Research Data Centre (FDZ)
of the German Federal
Employment Agency (BA)
at the Institute for
Employment Research (IAB)

FDZ-Datenreport

Documentation of labour market data

07/2016
EN

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

Datenreport Wave 9

Arne Bethmann,
Benjamin Fuchs,
Martina Huber,
Mark Trappmann,
Alice Reindl,
Marco Berg,
Ralph Cramer,
Christian Dickmann,
Reiner Gilberg,
Birgit Jesske,
Martin Kleudgen

Bundesagentur für Arbeit

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

Datenreport Wave 9

Arne Bethmann; Benjamin Fuchs; Martina Huber; Mark Trappmann; Alice Reindl (Institute for Employment Research - IAB)
Marco Berg; Ralph Cramer; Christian Dickmann; Reiner Gilberg; Birgit Jesske; Martin Kleudgen (infas Institut für angewandte Sozialwissenschaft GmbH)

FDZ-Datenreporte (FDZ data reports) describe FDZ data in detail. As a result, this series of reports has a dual function: on the one hand, those using the reports can ascertain whether the data offered is suitable for their research task; on the other, the data can be used to prepare evaluations. This data report documents the data preparation of the PASS wave 9 and is based upon the eights wave's data report: Marco Berg, Ralph Cramer, Christian Dickmann, Reiner Gilberg, Birgit Jesske, Martin Kleudgen, (all infas Institut für angewandte Sozialwissenschaft GmbH), Arne Bethmann, Benjamin Fuchs, Mark Trappmann, Martina Huber (all Institut für Arbeitsmarkt- und Berufsforschung (IAB)): Codebuch und Dokumentation des „Panel Arbeitsmarkt und soziale Sicherung"(PASS) volume I: Datenreport Welle 8, FDZ Datenreport, 06/2015 (de), Nuremberg.

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/ .

Contents

List of Figures

List of Tables

1 Introduction

1.1 The 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), creates a new empirical dataset for labour market, welfare state and poverty research and policy counseling in Germany. This study is conducted as part of IAB research on German Social Code Book II (SGB II)[1]. The IAB must fulfill a statutory mandate to study the effects of the benefits and services provided under SGB II, which are aimed at labour-market integration and subsistence benefits. However, due to its complex sampling design, this study also enables researchers to examine additional issues. The following five core questions, which are detailed in Achatz, Hirseland and Promberger (2007), influenced the development of this study.

1. What are the options for regaining financial independence from Unemployment Benefit (UB) II (Arbeitslosengeld II)?
2. How does a household’s social situation change when it receives benefits?
3. How do individuals who receive benefits cope with their situations? Do recipient atti-tudes toward the actions required to improve their situations change over time?
4. How does contact between benefit recipients and institutions that provide basic social security take place? What actual institutional procedures are applied in practice?
5. What employment history patterns or household dynamics lead to receiving Unemployment Benefit II?

This data report provides an overview of the ninth survey wave, for which 13,271 individuals in 8,921 households[2] were interviewed between February 2015 and September 2015. This sample included 11,659 individuals and 7,865 households that had previously been interviewed for PASS.

This data report[3] of wave 9 documents the wave-specific aspects of the study. Chapter 1.2 contains a short description of the instruments and the survey program. Following a

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

[2] These figures include evaluable interviews only. Additionally, repeatedly interviewed house-holds were considered even if only a household interview but no personal or senior citizen interview could be conducted.

[3] These reports were divided into the following two components for the first time in the wave 3 documentation: a wave-specific data report (including a codebook) and a cross-wave User Guide. The PASS project team at the IAB is responsible for creating the cross-wave User Guide. As of wave 3, infas has created the documentation for the wave-specific data report, which is based on the wave 2 data report. The cross-wave User Guide documents the entire study, details the objectives and 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.

short overview of the innovations and characteristics of wave 9 (Chapter 1.3), the data report provides key figures on the wave's sample and response rates (Chapter 2). The data preparation process is described (Chapter 5), and an overview of the variables generated is presented (Chapter 4). Additionally, the weighting procedure is presented (Chapter 6). Separate tables list the frequencies of all of the variables included in the scientific use file that were recorded in wave 8 by their respective datasets (Volumes II through V).

1.2 Instruments and interview program

The information in PASS is collected using separate questionnaires for the household and individual levels. First, a household interview is conducted. This interview gathers information about the entire household. The target person for this household interview[4] was selected during the contact phase preceding the interviews. Personal interviews of the household members follow the household interview. The aim is to conduct a personal interview of each individual living in the household who is 15 years of age or older. House-hold members who are 65 or older receive a shortened version of the questionnaire (the senior citizens' questionnaire), which excludes questions that are irrelevant to that age group.

The survey instruments and interview program for wave 9 are based on those used in wave 8. However, individual questions and modules have been revised or newly developed (see Chapter 1.3 for an overview).

The PASS survey instruments are designed to allow not only repeat interviews of individuals and households but also first-time interviews[5].

Since wave 3, dependent interviewing has been used for certain questions to update information that the respondent had previously provided to avoid seam effects[6] in the repeat interviews and to increase data quality. Information about constant characteristics was generally not gathered again. Additionally, since wave 4, an integrated questionnaire for repeatedly interviewed households (HHalt) and first-time interviewed households (HHneu) has been used[7].

[4] The target person for the household interview should know as much as possible about general household issues, and target selection was based on the rules documented in the methods reports (Jesske Quandt, 2011; Jesske Schulz 2012; Jesske Schulz 2013; Jesske Schulz (2014); Jesske Schulz (2015)).

[5] First-time interviewed households include the following groups: (1) households from the refreshment and replenishment samples of the current wave; and (2) households that split off from households interviewed during previous waves (split-off households). (For further explanation, please see the wave 4 methods report (Jesske Quandt, 2011).)

[6] In a panel data, the number of changes observed at the interface (seam) between interviews conducted in sequential panel waves is often considerably higher than the number of changes observed within an interview (see Jäckle 2008).

[7] In this survey, split-off households are treated like new households.

The cross-wave PASS User Guide elaborates the individual instruments and interview program. The following section reviews the characteristics and innovations of wave 9.

1.3 Characteristics and innovations of wave 9

At this point we outline the characteristics of the eighth wave for users who are already familiar with the data from previous PASS waves.

The characteristics and innovations of wave 9 affect the questions asked in the household and personal questionnaires (e.g., change of reference periods, modification of individual questions and new question modules)[8], sample and data preparation.

1.3.1 Individual Questionnaire

The personal questionnaire updates the employment history information gathered since wave 2[9]. Wave 9 maintains the chronological retrospective surveying introduced in wave 4 (see section 1.3.1 in Berg et al., FDZ Datenreport 08/2011).

For the personal questionnaire in wave 9, new modules were developed, whereas the single modules of the previous waves were used again, and extensions of the existing modules were made. The most important changes in the personal questionnaire relate to the development of a minimum wage module. The **minimum wage** module contains questions adressing attitudes towards minimum wage and work time changes *(PML0100-PML0700)*. Participants with an hourly wage of less than 13 Euro (so called low income groups) in at least one of their employments (mini-job or employment subject to social security) were interviewed about the (planned) minimum wage (to estimate the current hourly wage, a seperate variable for the filter control was generated). Additionally the initial questions were also surveyed by currently unemployed people and low income self-employed people. The findings of the pretest led to a small modification of the wage and social participation questions. The questions about work time changes *(PML0400, PML0600, PML0700)* were placed after the quality of employment module.module. Therefor the module got split off and the questions were placed in different parts of the questionnaire. Questions concerning attitudes towards minimum wage *(PML0100)* were placed after the social participation module, to avoid possible spill-over effects to the reservation.

[8] Not all of the minor changes to the questionnaire (adding, modifying or deleting individual questions) are listed.

[9] This information is gathered using the so-called dependent interviewing method. In dependent interviewing, information that was provided during previous interview waves is included in the interview text of the current interview to determine whether the information must be updated.

Six moduls were deleted.

The module **social media** *(PSM0100* and *PSM0200)* got deleted, but questions concerning the role of social media continue to be surveyed in the search for work module *(PAS0920)* and the employment biography module *(ET2410-ET2420)*. However the item I who-knows-who of the social network list got deleted from the search for work module, as this social network was stopped in 2014.

- the attitude (work)module (PEO0200*). (*PEO0200**).
- the attitude (role models) module (*PEO0400**).
- the attitude (living standard) (*PLS0100-PLS2600*) and attitude (financial) module (finance) (*PEF0100**).
- the sport module (*PG1500-PG130* + *PSB0100-PSB1500*).

Questions which recently got included in addition to the minimum wage module:

- Attitudes (reciprocity) [*PEO1500**]

The **networks** module is one of the main modules newly developed for wave 9. The relevant questions were placed in different parts of the questionnaire. Questions about the role/importance of personal contacts when starting a new employment (*ET4000-ET4090*) were surveyed in the employment module. Furthermore additional questions about the role/importance of personal contacts during job search (*PAS0930*) were asked in the search for work module. Within the network module general questions about close friends were modified or newly developed (*PSK0280+PSK0290*) and added to the questionnaire.

The module **attitude (working hours)** (*PEO1200-PEO1300*) was taken up again from wave 5, however the connecting text passage was changed for *PEO1200*.

In the **employment biography** module a new question about current participation of unemployed people in measures was added (*AL1400*)), to ensure a representative picture of the unemployment situation following the official definitions of the Federal Employment Agency. Since wave 9 the analysis of gaps in the employment biography module also accounts for gaps which lie in the beginning of the examined period.

The **minijob** module was extended in wave 9. Additional questions concerning working hours (contractual/actual), overtime hours, amount of mini-jobs, occupational activity and the professional sector of the mini-job (*PMJ0100-PMJ1400*) were included. In case that participants have several mini-jobs, the questions refer to the mini-job the participants would rate as their „main mini-job“ (also see module quality of employment).

The module **agency contacts** was modified, so that only the initial questions concerning contact and search commitment (*PTK0200-PTK0500*)) correspond to the former module. The questions *PTK0600-PTK1600* were deleted and replaced by newly developed questions (*PTK1700-PTK2500*):): promoting offers, requirements of the job center, inquiries about the integration agreement, evaluation of the job center.

In the **measures (one-euro-job)** module the questions about knowledge of one-euro jobs adressing all participants (*PEE0100 and PEE0200*) got deleted. The module now only addresses participants who receive Unemployment Benefit II in the reference period or participants who executed an one-euro job in the previous wave. The new special code „Participant doesnt know any one-euro-jobs“ got integrated in *PEE0300*. The following questions about initiation and (non-) participation in the one-euro-job remained.

Within the **health** module all questions that are thematically assigned to the „sport“ module (*PG1500-PG1530*) were deleted, as well as the actual sport module (*PSB0100-PSB1500*), as already mentioned above. Furthermore questions about interest and participation in health courses were newly asked (*PG1600-PG1650*). In addition a few questions from the health priority module from wave 6 were included again (*PG1205-PG1290*; without *PG1220 and PG1240*). The health module also contains a new question about the type of health impairment (*PG0950*) based on the previous question *PG0900*, but due to the extreme modifications a new variable was generated.

According to the rotation plan in wave 9 the **religion** module is only addressed to new participants.

Questions about the amount of hospital stays (*PG0400*), as well as presenteeism (*PG1400*) were deleted in the **basic module**.

In the **attitude (life)** module the self-efficacy scale (*PEO0100*) was deleted.

The initial filter of question *PEO1500* of the **attitude (leisure time children)** has been modified. In wave 9 the questions were only addressed to new participants with children under 18 years.

1.3.2 Senior citizens questionnaire

Out of the list of modifications realized for the personal questionnaire the following modifications were also implied for the senior citizens questionnaire:

- In the **attitude (life)** module the self-efficacy scale (*PEO0100*) was deleted.
- The **health** module also contains a new question about the type of health impairment (*PG0950*) based on the previous question *PG0900*, but due to the extreme modifications a new variable was generated.

1.3.3 Household questionnaire

In the household questionnaire of wave 9 only minor changes were made to the modules related to the use of charitable food banks and education and participation packages, as well as Unemployment Benefits II.

The module **use of charitable food banks** which was added in wave 8 was deleted again in wave 9 (*HTA0100-HTA0200*).

The **education and participation package** module was compressed. Questions about information sources and foreign-language information material (*HBT0200, HBT0210*) were omitted, as well as questions about reasons for not claiming benefits despite applying for it (*HBT0$25*) and participation before receiving financial support (*HBT0$20*). In addition the final evaluation questions about the education and participation package of BuT-beneficiaries or BuT-applicants (*HBT0900, HBT0950, HBT1200*) were deleted.

Entry filter *AL2050* (as well as exit filter *HA0250, AL20500*) changed:

In the **Unemployment Benefit II** module reasons for claiming Unemployment Benefits II were only collected once per Unemployment Benefit II episode. Participants who already provided information about reasons for claiming Unemployment Benefits II in the previous wave, skipped the question in the current wave.

1.3.4 Sample and data preparation

In wave 9, as in previous waves, a refreshment sample was drawn from the Federal Employment Agency (BA) subsample[10]. The aims are to guarantee the representativeness of the BA sample in the cross-section and to observe enough new transitions into benefits, that is, into UB II, over time. For the refreshment sample, benefit units were drawn receiving UB II in July 2014 but not on the sampling date of the first, second, third, fourth, fifth,

[10] Wave 1 of PASS includes two subsamples: (1) a sample of households receiving UB II, which was drawn from the Federal Employment Agency (BA) process data; and (2) a general popu-lation sample, stratified by status, drawn from a database provided by the commercial provider MICROM.

sixth and seventh waves (see Chapter 2.1 and, on the concept of the refreshment sample, Trappmann et al., 2009, page 11 ff.). All of the households that were surveyed for the first time during wave 9 can be identified via the sample indicator (**sample**).

The data preparation was performed in close cooperation with the IAB. Basic procedures, such as updating datasets and correcting problems in the household structures, were discussed during the preparation process. Final decisions were made by the IAB.

The integration of the spell datasets into the module **employment** and the necessary preparatory steps were discussed and determined in agreement with the IAB. That procedure is documented in Chapter 5.7.

2 Key figures

This chapter provides a brief overview of important figures in the study, such as sample sizes (gross and net) and response rates. The panel sample is represented over the course of the previous waves. Figures are reported not only for both the original and replenishment samples but also for the complete study.

- Subsample 1 (BA sample) refers to the sample of benefits 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 sample drawn from the SGB II inflow between waves 1 and 2.
- Refreshment sample 2 (BA sample) is the sample drawn from the SGB II inflow between waves 2 and 3.
- Refreshment sample 3 (BA sample) is the sample drawn from the SGB II inflow between waves 3 and 4.
- Refreshment sample 4 (BA sample) is the sample drawn from the SGB II inflow between waves 4 and 5.
- Panel replenishment/supplement 1 (municipal register sample) is the sample drawn from the registration office inflows in ten new postcode regions during wave 5.
- Panel replenishment/supplement 2 (BA sample) is the sample drawn from the SGB II inflows in 100 new postcode regions during wave 5.
- Refreshment sample 5 (BA sample) is the sample drawn from the SGB II inflow between waves 5 and 6.
- Refreshment sample 6 (BA sample) is the sample drawn from the SGB II inflow between waves 6 and 7.
- Refreshment sample 7 (BA sample) is the sample drawn from the SGB II inflow between waves 7 and 8.
- Refreshment sample 7 (BA sample) is the sample drawn from the SGB II inflow between waves 8 and 9.

2.1 Sample size

Each sample in a panel begins with the interviewed households from the first survey wave. In PASS, the gross panel sample contains the interviewed households from wave 1 and the HHneu from the refreshment samples in waves 2 to 8. Only those households being interviewed for the first time that are willing to participate in the panel and are available for

repeat interviews are considered[11]. Agreement to participate in the panel is only recorded during the first interview. Confirmation of these households' willingness in subsequent waves is not required. In addition to confirming willingness, access to the panel is induced during the first interview by general willingness to participate, that is, by providing an interview. Measures to ensure the best possible selection-free access to the panel as part of PASS are described in detail in the methods and field reports of waves 1 to 9[12].

Wave 1 of PASS included 12,794 household interviews, of which 12,000 households agreed to participate in the panel. These wave 1 households constitute the sample for the beginning of the first tracking survey.

The panel concept in PASS assumes that new or split-off households emerge as individuals move out of panel households, which are considered separate households as soon as a household interview is conducted.

This design results in a higher number of households compared to the original sample. Details about the procedures for the PASS panel concept can be found under „split-off households". In addition to the expansion of the panel, loss of households can occur due to panel mortality. Households in which all respondents passed away or moved abroad are removed from the gross panel in subsequent waves. Moreover, panel losses may occur if no household interview could be conducted for a household for two consecutive waves. This situation arose for the first time at the end of wave 3 and affected the gross panel in waves 4 to 9[13]. The gross sample used for wave 9 included 10,011 panel households. That includes additionally HHneu from the refreshment sample (n=3,408) and newly formed split-off households in wave 8[14] (n=216) and wave 9 (n=393)[15].

The case numbers for the gross sample size of the respective survey waves and subsamples[16] are reported in the following table. In wave 9, at least one interview could be conducted for 8,021 households in the panel sample. In addition, 900 first-time household

[11] Willingness to participate in the panel is confirmed by the household reference person and is thus valid for all household members. Households that were willing to participate in the panel have allowed their addresses to be stored for the purposes of this study's repeat interviews.

[12] Vgl. Hartmann et al. (2008); Büngeler et al. (2009); Büngeler et al. (2010), Jesske Quandt (2011), Jesske Schulz (2012), Jesske Schulz (2013), Jesske Schulz (2014), Jesske Schulz (2015), Jesske et al. (2016 im Erscheinen).

[13] The survey institute change also influenced the panel gross in wave 4 because transmitting participant addresses from the IAB to infas required the target person's permission. For details on this procedure and its results, please refer to the methods report for wave 4 (Jesske Quandt, 2011).

[14] Split-off households which could not be interviewed in the wave before, were considered like temporary drop outs and should be interviewed again in the following wave. Cases which could not be realized in the following wave were considered like final drop outs.

[15] Case numbers for the gross sample see Methodenbericht wave 9 (Jesske et al. 2016).

[16] The case numbers contain all cases of the register file. Deviations to the method data are possible because of subsequent data checks and cleaning procedures.

interviews were conducted from the refreshment sample, of which 827 were willing to participate in the panel. In addition, the households interviewed for the first time in wave 9 include 179 split-off households that arose because of the subsamples in waves 1 – 8.

Table 1: Panel sample at the household level by wave and subsample[17]

	n	BA	Microm	BA-Refresh-ment 1	BA-Refresh-ment 2	BA-Refresh-ment 3	BA-Refresh-ment 4	EWO-supple-ment	BA-supple-ment	BA-Refresh-ment 5	BA-Refresh-ment 6	BA-Refresh-ment 7	BA Refresh-ment 8	Total
Wave 1	HH-Interview real.	6.840	5.990											12.794
	HH panel participation	6.452	5.548											12.000
Wave 2	Panel-HH brutto	6.520	5.611											12.131
	HH-Interview real.	3.491	3.897	1.041										8.429
	HH panel participation	3.360	3.766	1.003										8.129
Wave 3	Panel-HH brutto	5.851	5.150	1.010										12.011
	HH-Interview real.	3.754	3.901	694	1.186									9.535
	HH panel participation	3.576	3.777	669	1.145									9.167
Wave 4*	Panel-HH brutto	3.926	3.628	863	1.069									9.486
	HH-Interview real.	2.815	2.977	563	745	748								7.848
	HH panel participation	2.754	2.933	554	7275	723								7.691
Wave 5**	Panel-HH brutto	3.392	3.334	676	960	727								9.089
	HH-Interview real.	2.382	2.680	464	608	517	753	1.510	1.321					10.235
	HH panel participation	2.347	2.633	456	598	512	702	1.415	1.257					9.920
Wave 6	Panel-HH brutto	2.902	3.021	576	768	687	653	1.324	1.185	961				12.077
	HH-Interview real.	2.109	2.539	398	532	466	497	1.103	908	961				9.513
	HH panel participation	2.078	2.503	389	519	460	492	1.087	890	919				9.337
Wave 7	Panel-HH brutto	2.540	2.797	484	658	553	626	1.270	1.137	930	949			11.944
	HH-Interview real.	1.984	2.409	359	505	414	413	996	798	682	949			9.509
	HH panel participation	1.954	2.383	357	502	412	407	969	783	671	914			9.352
Wave 8	Panel-HH brutto	2.231	2.608	429	572	466	494	1.085	924	875	919			10.603
	HH-Interview real.	1.738	2.194	324	431	359	348	883	678	571	677	795		8.998
	HH panel participation	1.718	2.158	314	425	355	344	863	670	562	659	755		8.823
Wave 9	Panel-HH brutto	2.010	2.388	365	508	415	427	958	778	698	859	761		10.167
	HH panel participation	1.569	2.029	285	379	308	305	795	606	503	585	534	827	8.725

Quelle: *HH-Register* and *PENDDAT*; SUF IAB

* Reduction of the gross sample due to objection procedures

** Expansion of the gross sample by supplementation

The 8,921 household interviews conducted in wave 9 correspond to 13,271 personal interviews. The following table lists the distribution of respondents across subsamples and survey waves.

[17] 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.

Table 2: Panel sample size at the individual level by wave and subsample

Personal interview realised	Wave 1	Wave 2	Wave 3	Wave 4*	Wave 5**	Wave 6	Wave 7	Wave 8	Wave 9
Stichprobe	abs.	abs.	abs.	abs.	abs.	abs.	abs.	abs.	abs.
BA	9.386	4.753	4.913	3.958	3.394	3.048	2.861	2.447	2.242
Microm	9.568	6.392	6.207	5.016	4.511	4.245	4.001	3.591	3.348
BA-Refreshment 1		1.342	898	786	653	558	505	450	402
BA-Refreshment 2			1.421	983	822	719	688	593	540
BA-Refreshment 3				1.025	760	679	590	512	459
BA-Refreshment 4					1.019	716	599	502	449
EWO supplement					2.589	1.990	1.784	1.533	1.406
BA supplement					1.859	1.350	1.182	999	912
BA-refreshment 5						1.314	975	821	733
BA-refreshment 6							1.264	932	838
BA-refreshment 7								1.080	760
BA-refreshment 8									1.182
Total	18.954	12.487	13.439	11.768	15.607	14.619	14.449	13.460	13.271

Source: *p_register*; SUF IAB

* Panel sample size at the individual level by wave and subsample

** Expansion of the gross sample by supplementation

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

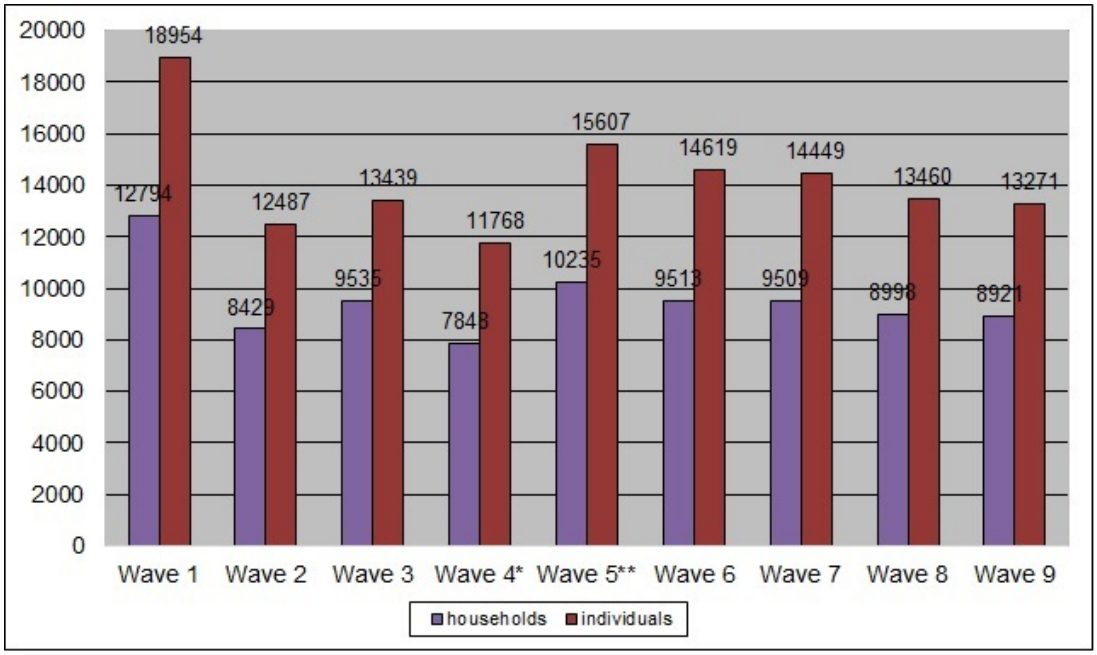
Table 3: Panel sample size of foreign-language interviews by wave

		Russian abs.	Turkish 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
Wave 6	Households	146	25
	Individuals	242	40
Wave 7	Households	145	29
	Individuals	245	43
Wave 8	Households	131	22
	Individuals	224	28
Welle 9	Households	111	14
	Individuals	187	27

Source: *PENDDAT*; SUF IAB

For the overall data pool of the realised panel sample, the following figure outlines households and individuals over the nine survey waves.

Figure 1: Realised panel sample for households and individuals by survey wave



2.2 Response rates

The response rate is calculated according to AAPOR standards (AAPOR, 2011). The response rate (RR1) is reported, which includes all cases of unknown eligibility in the denominator and therefore provides the minimum value of all response rates[18]. The response rate at the household level is calculated from the share of usable household interviews as a proportion of the total usable household interviews and non-neutral nonresponses. Only households in which all members have passed away or moved abroad permanently are considered cases of neutral nonresponse. Households are considered usable if at least one complete household interview is available. New households are considered usable if both the household interview and at least one complete personal interview are available. The following response rates were obtained at the household level for wave 9:

[18]This issue is addressed in very different ways in Germany. Frequently, a large number of individuals or households that were not interviewed are considered ineligible and are removed from the denominator when the response rate is calculated. When a sample is drawn from registers, 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 considered to have provided a neutral nonresponse. Moreover, the population of PASS is not restricted to German-speaking respondents or individuals who can be interviewed; therefore, the nonresponse reasons „does not speak German“or „respondent is sick/unable to be interviewed“cannot be considered cases of neutral nonresponse.

Table 4: Response rate for wave 8 at the household level by subsample

Wave 9	HH brutto abs. (%)	neutral nonresponse abs. (%)	HH brutto gross corrected* abs. (%)	HH-Interview realised* abs. (%)	of this HH willing to participate in panel abs. (%)
BA	2.231 (100)	6 (0,3)	2.225 (100)	1.586 (71,3)	1.569 (98,9)
Microm	2.608 (100)	8 (0,3)	2.600 (100)	2.063 (79,3)	2.029 (98,4)
BA-Refreshment 1	429 (100)	3 (0,7)	426 (100)	290 (68,1)	285 (98,3)
BA-Refreshment 2	572 (100)	3 (0,5)	569 (100)	387 (68)	379 (97,9)
BA-Refreshment 3	466 (100)	2 (0,4)	464 (100)	314 (67,7)	308 (98,1)
BA-Refreshment 4	494 (100)	3 (0,6)	491 (100)	313 (63,7)	305 (97,4)
EWO supplement	1.085 (100)	9 (0,8)	1.076 (100)	806 (74,9)	795 (98,6)
BA supplement	924 (100)	2 (0,2)	922 (100)	617 (66,9)	606 (98,6)
BA-Refreshment 5	875 (100)	5 (0,6)	870 (100)	507 (58,3)	503 (99,2)
BA-Refreshment 6	919 (100)	5 (0,5)	914 (100)	594 (65)	585 (98,5)
BA-Refreshment 7	3.134 (100)	5 (0,2)	3.129 (100)	544(17,4)	534 (98,2)
BA-Refreshment 8	3.408 (100)	32 (0,9)	3.376 (100)	900 (26,7)	827 (91,9)
Total	17.145 (100)	51 (0,3)	17.094 (100)	8.921 (52,2)	8.725 (97,8)

*HH brutto - neutral nonrespons (dead + moved to different country

Source: *HH-Register*; SUF IAB; for BA-Refreshment 8: Methods Data Set infas

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

The response rate within households indicates the average proportion of household members aged 15 or older within non valuable households for whom a complete personal interview is available.

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

Table 5: Average response rate among interviewed households by wave and subsample

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9
Sample	%	%	%	%	%	%	%	%	%
BA	85,6	85,5	83,1	88,4	88,7	89,3	89,2	89,3	88,9
Microm	84,2	85,1	83,6	88	88,4	88,6	88,4	88,6	88,0
BA-Refreshment 1		86,2	84,3	90,2	89,5	88,5	90,1	91	89,6
BA-Refreshment 2			84,2	88,3	89,3	88,5	88,8	88,3	88,7
BA-Refreshment 3				89,6	91,2	91,4	89,8	90,5	89,2
BA-Refreshment 4					89	92	90,6	91,3	90,2
EWO supplement					84,4	89,1	89,1	89	89,8
BA supplement					90	91,5	92	93,3	91,9
BA-Refreshment 5						89,9	90,7	91,3	91,4
BA-Refreshment 6							90,1	91,5	92,0
BA-Refreshment 7								90	91,3
BA-Refreshment 8									87,9
Total	84,9	85,4	83,5	88,5	88,3	89,5	89,5	89,9	89,4

Source: P-Register; SUF IAB

In addition to the between- and within-household response rates, the following table provides the repeat interview rate at the individual level. This value 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 through 8 with respond-ents who were willing to participate in the panel by subsample

			BA	Microm	BA-Refresh-ment 1	BA-Refresh-ment 2	BA-Refresh-ment3	BA-Refresh-ment 4	EWO supple-ment	BA supple-ment	BA-Refresh-ment 5	BA-Refresh-ment 6	BA-Refresh-ment 7	Total
Wave 2	Individ. willing panel part. W1	abs.	8.925	8.938										17.863
	re-interviewed individ. W2	abs.	4.274	5.828										10.102
	Share	%	47,9	65,2										56,6
Wave 3	Individ. willing panel part. W2	abs.	4.686	6.292	1.298									12.276
	re-interviewed individ. W3	abs.	3.365	4.955	820									9.140
	Share	%	71,8	78,8	63,2									56,6
Wave 4*	Individ. panel participation W3	abs.	4.844	6.100	894	1.380								13.218
	re-interviewed individ. W4	abs.	3.287	4.347	626	854								9.114
	Share	%	67,9	71,3	70	61,9								69
Wave 5	Individ. panel participation W4	abs.	3.946	5.004	785	979	993							11.707
	re-interviewed individ. W5	abs.	2.971	4.150	570	714	702							9.114
	Share	%	75,3	83	72,6	72,9	70,7							69
Wave 6	Individ. panel participation W5	abs.	3.378	4.468	645	819	756	957	2.439	1.786				15.248
	re-interviewed individ. W6	abs.	2.653	3.864	486	606	563	660	1.861	1.255				11.948
	Share	%	78,2	85,7	74,4	73,7	74,1	64,8	71,9	67,5				69
Wave 7	Individ. panel participation W6	abs.	3.034	4.216	555	711	667	712	1.973	1.337	1.264			14.469
	re-interviewed individ. W7	abs.	2.486	3.706	434	590	523	523	1.633	1.040	900			11.835
	Share	%	81,9	87,9	78,2	83	78,4	73,5	82,8	77,8	71,2			75,6
Wave 8	Individ. panel participation W7	abs.	2.837	3.979	504	687	588	597	1.745	1.167	970	1.219		14.293
	re-interviewed individ. W8	abs.	2.238	3.381	396	544	470	449	1.446	927	740	875		11.466
	Share	%	78,9	85	78,6	79,2	79,9	75,2	82,9	79,4	76,3	71,8		80,2
Wave 9	Individ. panel participation W8	abs.	2.439	3.547	442	589	509	499	1.508	994	818	916	1.027	13.288
	re-interviewed individ. W9	abs.	1.996	3.093	363	476	397	398	1.304	822	653	738	712	10.952
	Share	%	81,8	87,2	82,1	80,8	78,0	79,8	86,5	82,7	79,8	80,6	69,3	82,4

SOurce: *PENDDAT*; SUF IAB

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

2.3 Panel participation agreements, merging data and linking with process data

Respondent consent is always required to store addresses for repeat interviews in a subsequent wave and to merge survey data with the process data obtained from the Federal Employment Agency.

