectional information (see above) and information about new spells of benefit cuts. In addition to updating spells which were censored in the previous wave, new spells that were reported in wave 5 are merged with the spell dataset (3). These three variants are outlined briefly below:

1. *Cases in which the household in wave 5 contradicts an ongoing spell of Unemployment Benefit II receipt at the interview date in the previous wave.*

   If the household contradicted the information that there was an ongoing spell of Unemployment Benefit II receipt at the time of the previous wave, either explicitly or implicitly (by reporting an end date that preceded the interview date in the previous wave) in the update question, then *zensiert* was set to "2" (no). The information provided in the interview of the previous wave is assumed to be correct. As it is not possible to make any reliable statements about the continued duration of the benefit receipt beyond the date of the interview in the previous wave, it is assumed that the benefit receipt ended in the month of the interview in the previous wave. The reported and generated variables for the end date of the spell (*AL20300, AL20400* and *emonat, ejahr*) as well as the question whether a spell still continues (*AL20500*) remain un-

changed.[48] The generated end date of the Unemployment Benefit II spell (*emonat; ejahr*) was already set to the interview date of the previous wave in the previous wave.

2. *Cases in which the household reports the end date of a spell of benefit receipt that was still ongoing in the previous wave.*

If information about the end date of a spell of Unemployment Benefit II receipt that was censored in the previous wave is available in wave 5, then the spell which was censored in the previous wave was updated using the current information. First, the surveyed end date (*AL20300; AL20400*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the receipt of Unemployment Benefit II is still ongoing (*AL20500*) and the censoring indicator (zensiert) are overwritten with the information gathered in the previous wave. Furthermore, the spells of benefit cuts reported in wave 5 and the cross-sectional data referring to wave 5 (*AL20604; AL20704a to AL20704o, AL20804, AL20904*) were included.

3. *Spells of Unemployment Benefit II receipt reported for the first time in wave 5 which do not update any spells that were censored in the previous wave.*

Spells reported for the first time in wave 5 were added to the Unemployment Benefit II spell dataset. Then the spell counter was generated anew in order to create a variable *spellnr* without gaps.

5.7 Employment biographies

Employment, unemployment and gap periods at the individual level were recorded in spell form already in waves 2 and 3. This concept of modular spell survey was changed to an integrated survey of the employment biography in wave 4. For individuals that were asked for their employment biography for the first time in wave 5, the reference date for the start of the retrospective interval was adjusted. In wave 5, all spells of employment and unemployment since January 2009 were to be reported here. Individuals who were interviewed on their employment biography already in the previous wave, however, should report all new spells since the date of the last interview.

5.7.1 Concept for updating the spells that were still ongoing in the previous wave

Continuing ET, AL and gap spells were updated in wave 5. In order to update the spells which were still ongoing in the previous wave and were therefore right-censored in the spell dataset, dependent interviewing questions are included in the personal questionnaires.

---

[48]   The same applies here. Only the censoring indicator is changed. The reported end date, the question for continuing spells and the generated end date remain unchanged.

5.7.2 Structure of the BIO spell dataset

Regarding its structure, the BIO spell dataset has oriented itself on the modular ET, AL and LU spell datasets of waves 2 to 3 since wave 4 . ET-specific variables kept their name in the BIO spell dataset as compared to the ET SUF of wave 3, analogously also AL- and LU-specific variables. Variables which are the same in ET, AL and LU have been standardised (BIO0100, BIO0101, BIO0200, BIO0300, BIO0400, BIO0500, BIO0600) as of wave 4 or were already standardised in the original datasets of the SUF wave 3 (bmonat, bjahr, emonat, ejahr, zensiert). Furthermore, variables for the type of activity (spelltyp), for spell integration (spintegr) and a comprehensive spell number (spellnr) are available.

Due to the integration of the employment and unemployment spells reported in wave 5 into the BIO spell dataset, new ET- and AL-specific variables are added. Here it is necessary to distinguish between (1) new variables that refer to a particular wave and (2) new variables that do not refer to a particular wave.

1. The ET-specific variables in the BIO spell dataset *ET0600 to ET2200* are considered as wave-specific, cross-section information referring to wave 2; variables ET0601 to ET2201 as referring to wave 3, *ET0552 to ET2202* as referring to wave 4 and ET0553 to ET2203 as cross-section information referring to wave 5. Table 32 provides an overview of the ET-specific cross-section information in the BIO spell dataset.

Table 32:   ET-specific cross-section variables in the BIO spell dataset (bio_spells)

	Cross-sectional variable with information referring to...			
	Wave 2:	**Wave 3:**	**Wave 4:**	**Wave 5:**
Occupational status (simple and detailed classification)	*ET0600* *ET0700* *ET0800* *ET1000* *ET1100* *ET1200*	*ET0601* *ET0701* *ET0801* *ET1001* *ET1101* *ET1201*	*ET0552* *ET0602* *ET0702* *ET0802* *ET1002* *ET1102* *ET1202*	*ET0553* *ET0603* *ET0703* *ET0803* *ET1003* *ET1103* *ET1203*
Supervisory function; number of employees supervised	*ET1300* *ET1400*	*ET1301* *ET1401*	*ET1302* *ET1402*	*ET1303* *ET1403*
Cancellation of limitation of an initially limited employment	*ET1700*	*ET1701*	*ET1702*	*ET1703* *ET1753a* *ET1753b*
Working hours (contracted; actual; average for irregular working hours)	*ET2000* *ET2100* *ET2200*	*ET2001* *ET2101* *ET2201*	*ET1952* *ET2002* *ET2102* *ET2202*	*ET1953* *ET2003* *ET2103* *ET2203*

The BIO spell dataset furthermore includes an AL-specific variable which is to be understood as wave-specific cross-sectional information (AL1300 for wave 2; AL1301 for

wave 3, AL1302 for wave 4 and AL1003 for wave 5). Table 33 gives an overview of the cross-sectional information contained in the spell dataset.

Table 33: **AL-specific cross-section variables in the BIO spell dataset (bio_spells)**

Cross-sectional variable with information referring to …			
Wave 2:	**Wave 3:**	**Wave 4:**	**Wave 5:**
Amount of monthly UB I receipt?			
AL1300	AL1301	AL1302	AL1303

2. The non wave-specific ET variables *ET2800-ET3900* were surveyed first in wave 5 and were integrated in the BIO spell dataset.

5.7.3 Plausibility checks and corrections of the spell datasets

At the individual level, the plausibility checks and corrections orient themselves by wave 2 to wave 4. Similar to wave 4, the checks were made only within one spell type. Cross-spell type checks were not carried out. As with the spell data on Unemployment Benefit II receipt, corrections and recodings were only carried out in the generated date variables. Here, too, details on seasons were recoded into definite months, "-8" values were set for implausible responses and date information was replaced or rendered plausible. As only the generated date variables were edited, the original information gathered in the survey is available to the user in the date variables BIO0200-BIO0500 and AL0800-AL1100, thus permitting the user to conduct his/her own checks and corrections.

In addition, it seemed necessary to delete entire spells in some cases. Thus, for example, spells that were obviously recorded twice were removed. Spells that are completely outside the period surveyed but for which data were nonetheless collected were also deleted.

5.7.4 Update of spell datasets

After the spells that were reported in wave 5 had been converted into spell format and following the plausibility checks and corrections where inadmissible overlaps and spells with implausible dates were corrected, the spells which were still ongoing at the time of the interview in the previous wave were updated using the information recorded in wave 5.

Three variants are to be distinguished here. In the first (1), only the censoring indicator *zensiert* is changed. The second variant (2) is an update of the spell which was censored in the previous wave using information gathered in wave 5 in the narrow sense. Here, the censoring indicator is integrated into the spell which was still ongoing in the previous wave, as are the generated and recorded end dates and wave-specific cross-sectional information (see above).

In addition to updating spells which were censored in the previous wave, new spells that were reported in wave 5 are merged with the spell dataset (3). These three variants are outlined briefly below:

1. *Cases in which the individual in wave 5 contradicts an ongoing spell at the interview date in the previous wave.*

   If the individual contradicted the information that there was an ongoing spell at the time of the previous wave, either explicitly or implicitly (by reporting an end date that preceded the interview date in the previous wave) in the update question, then the censoring indicator *zensiert* was set to "2" (no). The information provided in the interview of the previous wave is assumed to be correct. As it is not possible to make any reliable statements about the continued duration of the spell beyond the date of the interview in the previous wave, it is assumed that the spell ended in the month of the interview in the previous wave. The reported and generated variables on the end date of the spell (BIO0400*, BIO0500 and emonat, ejahr*) as well as the question whether a spell still continues (BIO0600) remain unchanged[49]. The generated end date of the spell (*emonat; ejahr*) was already set to the interview date of the previous wave in the previous wave.

2. *Cases in which the individual reports the end date of a spell that was still ongoing in the previous wave.*

   If information about the end date of a spell that was censored in the previous wave is available in wave 5, then the spell which was censored in the previous wave was updated using the current information. For ET spells, the recorded end date (BIO0400*; BIO0500*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the spell was still ongoing (BIO0600), the reason for the cancellation of a work contract (*ET2300*), the generated variables on occupational status and weekly working hours (*stib, az1, az2*) and the censoring indicator (*zensiert*) were overwritten with the information gathered in wave 5. Furthermore, the cross-sectional data referring to wave 5 (*ET0553 to ET2203*) were included.

---

[49] Thus, the reported end date remains filled with the interview date of the wave in which the spell was censored or the special code "0" for continuing spells. Also the question whether the spell continued (for the case that the end date corresponds with the interview date) is not changed. The generated date variables continue to contain the last valid information, which here is the interview date of the wave in which the spell was censored.

For AL spells the recorded end date (BIO0400; *BIO0500*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the spell was still ongoing (BIO0600), the reason for the end of unemployment (*AL0600, AL0601*) and the censoring indicator (zensiert) were overwritten with the information gathered in wave 5. Furthermore, the cross-sectional data referring to wave 5 (*AL1303*) was included. AL spell data, moreover, feature the exception that the spell of Unemployment Benefit I (receipt of UB I) is recorded within an AL spell. Which information is updated depends on whether there already was a receipt of UB I in this spell of unemployment and whether this receipt was ongoing in the previous wave:

- If, in the previous wave, there also was an ongoing receipt of UB I in the AL spell to be updated, then the surveyed end date of the receipt (*AL1000, AL1100*), the indicator as to whether the spell is ongoing (*AL1200*), the generated end date of the receipt (*alg1em, alg1ej*) and the censoring indicator of the receipt (*alg1akt*) were overwritten with the information surveyed in wave 5.

- If no UB I was received in any previous waves in the AL spell to be updated, then the information on UB I receipt was overwritten with the information surveyed in wave 5. Besides the indicator as to whether UB I was received in the AL spell (*AL0700*), the reported start and end date (*AL0800, AL0900, AL1000, AL1100*), the indicator for ongoing receipt (*AL1200*) and the respective generated variables (*alg1bm, alg1bj, alg1em, alg1ej, alg1akt*) were replaced with the newly recorded information.

- If there was UB I receipt in the AL spell to be updated in the past, which, however, ended in the previous wave, no changes were made in these spells.

3. *Spells reported for the first time in wave 5 which do not update any spells that were censored in the previous wave.*

Spells reported for the first time in wave 5 were added to the BIO spell dataset. Then the spell counter was generated anew in order to create a variable *spellnr* without gaps.

Updating the spell datasets does not affect the spell numbers of the previous wave's SUF. Spells already included in the wave 4 SUF (spellnret, spellnral, spellnrlu, spellnr) maintain their spell number. The new spells from wave 5 are added to the respective dataset and the spell numbers are updated.

5.8   One-euro job spell dataset (ee_spells)

In wave 4, the concept for surveying participation in employment and training measures was thoroughly revised. The MN spell dataset has been replaced by the one-euro spell dataset (ee_spells) as of wave 4. This was updated for the first time in wave 5. The reference date as of which to consider one-euro jobs was January 2010 for wave 5.

5.8.1 Concept for updating the spells that were still ongoing in the previous wave

Continuing ET spells were updated in wave 5. In order to update the spells which were still ongoing in the previous wave and were therefore right-censored in the spell dataset, dependent interviewing questions are included in the personal questionnaires.

5.8.2 Structure of the EE spell dataset

By integrating the one-euro jobs (OEJ) reported in wave 5 in the OEJ spell dataset (ee_spells), new variables are added which refer to a specific wave. Table 34 gives an overview of the cross-sectional information contained in the EE spell dataset.

Table 34: Cross-sectional variables in the EE spell dataset (ee_spells)

Cross-sectional variable with information referring to...		
	Wave 4:	**Wave 5:**
Weekly working hours in the OEJ	*EE1100*	*EE1101*
OEJ is the same work permanent co-workers do	*EE1200*	*EE1201*
Which kind of training necessary for OEJ	*EE1300*	*EE1301*
Only work or also training/classes?	*EE1400*	*EE1401*
Assessment OEJ	*EE1500a-EE1500h*	*EE1501a-EE1501h*

For the OEJ spell dataset it must be considered that there are also spells if the OEJ was not performed, i.e. there was no participation.

5.8.3 Plausibility checks and corrections in the EEJ spell dataset

The OEJ spell dataset on the participation in one-euro jobs was checked for plausibility and corrected. The plausibility checks contained checks of dates, of the reference date for the newly integrated spells in wave 5 (1/2010) and checks for logical inconsistencies in case of respondents with several one-euro job spells.

Only the generated date variables (bmonat, bjahr, emonat, ejahr) were corrected and recoded. Details on seasons were recoded into definite months, "-8" values were allocated for implausible responses and date information was replaced or rendered plausible. Following that, a spell counter *spellnr* was generated. The generation was performed analogously to the chronological counters in the BIO spell datasets. Non-participating spells were not included in the sorting and thus kept their original *spellnr*. OEJ spells from wave 4 maintained their spell number from wave 4 also for the wave 5 SUF.

6 Weighting wave 5

The weighting concept for wave 5 generally follows the concept developed in previous waves (see Berg et.al, FDZ Datenreport 08/2011 for wave 4). The starting point for the weighting procedure for wave 5 and for the longitudinal section from wave 4 to wave 5 were the cross-sectional weights from wave 4 for households and individuals. The two weights of each household and the two weights of each individual were updated again. This chapter of the Datenreport documents the technical details and the exact models for the generation of the weights for wave 5. An overview of the weighting concept used in PASS can be found in chapter 8 (Trappmann, 2011) of the PASS User Guide (Bethmann & Gebhardt, 2011). Examples of how to use the weights can be found in Chapter 9.4 (Gebhardt & Trappmann, 2011) of the PASS User Guide.

6.1 Expansion of the wave 5 sample

PASS was supplemented with the following new samples in wave 5:

1.) A sample of SGB II inflows in July 2010

The procedure was as in the previous years. In the 300 originally selected postcode areas, new inflows were selected with the same sampling probability used to select the original households within this postcode area.

2.) replenishment samples for the existing panel samples

Since the 300 previously used points are severely exhausted, 100 new points for

i) a replenishment sample of the BA sample(s) and

ii) a replenishment sample of the population sample were drawn.

In doing so, highest possible comparability of the processes with the original sampling was aspired. Since the population sample was no longer drawn from the Microm addresses but from the registration offices' registers, the concept used in wave 1 (Rudolph & Trappmann, 2007) was modified. A disproportionate stratification of the population sample according to socio-economic status was waived. Furthermore, in particular two difficulties arose which prevented the complete transfer of the old concept to the new sample concept:

i) Municipal registers contain no information on households. Thus a person sample had to be drawn and then extended to the households these person live in.

ii) If selected postcode areas extend over several municipalities, then a decision had to be made from which municipality the addresses were to be drawn from.

6.1.1 Selection of other primary sampling units (PSUs)

Sampling points, short PSU (primary sampling units), in PASS are postcodes (PLZ). The PSUs are the same in both subsamples. The selection of the sampling points was made proportional to the size of the points (pps sampling) so that the size of the points is can-

celled out of the two-stage selection probability equation. Ideally, the size of a point would be defined as number of households within the point, since the units to be sampled in PASS are households. Since, however, only individuals (of 18 years and over) and no households can be selected from the municipal registers, individuals had to be used in this case (see below).

Furthermore, the sample was stratified at the first stage by municipal size classes (PASS uses the 10-level BIK municipality size class) and districts (Kreise). Each combination of municipal size and district received a weight , which equals the number of units in the cell. While every municipality in Germany can definitely be allocated to a cell in the municipal size × district matrix, this is not the case for the postcodes, which are the PSUs in PASS. Thus, each postcode must be allocated to a cell. This was made in such a way that the postcode was allocated to the cell which comprised most of the units. In order to allow an unbiased estimation of the design weights afterwards, it was, however, also necessary to allocate the units living there to the weight of this cell. It was thus necessary to know the number of units per postcode and municipality.

Since it is only possible to draw a sample of individuals of 18 years and over within the postcode area, the weight, i. e. the number of units, is ideally also defined as individuals of 18 years and over. For the drawing, however, the data provided by the ADM e.V. member institutions and the data provided to infas by infas Geodaten were used, which only include the number of individuals of 15 years and older per postcode and per municipality. Thus, the weight for each cell was determined using the number of 15-year-olds and older individuals. On this basis, 100 postcodes were selected which were distributed to the cells of the allocation matrix in proportion to the weights of the cells. Allocation was effected using the Cox algorithm (Cox 1987).

6.1.2 Selection of the municipalities for supplementing the population sample

Individuals were selected from the registers of the municipalities' registration offices. However, the PSUs in PASS are postcodes. It is impossible to draw a simple random sample of persons from these postcodes if postcodes exceed municipality limits. Drawing samples from all municipalities within a postcode in rural areas would, in some cases, involve a lot of effort. 19 of the 100 sampled postcodes extended over more than one municipality, sometimes over more than ten municipalities.

It was thus necessary to include another sampling stage. Since the distribution of individuals across the municipalities within the postcode area was known (this also required using the distribution of 15-year-olds and older individuals as approximate value since the distribution of 18-year-olds and older individuals was not available), a selection proportionate to the population in a municipality within a postcode area could be made. The selection probability of a municipality within the postcode area was thus proportionate to the share of individuals in this post-code area who reside in the given municipality The additional selection level impedes the variance estimation but maintains the low variance of the design weights without introducing a bias.

6.1.3 Selection of the households for supplementing the population sample

Individuals were selected in those municipalities in which the drawn postcode areas lie. To do so, the municipalities were contacted in writing and asked for a random sample of individuals of 18 years and older from the registration offices' registers solely from the selected postcode areas.

This results in an individual sample with the following selection probability:

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

p$_{ia}$ selection probability household a in point i

q$_{ib}$ selection probability individual b in point i

Q population of individuals of 18 years and older in population

Q$_i$ population of individuals of 18 years and older in point i

k number of sample points

q$_i$ sample size individuals of 18 years and older in point i

q$_{ia}$ number of individuals of 18 years and older in the household in point i

Q_i is cancelled from formula (1). This achieves the ideal of an equal selection probability of individuals in all points if the number of individuals selected q_i has the same size in each point.