Panel participation agreement was explained in detail in Chapter 2.1. HHneu[19] consent to participate in the panel is illustrated as follows:

Table 7: First-time interviewed households* consent to participate in the panel by wave**

	Realised HH interviews with first-time interviewed HH	Realised HH interviews with first-time interviewed HH willing to participate in panel	Share willing to participate in panel
	abs.	abs.	%
Wave 1	12.794	12.000	93,8
Wave 2	1.087	1.048	96,5
Wave 3	1.328	1.285	96,8
Wave 4*	903	866	95,9
Wave 5**	3.688	3.476	94,3
Wave 6	1.112	1.068	96
Wave 7	1.130	1.089	96,4
Wave 8	967	921	95,2
Wave 9	1.056	981	92,9

Source: *PENDDAT* and *HH-Register*; SUF IAB
* 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

The consent to participate in the panel is recorded following the first personal interview in a new household during each wave. The information provided by that individual is assumed to apply to the household. That is, if the individual consents to participate in the panel, the household is considered willing to participate in the panel and if the individual does not agree to participate in the panel, the household is considered unwilling to participate in the

[19] All households in wave 1 are HHneu. Subsequently, only households from the refreshment samples and split-off households participating for the first time are considered HHneu. Therefore, since wave 2, households interviewed for the first time have been in the minority - the majority of household interviews conducted in these waves were conducted previously.

panel (see also Chapter 2.1)[20].

In contrast, permission to merge process data from 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. Consent to merging of these data is not obtained again in each wave[21].

Table 8 provides an overview of obtained consent to merge data in each wave. Only interviews in which consent to merge data was requested in that wave as part of the personal questionnaire are listed.

Table 8: Consent to merge data in personal interviews (respondents aged 15-65 years) obtained by wave

	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
Wave 6	2.482	2.002	80,7
Wave 7	1.973	1.613	81,8
Wave 8	1.653	1.327	80,3
Wave 9	1.727	1.471	85,2

Source: *PENDDAT*; SUF IAB
* Reduction of the gross sample due to objection procedures
**Expansion of the gross sample by supplementation
Basis: individuals 15 to 64 years of age

[20] One individual confirms household willingness to participate in the panel. The information available on the household level was integrated into the individual dataset (*PENDDAT*) during data preparation. The individual respondents in the household were assigned the correspond-ing information available for that household. The same procedure was applied during wave 2. In wave 1; however, consent was recorded after each individual and senior citizen interview; therefore, data could vary within a household. 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 the following year. If the respondent did not explicitly object to this notification, the household was considered to agree to participate in the panel and the panel variable in the individual dataset (*PENDDAT*) was updated accordingly.

[21] Due to filtering modifications, there were cases in which permission to merge data was raised again in waves 2 and 3 if the respondent had not previously agreed to that during the previous waves.

2.4 Split-off households

PASS is designed as a dynamic panel. Individuals who join or are born into the household are interviewed if they are at least 15 years old. Individuals who move out of sample households for one year or more should continue to be interviewed; however, these individuals are considered new, split-off households. These split-off households also become sample households in PASS. All individuals 15 years of age or more living in these households become target persons for personal interviews. If part of this split-off house-hold in turn splits off in subsequent waves, then this new split-off household also becomes a PASS sample household regardless of whether that new household contains anyone from the original sample (see infinite degree contagion model, Rendtel Harms 2009, 267). However, individuals who have moved abroad are removed from the survey because they no longer belong to this population and research questions specific to SGB II no longer apply. Individuals who leave the household for less than one year continue to be considered household members.

There are 1.107 split-off households from waves 1 to 9, of which 617 could be interviewed during wave 9, including 118 newly split-off households from wave 9 and 61 HHneu that could be identified in wave 8. Please refer to the methods report for wave 9 for further information about split-off households (Jesske et al. 2016).

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*), sampling year (*jahrsamp*), primary sampling unit (*psu*) and stratification (*strpsu*) of their original household.

3 Dataset structure

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

This data structure allows the information to be stored using relatively little storage space. The variables for each year can be identified immediately when examining the datasets. Identifying the merged additional information via key variables, such as household or personal identification numbers, is also quite simple. However, this common panel data structure increases the difficulty of working with these datasets. If analyses are conducted not only cross-sectionally but also longitudinally, then first, all of the relevant variables from each wave dataset must be integrated into a common dataset and care must be taken to ensure that the constructs are comparable for each year. For typical longitudinal analyses, the cross-wave dataset created in this way then must be reshaped into the so-called long format. Unlike the wide format, which contains a data matrix with one row per observation unit (e.g., the household or individual) and several datasets for each survey wave, in the long format, all of the waves assigned to an observation unit are arranged below one another. Rather than arranging 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 for the observation units.

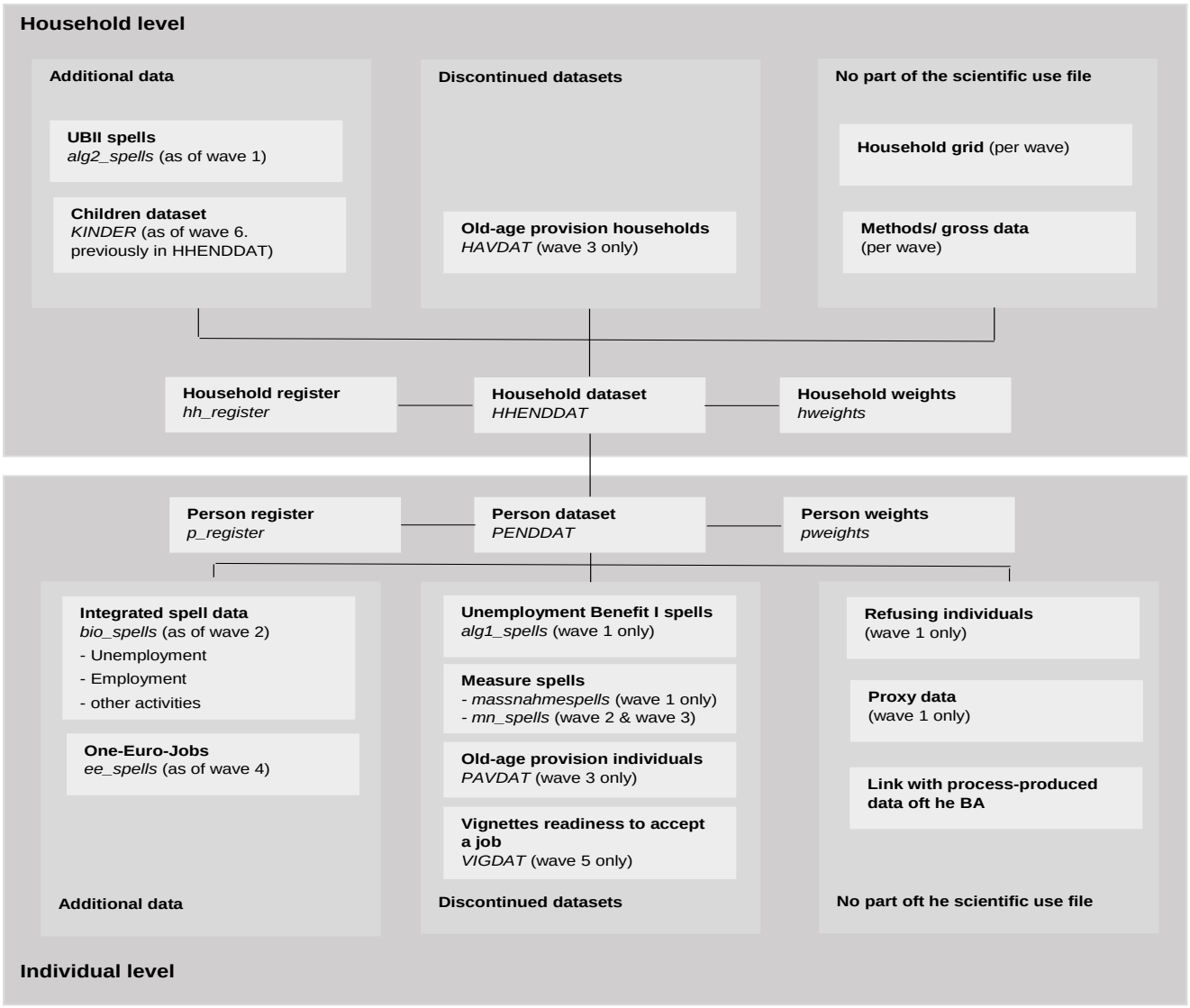
Reshaping the data into long format has both advantages and disadvantages. The decisive advantage of this variant is that this data structure is required for many longitudinal analyses (such as event history analyses). It is no longer necessary to invest additional time and effort creating a cross-wave file. The switch from long format to wide format is also quite easy to perform. STATA, for example, provides an option to switch between formats with little effort using the „reshape“command. Until a few years ago, the central argument against using this type of data structure was the significantly larger storage space required because even variables recorded in only one or a small number of survey waves require a complete column across all of the waves in the dataset. In addition, these long files become quite large with the increasing duration of the panel because all annual waves are appended, which significantly increases the storage space required and time needed to perform individual operations. The current wide availability of fast processors and large storage capacities even on simple desktop computers render this objection irrelevant. Another disadvantage occurs when merging additional data sources. Unlike datasets prepared in wide format, an additional variable is now required to identify an observation clearly. This variable may be a wave identifier in the household or individual datasets or the spell number in the spell datasets, which are also available in long format. Furthermore, it is not immediately apparent which variables were included in each wave because all variables are present in the dataset. These variables are assigned a special code (-9)

to identify waves during which they were not surveyed.

When the advantages and disadvantages of long format are weighed, the advantages of the long format clearly outweigh the disadvantages. Accordingly, household and individual PASS datasets (*HHENDDAT; PENDDAT*), corresponding weighting data (*hweights; pweights*) and a new dataset since wave 6 on children (*KINDER*) were prepared in long format.

At the household level, the scientific use file contains the data on household receipt of Unemployment Benefit II in spell form (*alg2_spells*). Since wave 4, the individual level has contained an integrated biographic spell dataset (*bio_spells*), that integrates and replaces the previous spell datasets *et_spells, al_spells* und *lu_spells*. Furthermore, a one Euro spell dataset (*ee_spells*) was introduced during wave 4. The household and person registers (*hh_register; p_register*) are available in wide format. During wave 5, the scientific use file was extended at the individual level by one dataset for the vignette module (*VIGDAT*) and was complemented by a dataset on resident children (*KINDER*), which includes household information. For further information on the structure of each dataset, please refer to the PASS User Guide (Fuchs 2013).

Figure 2: Dataset structure of PASS in wave 9



4 Generated variables

4.1 Coding 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, which differed from the original variable only insofar as the information from the open-ended responses could not be coded to the corresponding categories. Moreover, in some cases, new categories were created based on the information obtained 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” is replaced by “1.” The items on country of birth, nationality and parent/grandparent country of residence before migration were anonymised and assigned variable names[22]. Tables 9 and 10 provide an overview of the open-ended survey questions that were coded for wave9[23].

[22] *ogebland* (country of birth); *ostaatan* (nationality); *ozulanda* to *ozulandf* (parent/grandparent country of residence before migration).

[23] Variables for which information was obtained 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). Observations in waves without obtaining information on these variables were coded -9 (item not asked in wave) and documented in the survey wave data report.

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

Regular Variable name	Coded to variable	Dataset	Name
HD1100a-o	*HD1101a-o*	*HHENDDAT*	Other Employment status of HH members, proxy information, if necessary
HW0880a-i	*HW0881a-j*	*HHENDDAT*	Other reason for moving out, not listed
HT0510a-g	*HT0511a-g*	*KINDER*	Other type of group or club that a child is member of
AL20550a-h	*AL20551a-h*	*alg2_spells*	Other reasons for the beginning of UB II receipt
AL21300a-h to *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

Table 10: Coding responses to open-ended questions at the individual level in wave 9

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*	*PENDDAT*	Other foreign school qualification, not listed (first survey or not reported in previous wave)
PB1300a-j (Item I)	*PB1301a-j*	*PENDDAT*	Other German training qualifications not contained in the list (first survey or no statement in the previous wave)
PB1300a-j (Item J)	*PB1301a-j*	*PENDDAT*	Other foreign training qualifications not contained in the list (first survey or no statement in the previous wave)
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
ET2400	*ET2401*	*bio_spells*	Other source to get notice of a job
ET2420	*ET2421*	*bio_spells*	Other social network as source to get notice of a job
ET4020	*ET4021*	*bio_spells*	Different relationship to person acting as important source in job-search
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 reasons not contained in the list regarding why no job was searched
PTK1700a-i	*PTK1701a-i*	*PENDDAT*	Other support from job-center
PTK1800a-e	*PTK1801a-e*	*PENDDAT*	Other requirements for job center

Table 10: Coding responses to open-ended questions at the individual level in wave 9 (continued)

Regular Variable name	Coded to variable	Dataset	Name
PAS0900a-g	*PAS0901a-g PAS0901i*	*PENDDAT*	Other places where target pers. obtained information about job vacancies, not listed
PAS0950a-i	*PAS0951a-i*	*PENDDAT*	Other form of disability/impairment
PG1300	*PG1301*	*PENDDAT*	Other health insurance, not listed
PG1300a-e	*PG1301a-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
PA1100[24]	*freiz1-3*	*PENDDAT*	First to third leisure time activity
PA1200[25]	*frwunsch*	*PENDDAT*	Desired leisure time activity
PA1300a-g	*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

[25]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*.

[25]The variable *PA1200* 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 *PA1200* were included in the variable *frwunsch*.

4.2 Harmonisation

The survey instruments for some variables changed across waves. In particular, the integration of the module “employment biography” in wave 2 provided critical information on employment status, current main employment, status of economic inactivity and receipt of UB I in a different way than in wave 1. Since then, information has been collected not only for the date of the interview but also for particular periods.

To facilitate cross-wave analyses in such cases, variables are generated for important indicators, which are harmonised across waves. Harmonisation creates a special group within the generated variables (see Section 4.4) that is used to standardise indicators collected in different ways retrospectively.

Changes between the waves can affect the entire survey concept, categories and interviewed groups. Harmonised variables thus consider different source variables that result from changed survey concepts, categories or interviewed groups. This was an effort to standardise them across waves as much as possible before variables were generated.

Thus far, the simple classification for occupational status (*stibkz*) has been harmonised; however, the need harmonisation is expected to increase with the duration of the panel.

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

Variable	Subject area	Name
stibkiz	Employment	Current occupational status, simple classification, harmonisiert (anonymisiert)

Although explicitly harmonised variables also consider changes in categories and interviewed group across waves - in addition to changes in the survey concept - a second type of variable does not explicitly consider changes in the interviewed groups. These variables are generated for all waves but may contain information for different groups of respondents in each wave. These differences result from revisions to the filtering processes performed between waves and affect the source variables of generated variables.

Accordingly, cross-wave variables of this type apply in addition to harmonisations and standardise individual aspects across waves. In contrast to the harmonised variables, they are generated for each wave for all groups for which the corresponding source variables were collected. Thus, they can easily be used to evaluate 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.

Before working with cross-wave but not harmonised variables, it should be verified whether differences in the interviewed groups might cause problems in the evaluations, and it should be determined whether standardisation is necessary[26]. Subsequent cross-wave variables are different for the group for which they are generated.

[26] For example, in wave 1, the groups of respondents that were questioned about their employment were different from those questioned in the waves that followed. Accordingly, the respective groups that provided information about occupational status, occupational activities, working hours, fixed-term employment, etc., varied.

Table 12: Variables in the individual dataset (*PENDDAT*) 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 Portocarre-ro (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

At various times in both the household and personal interviews, information was gathered via dependent interviewing, i.e., interviews that were dependent on the responses provided during a previous wave. In this approach, data from the previous interview are used to control the filter questions or are integrated directly into the question text of the current interview.

Two main goals were pursued, utilising information from previous waves[27]. First, changes

[27] For example, individuals were only asked about their highest school qualification once. Only qualifications obtained since the previous interview were reported in subsequent waves.

that occurred since the previous wave were recorded, depending on the information available from the previous wave. At those points, information from previous waves was used to control the filter. Second, the respondent should have received information. In 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 clarify the definition of the reporting period[28]. In other places, especially where spell information was updated[29], the previous response was integrated into the question text to remind the respondent and prevent incorrect changes in status. Such changes are artifacts of the open-ended survey question arising out of inaccurate memories or imprecise information.

If information from a single wave in the dataset is reviewed, information is incomplete for some respondents due to dependent interviewing, which only represents the changes between survey dates. For respondents who are interviewed for the first time about a certain topic, complete information might be information available for that wave[30].

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

Table 13 and 14 provide a brief overview of the relevant updates to the questionnaires and indicate the variables for which updated information was obtained. Cases for which generated variables were updated or continued are listed in Chapter 4.4 of this data report.

[28] For example, if only new school qualifications were to be reported, the following question was asked: "Have you obtained a general school qualification since our last interview on [interview date of previous wave]?"

[29] Examples include updates of UB II receipts since the previous wave in the household interview or employment or unemployment updates in the individual interview.

[30] Individuals who were asked about their school qualifications for the first time reported their highest school qualification. Therefore, complete information on the highest school qualification is available for this wave in the recorded variable. In the subsequent wave, only newly obtained school qualifications are recorded. For example, if a school qualification is recorded, it is not clear whether it represents the individual's highest school qualification. In that sense, the information obtained in the subsequent wave is incomplete in its reported variables.

Table 13: Updated information in wave 9, household questionnaire

Construct	Q.No.	Note	Update in var.
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*: *bgbezs9; bgbezb9*

Table 14: Updated information since wave 9, personal questionnaire

Construct	Q.No.	Note	Updated in var.
Highest general school qualification	PB0220-PB1100	Updated in generated variable	*PENDDAT*: *schul1* (without responses to open-ended questions) *schul2* (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* (responses to open-ended questions)
Year of vocational qualification	PB1310a-k	Updated in generated variable	*berabj*
Periods of updated activities in the BIO spell dataset	BIO0600z1, BIO0600z2, BIO0400z, BIO0500z	Updated in the BIO spell dataset for attached spells	*bio_spells*: *BIO0400, BIO0500, BIO0600*
		Updated in the BIO spell dataset for attached spells	*bio_spells*: *ET2300, ET2700*
		Information on current employment, updated in generated variables	*PENDDAT*: *isco88; kldb; stib; stibkz; azhpt1; azhpt2; azges1; azges2; 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 Bene-fit 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 must be drawn between characteristics for which previously collected information is updated with information on changes between the survey dates and so-called constant characteristics that are not expected to change over time. Therefore, these characteristics are recorded only once in PASS, but in some cases, corrections are possible. Because information on these characteristics is usually only available for the surveyed variables during the first interview, they are subsequently provided in the form of generated variables (see Chapter 4.4, User Guide PASS Wave 6).

4.4 Simple generated variables

Simple generated variables include variables for which different items in a construct are surveyed separately for technical reasons and then aggregated. Alternatively, information from the current wave is combined with information from the previous wave (see Chapter 4.3), such as the highest educational qualification (see Chapter 4.3). Important information can also be obtained by merging partial datasets (e.g., indicators for current receipt of UB I or II).

The simple generated variables for households and individuals who are interviewed on a topic for the first time can always be generated based on information from the current wave. Households and individuals who provided information on a topic during a previous wave can be differentiated in the cross-section datasets (*HHENDDAT; PENDDAT*) to indicate the origin of the variables necessary for variable generation. The three different types of simple generated variables are provided in table 15.

Table 15: Simple generated variables in the cross-section datasets (*HHENDDAT; PENDDAT*) for households and individuals who previously provided information on the topic

Type	Generation based on source data from:		Description
	wave of the first survey of the topic for HH/individ.	current wave	
constant (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)
continued (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)
independent (new)	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)

Explanations that are more detailed must be provided on the type “unveränderlich (uv)” simple generated variables for *PENDDAT*. A first-time survey of a topic with an individual does not always take place during the first wave in which the individual provides an interview. Two groups of individuals are considered first-time interview respondents even if they provide a repeat interview.

The first group is individuals moving back into a household. Individuals who move out of their previous household to form a split-off household (see Chapter 2.4) take their preload information with them. Thus, they can be treated correctly as either first-time interviews or repeated interviews. However, if an individual returns from a split-off household into a panel household in which he/she lived during a previous wave, the preload of this individual is not transferred from the split-off household to the original household. Individuals returning home are treated as first-time interviewees. This situation has occurred since wave 3. The first move-outs of HHalt occurred during wave 2, and returns may occur by wave 3.

An individual preload for dependent interviewing is created for an individual (see Chapter 4.3) only if he/she provided an interview during one of the two preceding waves. The

context for this rule is that there is a point in time until which an individual is expected to remember the response in spell form. Individuals who last provided a personal or senior citizen interview during the third wave or earlier had passed this point. To reduce respondent stress and protect the validity of the information provided, which is presumably severely threatened beyond this limit, individuals whose reference date for information about spell results is before the relevant date are treated as first-time respondents[31]. This situation first occurred in wave 4 because that wave was the first time that a previous personal interview could have taken place more than two waves previously.

The information on which these generated variables are based is collected again for these two groups (e.g., in the module “social origin”) because they are treated as first-time interviews. Data preparation treats this survey information identically to the information from individuals engaged in actual first-time interviews within the PASS framework. These generated variables, e.g., the status of the mother and father, are thus based on information from the current wave. No transfer of information from previous waves takes place, and there is no attempt to make the data fit plausibly with previous information. We assume that the information provided by the target person, which is processed to become generated variables, is consistent with previous information in a repeated survey. However, deviations from previously obtained information in the previous waves cannot be generally excluded. Individuals included in either group are flagged in *PENDDAT* by the variable *altbefr* as first-time respondents (code “0” or “-9” for wave 1).

These simple generated variables are provided in tables 16 to 21. The tables include short descriptions of each variable. Furthermore, the source variables to generate the variable are indicated[32]. For the cross-section datasets (*HHENDDAT; PENDDAT*), additional information identifies the type of simple generated variable shown in Table 16 (uv; fs; neu). This division is not used for spell datasets because there are no wave-specific observations. Instead, variables are newly generated at the spell level if the spell was newly included in the wave or was updated with information obtained in the current wave. In addition, register datasets follow a different logic, and no further differentiation was made.

[31] Excluding previously granted consent to the merging of data. This preload information is generated regardless of when the previous personal interview was provided to avoid individuals negating question *RegP0100* and de facto withdrawing their consent. The option to with-draw consent to the merging of data remains unaffected by this decision.

[32] The data report documents how the variables in the cross-section datasets (*HHENDDAT; PENDDAT*) were generated for observations in previous waves. The documentation for specific 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 updated datasets. If a spell was not updated, the generated variables remain unchanged (with the exception that a special code was used in the censoring indicator if the spell could not be continued for technical reasons). If a spell was updated, then the most current information was used, i.e. the variables provided with information from the current wave or cross-section variables in the spells relevant for the current wave.

Table 16: Wave 9 simple generated variables in the household *HHENDDAT*) and *KINDER*) datasets (in alphabetical order)

Variable	Label and description	Source var. for gen. var wave 9
alg2abez	*Current receipt of UB II of the HH, generated*: Indicator for the household's current receipt of Unemployment Benefit II	*zensiert; AL20300; AL20400; AL20500 (alg2_spells)* information on further receipts of Unemployment Benefit II *(AL22700); hintjahr (HHENDDAT)*
anzgeschw	*Number of siblings in the household*: Indicator of an individual's number of siblings Parenthood and sibling status are surveyed separately. Individuals may share one parent but not call themselves siblings. Therefore in some cases, *anzgeschw* is not equivalent to sibling status, which can be generated through the parent indicator variable in *p_register*.	Information to relations in the household *household grid*
bik	*BIK region size classes (GKBIK10), generated*: The information on region size was generated by infas by converting the postcode from the address to *GKBIK10* (neu).	Supplied by survey institute
blneualt	*Western German States or Eastern German States, generated*: Divides the German states into the western states of the former FRG (excluding Berlin) and the eastern states of the former GDR (with Berlin). Infas determined the state based on the postcodes the address data (neu).	*bundesld* Information generated and supplied by the survey institute on the federal state in which the household is resident at the survey date.
butaber	*Eligibility for education package at point of interview*: This variable indicates that a household is eligible to draw benefits from the education and participation package if he draw one of the benefits like UB II, children's allowance, housing or social benefit since January of the year before the actual year of the survey (neu).	*AL20200; AL20400; AL20500 (alg2_spells); HA0250a-b; HW1800; HW1950; HEK0100; HEK0115; HEK1630; HEK1645 (HHENDDAT)*

Table 16: Wave 9 simple generated variables in the household (*HHENDDAT*) and *KINDER*-Datasets (in alphabetical order) (continued)

Variable	Label and description	Source var. for gen. var wave 9
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.*: This generated variable integrates information from categorised and openended survey questions on net household income (neu).	*HEK0600; HEK0700; HEK0800; HEK0900; HEK1000; HEK1100 (HHENDDAT)*
hintdat	*Date of household interview*: This generated variable indicates the date on which the household interview was conducted in the format YYMMDD (neu)	*hintjahr; hintmon; hinttag (HHENDDAT)*
hintnum	*interviewer in household interviews*: The artificial identifier indicates the interviewer who conducted the interview. This information is consistent between *PENDDAT* and *HHENDDAT* as well as across waves. A definite characteristic of the label always identifies the same interviewer (neu).	information that is generated and supplied by the survey institute
kindu4	*Control variable: child under the age of 4 in the HH*: A variable indicating 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 variable child under the age of 13 in the HH*: A variable indicating 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)*

Table 16: Wave 9 simple generated variables in the household (*HHENDDAT*) and *KINDER*-Datasets (in alphabetical order) (continued)

Variable	Label and description	Source var. for gen. var wave 9
kindu15	*Control variable: child under the age of 15 in the HH*: A variable indicating 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 irrele-vant 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)*)
kindu25	*Control variable: child under the age of 18 or pupils under the age of 25 in the HH.*: A variable indicating whether at least one individual in the household is under the age of 18 or that at least one individual is between the age of 18 and 25 and pupil. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this individual of the age group 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 used to generate the variable as well (neu).	*HD0200a - HD0200o;* categorical follow-up question about age group (in cases of no response in *HD0200*); *HD1100a-o (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 9 in the individual dataset (*PENDDAT*) (in alphabetical order)

Variable	Label and description	Source var. for gen. var wave 9
akt1euro	*Current part. in one-euro job, generated*: Indicator: respondent is participating in a one-euro job program at the time of the interview (neu).	*zensiert (ee_spells)*
alakt	*Currently reported as unemployed, generated (as of wave 2))*: Indicator: the TP was unemployed at the date of the personal interview of that wave (neu).	*zensiert; spintegr; BIO0101 (bio_spells)*
alg1abez	*Current receipt of UB I, generated*: Indicator: respondent is receiving Unemployment Benefit I at the interview date. In wave 6, the periods since January 2013 during which the respondent was unemployed were surveyed. For each spell, additional questions about whether and when the respondent received UB I (neu).	*AL0700; AL1000; AL1100; AL1200 (bio_spells)*
apartner	*Control variable: unmarried partner living in HH*: Indicator: respondent has a cohabitee or partner whose status is not specified in the household (neu).	Information on relationships between household members *(Haushaltsgrid); PD0500 - PD0800 (PENDDAT)*
azhpt1	*Current contractual working hrs. main employment (without marginal employment), gen* : Weekly contractual working hours provide the respondent's primary employment at the time of the interview. Generated from open-ended questions about working hours.	*ET2007 (bio_spells)*
azhpt2	*Act. effective working time main employment (without minijobs, incl. cat. statements), gen.*: Weekly effective working time of the main job that the respondent performed at the moment of the interview, which is generated using from open-ended questions about working hours and a categorical follow-up question in which irregular working hours were reported (neu).	*ET2107; ET2207 (bio_spells)*

Table 17: Simple generated variables for wave 9 in the individual dataset (*PENDDAT*) in alphabetical order (continued)

Variable	Label and description	Source var. for gen. var wave 9
azges1	*Current contractual working hrs. (without marginal employment), gen.*: Weekly contractual working hours for all positions held by the respondent at the time of the interview. Generated from open-ended questions about working hours.	*ET2007 (bio_spells)*
azges2	*Current total actual working hrs. (without marginal employment, incl. cat. info.), gen.* : Actual weekly working hours for all positions held by the respondent at the time of the interview. Generated from responses to open-ended questions on working hours and a categorical follow-up question in which irregular working hours were reported (neu).	*ET2107; ET2207 (bio_spells)*
befrist	*Current employment: limited contract? Generated (all waves)*: Indicator: the employment position held by the respondent at the interview date is on a limited contract (neu).	*PET2510a; PET2510b (PENDDAT)*
begjeewt	*Start year of first employment, generated*: The first year during which the respondent was employed in a regular position. To generate this variable, information about the first regular position was combined with information from the employment spells if the respondent had previously reported his/her first regular employment since January 2013 (uv).	For first survey: *bjahr (bio_spells); PET3200b (PENDDAT)* After first survey: *begjeewt* from previous wave *(PENDDAT)*
begjminj	*Start year of current mini-job, generated*: Year, since which participant is employed in current (main) mini-job (neu)	*PMJ0800b*
begmeewt	*Start month of first employment, generated*: The month during which the respondent first held regular employment (generated, see *begjeewt*) (uv).	For first survey: *bmonat (bio_spells); PET3200a (PENDDAT)*; After first survey: *begmeewt* from previous wave *(PENDDAT)*

Table 17: Simple generated variables for wave 9 in the individual dataset (*PENDDAT*) in alphabetical order (continued)

Variable	Label and description	Source var. for gen. var wave 9
begmminj	*Start month of current mini-job, generated*: Month, since which participant is employed in current (main) mini-job (neu).	*PMJ0800a*
berabj	*Year of the highest vocational qualification*: The year in which the respondent obtained his/her highest vocational qualification at the interview date (fs). *Note: The year in which the reported vocational qualifications reported in wave 1 but asked 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*: Identifies the highest vocational qualification obtained by the interview date by ranking 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*: Defined as in beruf1 with the following differences: 1. Inclusion of responses to open-ended questions; 2. Inclusion of foreign qualifications; and 3. Degrees are not distinguished by type of institution (e.g., university or other institution of higher education) but by level (Bachelor's degree; Master's degree; Ph.D.) (fs). .	For first survey: *PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601 (PENDDAT)* For repeated survey: *PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601 (PENDDAT)*
brges	*Current total gross income (without marginal employment, incl. cat. info.), gen.*: Contains the cumulative information on gross income from all employment ($>$ EUR 450). Generated from the answers provided in open-ended questions on gross income and categorical follow-up question when the "don't know" or "details refused" answers were provided to open-ended questions (neu).	*ET2804; ET2904; ET3004; ET3104; ET3204; ET3304 (bio_spells)*

Table 17: Simple generated variables for wave 9 in the individual dataset (*PENDDAT*) in alphabetical order (continued)

Variable	Label and description	Source var. for gen. var wave 9
brutto	*Gross income from the current main employment incl. categorised information, generated*: A generated variable integrating information from categorised and open-ended survey questions on gross income (neu).	*ET2804; ET2904; ET3004; ET3104; ET3204; ET3304 (bio_spells)*
bruttokat	*Categorised gross income from the current main employment, generated* : This variable aggregates the categorised information on gross income for a specific variable, which combines several items on income categories (neu).	*ET2804; ET2904; ET3004; ET3104; ET3204; ET3304 (bio_spells)*
emonlewt	*Time when last employment ended (month)*: Month in which the respondent was most recently employed. 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 2013 (fs).	For first survey: *PET1200b (PENDDAT); ejahr; emonat (bio_spells)* For repeated survey: textitejhrlewt from previous wave *(PENDDAT); ejahr; emonat (bio_spells)*
ejhrlewt	*Time when last employment ended (year)*: Year, in which the respondent was most recently employed. 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 2013 (fs).	For first survey: *PET1200b (PENDDAT); ejahr; emonat (bio_spells)* For repeated survey: *ejhrlewt* from previous wave *(PENDDAT) ejahr; emonat (bio_spells)*
ekin1517	*Control variable: own child aged between 15 and 17 in the household.*: A variable indicating whether 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)*

Table 17: Simple generated variables for wave 9 in the individual dataset (*PENDDAT*) in alphabetical order (continued)

Variable	Label and description	Source var. for gen. var wave 9
ekind	*Control variable: own child in HH*: A variable indicating whether 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*: A variable indicating whether 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).	Information on relationships between household members *(household grid)*
ekinu15	*Control variable: own child under the age of 15 in HH*: A variable indicating whether 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*: A variable indicating whether 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)*
epartner	*Control variable: spouse or registered partner in HH* : A variable indicating whether 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 450 per month), gen. (as of wave 2)*: A variable indicating whether the TP had an ongoing spell of employment at the time of the personal interview of the respective wave (i.e. employment earning >EUR 450) (neu).	*zensiert, spintegr, BIO0101 (bio_spells)*

Table 17: Simple generated variables for wave 9 in the individual dataset (*PENDDAT*) in alphabetical order (continued)

Variable	Label and description	Source var. for gen. var wave 9
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.*: A variable indicating 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 17: Simple generated variables for wave 9 in the individual dataset (*PENDDAT*) in alphabetical order (continued)

Variable	Label and description	Source var. for gen. var wave 9
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 about the mother's vocational qualification was asked only if the mother was not living in the survey household. If she was living in the household, this information was obtained from her personal interview.	For first survey: *PSH0300a-i (PENDDAT)* After first survey: *mberuf1* aus Vorwelle *(PENDDAT)*
mberuf2	*Highest vocational qualification attained by the mother, incl. mother in the HH, incl. information from open-ended survey questions, gen.*: Defined as in *mberuf1* except that responses to open-ended questions were also considered to generate *mberuf2* (uv).	For first survey: *PSH0301a-i (PENDDAT)* After first survey: *mberuf2* from previous wave *(PENDDAT)*
mhh	*Control variable: mother living in HH*: A variable indicating whether the respondent's biological mother, stepmother, adoptive mother or mother of non-specified status lives in the household (neu).	Information on relationships between household members *(household grid)*
migration	*Respondent's migration background, generated*: The following four categories were included in a generated variable for migration background: no migration background; personal migration (first generation); migration of at least one parent but no personal migration (second generation); migration of at least one grandparent but not the respondent or either parent (third generation) (uv). *Note: The concept for generating this variable has been revised as of wave 2. Previously, only the information on whether the respondent was born in Germany and which ancestor moved to Germany was collected. Now, information on whether an ancestor was born outside Germany and if applicable, which ancestor, is included. To guarantee consistency across waves, the variable for wave 1 was regenerated.*	For first survey: *PMI0100; PMI0700; PMI0800a-f; PMI0900a-f (PENDDAT)* After first survey: *migration* from previous wave *(PENDDAT)*

Table 17: Simple generated variables for wave 9 in the individual dataset (*PENDDAT*) in alphabetical order (continued)

Variable	Label and description	Source var. for gen. var wave 9
mschul1	*Highest general school qualification attained by the mother, incl. mother in HH, excl. information from open-ended questions, gen.*: In wave 1, the mother's highest academic qualification was inquired about only if the mother was not living within the survey household. If she was living in the household, this information was obtained from her personal interview (uv). As of wave 2, the mother's highest academic qualification has been asked of all newly interviewed individuals regardless of whether the mother was living in the survey household.	For first survey: *PSH0200 (PENDDAT)* After first survey: *mschul1* from previous wave *(PENDDAT)*
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 *mschul2* (uv).	For first survey: *PSH0201 (PENDDAT)* After first survey: *mschul2* from previous wave *(PENDDAT)*
mstib	*Mother's occupational status, code number, gen.*: The detailed occupational status of the mother was 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.*: This variable contains the accumulated information on net income from all employment positions ($>$ EUR 450), which is generated from the answers to open-ended questions on net income and a categorical follow-up question when respondents provided "don't know" or "details refused" answers to open-ended questions (neu).	*ET3404; ET3504; ET3604; ET3704; ET3804; ET3904 (bio_spells)*

Table 17: Simple generated variables for wave 9 in the individual dataset (*PENDDAT*) in alphabetical order (continued)

Variable	Label and description	Source var. for gen. var wave 9
netto	*Net income of the current main employment incl. categorised information, gen.*: A generated variable integrating information from categorised and open-ended survey questions on net income (neu).	*ET3404; ET3504; ET3604; ET3704; ET3804; ET3904 (bio_spells)*
nettokat	*Categorised net income from the current main employment, gen.*: This variable aggregates the categorised information on net income for a specific variable, which combines several items on income categories (neu).	*ET3404; ET3504; ET3604; ET3704; ET3804; ET3904 (bio_spells)*
palter	*Age (from PD0100), gen.*: The respondent's age is generated from the date of birth and date of the current personal interview (neu).	*PD0100; pintjahr, pintmon, pinttag (PENDDAT)*
panel	*Willingness to participate in the panel (neu)*: (neu).	Information supplied by the survey institute regarding the households' willingness to participate in the panel.
pintdat	*Date of personal interview*: This generated variable indicates the date on which the personal interview was conducted in the format YYMMDD (neu).	*pintjahr, pintmon, pinttag (PENDDAT)*
pintnum	*interviewer in personal interview*: The artificial identifier indicates the interviewer who conducted the interview. This information is consistent between *PENDDAT* and *HHENDDAT* as well as across waves. A definite characteristic of the label always identifies the same interviewer (neu).	Information that is generated and supplied by the survey institute.
schul1	*Highest school qualification, excl. foreign qualifications and information from open-ended survey questions*: This variable records the highest academic qualification. Equivalent Eastern and Western German qualifications were combined (e.g., EOS and Abitur), but information from open-ended questions was excluded (fs).	For first survey: *PB0200; PB0220; PB0230; PB0300; PB0400 (PENDDAT)* After repeated survey: *schul1* from previous wave ; *PB0200; PB0220; PB0230; PB0300; PB0400 (PENDDAT)*

Table 17: Simple generated variables for wave 9 in the individual dataset (*PENDDAT*) in alphabetical order (continued)

Variable	Label and description	Source var. for gen. var wave 9
schul2	*Highest school qualification, incl. foreign qualifications and information from open-ended survey questions*: Defined as in *schul1* with the following differences: 1. inclusion of responses to open-ended questions; and 2. inclusion of information about foreign qualifications (fs).	For first survey: *PB0200; PB0220; PB0231; PB0300; PB0401 (PENDDAT)* After 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 academic 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 9 was conducted before May 2015, it was assumed that the qualification was gained in 2014, if the interview was conducted later than May, the qualification was assumed to have been gained in 2015.*	For first survey: *PB0220; PB0230; PB0410; pintjahr; pintmon (PENDDAT)* After repeated survey: *schulabj* from previous wave ; *PB0220; PB0230; PB0410; pintjahr; pintmon (PENDDAT)*

Table 17: Simple generated variables for wave 9 in the individual dataset (*PENDDAT*) in alphabetical order (continued)

Variable	Label and description	Source var. for gen. var wave 9
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; azges2 (bio_spells)*
stib	*Occupational status, code number, generated*: A generated of the detailed code number for occupational status from the individual variables. A generated variable using information from the module "employment" (*ET060*-ET120**). 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).	*ET0607; ET0707; ET0807; ET0907; ET1007; ET1107; ET1207 (bio_spells)*
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 2013 (uv).	For first survey: *PET3300; PET3400; PET3500; PET3600; PET3700; PET3800; PET3900 (PENDDAT)* *ET0607; ET0707; ET0807; ET0907; ET1007; ET1107; ET1207 (bio_spells)* After first survey: *stibeewt* from previous wave *(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 2013 (fs).	For first survey: *PET1210; PET1220; PET1230; PET1240; PET1250; PET1260; PET1270 (PENDDAT)* *ET0607; ET0707; ET0807; ET0907; ET1007; ET1107; ET1207 (bio_spells)*

Table 17: Simple generated variables for wave 9 in the individual dataset (*PENDDAT*) in alphabetical order (continued)

Variable	Label and description	Source var. for gen. var wave 9
		After repeated survey: *stiblewt* from previous wave *(PENDDAT) ET0607; ET0707; ET0807; ET0907; ET1007; ET1107; ET1207 (bio_spells)*
vberuf1	*Highest vocational qualification attained by the father, incl. father in the HH, excl. open info., gen.*: A generated variable for father's highest vocational qualification analogous to *mberuf1* (uv).	For first survey: *PSH0600a-i (PENDDAT)* After first survey: *mberuf1* from previous wave *(PENDDAT)*
vberuf2	*Highest vocational qualification attained by the father, incl. father in the HH, incl. open info., gen.*: A generated variable for father's highest vocational qualification (incl. information from open-ended survey questions) analogous to *mberuf1* (uv).	For first survey: *PSH0601a-i (PENDDAT)* After first survey: *mberuf1* from previous wave *(PENDDAT)*
vhh	*Control variable: father living in HH*: Variable indicating that the respondent's natural father, stepfather, adoptive father or father of non-specified status is living in the household (neu).	Information on relationships between household members *(household grid)*
vschul1	*Highest general school qualification attained by the father, incl. father in HH, excl. information from* : A generated variable for father's highest general academic qualification analogous to *mschul1* (uv).	For first survey: *PSH0500 (PENDDAT)* After first survey: *vschul1* from previous wave *(PENDDAT)*
vschul2	*Highest general school qualification attained by the father, incl. father in household, incl. open info., gen.*: This generated variable records the father's highest general academic qualification (including information from open-ended survey questions) and is analogous to *mschul2* (uv).	For first survey: *PSH0501 (PENDDAT)* After first survey: *vschul2* from previous wave *(PENDDAT)*

Table 17: Simple generated variables for wave 9 in the individual dataset (*PENDDAT*) in alphabetical order (continued)

Variable	Label and description	Source var. for gen. var wave 9
vstib	*Father's occupational status, code number, generated*: The detailed occupational status of father is generated from individual variables (uv).	For first survey: *PSH0620; PSH0630; PSH0640; PSH0660; PSH0670; PSH0680 (PENDDAT)* After first survey: *vstib* from previous wave *(PENDDAT)*

Table 18: Wave 9 simple generated variables included in the spell dataset for Unemployment Benefit II *(alg2_spells)* (provided in the same order as in the dataset)

Variable	Label and description	Source var. for gen. var wave 9
bmonat	*Spell of UB II: start month, generated*: The month in which the spell of receiving Unemployment Benefit II began. If information was only available on the season when a spell began, the season was converted into a month to generate the variable. *Note:* The generated date variables were both checked for plausibility and corrected when necessary. The dates originally reported by the respondent have been included in the source variables as of wave 2. The season in which the spell began were recoded into months as follows: 21: 21 beginning of year/winter - January 24: 24 spring/Easter - April 27: 27 middle of year/summer - July 30: 30 autumn - October 32: 32 end of year - December	*AL20100 (alg2_spells)*

Table 18: Wave 9 simple generated variables included in the spell dataset for Unemployment Benefit II *(alg2_spells)* (provided in the same order as in the dataset) (continued)

Variable	Label and description	Source var. for gen. var wave 9
bjahr	*Spell of UB II: start year, generated*: The year during which the spell of receiving Unemployment Benefit II ended. Note: see *bmonat*	*AL20200 (alg2_spells)*
emonat	*Spell of UB II: end month, generated*: The month during which the spell of UB II receipts ended. To generate this variable, information about the season was converted into a month. For right-censored spells (i.e., spells that were 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*: The year during which the spell of Unemployment Benefit II ended. In the case of right-censored spells (i.e., spells that were ongoing when the household was interviewed), the interview year was entered. Note: see *bmonat*	*AL20400 (alg2_spells) hintjahr (HHENDDAT)*
alg2kbma-alg2kbmi	*UB II: 1st cut: start month, generated* to *UB II: 9th cut: start month, generated:* The month during which Unemployment Benefit II was reduced. To generate this variable, information about the season was converted into a month. *Note: These UB II reductions are embedded in spells of UB II receipts. Information on an individual benefit reduction 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 since wave 2.*	1st Benefit cut: *AL21000a (alg2_spells)* to 9th Benefit cut: *AL21000i (alg2_spells)*

Table 18: Wave 9 simple generated variables included in the spell dataset for Unemployment Benefit II *(alg2_spells)* (provided in the same order as in the dataset) (continued)

Variable	Label and description	Source var. for gen. var wave 9
alg2kbja-alg2kbji	*UB II: 1st cut: start year, generated* to *UB II: 9th cut: start year, generated:* The year during which the Unemployment Benefit II reduction began. *Note: see alg2kma - alg2kbmi*	1st Benefit cut: *AL21100a (alg2_spells)* to 9th Benefit cut: *AL21100i (alg2_spells)*
alg2kema-alg2kemi	*UB II: 1st cut: end month, generated* to *UB II: 9th cut: end month, generated:* The month during which the Unemployment Bene-fit II reduction ended. To generate this variable, information on the season was converted into a month. If the respondent reported the duration of the benefit reduction, this information was used to calculate the end date of the benefit cut based on the generated start date. *Note: see alg2kma - alg2kbmi*	1st Benefit cut: *alg2kbma; alg2kbja; AL21200a; AL21201a; AL21202a (alg2_spells)* to 9th Benefit cut: *alg2kbmi; alg2kbji; AL21200i; AL21201i; AL21202i (alg2_spells)*
alg2keja-alg2keji	*UB II: 1st cut: end year, generated* to *UB II: 9th cut: end year, generated:* Year in which the Unemployment Benefit II cut ended. If the respondent reported a duration for the benefit cut, this information was used to calculate the end date of the benefit cut *Note: see alg2kma - alg2kbmi*	1st Benefit cut: *alg2kbma; alg2kbja; AL21200a; AL21201a; AL21202a (alg2_spells)* to 9th Benefit cut: *alg2kbmi; alg2kbji; AL21200i; AL21201i; AL21202i (alg2_spells)*

Table 18: Wave 9 simple generated variables included in the spell dataset for Unemployment Benefit II *(alg2_spells)* (provided in the same order as in the dataset) (continued)

Variable	Label and description	Source var. for gen. var wave 9
AL22150a- *AL22150i*	*ALG2: 1st Benefit cut: which HH member's benefit was cut, gen.* to *ALG2: 9th Benefit cut: which HH member's benefit was cut, gen.*: This variable records which household members experienced reductions in Unemployment Benefit II. This is a string variable with 15 positions. Starting from the left, each position in this variable represents the position of one individual on 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 during the particular benefit reduction spell, the second position indicates whether the second individual's benefit was reduced, etc. Because source information for the generated variable was collected from wave 2 to wave 4, all 15 positions are coded "I" (i.e., item not asked in wave) for all benefit cuts reported during the first wave and since wave 5 (see below). Each of the 15 positions of this variable, which represent one of a maximum of 15 individuals in the household, is assigned one of the following codes indicating each individual' benefit status.	

Table 18: Wave 9 simple generated variables included in the spell dataset for Unemployment Benefit II *(alg2_spells)* (provided in the same order as in the dataset) (continued)

Variable	Label and description	Source var. for gen. var wave 9
	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 that had not been re-interviewed in the subsequent waves up to the current wave.* *(b) A household surveyed in wave 9 reports that a spell of UB II is still ongoing on the interview date in wave 9, or an end date is reported that is identical to the interview date in wave 9 and it is confirmed in the follow-up question that the benefit receipt is still currently ongoing.*	*AL20300; AL20400, AL20500 (alg2_spells)*

Table 19: Simple generated variables for wave9 in the BIO spell dataset *(bio_spells)* (in the same order presented in the dataset)

Variable	Label and description	Source var. for gen. var wave 9
bmonat	*Employment: start month, generated* The month during which the employment spell began. To generate the variable information on the season was converted into a 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 began 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)*

Table 19: Simple generated variables for wave9 in the BIO spell dataset *(bio_spells)* (in the same order presented in the dataset) (continued)

Variable	Label and description	Source var. for gen. var wave 9
bjahr	*Employment: start year, generated* The year during which the employment spell began. Note: see *bmonat*	*BIO0300 (bio_spells)*
emonat	*Employment: end month, generated* The month during which the employment spell ended. To generate the variable information on the season was converted into a month and for right-censored spells (i.e., spells that were ongoing when the individual was interviewed), the interview month was entered. Note: see *bmonat*	*BIO0400, BIO0600 (bio_spells); pintmon*
ejahr	*Employment: end year, generated* The year during which the employment spell ended. For right-censored spells (i.e., spells that were ongoing when the individual was interviewed), the interview month was entered. Note: see *bmonat*	*BIO0500, BIO0600 (bio_spells); pintjahr*
zensiert	*Employment: spell still currently ongoing (right censoring)* The censoring indicator shows whether a spell was 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 considered censored if one of the following conditions is met:* *(a) the individual reports an end date of the BIO spell that the employment is ongoing on the interview date.* *(b) Alternatively, when a reported end date is identical to the interview date, the follow-up question confirms that the activity is ongoing.*	

Table 20: Wave 9 simple generated variables included in the one-euro spell dataset *(ee_spells)* (in the same order presented in the dataset)

Variable	Label and description	Source var. for gen. var wave 9
bmonat	*Measure: start month, generated* The month during which the active labour market policy spell began. To generate this variable, information about the season was converted into a month. *Note*: The generated date variables were checked for plausibility and corrected if necessary. The dates reported by the respondent (excluding identified implausible values) are included in the source variables. Seasons during which the spell began 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	*EE0600a (ee_spells)*

Table 20: Wave 9 simple generated variables included in the one-euro spell dataset *(ee_spells)* (in the same order presented in the dataset) (continued)

Variable	Label and description	Source var. for gen. var wave 9
bjahr	*Measure: start year, generated* The year during which the active labour market policy spell began. Note: see *bmonat*	*EE0600b (ee_spells)*
emonat	*Measure: end month, generated* The month during which the active labour market policy ended. To generate the variable, information about the season was converted into a month. For right-censored spells (i.e., spells that were ongoing at the time of the interview), the interview date was entered. Note: see *bmonat*	*EE0600a; EE0600b; EE0700; EE0800a; EE0800b (ee_spells)*; *pintmon, pintjahr (PENDDAT)*
ejahr	*Measure: end year, generated* The year during which the active labour market policy spell ended. For right-censored spells (i.e., spells that were ongoing when the individual was interviewed), the interview date was entered. Note: see *bmonat*	*EE0600a; EE0600b; EE0700; EE0800a; EE0800b (ee_spells)*
zensiert	*Measure: spell still currently ongoing (right censoring)* The censoring indicator records whether a spell was ongoing at the time of the personal interview during the previous wave, i.e., whether this is a right-censored spell.	*EE0700 (ee_spells)*

Table 21: Wave 9 simple generated variables included in the person register dataset *(p_spells)* (in alphabetical order)

Variable	Label and description	Source var. for gen. var wave 9
alter9	*individual's age in wave 9 (2015)* A variable contains the best available information about an individual's age. This is either (a) the age calculated from the date of birth reported in wave 9 or (b) the age reported in the household interview if no date of birth is available from wave 9. The information from alter8 is transferred to the household dataset, which corresponds to the information in *HD0200a* to *HD0200o*. This procedure is consistent with conventions in the field. Even during the fieldwork, age was populated using the best available information. During fieldwork, the age variable is first populated using the age information obtained from the household interview. If a personal interview is conducted, this variable is overwritten in the database using the age calculated from the details obtained in the personal interview (date of birth, date of personal interview). The age information provided in the household and individual datasets are based on this variable. The best age information included in the household dataset for wave 9 was considered during the plausibility checks as well as generating the benefit unit and household type.	*PD0100; pintjahr; pintmon; pinttag (PENDDAT)*; *HD0200a* to *HD0200o (HHENDDAT)*

Table 21: Wave 9 simple generated variables included in the person register dataset *(p_* spells) (in alphabetical order) (continued)

Variable	Label and description	Source var. for gen. var wave 9
erwprox9	*Employment status according to HH interview in wave 9 (2015)* This variable is transferred unchanged as *HD1101** from the current wave from the *HHENDDAT* dataset.	*HD1101**
kinddat9	*Person included in the KINDER dataset in wave 9 (2015)* This variable indicates whether an individual is included in the *KINDER* dataset. Included in the *KINDER* dataset: All children aged under 15 years. Starting from wave 6 also all household members aged between 16 and under 25 years, for proxy variables surveyed in the modules social inclusion and education and participation packages.	*pnr (KINDER)*
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 their sex was corrected in the household interview.	*HD0100a* of all waves *HD0100o (HENDDAT)*
lastint	*Survey wave of last interview at individual level* This variable indicates the wave in which the last individual interview was conducted (personal or senior citizen interview).	Personal interviews from all waves *PENDDAT*
neuj9	*Year in which individual joined current HH, reported in wave 9 (2015)* This variable indicates the year during which an individual joined the current household of which he/she is a member reported during wave 9. Note: The wave 9 interview with the re-interviewed household provides that date when the individual moved or was born into the household since the previous wave.	Information on the date since which an individual has belonged to a household. Surveyed in the household grid
neum9	*Month in which individual joined current HH, reported in wave 9 (2015)* This variable indicates the month that the individual joined the household of which he/she is a current member. Note: see *neuj9*	Date an individual joined a household. Surveyed in the household grid.

Table 21: Wave 9 simple generated variables included in the person register dataset *(p_* spells) (in alphabetical order) (continued)

Variable	Label and description	Source var. for gen. var wave 9
wegj9	*Year since which individual has no longer been living in previous HH, reported in wave 9 (2015)* This variable indicates the year that the individual ceased to be a member of the household of the previous wave. Note: Information on the date comes from the wave 9 interview with the household in which the individual was living in the previous wave.	Date an individual ceased to belong to a household. Surveyed in the household grid.
zdub9	*Pointer: Personal identification no. of the individual doubled by the TP in wave 9 (2015)* 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: For matchings with the *p_register* via the personal identification number, one must first generate a match variable equalling *zdub**, if it exceeds 0, or otherwise equalling *pnr*. Chapter 5.4.1.2 of the data report for wave 5 of PASS provides a detailed explanation on the reasons for the introduction of this variable.	Information on all original household members of an original household and all of its split- off households are included in the household grid of the current and the previous waves.
zmhh9	*Pointer: Personal ID number of target person's mother in HH in wave 9 (2015)* Contains the personal identification number of the mother if she is living in the household. Biological mothers, stepmothers, adoptive or foster mothers and mothers whose status is not specified are considered mothers.	Relationships between household members (household grid).
zparthh9	*Pointer: personal ID number of target person's partner in HH in wave 9 (2015)* Contains the personal identification number of a partner living in the household. Spouses, registered partners, cohabitees and partners whose status is not specified are considered partners.	Relationships between house-hold 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.	The individuals living in a household across waves (household grid).

Table 21: Wave 9 simple generated variables included in the person register dataset *(p_spells)* (in alphabetical order) (continued)

Variable	Label and description	Source var. for gen. var wave 9
zvhh9	*Pointer: Personal ID number of target person's fa-ther in HH in wave 9 (2015)* Contains the personal identification number of the father if he lives in the household. Biological fathers, stepfathers, adoptive or foster fathers and fathers whose status is not specified are considered fathers.	Relationships between house-hold members (household grid).

The individual-level datasets contain a multitude of generated and constructed variables, including variables (e.g., occupational status) that are recorded in more than one dataset. Figure 3 provides an overview of both the simple and complex generated variables at the individual level.

Figure 3: Overview of generated variables for wave 9 at the individual level

	PENDDAT						BIO-Spells	EE_Spells
	Current status	Employment history		Social origin		€450 job	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	siopseewt	msiops	vsiops		siops	
Occupational status	stib	stiblewt	stibeewt	mstib	vstib		stib	
	stibkz							
Date of employment			begmeewt			begmminj	bmonat	bmonat
			begjeewt			begjminj	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	iscominj	isco88	
	kldb	kldblewt	kldbeewt	mkldb	vkldb	kldbminj	kldb	
Employed in which industry	branche					brancheminj	branche	

		PENDDAT					BIO-Spells	EE_Spells
		Current status	Employment history		Social origin		€450 job	Employment and unemployment biography	One-euro job participation
			last employment	first employment	mother	father			
Income	netges								
	brges								
	netto								
	nettokat								
	brutto								
	bruttokat								
Benefit receipt	alg1abez							alg1akt	
Household context and civil status	hhgr								
	famstand								
	vhh								
	mhh								
	apartner								
	epartner								
	ekind								
	ekin614								
	ekinu15								
	ekinu18								
	ekin1517								
	kindzges								
	kindzihh								
Migration background	ogebland								
	ostaatan								
	ozulanda								
	ozulandb								
	ozulandc								
	ozulandd								
	ozulande								
	ozulandf								
	migration								
Information on individual	gebhalbj								
	palter								
	zpalthh								
	zpsex								
Health	pcs								
	mcs								
General	altbefr								
	fb_vers								
	panel								
	pintdat								
	RegP0100								
	sample								
Leisure time behaviour	freiz1								
	freiz2								
	freiz3								
	frwunsch								

4.5 Constructed variables

Constructed variables are generated variables that require more extensive coding or recoding. In most cases, these variables have been empirically tested elsewhere and are based on theoretical concepts. At least some of these are standardized instruments used in social sciences or economics, such as the European Socio-economic Classification (ESeC), the International Standard Classification of Education (ISCED) or equivalised household income. This chapter provides detailed descriptions of the constructed variables made available in the PASS data, along with a short overview of the theoretical background and the most important references.

4.5.1 Individual Level

Education in years

Variable name	*bilzeit*
Variable label	Duration of school education and vocational training in years, generated
Source variables	*schul2; beruf2*
Type / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	For many statistical models, a linear variable for education and training is more appropriate than a categorical variable. For school qualifications, it is easy to convert categorical data to linear data. The linear value simply corresponds to the time spent in school until attainment of the final qualification. Care must be taken to ensure that equivalent qualifications are assigned identical durations. An upper secondary school certificate, for example, should always be labeled with the same duration regardless of whether it was obtained after twelve or thirteen years of education. Final qualifications were assigned the following durations: Lower secondary school certificate, lower secondary school certificate from the former GDR (POS) after completion of grade 8: 8 years intermediate secondary school certificate from the former GDR (POS) after completion of grade 10: 10 years Entrance qualification for university for applied sciences: 12 years General qualification for university or subject-specific higher education entrance (including EOS—similar qualification in the former GDR): 13 years

Education in years (continued)

	Vocational qualifications differ because of their numerous, different requirements and potentially large differences in income even for qualifications with similar training duration. 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 particular vocational qualification (see e.g., Helberger, 1988). This study adopts a similar approach. 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 education. Training as a semi-skilled worker: +1 year Apprenticeship, vocational school, school for health care occupations: +1.5 years Master craftsman certificate:+3 years Vocational academy: +3 years Applied sciences/Bachelor's degree: +3 years University/Master's degree: +5 years Ph.D.: +8 years Other German qualification: +1.5 years Other foreign qualification: +1.5 years
Literature:	Helberger (1988)

Education in years, mother

Variable name	*mbilzeit*
Variable label	Duration of school education and vocational training of mother 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 parents' years of education and training variables, the values added for vocational qualifications differ from those used to construct the corresponding variable for the respondents because information on vocational education/training was collected in less detail for parents (especially for tertiary education). The following values are assigned to particular courses of education/training: Training as a semi-skilled worker: +1 year

Education in years, mother (continued)

	Apprenticeship, vocational school, Health care occupations: +1.5 years Master craftsman certificate: +3 years Vocational academy: +3 years University, 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 of father 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 parents’ years of education and training variables, the values added for vocational qualifications differ from those used to construct the corresponding variable for the respondents because information on vocational education/training was collected in less detail for parents (especially for tertiary education). The following values are assigned to particular courses of education/training: Training as a semi-skilled worker: +1 year Apprenticeship, vocational school, Health care occupations: +1.5 years Master craftsman certificate: +3 years Vocational academy: +3 years University, 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 Müller,. 1987). An updated version is now available (Brauns	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 academic and vocational qualifications internationally (König, Lüttinger Steinmann, 1999). The procedures applied in the panel to recode qualifications according to the CASMIN classification, especially for problematic cases, follow the procedures described in Lechert, Schroedter and Lüttinger (2006) and Granato (2000). The slightly differing category values of the education variable in this dataset are considered. Details are presented in the table below. Cells containing valid CASMIN combinations are highlighted in light gray, whereas those containing missing values are dark grey.
Literature:	Brauns et al. (1999); Granato (2000); König et al. (1987); Lechert et al. (2006)

CASMIN (Befragte)

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

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 (above). Because the education variable has different category values for respondents and their parents, the coding pattern for *mcasmin* and *vcasmin* differs slightly from the pattern used in *casmin*. The following table details the differences.vgl. CASMIN
Literature:	Brauns et al. (1999); Granato (2000); König et al. (1987); Lechert et al. (2006)

CASMIN (Eltern)

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

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 (above). Because the education variable has different category values for respondents and their parents, the coding pattern for *mcasmin* and *vcasmin* differs slightly from the pattern used in *casmin*. The following table details the differences.
Literature:	Brauns et al. (1999); Granato (2000); König et al. (1987); Lechert et al. (2006)

CASMIN (Eltern)

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

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	The ISCED-97, (International Standard Classification of Education) developed by the OECD (OECD 1999; for an outline, see also BMBF, 2003), is an education classification alternative to CASMIN. Note that the coding for the ISCED-97 classification includes categories that cannot reasonably be assigned to these data. The ISCED values “0” (pre-primary education/kindergarten) and “1” (primary education) do not apply because the respondents are at least 15 years old. Instead, a separate group was created for individuals with an education below ISCED level 2 (ISCED 2 = lower or intermediate secondary school certificate). Therefore, only ISCED levels 2 to 6 are coded 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.
Literature:	BMBF (2003); OECD (1999)

ISCED 97 (Befragte)

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

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 variable generation details, see ISCED-97. In contrast to the ISCED-97 coding applied to respondent education, it is not possible to generate 6 ISCED levels for parents because data on the corresponding qualifications (i.e., Ph.D. or equivalent) were not collected for parents. Therefore, only ISCED levels 2 to 5 are coded in this dataset. The following table provides the coding details.
Literature:	BMBF (2003); OECD (1999)

ISCED 97 (Eltern)

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

VISCED 97

	VISCED 97
Variable name	*visced97*
Variable label	Education of father classified acc. to *isced97*, updated version, generated
Source variables	*vschul2; vberuf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	Zum theoretischen Hintergrund und zur Generierung vgl. ISCED-97. For the theoretical background and variable generation details, see ISCED-97. In contrast to the ISCED-97 coding applied to respondent education, it is not possible to generate 6 ISCED levels for parents because data on the corresponding qualifications (i.e., Ph.D. or equivalent) were not collected for parents. Therefore, only ISCED levels 2 to 5 are coded in this dataset. The following table provides the coding details.
Literature:	BMBF (2003); OECD (1999)

ISCED 97 (Eltern)

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

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

Generated:	Employment - Variable name - Source variables
	Current *(PENDDAT) - isco88 - ET2500* Spell data *(bio_spells) - isco88 - ET2500* first *(PENDDAT) - iscoeewt - ET2500, PET1280, PET3950* last *(PENDDAT) - iscolewt - ET2500, PET1280* of father *(PENDDAT) - visco - PSH0800* of mother *(PENDDAT) - misco - PSH0700* Minijob - *iscominj - PMJ0900*
Variable label:	Current Empl.: ISCO 88 (ZUMA-Coding), gen. Spell data: *(bio_spells)*: ISCO 88 (ZUMA-Coding), gen. first Empl.: ISCO 88 (ZUMA-Coding), first employment, gen. last Empl.: ISCO 88 (ZUMA-Coding), last employment, gen. Father: ISCO 88 (ZUMA-Coding) of the father, gen. Mother: ISCO 88 (ZUMA-Coding) of the mother, gen. Minijob: ISCO 88, current Minijob, gen.
Category / dataset	Occupation / individual-level data
Prepared by	Bernhard Christoph
Explanation	The International Standard Classification of Occupations (ISCO) was developed by the International Labour Organization (ILO) to allow international comparison. An advantage of the ISCO-88 is that in addition to the employment, the qualification level generally necessary to perform the job is also considered when assigning an occupation to a particular occupational code. This constitutes a major difference from 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)(KldB92)

Generated:	Employment - Variable name - Source variables
	Current *kldb - ET2500* Spell data *(bio_spells) - kldb - ET2500* first *(PENDDAT) - kldbeewt - ET2500, PET1280, PET3950* last *(PENDDAT) - kldblewt - ET2500, PET1280* of father *(PENDDAT) - vkldb - PSH0800* of mother *(PENDDAT) - mkldb - PSH0700* Minijob - *kldbminj - PMJ0900*
Variable label:	actual empl.: Classification of Occupations 1992, current employment Spell data: *(bio_spells)*: Classification of Occupations 1992, gen. first empl.: Classification of Occupations 1992, gen. last empl.: Classification of Occupations 1992, gen. Father: Classification of Occupations 1992, gen. Mother: Classification of Occupations 1992, gen. Minijob: Classification of Occupations 1992, gen.
Category / dataset	Occupation / individual-level data
Prepared by	Bernhard Christoph
Explanation	The KldB92 is the current version of the Classification of Occupations published by the German Federal Statistical Office (Statistisches Bundesamt). This classification system was developed to match the German occupational structure, which is based solely on employment.
Literature:	StBA (1992)

Erikson, Goldthorpe and Portocarrero (EGP) Class Scheme

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.