In order to determine the selection probability of a household information on the number of individuals of 18 years and older living in the household is required. This information can, however, only be determined ex-post from the survey, i. e. only for the realised cases.

The following applies:

$$(2)\ p_{ia} = \frac{Q_i}{Q} \times k \times \left(q \times \frac{q_{ia}}{Q_i} \right) - k \times \left(q_i \times \frac{q_{ia}}{Q} \right)$$

The effect of different household sizes q_{ia} increasing the variance of the the sampling probabilities cannot be avoided without knowing the grouping of individuals to households beforehand.

The mentioned number of individuals of 18 years and older per postcode x municipality was, however, not available so that instead the number of individuals of 15 years and older had to be used. However, the empirical correlation between the number of individuals of 18 years and older and the number of individuals of 15 years and older at the postcode level should be close enough to 1 so that the variance-reducing characteristic of the pps design remains largely intact.

6.1.4 Design weights for supplementing the population sample

The design weights for the replenishment sample for the general population sample (sample = 6) are defined as reciprocal value of the selection probability as described in 6.1.2 on the different levels of the sample design. The selection probabilities are defined via three selection stages. The selection probability of the PSU, the selection probability of the municipality in the cases in which a postcode extends over several municipalities (in all other cases = 1) and the selection probability of the individual in the PSU. By multiplying these three selection probabilities the selection probability of the selected individual in the gross sample is determined.

The transformation of the individual sample to a household sample is an additional step in the replenishment of the population sample which can only be taken for realised cases. This additional weighting step which corrects the different selection probabilities due to the different (reduced) household size, was made after the calculation of the participation propensities, i. e. after the transition from the gross sample to the net sample, by multiplying the selection probabilities of the individuals by the estimated participation propensity and the number of target persons in the household.

6.1.5 Selection of the BA replenishment sample

Parallel to the drawing of the registration offices' sample, the IAB drew a panel replenishment of the UB II samples in the same postcode areas (sample = 7). Since the selection probability of a postcode had already been determined and equalled the population size in this PSU, the number of benefit recipients in the PSU, however, varies additionally depending on the rate of people in need, significantly different selection probabilities would have occurred within the PSUs with identical q_i. In particular benefit recipients in areas with a high rate of people in need would have had low selection probabilities. This was not desired.

In Wave 1 the PASS team already developed a solution to this problem (see Rudolph & Trappmann 2007, 82). The number of benefit recipients to be selected per PSU is selected as product of the permanent $q_i=q$ (sample size individuals per PSU) in the population sample with the quotient from benefit recipient rate in the PSU and benefit recipient rate in all over Germany. Additionally, the database for the sampling of the benefit recipient sample provides a summary of the individuals in benefit units which equal households by approximation. Consequently, within each postcode area the units of interest could be sampled directly in this source. Formula 4.4 in Rudolph and Trappmann (2007) provides the selection probabilities of a benefit unit within this sample.

In order to understand the concept of a replenishment of the BA samples, it is beneficial to visualise the population from which the previous process data samples were selected: The BA sample from wave 1 is a random sample of households in receipt of Unemployment Benefit II on the sampling date in July 2006. The refreshments from waves 2, 3, 4 (and now 5, see above) are random samples of households receiving Unemployment Benefit II in July 2007, 2008, 2009 or 2010 in which no individual lived who received Un-

employment Benefit II on one of the previous sampling dates ("new inflows"). The total BA sample in wave 5 is thus a random sample of households which received Unemployment Benefit II in (at least) one July of the years 2006, 2007, 2008, 2009 or 2010.

Does it make sense to draw a sample from this combined population? This is certainly not the case. Benefit recipients from 2006 who are no longer in receipt are not of sufficient interest for SGB II research. Furthermore, they can be sufficiently examined with the original sample of over 6,000 households in receipt in July 2006.

Now a second characteristic of this specific refreshment concept becomes important. All households of the five mentioned BA samples (wave 1 plus refreshment samples) which were still receiving Unemployment Benefit II on the most recent sampling date (July 2010) are a random sample of all households receiving benefits at this time. This population was now refreshed with the panel replenishment in order to provide sufficient statistical power for analyses in particular regarding relevant subpopulations (e. g. young people (U25), individuals with a migration background, single parents, and replenishment benefit recipients) in benefit receipt in the future.

In analytical terms it makes sense for the calculation of the design weights to distinguish between cases from the population in the replenishment sample who belong to the panel population of wave 4 and those cases belonging to the population of the wave 5 refreshment (who entered receipt after wave 4). This differentiation was made by setting the cases from the panel replenishment which were inflows to sample = 8 (refreshment sample). They were treated like inflows in the design weighting.

6.2 Integration of the replenishment samples with the ongoing panel samples

Integrating the replenishment samples with the ongoing panel samples in terms of weighting is not trivial since weights must be integrated several times. This integration, which only became necessary in wave 5, was performed between steps 7 (8.2.7) and 8 (8.2.8).

In a first step cases in the ongoing panel that are not (or no longer) part of the population of the refreshments are identified. The weights of this subset are not affected by the replenishment samples. After this step, the remaining panel sample and the panel supplements are now two random samples of the same population with known inclusion probabilities. Thus, the concept of convex combination (Spieß &Rendtel, 2000) was applied to the fusion of these two samples and the combination with minimum variance was chosen.

The weights of the combined population samples are to project the initial general population sample (sampled from the Microm database) and the replenishment from municipal registers to all households in Germany. Thus, separate weights were calculated initially for the general population panel sample and the general population replenishment sample following the steps described in the PASS User Guide sections 8.1 and 8.2.1-8.2.7 (Trappmann, 2011). Then, the general population panel was integrated with the general

population replenishment (sample = 6) via a convex combination to obtain the population weight before calibration.

The weights of the combined BA samples are designed to project the original BA sample, the four refreshment samples with new entries to Unemployment Benefit II and the new BA replenishment sample to all households which received benefits at one of the reference days in July of the years 2006-2010. The subset still receiving Unemployment Benefit II at the reference date in July 2010, is to be projected to all recipients at that date.

Initially, separate weights were calculated according to the steps described in the PASS User Guide sections 8.1 and 8.2.1-8.2.7 (Trappmann, 2011) for the ongoing BA panel sample and the BA replenishment. The BA replenishment (sample = 7) then had to be integrated with recipients of UB II from all BA samples of waves 1-4 (convex combination). Cases from the BA samples from waves 1-4 which did no longer receive UB II maintain their weight in this case. This resulted in the new BA weight before calibration.

Since the new BA refreshment sample (Sample = 8) and the BA panel samples are disjoint, all cases maintain their weights during the integration. A fusion of the samples for the calculation of the BA weight before calibration was not necessary.

Following this, the general population weights (wwmihh) and BA weights (wqbahh) were integrated to generate the total weights as described in the PASS User Guide section 8.2.8 (Trappmann, 2011).

6.3 Design weights for the panel households in wave 4

New "household design weights" were generated for the fourth wave from the cross-sectional weights for households of the third wave, taking into account people moving into households from within Germany. This was again done by using the weight share procedure as described in wave 2 (see Gebhardt et.al, FDZ Datenreport 06/2009). Births, deaths or move-outs from households have no influence on the weight; moves into households from within Germany, on the other hand, increase the inclusion probability of a household as the individuals who moved into the household also had the chance of being included in the sample in wave 1, wave 2 or wave 3 (only refreshment sample BA). The new design weight for subsample i dw_ihh_4 is therefore calculated from the old cross-sectional weight wq_ihh_3:

$$1/dw_ihh_4 = 1/wq_ihh_3 + (n_{sample\ i}/n_{population\ i})$$

The new design weight is only an intermediate step and is therefore not included in the data supplied for wave 4.

6.4 Design weights for the refreshment sample in wave 5

In wave 5, the panel was again refreshed by sampling new households from the new inflows to benefit receipt. All households that were in receipt of benefits in July 2010 but had had no probability of being selected for the register data sample in the same month in 2009, 2008, 2007 and 2006 had a chance of being drawn. This refreshment could be

done by selecting only benefit units in which no member was receiving benefits in July of the previous years. The refreshment sample was drawn in the 300 points of the first wave and the 100 replenishment points of wave 5. Analogous with the special pps procedure used to draw the first register data sample, which is described in Rudolph and Trappmann (2007), the sample size was proportional to the share of new benefit recipients in the population in the sampling point (at the time when the sampling points were selected). The calculation of the design weights is also described in the same article. For cases with *sample = 8*, the design weight of the refreshment sample is included in the variable *dw_ba*.

6.5 Propensity to participate again - households

In this step, again similar to the procedure in wave 4, the probability of re-participation in wave 5 was estimated for each household that participated in wave 4 on the basis of logit models for willingness to participate in the panel, availability and participation. Also households that participated in wave 3 but not in wave 4 (temporary non-responses) were considered in the modelling for wave 5. In addition to variables from the household interview and the personal interview with the head of the household in the previous wave, also other variables were included which are associated with the fieldwork, e. g. number of contact attempts. The estimated propensities of all three models were multiplied. The reciprocal value of this product can be found in the variable *hpbleib* for each wave. The longitudinal weight for a household from one of the samples of wave 1 for the total period possible $[t_1, t_2, t_3, t_4, t_5]$ across all five waves can be obtained as the product of the cross-sectional weight to t_1, *hpbleib* (wave 1 to wave 2) and *hpbleib* (wave 2 to wave 3, etc.) (see also the PASS User Guide section 9.4 (Bethmann & Gebhardt, 2011)).

Table 35: Variable overview, codes and reference categories for the logit models of the re-participating households

Variable code and reference category	Explanation
alter_1	Household reference person (HRP) younger than 30 years
alter_2	HRP 30-39 years of age
alter_4	HRP 50-64 years of age
alter_5	HRP 65 years and older
Reference category	HRP 40-49 years of age
sex_1	HRP male
Reference category	HRP female
nichtdeutsch	HRP has nationality other than German
Reference category	HRP has German nationality or missing information
schulbil_1	School qualification HRP: no qualification
schulbil_2	School qualification HRP: lower secondary school
schulbil_4	School qualification HRP: college/university qualification
Reference category	School qualification HRP: intermediate secondary school/still pupil
gesundheit_3	Subjective evaluation of the health state of the HRP: satisfactory
gesundheit_4	Subjective evaluation of the health state of the HRP: not so good
gesundheit_5	Subjective evaluation of the health state of the HRP: bad
Reference category	Subjective evaluation of the health state of the HRP: very good to good
zufrieden_1	General life satisfaction HRP: scale value 0-2
zufrieden_2	General life satisfaction HRP: scale value 3-5
zufrieden_3	General life satisfaction HRP: scale value 6-8
Reference category	General life satisfaction HRP: scale value 9-10
anz_0_3	Number of individuals in the household aged 0-3 years
anz_4_6	Number of individuals in the household aged 4-6 years
anz_7_14	Number of individuals in the household aged 7-14 years
anz_65	Number of individuals in the household aged 65 years and older
Reference category	Number of individuals in the household aged 15-64 years
eigentum	Type of residential property: proprietor
Reference category	Type of residential property: tenant, missing information
wnka_1	Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: none
wnka_3	Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: 11 and more
Reference category	Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: 1-10
hhincome_1	Household income: up to EUR 870
hhincome_2	Household income: EUR 871-1,400
hhincome_4	Household income: more than EUR 2,200
Reference category	Household income: EUR 1,401-2,200
alg2_1	UB II receipt of the household: current receipt of UB II
Reference category	UB II receipt of the household: no current receipt of UB II
stichprobe_ba	BA samples
Reference category	Microm sample
anzkon_1	Number of contact attempts CATI/CAPI: 1 contact attempt
anzkon_3	Number of contact attempts CATI/CAPI: 4-9 contact attempts
anzkon_4	Number of contact attempts CATI/CAPI: 10 and more contact attempts
Reference category	Number of contact attempts CATI/CAPI: 2-3 contact attempts

Table 35: **Variable overview, codes and reference categories for the logit models of the re-participating households (continued)**

Variable code and reference category	Explanation
blneualt_2	New federal states
Reference category	Old federal states
bundesld_1	Federal state: Schleswig-Holstein
bundesld_2	Federal state: Hamburg
bundesld_3	Federal state: Lower-Saxony
bundesld_4	Federal state: Bremen
bundesld_6	Federal state: Hesse
bundesld_7	Federal state: Rhineland-Palatinate
bundesld_8	Federal state: Baden-Wuerttemberg
bundesld_9	Federal state: Bavaria
bundesld_10	Federal state: Saarland
bundesld_11	Federal state: Berlin
bundesld_12	Federal state: Brandenburg
bundesld_13	Federal state: Mecklenburg-Vorpommern
bundesld_14	Federal state: Saxony
bundesld_15	Federal state: Saxony-Anhalt
bundesld_16	Federal state: Thuringia
Reference category	Federal state: North Rhine-Westphalia
bik_1	BIK size class of municipality: population of less than 2,000
bik_2	BIK size class of municipality: population of 2,000 to under 5,000
bik_3	BIK size class of municipality: population of 5,000 to under 20,000
bik_4	BIK size class of municipality: population of 20,000 to under 50,000
bik_5	BIK size class of municipality: population of 50,000 to under 100,000 STYP 2/3/4
bik_6	BIK size class of municipality: population of 50,000 to under 100,000 STYP 1
bik_7	BIK size class of municipality: population of 100,000 to under 500,000 STYP 2/ 3/ 4
bik_8	BIK size class of municipality: population of 100,000 to under 500,000 STYP 1
bik_9	BIK size class of municipality: population of 500,000 and more STYP 2/ 3/ 4
Reference category	BIK size class of municipality: population of 500,000 and more STYP 1

Table 36: **Logit models on re-participation for willingness to participate in a panel, availability and participation**

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

Table 36: Logit models on re-participation for willingness to participate in the panel, availability and participation (continued)

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

6.6 Propensity to participate – first-time interviewed split-off households

This step calculated the propensities to participate for new split-off households, i.e. households that are included in the panel due to the relocation of one individual of the panel sample in a new household. Here, only split-off households were considered that had not been interviewed in the previous waves. This means that the participation propensities for first-time participating split-off households were modelled separately following the criterion of split-off households originated in wave 4 (split-off W4 households) and split-off households originated in wave 5 (split-off W5 households). The probability of re-participation was estimated via logit models for availability and participation. Missing time-stable information on the household reference person (HRP) was added from the previous wave if necessary. The estimated propensities of the two models were multiplied. The reciprocal value of the product for the split-off households can also be found in the variable *hpbleib*.

Table 37: Variable overview, codes and reference categories for the logit models of the split-off households participating for the first time (wave 4 and wave 5)

Variable code and reference category	Explanation
alter_1	Household reference person (HRP) younger than 30 years
alter_2	HRP 30-39 years of age
alter_4	HRP 50-64 years of age
alter_5	HRP 65 years and older
Reference category	HRP 40-49 years of age
sex_1	HRP male
Reference category	HRP female
nichtdeutsch	HRP has nationality other than German
Reference category	HRP has German nationality or missing information
schulbil_1	School qualification HRP: no qualification
schulbil_2	School qualification HRP: lower secondary school
schulbil_4	School qualification HRP: college/university qualification
Reference category	School qualification HRP: intermediate secondary school/still pupil
stichprobe1	BA sample
stichprobe3	Refreshment sample (BA) wave 2
stichprobe4	Refreshment sample (BA) wave 3
stichprobe5	Refreshment sample (BA) wave 4
Reference category	Microm sample
anzkon_1	Number of contact attempts CATI/CAPI: 1 contact attempt
anzkon_3	Number of contact attempts CATI/CAPI: 4-9 contact attempts
anzkon_4	Number of contact attempts CATI/CAPI: 10 and more contact attempts
Reference category	Number of contact attempts CATI/CAPI: 2-3 contact attempts

Table 38: Logit models on the first participation of split-off wave 4 households for availability and participation

	Contact		Participation	
	Coef.	**p**	**Coef.**	**p**
alter_1	-.3169316	0.624	-.2019633	0.781
alter_2	.807183	0.350	-.597887	0.474
alter_4	.7431398	0.148	-1.09883	0.081
alter_5	.3700588	0.772	-.0205646	0.986
sex_1	-1.3948	0.004	-.1425814	0.806
nichtdeutsch	1.206919	0.304		
schulbil_1	-.7347448	0.446	.3429	0.773
schulbil_2	-.2239878	0.652	.1362247	0.809
schulbil_4	.5523845	0.384	-.5304055	0.471
anzkon_1	-1.720018	0.001		
anzkon_3	.9066009	0.234		
anzkon_4	-.7229558	0.561		
stichprobe1	-.9706771	0.055	-.3784742	0.477
stichprobe3	.4640556	0.686	-.1573479	0.858
stichprobe4	-1.772916	0.070		
cons	3.069574	0.000	-1.177896	0.024
n	208		154	
Log likelihood	-75.427106		-58.556792	
Pseudo R^2	0.1858		0.0454	

Table 39: Logit models on the first participation of split-off wave 5 households for availability and participation

	Contact		Participation	
	Coef.	**p**	**Coef.**	**p**
alter_1	-.169106	0.798	-.7357389	0.130
alter_2	-.2135482	0.731	-.2502378	0.553
alter_4	-.4650614	0.358	-.1883364	0.550
alter_5	-.8130421	0.372	-.8585624	0.305
sex_1	.8770132	0.062	-.1019223	0.718
nichtdeutsch			-.1523006	0.802
schulbil_1	-1.65523	0.024	.8945813	0.169
schulbil_2	-.3336857	0.492	.4011616	0.199
schulbil_4	-.6393656	0.281	.4509772	0.231
anzkon_1	-1.277773	0.007		
anzkon_3	-.8179033	0.135		
anzkon_4	1.002637	0.362		
stichprobe1	-.1322999	0.790	-.2529847	0.415
stichprobe3	-.0239096	0.978	-.4737019	0.443
stichprobe4	.4028368	0.651	-.3263525	0.551
stichprobe5	-1.185011	0.097	-.2990074	0.637
cons	3.313871	0.000	-.849377	0.008
n	338		307	
Log likelihood	-92.172315		-175.50534	
Pseudo R^2	0.1103		0.0205	

6.7 Non-response weighting for households from the BA refreshment sample and the BA panel replenishment sample of wave 5

Again a tow-stage nonresponse modelling for the households from the refreshment sample of BA new inflows into UB II receipt (sample = 8) and for the first time also for households of a replenishment sample to the BA panel (sample = 7) was performed (availability and participation) similar to the wave 4 refreshment sample. The participation probability derived from this can be found in variable *prop_t0*. Since the set of variables and the origin of both subsamples is the same, only one joint non-response model was calculated with a subsample code as control variable.