Erikson, Goldthorpe and Portocarrero (EGP) Class Scheme (continued)

	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 among the most common instruments for operationalising class. For this variable, data are coded by ISCO-88 occupational classification and occupational status. The coding procedure is based on an earlier approach elaborated by Christoph et al. (2005), who provide a detailed description of the procedure. Here, in contrast, unpaid family workers were not coded as self-employed but as individuals in dependent employment consistent with the coding applied in the European Socio-Economic Classification (ESeC), which is described in the next section. One difference between the EGP coding applied here and the ESeC coding is that in the EGP coding procedure, cases are “missing” (-7) in which the occupational activity seemed incompatible with occupational status (e.g., “directors and chief executives” [ISCO=1210] who reported that they were “employees performing simple duties” [StiB=51]). To ensure compatibility with the standardised coding procedure we adopted, we did not apply a comparable revision procedure using the EseC codes.
Literature:	Christoph et al. (2005); Erikson & 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, ET1103,ET1104,ET1105, ET1300,ET1301,ET1302, ET1303,ET1304,ET1305* first - *eseceewt - iscoeewt, stibeewt, PET1261* last - *eseclewt - iscolewt, stiblewt, PET3801* of father - *vesec - visco, vstib, PSH0670* of mother - *mesec - misco, mstib, PSH0370*

European Socio-economic Classification (ESeC) (continued)

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	The European Socio-economic Classification is largely based on the EGP class scheme. Unlike the latter, great importance was attached to international comparability of the operationalisation and validation of the classification (for a general description, see Rose & Harrison, 2007; for Germany, see Müller et al. 2006, 2007). The Stata do-file required to generate the ESeC was kindly provided by Heike Wirth from GESIS-ZUMA (Fischer & Wirth 2007). We simply adjusted the file to meet the requirements of this study. This do-file, originally written in standard SPSS syntax by Harrison and Rose (2006) as a standard program to generate the ESeC, was converted into Stata.
Literature:	Fischer & Wirth (2007); Harrison & Rose (2006); Müller et al. (2006, 2007); Rose & 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 empl. gen. Spell data *(bio_spells)*: Magnitude-Prestige Scale , gen. first empl..: Magnitude-Prestige Scale , first employment, gen. last empl.: Magnitude-Prestige Scale , last employment, gen. father: Magnitude-Prestige Scale , occupation of father, gen.

Magnitude-Prestige Scale (MPS) (continued)

	mother: Magnitude-Prestige Scale , occupation of mother, gen.
Category / dataset	socio-economic position / individual-level data
Prepared by	Bernhard Christoph
Explanation :	The MPS (Wegener, 1985, 1988) is the only Germany-specific instrument available to operationalize social prestige based on detailed occupation information. The scale was originally developed for the 1968 version of the International Standard Classification of Occupations (ISCO-68). Because occupation codes in this study were 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 adapted to the 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 datea *(bio_spells)* - *siops* - *isco88* first - *siopseewt* - *iscoeewt* last - *siopslewt* - *iscoeewt* of father - *vsiops* - *visco* of mother - *msiops* - *misco*
Variable label:	aktuelle Ewt.: Standard International Occupational Prestige Scale, current empl., gen. Spell data *(bio_spells)*: Standard International Occupational Prestige Scale, gen. first empl..: Standard International Occupational Prestige Scale, first empl., gen. last empl.: Standard International Occupational Prestige Scale, last empl., 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
Prepared by	Bernhard Christoph

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

Explanation:	The Treiman Prestige Scale, which was originally constructed by Treiman (1977) for ISCO-68, is the first and only prestige scale available for international comparative research on occupations. Since its adaptation to the ISCO-88 (Ganzeboom & Treiman, 1996, 2003), the scale has commonly been called the “Standard International Occupational Prestige Scale”. Infas merged the data as part of the occupational coding procedure.
Literature:	Ganzeboom & 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* - *iscoeewt* of father - *visei* - *visco* of mother - *misei* - *misco*
Variable label:	aktuelle Ewt.: International Socio-Economic Index, current empl., gen. Spell data *(bio_spells)*: International Socio-Economic Index, gen. 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
Prepared by	Bernhard Christoph
Explanation:	The ISEI is among the most common indices of this kind, in part, due to the fact that, unlike most other SEIs, the ISEI is based on an original theoretical concept that considers the occupation and its socio-economic status as an intervening var-iable in the relationship between education and income. The ISEI was developed for the ISCO-68 (Ganzeboom, De Graaf & Treiman, 1992); it 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 & Treiman (1996, 2003)

Classification of Economic Activities 2003 (WZ2003)

Generated:	Employment - Variable name - Source variables
	current - *branche - ET2600* Spell data *(bio_spells) - branche - ET2600* Minijob - *brancheminj - PMJ1300*
Variable label:	Current empl.: Current activity: economic sector/industry (WZ2003) Spell data *(bio_spells)*: economic sector/industry (WZ2003), generated Minijob: Wirtschaftszweig/Branche, current Minijob (WZ 2003)
Category / dataset	socio-economic position / individual-level data
Prepared by	Bernhard Christoph
Explanation :	The information obtained from the open-ended survey question about the sec-tor/industry in which the respondent is employed was coded using the 2-digit Classification of Economic Activities of the Federal Statistical Office (WZ2003) code. At the two-digit level, this classification largely corresponds to the European Nomen-clature générale des Activités économiques dans les Communautés Européennes (NACE) in revision 1.1.
Literature:	StaBA (2002); EG (2002)

Physiological scale of SF12v2 (SOEP-Version, NBS)

Variable name	*pcs*
Variable label	Physiological scale of SF12v2 (SOEP-Version, NBS), generated
Source variables	*PG1200; PG1205; PG1210; PG1215**
Category / dataset	Health / individual-level data
Prepared by	Christian Dickmann
Explanation	The SF12 Questionnaire is an abbreviated version of the SF36 Questionnaire for measuring health-related quality of life. Since 2002 internationally renowned and applied SF12 indicators (version 2 – SF12v2) are used at SOEP. The SOEP-version of the questionnaire, however, differs from the original SF12v2 within formulation, order and layout of the questions. The SF12-indicators of PASS were surveyed analogous to SOEP. The generated pcs variable of PASS is based on the reproduced SPSS-Syntax of Nübling et al. (2006). So far the SF12-indicators were surveyed in waves 3,6, and 9 of PASS.
Literatur:	Nübling et al. (2006); Andersen et al. (2007)

Psychological scale of SF12v2 (SOEP-Version, NBS)

Variable name	*mcs*
Variable label	Physiological scale of SF12v2 (SOEP-Version, NBS), generated
Source variables	*PG1200; PG1205; PG1210; PG1215**
Category / dataset	Health / individual-level data
Prepared by	Christian Dickmann
Explanation	The SF12 Questionnaire is an abbreviated version of the SF36 Questionnaire for measuring health-related quality of life. Since 2002 internationally renowned and applied SF12 indicators (version 2 – SF12v2) are used at SOEP. The SOEP-version of the questionnaire, however, differs from the original SF12v2 within formulation, order and layout of the questions. The SF12-indicators of PASS were surveyed analogous to SOEP. The generated mcs variable of PASS is based on the reproduced SPSS-Syntax of Nübling et al. (2006). So far the SF12-indicators were surveyed in waves 3,6, and 9 of PASS.
Literatur:	Nübling et al. (2006); Andersen et al. (2007)

Leisure activities pursued and desired 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: eisure time activity, desired
Source variables	*PA1100 (für freiz1-freiz3); PA1200 (für 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 newly developed categories for youth leisure activities. This scheme originates in the three most popular (PA1100) and desired (PA1200) leisure activities obtained through open-ended questions. The most popular leisure activities were converted into three individual variables according to the question text. Only one desired leisure activity was considered. Additional responses were not included in the coding. The scheme was developed inductively based on corrected information. To achieve comparability among waves, the new scheme includes all leisure activities that were asked in restricted questions during previous waves. Furthermore, the scheme is designed to allow expansion, if necessary, over subsequent waves with new (sub)categories.

Leisure activities pursued and desired by young people (continued)

	The scheme includes not only 16 main categories but also categories for no leisure activities and information that could not be assigned. The ranking of the 16 main categories results from the frequency with which they were mentioned. The main categories can be differentiated into 77 subcategories. Code - Main Category - 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, television and events 8 7000 Creative hobbies, crafts, cooking and baking 11 8000 Going out, partying, nightlife 3 9000 Hanging out, relaxing - 10000 Shopping - 11000 Traveling, trips, tours and being mobile 3 12000 Spending time with pets - 13000 Volunteer work 4 14000 Learning and education - 15000 Games and mental exercise 2 16000 Side job - 99998 No leisure activity - 99999 Information cannot be assigned -
Literature:	Johanna Eckert, Arne Bethmann, Claudia Wenzig (geplant): Manual coding “Pursued and desired leisure time activities by young people”. PASS wave 5 (2011).

4.5.2 Household or benefit unit level

Equivalised household income, previous 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	Equivalised household income considers the savings achievable through joint housekeeping in multiindividual households compared to single households. The per-capita income of the household is not divided by the actual number of individuals but by a divisor, which is usually less than this figure, and is calculated based on the assumed needs of household members (equivalised household size). According to the previous OECD scale, only the first household member (15 or older) is assigned a weighting factor of 1.0. Household members at least 15 years of age are assigned a weighting factor of 0.7, and children up to age 14 are assigned a weighting factor of 0.5 to calculate equivalised household size.
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, previous OECD weighting (above). The modified OECD equivalence scale assumes a weighting factor of 1.0 only for the first household member (15 or older). Household members at least 15 years old are assigned a weighting factor of 0.5, and children up to age 14 are assigned a weighting factor of 0.3 to calculate 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

Variable name	*depindug2*
Variable label	All waves: deprivation index, unweighted (item total: 23) .
Source variables	*HLS0100a-HLS0400a; HLS0100b-HLS0400b; HLS0600a-HLS1200a; HLS0600b-HLS1200b; HLS1400a-HLS2500a; HLS1400b-HLS2500b;*
Category / dataset	material situation / household-level data
Prepared by	Bernhard Christoph
Explanation	Following Ringen (1988), poverty researchers usually distinguish between direct and indirect measures of poverty. Indirect measurement focuses on the resources available to attain a particular standard of living, especially (equivalised household) income. This method is also called the resource-based approach to measuring poverty. In contrast, direct measurement attempts to record the household's ownership of goods and to determine the extent to which the households cannot afford certain goods or activities that are considered relevant. This method is also called the deprivation approach (see, e.g., Halleröd 1995). Previous scientific research suggests that the population classified as poor by the resource-based approach is not always identical to that identified by the deprivation approach. To define with precision who is to be considered poor, combining measures of resource poverty and deprivation is often been suggested i.e., to classify as poor only those individuals identified by both approaches (see Halleröd 1995; Nolan & Whelan 1996; Andreß & Lipsmeier 2001). The deprivation index is based on a list of 23 goods or activities. The surveyed households are asked to indicate whether they possessed these goods or participated in the activities mentioned. The unweighted index simply adds the number of items that respondents indicated they did not possess or in which they did not participate. However, only items that are missing for financial reasons are counted to prevent consumer preferences (e.g., a household choosing not to own a car or television) from being misinterpreted as a reduced standard of living.

Deprivation index, unweighted (continued)

	Additionally, an item was only accepted as missing for financial reasons if explicitly confirmed in the answers to both questions. “Don’t know” or “details refused” answers were considered available goods or missing for a non-financial reason. This assumption does not apply to all cases. Alternatively, an index value for households that failed to answer a question for (at least) one particular good could be excluded (through listwise deletion). Of the 23 goods and activities surveyed, however, this method would quickly lead to a large number of missing index values. Therefore, the first method described was selected. Nevertheless, compared to the listwise deletion procedure, there is a risk that the number of goods missing for financial reasons is underestimated by this method. For waves 1 through 4, the variable *depindug* provides a version of the un-weighted deprivation index based on 26 items, i.e., adding to the items mentioned above *HLS0500*, HLS1300** and *HLS2600**. These three items have not been asked since wave 5. Thus, depindug2 was newly integrated into the dataset and has been generated retroactively since wave 1.
Literature:	Andreß & Lipsmeier (2001); Halleröd (1995); Nolan & Whelan (1996); Ringen (1988)

Deprivation index, weighted

Variable name	*depindg2*
Variable label	All waves: deprivation index, weighted (item total until W7: 11.08, since W8: 10.59)
Source variables	*HLS0100a-HLS0400a; HLS0100b-HLS0400b;* *HLS0600a-HLS1200a; HLS0600b-HLS1200b;* *HLS1400a-HLS2500a; HLS1400b-HLS2500b;* *PLS0100-PLS0400; PLS0600-PLS1200;* *PLS1400-PLS2500;*
Category / dataset	material situation / householdltsdaten
Prepared by	Bernhard Christoph
Explanation:	For a general description: see deprivation index, unweighted (above).

Unweighted indices, such as the one described above, are often criticised for assigning all items included identical weightings. For example, the difference in asking whether a dwelling has an indoor toilet or whether there is a VCR/DVD player in the household immediately reveals the vast difference in the reduction of household's standard of living caused by the lack of an item. It therefore seems reasonable to weight the items. However, empirical research indicates that in most cases, weighted and unweighted index variants do not yield significantly different results (see Lipsmeier, 1999).

For this survey, we weighted items according to the proportion of respondents who considered a particular item as necessary. We selected this procedure not only because it is conceptually convincing and commonly used (applied by Halleröd 1995, for example) but also because it can be implemented without unreasonable costs. The deprivation weightings determined for the individual questionnaire items are assumed highly stable over time, and these items only need to be administered once or in long intervals. Moreover, the large PASS sample allowed us to split the sample into several randomly selected subsamples, each of which classified only some items. Alternative weighting methods, such as restricting the indices to items that are considered necessary by a minimum proportion of the respondents (e.g., Andreß & Lipsmeier 1995, Andreß et al. 1996) or theoretically restricting the indices to a few fundamental items (e.g., Nolan & Whelan 1996), were not utilised in this survey but can be generated, if necessary, from the data provided. A discussion of the different methods of index weighting can be found in Andreß and Lipsmeier (2001, esp. p. 28 ff.).

For waves 1 through 4, the variable *depindg* provides a version of the weighted deprivation index based on 26 rather than 23 items, i.e., in addition to the items mentioned above, it includes the following items: *HLS0500*; HLS1300** and *HLS2600**; and *PLS0500, PLS1300* and *PLS2600*. These three HLS items have not been asked since wave 5. Thus, depindg2 is newly integrated into the dataset and has been generated retroactively since wave 1.

Deprivation index, weighted (continued)

	The questions about the necessity of the deprivation index were surveyed again in wave 9. The weighting of the deprivation index for waves 1 through 8 bases on the data of wave 1 and since wave 9 on the data of wave 8.
Literature:	Andreß & Lipsmeier (1995, 2001); Andreß et al. (1996); Halleröd (1995); Lipsmeier (1999); Nolan & 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	Category / dataset Household structure / household data
Prepared by	Daniel Gebhardt
Explanation	Various household typologies exist (see, e.g., Lengerer, Bohr & Jansen, 2005 for the Microcensus 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 differentiation criteria are existing partnerships, number and age of children and existing generational relation-ships. Whereas the SOEP typology is based on the relationship of the household members to the head of the household, PASS uses information on the relationships among all household members. The PASS typology includes the ages of household members as indicated in the household interview and household size.

Household typology (continued)

Definition of relationships for generating the household type:

- married couples, registered partnerships, nonmarried partnerships and partner-ships whose status is not specified (missing value for the follow-up question about the type of partnership).
- Child of an individual: biological child, stepchild, adopted/foster child or child whose status is not specified (missing value for the follow-up question about type of relationship to the child).
- Parent of an individual: biological parent, stepparent, adoptive/foster parent or parent whose status is not specified (missing value in follow-up question about type of parenthood).

Definition of household type:

- One-person household: A household consisting of only one individual.
- Couple without children: A household consisting of two individuals living as a couple.
- One-parent household: A household consisting solely of one parent and his/her children. No restrictions apply to children's ages.
- Couple with children under the age of 16: A household consisting of two individuals living as a couple and their respective and/or mutual children. All of the children are younger than 16.
- Couple with children aged 16 or over: A household consisting 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 both under and over 16: A household consisting of two individuals living as a couple and their respective and/or mutual children. Some children living in the household are younger than 16 and others are older than 16.

Household typology (continued)

	▪ Multigeneration household: A household consisting of members of at least three generations in linear succession. The core of the household is multigenerational, i.e., at least one individual in the household is both a child and a parent of another member of the household. Other people living in the household include parents, children, siblings, the central member's partner or a partner's siblings. ▪ Other household: A household that could not be assigned to another household type. ▪ • Generation not possible (missing values): All households with at least one miss-ing value (-1, -2, -4) or implausible value (-8) in the main category of a relationship or age variable (except for households with three or fewer members in unambiguous relationship constellations for which the household type was generated even if ages were missing).
Literature	Beckmann & Trometer (1991); Frick et al. (o.J.); Lengerer et al. (2005); Porst (1984)

Wave 9 benefit unit ID

Variable name	*bgnr9*
Variable label	Benefit unit ID in wave 9 (2015)
Source variables	Household information on age and relationships between household members
Category / dataset	Benefit unit / person register
Prepared by	Gerrit Müller
Explanation	The *bgnr9* variable is created at the individual level. It assigns an identification number to each household member that indicates the individual's relationship to a particular benefit unit. Consequently, household members with the same identification number constitute a benefit unit. The bgnr8 variable is composed of the known household number and a two-digit indicator to identify the benefit unit with-in the household.

Wave 9 benefit unit ID (continued)

The identification of a household member's relationship to a benefit unit is based solely on information about the relationships between household members from the household grid along with the ages obtained from the household interview. Therefore, the benefit units identified in this way are considered synthetic benefit units. The identification process does not consider information about actual benefits received, individual members' ability to work or qualification status, but it does identify groups of individuals in the same household who are or would be considered benefit units in jointly receiving benefits according to the provisions of Book II of the German Social Code in the event that such benefits are needed. This artificial allocation procedure is necessary because information about the existence of a benefit unit and the identification of individuals affiliated with that unit cannot be collected directly in the context of an interview.

The allocation of an individual to a benefit unit is based on the latest version of the German Social Code, Book II, Section 7, Subsection 3 (last amended on 21 March 2013). Each individual ages 25- 65 constitutes a separate benefit unit unless he or she is living in a partnership and/or has a child/children younger than 25 who has/have no partner/children of their own. In the latter case, the benefit unit consists of the individual, his/her partner and child(ren). If two individuals live in the same household with a mutual child but do not indicate that they are living in a partnership, a partnership is nevertheless assumed to exist according to Section 7, Subsection 3a. The corresponding individuals and their child(ren) are assigned to the same benefit unit. Individuals who are between the ages of 15 and 25 are generally assigned to their parents unless they are already living with a partner (or a child of their own) in a joint household. Individuals between the ages of 15 and 25 who live without their parents, partner or children constitute a separate benefit unit.

Wave 9 benefit unit ID (continued)

	Individuals older than 65 are not covered by Book II of the German Social Code and are therefore not considered members of a benefit unit (coded 0) unless they live with a partner who is under 65 (or a child under 25). Likewise, children who have not reached age 15 who live in a household without their parents are not considered members of a benefit unit (code 0) because they are covered by the provisions of German Social Code Book XII. Benefit units were not assigned to households with missing information on relationships or the age of certain house-hold members. Instead, all members of these households were assigned code 99. By approximation, such households are interpreted as households consisting of only one benefit unit.
Literature:	German Social Code Book II – basic security for job-seekers (Sozialgesetzbuch, Zweites Buch - Grundsicherung für Arbeitssuchende (SGB II))

Wave 9 benefit unit typology

Variable name	*bgtyp9*
Variable label	Type of benefit unit in wave 9 (2015)
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 as the synthetic benefit unit used for variable *bgnr8*. Until age 25, children are considered members of their parents' benefit unit unless they themselves have a partner or child. BA statistics typologies are often still established based on reaching legal age (the 18th birthday). For example, according to our typology, households in which the youngest child is between 18 and 24 years old and that are classified as one-parent benefit units are considered single households in BA statistics. This difference must be noted 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 Social Code Book II (see also code 0 for *bgnr9)*. Code 5, generation impossible (missing values), was assigned to households with missing information on relationships or the ages of individual household members (see code 99 for bgnr8).
Literature:	-

Benefit unit receiving Unemployment Benefit II on the wave 9 sampling date probenziehung,

Variable name	*bgbezs9*
Variable label	Benefit unit in receipt of UB II on the sampling date in wave 9 (2015)
Source variables	*HA0250*, HA0300, AL20100, AL20200, AL20300, AL20400, AL20608, AL20708*, HA0400, sample, hnr, bgnr9, hhgr*
Category / dataset	Benefit unit / person register
Prepared by	Mark Trappmann
Explanation	For each benefit unit that was identified according to the procedure described for variable *bgnr9*, this variable indicates whether the benefit unit was actually receiving Unemployment Benefit II on the sampling date of wave 9.
Literature:	-

Benefit unit receiving Unemployment Benefit II on the wave 9 survey date

Variable name	*bgbezs9*
Variable label	Benefit unit in receipt of UB II on the survey date in wave 9 (2015)
Source variables	*AL20608, AL20708*, zensiert (alg2_spells), sample, hhgr, bgnr9*
Category / dataset	Benefit unit / person register
Prepared by	Daniel Gebhardt
Explanation	For each benefit unit that was identified according to the procedure described for variable *bgnr9*, this variable indicates whether the benefit unit was actually receiving Unemployment Benefit II on the wave 9 survey date.
Literature:	-

Number of benefit units within the household

Variable name	*anzbg*
Variable label	Number of synthetic benefit units in the HH, generated
Source variables	*bgnr9, 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 according to the procedure to generate the variable *bgnr9*.
Literature:	-

Number of benefit units in the household 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	*bgbezs9, bgnr9, hnr*
Category / dataset	Benefit unit / household dataset
Prepared by	Daniel Gebhardt
Explanation	This variable indicates the number of benefit units within a household that were receiving benefits according to Social Code Book II on the sampling date. The value was calculated via the household number by aggregating the benefit units within a household that were actually receiving benefits according to variable *bgnr9* from the person register.
Literature:	-

5 Datenaufbereitung

Since wave 3, infas, not the IAB, has been responsible for preparing the data. To guarantee consistent data preparation in the longitudinal section, infas was provided with the relevant syntax files for data preparation from wave 2, necessary sources, intermediary datasets and documentation of individual operations. Important decisions, such as the correction of structural problems in participating households or the development of the *bio_spells* dataset, which was first developed in wave 4, were made with the IAB. The IAB was also available for questions during data preparation.

The information gathered in the wave 9 interviews is available from infas as ASCII data. First, infas prepared the following datasets from the raw data[33]:

- Household dataset for the cross-section, including the spell-reshaped questions for the modules “childcare”, “social participation” and “educational package”
- Household dataset for the longitudinal section (module “Unemployment Benefit II”)
- Dataset updating household composition (matrix)
- Dataset updating family relationships in the household (relationship matrix)
- Individual/senior citizen dataset for the cross-section
- Individual dataset for longitudinal section I (module “employment biography [spells]”)
- Individual dataset for longitudinal section II (module “measures”)
- Dataset for open texts (across household, personal and senior citizen interviews)

Second, a more detailed, formal and content-oriented verification of the data was performed. These data were then prepared as the scientific use file. Furthermore, infas provides a gross dataset along with special datasets that are not derived directly from the actual survey instruments.

The data checks conducted at infas can be divided into three steps, which are detailed in the following sections. First, the household structure of the re-interviewed households was reviewed and when necessary, corrected. If serious problems were identified in the structure, the corresponding interviews were removed (see Chapter 5.1 on this issue). This step was followed by a detailed review of the filter questions (applying corrections if necessary). Filter errors were marked and specific codes were set for missing values (see Chapter 5.2 on this issue). Next, selected items were verified for plausibility. Clearly implausible or contradictory responses were marked by a specific missing code. However, such data

[33] The software packages Stata (versions 11 and 13) and PASW (version 18) were used for data preparation.

corrections were limited.

The following table reviews the steps of the data preparation:

[label=(21.0)]

Table 22: Overview of the steps involved in preparing the data of wave 9 of PASS

No.	Procedure
1	Import the raw data into working datasets
2	Check the household structure (see Chapter 5.1)
3	Remove problematic interviews (household and/or individual levels) (see Chapter 5.1)
4	Integrate individual and senior citizen datasets
5	Correct the household structure of re-interviewed households (see Chapter 5.1)
6	Filter checks at the household level (see Chapter 5.2)
7	Construct a household grid dataset and perform plausibility checks (see Chapter 5.3)
8	Generate synthetic benefit units (see description of variables, Chapter 4.5)
9	Generate new control variables based on the household data after filter checks, household grid dataset and plausibility checks
10	Filter checks at the individual level (see Chapter 5.2)
11	Code information from open-ended survey questions (see Chapter 4.1)
12	Plausibility checks of household and individual-level data (excluding spell data) (see Chapter 5.3)
13	Prepare, plausibility check and construct spell datasets (see Chapters 5.6 to 5.8 and Chapter 5.3)
14	Simple generated variables (see Chapter 4.4)
15	Complex generated variables (see Chapter 4.5)
16	Generation of the data structure for the scientific use file (household, individual and register datasets)
17	Anonymisation (see Chapter 5.5)

5.1 Structure checks and removing interviews

A structure check was conducted before the filter checks. Here, interviews that were not considered successful were to be identified and if necessary, removed from the datasets. In addition, the structure of re-interviewed households was compared with the structure reported during the previous wave to identify and if necessary, to correct implausible or problematic changes in household composition and errors in the allocation of the personal interviews to their respective positions in the household. To observe households in the longitudinal section, it is essential that the individuals be assigned consistently to their position in the household and the respondents can be identified clearly across waves. A personal identification number must not be assigned to different individuals in different waves. If the correct household composition was unclear, all of the interviews conducted with this household in wave 9 were removed from the dataset. If a personal interview was conducted with the wrong individual without further problems in household composition, then

only the personal interview was removed.

Different processes identified problematic cases. The relevant cases were discussed as part of a formal procedure between infas and the IAB. The final decision on how to proceed with these cases was made by the IAB. The following specifies the extent of the checks conducted. Not every check in every wave identifies problems. The result of a check is usually that an issue occurs in few cases. Furthermore, known error sources are absorbed during the interviews. For example, the intention of the survey instrument is that not all known target persons can move out of a panel household at the same time and that at least one remaining individual is at least 15 years old.

- By comparing the first names reported in the current and previous waves, changes in household composition that had not been recorded correctly were identified. Instead of recording moves into and out of a household in the relevant places during the house-hold interview, interviewers sometimes renamed household members or changed their age or sex. All cases in which a first name had been changed that could not be attributed to correcting the spelling and for which the year of birth reported in the previous wave differed by more than one year from that reported in the current wave were reviewed individually. A decision was made as to whether the interviewer made a simple change requiring correction of the first name, age or sex or an inadmissible change to the household structure.
- Furthermore, whether more than one individual with the same date of birth was living in the household was reviewed. Whether these cases were plausible was decided in the context of the household, using two waves. The remaining cases then underwent an-other review. Households in which a date of birth was reported in the current and previous waves by individuals in different positions in the household structure were identified. Here, it seemed reasonable to suspect that a different individual provided the personal interview in the current wave. In the context of the household and individual-level data of the current and previous wave, individual decisions were made for each household and personal interview.
- In general, the date of birth from the personal/senior citizen interview of the current wave displaces all other age information on that individual, e.g., from the household grid, and is the basis for all generated variables utilising age. The date of birth is corrected in *PD0100*. If an individual's year of birth changes significantly according to *PD0100* but the day and month stay the same, the previously known date of birth has never changed according to *PD0100*, and at least two pieces of information about the date of birth from *PD0100* are available from previous waves, then the year of birth is reset to the value from the previous waves considering the whole household. Consider a hypothetical individual whose date of birth is recorded as February 1, 1972 in at least two previous waves and whose date of birth is now recorded as February 1, 1992. This date of birth would make this individual younger than the other children in the household. Without a correction, such an arrangement

leads to an implausible relationship structure, which would consequently mean that synthetic benefit units could not be generated. Hence, in the example above, the date is corrected to February 1, 1972 in the current wave.

- To identify households that are considered not successfully surveyed, the datasets at the household and individual level are merged. Personal interviews without a full household interview and household interviews for which no individual interview was available were marked[34].

- Moves into and out of a household are another important factor. Panel households with reported move-outs were generally inspected and correlated with the split-off households. Evaluations were made as to whether the remaining household of the panel household is plausible. Interviews from panel households in which all household members leave except individual children under 15 years old were discarded for the panel and split-off households. If more than one individual moved, whether these individuals formed a joint split-off or several different households was considered and whether this is plausible was determined. For instance, cases in which one partner left the panel household with young children but the children formed several split-off households were considered implausible. In cases of a non-realised split-off household, move-outs were considered plausible, but all individuals who moved out were remerged into one joint split-off household.

- Individual cases occurred in which the panel household indicates that individuals formed a split-off household, but all members could be identified in the split-off household. Alternatively, not all members of the panel household live in the split-off household, and at least one member of the panel household was not reported as having moved out or moved to a split-off household other than the one observed. 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 who had allegedly moved out was retroactively re-integrated into the household panel.

- In split-off households, individuals who are not known from the panel household but who join PASS through the split-off household might still originate from the panel household. Two situations promote these cases. The first situation arises when a panel household reports several individuals moving out and the split-off individuals formed more than one household. In that case, a dynamic preload is created for the current file for all split-off households identified through the panel household. If, however, individuals who, according to the panel household, live in various split-off households are actually sharing a split-off 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 and are included as new individuals.

[34] New sample households for which a household interview but no valid personal interview was available 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.

- It is possible that individuals from a panel household move out of or into a household that was formed as split-off household during a previous wave and that was successfully surveyed at that time. 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 split-off household emerged was successfully surveyed during the wave of the move, 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-offs from this panel household) as a preload. The few cases in which such a situation might occur do not justify such efforts in the field. Instead, these cases must be found during the structure checks. Note that in this context, split-off households must be considered in the waves following their first successful survey even if they are considered panel households in field control. In both cases, the personal identification numbers *pnr* of the individuals in the split-off household are corrected retrospectively. It must also be considered that these individuals are treated as new respondents in the personal/senior citizen interview although they might have already participated in an interview. This deviation is generally not corrected (see also Chapter 4.4).

- In panel households that reported a move-out as of wave 2, a return to the household can also occur as of wave 3. Recognising these individuals as moving back in and assigning them their former household position instead of a new household position is a function of the household grid. Whether these requirements were met in the field in all cases was also evaluated. For individuals who were identified in the current wave as moving back in by comparing the first name, age and sex with the members who previously moved out of the household, the household structure must be changed. These changes led to retroactive changes of the personal identification number of the individual and the individual information in the household interview - e.g., information about childcare or the reasons for a cut in Unemployment Benefit II - to the correct position within the structural check. Whether an individual who is marked in the field as moving back in is the same individual who moved out during a previous wave was also verified. If not, this change represents an individual who is new to PASS. Changes to the household structure are also made in this case.

- In case of moves back into a household, whether the split-off household in which the individual lived was successfully surveyed during the current wave and whether the split-off household reported that the individual moved out were verified. In addition, the status of individuals who moved back into their panel household during a previous wave must continue to be verified with the split-off household provided 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 as to whether this was plausible or whether either household structure should be corrected.

- Returns are not the only cases of individuals being considered current household

members of several households. This situation can also occur when a member of a split-off household is not recorded as having moved out of the panel household. Individual cases can be acknowledged as plausible after examination of both household structures. These cases are documented in the {zdub* variables in the person register. For further explanation, please refer to Chapters 4.4 and 5.4.1.2 of the data report for Wave 5 of PASS (Berg et. al., 2012).

- Other issues concerning the relationship of a panel household and its split-off households can also arise. Individuals who joined PASS via a split-off household might move to the panel household. Another possibility is that individuals move from one split-off household to another. Generally, all individuals in a panel household and all of its split-off households must be considered a network. The structure checks are designed so that individual moves among the households of such a network are detected regardless of the direction in which an individual moves.

- Household structure verification generally evaluates the changes between waves, not the plausibility of the structure. Therefore, the household structure first-time interviews can only be verified to a limited extent. For first-time households, information concerning first name, age and sex is reviewed to determine whether individual household members are listed multiple times. In this case, only the initially reported household position is maintained. This situation 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 in positions 2 and 3 are identical, individual 3 is removed and individual 4 is retroactively moved to position 3. As a rule, in a household interviewed for the first time with X household members, positions 1 to X are to be filled without gaps. Someone retroactively recognised as moving back through a subsequent change in his or her personal identification number also makes it necessary to move the individual information in the household interview.

- Thanks to feedback provided by a field interviewer, a household that was included twice in the panel sample during wave 4 was detected. Household 10015439 had been included in the sample as the identical household 15044862 since wave 1. Both households were successfully surveyed during waves 1 and 3 and not surveyed during wave 2. In wave 4, household 10015439 was successfully surveyed. This duplicate was detected because “both” households were assigned to the CAPI interviewer for that point. The household composition 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 because to do so would affect weighting. The duplicate household is coded 26 in the *hnettod4* variable in *hh_register*, which identifies the reason for non-surveying. All household members of the duplicate household are coded 56 in the *pnettod4* variable in *p_register*.

- Individual decisions were also made to address cases that proved to be problematic during the structure checks. Here, the seriousness of the particular problem was significant. In cases in which the correct household composition in wave 9 was unclear, all of the interviews from wave 9 were removed. In wave 9, these households will be

treated as households that did not participate in wave 9. If in retroactively removed household interviews moves-out were reported, the split-off households were discarded. This removal affected both the interviews conducted in the current wave in these split-off households and 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 contribute to the panel sample of the subsequent wave. If there was merely a problem in assigning individuals to their respective positions in the household, i.e., if it was suspected that a personal interview had been conducted with the wrong individual in wave 9, then only that personal or senior citizen interview was removed. Structural problems with no serious consequences that could be solved, for example, by removing a personal interview, first name, age and sex were made at the household level. The incorrect information concerned was replaced with the last valid value from the previous wave or the value from the previous wave added to the number of years since the last valid interview.

In addition, all interviews with individuals for households with no complete household interview 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. Households from the refreshment sample that were 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 *(hnettok9, hnettod9, pnettok9, pnettod9)* in the household and person register datasets indicate removed interviews. Through the corresponding variables in the household register, it is possible to trace the re-interviewed households whose household interviews were later removed. Net variables in the person register allow for tracing the cases in which only single individual-level interviews or all of the interviews in the household were deleted. In the case of households from the refreshment sample of wave 9 without at least one valid household and personal interview, it is not possible to trace deleted interviews in the register datasets because 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 instruments was verified using a statistical program. If certain questions were asked when the value of the relevant filter variable would have required something else (for example, if detailed information was requested about vocational training although the respondent had stated that he/she did not have any vocational qualification), these variables were set to missing code “-3” (not applicable), which they would also have received through correct use of

the filters[35]. Moreover, some items were not asked in individual cases when those questions would have been necessary according to the filter (e.g., if no further information was recorded about vocational training although the respondent had stated that he/she had under-gone such training). In these cases, the missing code “-4” (question mistakenly not asked) was assigned. An assignment of code“-4” can also be based on the household structure evaluation described in Chapter 5.1. If an individual’s move-out is retroactively discarded as implausible and the individual is retroactively classified as belonging to his or her former household, then individual information about 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 code “-4” is assigned to a question that is relevant for filtering subsequent questions, then the subsequent questions are also coded “-4” in case these subsequent questions are not asked. If these questions were asked because, for instance, several filter questions linked to this subsequent question and another filter question triggered the question correctly, the value recorded there remains.

In an additional step, the missing codes assigned by the field institute and system missing codes were replaced by standard values for all variables. Table 23 provides an overview of the assigned values. Codes “-1” and “-2” are the standard “don’t know” and “details refused” answers recorded during the survey, respectively. Code “-3” is the general “not applicable” code for questions not asked due to filters. As described above, code “-4” was as-signed if a question was not asked because of a filter error. Codes “-5” through “-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 of greater than 99,999 in the open question on income). These codes were only assigned as required.

[35] As is customary in such cases, the filter checks were conducted beginning with the items that were asked first.

Table 23: 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 have been asked)
-5	question-specific code number 1, only assigned as required
-6	question-specific code number 2, only assigned as required
-7	question-specific code number 3, only assigned as required
-8	“implausible value”
-9	“item not surveyed in wave”
-10	“item not surveyed in questionnaire version” [36]

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 an item was not asked during a specific wave.

Because the dataset is prepared in long format, as was described above, variables that were no longer asked in any version of the questionnaire as of wave 2 are coded “-9” for the observations in this wave. Variables included 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 questionnaires, that is, between the personal questionnaire and senior citizen questionnaire or between 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. The checks conducted during the previous waves were adapted and extended for the current wave. Furthermore, the household structure and spell data were checked for plausibility - especially for inadmissible overlaps within the individual spell types. Generally, only the data gathered in the cross-section of wave 9 were verified. No checks were conducted in the longitudinal section, that is, to compare the information provided in the current wave with that provided in the previous wave.

In detail, the following steps were conducted:

- Contradiction check: In general, contradictions were only corrected either if the implausibility could be defined as particularly serious and/or if the alteration was considered minor. The latter applied, for example, if only a small number of cases

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

were affected or if one missing code (e.g., “-3”) was replaced by another (e.g., “-8”). Two strategies were used to filter implausible statements. Either the implausible responses were corrected directly, or they were assigned a specific missing code.

- Implausible responses were only corrected if it was highly probable that the interviewer had entered information incorrectly: for example, if the interviewer entered 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 certainty. In most cases, it was only possible to establish a contradiction between two statements but not to identify specific incorrect entries that had led to the implausible statement. Therefore, in these cases, no corrections were made, and the specific missing value code “-8” was assigned instead. It was decided on an individual basis whether the code was assigned to one of the two variables involved in the contradiction or to both of them.

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

- To identify implausible household structures, the information on relationships was first combined with the demographic information for individual household members. For the households that were identified as implausible during these checks, individual decisions were made considering 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 corrected based on additional information on the household context if it was highly probable that an error had occurred. For example, in the case of two people of the same sex who were both biological parents of a third member of the household, the sex was corrected based on the first name. If the first names also indicated 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 conducted comparing sets of three family relationships for plausibility. The following provides an example of a relationship structure that would be classified as implausible: individual A is individual B’s spouse. Individual A is the biological parent of individual C. Individual C is a sibling of individual B. If such a combination or similarly implausible combination of relationships was identified, an attempt was made to make the relationship plausible based on the household

context. In the case described, the relationship data were corrected by coding individual C as a child of individual B, whose status was not specified. The aim was to correct as many of the implausible entries as possible because a plausible and complete set of relationships is necessary to generate the benefit unit.

- In addition, the spell datasets were subjected to a number of plausibility checks, as detailed in Chapters 5.6 through 5.8.

5.4 Retroactive changes in waves 1 to 8

During the data preparation process for the scientific use file for wave 9, some changes were also made to the waves that had already been delivered. These changes included corrections of errors that were detected after the completion of the scientific use file of wave 8. The corrected data can now be used in the SUF datasets of the current wave, wave 9. Tables 24 through 28 provide an overview of the retroactive changes to the delivered waves of PASS[37].

[37] Adjustments to value or variable labels are only considered here if this changes the interpretation of variables or values.

Table 24: Overview of retroactive changes to the household dataset *(HHENDDAT, KINDER)*

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
*HD0200**	*HHENDDAT*	6-8	Correction	Persons, for whom no exact age was specified, except the information that they are under 15 years, mistakenly got assigned the value 0 in variable HD0200* instead oft he correct values -1 or -2. One case each was affected in waves 6 and 8 and three cases in wave 7.
alter *HKI0205* *HKI0250* *HKI0260** *HT0100* *HT0200* *HT0300* *HT05** *HBT03** *HBT04** *HBT05** *HBT08**	*KINDER*	6-8	Correction	Persons, for whom no exact age was specified except the information that they are under 15 years, mistakenly got assigned the value 0 in variable alter instead oft he correct values -1 or -2. Additionally some of the variables of these cases were recoded from -4 to -3. One case each was affected in waves 6 and 8 and three cases in wave 7.

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

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
PEO1000 PEO1100**	*PENDDAT*	5	Correction	The variable labels were misleading as in PEO1000* it refered to an employment of under 30 working hours instead of an employment under 15 working hours and for PEO1000* the labels mistakenly refered to an employment over 15 working hours instead of an employment of under 30 working hours.
azhpt1 azges1	*PENDDAT*	3-8	Correction	116 cases, namely 22 cases of wave 3, 11 cases of wave 4, 18 cases of wave 5, 22 cases of wave 6, 23 cases of wave 7 and 20 cases of wave W8 were recoded to -3. If the main employment in the previous wave was not equivalent to a self-employment and continues in the current wave, but changed the occupational status in the current wave to Self-Employed/Helping family member, Not specified or Don't know, then both variables including the contractual working hours were mistakingly filled with information instead of assigning the correct value -3.

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

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
kldbeewt	*PENDDAT*	5-8	Correction	11 cases, namely 1 case each in waves 5 and 6, 3 cases in wave 7 and 6 cases in wave 8 were coded disaccordingly to the provided KldB-coding scheme.
iscoeewt *mpseewt* *siopseewt* *iseieewt* *egpeewt* *eseceewt* *kldbeewt*	*PENDDAT*	4-8	Correction	Information about the participants first employment are retrieved from the *bio_spells* or the informations about his/her last employment if participant states at the beginning of the employment questions set, that he/she has already given information about his/her first employment. Information are retrieved from the *bio_spells* if participant was employed since the reference date. If the participant was unemployed since the reference date, information are retrieved from the statements of last employment. However, in the affected waves this procedure was only implemented for stibeewt and not for the here mentioned variables. The variables are affected as follows (number of cases) : *iscoeewt* 1979 x, *mpseewt*, *siopseewt* and *iseieewt* each 1946 x, *egpeewt* 1931 x, *eseceewt* 1935 x and *kldbeewt* 1980 x. So far all affected cases were included with negative codes, which indicated, that no in-formation was available for these cases. Within the correction procedure these negative codes were replaced with contentrelated information. The high number of cases affected occurs, because the information was only collected in the first survey and then updated in the following waves. These updates were corrected as well.

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

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
PET1300 PET1450	*PENDDAT*	3	Correction	PET1300 and PET1450 contained information for 5 participants about a second employment spell which was not the main employment and contains information about irregular working time.

Table 26: Overview of retroactive corrections to spell datasets *(bio_spells, alg2_-spells, ee_spells)*

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
az1	*bio_spells*	3-8	Correction	So far, az1 included the last surveyed amount of working hours from variables ET2000-ET2007 (of former dependent employment) for all ET-spells, which originally derived from a dependant employment and then got updated to Employed/Helping family member, Not specified or Don't know according to variables ET0600-ET0607. Instead the latest value from ET2000-ET2007 should be used, which equals -3 for all affected 35 spells, as these cases did not get the question.

Table 26: Overview of retroactive corrections to spell datasets *(bio_spells, alg2_spells, ee_spells)* (continued)

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
*kuerz** *alg2kbm** *alg2kbj** *alg2kem** *alg2kej**	*alg2_spells*	2-8	Correction	So far there existed no checks whether the ending date is before the beginning date. This check had been made retroactive. For eight spells the generated variables begin and end were coded to -8. In wave 2 and 7 two cases were affected and in wave 4, 5, 6, and 8 one case was affected.

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

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
hhgr3 *nweg3* *nneu4* *nneu5*	*hh_register*	3-5	Correction	Persons who leave their household less than a year, are usually still counted as members of their household in the *p_register*. In wave 3 the affected persons were counted as having left the household. This was the case for three persons, whose households had to be corrected for *p_register hhgr3* and *nweg3*. This also accounted for one case each in *nneu4* and *nneu5*.
hnr3 *zplfd3* *alter3* *bgnr3* *bgtyp3* *bgbezs3* *bgbezb3* *pnettok3* *pnettod3* *wegm3* *wegj3* *zmhh3* *zvhh3* *zparthh3* *zdub3* *kinddat3*	*p_register*	3	Correction	Persons who leave their household less than a year, are usually still counted as members of their household in the *p_register*. In wave 3 the affected persons were counted as having left the household. This was the case for three persons. All variables related to wave 3 except *neum3* and *neuj3* were newly generated for these cases. In one household *bgtyp3*, *zvhh3* and *zparthh3* also had to be regenerated for the rest of the persons belonging to the household.

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

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
*pnetto*6* *weg*6* *neu*6* *pnetto*7* *weg*7* *neu*7*	*p_register*	6-7	Correction	Persons, who moved to a Split-Household according to the information about their previous household, but did not arrive there according to the information of the other members of the Split-Household, receive the value “adress unknown“ (*pnettod**=41). One person in wave 6 and three persons in wave 7 mistakenly received the value *pnettod**=60. These persons were falsely marked as not belonging to the sample anymore. According to that procedure also the *weg**- and the *neu**-variables have to be corrected.
*pnetto*8*	*p_register*	8	Correction	Two persons in wave 8 were marked as not belonging to the sample anymore. A few waves ago they moved to a split-household and later to another split household. The current split-household was not included as the former split-household was excluded from the sample. The persons were falsely assigned to the former split-household instead of the current split-household. *pnettok8* was changed from code 6 to code 5 and pnettod8 was corrected accordingly.

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

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
alter6 *alter7* *alter8*	*p_register*	6-8	Correction	Persons, for whom no exact age was specified, except the information that they are under 15 years, mistakenly got assigned the value 0 in variable *alter** instead of the correct values -1 or -2. One case each was affected in waves 6 and 8 and three cases in wave 7.
hnettod6 *hnettod7* *hnettod8*	*hh_register*	6-8	Correction	For a few cases refusals on the household level were coded with 20 “HH not reached“ instead of the correct code 21 “HH refusal“. The correction affected 7 households in wave 6, 8 households in wave 7 and two households in wave 8.
pnettod6 *pnettod7* *pnettod8*	*p_register*	6-8	Correction	In a few cases refusals on the household level were coded with 50 “HH not reached“ instead of the correct code 51 “HH refusal“. The correction affected 9 persons in wave 6, 18 persons in wave 6 and two persons in wave 8.

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

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

5.5 Anonymisation

All data obtained by the IAB, a special department of the Federal Employment Agency (BA), are social data, which places high demands on data protection. It was therefore nec-

essary to include some of the variables in the scientific use file in simplified form. These variables are generally labeled with the flag “anonymised” in the variable label. For the same reason, it was also necessary to exclude available regional information, excluding the German states and information about East/West Germany. To protect the data, neither 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, the household and benefit unit type (*hhtyp*[38], *bgtyp*[39]), indicator variables on partners in the household (*apartner; epartner*[40]), indicator variables pointing to parents, partners in the household (*zmhh; zvhh; zparthh*[41]) and various indicator variables for parents (*mhh; vhh*[42]) or children of the target person (e.g. *ekind*[43]) living in the household are provided. Table 29 provides an overview of the variables concerned and the process of anonymisation[44] in each dataset. Table 30 provides the anonymised variables for the employment spell dataset.

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

Varname	Variable label	Procedure
PD0100	Year of birth (date of birth, anon.)	The precise date of birth was shortened to 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 (anon.)	For technical reasons, professional and regular soldiers were recorded separately. Due to the few case numbers and because this group is not usually asked about occupational status, this group was merged with civil servants and judges.
PET1250	Last occup. status civil servant: detailed info., incl. soldiers (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.

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

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

Varname	Variable label	Procedure
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 were assigned to the corresponding civil servant category.
PET1510	Current occup. status, simple classification, surv. as of wave 2 (anon.)	Procedure as for *PET1210*.
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*.

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

Varname	Variable label	Procedure
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*.
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*.

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

Varname	Variable label	Procedure
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*.
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 individual dataset *(bio_spells)* in wave 9

Varname	Variable label	Procedure
ET0607	Wave 9, Occup. status, simple classification (anon.)	Procedure as for *PET1210*.
ET1007	Wave 9, Occ. status: civil servant/ judge/soldier, detailed information (anon.)	Procedure as for *PET1250*. The variable for professional and and regular soldiers is not supplied.
stib	Occ. status, code number, gen.	Procedure as for *stiblewt*.

5.6 Receipt of Unemployment Benefit II

UB II is recorded at the household level in spell form in waves 1 to 8. This concept was continued in wave 9 but with a slightly revised set of questions.

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

To update spells for which UB II was ongoing during the previous wave and therefore were 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 that were interviewed for the first time in wave 9 were asked about their receipt of UB II during the period since the last change in the household composition. If this change was before January 2013 or if no information was provided about changes in the household, then the household's receipt of UB II from January 2013 on was recorded.

5.6.2 Structure of the Unemployment Benefit II spell dataset

The structure and contents of the spell dataset on UB II change due to the integration of the spells of UB II reported in wave 9. Here, it is necessary to distinguish among (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 asked in wave 9.

1. Additionally, in wave 9, new wave-specific, cross-sectional variables were included in the UB II spell dataset. These variables include *AL20608, AL20708a* to *AL20708o, AL20808* and *AL 20908*. These variables refer to the interview date in wave 9. Cross-sectional variables also ex-ist for the interview dates of the previous waves that contain the analogous information referring to the respective wave. Table 31 provides an overview of the cross-sectional information contained in the UB II spell dataset.

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

	Wave 1	Wave 2	Wave 3	...	Wave 9
Does the HH receive UB II for all HH members?	*AL20600*	*AL20601*	*AL20602*	...	*AL20608*
Does the HH receive UB II for individuals 1 to 15?	*AL20700a-AL20700o*	*AL20701a-AL20701o*	*AL20702a-AL20702o*	...	*AL20708a-AL20708o*
Amount of monthly UB II receipt?	*AL20800*	*AL20801*	*AL20802*	...	*AL20808*
Has a cut of UB II begun?	*AL20900*	*AL20901*	*AL20902*	...	*AL20908*

2. Not available in wave 9 compared to wave 8.

3. Not available in wave 9 compared to wave 8.

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

As in waves 1 to 8, the information on UB II was also subjected to a number of plausibility checks in wave 9. Inadmissible overlaps and dates of spells of UB II or benefit cuts were corrected when necessary. In principle, changes were only made to the generated date variables (*bmonat; bjahr; emonat; ejahr*) of the spell of UB II receipt, the spells of benefit cuts (*alg2kbm*; alg2kbj*; alg2kem*; alg2kej**) *) and the censoring indicator of the spell of UB II receipt (*zensiert*). If it was not possible to remove implausible data by correcting the dates, then in a small number of cases, spells of UB II receipt or cuts were merged or deleted.

5.6.4 Updating the Unemployment Benefit II spell dataset

After the spells of Unemployment Benefit II reported in wave 9 had been converted into spell format, and after inadmissible overlaps and implausible dates were corrected following the plausibility checks and corrections, the spells of UB II that were ongoing at the time of the interview in the previous wave were updated using the information gathered in wave 9. 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 that was censored during the previous wave using information gathered in wave 9. Here, the censoring indicator is integrated into the spell of receiving UB II, which was ongoing during the previous

wave, as are the generated and recorded end dates, wave-specific cross-sectional information (see above) and new spells of benefit cuts. In addition to updating spells that were censored during the previous wave, new spells that were reported in wave 9 are merged with the spell dataset (3). These three variants are outlined briefly below:

1. *Cases in which the household in wave 9 contradicts an ongoing spell of receiving UB II at the interview date in the previous wave.*
   If the household contradicted an ongoing spell of receiving UB II 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 correct. Because it is not possible to make 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 during 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*) along with the question of whether a spell continues (*AL20500*)remain unchanged[45]. The generated end date of the UB II spell (*emonat; ejahr*)) had been 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 ongoing in the previous wave.*
   If information about the end date of a spell of UB II receipt that was censored in the previous wave is available in wave 9, then the spell that was censored in the previous wave was updated using the current information. First, the recorded end date (*AL20300; AL20400*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the receipt of UB II is 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 9 and the cross-sectional data referring to wave 9 (*AL20608; AL20708a to AL20708o, AL20808, AL20908*) were included.

3. *Spells of UB II receipt reported for the first time during wave 9 that do not update any spells that were censored in the previous wave.*
   Spells reported for the first time during wave 9 were added to the UB II spell dataset. Next, the spell counter was generated new to create a variable *spellnr* without gaps.

[45] 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 Employment biographies

Employment, unemployment and gap periods at the individual level were recorded in spell form in waves 2 and 3. This concept of a modular spell survey was changed to an integrated survey of the employment biography in wave 4. For individuals who were asked for their employment biography for the first time in wave 9, the reference date for the start of the retrospective interval was adjusted. In wave 9, all spells of employment and unemployment since January 2013 were to be reported here. Individuals who were interviewed about their employment biography during 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 ongoing in the previous wave

Continuing ET, AL and gap spells were updated in wave 9. To update the spells that were ongoing during the previous wave and were therefore right-censored in the spell dataset, dependent interviewing questions are included in the personal questionnaires.

5.7.2 Structure of the BIO spell dataset

With respect to 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 names in the BIO spell dataset compared to the ET SUF of wave 3, analogous to the 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 type of activity (*spelltyp*), spell integration (*spintegr*) and comprehensive spell number (*spellnr*) are available.

Due to the integration of the employment and unemployment spells reported in wave 9 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. 1. The ET-specific variables in the BIO spell dataset *ET0600* to *ET2200* are considered wave-specific, cross-section information that refer to wave 2; variables *ET0601* to *ET2201* refer to wave 3, *ET0552* to *ET2202* refer to wave 4, *ET0553* to *ET2203* refer to wave 5, *ET0554* to *ET2204* refer to wave 6, *ET0555* to *ET2205* refer to wave 7, *ET0556* to *ET2206* refer to wave 8 and *ET0557* to *ET2207* are cross-section information that refers to wave 9. 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*)

	Wave 1	Wave 2	Wave 3	...	Wave 9	
Occupational status			*ET0552*	*ET0553*	...	*ET0557*
(simple and detailed	*ET0600*	*ET0601*	*ET0602*	*ET0603*		*ET0607*
classification)	*ET0700*	*ET0701*	*ET0702*	*ET0703*		*ET0707*
	ET0800	*ET0801*	*ET0802*	*ET0803*		*ET0807*
	ET1000	*ET1001*	*ET1002*	*ET1003*		*ET1007*
	ET1100	*ET1101*	*ET1102*	*ET1103*		*ET1107*
	ET1200	*ET1201*	*ET1202*	*ET1203*		*ET1207*
Supervisory function;	*ET1300*	*ET1301*	*ET1302*	*ET1303*	...	*ET1307*
number of employees	*ET1400*	*ET1401*	*ET1402*	*ET1403*		*ET1407*
supervised						
Cancellation of limi-	*ET1700*	*ET1701*	*ET1702*	*ET1703*	...	*ET1707*
tation of an initially				*ET1753a*		*ET1757a*
limited employment				*ET1753b*		*ET1757b*
Working hours			*ET1952*	*ET1953*	...	*ET1957*
(contracted; actual;	*ET2000*	*ET2001*	*ET2002*	*ET2003*		*ET2007*
average for irregular	*ET2100*	*ET2101*	*ET2102*	*ET2103*		*ET2107*
working hours)	*ET2200*	*ET2201*	*ET2202*	*ET2203*		*ET2207*
Income for current				*ET2800-*	...	*ET2804-*
ongoing spells				*ET3900*		*ET3904*
Overtime						*ET4100*
						ET4200

The BIO spell dataset also includes an AL-specific variable which is understood as wave-specific cross-sectional information (*AL1300* for wave 2; *AL1301* for wave 3, *AL1302* for wave 4, *AL1303* for wave 5, *AL1304* for wave 6, *AL1305* for wave 7, *AL1306* for wave 8 and *AL1307* for wave 9).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*)

	Wave 1	Wave 2	Wave 3	...	Wave 9	
Amount of monthly UB I receipt?	*AL1300*	*AL1301*	*AL1302*	*AL1303*	...	*AL1307*

2. The wave-specific Variables that were surveyed for the first time in wave 9 *ET4000, ET4010a, ET4010b, ET4010c, ET4010d, ET4010e, ET4020, ET4021, ET4030a, ET4030b, ET4040, ET4050, ET4060, ET4070, ET4080, ET4090, ET4300* and *AL1400* were integrated in the 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. As in wave 4, checks were made only within one spell type. Cross-spell type checks were not conducted. As with the spell data on receiving UB II, correction and recoding were only conducted for the generated date variables. Here, details on seasons were recoded into months, “-8” values were set for implausible responses and date information was replaced or rendered plausible. Because 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, in some cases it was necessary to delete entire spells. For example, spells that were obviously recorded twice were removed. Spells that are completely outside the survey period but for which data were nonetheless collected were also deleted.

5.7.4 Update of spell datasets

After the spells reported in wave 9 had been converted into spell format, plausibility checks and corrections for inadmissible overlaps and spells with implausible dates were corrected. The spells that were ongoing at the time of the previous interview wave were updated using the information recorded in wave 9.

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 that was censored in the previous wave using information gathered in wave 9 in the narrow sense. Here, the

censoring indicator is integrated into the spell that was ongoing during the previous wave, as are the generated and recorded end dates and wave-specific cross-sectional information (see above).

In addition to updating spells that were censored during the previous wave, new spells reported in wave 9 are merged with the spell dataset (3). These three variants are outlined briefly below:

1. *Cases in which the individual in wave 9 contradicts an ongoing spell on 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 correct. Because 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 during 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*), along with the question of whether a spell continues (*BIO0600*) remain unchanged[46]. 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 ongoing in the previous wave.*
   If information about the end date of a spell that was censored during the previous wave is available in wave 9, then the spell that was censored 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 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 9. Furthermore, the cross-sectional data referring to wave 9 (*ET0557* to *ET2207*) were included.

   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 ongoing (*BIO0600*), the reason for the end of unemployment (*AL0600, AL0601*) and the censoring indicator (*zensiert*) were overwritten with the information gathered in wave 9. Furthermore,

[46] Thus, the reported end date remains completed with the interview date of the wave in which the spell was censored or the special code "0" for continuing spells. In addition, the question about 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 in-formation, which here is the interview date for the wave in which the spell was censored.

the cross-sectional data referring to wave 9 (*AL1307*) were included. AL spell data, moreover, feature the exception that the spell of UB I (receipt of UB I) is recorded within an AL spell. Which information is updated depends on whether UB I was already received during this spell of unemployment and whether this benefit was ongoing during the previous wave.

If, in the previous wave, there was also an ongoing receipt of UB I in the AL spell to be updated, then the recorded 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 obtained in wave 9.

If no UB I was received in previous waves in the AL spell to be updated, then the information on UB I receipt was overwritten with the information obtained in wave 9. In addition to 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 but that ended in the previous wave, no changes were made to these spells.

3. *3. Spells reported for the first time in wave 9 that do not update any spells that were censored in the previous wave.*
   Spells reported for the first time in wave 9 were added to the BIO spell dataset. Next, the spell counter was generated anew 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 7 SUF (*spellnret, spellnral, spellnrlu, spellnr*) maintain their spell number. The new spells from wave 9 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 in wave 9. The reference date as of which to consider one-Euro jobs was January 2014 for wave 9.

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

Continuing *ee_spells* were updated in wave 9. To update the spells that were 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 9 in the OEJ spell dataset (*ee_spells*), new variables are added that refer to a specific wave. Table 34 gives an over-view of the cross-sectional information contained in the EE spell dataset.

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

	Wave 4	Wave 4	...	Wave 9
Weekly working hours in the OEJ	*EE1100*	*EE1101*	...	*EE1105*
OEJ is the same work permanent co-workers do	*EE1200*	*EE1201*	...	*EE1205*
Which kind of training necessary for OEJ	*EE1300*	*EE1301*	...	*EE1305*
Only work or also training/ classes?	*EE1400*	*EE1401*	...	*EE1405*
Assessment OEJ	*EE1500a-EE1500h*	*EE1501a-EE1501h*	...	*EE1505a-EE1505h*

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

5.8.3 Plausibility checks and corrections in the EEJ spell dataset

The OEJ spell dataset on the participation in OEJ was both checked for plausibility and corrected. The plausibility checks contained checks for dates, for the reference date for the newly integrated spells in wave 9 (January 2014) and for logical inconsistencies in cases of respondents with several OEJ spells.

Only the generated date variables (*bmonat, bjahr, emonat, ejahr*) were corrected and recoded. Details on seasons were recoded into months, “-8” values were assigned for im-

plausible responses and date information was replaced or rendered plausible. Next, a spell counter *spellnr* was generated. The variable 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 position within the survey wave. Spells from wave 8 maintained their spell number for the wave 9 SUF.

6 Weighting Welle 9

The weighting concept for wave 9 generally follows the concepts developed in previous waves (see Berg et al., 2015). The starting point for the wave 9 weighting procedure and for the longitudinal section from wave 8 to wave 8 were the cross-sectional weights from wave 8 for households and individuals. The two weights for each household and two weights for each individual were updated. This chapter of the data report documents the technical details and exact models used to generate the weights for wave 9. An overview of the weighting concept used in PASS can be found in chapter 8 (Trappmann, 2013a) of the PASS User Guide (Bethmann, Fuchs, and Wurdack, 2013). Examples of how to use the weights can be found in Chapter 12 (Trappmann, 2013b).

6.1 Design weights for the panel households in wave 9

New "household design weights" were generated for wave 9 from the cross-sectional weights for households of wave 8, taking into account people moving into households from within Germany. This step was performed by using the weight share procedure as described in wave 2 (see Gebhardt et al., 06/2009). Births, deaths or move-outs from households have no influence on weight; moves into households from within Germany, however, increase the inclusion probability of a household because the individuals who moved into the household also had the opportunity to be included in the sample in waves 1 to 8. The new design weight for subsample i *dwihh9* is therefore calculated from the old cross-sectional weight *wqihh8*:

$$1/dw_ihh_9 = 1/wq_ihh_8 + (n_{samplei}/n_{populationi})$$

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

6.2 Design weights for the refreshment sample in wave 9

In wave 9 the panel was refreshed by sampling new households from new inflows to benefit receipt. All households that were receiving benefits in July 2014 but had had no probability of being selected for the register data sample in the same month in 2013, 2012, 2011, 2010, 2009, 2008, 2007 and 2006 had a likelihood of being selected. This refreshment could be achieved by selecting only benefit units in which no member was receiving benefits in July of the previous years. The refreshment sample was drawn from the 300 points of the first wave and the 100 replenishment points of wave 5. Analogous to 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* = 12, the design weight of the refreshment sample is included in the variable *dw_ba*.

6.3 Propensity to participate again - households

In this step, again similar to the procedure in wave 8, the probability of re-participation in wave 9 was estimated for each household that participated in wave 8 based on logit models for willingness to participate in the panel, availability and participation. Additionally, households that participated in wave 7 but not in wave 8 (temporary nonresponses) were considered in the modeling for wave 9. In addition to variables from the household and personal interviews with the head of the household conducted during the previous wave, other fieldwork variables were included, 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 [t1, t2, t3, t4, t5, t6, t7, t8, t9] across all nine waves can be obtained as the product of the cross-sectional weight to t1, *hpbleib* (wave 1 to wave 2) and hpbleib (wave 2 to wave 3, etc.) (see also the PASS User Guide section 12 (Trappmann, 2013b)).