Table 40: Variable overview, codes and reference categories for the logit models of the BA refreshment sample and BA replenishment sample of wave 5

Variable code and reference category	Explanation
alter_1	Household reference person (HRP) younger than 30 years
alter_2	HRP 30-39 years of age
alter_4	HRP 50-64 years of age
Reference category	HRP 40-49 years of age
sex_1	HRP male
Reference category	HRP female
nichtdeutsch	HRP has nationality other than German
Reference category	HRP has German nationality or missing information
stichprobe2	Refreshment sample (BA new inflows) wave 5
Reference category	Replenishment sample (BA panel) wave 5
schulbil_1	School qualification HRP: no qualification
schulbil_2	School qualification HRP: lower secondary school
schulbil_4	School qualification HRP: college/university qualification
schulbil_5	School qualification HRP: Details refused
Reference category	School qualification HRP: intermediate secondary school/still pupil
anz_persBG_2	Number of individuals in the benefit unit: 2 individuals
anz_persBG_3	Number of individuals in the benefit unit: 3 and more individuals
Reference category	Number of individuals in the benefit unit: 1 individual
anz_verwfBG_1	Number of individuals capable of work in the benefit unit: none
anz_verwfBG_3	Number of individuals capable of work in the benefit unit: 2 individuals
anz_verwfBG_4	Number of individuals capable of work in the benefit unit: 3 and more individuals
Reference category	Number of individuals capable of work in the benefit unit: 1 individual
BG_typ_2	Type of benefit unit: single parent
BG_typ_3	Type of benefit unit: couple without children
BG_typ_4	Type of benefit unit: couple with children under the age of 18
BG_typ_5	Type of benefit unit: other benefit unit
Reference category	Type of benefit unit: single
famstand_2	Marital status: married
famstand_3	Marital status: widowed
famstand_4	Marital status: divorced
famstand_5	Marital status: separated
famstand_6	Marital status: relationship similar to a marriage or registered partnership
Reference category	Marital status: single

Table 40: Variable overview, codes and reference categories for the logit models of the BA refreshment sample and BA replenishment sample of wave 5 (continued)

Variable code and reference category	Explanation
bundesld_1	Federal state: Schleswig-Holstein
bundesld_2	Federal state: Hamburg
bundesld_3	Federal state: Lower-Saxony
bundesld_4	Federal state: Bremen
bundesld_6	Federal state: Hesse
bundesld_7	Federal state: Rhineland-Palatinate
bundesld_8	Federal state: Baden-Wuerttemberg
bundesld_9	Federal state: Bavaria
bundesld_10	Federal state: Saarland
bundesld_11	Federal state: Berlin
bundesld_12	Federal state: Brandenburg
bundesld_13	Federal state: Mecklenburg-Vorpommern
bundesld_14	Federal state: Saxony
bundesld_15	Federal state: Saxony-Anhalt
bundesld_16	Federal state: Thuringia
Reference category	Federal state: North Rhine-Westphalia
bik_1	BIK size class of municipality: population of less than 2,000
bik_2	BIK size class of municipality: population of 2,000 to under 5,000
bik_3	BIK size class of municipality: population of 5,000 to under 20,000
bik_4	BIK size class of municipality: population of 20,000 to under 50,000
bik_5	BIK size class of municipality: population of 50,000 to under 100,000 STYP 2/3/4
bik_6	BIK size class of municipality: population of 50,000 to under 100,000 STYP 1
bik_7	BIK size class of municipality: population of 100,000 to under 500,000 STYP 2/3/4
bik_8	BIK size class of municipality: population of 100,000 to under 500,000 STYP 1
bik_9	BIK size class of municipality: population of 500,000 and more STYP 2/ 3/ 4
Reference category	BIK size class of municipality: population of 500,000 and more STYP 1
anzkon_1	Number of contact attempts CATI/CAPI: 1 contact attempt
anzkon_3	Number of contact attempts CATI/CAPI: 4-9 contact attempts
anzkon_4	Number of contact attempts CATI/CAPI: 10 and more contact attempts
Reference category	Number of contact attempts CATI/CAPI: 2-3 contact attempts

Table 41: Logit models on the first participation for availability and participation of the BA refreshment sample and BA replenishment sample of wave 5

	Contact		Participation	
	Coef.	**p**	**Coef.**	**p**
alter_1	-.0015206	0.992	-.0692774	0.402
alter_2	.2276912	0.139	-.0437775	0.582
alter_4	.246158	0.126	.1627027	0.037
sex_1	-.171073	0.142	-.010268	0.863
nichtdeutsch	-.2467698	0.079	-.3465132	0.000
stichprobe2	.3263159	0.005	.1669613	0.004
schulbil_1	-.4839556	0.005	-.283361	0.002
schulbil_2	-.3921248	0.009	-.1049563	0.141
schulbil_4	-.1269403	0.525	.1992542	0.031
schulbil_5	-.404894	0.040	-.1888125	0.055
anz_persBG_2	1.137636	0.021	.1648316	0.474
anz_persBG_3	.9215946	0.090	.3766071	0.145
anz_verwfBG_1	.1733314	0.833	-.3751382	0.510
anz_verwfBG_3	.2088258	0.494	-.0132664	0.916
anz_verwfBG_4	.6764131	0.198	.0427438	0.824
BG_typ_2	-.5693997	0.265	.0542831	0.819
BG_typ_3	-.7152423	0.168	-.0431914	0.852
BG_typ_4	-.2865151	0.603	-.0150196	0.951
BG_typ_5	-.4500966	0.230	-.0384718	0.861
famstand_2	.3633272	0.113	-.1309437	0.233
famstand_3	.7126889	0.244	-.0205194	0.927
famstand_4	.3735799	0.032	.1744604	0.039
famstand_5	.1017989	0.585	-.2219197	0.031
famstand_6	-.2185996	0.425	-.0498226	0.726
bundesld_1	-.441315	0.102	.0290337	0.858
bundesld_2	-.2615607	0.294	.0738912	0.623
bundesld_3	-.046195	0.823	-.0430577	0.675
bundesld_4	.8261968	0.088	-.1091292	0.636
bundesld_6	-.8061113	0.000	-.2272539	0.084
bundesld_7	-.9465943	0.000	.20312	0.172
bundesld_8	.115628	0.605	-.0744395	0.491
bundesld_9	.0074188	0.970	-.0037252	0.971
bundesld_10	.523671	0.479	.8561815	0.001
bundesld_11	.2789074	0.188	-.3266551	0.007
bundesld_12	.6622326	0.053	-.0233711	0.847
bundesld_13	-.5655013	0.079	-.0466962	0.807
bundesld_14	.5755524	0.098	.2690728	0.029
bundesld_15	.0071509	0.981	.0044224	0.975
bundesld_16	-.472969	0.101	.0559909	0.703
bik10_1	2.328615	0.024	-.2967277	0.295
bik10_2	.2433743	0.509	.0324813	0.877
bik10_3	.3823661	0.129	.1538863	0.197
bik10_4	1.059255	0.000	.0331483	0.741
bik10_5	.2207301	0.361	.4259753	0.001
bik10_6	.7946102	0.032	.6038444	0.000
bik10_7	.0115889	0.949	.1962604	0.060
bik10_8	.4518842	0.005	.0162977	0.849
bik10_9	.7751251	0.008	.2964757	0.011
anzkon_1	-1.814473	0.000		
anzkon_3	-.7893974	0.001		
anzkon_4	-2.369228	0.000		
cons	4.057722	0.000	-1.13626	0.000
n	8220		7763	
Log likelihood	-1500.1312		-4406.527	
Pseudo R^2	0.1499		0.0220	

6.8 Non-response weighting for households from the wave 5 EWO replenishment sample

A two-stage non-response modelling (availability and participation) was done for the households from the EWO replenishment sample for the general population drawn for the first time (sample = 6). The participation probability derived from this can be found in variable *prop_t0*.

Table 42: Variable overview, codes and reference categories for the logit models of the EWO replenishment sample of wave 5

Variable code and reference category	Explanation
alter_1 alter_2 alter_4 alter_5 Reference category	15-29 years of age 30-39 years of age 50-64 years of age 65+ years of age 40-49 years of age
sex_1 Reference category	HRP male HRP female
nichtdeutsch Reference category	HRP has nationality other than German HRP has German nationality or missing information
blneualt_2 Reference category	New federal states Old federal states
bundesld_1 bundesld_2 bundesld_3 bundesld_4 bundesld_6 bundesld_7 bundesld_8 bundesld_9 bundesld_10 bundesld_11 bundesld_12 bundesld_13 bundesld_14 bundesld_15 bundesld_16 Reference category	Federal state: Schleswig-Holstein Federal state: Hamburg Federal state: Lower-Saxony Federal state: Bremen Federal state: Hesse Federal state: Rhineland-Palatinate Federal state: Baden-Wuerttemberg Federal state: Bavaria Federal state: Saarland Federal state: Berlin Federal state: Brandenburg Federal state: Mecklenburg-Vorpommern Federal state: Saxony Federal state: Saxony-Anhalt Federal state: Thuringia Federal state: North Rhine-Westphalia
anzkon_1 anzkon_3 anzkon_4 Reference category	Number of contact attempts CATI/CAPI: 1 contact attempt Number of contact attempts CATI/CAPI: 4-9 contact attempts Number of contact attempts CATI/CAPI: 10 and more contact attempts Number of contact attempts CATI/CAPI: 2-3 contact attempts

Table 43: **Logit models on the first participation for availability and participation of the wave 5 EWO replenishment sample**

	Contact		Participation	
	Coef.	**p**	**Coef.**	**p**
alter_1	-.1734983	0.252	-.2048812	0.053
alter_2	-.2268192	0.147	-.4239397	0.000
alter_4	.2374297	0.104	.2044171	0.021
alter_5	.7053023	0.000	-.0647211	0.477
sex_1	-.050245	0.609	.1124317	0.065
nichtdeutsch	-.4406449	0.002	-.5264382	0.000
blneualt_2	-.3393163	0.003		
bundesld_1			-.5270543	0.039
bundesld_2			.1204216	0.629
bundesld_3			.1418745	0.204
bundesld_4			-.4679943	0.169
bundesld_6			-.2911917	0.038
bundesld_7			.2796866	0.048
bundesld_8			-.2024863	0.060
bundesld_9			.2596503	0.009
bundesld_10			.5289703	0.051
bundesld_11			-.2185617	0.217
bundesld_12			-.1186459	0.468
bundesld_13			-.90868	0.004
bundesld_14			-.0081744	0.960
bundesld_15			.4191263	0.016
bundesld_16			.133729	0.351
anzkon_1	-1.337765	0.000		
anzkon_3	-.7611378	0.000		
anzkon_4	-.1804869	0.401		
cons	3.201539	0.000	-1.030231	0.000
n	6,237		5,769	
Log likelihood	-1569.4437		-3249.2651	
Pseudo R^2	0.0557		0.0203	

6.9 Propensity to participate again – individuals

The decisive longitudinal weight is not the one at the household level but the one at the individual level, as the units here are stable over time. The propensities to participate again for individuals in wave 5 were estimated including additional personal characteristics via logit models for willingness to participate in the panel, availability and participation. The dependence of the personal sample conveyed via the household context and correction of the estimation of standard errors made necessary by it was considered in these models by clustering the error terms at the household level. The predicted propensities of the models were again multiplied. The reciprocal value of this product can be found in variable *ppbleib*. The longitudinal weight for an individual for the period $[t_1; t_2; t_3; t_4; t_5]$ across all five waves can be obtained as the product of the cross-sectional weight to t_1, *ppbleib* (wave 1 to wave 2) and *ppbleib* (wave 2 to wave 3, etc.).

Table 44: Variable overview, codes and reference categories for the logit models of re-participating individuals

Variable code and reference category	Explanation
alter_1	Individual younger than 30 years
alter_2	Individual 30-39 years of age
alter_4	Individual 50-64 years of age
alter_5	Individual 65 years and older
Reference category	Individual 40-49 years of age
sex_1	Individual male
Reference category	Individual female
nichtdeutsch	Individual has nationality other than German
Reference category	Individual has German nationality or missing information
schulbil_1	School qualification individual: no qualification
schulbil_2	School qualification individual: lower secondary school
schulbil_4	School qualification individual: college/university qualification
Reference category	School qualification individual: intermediate secondary school/still pupil
gesundheit_3	Subjective evaluation of the health state of the individual: satisfactory
gesundheit_4	Subjective evaluation of the health state of the individual: not so good
gesundheit_5	Subjective evaluation of the health state of the individual: bad
Reference category	Subjective evaluation of the health state of the individual: very good to good
zufrieden_1	General life satisfaction of the individual: scale value 0-2
zufrieden_2	General life satisfaction of the individual: scale value 3-5
zufrieden_3	General life satisfaction of the individual: scale value 6-8
Reference category	General life satisfaction of the individual: scale value 9-10
anz_0_3	Number of individuals in the household aged 0-3 years
anz_4_6	Number of individuals in the household aged 4-6 years
anz_7_14	Number of individuals in the household aged 7-14 years
anz_65	Number of individuals in the household aged 65 years and older
Reference category	Number of individuals in the household aged 15-64 years
eigentum	Type of residential property: proprietor
Reference category	Type of residential property: tenant, missing information
wnka_1	Number of "don't know" and "details refused" responses in household and personal interviews of the individual: none
wnka_3	Number of "don't know" and "details refused" responses in household and personal interviews of the individual: 11 and more
Reference category	Number of "don't know" and "details refused" responses in household and personal interviews of the individual: 1-10
hhincome_1	Household income: up to EUR 870
hhincome_2	Household income: EUR 871-1,400
hhincome_4	Household income: more than EUR 2,200
Reference category	Household income: EUR 1,401-2,200
alg2_1	UB II receipt of the household: current receipt of UB II
Reference category	UB II receipt of the household: no current receipt of UB II
stichprobe1	BA sample
stichprobe3	Refreshment sample (BA) wave 2
stichprobe4	Refreshment sample (BA) wave 3
stichprobe5	Refreshment sample (BA) wave 4
Reference category	Microm sample
anzkon_1	Number of contact attempts CATI/CAPI: 1 contact attempt
anzkon_3	Number of contact attempts CATI/CAPI: 4-9 contact attempts
anzkon_4	Number of contact attempts CATI/CAPI: 10 and more contact attempts
Reference category	Number of contact attempts CATI/CAPI: 2-3 contact attempts

Table 44: Variable overview, codes and reference categories for the logit models of the re-participants (continued)

Variable code and reference category	Explanation
blneualt_2	New federal states
Reference category	Old federal states
bundesld_1	Federal state: Schleswig-Holstein
bundesld_2	Federal state: Hamburg
bundesld_3	Federal state: Lower-Saxony
bundesld_4	Federal state: Bremen
bundesld_6	Federal state: Hesse
bundesld_7	Federal state: Rhineland-Palatinate
bundesld_8	Federal state: Baden-Wuerttemberg
bundesld_9	Federal state: Bavaria
bundesld_10	Federal state: Saarland
bundesld_11	Federal state: Berlin
bundesld_12	Federal state: Brandenburg
bundesld_13	Federal state: Mecklenburg-Vorpommern
bundesld_14	Federal state: Saxony
bundesld_15	Federal state: Saxony-Anhalt
bundesld_16	Federal state: Thuringia
Reference category	Federal state: North Rhine-Westphalia
bik_1	BIK size class of municipality: population of less than 2,000
bik_2	BIK size class of municipality: population of 2,000 to under 5,000
bik_3	BIK size class of municipality: population of 5,000 to under 20,000
bik_4	BIK size class of municipality: population of 20,000 to under 50,000
bik_5	BIK size class of municipality: population of 50,000 to under 100,000 STYP 2/3/4
bik_6	BIK size class of municipality: population of 50,000 to under 100,000 STYP 1
bik_7	BIK size class of municipality: population of 100,000 to under 500,000 STYP 2/3/4
bik_8	BIK size class of municipality: population of 100,000 to under 500,000 STYP 1
bik_9	BIK size class of municipality: population of 500,000 and more STYP 2/ 3/ 4
Reference category	BIK size class of municipality: population of 500,000 and more STYP 1

Table 45: Logit models on re-participation for willingness to participate in a panel, availability and participation

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

| zufrieden_3 | -.6356465 | 0.417 | -.2344541 | 0.129 | -.0029418 | 0.967 |

Table 45: Logit models on re-participation for willingness to participate in the panel, availability and participation (continued)

	Willingness to participate in the panel		Contact		Participation	
	Coef.	p	Coef.	p	Coef.	p
anz_0_3	-.125206	0.793	-.1034997	0.527	-.0185916	0.829
anz_4_6	-.3209503	0.569	.0826405	0.678	.0741661	0.417
anz_7_14	-.3420701	0.410	-.0797108	0.503	.0125192	0.823
anz_65	-1.270068	0.007	.4417246	0.171	.140469	0.241
eigentum	-.6838766	0.242	.7462691	0.001	.0186042	0.831
wnka_1	-.5217555	0.255	.0493292	0.660	.2954034	0.000
wnka_3	-1.547987	0.000	.0660804	0.677	-.1146345	0.163
hhincome_1	-.958492	0.181	-.2721554	0.127	.1612614	0.106
hhincome_2	-.5737491	0.344	-.0398392	0.815	.0505144	0.562
hhincome_4	.9467236	0.246	.34564	0.124	.2493001	0.008
alg2_1	-.6571527	0.299	.2022105	0.161	.0431921	0.584
stichprobe1			-.6368302	0.001	-.1063695	0.252
stichprobe3			-.8485139	0.000	-.1634342	0.231
stichprobe4			-.7266907	0.002	-.1135386	0.373
stichprobe5			-.5145691	0.036	-.3092601	0.011
blneualt_2	1.326077	0.010				
anzkon_1			-.5881563	0.014		
anzkon_3			-.3647335	0.048		
anzkon_4			-1.366091	0.000		
bundesld_1			-.8153592	0.002	-.6512667	0.000
bundesld_2			-.2045841	0.626	.0130608	0.956
bundesld_3			.0135992	0.961	-.3224352	0.005
bundesld_4			-.0781003	0.891	-.10384	0.738
bundesld_6			-.5307905	0.041	-.2260133	0.094
bundesld_7			.1321791	0.731	-.5868427	0.000
bundesld_8			-.4056052	0.089	-.4287236	0.000
bundesld_9			-.1506979	0.523	-.2695108	0.015
bundesld_10			.2442566	0.653	-.3633051	0.181
bundesld_11			-.8108077	0.003	-.131471	0.384
bundesld_12			.0565758	0.881	-.1744911	0.291
bundesld_13			-.0330389	0.939	.2577315	0.250
bundesld_14			-.4681859	0.112	.4859923	0.006
bundesld_15			-.2825624	0.351	-.3608326	0.018
bundesld_16			.3102048	0.446	-.0823932	0.656
bik_1			.5100487	0.630	.5687608	0.219
bik_2			.5193127	0.297	.221327	0.252
bik_3			-.0797132	0.786	.1251445	0.347
bik_4			-.5064409	0.034	.3166536	0.013
bik_5			-.3559797	0.180	.3606542	0.006
bik_6			-.2955571	0.412	.174933	0.358
bik_7			-.2154511	0.362	.3458393	0.003
bik_8			-.2644422	0.187	.1943944	0.053
bik_9			-.0031591	0.993	.2181262	0.107
cons	8.899969	0.000	4.885233	0.000	1.442663	0.000
n	11768		11730		11252	
Log likelihood	-224.70204		-1772.3739		-5215.4557	
Pseudo R^2	0.1219		0.1129		0.0465	

Note: The correction of standard errors was made by means of an estimation clustered across households.

6.10 Integration of the weights to yield the total weight before calibration

This step again involved combining the household weights of the new replenishment and panel household samples (including the refreshments from waves 2 to 4) which were modified by the non-response modelling. The integration of the panel supplements in the respective panels thus occurred by a convex combination of the respective samples. The multiple selection probability of a sampled benefit recipient who was living in the same household as benefit recipient in the previous years but without being a member of the benefit unit himself/herself was ignored again. The new design weights of the benefit recipient sample are projected in the cross-section to all individuals who were living in a household including at least one benefit unit in either 7/2006, in 7/2007, in 7/2008, in 7/2009 or in 7/2010. It is only when calculating new weights for the total sample that it becomes necessary to adjust the weights for all households in receipt of benefits in 7/2010. For this adjustment the inclusion probability in the respective other sample was estimated for cases from the Microm sample (wave 1) and the refreshment sample (wave 5). For cases from the refreshment sample, the mean wave 1 selection probability in the Microm sample in the respective postcode area and the average participation probability (for wave 1, wave 2, wave 3, wave 4 and wave 5) in that sample were assumed. For cases from the Microm sample, if they are (according to survey data) new recipients of Unemployment Benefit II who first received the benefit between the last four sampling dates (wave 2, wave 3, wave 4 and wave 5), the mean selection probability of a household in the refreshment sample in the respective postcode area and the average participation probability in that sample were assumed. The two weights were then integrated to form a new total weight.

6.11 Integration of temporary non-responses (households)

Households that skipped one wave, i.e. did not participate (temporary non-responses), could participate again in wave 5 as was possible in wave 4. No longitudinal weights are calculated for these households, i.e. (weighted) longitudinal evaluations can only be made with participants across all waves in question. Non-participation of a household can only occur in one wave; if a household skips two consecutive waves, it will no longer be contacted. In order to calculate mutual cross-sectional weights including the temporary non-responses, there was a convex combination of the modified household weights of the temporary non-responses and the modified household weights of the panel household sample (not of the refreshment sample) before calibration. The convex combination of the household weights was hence made before calibration; the calibration was then made with the new combined household weights.

Although the household weights modified by non-response modelling already serve as projection factors for the panel and refreshment sample, it was necessary to calculate such modified household weights as estimator for the respective population again for the temporary non-responses. The starting point was the calibrated household weights of wave 3 (wave 4 is the temporary non-response).

For temporary non-responses the probability of non-participation in wave 4 in case of participation in wave 3 (non-participation propensities wave 4) and the probability of participa-

tion in wave 5 in case of a non-participation in wave 4 (participation propensities wave 5) was determined. The probability of non-participation in wave 4 is calculated from 1– participation probability in wave 4.

The described propensities for participation and non-participation were estimated via logit models. The estimated probabilities of the respective models were multiplied. The modified household weight of the temporary non-responses was then calculated by multiplying the calibrated household weights of wave 2 by the reciprocal value of this product.

Table 46: **Variable overview, codes and reference categories for the logit models of the temporary non-responses**

Variable code and reference category	Explanation
panelber	Willingness to participate in the panel (objection proceedings from W3 to W4 considered)
alter_1	Household reference person (HRP) younger than 30 years
alter_2	HRP 30-39 years of age
alter_4	HRP 50-64 years of age
alter_5	HRP 65 years and older
Reference category	HRP 40-49 years of age
sex_1	HRP male
Reference category	HRP female
nichtdeutsch	HRP has nationality other than German
Reference category	HRP has German nationality or missing information
sprache_nichtdeutsch	Language HRP: primarily spoken language in the household is not German
Reference category	Language HRP: primarily spoken language in the household is German
schulbil_1	School qualification HRP: no qualification
schulbil_2	School qualification HRP: lower secondary school
schulbil_4	School qualification HRP: college/university qualification
Reference category	School qualification HRP: intermediate secondary school/still pupil
gesundheit_3	Subjective evaluation of the health state of the HRP: satisfactory
gesundheit_4	Subjective evaluation of the health state of the HRP: not so good
gesundheit_5	Subjective evaluation of the health state of the HRP: bad
Reference category	Subjective evaluation of the health state of the HRP: very good to good
zufrieden_1	General life satisfaction HRP: scale value 0-2
zufrieden_2	General life satisfaction HRP: scale value 3-5
zufrieden_3	General life satisfaction HRP: scale value 6-8
Reference category	General life satisfaction HRP: scale value 9-10
anz_0_3	Number of individuals in the household aged 0-3 years
anz_4_6	Number of individuals in the household aged 4-6 years
anz_7_14	Number of individuals in the household aged 7-14 years
anz_65	Number of individuals in the household aged 65 years and older
DinvalidAge	Age responses that cannot be evaluated
Reference category	Number of individuals in the household aged 15-64 years
eigentum	Type of residential property: proprietor
Reference category	Type of residential property: tenant, missing information

Table 46: **Variable overview, codes and reference categories for the logit models of the temporary non-responses (continued)**

Variable code and reference category	Explanation
wnka_1	Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: none
wnka_3	Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: 11 and more
Reference category	Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: 1-10
hhincome_1	Household income: up to EUR 870
hhincome_2	Household income: EUR 871-1,400
hhincome_4	Household income: more than EUR 2,200
Reference category	Household income: EUR 1,401-2,200
alg2_1	UB II receipt of the household: current receipt of UB II
Reference category	UB II receipt of the household: no current receipt of UB II
bundesld_1	Federal state: Schleswig-Holstein
bundesld_2	Federal state: Hamburg
bundesld_3	Federal state: Lower-Saxony
bundesld_4	Federal state: Bremen
bundesld_6	Federal state: Hesse
bundesld_7	Federal state: Rhineland-Palatinate
bundesld_8	Federal state: Baden-Wuerttemberg
bundesld_9	Federal state: Bavaria
bundesld_10	Federal state: Saarland
bundesld_11	Federal state: Berlin
bundesld_12	Federal state: Brandenburg
bundesld_13	Federal state: Mecklenburg-Vorpommern
bundesld_14	Federal state: Saxony
bundesld_15	Federal state: Saxony-Anhalt
bundesld_16	Federal state: Thuringia
Reference category	Federal state: North Rhine-Westphalia
bik_1	BIK size class of municipality: population of less than 2,000
bik_2	BIK size class of municipality: population of 2,000 to under 5,000
bik_3	BIK size class of municipality: population of 5,000 to under 20,000
bik_4	BIK size class of municipality: population of 20,000 to under 50,000
bik_5	BIK size class of municipality: population of 50,000 to under 100,000 STYP 2/3/4
bik_6	BIK size class of municipality: population of 50,000 to under 100,000 STYP 1
bik_7	BIK size class of municipality: population of 100,000 to under 500,000 STYP 2/3/4
bik_8	BIK size class of municipality: population of 100,000 to under 500,000 STYP 1
bik_9	BIK size class of municipality: population of 500,000 and more STYP 2/ 3/ 4
Reference category	BIK size class of municipality: population of 500,000 and more STYP 1

Table 47:   Logit models of temporary non-responses

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

Table 47: **Logit models of temporary non-responses (continued)**

	Re-participation in wave 4 to determine the W4 non-participation probability (1-participation probability W4)		Re-participation in wave 5 in case of non-participation in wave 4	
	Coef.	**p**	**Coef.**	**p**
bik_1	.1435061	0.610	-.3076576	0.632
bik_2	.0894075	0.616	-.3209982	0.490
bik_3	.0932077	0.451	.2167444	0.422
bik_4	.0531367	0.649	-.2123762	0.453
bik_5	.0524389	0.672	.2087075	0.459
bik_6	.2081499	0.246	-.3048071	0.495
bik_7	-.1223254	0.232	.337995	0.127
bik_8	-.0280089	0.754	-.1502369	0.467
bik_9	.1887004	0.154	-.3209531	0.349
cons	-4.947284	0.000	-2.283381	0.000
n	9535		2892	
Log likelihood	-4168.0111		-890.03657	
Pseudo R^2	0.2875		0.0461	

The convex combination of the weights of the participants across all waves (panel household sample) and the temporary non-responses was made for the weights of all three subsamples i (Microm, BA and total) by multiplying the respective modified household weights by the share of the panel household sample or the temporary non-responses from the total sample, i.e. the sum of the panel household sample and temporary non-responses:

$$dw_i hh_{temp.non\ reps.} \times \frac{n_{temp.\ non\ resp.i}}{n_{temp.non\ resp.\ i} + n_{panel\ household\ sample\ i}}$$ for temporary non-responses and

$$dw_i hh_{panel\ household\ sample} \times \frac{n_{panel\ household\ sample\ i}}{n_{temp.non\ resp.\ i} + n_{panel\ household\ sample\ i}}$$ for the panel household sample.

6.12 Calibration to the household weight, wave 5, cross-section

Following that came another calibration of the modified design weights including the non-response weighting at the household level using the GREG procedure to the benchmark values of the Federal Statistical Office for 2010. For households in receipt of benefits the weights were adjusted to the statistics of the Federal Employment Agency for July 2010. As in the previous year, also the increase in Unemployment Benefit II receipt since the previous year at the level of benefit units (367,870) was also included as an additional benchmark value in the total sample. Those cases in the previous samples from wave 1 to wave 5 which, according to wave 5 of the survey, were receiving Unemployment Benefit II in July 2010 will be projected to the benchmark statistics of the Federal Employment Agency on receipt of Unemployment Benefit II.

The main objective of weighting is to balance distortions arising from the sample design (with different selection probabilities) and through selective participation or non-

participation. By using the weights, population values from the sample can be estimated in an unbiased way. If the weights show a high variance, this can lead to a large variance of the estimation functions. This is the trade-off between bias and variance so typical for statistics. The weighting reduces the bias; however, a too severe increase in the variance caused by weighting is to be avoided, too. Therefore, attempts are made to avoid very large weighting factors (and subsequently also very small factors) whenever possible and make appropriate corrections to the weights if necessary. Within the framework of the calibration at hand, this was made in two points:

- The input weights for the calibration (the modified design weights after considering non-response analyses) were trimmed before calibration, i.e. they were replaced by new input weights. The maximum and minimum of the trimmed design weights was determined by using certain percentiles of the distribution depending on the distribution of the design weights.

- Also the interval of weights was limited during calibration, i.e. a maximum and a minimum limit for weights was determined. Here also the total width of the weights was determined; the range of the pure calibration weights can be calculated from the relation of original weights to the trimmed input weight. It had to be observed here that narrower limits for the weights result in less variance of the weights and thus less variance of the estimations; too narrow limits can, however, make the calibration of all benchmark values impossible.

To evaluate the weights, the following describes besides the average value and the standard deviation also the efficiency measure (E). The efficiency measure E is based on the variance of the weighting factor. The efficiency measure indicates in percent of the conducted case number how large the effective case number of a passive characteristic which does not correlate with active characteristics is when using the weight. The effective case number is the number of respondents who would have produced the same sample error in an unlimited random sample given the variance of the characteristic in the sample. The efficiency measure expresses the relation of n to n' as percentage.

6.12.1 Calibration of the BA sample

The population of the cumulated BA sample of all five waves consists of all households in Germany with at least one benefit unit receiving benefits in accordance with SGB II at one of the, up to now, five drawing dates (in July 2006, July 2007, July 2008, July 2009 or July 2010). In wave 5, only the benchmark values of the BA statistics from July 2010 are calibrated. The calibration thus only influences the weights of the households from the BA sample in which at least one benefit unit receiving benefits in accordance with SGB II was living in July 2010. Starting point for the calibration were modified design weights including the non-response weighting. The modified design weights were trimmed at the 5 % percentile and the 95 % percentile of their distribution and after that rescaled in such a way that their total again resulted in the total of the untrimmed design weights. The projection factors of the trimmed design weights range from 242.61 to 1976.16. The relation between the total projection factors after calibration and the trimmed design weights was limited

downwards to 0.3 and upwards to 2.0. Thus, the total projection factors after calibration lie between min. 72.78 and max. 3014.12.

A calibration was made for the following characteristics:

Benefit units basis BA statistics:

- Increase in BU Unemployment Benefit II recipients
- Number BCs receiving benefits in accordance with SGB II by federal states
- Number of BCs receiving benefits in accordance with SGB II by number of individuals under 65 years of age in the benefit unit, by west/east
- Number of BCs receiving benefits in accordance with SGB II by number of children under 15 years of age in the benefit unit, by west/east
- Number of BCs receiving benefits in accordance with SGB II consisting of a single parent with child(ren), by west/east

As in the previous year, an additional benchmark was included; this is the increase in Unemployment Benefit II recipients since the previous year at the level of benefit units (367,870).

For the calibration, each benchmark variable for each household must have a valid value. Therefore, the very low non-response item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Since the imputation only serves the feasibility of the calibration, the imputed values were set back to missing values after the calibration. A projection with the calibrated weights without considering the non-response item thus leads to slight deviations from the values as presented in the following.

Table 48: Nominal distributions and distributions after calibration (BA sample, households)

Benchmark figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA statistics	Distribution with calibrated weights
Number BCs receiving benefits in accordance with SGB II by federal states (16 categories)	Number BCs Schleswig-Holstein	161	122,816	122,816
	Number BCs Hamburg	138	108,473	108,473
	Number BCs Lower-Saxony	436	334,073	334,073
	Number BCs Bremen	51	51,949	51,949
	Number BCs North Rhine-Westphalia	1,080	849,532	849,532
	Number BCs Hesse	217	219,564	219,564
	Number BCs Rhineland-Palatinate	151	123,522	123,522
	Number BCs Baden-Wuerttemberg	314	256,618	256,618
	Number BCs Bavaria	387	267,375	267,375
	Number BCs Saarland	71	44,787	44,787
	Number BCs Berlin	303	334,106	334,106
	Number BCs Brandenburg	272	167,160	167,160
	Number BCs Mecklenburg-Vorpommern	129	127,174	127,174
	Number BCs Saxony	339	280,039	280,039
	Number BCs Saxony-Anhalt	261	186,400	186,400
	Number BCs Thuringia	174	130,035	130,035
Number of BCs receiving benefits in accordance with SGB II by number of individuals under 65 years of age in the benefit unit (1, 2, 3, 4, and "5 or more") and by west/east (10 categories)	Number BCs with 1 individual under 65 (west)	1,225	1,287,479	1,287,479
	Number BCs with 2 individuals under 65 (west)	792	480,035	480,035
	Number BCs with 3 individuals under 65 (west)	506	297,590	297,590
	Number BCs with 4 individuals under 65 (west)	285	185,355	185,355
	Number BCs with 5 or more individuals under 65 (west)	198	128,250	128,250
	Number BCs with 1 individual under 65 (east)	682	714,572	714,572
	Number BCs with 2 individuals under 65 (east)	422	264,338	264,338
	Number BCs with 3 individuals under 65 (east)	221	135,917	135,917
	Number BCs with 4 individuals under 65 (east)	90	71,425	71,425
	Number BCs with 5 or more individuals under 65 (east)	63	38,662	38,662

Table 48: Nominal distributions and distributions after calibration (BA sample, house-holds) (continued)

Benchmark figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA statistics	Distribution with calibrated weights
Number of BCs receiving benefits in accordance with SGB II by number of individuals under 15 years of age in the benefit unit (0, 1, 2, 3, "4 or more") and by west/east (10 categories)	Number BCs without children under 15 years of age (west)	1,997	1,629,124	1,629,124
	Number BCs with 1 child under 15 years of age (west)	550	406,480	406,480
	Number BCs with 2 children under 15 years of age (west)	325	230,977	230,977
	Number BCs with 3 children under 15 years of age (west)	93	80,359	80,359
	Number BCs with 4 or more children under 15 years of age (west)	41	31,769	31,769
	Number BCs without children under 15 years of age (east)	1,095	909,771	909,771
	Number BCs with 1 child under 15 years of age (east)	247	188,910	188,910
	Number BCs with 2 children under 15 years of age (east)	100	90,185	90,185
	Number BCs with 3 children under 15 years of age (east)	23	25,972	25,972
	Number BCs with 4 or more children under 15 years of age (east)	13	10,076	10,076
Number BCs receiving benefits in accordance with SGB II consisting of a single parent with children by west/east (4 categories)	Number BCs with a single parent (west)	715	450,759	450,759
	Rest BCs without a single parent (west)	2,291	1,927,950	1,927,950
	Number BCs with a single parent (east)	258	189,640	189,640
	Rest BCs without a single parent (east)	1,220	1,035,274	1,035,274

Table 49:   Parameters of distribution of weights

1% percentile	105.1712
5% percentile	173.7548
10% percentile	239.6466
25% percentile	372.7158
50% percentile	682.7329
75% percentile	1,085.275
90% percentile	1,588.228
95% percentile	2,006.884
99% percentile	2,652.644
Mean	813.7423
Standard deviation	567.7689
Minimum	72.78169
Maximum	3,014.117
Number of observations	4,482
Efficiency measure	**66.8%**

6.12.2 Population sample

All private households in Germany form the population. Starting point for the calibration were modified design weights including the non-response weighting. The modified design weights were trimmed at the 5 % percentile and the 95 % percentile of their distribution and after that rescaled in such a way that their total again resulted in the total of the un-trimmed design weights. The projection factors of the trimmed design weights range from 2,775.61 to 2,7305.53. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.1 and upwards to 4.0. Thus, the total projection factors after calibration lie between minimal 277.56 and maximal 79,154.33.

A calibration was made for the following characteristics:

Benefit units: basis BA statistics:

• Number BCs receiving benefits in accordance with SGB II by federal states

• Number of BCs receiving benefits in accordance with SGB II by number of individuals under 65 years of age in the benefit unit, by west/east

• Number of BCs receiving benefits in accordance with SGB II by number of children under 15 years of age in the benefit unit, by west/east

• Number of BCs receiving benefits in accordance with SGB II consisting of a single parent with child(ren), by west/east

Households: basis Mikrozensus 2010:

- Number of households by federal state and BIK type

- Number of households by household size and west/east

- Number of households by "children under 15 years of age in the household yes/no" and west/east

For the calibration, each benchmark variable for each household must have a valid value. Therefore, the very low non-response item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Since the imputation only serves the feasibility of the calibration, the imputed values were set back to missing values after the calibration. A projection with the calibrated weights without considering the non-response item thus leads to slight deviations from the values as presented in the following.