Table 35: Variable overview, codes and reference categories for logit models of 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 nationality other than German
Reference category	HRP 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/pupil

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

Variable code and reference category	Explanation
gesundheit_1	Subjective evaluation of the health state of the HRP: very good
gesundheit_2	Subjective evaluation of the health state of the HRP: good
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: satisfactory
zufrieden_1	General life satisfaction HRP: scale value 0-2
zufrieden_2	General life satisfaction HRP: scale value 3-5
zufrieden_4	General life satisfaction HRP: scale value 9-1
Reference category	General life satisfaction HRP: scale value 6-8
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_15_64	Number of individuals in the household aged 65 years and older
anz_65	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 category	UB II receipt of the household: no current receipt of UB II

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

Variable code and reference category	Explanation
stichprobe1	BA sample
stichprobe3	Refreshment sample (BA) wave 2
stichprobe4	Refreshment sample (BA) wave 3
stichprobe5	Refreshment sample (BA) wave 4
stichprobe6	Replenishment sample (EWO) wave 5
stichprobe7	Replenishment sample (BA) wave 5
stichprobe8	Refreshment sample (BA) wave 5
stichprobe9	Refreshment sample (BA) wave 6
stichprobe10	Refreshment sample (BA) wave 7
stichprobe11	Refreshment sample (BA) wave 8
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
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

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

Variable code and reference category	Explanation
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 Contact		Participation Participation	
	Coef.	p	Coef	p	Coef.	p
alter_1	-.3866829	0.120	-1.370016	0.000	-.7867117	0.000
alter_2	.3145232	0.279	-.7761645	0.005	-.4222127	0.000
alter_3	-.2079536	0.351	.1064895	0.741	-.2643854	0.006
alter_5	-.5569189	0.193	-2.519568	0.004	.3537929	0.118
sex_1	.2944272	0.062	-.0110032	0.951	-.0389449	0.552
nichtdeutsch	-.4654752	0.066	-.8832104	0.000	-.3245484	0.004
schulbil_1	.1897246	0.628	.9001152	0.094	-.2689785	0.069

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

	Willingness to participate in the panel		Contact		Participation	
	Coef.	p	Coef	p	Coef.	p
schulbil_2	.0037373	0.985	.2090964	0.337	-.0348199	0.668
schulbil_4	-.1861201	0.338	.071271	0.738	.1314541	0.115
gesundheit_1	-.1751737	0.525	.2978212	0.404	.0357796	0.769
gesundheit_2	-.054073	0.777	.0830684	0.713	-.0377541	0.638
gesundheit_4	.1439384	0.531	-.2450635	0.299	.0540237	0.566
gesundheit_5	-.0633256	0.841	-.2800593	0.419	-.0437639	0.746
zufrieden_1	-.1533756	0.755	.5001694	0.416	.1844538	0.372
zufrieden_2	-.2038934	0.323	-.127826	0.560	.0212367	0.806
zufrieden_4	-.2071745	0.326	-.1328277	0.615	-.0110766	0.908
anz_0_3	.7826537	0.022	.3724317	0.179	.0886097	0.342
anz_4_6	-.3191938	0.184	.7096552	0.065	-.0067116	0.949
anz_7_14	.0022726	0.988	.0154742	0.937	-.062216	0.316
anz_15_64	-.0171884	-0.874	.1286946	0.379	.0112485	0.814
anz_65	-.1125482	0.665	1.617795	0.029	.0036391	0.979
eigentum	-.1125482	0.665	1.617795	0.029	.0036391	0.979
wnka_1	.4313943	0.014	.2140892	0.256	.0811439	0.242
wnka_3	-.7038993	0.001	.1272896	0.642	-.0592294	0.558
hhincome_1	.3298667	0.234	-.2711439	0.335	-.095104	0.386
hhincome_2	-.0433029	0.843	-.0335898	0.894	-.0009574	0.992
hhincome_4	.0224275	0.921	.2731451	0.383	.1037942	0.291
alg2_1	-.2083662	0.305	-.0249373	0.905	.098082	0.228
stichprobe1					-.3004804	0.009
stichprobe3					-.1974554	0.306
stichprobe4					-.4036901	0.013
stichprobe5					-.3793223	0.029
stichprobe6					-.2099766	0.124
stichprobe7					-.2671077	0.074
stichprobe8					-.3391149	0.054
stichprobe9					-.3974861	0.007
stichprobe10					-.3171113	0.027
stichprobe11					-.800467	0.000

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

	Willingness to participate in the panel		Contact		Participation	
	Coef.	p	Coef	p	Coef.	p
stichprobe_ba	.1493704	0.465	-.7000264	0.013		
blneualt_2	.0970008	0.585	.443215	0.033		
bundesld_1					-.1819051	0.333
bundesld_2					.2863135	0.260
bundesld_3					-.2406312	0.040
bundesld_4					.0123228	0.973
bundesld_6					-.2292433	0.125
bundesld_7					-.3377148	0.040
bundesld_8					-.0207589	0.872
bundesld_9					.0310587	0.792
bundesld_10					.1301885	0.674
bundesld_11					-.0582538	0.699
bundesld_12					.0311493	0.854
bundesld_13					.1977306	0.398
bundesld_14					.3163393	0.042
bundesld_15					.318715	0.084
bundesld_16					-.1810984	0.272
bik_1					-.4283747	0.175
bik_2					-.0209543	0.922
bik_3					.0886966	0.501
bik_4					.039898	0.756
bik_5					.3626697	0.019
bik_6					-.057666	0.767
bik_7					.0431659	0.716
bik_8					.2829606	0.012
bik_9					.1288702	0.333
anzkon_1			-1.624666	0.000		
anzkon_3			-.2618605	0.311		
anzkon_4			-1.488486	0.000		
cons	3.96387	0.000	5.334668	0.000	2.195595	0.000
n	8998		8816		8671	
Log likelihood	-860.89256		-637.39768		-3351.5607	
Pseudo R^2	0.0328		0.1380		0.0456	

6.4 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 that had not been interviewed in the previous waves were considered. This condition means that the participation propensities for first-time participating split-off households were modeled separately following the criterion of originating in wave 8 (split-off W8 households) or originating in wave 9 (split-off W8 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 when 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 (waves 8 and 9)

Variable code and reference category	Explanation
alter_1(Split W8)	Household reference person (HRP) younger than 30 years
alter_2(Split W8)	HRP 30-39 years of age
alter_4(Split W8)	HRP 50-59 years of age
alter_5(Split W8)	HRP 60 years and older
Reference category (Split W8)	HRP 40-49 years of age
alter_1(Split W9)	Household reference person (HRP) younger than 30 years
alter_2(Split W9)	HRP 30-39 years of age
alter_3(Split W9)	HRP 40-49 years of age
Reference category (Split W8)	HRP 50 years or older
sex_1	HRP male
Reference category	HRP female
nichtdeutsch	HRP has nationality other than German
Reference category	HRP has German nationality or missing information

Table 37: Variable overview, codes and reference categories for the logit models of the split-off households participating for the first time (waves 8 and 9) (continued)

Variable code and reference category	Explanation
schulbil_1	School qualification HRP: no qualification, lower secondary school
schulbil_3	School qualification HRP: college/university qualification
Reference category	School qualification HRP: intermediate secondary school
stichprobe_ba	BA samples (incl. BA refreshment samples and BA replenishment sample)
Reference category	Microm sample (incl. EWO replenishment sample)
anzkon_1	Number of contact attempts CATI/CAPI: 1 contact attempt
anzkon_1	Number of contact attempts CATI/CAPI: 4-9 contact attempts
anzkon_1	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 8 households for participation

	Participation	
	Coef.	p
alter_1	-.5381979	0.373
alter_2	.3423103	0.578
alter_4	.4657316	0.324
alter_5	-1.058713	0.213
sex_1	-.3246385	0.428
nichtdeutsch	.3067525	0.599
schulbil_1	.9621993	0.059
schulbil_3	1.425092	0.007
stichprobe_ba	-.281692	0.512
cons	-2.124076	0.000
n	216	
Log likelihood	-92.475775	
Pseudo R^2	0.0796	

Table 39: Logit models on the first participation of split-off wave 9 households for availability and participation

	Contact		Participation	
	Coef.	p	Coef	p
alter_1	1.525588	0.156	-1.122849	0.005
alter_2	.3703229	0.660	-.2655897	0.519
alter_3	-.8855558	0.056	.0930679	0.735
sex_1	.1789398	0.680	-.4294758	0.071
nichtdeutsch	-.271658	0.693	-.3965396	0.364
schulbil_1	.5106991	0.318	-.0938394	0.732
schulbil_3	-.0306574	0.954	.1307488	0.651
anzkon_1	-1.812903	0.001		
anzkon_3	.8917947	0.433		

Table 39: Logit models on the first participation of split-off wave 9 households for availability and participation (continued)

	Contact		Participation	
	Coef.	p	Coef	p
anzkon_4	.4763812	0.677		
stichprobe_ba	.0782445	0.862	-.3130844	0.205
cons	3.386681	0.000	-.1964116	0.462
n	393		365	
Log likelihood	-83.562315		-221.10221	
Pseudo R^2	0.1722		0.0374	

6.5 Nonresponse weighting for households from the BA refreshment sample and the BA panel replenishment sample of wave 9

Again, a nonresponse modeling for the households from the refreshment sample of BA new inflows into UB II receipt (sample = 12) was performed (participation) similar to the wave 8 refreshment sample. As the participation of the refreshment sample was very high (only n=52 cases did not participate), no separate modelling of the contact was implemented. The participation probability derived from this procedure can be found in variable *propt0*.

Table 40: Variable overview, codes and reference categories for the logit models of the BA refreshment sample of wave 9

Variable code and reference category	Explanation
alter_2	HRP 30-39 years of age
alter_3	HRP 40-49 years of age
alter_4	HRP 50-65 years of age
Reference category	Household reference person (HRP) younger than 30 years
sex_1	HRP male
Reference category	HRP female
nichtdeutsch	HRP has nationality other than German

Table 40: Variable overview, codes and reference categories for the logit models of the BA refreshment sample of wave 9 (continued)

Variable code and reference category	Explanation
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
schulbil_5	School qualification HRP: Details refused
Reference category	School qualification HRP: intermediate secondary school
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 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
Reference category	Marital status: single
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

Table 40: Variable overview, codes and reference categories for the logit models of the BA refreshment sample of wave 9 (continued)

Variable code and reference category	Explanation
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 to under 5,000 (BIK-Region size classes 1 and 2 combined)
bik_2	BIK size class of municipality: population of 5,000 to under 20,000
bik_3	BIK size class of municipality: population of 20,000 to under 50,000
bik_4	BIK size class of municipality: population of 50,000 to under 100,000 STYP 2/3/4
bik_5	BIK size class of municipality: population of 50,000 to under 100,000 STYP 1
bik_6	BIK size class of municipality: population of 100,000 to under 500,000 STYP 2/3/4
bik_7	BIK size class of municipality: population of 100,000 to under 500,000 STYP 1
bik_8	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 41: Logit models on the first participation for availability and participation of the BA refreshment sample and BA replenishment sample of wave 9

	Contact	
	Coef.	p
alter_2	-.1901391	0.091
alter_3	-.160985	0.205
alter_4	.0800658	0.530
sex_1	-.0895635	0.311
nichtdeutsch	-.2497015	0.010
schulbil_1	-.4374738	0.047
schulbil_2	-.113186	0.407
schulbil_4	.0887874	0.551
schulbil_5	-.2139982	0.076
anz_persBG_2	1.982166	0.022
anz_persBG_3	1.877029	0.035
anz_verwfBG_1	.9236316	0.082
anz_verwfBG_3	-.3499395	0.167
BG_typ_2	-1.840855	0.036
BG_typ_3	-1.560693	0.067
BG_typ_4	-1.274595	0.142
BG_typ_5	-1.797158	0.028
famstand_2	.0612221	0.678
famstand_3	.1975231	0.158
famstand_4	-.0762004	0.636
bundesld_1	-.0040559	0.987
bundesld_2	-.3080527	0.266
bundesld_3	-.2266489	0.143
bundesld_4	-1.134361	0.034
bundesld_6	.0212446	0.894
bundesld_7	.3122807	0.104
bundesld_8	-.1177594	0.455
bundesld_9	.0680577	0.618
bundesld_10	-.0543557	0.876

Table 41: Logit models on the first participation for availability and participation of the BA refreshment sample and BA replenishment sample of wave 9 (continued)

	Contact	
	Coef.	p
bundesld_11	.1084046	0.521
bundesld_12	-.0378562	0.879
bundesld_13	-.1045887	0.773
bundesld_14	.5494685	0.009
bundesld_15	-.1442964	0.596
bundesld_16	-.1355983	0.624
bik10_1	-.0159311	0.956
bik10_2	.1261773	0.464
bik10_3	.1131705	0.511
bik10_4	.0839847	0.671
bik10_5	-.4066891	0.131
bik10_6	.2036228	0.161
bik10_7	.0264447	0.833
bik10_8	.356567	0.023
cons	-.8693584	0.000
n	3408	
Log likelihood	-1925.1568	
Pseudo R^2	0.0215	

6.6 Propensity to participate again - individuals

The decisive longitudinal weight is not the household but the individual-level weight because these units are stable over time. The propensities to participate again for individuals in wave 9 were estimated using 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 were considered in these models by clustering the error terms at the household level. The predicted propensities of the models were multiplied. The reciprocal value of this product can be found in variable *ppbleib*. The longitudinal weight for an individual for the period [t1; t2; t3; t4; t5; t6, t7, t8, t9] across all six waves can be obtained as the product of the cross-sectional weight to t1, *ppbleib* (wave 1 to wave 2) and *ppbleib* (wave 2 to wave 3, etc.).

Table 42: 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_1	Subjective evaluation of the health state of the individual: very good
gesundheit_2	Subjective evaluation of the health state of the individual: good

Table 42: Variable overview, codes and reference categories for the logit models of re-participating individuals (continued)

Variable code and reference category	Explanation
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: satisfactory
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_4	General life satisfaction of the individual: scale value 9-10
Reference category	General life satisfaction of the individual: scale value 6-8
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_15_64	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

Table 42: Variable overview, codes and reference categories for the logit models of re-participating individuals (continued)

Variable code and reference category	Explanation
stichprobe3	Refreshment sample (BA) wave 2
stichprobe4	Refreshment sample (BA) wave 3
stichprobe5	Refreshment sample (BA) wave 4
stichprobe6	Replenishment sample (EWO) wave 5
stichprobe7	Replenishment sample (BA) wave 5
stichprobe8	Refreshment sample (BA) wave 5
stichprobe9	Refreshment sample (BA) wave 6
stichprobe10	Refreshment sample (BA) wave 7
stichprobe11	Refreshment sample (BA) wave 8
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
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

Table 42: Variable overview, codes and reference categories for the logit models of re-participating individuals (continued)

Variable code and reference category	Explanation
bik_1	BIK size class of municipality: population of less than 2,000
bik_2	BIK size class of municipality: population 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 43: 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	.1646478	0.474	-1.294856	0.000	-.5226136	0.000
alter_2	.4284604	0.113	-.7988951	0.015	-.1185355	0.180
alter_4	-.1376607	0.534	-.1935079	0.584	.3033766	0.000
alter_5	-.6690164	0.031	-2.014648	0.005	.337171	0.027
sex_1	.1850045	0.063	-.074714	0.605	-.0560105	0.192
nichtdeutsch	.1302709	0.643	-.6715557	0.020	-.2488936	0.011
schulbil_1	.2826818	0.424	.6930433	0.124	-.2915107	0.011
schulbil_2	.0879565	0.624	-.0245782	0.909	-.0488271	0.451

Table 43: Logit models on re-participation for willingness to participate in a panel, availability and participation (continued)

	Willingness to participate in the panel		Contact		Participation	
	Coef.	p	Coef	p	Coef.	p
schulbil_4	.0038316	0.984	-.0066212	0.975	.1138425	0.093
gesundheit_1	-.1523519	0.513	.5499988	0.107	-.1011758	0.265
gesundheit_2	-.0933949	0.586	.1155959	0.568	-.0616845	0.332
gesundheit_4	.1248093	0.550	-.2781175	0.214	.0051302	0.946
gesundheit_5	-.1594422	0.569	-.3700654	0.266	-.1399806	0.206
zufrieden_1	-.0067287	0.988	.0665333	0.892	-.0708894	0.661
zufrieden_2	-.4303311	0.033	-.1232092	0.521	-.0009525	0.990
zufrieden_4	-.2369665	0.223	-.0761642	0.768	.0348539	0.634
anz_0_3	.9229451	0.012	.2911601	0.265	.0658427	0.471
anz_4_6	-.4460809	0.076	.7042901	0.069	-.0087004	0.930
anz_7_14	-.0394324	0.820	-.1105583	0.707	-.0457342	0.448
anz_15_64	.081635	0.470	.1998433	0.194	-.0823238	0.034
anz_65	.1307198	0.528	1.309645	0.021	-.0901588	0.351
eigentum	.3151747	0.185	.4702481	0.288	.0983981	0.267
wnka_1	.5033232	0.003	.0208692	0.905	.0952009	0.097
wnka_3	-.4604173	0.021	.0946789	0.700	-.0837971	0.303
hhincome_1	.3781639	0.233	-.2504535	0.428	-.1602231	0.133
hhincome_2	-.097033	0.681	-.1166503	0.681	-.0679374	0.448
hhincome_4	-.1630718	0.492	.3890289	0.264	-.0299932	0.741
alg2_1	-.0331612	0.872	-.1174222	0.597	.0445435	0.577
stichprobe1	1.126207	0.001	-.6120877	0.093	-.3134851	0.003
stichprobe3	-.1655111	0.689	-.5199342	0.355	-.3129791	0.086
stichprobe4	.8286138	0.089	-.7284642	0.156	-.3551866	0.019
stichprobe5	.045512	0.931	-1.574942	0.002	-.3534101	0.035
stichprobe6	-.3922883	0.154	.1074404	0.844	-.0544099	0.672
stichprobe7	1.094644	0.014	-.0465077	0.936	-.2466129	0.083
stichprobe8	.3177506	0.462	-.8328183	0.111	-.3392883	0.050
stichprobe9	.4830901	0.272	-.3272227	0.507	-.319326	0.030
stichprobe10	-.0448332	0.902	-.8938388	0.036	-.228335	0.099
stichprobe11	-1.064421	0.000	-1.238166	0.005	-.7548493	0.000
blneualt_2	.0788375	0.690	.3328125	0.153		

Table 43: Logit models on re-participation for willingness to participate in a panel, availability and participation (continued)

	Willingness to participate in the panel		Contact		Participation	
	Coef.	p	Coef	p	Coef.	p
bundesld_1					-.1694366	0.368
bundesld_2					.1199952	0.633
bundesld_3					-.2274303	0.049
bundesld_4					.1214894	0.734
bundesld_6					-.2623068	0.068
bundesld_7					-.2747483	0.091
bundesld_8					-.1449369	0.234
bundesld_9					.0489262	0.663
bundesld_10					.3326147	0.271
bundesld_11					-.1334251	0.340
bundesld_12					-.003701	0.982
bundesld_13					.0611443	0.783
bundesld_14					.29045	0.048
bundesld_15					.0866817	0.583
bundesld_16					-.2069082	0.204
bik_1			.0872487	0.939	-.6613138	0.024
bik_2			1.51142	0.149	.0486796	0.798
bik_3			.2177005	0.542	.0778658	0.531
bik_4			.3657613	0.374	.151768	0.225
bik_5			-.1181657	0.761	.2131477	0.129
bik_6			.0337765	0.966	-.0677938	0.708
bik_7			.1264536	0.699	.0216642	0.846
bik_8			.3008612	0.255	.2852531	0.008
bik_9			.0351565	0.921	.1485125	0.255
anzkon_1			-1.646939	0.000		
anzkon_3			-.0420977	0.880		
anzkon_4			-1.487943	0.000		
cons	3.656049	0.000	5.52969	0.000	2.144722	0.000
n	13460		13211		13030	
Log likelihood	-1156.2061		-809.64338		-5473.5543	
Pseudo R^2	0.0677		0.1534		0.0427	

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

6.7 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 8) that were modified by the nonresponse modeling. The multiple selection probability of a sampled benefit recipient living in the same household as a benefit recipient in previous years without being a member of the benefit unit himself/herself was ignored. The new design weights of the benefit recipient sample are projected in the cross-section to all individuals who were living in a household that included at least one benefit unit in either July 2006, in July 2007, in July 2008, in July 2009, in July 2010, in July 2011, in July 2012, in July 2013 or in July 2014. It is only when calculating new weights for the total sample that it becomes necessary to adjust the weights for all households receiving benefits in July 2013. For this adjustment, the inclusion probability in the other sample was estimated for cases from the Microm sample (wave 1), EWO replenishment sample (wave 5) and new refreshment sample (wave 9). For cases from the refreshment sample, the mean wave 1 selection probability in the Microm sample respectively, the mean wave 5 selection probability of EWO refreshment in the respective postcode area and the average participation probability (for waves 1 to 9) in that sample were assumed. For cases from the Microm sample, if they are (according to survey data) new recipients of UB II who first received the benefit between the last five sampling dates (waves 2, 3, 4, 5, 6, 7, 8 and 9), 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.

6.8 Integration of temporary non-responses (households)

Households that skipped one wave - i.e., did not participate (temporary nonresponses) - could participate again in wave 9, as was possible in wave 8. 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. To calculate mutual cross-sectional weights including the temporary nonresponses, there was a convex combination of the modified household weights of the temporary nonresponses and the modified household weights of the panel household sample (not of the refreshment sample) before calibration. Thus, the convex combination of the household weights was made before calibration; the calibration was then made with the new combined household weights.

Although the household weights modified by nonresponse modeling already serve as projection factors for the panel and refreshment sample, it was necessary to calculate such modified household weights as an estimator for the respective population again for the temporary nonresponses. The starting point was the calibrated household weights of wave

7 (wave 8 is the temporary non-response).

For temporary nonresponses, the probability of non-participation in wave 8 in case of participation in wave 7 (non-participation propensities wave 8) and the probability of participation in wave 9 in case of a non-participation in wave 8 (participation propensities wave 9) was determined. The probability of non-participation in wave 8 is calculated from 1– participation probability in wave 8.

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 nonresponses was then calculated by multiplying the calibrated household weights of wave 7 by the reciprocal value of this product.

Table 44: Variable overview, codes and reference categories for the logit models of the temporary nonresponses

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_3	HRP 40-49 years of age
alter_5	HRP 65 years and older
Reference category	HRP 50-64 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_1	Subjective evaluation of the health state of the HRP: very good
gesundheit_3	Subjective evaluation of the health state of the HRP: satisfactory

Table 44: Variable overview, codes and reference categories for the logit models of the temporary nonresponses (continued)

Variable code and reference category	Explanation
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: good
zufrieden_1	General life satisfaction HRP: scale value 0-2
zufrieden_2	General life satisfaction HRP: scale value 3-5
zufrieden_4	General life satisfaction HRP: scale value 9-10
Reference category	General life satisfaction HRP: scale value 6-8
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_15_64	Number of individuals in the household aged 15 - 64 years
anz_65	Number of individuals in the household aged 65 years and older
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
bundesld_1	Federal state: Schleswig-Holstein
bundesld_2	Federal state: Hamburg
bundesld_3	Federal state: Lower-Saxony

Table 44: Variable overview, codes and reference categories for the logit models of the temporary nonresponses (continued)

Variable code and reference category	Explanation
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 of temporary nonresponses

	Re-participation in wave 8 to determine the W8 non-participation probability (1-participation probability W8)		Re-participation in wave 9 in case of non-participation in wave 8	
	Coef.	p	Coef	p
alter_1	-.924089	0.000	-.3926952	0.037
alter_2	-.6455395	0.000	.0135084	0.941
alter_3	-.3928697	0.000	-.0018142	0.991
alter_5	.035986	0.852	-.2264568	0.629
sex_1	-.0634574	0.258	-.1234588	0.308
nichtdeutsch	-.3809893	0.000	-.1875375	0.348
schulbil_1	-.2053132	0.106	-.3704456	0.199
schulbil_2	-.079217	0.248	.0507039	0.731
schulbil_4	.1530206	0.035	.1023795	0.513
gesundheit_1	-.1175843	0.248	.1087506	0.614
gesundheit_3	.0269191	0.696	.4402197	0.003
gesundheit_4	.2218043	0.007	.2347918	0.195
gesundheit_5	.1075586	0.359	.1145779	0.669
zufrieden_1	-.2367545	0.131	-.3943067	0.269
zufrieden_2	.0288349	0.695	-.4009352	0.015
zufrieden_4	.0582089	0.468	.1987837	0.238
anz_0_3	-.0056469	0.942	.1851215	0.255
anz_4_6	.1160996	0.195	.311096	0.077
anz_7_14	-.0104847	0.839	.0331403	0.746
anz_15_64	-.1050558	0.007	.0323568	0.703
anz_65	.0923899	0.429	-.348911	0.240
eigentum	.0612504	0.436	.044502	0.794
wnka_1	.2083306	0.000	.1470659	0.249
wnka_3	-.1771666	0.036	.0059751	0.974
hhincome_1	-.2254078	0.019	.3315647	0.112

Table 45: Logit models of temporary nonresponses (continued)

	Re-participation in wave 8 to determine the W8 non-participation probability (1-participation probability W8)		Re-participation in wave 9 in case of non-participation in wave 8	
	Coef.	p	Coef	p
hhincome_2	-.2470782	0.003	.2049333	0.250
hhincome_4	.0376452	0.652	.1261037	0.484
alg2_1	.0118046	0.865	-.0276084	0.847
bundesld_1	-.0234796	0.881	.1695819	0.601
bundesld_2	-.2492006	0.184	.5585375	0.111
bundesld_3	.0643776	0.537	-.1478086	0.524
bundesld_4	.3150956	0.321	-.2157744	0.752
bundesld_6	-.2386218	0.048	.056217	0.826
bundesld_7	-.2559816	0.068	.0274748	0.925
bundesld_8	-.1083103	0.304	.0376094	0.865
bundesld_9	.0823783	0.398	.1994604	0.326
bundesld_10	.1330134	0.601	-.6488084	0.321
bundesld_11	.2593866	0.057	.0656259	0.820
bundesld_12	.4210398	0.006	-.4634187	0.245
bundesld_13	.0136652	0.942	-.13475	0.745
bundesld_14	.5240086	0.000	.1831355	0.538
bundesld_15	.3334854	0.024	.2002333	0.523
bundesld_16	.3320182	0.036	-.4645404	0.233
bik_1	.3484311	0.319	.9611666	0.157
bik_2	.3520889	0.071	.3226434	0.407
bik_3	.1910563	0.091	-.1859611	0.446
bik_4	-.0356301	0.756	-.1403298	0.568
bik_5	.1040279	0.386	-.3776063	0.166
bik_6	-.0854069	0.580	-.1155159	0.723
bik_7	.1767652	0.085	-.1669938	0.454
bik_8	.1305952	0.161	.0611594	0.749
bik_9	.0735452	0.514	-.3018147	0.236
cons	1.873077	0.000	-1.437973	0.000
n	9509		1751	
Log likelihood	-4341.8591		-917.39726	
Pseudo R^2	0.0440		0.0349	

The convex combination of the weights of the participants across all waves (panel household sample) and the temporary nonresponses 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 nonresponses from the total sample, i.e., the sum of the panel household sample and temporary nonresponses:

$dwihh_{temp.Ausfall} * (n_{temp.Ausfalli}/(n_{temp.Ausfalli} + n_{Bestandi}))$ for temporary nonresponses and

$dwihh_{Bestand} * (n_{Bestandi}/(n_{temp.Ausfalli} + n_{Bestandi}))$ for the panel household sample.

6.9 Calibration to the household weight, wave 9, cross-section

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 2014, followed. For households receiving benefits the weights were adjusted to the statistics of the Federal Employment Agency for July 2014. As in the previous year, the increase in UB II receipt since the previous year at the level of benefit units (269,085) was also included as an additional benchmark value in the total sample. Cases in the previous samples from waves 1 to 9 that, according to wave 9 of the survey, were receiving UB II in July 2014, will be projected to the benchmark statistics of the Federal Employment Agency on UB 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, a large variance of the estimation functions can result. 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 also to be avoided. Therefore, attempts are made to avoid very large weighting factors (and subsequently, very small factors) whenever possible and to make appropriate corrections to the weights if necessary. Within the framework of the calibration at hand, these corrections are made at 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 were determined by using particular percentiles of the distribution depending on the distribution of the design weights.
- In addition, the interval of weights was limited during calibration, i.e., a maximum and a minimum limit for weights was determined. Here, 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. Notably, 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, in addition to the average value and the standard deviation, the efficiency measure (E) is described as follows. The efficiency measure E is based on the variance of the weighting factor. The efficiency measure indicates the size of the effective case number of a passive characteristic that does not correlate with active characteristics 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.

6.10 Calibration of the BA sample

The population of the cumulated BA sample of all nine waves consists of all of the households in Germany with at least one benefit unit receiving benefits according to SGB II at one of the (until now) nine drawing dates (July 2006, July 2007, July 2008, July 2009, July 2010, July 2011, July 2012, July 2013 or July 2014). In wave 9, only the benchmark values of the BA statistics from July 2014 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 according to SGB II was living in July 2014. The starting points for the calibration were modified design weights, including the nonresponse weighting. The modified design weights were trimmed at the fifth and ninety-fifth percentiles of their distribution and then rescaled so that they totaled the untrimmed design weights. The projection factors of the trimmed design weights range from 197,22 to 4315,7. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.3 and upwards to 2.5. Thus, the total projection factors after calibration lie between a minimum of 59,17 and a maximum of 6267,63.

A calibration was made for the following characteristics:

Benefit unit basis BA statistics:

- Increase in BU UB II recipients
- Number of BCs receiving benefits according to SGB II by federal states
- Number of BCs receiving benefits according to SGB II by number of individuals under 65 years of age in the benefit unit and by west/east
- Number of BCs receiving benefits according to SGB II by number of children under 15 years of age in the benefit unit and by west/east

- Number of BCs receiving benefits according to 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 UB II recipients since the previous year at the level of benefit units (269.085).

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

Because 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 nonresponse item thus leads to slight deviations from the values as presented in the following.

Table 46: 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	105	118.163	118.163
	Number BCs Hamburg	74	100.471	100.471
	Number BCs Lower-Saxony	311	304.558	304.558
	Number BCs Bremen	37	50.843	50.843
	Number BCs North Rhine-Westphalia	835	844.110	844.110
	Number BCs Hesse	196	211.160	211.160
	Number BCs Rhineland-Palatinate	137	116.049	116.049
	Number BCs Baden-Wuerttemberg	234	230.451	230.451

Table 46: Nominal distributions and distributions after calibration (BA sample, house-) holds (continued)

Benchmark Figure	Characteristics benchbenchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number BCs Bavaria	301	233.231	233.231
	Number BCs Saarland	53	43.072	43.072
	Number BCs Berlin	245	313.357	313.357
	Number BCs Brandenburg	151	144.807	144.807
	Number BCs Mecklenburg-Vorpommern	77	108.627	108.627
	Number BCs Saxony	225	229.626	229.626
	Number BCs Saxony-Anhalt	173	160.895	160.895
	Number BCs Thuringia	119	103.461	103.461
Number 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)	976	1.249.634	1.249.634
	Number BCs with 2 individual under 65 (west)	579	444.020	444.020
	Number BCs with 3 individual under 65 (west)	362	272.340	272.340
	Number BCs with 4 individual under 65 (west)	195	166.625	166.625
	Number BCs with 5 or more individuals under 65 (west)	171	119.489	119.489
	Number BCs with 1 individual under 65 (east)	503	641.158	641.158

Table 46: Nominal distributions and distributions after calibration (BA sample, house-) holds (continued)

Benchmark Figure	Characteristics bench-benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number BCs with 2 individual under 65 (east)	252	212.868	212.868
	Number BCs with 3 individual under 65 (east)	143	109.843	109.843
	Number BCs with 4 individual under 65 (east)	47	60.247	60.247
	Number BCs with 5 or more individuals under 65 (east)	45	36.657	36.657
Number BCs receiving benefits in accordance with SGB II by number of children under 15 years of age in the benefit unit (1, 2, 3, "4 or more") and by west/ east (10 categories)	Number BCs without children under 15 years (west)	1.559	1.545.685	1.545.685
	Number BCs with 1 child under 15 years (west)	416	379.037	379.037
	Number BCs with 2 children under 15 years (west)	203	217.527	217.527
	Number BCs with 3 children under 15 years (west)	85	79.100	79.100
	Number BCs with 4 or more children under 15 years (west)	20	30.759	30.759
	Number BCs without children under 15 years (east)	759	779.087	779.087
	Number BCs with 1 child under 15 years (east)	135	160.802	160.802
	Number BCs with 2 children under 15 years (east)	70	83.895	83.895

Table 46: Nominal distributions and distributions after calibration (BA sample, house-) holds (continued)

Benchmark Figure	Characteristics benchbenchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number BCs with 3 children under 15 years (east)	22	26.312	26.312
	Number BCs with 4 or more children under 15 years (east)	4	10.677	10.677
Number BCs receiving benefits in accordance with SGB II constisting of a single parent with children by west/east (4 categories)	Number BCs with a single parent (west)	492	448.398	448.398
	Rest BCs without a single parent (west)	1.791	1.803.710	1.803.710
	Number BCs with a single parent (east)	154	183.565	183.565
	Rest BCs without a single parent (east)	836	877.208	877.208

Table 47: Parameters of distribution of weights (BA-sample, households)

1%-percentile	125,8683
5%-percentile	162,3364
10%-percentile	187,5593
25%-percentile	265,271
50%-percentile	470,1457
75%-percentile	1293,411
90%-percentile	3054,19
95%-percentile	3991,468
99%-percentile	4794,326
Mean	1038,156
Standard deviation	1206,57
Minimum	59,16505
Maximum	6267,625
Number of observations	3.149
Efficiency measure	**42,6%**

6.11 Population sample

All private households in Germany form the population. The starting points for the calibration were modified design weights, including the nonresponse weighting. The modified de-sign weights were trimmed at the fifth and ninety-fifth percentiles of their distribution and after that rescaled so that they totaled the untrimmed design weights. The projection factors of the trimmed design weights range from 2851,14 to 40983,21.. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.1 and upwards to 3.0. Thus, the total projection factors after calibration lie between minimal 2383 and maximal 63175,04.