Table 50: Nominal distributions and distributions after calibration (population sample, households)

Benchmark fig-ure	Characteristics bench-mark figure from BA sta-tistics and Mikrozensus 2010	Unweighted dis-tribution	Nominal values	Distribution with calibrat-ed weights
Number BCs re-ceiving benefits in accordance with SGB II by federal states (16 catego-ries)	Number BCs Schleswig-Holstein	9	122,816	123,812
	Number BCs Hamburg	6	108,473	108,473
	Number BCs Lower-Saxony	32	334,073	334,073
	Number BCs Bremen	6	51,949	51,949
	Number BCs North Rhine-Westphalia	93	849,532	848,046
	Number BCs Hesse	7	219,564	219,564
	Number BCs Rhineland-Palatinate	11	123,522	124,333
	Number BCs Baden-Wuerttemberg	10	256,618	256,618
	Number BCs Bavaria	26	267,375	267,375
	Number BCs Saarland	7	44,787	44,787
	Number BCs Berlin	17	334,106	333,949
	Number BCs Brandenburg	24	167,160	167,160
	Number BCs Mecklenburg-Vorpommern	4	127,174	127,174
	Number BCs Saxony	18	280,039	280,039
	Number BCs Saxony-Anhalt	12	186,400	186,400
	Number BCs Thuringia	15	130,035	129,872

Table 50: Nominal distributions and distributions after calibration (population sample, households) (continued 1)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number of BCs receiving benefits in accordance with SGB II by number of individuals under 65 years of age in the benefit unit (1, 2, 3, 4, and "5 or more") and by west/east (10 categories)	Number BCs with 1 individual under 65 (west)	72	1,287,479	1,287,479
	Number BCs with 2 individuals under 65 (west)	50	480,035	479,859
	Number BCs with 3 individuals under 65 (west)	35	297,590	297,090
	Number BCs with 4 individuals under 65 (west)	29	185,355	185,355
	Number BCs with 5 or more individuals under 65 (west)	21	128,250	129,246
	Number BCs with 1 individual under 65 (east)	28	714,572	714,572
	Number BCs with 2 individuals under 65 (east)	27	264,338	264,338
	Number BCs with 3 individuals under 65 (east)	22	135,917	135,597
	Number BCs with 4 individuals under 65 (east)	7	71,425	71,425
	Number BCs with 5 or more individuals under 65 (east)	6	38,662	38,662
Number of BCs receiving benefits in accordance with SGB II by number of individuals under 15 years of age in the benefit unit (0, 1, 2, 3, "4 or more") and by west/east (10 categories)	Number BCs without children under 15 years of age (west)	139	1,629,124	1,628,448
	Number BCs with 1 child under 15 years of age (west)	27	406,480	406,480
	Number BCs with 2 children under 15 years of age (west)	26	230,977	231,973
	Number BCs with 3 children under 15 years of age (west)	8	80,359	80,359
	Number BCs with 4 or more children under 15 years of age (west)	7	31,769	31,769
	Number BCs without children under 15 years of age (east)	65	909,771	909,614
	Number BCs with 1 child under 15 years of age (east)	11	188,910	188,747
	Number BCs with 2 children under 15 years of age (east)	11	90,185	90,185
	Number BCs with 3 children under 15 years of age (east)	2	25,972	25,972
	Number BCs with 4 or more children under 15 years of age (east)	1	10,076	10,076

Table 50: Nominal distributions and distributions after calibration (population sample, households) (continued 2)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number BCs receiving benefits in accordance with SGB II consisting of a single parent with children by west/east (4 categories)	Number BCs with a single parent (west)	44	450,759	451,755
	Rest BCs without a single parent (west)	163	1,927,950	1,927,274
	Number BCs with a single parent (east)	11	189,640	189,640
	Rest BCs without a single parent (east)	79	1,035,274	1,034,954
Number of households by federal state and BIK type (spelling: "Federal state.BIK type")	1.1 to 1.4	12	308,000	308,000
	1.5 to 1.6	21	139,000	139,000
	1.7 to 1.8	37	518,000	518,000
	1.9	15	179,000	179,000
	1.10	32	221,000	221,000
	2.10	55	972,000	972,000
	3.2 to 3.3	59	500,000	500,000
	3.4	47	442,000	442,000
	3.5	47	458,000	458,000
	3.7	131	852,000	852,000
	3.8	62	553,000	553,000
	3.9	84	623,000	623,000
	3.10	39	393,000	393,000
	4.8 to 4.10	31	358,000	358,000
	5.2 to 5.3	55	356,000	356,000
	5.4	97	1,038,000	1,038,000
	5.5	91	640,000	640,000
	5.6	39	355,000	355,000
	5.7	91	741,000	741,000
	5.8	199	2,143,000	2,143,000
	5.9	79	457,000	457,000
	5.10	289	2,821,000	2,821,000
	6.2	11	65,000	65,000
	6.3	50	319,000	319,000
	6.4 to 6.5	19	322,000	322,000
	6.7	71	565,000	565,000
	6.8	24	463,000	463,000
	6.9	62	360,000	360,000
	6.10	56	794,000	794,000
	7.1 to 7.2	37	305,000	305,000
	7.3	15	187,000	187,000
	7.4	33	178,000	178,000
	7.5 to 7.6	19	222,000	222,000
	7.7	67	394,000	394,000
	7.8	32	321,000	321,000
	7.9 to 7.10	11	258,000	258,000
	8.2 to 8.3	60	640,000	640,000
	8.4	71	517,000	517,000
	8.5 to 8.6	28	475,000	475,000

Table 50: Nominal distributions and distributions after calibration (population sample, households) (continued 3)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number of house-holds by federal state and BIK type (spelling: "Federal state.BIK type")	8.7	77	940,000	940,000
	8.8	75	691,000	691,000
	8.9	52	491,000	491,000
	8.10	106	1,195,000	1,195,000
	9.1 to 9.2	27	358,000	358,000
	9.3	85	492,000	492,000
	9.4	84	692,000	692,000
	9.5	77	370,000	370,000
	9.6 to 9.7	136	1,096,000	1,096,000
	9.8	52	643,000	643,000
	9.9	99	723,000	723,000
	9.10	127	1,499,000	1,499,000
	10.3 to 10.4	14	155,000	155,000
	10.7 to 10.8	39	330,000	330,000
	11.10	127	1,974,000	1,974,000
	12.1 to 12.3	23	270,000	270,000
	12.4	28	239,000	239,000
	12.5 to 12.6	33	174,000	174,000
	12.7	13	75,000	75,000
	12.8	16	149,000	149,000
	12.9 to 12.10	38	332,000	332,000
	13.1 to 13.4	23	360,000	360,000
	13.5 to 13.6	11	161,000	161,000
	13.7	4	100,000	100,000
	13.8	21	223,000	223,000
	14.1	13	19,000	19,000
	14.2	32	132,000	132,000
	14.3 to 14.4	24	470,000	470,000
	14.5	11	164,000	164,000
	14.6	15	118,000	118,000
	14.7 to 14.8	26	376,000	376,000
	14.9 to 14.10	76	912,000	912,000
	15.1 to 15.2	16	111,000	111,000
	15.3 to 15.4	27	292,000	292,000
	15.5 to 15.6	42	275,000	275,000
	15.7	23	234,000	234,000
	15.8	42	282,000	282,000
	16.1 to 16.2	36	146,000	146,000
	16.3 to 16.4	44	361,000	361,000
	16.5	34	242,000	242,000
	16.6	31	53,000	53,000
	16.7 to 16.8	33	305,000	305,000

Table 50:   Nominal distributions and distributions after calibration (population sample, households) (continued 4)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number of households by household size (1,2,3,4,"5 and more individuals") and west/east (10 categories)	Number households with 1 individual (west)	797	12,066,000	12,066,000
	Number households with 2 individuals (west)	1,272	10,591,000	10,591,000
	Number households with 3 individuals (west)	514	3,971,000	3,971,000
	Number households with 4 individuals (west)	532	3,273,000	3,273,000
	Number households with 5 or more individuals (west)	213	1,226,000	1,226,000
	Number households with 1 individual (east)	254	3,651,000	3,651,000
	Number households with 2 individuals (east)	359	3,087,000	3,087,000
	Number households with 3 individuals (east)	147	1,099,000	1,099,000
	Number households with 4 individuals (east)	71	563,000	563,000
	Number households with 5 and more individuals (east)	31	149,000	149,000
Number of households by "children under 15 years of age in the household yes/no" and west/east	Number households with children under 15 (west)	759	5,640,000	5,640,000
	Number households without children under 15 (west)	2,569	25,487,000	25,487,000
	Number households with children under 15 (east)	134	1,232,000	1,232,000
	Number households without children under 15 (east)	728	7,317,000	7,317,000

Table 51: **Parameters of distribution of weights**

1% percentile	705.2742
5% percentile	2,162.21
10% percentile	2,903.944
25% percentile	4,492.12
50% percentile	7,049.734
75% percentile	11,456.63
90% percentile	20,196.32
95% percentile	26,008.26
99% percentile	37,335.97
Mean	9,469.212
Standard deviation	7867.704
Minimum	277.5614
Maximum	79,154.33
Number of observations	4,190
Efficiency measure	**59.2%**

6.12.3 Total sample

All private households in Germany form the population. Starting point for the calibration were modified design weights including the non-response weighting. The modified design weights were trimmed at the 5 % percentile and the 95 % percentile of their distribution and after that rescaled in such a way that their total again resulted in the total of the un-trimmed design weights. The projection factors of the trimmed design weights range from 209.1 to 18628.5. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.1 and upwards to 5.0. Thus, the total projection factors after calibration lie between min. 20.9 and max. 46186.

A calibration was made for the following characteristics:

Benefit units basis BA statistics:

- Number BCs receiving benefits in accordance with SGB II by federal states

- Number of BCs receiving benefits in accordance with SGB II by number of individuals under 65 years of age in the benefit unit, by west/east

- Number of BCs receiving benefits in accordance with SGB II by number of children under 15 years of age in the benefit unit, by west/east

- Number of BCs receiving benefits in accordance with SGB II consisting of a single parent with child(ren), by west/east

Households basis Mikrozensus 2010:

- Number of households by federal state and BIK type

- Number of households by household size and west/east

- Number of households by "children under 15 years of age in the household yes/no" and west/east

Besides that also the increase in Unemployment Benefit II recipients since the previous year at the level of benefit units (367,870) was included as an additional benchmark value in the total sample.

For the calibration, each benchmark variable for each household must have a valid value. Therefore, the very low non-response item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Since the imputation only serves the feasibility of the calibration, the imputed values were set back to missing values after the calibration. A projection with the calibrated weights without considering the non-response item thus leads to slight deviations from the values as presented in the following.

Table 52: **Nominal distributions and distributions after calibration (total sample, households)**

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikro-zensus 2010	Unweighted distribution	Nominal values	Distribution with cali-brated weights
Number BCs re-ceiving benefits in accordance with SGB II by federal states (16 categories)	Number BCs Schleswig-Holstein	170	122,816	122,795
	Number BCs Hamburg	144	108,473	108,494
	Number BCs Lower-Saxony	468	334,073	333,999
	Number BCs Bremen	57	51,949	51,948
	Number BCs North Rhine-Westphalia	1,173	849,532	849,304
	Number BCs Hesse	224	219,564	219,541
	Number BCs Rhineland-Palatinate	162	123,522	123,634
	Number BCs Baden-Wuerttemberg	324	256,618	256,578
	Number BCs Bavaria	413	267,375	268,778
	Number BCs Saarland	78	44,787	44,749
	Number BCs Berlin	320	334,106	333,546
	Number BCs Brandenburg	296	167,160	166,990
	Number BCs Mecklenburg-Vorpommern	133	127,174	127,127
	Number BCs Saxony	357	280,039	279,896
	Number BCs Saxony-Anhalt	273	186,400	186,234
	Number BCs Thuringia	189	130,035	129,989

Table 52: Nominal distributions and distributions after calibration (total sample, households) (continued 1)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number of BCs receiving benefits in accordance with SGB II by number of individuals under 65 years of age in the benefit unit (1, 2, 3, 4, and "5 or more") and by west/east (10 categories)	Number BCs with 1 individual under 65 (west)	1,297	1,287,479	1,287,478
	Number BCs with 2 individuals under 65 (west)	842	480,035	481,373
	Number BCs with 3 individuals under 65 (west)	541	297,590	297,435
	Number BCs with 4 individuals under 65 (west)	314	185,355	185,402
	Number BCs with 5 or more individuals under 65 (west)	219	128,250	128,132
	Number BCs with 1 individual under 65 (east)	710	714,572	714,572
	Number BCs with 2 individuals under 65 (east)	449	264,338	263,868
	Number BCs with 3 individuals under 65 (east)	243	135,917	135,359
	Number BCs with 4 individuals under 65 (east)	97	71,425	71,465
	Number BCs with 5 or more individuals under 65 (east)	69	38,662	38,520
Number of BCs receiving benefits in accordance with SGB II by number of individuals under 15 years of age in the benefit unit (0, 1, 2, 3, "4 or more") and by west/east (10 categories)	Number BCs without children under 15 years of age (west)	2,136	1,629,124	1,630,364
	Number BCs with 1 child under 15 years of age (west)	577	406,480	406,504
	Number BCs with 2 children under 15 years of age (west)	351	230,977	230,817
	Number BCs with 3 children under 15 years of age (west)	101	80,359	80,376
	Number BCs with 4 or more children under 15 years of age (west)	48	31,769	31,759
	Number BCs without children under 15 years of age (east)	1,160	909,771	908,804
	Number BCs with 1 child under 15 years of age (east)	258	188,910	188,887
	Number BCs with 2 children under 15 years of age (east)	111	90,185	90,044
	Number BCs with 3 children under 15 years of age (east)	25	25,972	25,972
	Number BCs with 4 or more children under 15 years of age (east)	14	10,076	10,076

Table 52: Nominal distributions and distributions after calibration (total sample, households) (continued 2)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted dis-tribution	Nominal val-ues	Distribution with cali-brated weights
Number BCs receiv-ing benefits in ac-cordance with SGB II consisting of a single parent with children by west/east (4 catego-ries)	Number BCs with a single parent (west)	759	450,759	450,656
	Rest BCs without a single parent (west)	2,454	1,927,950	1,929,164
	Number BCs with a single parent (east)	269	189,640	189,580
	Rest BCs without a single parent (east)	1,299	1,035,274	1,034,203
Number of house-holds by federal state and BIK type (spelling: "Federal state.BIK type")	1.1 to 1.4	48	308,000	308,000
	1.5 to 1.6	45	139,000	139,000
	1.7 to 1.8	134	518,000	518,000
	1.9	35	179,000	179,000
	1.10	78	221,000	221,000
	2.10	226	972,000	972,000
	3.2 to 3.3	129	500,000	500,000
	3.4	102	442,000	442,000
	3.5	67	458,000	458,000
	3.7	261	852,000	852,000
	3.8	193	553,000	553,000
	3.9	155	623,000	623,000
	3.10	135	393,000	393,000
	4.8 to 4.10	96	358,000	358,000
	5.2 to 5.3	126	356,000	356,000
	5.4	228	1,038,000	1,038,000
	5.5	231	640,000	640,000
	5.6	103	355,000	355,000
	5.7	204	741,000	741,000
	5.8	535	2,143,000	2,143,000
	5.9	152	457,000	457,000
	5.10	770	2,821,000	2,821,000
	6.2	18	65,000	65,000
	6.3	93	319,000	319,000
	6.4 to 6.5	46	322,000	322,000
	6.7	122	565,000	565,000
	6.8	83	463,000	463,000
	6.9	105	360,000	360,000
	6.10	131	794,000	794,000
	7.1 to 7.2	59	305,000	305,000
	7.3	23	187,000	187,000
	7.4	53	178,000	178,000
	7.5 to 7.6	61	222,000	222,000
	7.7	86	394,000	394,000
	7.8	87	321,000	321,000
	7.9 to 7.10	39	258,000	258,000
	8.2 to 8.3	96	640,000	640,000
	8.4	114	517,000	517,000
	8.5 to 8.6	42	475,000	475,000

Table 52: Nominal distributions and distributions after calibration (total sample, households) (continued 3)

Benchmark fig-ure	Characteristics bench-mark figure from BA statistics and Mikro-zensus 2010	Unweighted distribution	Nominal values	Distribution with calibrat-ed weights
Number of households by federal state and BIK type (spelling: "Feder-al state.BIK type")	8.7	154	940,000	940,000
	8.8	154	691,000	691,000
	8.9	92	491,000	491,000
	8.10	240	1,195,000	1,195,000
	9.1 to 9.2	40	358,000	358,000
	9.3	122	492,000	492,000
	9.4	127	692,000	692,000
	9.5	128	370,000	370,000
	9.6 to 9.7	218	1,096,000	1,096,000
	9.8	151	643,000	643,000
	9.9	140	723,000	723,000
	9.10	305	1,499,000	1,499,000
	10.3 to 10.4	39	155,000	155,000
	10.7 to 10.8	103	330,000	330,000
	11.10	556	1,974,000	1,974,000
	12.1 to 12.3	108	270,000	270,000
	12.4	86	239,000	239,000
	12.5 to 12.6	126	174,000	174,000
	12.7	44	75,000	75,000
	12.8	33	149,000	149,000
	12.9 to 12.10	107	332,000	332,000
	13.1 to 13.4	102	360,000	360,000
	13.5 to 13.6	44	161,000	161,000
	13.7	39	100,000	100,000
	13.8	63	223,000	223,000
	14.1	19	19,000	19,000
	14.2	104	132,000	132,000
	14.3 to 14.4	82	470,000	470,000
	14.5	40	164,000	164,000
	14.6	96	118,000	118,000
	14.7 to 14.8	68	376,000	376,000
	14.9 to 14.10	253	912,000	912,000
	15.1 to 15.2	62	111,000	111,000
	15.3 to 15.4	69	292,000	292,000
	15.5 to 15.6	105	275,000	275,000
	15.7	101	234,000	234,000
	15.8	158	282,000	282,000
	16.1 to 16.2	73	146,000	146,000
	16.3 to 16.4	115	361,000	361,000
	16.5	89	242,000	242,000
	16.6	59	53,000	53,000
	16.7 to 16.8	110	305,000	305,000

Table 52: **Nominal distributions and distributions after calibration (total sample, households) (continued 4)**

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with cali- brated weights
Number of households by household size (1,2,3,4,"5 and more individu-als") and west/east (10 categories)	Number households with 1 individual (west)	2,310	12,066,000	12,066,000
	Number households with 2 individuals (west)	2,392	10,591,000	10,591,000
	Number households with 3 individuals (west)	1,202	3,971,000	3,971,000
	Number households with 4 individuals (west)	946	3,273,000	3,273,000
	Number households with 5 or more indi-viduals (west)	474	1,226,000	1,226,000
	Number households with 1 individual (east)	1,144	3,651,000	3,651,000
	Number households with 2 individuals (east)	958	3,087,000	3,087,000
	Number households with 3 individuals (east)	484	1,099,000	1,099,000
	Number households with 4 individuals (east)	213	563,000	563,000
	Number households with 5 and more in-dividuals (east)	112	149000	149000
Number of households by "children under 15 years of age in the house-hold yes/no" and west/east	Number households with children under 15 (west)	2,086	5,640,000	5,640,000
	Number households without children un-der 15 (west)	5,238	25,487,000	25,487,000
	Number households with children under 15 (east)	647	1,232,000	1,232,000
	Number households without children un-der 15 (east)	2,264	7,317,000	7,317,000

Table 53:   Parameters of distribution of weights

1% percentile	42.12403
5% percentile	108.0065
10% percentile	174.8335
25% percentile	402.8683
50% percentile	1049.692
75% percentile	5171.642
90% percentile	11155.47
95% percentile	17561.72
99% percentile	25649.94
Mean	3876.502
Standard deviation	5832.188
Minimum	20.90576
Maximum	46,186
Number of observations	10,235
Efficiency measure	**30.6%**

6.13 Calibration to the person weight, wave 5, cross-section

As in the previous waves, the person weights were calibrated under the restriction that they differ as little as possible from the calibrated household weights. The calibrated household weights were quasi inherited by the individual household members. Following this, these input weights were calibrated at the individual level.