A calibration was made for the following characteristics:

1. Benefit units based on BA statistics:
    - Number of BCs receiving benefits according to SGB II by federal states
    - Number of BCs receiving benefits according to SGB II by number of individuals under 65 years of age in the benefit unit and by west/east
    - Number of BCs receiving benefits according to SGB II by number of children under 15 years of age in the benefit unit and by west/east

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

2. Households based on Mikrozensus 2014:
   - 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 nonresponse item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Because 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 nonresponse item thus leads to slight deviations from the values as presented in the following.

Table 48: Nominal distributions and distributions after calibration (population 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 west/ east (2 categories)	Number BGs west	88	2.252.108	2.252.109
	Number BGs east	42	1.060.773	1.060.773
Number BCs receiving benefits in accordance with SGB II by number of individuals under 65 years of age in the benefit unit (4 categories)	Number BCs with 1 individual under 65	57	1.890.792	1.890.792
	Number BCs with 2 individual under 65	30	656.888	656.888
	Number BCs with 3 individual under 65	23	382.183	382.183
	Number BCs with 4 individ-	20	383.018	383.019

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	ual under 65			
Number BCs receiving benefits in accordance with SGB II by number of children under 15 years of age in the benefit unit (2 categories)	Number BCs without children under 15 years (west)	93	2.324.772	2.324.773
	Rest BCs with 1 child or more under 15 years (west)	37	988.109	988.109
Number BCs receiving benefits in accordance with SGB II constisting of a single parent with children (2 categories)	Number BCs with a single parent (west)	28	631.963	631.964
	Rest BCs without a single parent (west)	102	2.680.918	2.680.918
Number of households by federal state and BIK type (spelling: “Federal state. BIK type”; 38 categories)	1.1 to 1.6	28	476.000	476.000
	1.7 to 1.10	54	911.000	911.000
	2.10	38	966.000	966.000
	3.1 to 3.5	96	1.374.000	1.374.000
	3.7 to 3.8	123	1.381.000	1.381.000
	3.9 to 3.10	81	1.023.000	1.023.000
	4.8 to 4.10	19	359.000	359.000
	5.2 to 5.4	95	1.129.000	1.129.000
	5.5 to 5.6	80	942.000	942.000
	5.7 to 5.8	197	2.917.000	2.917.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	5.9 to 5.10	247	3.520.000	3.520.000
	6.1 to 6.4	61	647.000	647.000
	6.5 to 6.8	71	1.065.000	1.065.000
	6.9 to 6.10	78	1.186.000	1.186.000
	7.1 to 7.6	60	928.000	928.000
	7.7 to 7.10	67	952.000	952.000
	8.1 to 8.4	104	1.226.000	1.226.000
	8.5 to 8.8	112	2.140.000	2.140.000
	8.9 to 8.10	110	1.612.000	1.612.000
	9.1 to 9.4	138	1.571.000	1.571.000
	9.5 to 9.7	141	1.473.000	1.473.000
	9.8 to 9.9	113	1.415.000	1.415.000
	9.10	103	1.593.000	1.593.000
	10.3 to 10.8	37	490.000	490.000
	11.10	90	1.949.000	1.949.000
	12.1 to 12.4	34	459.000	459.000
	12.5 to 12.7	29	261.000	261.000
	12.9 to 12.10	29	509.000	509.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	13.1 to 13.6	26	514.000	514.000
	13.7 to 13.9	21	308.000	308.000
	14.1 to 14.4	64	569.000	569.000
	14.5 to 14.8	24	640.000	640.000
	14.9 to 14.10	60	923.000	923.000
	15.1 to 15.4	58	393.000	393.000
	15.5 to 15.7	29	485.000	485.000
	15.8 to 15.9	27	274.000	274.000
	16.1 to 16.4	70	529.000	529.000
	16.5 to 16.8	55	563.000	563.000
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)	629	12.361.000	12.361.000
	Number households with 2 individuals (west)	921	10.706.000	10.706.000
	Number households with 3 individuals (west)	311	3.943.000	3.943.000
	Number households with 4 individuals (west)	286	3.119.000	3.119.000
	Number households with 5 or more individuals (west)	106	1.167.000	1.167.000
	Number households with	202	3.636.000	3.636.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	1 individual (east)			
	Number households with 2 individuals (east)	272	3.026.000	3.026.000
	Number households with 3 individuals (east)	74	1.007.000	1.007.000
	Number households with 4 individuals (east)	45	545.000	545.000
	Number households with 5 or more individuals(east)	23	162.000	162000
Number of households by “children under 15 years of age in the household "yes/no” and west/east (4 categories)	Number households with children under 15 years (west)	412	5.481.000	5.481.000
	Number households without children under 15 years (west)	1.841	25.815.000	25.815.000
	Number households with children under 15 years (east)	88	1.319.000	1.319.000
	Number households without children under 15 years (east)	528	7.057.000	7.057.000

Table 49: Parameters of distribution of weights (Population sample, households)

1%-percentile	2709,679
5%-percentile	3019,569
10%-percentile	3578,792
25%-percentile	5531,814
50%-percentile	9426,067
75%-percentile	18385,89
90%-percentile	32744,68
95%-percentile	38690,99
99%-percentile	49918,72
Mean	13827,81
Standard deviation	11500,43
Minimum	2382,988
Maximum	63175,04
Number of observations	2.869
Efficiency measure	**59,1%**

6.12 Total sample

All of the private households in Germany form the population. The starting points for the calibration were modified design weights, including the non-response weighting. The modified design weights were trimmed at the fifth and ninety-fifth percentiles of their distribution and after that rescaled so that they totaled the untrimmed design weights. The projection factors of the trimmed design weights range from 168,7 to 23975,6. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.2 and upwards to 5.0. Thus, the total projection factors after calibration lie between min. 54,8 and max. 30683.

A calibration was made for the following characteristics:

1. Benefit unit basis BA statistics:
    - Number of BCs receiving benefits according to SGB II by federal states
    - Number of BCs receiving benefits according to SGB II by number of individuals under 65 years of age in the benefit unit and by west/east
    - Number of BCs receiving benefits according to SGB II by number of children under 15 years of age in the benefit unit and by west/east

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

2. Household basis Mikrozensus 2014:

   - 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

In addition, the increase in UB II recipients since the previous year at the level of benefit units (269.085) 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.

Because 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 below.

Table 50: Nominal distributions and distributions after calibration (total 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	108	118.163	118.163
	Number BCs Hamburg	76	100.471	100.471
	Number BCs Lower-Saxony	328	304.558	304.557
	Number BCs Bremen	38	50.843	50.843
	Number BCs North Rhine-Westphalia	879	844.110	844.101
	Number BCs Hesse	196	211.160	211.153
	Number BCs Rhineland-Palatinate	143	116.049	116.045
	Number BCs Baden-Wuerttemberg	237	230.451	230.450
	Number BCs Bavaria	312	233.231	233.230
	Number BCs Saarland	54	43.072	43.071
	Number BCs Berlin	251	313.357	313.355
	Number BCs Brandenburg	160	144.807	144.806
	Number BCs Mecklenburg-Vorpommern	80	108.627	108.624
	Number BCs Saxony	231	229.626	229.625
	Number BCs Saxony-Anhalt	183	160.895	160.925

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number BCs Thuringia			
Number 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)	1.014	1.249.634	1.249.634
	Number BCs with 2 individual under 65 (west)	597	444.020	444.012
	Number BCs with 3 individual under 65 (west)	377	272.340	272.333
	Number BCs with 4 individual under 65 (west)	201	166.625	166.619
	Number BCs with 5 or more individuals under 65 (west)	182	119.489	119.484
	Number BCs with 1 individual under 65 (east)	522	641.158	641.158
	Number BCs with 2 individual under 65 (east)	264	212.868	212.899
	Number BCs with 3 individual under 65 (east)	151	109.843	109.838
	Number BCs with 4 individual under 65 (east)	47	60.247	60.247
	Number BCs with 5 or more individuals under 65 (east)	48	36.657	36.656
Number BCs receiving benefits in accordance with SGB II by number of children under 15	Number BCs without children under 15 years (west)	1.619	1.545.685	1.545.666
	Number BCs with 1 child	429	379.037	379.033

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
years of age in the benefit unit (1, 2, 3, "4 or more") and by west/ east (10 categories)	under 15 years (west)			
	Number BCs with 2 children under 15 years (west)	212	217.527	217.525
	Number BCs with 3 children under 15 years (west)	89	79.100	79.099
	Number BCs with 4 or more children under 15 years (west)	22	30.759	30.759
	Number BCs without children under 15 years (east)	792	779.087	779.113
	Number BCs with 1 child under 15 years (east)	140	160.802	160.802
	Number BCs with 2 children under 15 years (east)	72	83.895	83.894
	Number BCs with 3 children under 15 years (east)	23	26.312	26.311
	Number BCs with 4 or more children under 15 years (east)	5	10.677	10.677
Number BCs receiving benefits in accordance with SGB II constisting of a single parent with children by west/east (4 categories)	Number BCs with a single parent (west)	512	448.398	448.395
	Rest BCs without a single parent (west)	1.859	1.803.710	1.803.687
	Number BCs with a single parent (east)	162	183.565	183.565

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Rest BCs without a single parent (east)	870	877.208	877.233
Number of households by federal state and BIK type (spelling: “Federal state. BIK type”; 38 categories)	1.1 to 1.6	70	476.000	476.000
	1.7 to 1.10	203	911.000	911.000
	2.10	181	966.000	966.000
	3.1 to 3.5	248	1.374.000	1.374.000
	3.7 to 3.8	390	1.381.000	1.381.000
	3.9 to 3.10	237	1.023.000	1.023.000
	4.8 to 4.10	70	359.000	359.000
	5.2 to 5.4	293	1.129.000	1.129.000
	5.5 to 5.6	259	942.000	942.000
	5.7 to 5.8	698	2.917.000	2.917.000
	5.9 to 5.10	830	3.520.000	3.520.000
	6.1 to 6.4	140	647.000	647.000
	6.5 to 6.8	189	1.065.000	1.065.000
	6.9 to 6.10	184	1.186.000	1.186.000
	7.1 to 7.6	150	928.000	928.000
	7.7 to 7.10	202	952.000	952.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	8.1 to 8.4	207	1.226.000	1.226.000
	8.5 to 8.8	305	2.140.000	2.140.000
	8.9 to 8.10	285	1.612.000	1.612.000
	9.1 to 9.4	258	1.571.000	1.571.000
	9.5 to 9.7	283	1.473.000	1.473.000
	9.8 to 9.9	281	1.415.000	1.415.000
	9.10	287	1.593.000	1.593.000
	10.3 to 10.8	119	490.000	490.000
	11.10	516	1.949.000	1.949.000
	12.1 to 12.4	151	459.000	459.000
	12.5 to 12.7	134	261.000	261.000
	12.9 to 12.10	119	509.000	509.000
	13.1 to 13.6	124	514.000	514.000
	13.7 to 13.9	84	308.000	308.000
	14.1 to 14.4	219	569.000	569.000
	14.5 to 14.8	144	640.000	640.000
	14.9 to 14.10	250	923.000	923.000
	15.1 to 15.4	176	393.000	393.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	15.5 to 15.7	135	485.000	485.000
	15.8 to 15.9	126	274.000	274.000
	16.1 to 16.4	223	529.000	529.000
	16.5 to 16.8	151	563.000	563.000
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)	2.294	12.361.000	12.361.000
	Number households with 2 individuals (west)	2.085	10.706.000	10.706.000
	Number households with 3 individuals (west)	962	3.943.000	3.943.000
	Number households with 4 individuals (west)	662	3.119.000	3.119.000
	Number households with 5 or more individuals (west)	366	1.167.000	1.167.000
	Number households with 1 individual (east)	1.081	3.636.000	3.636.000
	Number households with 2 individuals (east)	840	3.026.000	3.026.000
	Number households with 3 individuals (east)	360	1.007.000	1.007.000
	Number households with 4 individuals (east)	179	545.000	545.000
	Number households with	92	162.000	162000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	5 or more individuals(east)			
Number of households by “children under 15 years of age in the household "yes/no” and west/east (4 categories)	Number households with children under 15 years (west)	1.621	5.481.000	5.481.000
	Number households without children under 15 years (west)	4.748	25.815.000	25.815.000
	Number households with children under 15 years (east)	529	1.319.000	1.319.000
	Number households without children under 15 years (east)	2.023	7.057.000	7.057.000

Table 51: Parameters of distribution of weights (Total sample, households)

1%-percentile	127,5883
5%-percentile	169,69
10%-percentile	200,6964
25%-percentile	338,3151
50%-percentile	1009,789
75%-percentile	5276,326
90%-percentile	15954,85
95%-percentile	22693,44
99%-percentile	25761,41
Mean	4447,035
Standard deviation	6824,17
Minimum	54,8359
Maximum	30682,95
Number of observations	8.921
Effiency measure	**29,8%**

6.13 Calibration of the person weight, wave 9, cross-section

As in 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. These input weights were calibrated at the individual level.

As in the previous year, the increase in UB II recipients since the previous year at the level of individuals between 15 and 64 years (350.298) was also included as an additional benchmark value in the total sample. Again, those cases in the previous samples from waves 1 to 7 of the survey who were receiving UB II in July 2014 are projected to the benchmark statistics of the Federal Employment Agency on receipt of UB II.

Before calibration, the calibrated household weights that formed the input weight were also trimmed. For the calibration of person weights, the range of weights was determined to a certain interval.

6.14 BA sample

The population of the cumulated BA sample of all nine 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 according to SGB II at one of the (until now) seven drawing dates (in July 2006, July 2007, July 2008, July 2009, July 2010, July 2011, July 2012, July 2013 or July 2014). Only those individuals aged 15 and over who were living in a benefit unit that received benefits according to SGB II in July 2014 were considered for calibration. Individuals living in a household that did not receive benefits and individuals living in a household with at least one benefit unit according to SGB II but who were not 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 fifth and ninety-fifth percentiles of their distribution and then re-scaled so that they totaled the untrimmed calibrated household weights. The trimmed projection factors range from 362,45 to 9388,76. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.11 and upwards to 3.0. Thus, the total projection factors after calibration lie between a minimum of 53,46 and a maximum of 9345,04.

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 according to SGB II, by federal states
- Number of individuals in benefit units receiving benefits according to SGB II, by age (15-24 and 25-64)
- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II by sex and by west/east
- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by “single parent yes/no” and by west/east
- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by nationality (German/non-German)

As in the previous year, the increase in UB II recipients since the previous year at the level of individuals between 15 and 64 years (350.298) 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. Because 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 nonresponse item thus leads to slight deviations from the values as presented below.

Table 52: Nominal distributions and distributions after calibration (BA sample, individuals)

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by federal state (16 categories	Number individuals in BCs Schleswig-Holstein	131	159.913	159.913
	Number individuals in BCs Hamburg	90	133.409	133.409
	Number individuals in BCs Lower Saxony	399	419.975	419.975
	Number individuals in BCs Bremen	43	68.591	68.591
	Number individuals in BCs North Rhine - Westphalia	1.106	1.183.508	1.183.508
	Number individuals in BCs Hesse	247	297.701	297.701
	Number individuals in BCs Rhineland-Palatinate	189	159.681	159.681
	Number individuals in BCs Baden-Wuerttemberg	289	310.103	310.103
	Number individuals in BCs Bavaria	362	305.503	305.503
	Number individuals in BCs	72	57.765	57.765

Table 52: Nominal distributions and distributions after calibration (BA sample, individuals) (continued)

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Saarland			
	Number individuals in BCs Berlin	283	420.240	420.240
	Number individuals in BCs Brandenburg	192	186.660	186.660
	Number individuals in BCs Mecklenburg-Vorpommern	92	140.932	140.932
	Number individuals in BCs Saxony	290	295.680	295.680
	Number individuals in BCs Saxony-Anhalt	225	209.745	209.745
	Number individuals in BCs Thuringia	156	132.800	132.800
Number of individuals in benefit untits receiving benefits in accordance with SGB II by age (15-24 and 25-64; 2 categories)	Number individuals in BCs aged 15-24	584	780.143	780.143
	Number individuals in BCs aged 25-64	3.582	3.702.063	3.702.063
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)	1.348	1.476.850	1.476.850
	Number women in BCs (west)	1.580	1.619.299	1.619.299
	Number men in BCs (east)	630	693.930	693.930

Table 52: Nominal distributions and distributions after calibration (BA sample, individuals) (continued)

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number women in BCs (east)	608	692.127	692.127
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by “single parent yes/no” and west/east (4 categories)	Number non single parents in BCs (west)	2.432	2.647.751	2.647.751
	Number single parents in BCs (west)	496	448.398	448.398
	Number non single parents in BCs(east)	1.078	1.202.492	1.202.492
	Number single parents in BCs(east)	160	183.565	183.565
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by nationality (german/non-german; 2 categories)	Number non-german individuals in BCs	740	1.041.091	1.041.091
	Number german individuals in BCs	3.426	3.441.115	3.441.115

Table 53: Parameters of distribution of weights (BA-sample, individuals)

1%-percentile	84,10716
5%-percentile	115,5764
10%-percentile	149,0444
25%-percentile	259,9994
50%-percentile	521,3685
75%-percentile	1342,937
90%-percentile	2982,618
95%-percentile	3850,562
99%-percentile	6785,776
Mean	1076,16
Standard deviation	1324,547
Minimum	53,45628
Maximum	9345,035
Number of observations	4.165
Efficiency measure	**39,8%**

6.15 Population sample

All individuals over 14 years of age in private households in Germany form the basic population. The starting points for the calibration were calibrated household weights of the population sample. These weights were trimmed at the fifth and ninety-fifth percentiles of their distribution and after that rescaled so that they totaled the untrimmed calibrated household weights. The trimmed projection factors lie between a minimum of 3558,7 to a maximum of 44838,5. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.2 and upwards to 5.0. Thus, the total projection factors after calibration lie between a minimum of 711,7 and a maximum of 162379,4.

A calibration was made for the following characteristics:

1. Benefit recipients basis BA statistics:
    - Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by federal states
    - Number of individuals in benefit communities receiving benefits according to SGB II, by age (15-24 and 25-64)

- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II by sex and by west/east
- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by “single parent yes/no” and by west/east
- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by nationality (German/non-German)

2. Population based on Mikrozensus 2014:

   - 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 region
   - Number of individuals aged 15 and over in private households, by household size and west/east region
   - Number of individuals aged 15 and over in private households, by academic qualifica-tions and west/east region
   - Number of individuals aged 15 and over in private households, by marital status and west/east region
   - Number of individuals aged 15 and over in private households, by nationality

3. Population based on BA statistics:

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

The source for the benchmark value of employment status was the BA statistics because the definition of unemployment and employment subject to social insurance in PASS does not correspond to the ILO.

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

Because the imputation only serves the feasibility of the calibration, the imputed values were set to missing values after the calibration. A projection with the calibrated weights without considering the nonresponse item therefore leads to slight deviations from the values as presented below.

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
Number of individuals in benefit untits receiving benefits in accordance with SGB II by west/east (2 categories)	Number individuals in BCs west	135	3.096.149	3.096.149
	Number individuals in BCs east	54	1.386.057	1.386.057
Number of individuals in benefit untits receiving benefits in accordance with SGB II by age (15-24 and 25-64; 2 categories)	Number individuals in BCs aged 15-24	31	780.143	780.143
	Number individuals in BCs aged 25-64	156	3.702.063	3.702.063
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by sex (2 categories)	Number men in BCs	85	2.170.780	2.170.780
	Number women in BCs	102	2.311.426	2.311.426
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by “single parent yes/no (2 categories)	Number non single parents in BCs	159	3.850.243	3.850.243
	Number single parents in BCs	28	631.963	631.963
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by	Number non-german individuals in BCs	19	1.041.091	1.041.091
	Number german individuals in BCs	168	3.441.115	3.441.115

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
nationality (german/ /non-german; 2 categories)				
Number of individuals aged 15 and over in private households (PH) by federal state (16 categories)	Number individuals in private households Schleswig-Holstein	132	2.409.000	2.409.000
	Number individuals in private households Hamburg	54	1.515.000	1.515.000
	Number individuals in private households Lower Saxony	488	6.669.000	6.669.000
	Number individuals in private households Bremem	27	575.000	575.000
	Number individuals in private households North Rhine-Westphalia	1.003	15.082.000	15.082.000
	Number individuals in private households Hesse	377	5.192.000	5.192.000
	Number individuals in private households Rhineland-Palatinate	227	3.433.000	3.433.000
	Number individuals in private households Baden-Wuerttemberg	534	9.089.000	9.089.000
	Number individuals in private households Bavaria	864	10.808.000	10.808.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number individuals in private households Saarland	69	863.000	863.000
	Number individuals in private households Berlin	124	2.974.000	2.974.000
	Number individuals in private households Brandenburg	147	2.122.000	2.122.000
	Number individuals in private households Mecklenburg-Vorpommern	76	1.390.000	1.390.000
	Number individuals in private households Saxony	236	3.504.000	3.504.000
	Number individuals in private households Saxony-Anhalt	191	1.940.000	1.940.000
	Number individuals in private households Thuringia	205	1.873.000	1.873.000
Number of individuals aged 15 and over in private households (PH) by age (in 5-year classes) gender and west/ east (56 categories)	Number men in PH (west) 15-19 years	127	1.761.000	1.761.000
	Number men in PH (west) 20-24 years	110	1.945.000	1.945.000
	Number men in PH (west) 25-29 years	106	1.973.000	1.973.000
	Number men in PH (west) 30-34 years	73	1.994.000	1.994.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number men in PH (west) 35-39 years	65	1.936.000	1.936.000
	Number men in PH (west) 40-44 years	103	2.227.000	2.227.000
	Number men in PH (west) 45-49 years	164	2.728.000	2.728.000
	Number men in PH (west) 50-54 years	194	2.606.000	2.606.000
	Number men in PH (west) 55-59 years	179	2.210.000	2.210.000
	Number men in PH (west) 60-64 years	160	1.929.000	1.929.000
	Number men in PH (west) 65-69 years	152	1.548.000	1.548.000
	Number men in PH (west) 70-74 years	153	1.675.000	1.675.000
	Number men in PH (west) 75-79 years	142	1.412.000	1.412.000
	Number men in PH (west) 80+ years	77	1.205.000	1.205.000
	Number women in PH (west) 15-19 years	120	1.665.000	1.665.000
	Number women in PH (west) 20-24 years	129	1.814.000	1.814.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number women in PH (west) 25-29 years	71	1.944.000	1.944.000
	Number women in PH (west) 30-34 years	65	2.023.000	2.023.000
	Number women in PH (west) 35-39 years	102	1.927.000	1.927.000
	Number women in PH (west) 40-44 years	120	2.216.000	2.216.000
	Number women in PH (west) 45-49 years	203	2.669.000	2.669.000
	Number women in PH (west) 50-54 years	246	2.608.000	2.608.000
	Number women in PH (west) 55-59 years	207	2.273.000	2.273.000
	Number women in PH (west) 60-64 years	194	2.035.000	2.035.000
	Number women in PH (west) 65-69 years	161	1.667.000	1.667.000
	Number women in PH (west) 70-74 years	163	1.888.000	1.888.000
	Number women in PH (west) 75-79 years	122	1.769.000	1.769.000
	Number women in PH (west) 80+ years	67	1.988.000	1.988.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number men in PH (east) 15-19 years	19	296.000	296.000
	Number men in PH (east) 20-24 years	23	364.000	364.000
	Number men in PH (east) 25-29 years	30	545.000	545.000
	Number men in PH (east) 30-34 years	24	534.000	534.000
	Number men in PH (east) 35-39 years	20	471.000	471.000
	Number men in PH (east) 40-44 years	27	535.000	535.000
	Number men in PH (east) 45-49 years	33	631.000	631.000
	Number men in PH (east) 50-54 years	58	662.000	662.000
	Number men in PH (east) 55-59 years	63	607.000	607.000
	Number men in PH (east) 60-64 years	42	551.000	551.000
	Number men in PH (east) 65-69 years	40	375.000	375.000
	Number men in PH (east) 70-74 years	42	490.000	490.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number men in PH (east) 75-79 years	32	381.000	381.000
	Number men in PH (east) 80+ years	22	303.000	303.000
	Number women in PH (east) 15-19 years	15	264.000	264.000
	Number women in PH (east) 20-24 years	22	353.000	353.000
	Number women in PH (east) 25-29 years	22	510.000	510.000
	Number women in PH (east) 30-34 years	27	493.000	493.000
	Number women in PH (east) 35-39 years	15	453.000	453.000
	Number women in PH (east) 40-44 years	32	477.000	477.000
	Number women in PH (east) 45-49 years	35	594.000	594.000
	Number women in PH (east) 50-54 years	54	678.000	678.000
	Number women in PH (east) 55-59 years	62	610.000	610.000
	Number women in PH (east) 60-64 years	66	594.000	594.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number women in PH (east) 65-69 years	45	417.000	417.000
	Number women in PH (east) 70-74 years	50	574.000	574.000
	Number women in PH (east) 75-79 years	42	509.000	509.000
	Number women in PH (east) 80+ years	17	532.000	532.000
Number of individuals aged 15 and over in private households (PH) by household size (1,2, 3,4,"5 or more individuals") and west/east (10 categories)	Number individuals in PH 1 individual (west)	625	12.361.000	12.361.000
	Number individuals in PH 2 individuals (west)	1.555	20.948.000	20.948.000
	Number individuals in PH 3 individuals (west)	649	9.717.000	9.717.000
	Number individuals in PH 4 individuals (west)	643	8.695.000	8.695.000
	Number individuals in PH 5 or more individuals (west)	303	3.914.000	3.914.000
	Number individuals in PH 1 individual (east)	200	3.636.000	3.636.000
	Number individuals in PH 2 individuals (east)	474	5.858.000	5.858.000
	Number individuals in PH 3 individuals (east)	153	2.412.000	2.412.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number individuals in PH 4 individuals (east)	93	1.409.000	1.409.000
	Number individuals in PH 5 or more individuals (east)	59	488.000	488.000
Number of individuals aged 15 and over in private households (PH) by highest school qualification and west/east (12 categories)	Number individuals in PH with highest school qualification: still pupil (west)	140	2.097.000	2.097.000
	Number individuals in PH with highest school qualification: no qualification (west)	100	2.090.000	2.090.000
	Number individuals in PH with highest school qualification: lower secondary school (west)	1.166	20.743.000	20.743.000
	Number individuals in PH with highest school qualification: intermediate secondary school; intermediate secondary school in the former GDR (west)	1.065	14.291.000	14.291.000
	Number individuals in PH with highest school qualification: university (of applied sciences) qualification (west)	1.304	16.414.000	16.414.000
	Number individuals in PH with highest school quali-	26	377.000	377.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	fication: still pupil (east)			
	Number individuals in PH with highest school qualification: no qualification (east)	16	303.000	303.000
	Number individuals in PH with highest school qualification:lower secondary school	219	2.588.000	2.588.000
	Number individuals in PH with highest school qualification:intermediate secondary school; intermediate secondary school in the former GDR (east)	427	6.741.000	6.741.000
	Number individuals in PH with highest school qualification:university (of applied sciences) qualification (east)	291	3.794.000	3.794.000
Number of individuals aged 15 and over in private households (PH) by marital status and west/east (8 categories)	Number individuals in PH with marital status: single (west)	908	10.853.000	10.853.000
	Number individuals in PH with marital status: married, civil partnership (west)	2.356	34.868.000	34.868.000
	Number individuals in PH with marital status: divorced (west)	284	5.219.000	5.219.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number individuals in PH with marital status: widowed (west)	227	4.695.000	4.695.000
	Number individuals in PH with marital status: single (east)	210	3.557.000	3.557.000
	Number individuals in PH with marital status: married, civil partnership (east)	599	7.602.000	7.602.000
	Number individuals in PH with marital status: divorced (east)	86	1.405.000	1.405.000
	Number individuals in PH with marital status: widowed (east)	84	1.239.000	1.239.000
Number of individuals aged 15 and over in private households (PH) by nationality (2 categories)	Number individuals in PH non-germans	131	6.513.000	6.513.000
	Number individuals in PH german	4.623	62.925.000	62.925.000
Unemployed individuals incl. participants in measures west/east (4 categories)	Not unemployed (west)	3.645	53.020.38	53.020.38
	Unemployed individuals incl. participants in measures (west)	130	2.614.618	2.614.618
	not unemployed (east)	934	12.788.33	12.788.330
	Unemployed individuals incl. participants in measures (east)	45	1.014.670	1.014.670

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
Employees subject to social security contributions west/east (2 categories)	Employees not subject to security contributions (west)	30.774.94	30.774.943	
	Employees subject to security contributions (west)	1.501	24.860.05	24.860.057
	Employees not subject to security contributions (east)	547	8.095.163	8.095.163
	Employees subject to security contributions (east)	432	5.707.837	5.707.837

Table 55: Parameters of distribution of weights (Population sample, individuals)

1%-percentile	1166,853
5%-percentile	2164,635
10%-percentile	2900,455
25%-percentile	4648,73
50%-percentile	8821,889
75%-percentile	17431,82
90%-percentile	33629,39
95%-percentile	48278,53
99%-percentile	87571,02
Mean	16950,57
Standard deviation	16935,81
Minimum	711,7394
Maximum	162178,2
Number of observations	4.754
Efficiency measure	**42,6%**

6.16 Total sample

All individuals aged 15 and over in private households in Germany form the population. The starting point for the calibration was the calibrated household weight of the total sample. That weight was trimmed at the fifth and ninety-fifth percentiles of their distribution and then rescaled so that they totaled the untrimmed calibrated household weights. The trimmed projection factors range from 195,02 to 26785,14. 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 a minimum of 19,5 and a maximum of 107140,6.

A calibration was made for the following characteristics:

1. Benefit recipients basis BA statistics:
    - Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by federal states
    - Number of individuals in benefit units receiving benefits according to SGB II, by age (15-24 and 25-64)
    - Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by sex and by west/east

- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by “single parent yes/no” and by west/east
- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by nationality (German/non-German)

2. Population based on Mikrozensus 2014:
    - 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 academic qualifications 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
3. Population based on 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 for the benchmark value of employment status was the BA statistics because the definition of unemployment and employment subject to social insurance in PASS does not correspond to the ILO concept.

In addition, the increase in UB II recipients since the previous year at the level of individuals between 15 and 64 years of age (350.298) 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.

Because the imputation is only required for 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 therefore leads to slight deviations from the values, as presented below.