As in the previous year, also the increase in Unemployment Benefit II recipients since the previous year at the level of individuals between 15 and 64 years (499,235) was also included as an additional benchmark value in the total sample. Again, those cases in the previous samples from wave 1, wave 2, wave 3 and wave 4 which, according to wave 5 of the survey, were receiving Unemployment Benefit II in July 2010 will be projected to the benchmark statistics of the Federal Employment Agency on receipt of Unemployment Benefit II.

Before calibration, the calibrated household weights that formed the input weight were trimmed, too. Also for the calibration of person weights, additionally the range of weights was determined to a certain interval.

6.13.1 BA sample

The population of the cumulated BA sample of all five waves consists of all individuals aged 15 and over who are living in a household in which there was at least one benefit unit receiving benefits in accordance with SGB II at one of the, up to now, five drawing dates (in July 2006, July 2007, July 2008, July 2009 or July 2010). Only those individuals aged 15 and over who were living in a benefit unit receiving benefits in accordance with SGB II in July 2010 were considered for the calibration. Individuals living in a household

that does not receive benefits and individuals living in a household with at least on benefit unit in accordance with SGB II but are no part of a benefit unit themselves were removed from the dataset for the calibration. The weighting of these individuals was calculated in a different way (see below).

The starting point for the calibration is the calibrated household weights of the BA sample. They were trimmed at the 5 % percentile and the 95 % percentile of their distribution and after that rescaled in such a way that their total again resulted in the total of the untrimmed calibrated household weights. The trimmed projection factors range from 284.1 to 3407.7. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.2 and upwards to 3.5. Thus, the total projection factors after calibration lie between minimal 78.9 and maximal 4699.1.

A calibration was made for the following characteristics:

Benefit recipients basis BA statistics:

- Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by federal states

- Number of individuals in benefit units receiving benefits in accordance with SGB II by age (15-24 and 25-64)

- Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by sex, by west/east

- Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by "single parent yes/no", by west/east

- Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by nationality (German/non-German)

As in the previous year, also the increase in Unemployment Benefit II recipients since the previous year at the level of individuals between 15 and 64 years (499,235) was also included as an additional benchmark value in the total sample.

For the calibration, each benchmark variable for each individual must have a valid value. Therefore, the very low non-response item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Since the imputation only serves the feasibility of the calibration, the imputed values were set back to missing values after the calibration. A projection with the calibrated weights without considering the non-response item thus leads to slight deviations from the values as presented in the following.

Table 54: Nominal distributions and distributions after calibration (BA sample, individuals)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by federal states (16 categories)	Number individuals in BCs Schleswig-Holstein	203	170,430	170,430
	Number individuals in BCs Hamburg	173	146,202	146,202
	Number individuals in BCs Lower-Saxony	562	472,434	472,434
	Number individuals in BCs Bremen	63	70,746	70,746
	Number individuals in BCs North Rhine-Westphalia	1,470	1,217,381	1,217,381
	Number individuals in BCs Hesse	284	315,063	315,063
	Number individuals in BCs Rhineland-Palatinate	198	175,315	175,315
	Number individuals in BCs Baden-Wuerttemberg	403	355,805	355,805
	Number individuals in BCs Bavaria	493	358,390	358,390
	Number individuals in BCs Saarland	91	61,906	61,906
	Number individuals in BCs Berlin	357	450,129	450,129
	Number individuals in BCs Brandenburg	358	225,513	225,513
	Number individuals in BCs Mecklenburg-Vorpommern	156	171,727	171,727
	Number individuals in BCs Saxony	441	379,470	379,470
	Number individuals in BCs Saxony-Anhalt	373	256,044	256,044
	Number individuals in BCs Thuringia	234	175,557	175,557
Number of individuals in benefit units receiving benefits in accordance with SGB II by age (15-24 and 25-64; 2 categories)	Number individuals in BCs aged 15-24	906	939,152	939,152
	Number individuals in BCs aged 25-64	4,953	4,062,960	4,062,960
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by sex and west/east (4 categories)	Number men in BCs (west)	1,792	1,637,643	1,637,643
	Number women in BCs (west)	2,148	1,706,029	1,706,029
	Number men in BCs (east)	923	854,217	854,217
	Number women in BCs (east)	996	804,223	804,223
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by "single parent yes/no", sex and west/east (8 categories)	Number non single parents in BCs (west)	3,210	2,892,913	2,892,913
	Number single parents in BCs (west)	730	450,759	450,759
	Number non single parents in BCs (east)	1,646	1,468,800	1,468,800
	Number single parents in BCs (east)	273	189,640	189,640
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by nationality (German/non-German)	Number non-German individuals in BCs	678	990,860	990,860
	Number German individuals in BCs	5,181	4,011,252	4,011,252

Table 55:   Parameters of distribution of weights

1% percentile	118.4711
5% percentile	168.7721
10% percentile	225.2746
25% percentile	357.9115
50% percentile	669.5241
75% percentile	1,152.045
90% percentile	1,825.846
95% percentile	2,114.982
99% percentile	3,300.052
Mean	854.1858
Standard deviation	655.8097
Minimum	78.91148
Maximum	4,699.068
Number of observations	5,856
Efficiency measure	**62.9%**

6.13.2 Population sample

All individuals over 14 years of age in private households in Germany form the population. Starting point for the calibration were calibrated household weights of the population sample. They were trimmed at the 5 % percentile and the 95 % percentile of their distribution and after that rescaled in such a way that their total again resulted in the total of the untrimmed calibrated household weights. The trimmed projection factors range from 2,358.6 to 28,801. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.2 and upwards to 3.5. Thus, the total projection factors after calibration lie between minimal 471.7 and maximal 79,063.4.

A calibration was made for the following characteristics:

Benefit recipients basis BA statistics:

- Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by federal states

- Number of individuals in benefit communities receiving benefits in accordance with SGB II by age (15-24 and 25-64)

- Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by sex, by west/east

- Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by "single parent yes/no", by west/east

- Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by nationality (German/non-German)

Population basis Mikrozensus 2010:

- Number of individuals aged 15 and over in private households by federal state
- Number of individuals aged 15 and over in private households by age, sex and west/east
- Number of individuals aged 15 and over in private households by household size and west/east
- Number of individuals aged 15 and over in private households by school qualification and west/east
- Number of individuals aged 15 and over in private households by marital status and west/east
- Number of individuals aged 15 and over in private households by nationality

Population basis BA statistics:

- Number of unemployed individuals including participants in measures by west/east
- Number of employees subject to social security by west/east

The source used for the benchmark value of the employment status was the BA statistics since the definition of unemployment and employment subject to social insurance in PASS does not correspond to the ILO concept of the Federal Statistical Office but can be taken from the statistics of the BA.

For the calibration, each benchmark variable for each individual must have a valid value. Therefore, the very low non-response item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Since the imputation only serves the feasibility of the calibration, the imputed values were set back to missing values after the calibration. A projection with the calibrated weights without considering the non-response item thus leads to slight deviations from the values as presented in the following.

Table 56: Nominal distributions and distributions after calibration (population sample, individuals)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by federal states (16 categories)	Number individuals in BCs Schleswig-Holstein	10	170,430	170,430
	Number individuals in BCs Hamburg	8	146,202	146,202
	Number individuals in BCs Lower-Saxony	45	472,434	472,434
	Number individuals in BCs Bremen	9	70,746	70,746
	Number individuals in BCs North Rhine-Westphalia	143	1,217,381	1,217,381
	Number individuals in BCs Hesse	15	315,063	315,063
	Number individuals in BCs Rhineland-Palatinate	18	175,315	175,315
	Number individuals in BCs Baden-Wuerttemberg	19	355,805	355,805
	Number individuals in BCs Bavaria	35	358,390	358,390
	Number individuals in BCs Saarland	10	61,906	61,906
	Number individuals in BCs Berlin	19	450,129	450,129
	Number individuals in BCs Brandenburg	34	225,513	225,513
	Number individuals in BCs Mecklenburg-Vorpommern	8	171,727	171,727
	Number individuals in BCs Saxony	27	379,470	379,470
	Number individuals in BCs Saxony-Anhalt	19	256,044	256,044
	Number individuals in BCs Thuringia	21	175,557	175,557
Number of individuals in benefit units receiving benefits in accordance with SGB II by age (15-24 and 25-64; 2 categories)	Number individuals in BCs aged 15-24	76	939,152	939,152
	Number individuals in BCs aged 25-64	364	4,062,960	4,062,960
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by sex and west/east (4 categories)	Number men in BCs (west)	148	1,637,643	1,637,643
	Number women in BCs (west)	164	1,706,029	1,706,029
	Number men in BCs (east)	65	854,217	854,217
	Number women in BCs (east)	63	804,223	804,223
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by "single parent yes/no", sex and west/east (8 categories)	Number non single parents in BCs (west)	267	2,892,913	2,892,913
	Number single parents in BCs (west)	45	450,759	450,759
	Number non single parents in BCs (east)	115	1,468,800	1,468,800
	Number single parents in BCs (east)	13	189,640	189,640
Number of individuals aged 15 and over in benefit units receiving	Number non-German individuals in BCs	59	990,860	990,860
	Number German individuals in	381	4,011,252	4,011,252

benefits in accordance with SGB II by nationality (German/non-German)	BCs				

Table 56: Nominal distributions and distributions after calibration (population sample, individuals) (continued 1)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number of individuals aged 15 and over in private households by federal state (16 categories)	Number individuals in private households Schleswig-Holstein	188	2,398,000	2,398,000
	Number individuals in private households Hamburg	82	1,550,000	1,550,000
	Number individuals in private households Lower-Saxony	789	6,736,000	6,736,000
	Number individuals in private households Bremen	43	577,000	577,000
	Number individuals in private households North Rhine-Westphalia	1,560	15,266,000	15,266,000
	Number individuals in private households Hesse	532	5,201,000	5,201,000
	Number individuals in private households Rhineland-Palatinate	360	3,437,000	3,437,000
	Number individuals in private households Baden-Wuerttemberg	818	9,138,000	9,138,000
	Number individuals in private households Bavaria	1,214	10,646,000	10,646,000
	Number individuals in private households Saarland	96	885,000	885,000
	Number individuals in private households Berlin	177	3,017,000	3,017,000
	Number individuals in private households Brandenburg	253	2,207,000	2,207,000
	Number individuals in private households Mecklenburg-Vorpommern	102	1,450,000	1,450,000
	Number individuals in private households Saxony	319	3,673,000	3,673,000
	Number individuals in private households Saxony-Anhalt	262	2,061,000	2,061,000
	Number individuals in private households Thuringia	305	1,961,000	1,961,000
Number of individuals aged 15 and over in private households by age (in 5-year classes), gender and west/east (56 categories)	Number men in private households (west), 15-19 years	195	1,884,000	1,884,000
	Number men in private households (west), 20-24 years	191	1,939,000	1,939,000
	Number men in private households (west), 25-29 years	110	1,948,000	1,948,000
	Number men in private households (west), 30-34 years	123	1,908,000	1,908,000
	Number men in private households (west), 35-39 years	147	2,062,000	2,062,000
	Number men in private households (west), 40-44 years	252	2,826,000	2,826,000
	Number men in private households (west), 45-49 years	284	2,736,000	2,736,000
	Number men in private households (west), 50-54 years	299	2,383,000	2,383,000
	Number men in private households (west), 55-59 years	237	2,068,000	2,068,000

Table 56: Nominal distributions and distributions after calibration (population sample, individuals) (continued 2)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with cali-brated weights
Number of indi-viduals aged 15 and over in pri-vate households by age (in 5-year classes), gender and west/east (56 categories)	Number men in private households (west), 60-64 years	232	1,781,000	1,781,000
	Number men in private households (west), 65-69 years	208	1,749,000	1,749,000
	Number men in private households (west), 70-74 years	241	1,787,000	1,787,000
	Number men in private households (west), 75-79 years	100	1,109,000	1,109,000
	Number men in private households (west), 80+ years	83	1,048,000	1,048,000
	Number women in private households (west), 15-19 years	215	1,768,000	1,768,000
	Number women in private households (west), 20-24 years	148	1,896,000	1,896,000
	Number women in private households (west), 25-29 years	122	1,944,000	1,944,000
	Number women in private households (west), 30-34 years	144	1,916,000	1,916,000
	Number women in private households (west), 35-39 years	200	2,052,000	2,052,000
	Number women in private households (west), 40-44 years	292	2,722,000	2,722,000
	Number women in private households (west), 45-49 years	359	2,678,000	2,678,000
	Number women in private households (west), 50-54 years	324	2,426,000	2,426,000
	Number women in private households (west), 55-59 years	300	2,120,000	2,120,000
	Number women in private households (west), 60-64 years	260	1,841,000	1,841,000
	Number women in private households (west), 65-69 years	203	1,894,000	1,894,000
	Number women in private households (west), 70-74 years	222	2,037,000	2,037,000
	Number women in private households (west), 75-79 years	118	1,375,000	1,375,000
	Number women in private households (west), 80+ years	73	1,937,000	1,937,000
	Number men in private households (east), 15-19 years	34	317,000	317,000
	Number men in private households (east), 20-24 years	51	556,000	556,000
	Number men in private households (east), 25-29 years	43	550,000	550,000
	Number men in private households (east), 30-34 years	41	509,000	509,000
	Number men in private households (east), 35-39 years	33	507,000	507,000
	Number men in private households (east), 40-44 years	53	655,000	655,000
	Number men in private households (east), 45-49 years	84	718,000	718,000
	Number men in private households (east), 50-54 years	77	635,000	635,000
	Number men in private households (east), 55-59 years	62	621,000	621,000

Number men in private households (east), 60-64 years	58	438,000	438,000
Number men in private households (east), 65-69 years	54	504,000	504,000
Number men in private households (east), 70-74 years	45	502,000	502,000
Number men in private households (east), 75-79 years	37	301,000	301,000
Number men in private households (east), 80+ years	16	237,000	237,000

Table 56: Nominal distributions and distributions after calibration (population sample, individuals) (continued 3)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikro-zensus 2010	Unweighted distribution	Nominal values	Distribution with cali-brated weights
Number of indi-viduals aged 15 and over in pri-vate households by age (in 5-year classes), gender and west/east (56 categories)	Number women in private house-holds (east), 15-19 years	28	279,000	279,000
	Number women in private house-holds (east), 20-24 years	38	497,000	497,000
	Number women in private house-holds (east), 25-29 years	37	491,000	491,000
	Number women in private house-holds (east), 30-34 years	29	461,000	461,000
	Number women in private house-holds (east), 35-39 years	31	475,000	475,000
	Number women in private house-holds (east), 40-44 years	56	624,000	624,000
	Number women in private house-holds (east), 45-49 years	88	681,000	681,000
	Number women in private house-holds (east), 50-54 years	78	642,000	642,000
	Number women in private house-holds (east), 55-59 years	88	636,000	636,000
	Number women in private house-holds (east), 60-64 years	71	454,000	454,000
	Number women in private house-holds (east), 65-69 years	54	571,000	571,000
	Number women in private house-holds (east), 70-74 years	68	604,000	604,000
	Number women in private house-holds (east), 75-79 years	38	404,000	404,000
	Number women in private house-holds (east), 80+ years	26	500,000	500,000
Number of indi-viduals aged 15 and over in pri-vate households by household size (1, 2, 3, 4, "5 or more indi-viduals") and west/east (10 categories)	Number individuals in private house-holds with 1 individual (west)	792	12,066,000	12,066,000
	Number individuals in private house-holds with 2 individuals (west)	2,144	20,714,000	20,714,000
	Number individuals in private house-holds with 3 individuals (west)	1,005	9,811,000	9,811,000
	Number individuals in private house-holds with 4 individuals (west)	1,177	9,128,000	9,128,000
	Number individuals in private house-holds with 5 or more individuals (west)	564	4,115,000	4,115,000
	Number individuals in private house-holds with 1 individual (east)	253	3,651,000	3,651,000
	Number individuals in private house-holds with 2 individuals (east)	633	5,998,000	5,998,000
	Number individuals in private house-holds with 3 individuals (east)	304	2,721,000	2,721,000
	Number individuals in private house-holds with 4 individuals (east)	147	1,527,000	1,527,000
	Number individuals in private house-holds with 5 or more individuals (east)	81	472,000	472,000

Table 56: Nominal distributions and distributions after calibration (population sample, individuals) (continued 4)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Un-weighted distribu-tion	Nominal values	Distribution with calibrat-ed weights
Number of individuals aged 15 and over in private households by highest school qualification and west/east (12 categories)	Number individuals in private households with highest school qualification: still pupil (west)	245	2,273,000	2,273,000
	Number individuals in private households with highest school qualification: no qualification (west)	161	2,336,000	2,336,000
	Number individuals in private households with highest school qualification: lower secondary school (west)	1,903	22,759,000	22,759,000
	Number individuals in private households with highest school qualification: intermediate secondary school; intermediate secondary school in the former GDR (west)	1,626	13,732,000	13,732,000
	Number individuals in private households with highest school qualification: university (of applied sciences) qualification (west)	1,747	14,734,000	14,734,000
	Number individuals in private households with highest school qualification: still pupil (east)	27	343,000	343,000
	Number individuals in private households with highest school qualification: no qualification (east)	26	361,000	361,000
	Number individuals in private households with highest school qualification: lower secondary school (east)	282	3,270,000	3,270,000
	Number individuals in private households with highest school qualification: Intermediate secondary school; intermediate secondary school in the former GDR (east)	697	6,752,000	6,752,000
	Number individuals in private households with highest school qualification: university (of applied sciences) qualification (east)	386	3,643,000	3,643,000
Number of individuals aged 15 and over in private households by marital status and west/east (10 categories)	Number individuals in private households with marital status: single (west)	1,311	10,125,000	10,125,000
	Number individuals in private households with marital status: married, civil partnership (west)	3,654	36,063,000	36,063,000
	Number individuals in private households with marital status: divorced (west)	406	4,883,000	4,883,000
	Number individuals in private households with marital status: widowed (west)	311	4,763,000	4,763,000
	Number individuals in private households with marital status: single (east)	339	3,351,000	3,351,000
	Number individuals in private households with marital status: married, civil partnership (east)	856	8,264,000	8,264,000
	Number individuals in private households with marital status: divorced (east)	117	1,466,000	1,466,000

	Number individuals in private households with marital status: widowed (east)	106	1,288,000	1,288,000
Number of individuals aged 15 and over in private households by nationality	Number individuals in private households non-German	290	6,340,000	6,340,000
	Number individuals in private households German	6,810	63,863,000	63,863,000

Table 56: Nominal distributions and distributions after calibration (population sample, individuals) (continued 5)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikro-zensus 2010	Unweighted distribution	Nominal val-ues	Distribu-tion with calibrated weights
Unemployed indi-viduals incl. partici-pants in measures west/east	Not unemployed west	5,409	53,032,681	53,032,681
	Unemployed individuals incl. partici-pants in measures west	273	2,801,319	2,801,319
	Not unemployed east	1,309	13,084,328	13,084,328
	Unemployed individuals incl. partici-pants in measures east	109	1,284,672	1,284,672
Employees subject to social security contributions west/east	Employees not subject to social se-curity contributions west	3,431	32,861,296	32,861,296
	Employees subject to social security contributions west	2,251	22,972,704	22,972,704
	Employees not subject to social se-curity contributions east	790	9,000,117	9,000,117
	Employees subject to social security contributions east	628	5,368,883	5,368,883

Table 57: Parameters of distribution of weights

1% percentile	962.9659
5% percentile	1,799.035
10% percentile	2,360.424
25% percentile	3,893.852
50% percentile	6,806.903
75% percentile	12,181.16
90% percentile	21,342.62
95% percentile	29,527.5
99% percentile	47,323.72
Mean	9,887.747
Standard deviation	9,479.726
Minimum	471.7283
Maximum	79,063.38
Number of observations	7,100
Efficiency measure	**52.1%**

6.13.3 Total sample

As for the population sample, all individuals of aged 15 and over in private households in Germany form the population. Starting point for the calibration were calibrated household weights of the total sample. They were trimmed at the 5 % percentile and the 95 % percentile of their distribution and after that rescaled in such a way that their total again resulted in the total of the untrimmed calibrated household weights. The trimmed projection factors range from 121.5 to 20,778.5. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.2 and upwards to 3.5. Thus, the total projection factors after calibration lie between minimal 24.3 and maximal 65,767.4.

A calibration was made for the following characteristics:

Benefit recipients basis BA statistics:

- Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by federal states

- Number of individuals in benefit units receiving benefits in accordance with SGB II by age (15-24 and 25-64)

- Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by sex, by west/east

- Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by "single parent yes/no", by west/east

- Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by nationality (German/non-German)

Population basis Mikrozensus 2010:

- Number of individuals aged 15 and over in private households by federal state

- Number of individuals aged 15 and over in private households by age, sex and west/east

- Number of individuals aged 15 and over in private households by household size and west/east

- Number of individuals aged 15 and over in private households by school qualification and west/east

- Number of individuals aged 15 and over in private households by marital status and west/east

- Number of individuals aged 15 and over in private households by nationality

Population basis BA statistics:

- Number of unemployed individuals including participants in measures by west/east
- Number of employees subject to social security by west/east

The source used for the benchmark value of the employment status was the BA statistics since the definition of unemployment and employment subject to social insurance in PASS does not correspond to the ILO concept of the Federal Statistical Office but can be taken from the statistics of the BA.

Besides that, also the increase in Unemployment Benefit II recipients since the previous year at the level of individuals between 15 and 64 years (499,235) was included as an additional benchmark value in the total sample.

For the calibration, each benchmark variable for each individual must have a valid value. Therefore, the very low non-response item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Since the imputation only serves the feasibility of the calibration, the imputed values were set back to missing values after the calibration. A projection with the calibrated weights without considering the non-response item thus leads to slight deviations from the values as presented in the following.

Table 58: Nominal distributions and distributions after calibration (total sample, individuals)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Un-weighted distribution	Nominal values	Distribution with calibrated weights
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by federal states (16 categories)	Number individuals in BCs Schleswig-Holstein	213	170,430	170,430
	Number individuals in BCs Hamburg	181	146,202	146,202
	Number individuals in BCs Lower-Saxony	607	472,434	472,434
	Number individuals in BCs Bremen	72	70,746	70,746
	Number individuals in BCs North Rhine-Westphalia	1,613	1,217,381	1,217,381
	Number individuals in BCs Hesse	299	315,063	315,063
	Number individuals in BCs Rhine-land-Palatinate	216	175,315	175,315
	Number individuals in BCs Baden-Wuerttemberg	422	355,805	355,805
	Number individuals in BCs Bavaria	528	358,390	358,390
	Number individuals in BCs Saar-land	101	61,906	61,906
	Number individuals in BCs Berlin	376	450,129	450,129
	Number individuals in BCs Brandenburg	392	225,513	225,513
	Number individuals in BCs Mecklenburg-Vorpommern	164	171,727	171,727
	Number individuals in BCs Saxony	468	379,470	379,470
	Number individuals in BCs Saxony-Anhalt	392	256,044	256,044
	Number individuals in BCs Thuringia	255	175,557	175,557
Number of individuals in benefit units receiving benefits in accordance with SGB II by age (15-24 and 25-64; 2 categories)	Number individuals in BCs aged 15-24	982	939,152	939,152
	Number individuals in BCs aged 25-64	5,317	4,062,960	4,062,960
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by sex and west/east (4 categories)	Number men in BCs (west)	1,940	1,637,643	1,637,643
	Number women in BCs (west)	2,312	1,706,029	1,706,029
	Number men in BCs (east)	988	854,217	854,217
	Number women in BCs (east)	1,059	804,223	804,223
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by "single parent yes/no", sex and west/east (8 categories)	Number non single parents in BCs (west)	3,477	2,892,913	2,892,913
	Number single parents in BCs (west)	775	450,759	450,759
	Number non single parents in BCs (east)	1,761	1,468,800	1,468,800
	Number single parents in BCs (east)	286	189,640	189,640
Number of individuals aged 15 and over in benefit units receiving benefits in accordance	Number non-German individuals in BCs	737	990,860	990,860
	Number German individuals in BCs	5,562	4,011,252	4,011,252

with SGB II by nationality (German/non-German)				

Table 58: Nominal distributions and distributions after calibration (total sample, individuals) (continued 1)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number of individuals aged 15 and over in private households by federal state (16 categories)	Number individuals in private households Schleswig-Holstein	484	2,398,000	2,398,000
	Number individuals in private households Hamburg	310	1,550,000	1,550,000
	Number individuals in private households Lower-Saxony	1,587	6,736,000	6,736,000
	Number individuals in private households Bremen	129	577,000	577,000
	Number individuals in private households North Rhine-Westphalia	3,643	15,266,000	15,266,000
	Number individuals in private households Hesse	949	5,201,000	5,201,000
	Number individuals in private households Rhineland-Palatinate	628	3,437,000	3,437,000
	Number individuals in private households Baden-Wuerttemberg	1,397	9,138,000	9,138,000
	Number individuals in private households Bavaria	1,973	10,646,000	10,646,000
	Number individuals in private households Saarland	227	885,000	885,000
	Number individuals in private households Berlin	710	3,017,000	3,017,000
	Number individuals in private households Brandenburg	750	2,207,000	2,207,000
	Number individuals in private households Mecklenburg-Vorpommern	356	1,450,000	1,450,000
	Number individuals in private households Saxony	985	3,673,000	3,673,000
	Number individuals in private households Saxony-Anhalt	790	2,061,000	2,061,000
	Number individuals in private households Thuringia	689	1,961,000	1,961,000
Number of individuals aged 15 and over in private households by age (in 5-year classes), gender and west/east (56 categories)	Number men in private households (west), 15-19 years	412	1,884,000	1,884,000
	Number men in private households (west), 20-24 years	387	1,939,000	1,939,000
	Number men in private households (west), 25-29 years	373	1,948,000	1,948,000
	Number men in private households (west), 30-34 years	384	1,908,000	1,908,000
	Number men in private households (west), 35-39 years	381	2,062,000	2,062,000
	Number men in private households (west), 40-44 years	539	2,826,000	2,826,000
	Number men in private households (west), 45-49 years	590	2,736,000	2,736,000

Table 58: Nominal distributions and distributions after calibration (total sample, individuals) (continued 2)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number of individuals aged 15 and over in private households by age (in 5-year classes), gender and west/east (56 categories)	Number men in private households (west), 50-54 years	553	2,383,000	2,383,000
	Number men in private households (west), 55-59 years	513	2,068,000	2,068,000
	Number men in private households (west), 60-64 years	428	1,781,000	1,781,000
	Number men in private households (west), 65-69 years	278	1,749,000	1,749,000
	Number men in private households (west), 70-74 years	257	1,787,000	1,787,000
	Number men in private households (west), 75-79 years	107	1,109,000	1,109,000
	Number men in private households (west), 80+ years	84	1,048,000	1,048,000
	Number women in private households (west), 15-19 years	451	1,768,000	1,768,000
	Number women in private households (west), 20-24 years	415	1,896,000	1,896,000
	Number women in private households (west), 25-29 years	425	1,944,000	1,944,000
	Number women in private households (west), 30-34 years	503	1,916,000	1,916,000
	Number women in private households (west), 35-39 years	522	2,052,000	2,052,000
	Number women in private households (west), 40-44 years	633	2,722,000	2,722,000
	Number women in private households (west), 45-49 years	726	2,678,000	2,678,000
	Number women in private households (west), 50-54 years	638	2,426,000	2,426,000
	Number women in private households (west), 55-59 years	580	2,120,000	2,120,000
	Number women in private households (west), 60-64 years	447	1,841,000	1,841,000
	Number women in private households (west), 65-69 years	255	1,894,000	1,894,000
	Number women in private households (west), 70-74 years	237	2,037,000	2,037,000
	Number women in private households (west), 75-79 years	126	1,375,000	1,375,000
	Number women in private households (west), 80+ years	83	1,937,000	1,937,000
	Number men in private households (east), 15-19 years	116	317,000	317,000
	Number men in private households (east), 20-24 years	161	556,000	556,000
	Number men in private households (east), 25-29 years	207	550,000	550,000
	Number men in private households (east), 30-34 years	173	509,000	509,000

Number men in private households (east), 35-39 years	141	507,000	507,000
Number men in private households (east), 40-44 years	175	655,000	655,000

Table 58: Nominal distributions and distributions after calibration (total sample, individuals) (continued 3)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number of individuals aged 15 and over in private households by age (in 5-year classes), gender and west/east (56 categories)	Number men in private households (east), 45-49 years	248	718,000	718,000
	Number men in private households (east), 50-54 years	228	635,000	635,000
	Number men in private households (east), 55-59 years	232	621,000	621,000
	Number men in private households (east), 60-64 years	194	438,000	438,000
	Number men in private households (east), 65-69 years	76	504,000	504,000
	Number men in private households (east), 70-74 years	53	502,000	502,000
	Number men in private households (east), 75-79 years	38	301,000	301,000
	Number men in private households (east), 80+ years	17	237,000	237,000
	Number women in private households (east), 15-19 years	108	279,000	279,000
	Number women in private households (east), 20-24 years	164	497,000	497,000
	Number women in private households (east), 25-29 years	188	491,000	491,000
	Number women in private households (east), 30-34 years	187	461,000	461,000
	Number women in private households (east), 35-39 years	143	475,000	475,000
	Number women in private households (east), 40-44 years	205	624,000	624,000
	Number women in private households (east), 45-49 years	309	681,000	681,000
	Number women in private households (east), 50-54 years	255	642,000	642,000
	Number women in private households (east), 55-59 years	253	636,000	636,000
	Number women in private households (east), 60-64 years	181	454,000	454,000
	Number women in private households (east), 65-69 years	79	571,000	571,000
	Number women in private households (east), 70-74 years	77	604,000	604,000
	Number women in private households (east), 75-79 years	41	404,000	404,000
	Number women in private households (east), 80+ years	31	500,000	500,000
Number of individuals aged 15 and over in private households by household size (1, 2, 3, 4, "5 or more individuals") and west/east (10 categories)	Number individuals in private households with 1 individual (west)	2,300	12,066,000	12,066,000
	Number individuals in private households with 2 individuals (west)	3,741	20,714,000	20,714,000
	Number individuals in private households with 3 individuals (west)	2,158	9,811,000	9,811,000
	Number individuals in private households with 4 individuals (west)	1,951	9,128,000	9,128,000
	Number individuals in private households with 5 or more individuals (west)	1,177	4,115,000	4,115,000

	Number individuals in private house-holds with 1 individual (east)	1,139	3,651,000	3,651,000
	Number individuals in private house-holds with 2 individuals (east)	1,533	5,998,000	5,998,000
	Number individuals in private house-holds with 3 individuals (east)	903	2,721,000	2,721,000
	Number individuals in private house-holds with 4 individuals (east)	426	1,527,000	1,527,000
	Number individuals in private house-holds with 5 or more individuals (east)	279	472,000	472,000

Table 58: Nominal distributions and distributions after calibration (total sample, individuals) (continued 4)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number of individuals aged 15 and over in private households by highest school qualification and west/east (12 categories)	Number individuals in private households with highest school qualification: still pupil (west)	537	2,273,000	2,273,000
	Number individuals in private households with highest school qualification: no qualification (west)	628	2,336,000	2,336,000
	Number individuals in private households with highest school qualification: lower secondary school (west)	4,130	22,759,000	22,759,000
	Number individuals in private households with highest school qualification: intermediate secondary school; intermediate secondary school in the former GDR (west)	3,151	13,732,000	13,732,000
	Number individuals in private households with highest school qualification: university (of applied sciences) qualification (west)	2,881	14,734,000	14,734,000
	Number individuals in private households with highest school qualification: still pupil (east)	111	343,000	343,000
	Number individuals in private households with highest school qualification: no qualification (east)	148	361,000	361,000
	Number individuals in private households with highest school qualification: lower secondary school (east)	912	3,270,000	3,270,000
	Number individuals in private households with highest school qualification: Intermediate secondary school; intermediate secondary school in the former GDR (east)	2,249	6,752,000	6,752,000
	Number individuals in private households with highest school qualification: university (of applied sciences) qualification (east)	860	3,643,000	3,643,000
Number of individuals aged 15 and over in private households by marital status and west/east (10 categories)	Number individuals in private households with marital status: single (west)	3,537	10,125,000	10,125,000
	Number individuals in private households with marital status: married, civil partnership (west)	5,739	36,063,000	36,063,000
	Number individuals in private households with marital status: divorced (west)	1,592	4,883,000	4,883,000
	Number individuals in private households with marital status: widowed (west)	459	4,763,000	4,763,000
	Number individuals in private households with marital status: single (east)	1,585	3,351,000	3,351,000
	Number individuals in private households with marital status: married, civil partnership (east)	1,854	8,264,000	8,264,000

Number individuals in private house-holds with marital status: divorced (east)	663	1,466,000	1,466,000
Number individuals in private house-holds with marital status: widowed (east)	178	1,288,000	1,288,000

Table 58:   Nominal distributions and distributions after calibration (total sample, individuals) (continued 5)

Benchmark figure	Characteristics benchmark figure from BA statistics and Mikrozensus 2010	Unweighted distribution	Nominal values	Distribution with calibrated weights
Number of individuals aged 15 and over in private households by nationality	Number individuals in private households non-German	1,166	6,340,000	6,340,000
	Number individuals in private households German	14,441	63,863,000	63,863,000
Unemployed individuals incl. participants in measures west/east	Not unemployed west	8,918	53,032,681	53,032,681
	Unemployed individuals incl. participants in measures west	2,409	2,801,319	2,801,319
	Not unemployed east	2,996	13,084,328	13,084,328
	Unemployed individuals incl. participants in measures east	1,284	1,284,672	1,284,672
Employees subject to social security contributions west/east	Employees not subject to social security contributions west	7,325	32,861,296	32,861,296
	Employees subject to social security contributions west	4,002	22,972,704	22,972,704
	Employees not subject to social security contributions east	2,655	9,000,117	9,000,117
	Employees subject to social security contributions east	1,625	5,368,883	5,368,883

Table 59:   Parameters of distribution of weights

1% percentile	39.67916
5% percentile	94.32583
10% percentile	141.2273
25% percentile	353.4838
50% percentile	1,387.97
75% percentile	5,820.786
90% percentile	13,003.01
95% percentile	19,010.12
99% percentile	31,854.19
Mean	4,498.174
Standard deviation	6,954.562
Minimum	24.30088
Maximum	65,767.37
Number of observations	15,607
Efficiency measure	**29.5%**

6.14 Estimating the BA cross-sectional weights for households and individuals not in receipt of Unemployment Benefit II

Finally, also in wave 5 some households and individuals remained that could not be assigned a BA cross-sectional household weight or a BA cross-sectional person weight by means of calibration. The number of these households is now larger in wave 5 than in wave 4 since a larger part of the BA sample of waves 1 to 4 has meanwhile withdrawn from receiving benefits. These are the following three groups which were not receiving benefits in July 2010 but which belong to the population of the BA sample (households in receipt of Unemployment Benefit II in 7/2006 or 7/2007 or 7/2008 or 7/2009 or 7/2010 and individuals in households in receipt of Unemployment Benefit II in 7/2006 or 7/2007 or 7/2008 or 7/2009 or 7/2010).

- From the refreshment sample: Individuals in the household who are not members of a benefit unit: Here, the person weight was obtained from the BA household weight in wave 5 after calibration (*wqbahh*) by dividing it by the proportion of these individuals who gave a personal or senior citizens' interview – provided that their household was participating.

- Panel households in which nobody was in receipt of Unemployment Benefit II any longer in July 2010: The household retains the BA weight before calibration. Individuals in these households with interviews in both waves were given a new BA person weight which is obtained by multiplying their old BA person weight from the previous wave by the reciprocal re-participation probability *ppbleib*. Individuals in these households who did not provide a personal interview in wave 4 are given a new BA person weight calculated by dividing the BA household weight of their household for wave 5 by the proportion of such individuals who participate provided that their household is taking part.

- Individuals who are not members of a benefit unit in panel households that were still in receipt of Unemployment Benefit II in July 2010: Individuals in these households with interviews in both waves were given a new BA person weight which is obtained by multiplying their BA person weight from the previous wave by the reciprocal re-participation probability *ppbleib*.

For the first time in wave 5, the individuals and households were additionally adjusted to a benchmark figure for the individuals or benefit units which were no longer in receipt of Unemployment Benefit II. The exact population size of this group is unknown, but can be determined approximately from the total of all cumulated BA subsamples minus the individuals or benefit units currently in receipt of benefits. The respective number of individuals who are no longer in receipt of Unemployment Benefit II is 3,185,162, the number of benefit units which are no longer in receipt of Unemployment Benefit II is 2,368,760.

7   Appendix: Brief description of the dataset

Content characteristics

Categories	Comments
Topics/ characteristics categories	**Socio-demographic characteristics:** artificial individual ID; sex; year of birth; age; marital status; number of children living in and outside the household; nationality; country of origin and migration background; command of the German language and language spoken in the household/among friends; school and vocational qualifications (incl. generated scales: CASMIN, ISCED-97, number of years of schooling and vocational training), parents' school and vocational qualifications; health indicators; religious denomination; social contacts; leisure time activities of young people; childcare and school attendance of children; caretaking activities; household income (incl. individual components and equivalised household income); basic information on assets and liabilities; household equipment (deprivation index); housing and residential environment; detailed information on the topic of old age benefits (only wave 3); **Employment-related characteristics:** employment status / economic inactivity status; marginal employment; working hours; occupational status (detailed); employment (ISCO-88 and KldB-92); ISCO-based measures of occupational status and prestige (ISEI, SIOPS, MPS, EGP, ESeC); earned income (gross and net); employment biographies with employment/unemployment spells and periods of economic inactivity since 01/2005 (from wave 2 onwards); limited-term employment and cancellation of limitation of employment; supervisory function; employer: public service/private industry; employer: number of employees; other employment; pooled information on the employment and unemployment history; detailed information on the subject of job-search; reservation wage; vignettes module on readiness to accept a job (wave 5 only); **Characteristics on receiving benefits:** Unemployment Benefit I: start and end dates of the spell(s) of benefit receipt since 01/2005 (wave 1 only); information on periods of Unemployment Benefit I receipt in the context of registered unemployment since 01/2005 (from wave 2 onwards); amount of benefit; reason for end; Unemployment Benefit II: start and end dates of the spell(s) of benefit receipt since 01/2005; reason for start and end; identification of household members receiving benefits; amount of benefits received; benefit cuts (start date, duration, reasons, which household members' benefit cut); Measurement participation (until wave 3): type of measure; start and end dates of measure; indicator of dropout; reasons for dropout; type of access to measure; assessment of measure; working hours in measure; comparison to regular employment; economic sector/industry; from wave 4 onwards only one-euro job; One-euro job participation (as of wave 4): knowledge, discussions about one-euro jobs; reasons for non-participation; start and end of a one-euro job; reasons for premature ending; working hours; comparison of activity with regular employment; assessment Contacts with Unemployment Benefit II institutions: number and type of contacts; contents of discussion; offers; integration agreement; assessment of institution;

Categories	Comments
Topics/ characteristics catego- ries (continued)	**Subjective indicators:** satisfaction; fears and problems; employment orientation; education as- piration; sex role orientation; subjective social position (top-bottom scale); subjective assessment of health state; personality scale "big five"
Data unit	Individuals and households in receipt of Unemployment Benefit II in 7/2006 (sample I) Individuals and households in the resident population of Germany (sample II) Individuals and households in receipt of Unemployment Benefit II in 7/2007 but without receipt in 7/2006 (sample III; refreshment sample 1) Individuals and households in receipt of Unemployment Benefit II in 7/2008 but without receipt in 7/2006 or 7/2007 (sample IV; refresh- ment sample 2) Individuals and households in receipt of Unemployment Benefit II in 7/2009 but without receipt in 7/2006, 7/2007 or 7/2008 (sample V; refreshment sample 3) Individuals and households in receipt of Unemployment Benefit II in 7/2010 but without receipt in 7/2006, 7/2007, 7/2008 or 7/2009 (sample VIII; refreshment sample 4) Individuals and households of the resident German population (sample VI, panel refreshment/replenishment sample) Individuals and households in receipt of UB II in 7/2010 (sample VII, panel refreshment/replenishment sample) Note: individuals aged 65 and over are interviewed using a shorter version of the questionnaire

Categories	Comments
Case numbers	Wave 1: sample I: 9,386 individuals (living in 6,804 households) Sample II: 9,568 individuals (living in 5,990 households) Wave 2: sample I: 4,753 individuals (living in 3,491 households) Sample II: 6,392 individuals (living in 3,897 households) Sample III: 1,342 individuals (living in 1,041 households) Wave 3: sample I: 4,913 individuals (living in 3,754 households) Sample II: 6,207 individuals (living in 3,901 households) Sample III: 898 individuals (living in 694 households) Sample IV: 1,421 individuals (living in 1,186 households) Wave 4: sample I: 3,958 individuals (living in 2,815 households) Sample II: 5,016 individuals (living in 2,977 households) Sample III: 786 individuals (living in 563 households) Sample IV: 983 individuals (living in 745 households) Sample V: 1,025 individuals (living in 748 households) Wave 5: Sample I: 3,394 individuals (in 2,382 households) Sample II: 4,511 individuals (in 2,680 households) Sample III: 653 individuals (living in 464 households) Sample IV: 822 individuals (living in 608 households) Sample V: 760 individuals (in 517 households) Sample VI: 2,589 individuals (in 1,510 households) Sample VII: 1,859 individuals (in 1,321 households) Sample VIII: 1,019 individuals (living in 753 households)

Categories	Comments
Data collection mode	CATI and CAPI CAPI interviews were conducted when a sample household could not be reached by telephone or when a personal interview was requested. Wave 1: N (CATI): 12,414 individuals (8,445 households) N (CAPI): 6,540 individuals (4,339 households) Wave 2: N (CATI): 7,888 individuals (5,378 households) N (CAPI): 4,599 individuals (3,051 households) Wave 3: N (CATI): 7,776 individuals (5,664 households) N (CAPI): 5,663 individuals (3,871 households) Wave 4: n (CATI): 6,913 individuals (4,669 households) n (CAPI): 4,855 individuals (3,179 households) Wave 5: n (CATI): 7,358 individuals (4,987 households) n (CAPI): 8,249 individuals (5,248 households)
Interview languages	Wave 1: German: 18,205 individuals (12,347 households) Russian: 432 individuals (275 households) Turkish: 305 individuals (163 households) English: 12 individuals (9 households) Wave 2: German: 12,237 individuals (8,234 households) Russian: 219 individuals (156 households) Turkish: 31 individuals (39 households) English: no longer offered in wave 2 due to the low case numbers in wave 1 Wave 3: German: 13,000 individuals (9,256 households) Russian: 330 individuals (210 households) Turkish: 109 individuals (69 households) Wave 4: German: 11,405 individuals (7,627 households) Russian: 285 individuals (179 households) Turkish: 78 individuals (42 households) Wave 5: German: 15,290 individuals (10,040 households) Russian: 259 individuals (159 households) Turkish: 58 individuals (36 households)

Categories	Comments
Response rates	Wave 1: Sample I: 35.1 % Sample II: 26.6 % Total: 30.5 % Wave 2: Sample I (HHs agreeing to participate only): 51.1 % Sample II (HHs agreeing to participate only): 64.7 % Sample III: 26.3 % Split-off households (from samples I and II): 13.4 % Total: 45.0 % Wave 3: Sample I (HHs agreeing to participate only): 64.5 % Sample II (HHs agreeing to participate only): 76.4 % Sample II (HHs agreeing to participate only): 69.0 % Sample IV: 31.2% Total: 60.6 % Wave 4: Sample I (HHs agreeing to participate only): 72.1 % Sample II (HHs agreeing to participate only): 82.4 % Sample III (HHs agreeing to participate only): 65.6 % Sample IV (HHs agreeing to participate only): 68.2 % Sample V: 30.9 % Total: 59.5 % Wave 5: Sample I (HHs agreeing to participate only): 71.1 % Sample II (HHs agreeing to participate only): 81.3 % Sample III (HHs agreeing to participate only): 69.2 % Sample IV (HHs agreeing to participate only): 63.7 % Sample V: (HHs agreeing to participate only): 71.5 % Sample VI: 24.5 % Sample VII: 24.5 % Sample VIII: 27.1 % Total: 43.9 %

Categories	Comments
Response rates within households	Stage 1: sample I: 85.6 % Sample II: 84.3 % Total: 85.0 % Wave 2: Sample I (re-interviewed households only): 85.5 % Sample II (re-interviewed households only): 85.1 % Sample III: 86.2 % Split-off households (from samples I and II): 88.3 % Total: 85.4 % Wave 3: Sample I (re-interviewed households only): 83.1 % Sample I (re-interviewed households only): 83.6 % Sample III (re-interviewed households only): 84.3 % Sample IV: 84.2 % Split-off households (from samples I - II): 84.2 % Total: 83.5 % Wave 4: Sample I (re-interviewed households only): 88.4 % Sample I (re-interviewed households only): 88.0 % Sample III (re-interviewed households only): 90.2 % Sample IV (re-interviewed households only): 88.3 % Sample V: 89.6 % Split-off households (from samples I - IV): 86.4 % Total: 88.5 % Wave 5: Sample I (re-interviewed households only): 88.7 % Sample I (re-interviewed households only): 88.3 % Sample III (re-interviewed households only): 89.5 % Sample IV (re-interviewed households only): 89.3 % Sample V (re-interviewed households only): 91.2 % Sample VI: 84.4 % Sample VII: 90.0 % Sample VIII: 88.9 % Split-off households (from samples I - V): 89.9 % Total: 88.3 %
Fieldwork period	Wave 1: December 2006-June 2007 Wave 2: December 2007-July 2008 Wave 3: December 2008-August 2009 Wave 4: February 2010-September 2010 Wave 5: February 2011-September 2011

Categories	Comments
Period	Wave 1: fieldwork period and retrospective spell data as of 01/2005 Wave 2: fieldwork period and retrospective spell data as of 01/2005 or the respective reference period of the spell type Wave 3: fieldwork period and retrospective spell data as of 01/2006 or the respective reference period of the spell type Wave 4: fieldwork period and retrospective spell data as of 01/2008 or the respective reference period of the spell type Wave 5: fieldwork period and retrospective spell data as of 01/2009 or the respective reference period of the spell type
Time reference	Repeat interview (household panel)
Regional structure	German federal state, east/west Germany (Further regional information is available but is not contained in the scientific use file for data protection reasons. Detailed information available on request)
Territorial allocation	At the survey date

Methodological characteristics

Categories	Comments
Survey design	Original sample wave 1: two-stage random sample with two sub-populations
	Stage 1: selection of 300 postcode sectors as primary sampling units (PSU) for both subsamples. The sampling probability of the individual postcode areas depended on the particular size of the area in terms of the number of residents (probability proportional to size/pps).
	Stage 2, sample I: drawing of benefit units from the register data of the Federal Employment Agency with sampling date July 2006. The number of the gross sample drawn per PSU depended on the PSU size in terms of the relative proportion of benefit recipients within the respective postcode sector (probability proportional to size/pps). The average size of the gross sample was N=100 per postcode area.
	Stage 2, sample II: for sample II, first a sample of residential buildings was drawn from a commercial database (Microm mosaic). This was then stratified by a stratification index contained in the database at a ratio of 4:2:1 for households with a low, medium or high status respectively. Interviewers from the surveying institute visited the selected buildings. In the event that a building accommodated several households, this was noted and then one of the households was selected by the institute as the household to be interviewed. The gross sample comprised N=100 households per postcode area.
	BA refreshment sample in wave 2 (sample III): In addition to continuing sample I which was drawn for wave 1, in the second wave a refreshment sample was drawn from the register data of the Federal Employment Agency. For this, benefit units which were in receipt of Unemployment Benefit II in July 2007 but not in July 2006 were selected, i.e. benefit units which are new benefit recipients. The sample was drawn in the postcode areas selected for wave 1 following the procedure used in the first wave.
	BA refreshment sample in wave 3 (sample IV): Also in wave 3 a refreshment sample for sample I was drawn from the register data of the Federal Employment Agency. For this, benefit units which were in receipt of Unemployment Benefit II in July 2008 but not in July 2006 or July 2007 were selected, i.e. benefit units which are new benefit recipients. The sample was drawn in the postcode sectors selected for wave 1 following the procedure used in the first wave.
	BA refreshment sample in wave 4 (sample V): Also in wave 4 a refreshment sample for sample I was drawn from the register data of the Federal Employment Agency. For this, benefit units

which were in receipt of Unemployment Benefit II in July 2009 but not in July 2006, July 2007 or July 2008 were selected. These benefit units thus depict the inflows to benefit receipt. The sample was drawn in the postcode sectors selected for wave 1 following the procedure used in the first wave.

In wave 5, the panel of the original sample was refreshed with two replenishment samples based on a two-staged random sample with two subpopulations.

Stage 1: selection of 100 postcode sectors as primary sampling units (PSU) for both subsamples. The sampling probability of the individual postcode sectors depended on the particular size of the sector in terms of the number of residents (probability proportional to size/pps).

Stage 2, sample VI: in sample VI the individuals were drawn from the registration offices' registers. For this, 96 municipalities were allocated to the 100 postcode areas. The drawing of the personal addresses from the possible choices in the municipalities was made by systematic random sampling (interval sampling). Sampling of addresses from the registration offices' registers was made from year of birth 1992 and earlier. 144 addresses were drawn from the municipalities' registers in each sample point.

Stage 2, sample VII: drawing of benefit units from the register data of the Federal Employment Agency with sampling date July 2010. The number of benefit recipients to be selected per point was selected as the product of the permanent sample size (sample size individuals per point) in the population sample with the quotient from benefit recipient rate in the point and benefit recipient rate in all over Germany.

BA refreshment sample in wave 5 (Sample VIII):
In wave 5 a refreshment sample for sample I was again drawn from the register data of the Federal Employment Agency. For this, benefit units which were in receipt of Unemployment Benefit II in July 2010 but not in July 2006, July 2007, July 2008 or July 2009 were selected, i.e. benefit units which are new benefit recipients. The sample was drawn in the postcode sectors selected for wave 1 following the procedure used in the first wave.

Categories	Comments
Institutions involved in survey	Institute for Employment Research (IAB); TNS Infratest Sozialforschung (waves 1 to 3), infas Institut für angewandte Sozialwissenschaft GmbH (as of wave 4)
Frequency of data collection	Annually (panel)
File format and size	STATA, SPSS (several files)
File architecture	Household dataset: HHENDDAT.dta/.sav
	Individual dataset: *PENDDAT*.dta/.sav
	Spell data Unemployment Benefit I: alg1_spells.dta/.sav (wave 1 only)
	Spell data Unemployment Benefit II: alg2_spells.dta/.sav
	Spell data unemployment: al_spells.dta/.sav (waves 2 and 3)
	Spell data employment: et_spells.dta/.sav (waves 2 and 3)
	Spell data gaps: lu_spells.dta/.sav (waves 2 and 3)
	from wave 4 onwards: spell data on employment, unemployment and gaps integrated: bio_spells.dta/.sav
	Spell data measures: mn_spells.dta/.sav (from wave 2 onwards)
	Spell data participation in measures: massnahmespells.dta/.sav (wave 1 only)
	Register data on households: hh_register.dta/.sav
	Register data on individuals: p_register.dta/.sav
	Weighting data on households: hweights.dta/.sav
	Weighting data on individuals: pweights.dta/.sav
	Old-age provision household level: HAVDAT.dta/.sav (wave 3 only)
	Old-age provision individual level: PAVDAT.dta/.sav (wave 3 only)
	Vignette data: VIGDAT.dat/.sav (wave 5 only)

Data access

Categories	Comments
Data access	Scientific use file (SUF)
Degree of anonymisation	Factually anonymised
Sensitive characteristics	none

Literature:

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

Achatz, J., Hirseland, A. & Promberger, M. (2007). Rahmenkonzept für das IAB-Panel "Arbeitsmarkt und Soziale Sicherung". In M. Promberger (Ed.), Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung "Arbeitsmarkt und Soziale Sicherung", *IAB-Forschungsbericht 12/2007* (pp. 11-32), Nuremberg.

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

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

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

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

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

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

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

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

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

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

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

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

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

Erikson, R., Goldthorpe, J. & Portocarero, L. (1979). Intergenerational Class Mobility in Three Western Societies: England, France and Sweden. *British Journal of Sociology 30.* pp. 415-441.

Erikson, R., Goldthorpe, J. & Portocarero, L. (1982): Social Fluidity in Industrial Nations: England, France and Sweden. *British Journal of Sociology 33.* pp. 1-34.

Fischer, A. & Wirth, H. (2007): *Constructing Version 4 of ESEC Classes from 3-digit ISCO (Stata-do file).* Mannheim: Gesis-ZUMA.

Frick, J., Göbel, J. & Krause, P. (n.d.). $HGEN: Generated Household Level Variables. [http://www.diw.de/documents/dokumentenarchiv/17/60053/hgen.pdf (8.11.2007)].

Ganzeboom, H. & Treiman, D. (1996). Internationally Comparable Measures for Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research 25.* pp. 201-239.

Ganzeboom, H. & Treiman, D. (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. In H. Jürgen, P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables (pp. 159-193), New York: Kluwer Academic / Plenum Publishers.

Ganzeboom, H., De Graaf, P. & Treiman, D. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research 21.* pp. 1-56.

Gebhardt, D., Müller, G., Bethmann, A., Trappmann, M., Christoph, B., Gayer, C., Müller, B., Tisch, A., Siflinger, B., Kiesl, H., Huyer-May, B., Achatz, J., Wenzig, C., Rudolph, H., Graf, T. & Biedermann, A. (2009). Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS). Datenreport Welle 2 (2007/2008). *FDZ Datenreport 06/2009.* Nuremberg.

Gebhardt, D. & Trappmann, M. (2011): Using the datasets. In A. Bethmann & D. Gebhardt (Hrsg.), User Guide "Panel Study Labour Market and Social Security" (PASS). Wave 3, *FDZ Datenreport, 04/2011* (pp. 62-107), Nuremberg.

Granato, N. (2000). Mikrodaten-Tools: CASMIN-Bildungsklassifikation. Eine Umsetzung mit dem Mikrozensus 1996. *ZUMA-Technischer Bericht 2000/12.* Mannheim.

Hagenaars, A., de Vos, K. & Zaidi, M. (1994). *Poverty Statistics in the Late 1980s: Research Based on Micro-data.* Luxembourg: Office for Official Publications of the European Communities.

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

Harrison, E. & Rose, R. (2006). ESeC User Guide, Appendix 6 (SPSS-Syntax: Esec Full) [http://www.iser.essex.ac.uk/esec/guide/docs/ Appendix6.sps (31.11.2007)]

Hartmann, J., Brink, K., Jäckle, R. & Tschersich, N. (2008). IAB-Haushaltspanel im Niedrigeinkommensbereich. Methoden- und Feldbericht. *FDZ Methodenreport 07/2008.* Nuremberg.

Hauser, R. (1996). Zur Messung individueller Wohlfahrt und Ihrer Verteilung. In Statistisches Bundesamt (Ed.), *Wohlfahrtsmessung. Aufgabe der Statistik im gesellschaftlichen Wandel* (pp. 13-38), Stuttgart: Metzler-Poeschel.

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

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

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

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

Jesske, B. & Schulz, S. (2012). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 5. Erhebungswelle – 2011 (Haupterhebung), *FDZ Methodenreport 11/2012*, Nuremberg.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Imprint

Corresponding author:

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