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by federal state (16 categories	Number individuals in BCs Schleswig-Holstein	136	159.913	159.913
	Number individuals in BCs Hamburg	93	133.409	133.409
	Number individuals in BCs Lower Saxony	424	419.975	419.975
	Number individuals in BCs Bremen	44	68.591	68.591
	Number individuals in BCs North Rhine - Westphalia	1.170	1.183.508	1.183.508
	Number individuals in BCs Hesse	247	297.701	297.701
	Number individuals in BCs Rhineland-Palatinate	196	159.681	159.681
	Number individuals in BCs Baden-Wuerttemberg	292	310.103	310.103
	Number individuals in BCs Bavaria	382	305.503	305.503
	Number individuals in BCs Saarland	75	57.765	57.765
	Number individuals in BCs Berlin	289	420.240	420.240
	Number individuals in BCs Brandenburg	204	186.660	186.660

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number individuals in BCs Mecklenburg-Vorpommern	97	140.932	140.932
	Number individuals in BCs Saxony	297	295.680	295.680
	Number individuals in BCs Saxony-Anhalt	239	209.745	209.745
	Number individuals in BCs Thuringia	168	132.800	132.800
Number of individuals in benefit untits receiving benefits in accordance with SGB II by age (15-24 and 25-64; 2 categories)	Number individuals in BCs aged 15-24	615	780.143	780.143
	Number individuals in BCs aged 25-64	3.738	3.702.063	3.702.063
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)	1.408	1.476.850	1.476.850
	Number women in BCs (west)	1.651	1.619.299	1.619.299
	Number men in BCs (east)	655	693.930	693.930
	Number women in BCs (east)	639	692.127	692.127
Number of individuals aged 15 and over in benefit units receiving benefits in	Number non single parents in BCs (west)	2.543	2.647.751	2.647.751
	Number single parents in	516	448.398	448.398

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
accordance with SGB II by "single parent yes/ no" and west/east (4 categories)	BCs (west)			
	Number non single parents in BCs(east)	1.126	1.202.492	1.202.492
	Number single parents in BCs(east)	168	183.565	183.565
Number of individuals aged 15 and over in benefit units receiving benefits in accordance with SGB II by nationality (german/non-german; 2 categories)	Number non-german individuals in BCs	759	1.041.091	1.041.091
	Number german individuals in BCs	3.594	3.441.115	3.441.115
Number of individuals aged 15 and over in private households (PH) by federal state (16 categories)	Number individuals in private households Schleswig-Holstein	391	2.409.000	2.409.000
	Number individuals in private households Hamburg	242	1.515.000	1.515.000
	Number individuals in private households Lower Saxony	1.281	6.669.000	6.669.000
	Number individuals in private households Bremem	98	575.000	575.000
	Number individuals in private households North Rhine-Westphalia	3.101	15.082.000	15.082.000
	Number individuals in pri-	792	5.192.000	5.192.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	vate households Hesse			
	Number individuals in private households Rhineland-Palatinate	578	3.433.000	3.433.000
	Number individuals in private households Baden-Wuerttemberg	1.194	9.089.000	9.089.000
	Number individuals in private households Bavaria	1.728	10.808.000	10.808.000
	Number individuals in private households Saarland	192	863.000	863.000
	Number individuals in private households Berlin	661	2.974.000	2.974.000
	Number individuals in private households Brandenburg	585	2.122.000	2.122.000
	Number individuals in private households Mecklenburg-Vorpommern	293	1.390.000	1.390.000
	Number individuals in private households Saxony	919	3.504.000	3.504.000
	Number individuals in private households Saxony-Anhalt	650	1.940.000	1.940.000
	Number individuals in private households Thuringia	566	1.873.000	1.873.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
Number of individuals aged 15 and over in private households (PH) by age (in 5-year classes) gender and west/ east (56 categories)	Number men in PH (west) 15-19 years	324	1.761.000	1.761.000
	Number men in PH (west) 20-24 years	290	1.945.000	1.945.000
	Number men in PH (west) 25-29 years	378	1.973.000	1.973.000
	Number men in PH (west) 30-34 years	385	1.994.000	1.994.000
	Number men in PH (west) 35-39 years	296	1.936.000	1.936.000
	Number men in PH (west) 40-44 years	337	2.227.000	2.227.000
	Number men in PH (west) 45-49 years	405	2.728.000	2.728.000
	Number men in PH (west) 50-54 years	494	2.606.000	2.606.000
	Number men in PH (west) 55-59 years	462	2.210.000	2.210.000
	Number men in PH (west) 60-64 years	393	1.929.000	1.929.000
	Number men in PH (west) 65-69 years	303	1.548.000	1.548.000
	Number men in PH (west) 70-74 years	204	1.675.000	1.675.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number men in PH (west) 75-79 years	154	1.412.000	1.412.000
	Number men in PH (west) 80+ years	77	1.205.000	1.205.000
	Number women in PH (west) 15-19 years	333	1.665.000	1.665.000
	Number women in PH (west) 20-24 years	377	1.814.000	1.814.000
	Number women in PH (west) 25-29 years	386	1.944.000	1.944.000
	Number women in PH (west) 30-34 years	429	2.023.000	2.023.000
	Number women in PH (west) 35-39 years	410	1.927.000	1.927.000
	Number women in PH (west) 40-44 years	419	2.216.000	2.216.000
	Number women in PH (west) 45-49 years	499	2.669.000	2.669.000
	Number women in PH (west) 50-54 years	588	2.608.000	2.608.000
	Number women in PH (west) 55-59 years	510	2.273.000	2.273.000
	Number women in PH (west) 60-64 years	448	2.035.000	2.035.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number women in PH (west) 65-69 years	286	1.667.000	1.667.000
	Number women in PH (west) 70-74 years	198	1.888.000	1.888.000
	Number women in PH (west) 75-79 years	136	1.769.000	1.769.000
	Number women in PH (west) 80+ years	76	1.988.000	1.988.000
	Number men in PH (east) 15-19 years	97	296.000	296.000
	Number men in PH (east) 20-24 years	95	364.000	364.000
	Number men in PH (east) 25-29 years	178	545.000	545.000
	Number men in PH (east) 30-34 years	174	534.000	534.000
	Number men in PH (east) 35-39 years	149	471.000	471.000
	Number men in PH (east) 40-44 years	103	535.000	535.000
	Number men in PH (east) 45-49 years	133	631.000	631.000
	Number men in PH (east) 50-54 years	220	662.000	662.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number men in PH (east) 55-59 years	207	607.000	607.000
	Number men in PH (east) 60-64 years	210	551.000	551.000
	Number men in PH (east) 65-69 years	122	375.000	375.000
	Number men in PH (east) 70-74 years	58	490.000	490.000
	Number men in PH (east) 75-79 years	40	381.000	381.000
	Number men in PH (east) 80+ years	23	303.000	303.000
	Number women in PH (east) 15-19 years	76	264.000	264.000
	Number women in PH (east) 20-24 years	97	353.000	353.000
	Number women in PH (east) 25-29 years	168	510.000	510.000
	Number women in PH (east) 30-34 years	177	493.000	493.000
	Number women in PH (east) 35-39 years	128	453.000	453.000
	Number women in PH (east) 40-44 years	124	477.000	477.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number women in PH (east) 45-49 years	171	594.000	594.000
	Number women in PH (east) 50-54 years	237	678.000	678.000
	Number women in PH (east) 55-59 years	238	610.000	610.000
	Number women in PH (east) 60-64 years	200	594.000	594.000
	Number women in PH (east) 65-69 years	109	417.000	417.000
	Number women in PH (east) 70-74 years	71	574.000	574.000
	Number women in PH (east) 75-79 years	47	509.000	509.000
	Number women in PH (east) 80+ years	22	532.000	532.000
Number of individuals aged 15 and over in private households (PH) by household size (1,2, 3,4,"5 or more individuals") and west/east (10 categories)	Number individuals in PH 1 individual (west)	2.287	12.361.000	12.361.000
	Number individuals in PH 2 individuals (west)	3.288	20.948.000	20.948.000
	Number individuals in PH 3 individuals (west)	1.785	9.717.000	9.717.000
	Number individuals in PH 4 individuals (west)	1.333	8.695.000	8.695.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number individuals in PH 5 or more individuals (west)	904	3.914.000	3.914.000
	Number individuals in PH 1 individual (east)	1.073	3.636.000	3.636.000
	Number individuals in PH 2 individuals (east)	1.359	5.858.000	5.858.000
	Number individuals in PH 3 individuals (east)	690	2.412.000	2.412.000
	Number individuals in PH 4 individuals (east)	335	1.409.000	1.409.000
	Number individuals in PH 5 or more individuals (east)	217	488.000	488.000
Number of individuals aged 15 and over in private households (PH) by highest school qualification and west/east (12 categories)	Number individuals in PH with highest school qualification: still pupil (west)	394	2.097.000	2.097.000
	Number individuals in PH with highest school qualification: no qualification (west)	486	2.090.000	2.090.000
	Number individuals in PH with highest school qualification: lower secondary school (west)	3.239	20.743.000	20.743.000
	Number individuals in PH with highest school qualification: intermediate secondary school; intermediate secondary school in the former GDR (west)	2.658	14.291.000	14.291.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
	Number individuals in PH with highest school qualification: university (of applied sciences) qualification (west)	2.820	16.414.000	16.414.000
	Number individuals in PH with highest school qualification: still pupil (east)	101	377.000	377.000
	Number individuals in PH with highest school qualification: no qualification (east)	100	303.000	303.000
	Number individuals in PH with highest school qualification:lower secondary school	859	2.588.000	2.588.000
	Number individuals in PH with highest school qualification:intermediate secondary school; intermediate secondary school in the former GDR (east)	1.741	6.741.000	6.741.000
	Number individuals in PH with highest school qualification:university (of applied sciences) qualification (east)	873	3.794.000	3.794.000
Number of individuals aged 15 and over in pri-	Number individuals in PH with marital status:	3.158	10.853.000	10.853.000

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
vate households (PH) by marital status and west/east (8 categories)	single (west)			
	Number individuals in PH with marital status: married, civil partnership (west)	4.622	34.868.000	34.868.000
	Number individuals in PH with marital status: divorced (west)	1.423	5.219.000	5.219.000
	Number individuals in PH with marital status: widowed (west)	394	4.695.000	4.695.000
	Number individuals in PH with marital status: single (east)	1.383	3.557.000	3.557.000
	Number individuals in PH with marital status: married, civil partnership (east)	1.542	7.602.000	7.602.000
	Number individuals in PH with marital status: divorced (east)	586	1.405.000	1.405.000
	Number individuals in PH with marital status: widowed (east)	163	1.239.000	1.239.000
Number of individuals aged 15 and over in private households (PH) by nationality (2 categories)	Number individuals in PH non-germans	1.215	6.513.000	6.513.000
	Number individuals in PH german	12.056	62.925.000	62.925.000
Unemployed individ-	Not unemployed (west)	7.844	53.020.382	53.020.382

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

Benchmark Figure	Characteristics benchmark figure from BA statistics	Unweighted distribution	Nominal values from BA-statistics	Distribution with calibrated weights
uals incl. participants in measures west/east (4 categories)	Unemployed individuals incl. participants in measures (west)	1.753	2.614.618	2.614.618
	not unemployed (east)	2.874	12.788.330	12.788.330
	Unemployed individuals incl. participants in measures (east)	800	1.014.670	1.014.670
Employees subject to social security contributions west/east (2 categories)	Employees not subject to security contributions (west)	6.008	30.774.943	30.774.943
	Employees subject to security contributions (west)	3.589	24.860.057	24.860.057
	Employees not subject to security contributions (east)	2.119	8.095.163	8.095.163
	Employees subject to security contributions (east)	1.555	5.707.837	5.707.837

Table 57: Parameters of distribution of weights (Total sample, individuals)

1%-percentile	26,02415
5%-percentile	88,80314
10%-percentile	160,281
25%-percentile	349,0301
50%-percentile	1223,995
75%-percentile	5797,29
90%-percentile	16213,96
95%-percentile	24464,85
99%-percentile	43148,79
Mean	5232,311
Standard deviation	9117,203
Minimum	19,50215
Maximum	107140,6
Number of observations	13.271
Efficiency measure	**24,8%**

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

Finally, in wave 9, 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 larger in wave 9 than in wave 8 because a larger part of the BA sample of waves 1 to 8 has withdrawn from benefits. These are the following three groups that were not receiving benefits in July 2014 but that belong to the population of the BA sample (households receiving UB II in July 2006, July 2007, July 2008, July 2009, July 2010, July 2011 or July 2012 and individuals in households receiving UB II in July 2006, July 2007, July 2008, July 2009, July 2010, July 2011, July 2012, July 2013 or July 2014).

- 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 9 after calibration (*wqbahh*) by dividing it by the proportion of these individuals who gave a personal or senior citizen interview - provided that their household was participating.
- Panel households in which nobody continued to receive UB II in July 2014: The house-hold retains the BA weight before calibration. Individuals in households with

interviews in both waves were assigned a new BA person weight, which is obtained by multiplying their old BA person weight by the reciprocal re-participation probability *ppbleib*. Individuals in these households who did not provide a personal interview in wave 8 are as-signed a new BA person weight calculated by dividing the BA household weight of their household for wave 9 by the proportion of such individuals who participate if their household is taking part.

- Individuals who are not members of a benefit unit in panel households that continued to receive UB II in July 2014: Individuals in these households with interviews in both waves were assigned 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*.
- The individuals and households were also adjusted to a benchmark figure for the individuals or benefit units that did not continue to receive UB II. The exact population of this group is unknown but can be approximated from the total of all cumulated BA subsamples minus the individuals or benefit units currently receiving benefits. The number of individuals who are no longer receiving UB II at wave 9 is 5.130.027. The number of benefit units that are no longer receiving UB II is 3.760.520.

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; 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; childcare and school attendance of children; 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 January 2005 (from wave 2 onwards); limited-term 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; **Characteristics on receiving benefits:** Unemployment Benefit I: start and end dates of the spell(s) of benefit receipt since January 2005 (wave 1 only); information on periods of Unemployment Benefit I receipt in the context of registered unemployment since January 2005 (from wave 2 onwards); amount of benefit; reason for end;

Categories	Comments
	Unemployment Benefit II: start and end dates of the spell(s) of benefit receipt since January 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); Contacts with Unemployment Benefit II institutions: number and type of contacts; contents of discussion; offers; integration agreement; assessment of institution; **Subjective indicators:** satisfaction; fears and problems; employment orientation; education aspiration; sex role orientation; subjective social position (top-bottom scale); subjective assessment of health state

Categories	Comments
Data Unit	Individuals and households receiving Unemployment Benefit II in July 2006 (sample I) Individuals and households in the resident population of Germany (sample II) Individuals and households receiving Unemployment Benefit II in July 2007 but without receipt in July 2006 (sample III; refreshment sample 1) Individuals and households receiving Unemployment Benefit II in July 2008 but without receipt in July 2006 or July 2007 (sample IV; refreshment sample 2) Individuals and households receiving Unemployment Benefit II in July 2009 but without receipt in July 2006, July 2007 or July 2008 (sample V; refreshment sample 3) Individuals and households receiving Unemployment Benefit II in July 2010 but without receipt in July 2006, July 2007, July 2008 or July 2009 (sample VI; refreshment sample 4) Individuals and households of the resident German population (sample VII, panel refreshment/replenishment sample) Individuals and households receiving UB II in July 2010 (sample VIII, panel refreshment/replenishment sample) Individuals and households receiving Unemployment Benefit II in July 2011 but without receipt in July 2006, July 2007, July 2008, July 2009 or July 2010 (sample IX; refreshment sample 5) Individuals and households receiving Unemployment Benefit II in July 2012 but without receipt in July 2006, July 2007, July 2008, July 2009, July 2010 or July 2011 (sample X; refreshment sample 6) Individuals and households receiving Unemployment Benefit II in July 2013 but without receipt in July 2006, July 2007, July 2008, July 2009, July 2010, July 2011 or July 2012 (sample XI; refreshment sample 7) Individuals and households receiving Unemployment Benefit II in July 2014 but without receipt in July 2006, July 2007, July 2008, July 2009, July 2010, July 2011, July 2012 or July 2013 (sample XII; refreshment sample 8) 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	(in 6.804 households)
	Sample II: 9.568 Individuals	(in 5.990 households)
	Wave 2:	
	Sample I: 4.753 Individuals	(in 3.491 households)
	Sample II: 6.392 Individuals	(in 3.897 households)
	Sample III: 1.342 Individuals	(in 1.041 households)
	Wave 3:	
	Sample I: 4.913 Individuals	(in 3.754 households)
	Sample II: 6.207 Individuals	(in 3.901 households)
	Sample III: 898 Individuals	(in 694 households)
	Sample IV: 1.421 Individuals	(in 1.186 households)
	Wave 4:	
	Sample I: 3.958 Individuals	(in 2.815 households)
	Sample II: 5.016 Individuals	(in 2.977 households)
	Sample III: 786 Individuals	(in 563 households)
	Sample IV: 983 Individuals	(in 745 households)
	Sample V: 1.025 Individuals	(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	(in 464 households)
	Sample IV: 822 Individuals	(in 608 households)
	Sample V: 760 Individuals	(in 517 households)
	Sample VI: 1.019 Individuals	(in 753 households)
	Sample VII: 2.589 Individuals	(in 1.510 households)
	Sample VIII: 1.859 Individuals	(in 1.321 households)
	Wave 6:	
	Sample I: 3.048 Individuals	(in 2.109 households)
	Sample II: 4.245 Individuals	(in 2.539 households)
	Sample III: 558 Individuals	(in 398 households)
	Sample IV: 719 Individuals	(in 532 households)
	Sample V: 679 Individuals	(in 466 households)
	Sample VI: 716 Individuals	(in 497 households)

Categories	Comments	
	Sample VII: 1.990 Individuals	(in 1.103 households)
	Sample VIII: 1.350 Individuals	(in 908 households)
	Sample IX: 1.314 Individuals	(in 961 households)
	Wave 7:	
	Sample I: 2.861 Individuals	(in 1.984 households)
	Sample II: 4.001 Individuals	(in 2.409 households)
	Sample III: 505 Individuals	(in 359 households)
	Sample IV: 688 Individuals	(in 505 households)
	Sample V: 590 Individuals	(in 414 households)
	Sample VI: 599 Individuals	(in 413 households)
	Sample VII: 1.784 Individuals	(in 996 households)
	Sample VIII: 1.182 Individuals	(in 798 households)
	Sample IX: 975 Individuals	(in 682 households)
	Sample X: 1.264 Individuals	(in 949 households)
	Wave 8:	
	Sample I: 2.447 Individuals	(in 1.738 households)
	Sample II: 3.591 Individuals	(in 2.194 households)
	Sample III: 450 Individuals	(in 324 households)
	Sample IV: 593 Individuals	(in 431 households)
	Sample V: 512 Individuals	(in 359 households)
	Sample VI: 502 Individuals	(in 348 households)
	Sample VII: 1.533 Individuals	(in 883 households)
	Sample VIII: 999 Individuals	(in 687 households)
	Sample IX: 821 Individuals	(in 571 households)
	Sample X: 932 Individuals	(in 677 households)
	Sample XI: 1.080 Individuals	(in 795 households)
	Wave 9:	
	Sample I: 2242 Individuals	(in 1586 households)
	Sample II: 3348 Individuals	(in 2063 households)
	Sample III: 402 Individuals	(in 290 households)
	Sample IX: 540 Individuals	(in 387 households)
	Sample V: 459 Individuals	(in 314 households)
	Sample VI: 449 Individuals	(in 313 households)
	Sample VII: 1406 Individuals	(in 806 households)
	Sample VIII: 912 Individuals	(in 617 households)
	Sample IX: 733 Individuals	(in 507 households)
	Sample X: 838 Individuals	(in 594 households)

Categories	Comments	
	Sample XI: 760 Individuals	(in 544 households)
	Sample XII: 1182 Individuals	(in 900 households)

Categories	Comments
Data collection mode	CATI and CAPI CAPI interviews were conducted when a sample household couldnot 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) Wave 6: n (CATI): 6.069 Individuals (4.058 households) n (CAPI): 8.550 Individuals (5.455 households) Wave 7: n (CATI): 5.779 Individuals (3.874 households) n (CAPI): 8.670 Individuals (5.635 households) Wave 8: n (CATI): 5.074 Individuals (3.454 households) n (CAPI): 8.386 Individuals (5.544 households) Wave 9: n (CATI): 4.416 Individuals (3.039 households) n (CAPI): 8.855 Individuals (5.882 households)

Categories	Comments
Interview languages	Wave 1: German: 18.205 Individuals (12.347 households) Russian: 432 Individuals (275 households) Turkish: 305 Individuals (163 households) Englisch: 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) Wave 6: German: 14.337 Individuals (9.342 households) Russian: 242 Individuals (146 households) Turkish: 40 Individuals (25 households) Wave 7: German: 14.161 Individuals (9.335 households) Russian: 245 Individuals (145 households) Turkish: 43 Individuals (29 households)

Categories	Comments	
	Wave 8:	
	German: 13.208 Individuals	(8.845 households)
	Russian: 224 Individuals	(131 households)
	Turkish: 28 Individuals	(22 households)
	Wave 9:	
	German: 13.057 Individuals	(8.796 households)
	Russian: 187 Individuals	(111 households)
	Turkish: 27 Individuals	(14 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 III (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	
	Wave 6:	
	Sample I (HHs agreeing to participate only):	73,3 %
	Sample II (HHs agreeing to participate only):	85,1 %
	Sample III (HHs agreeing to participate only):	70,2 %
	Sample IV (HHs agreeing to participate only):	69,9 %
	Sample V (HHs agreeing to participate only):	68,4 %
	Sample VI (HHs agreeing to participate only):	78,4 %
	Sample VII (HHs agreeing to participate only):	84,1 %
	Sample VIII (HHs agreeing to participate only):	77,1 %
	Sample IX:	30,8 %
	Total:	67,4 %
	Wave 7:	
	Sample I (HHs agreeing to participate only):	79,1 %
	Sample II (HHs agreeing to participate only):	86,8 %
	Sample III (HHs agreeing to participate only):	75,3 %
	Sample IV (HHs agreeing to participate only):	77,5 %
	Sample V (HHs agreeing to participate only):	76,4 %
	Sample VI (HHs agreeing to participate only):	66,6 %
	Sample VII (HHs agreeing to participate only):	79,3 %
	Sample VIII (HHs agreeing to participate only):	70,8 %
	Sample IX (HHs agreeing to participate only):	74,2 %
	Sample X:	32,1 %
	Total:	68,7 %
	Wave 8:	
	Sample I (HHs agreeing to participate only):	78,2 %
	Sample II (HHs agreeing to participate only):	84,7 %
	Sample III (HHs agreeing to participate only):	76,1 %
	Sample IV (HHs agreeing to participate only):	75,7 %
	Sample V (HHs agreeing to participate only):	77,0 %
	Sample VI (HHs agreeing to participate only):	71,0 %
	Sample VII (HHs agreeing to participate only):	81,8 %
	Sample VIII (HHs agreeing to participate only):	74,1 %
	Sample IX (HHs agreeing to participate only):	65,6 %
	Sample X (HHs agreeing to participate only):	74,0 %
	Sample XI:	25,6 %
	Total:	65,9 %
	Wave 9:	

Categories	Comments	
	Sample I (HHs agreeing to participate only):	71,3 %
	Sample II (HHs agreeing to participate only):	79,3 %
	Sample III (HHs agreeing to participate only):	68,1 %
	Sample IV (HHs agreeing to participate only):	68,0 %
	Sample V (HHs agreeing to participate only):	67,7 %
	Sample VI (HHs agreeing to participate only):	63,7 %
	Sample VII (HHs agreeing to participate only):	74,9 %
	Sample VIII (HHs agreeing to participate only):	66,9 %
	Sample IX (HHs agreeing to participate only):	58,3 %
	Sample X (HHs agreeing to participate only):	65,0 %
	Sample XI (HHs agreeing to participate only):	17,4 %
	Sample XII:	26,7 %
	Total:	52,2 %

Categories	Comments	
Response rates within households	Wave 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-III):	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 %

Categories	Comments	
	Total:	88,3 %
	Wave 6:	
	Sample I (re-interviewed households only):	89,3 %
	Sample I (re-interviewed households only):	88,6 %
	Sample III (re-interviewed households only):	88,5 %
	Sample IV (re-interviewed households only):	88,5 %
	Sample V (re-interviewed households only):	91,4 %
	Sample VI (re-interviewed households only):	92,0 %
	Sample VII (re-interviewed households only):	89,1 %
	Sample VIII (re-interviewed households only):	91,5 %
	Sample IX:	89,9 %
	Split-off households (from Samples I-VIII):	91,7 %
	Total:	89,5 %
	Wave 7:	
	Sample I (re-interviewed households only):	89,2 %
	Sample I (re-interviewed households only):	88,4 %
	Sample III (re-interviewed households only):	90,1 %
	Sample IV (re-interviewed households only):	88,8 %
	Sample V (re-interviewed households only):	89,8 %
	Sample VI (re-interviewed households only):	92,6 %
	Sample VII (re-interviewed households only):	89,1 %
	Sample VIII (re-interviewed households only):	92,0 %
	Sample IX (re-interviewed households only):	90,7 %
	Sample X:	90,1 %
	Split-off households (from Samples I-IX):	90,3 %
	Total:	89,5 %
	Wave 8:	
	Sample I (re-interviewed households only):	89,3 %
	Sample I (re-interviewed households only):	88,6 %
	Sample III (re-interviewed households only):	91,0 %
	Sample IV (re-interviewed households only):	88,3 %
	Sample V (re-interviewed households only):	90,5 %
	Sample VI (re-interviewed households only):	91,3 %
	Sample VII (re-interviewed households only):	89,0 %
	Sample VIII (re-interviewed households only):	93,3 %
	Sample IX (re-interviewed households only):	91,3 %
	Sample X: (re-interviewed households only):	91,5 %

Categories	Comments	
	Sampe XI:	90,0 %
	Split-off households (from Samples I-X):	90,0 %
	Total:	89,9 %
	Wave 9:	
	Sample I (re-interviewed households only):	88,9 %
	Sample I (re-interviewed households only):	88,0 %
	Sample III (re-interviewed households only):	89,6 %
	Sample IV (re-interviewed households only):	88,7 %
	Sample V (re-interviewed households only):	89,2 %
	Sample VI (re-interviewed households only):	90,2 %
	Sample VII (re-interviewed households only):	89,8 %
	Sample VIII (re-interviewed households only):	91,9 %
	Sample IX (re-interviewed households only):	91,4 %
	Sample X (re-interviewed households only):	92,0 %
	Sampe XI (re-interviewed households only):	91,3 %
	Sample XII:	87,9 %
	Split-off households (from Samples I-XI):	90,2 %
	Total:	89,4 %

Categories	Comments
Fieldwork period	Wave 1: December 2006-June 2007 Wave 2: December 2007-July 2008 Wave 3: December 2008-August 2009 Wave 4: Februar 2010-September 2010 Wave 5: February 2011-September 2011 Wave 6: February 2012-September 2012 Wave 7: February 2013-September 2013 Wave 8: February 2014-September 2014 Wave 9: February 2015-September 2015
Period	Wave 1: fieldwork period and retrospective spell data as of January 2005 Wave 2: Wave 2: fieldwork period and retrospective spell data as of January 2005 or the respective reference period of the spell type Wave 3: Wave 2: fieldwork period and retrospective spell data as of January 2006 or the respective reference period of the spell type Wave 4: Wave 2: fieldwork period and retrospective spell data as of January 2008 or the respective reference period of the spell type Wave 5: Wave 2: fieldwork period and retrospective spell data as of January 2009 or the respective reference period of the spell type Wave 6: Wave 2: fieldwork period and retrospective spell data as of January 2010 or the respective reference period of the spell type Wave 7: Wave 2: fieldwork period and retrospective spell data as of January 2011 or the respective reference period of the spell type

Categories	Comments
	Wave 8: Wave 2: fieldwork period and retrospective spell data as of January 2012 or the respective reference period of the spell type Wave 9: Wave 2: fieldwork period and retrospective spell data as of January 2013 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 is available on request.
Territorial allocation	On 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. The number of the gross sample drawn per PSU depended on the PSU size in terms of the relative proportion of bene-fit recipients within the respective postcode sector (probability proportional to size/pps). The average size of the gross sample was N=100 per post-code 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 using a stratification index contained in the database at a ratio of 4:2:1 for low-, medium- or high-status households, respectively. Interviewers from the surveying institute visited the selected buildings. In the event that a building accommodated several households, this fact 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. Refreshment sample I for sample I in wave 2: 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. Benefit units that received Unemployment Benefit II in July 2007 but not in July 2006 were selected, i.e., new recipients. The sample was drawn in the postcode areas selected for wave 1 following the procedure used in wave 1.

Categories	Comments
	Refreshment sample 2 for sample I in wave 3: Also in wave 3, a refreshment sample for sample I was drawn from the register data of the Federal Employment Agency. To do so, benefit units that received Unemployment Benefit II in July 2008 but not in July 2006 or July 2007 were selected, i.e., new benefit recipients. The sample was drawn in the postcode sectors selected for wave 1 following the procedure used in wave 1. Refreshment sample 3 for sample I in wave 4: Also in wave 4, a refreshment sample for sample I was drawn from the register data of the Federal Employment Agency. Benefit units that were receiving Unemployment Benefit II in July 2009 but not in July 2006, July 2007, July 2008 or July 2009 were selected. These benefit units thus depict the in-flows to benefit receipt. The sample was drawn in the postcode sectors selected for wave 1 following the procedure used in wave 1. Refreshment sample 4 for sample I in wave 5: Also in wave 5, a refreshment sample for sample I was drawn from the register data of the Federal Employment Agency. Benefit units that were receiving Unemployment Benefit II in July 2010 but not in July 2006, July 2007, July 2008 or July 2009 were selected. These benefit units thus depict the inflows to benefit receipt. The sample was drawn in the postcode sec-tors selected for wave 1 following the procedure used in wave 1. 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).

Categories	Comments
	Stage 2, sample VIII: 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 across Germany. Stage 2, sample VII: in sample VII, the individuals were drawn from the registration offices' registers. To do so, 96 municipalities were assigned 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 for birth years of 1992 and earlier. One hundred forty-four addresses were drawn from the municipalities' registers in each sample point. Refreshment sample 5 for sample I in wave 6: In wave 6, a refreshment sample for sample I was again drawn from the register data of the Federal Employment Agency. Benefit units that were receiving Unemployment Benefit II in July 2011 but not in July 2006, July 2007, July 2008, July 2009 or July 2010 were selected, i.e., new benefit recipients. The sample was drawn in the postcode sectors selected for wave 1 following the procedure used in wave 1. Refreshment sample 6 for sample I in wave 7: In wave 7, a refreshment sample for sample I was again drawn from the register data of the Federal Employment Agency. Benefit units that were receiving Unemployment Benefit II in July 2012 but not in July 2006, July 2007, July 2008, July 2009, July 2010 or July 2011 were selected, i.e., new benefit recipients. The sample was drawn in the postcode sectors selected for wave 1 following the procedure used in wave 1.

Categories	Comments
	Refreshment sample 7 for sample I in wave 8: In wave 8, a refreshment sample for sample I was again drawn from the register data of the Federal Employment Agency. Benefit units that were receiving Unemployment Benefit II in July 2013 but not in July 2006, July 2007, July 2008, July 2009, July 2010, July 2011 or July 2012 were selected, i.e., new benefit recipients. The sample was drawn in the postcode sec-tors selected for wave 1 following the procedure used in wave 1. Refreshment sample 8 for sample I in wave 9: In wave 8, a refreshment sample for sample I was again drawn from the register data of the Federal Employment Agency. Benefit units that were receiving Unemployment Benefit II in July 2014 but not in July 2006, July 2007, July 2008, July 2009, July 2010, July 2011, July 2012 or July 2013 were selected, i.e., new benefit recipients. The sample was drawn in the postcode sec-tors selected for wave 1 following the procedure used in wave 1.
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	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 (nur Welle 1) Spell data Unemployment Benefit II: alg2_spells.dta/.sav Spell data unemployment: al_spells.dta/.sav (Wellen 2 und 3) Spell data employment: et_spells.dta/.sav (Wellen 2 und 3)

Categories	Comments
	Spell data gaps: lu_spells.dta/.sav (Wellen 2 und 3) from wave 4 onwards: spell data on employment, unemployment and gaps integrated: bio_spells.dta/.sav Spell data measures: mn_spells.dta/.sav (ab Welle 2) Spell data participation in measures: massnahmespells.dta/.sav (nur Welle 1) 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 (nur Welle 3) Old-age provision individual level: PAVDAT.dta/.sav (nur Welle 3) Vignette data: VIGDAT.dat/.sav (nur Welle 5) Children data: KINDER.dta/.sav (ab Welle 6) Interviewer follow-up data: PINTDAT-1015.dta/.sav

Data access

Categories	Comments
Data access	Scientific Use File (SUF)
Degree of anonymisation	Factually anonymised
Sensitive characteristics	None

Literature

AAPOR (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 7. Auflage. Lanexa: AAPOR.

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

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

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

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

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

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

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, M., Bethmann, A., Fuchs, B., Huber, M. & Trappmann, M. (2014). Codebuch und Dokumentation des „Panel Arbeitsmarkt und soziale Sicherung" (PASS). Bd. 1: Datenreport Welle 7. *FDZ Datenreport 02/2014*. Nürnberg.

Bethmann, A., Fuchs, B. & Wurdack, A. (Hrsg.)(2013). User Guide "Panel Labour Market and Social Security" (PASS). Wave 6. *FDZ Datenreport 07/2013*. Nürnberg.

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

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

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

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

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

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

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

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

Erikson, R., Goldthorpe, J. & Portocarero, L. (1982). Social Fluidity in Industrial Nations: England, France and Sweden. *British Journal of Sociology, 33*, S. 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. (o.J.). $HGEN: Generated Household Level Variables. [http://www.diw.de/documents/dokumentenarchiv/17/60053/hgen.pdf (8.11.2007)].

Fuchs, B. (2013). Structure of the scientific use file and its datasets. In: Bethmann, A., Fuchs, B. & Wurdack, A. (Hrsg.)(2013). User Guide "Panel Labour Market and Social Security" (PASS). Wave 6. *FDZ Datenreport 07/2013* (S. 27-47). Nürnberg.

Ganzeboom, H. & Treiman, D. (1996). Internationally Comparable Measures for Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research, 25*, S. 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 (Hrsg.), *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables* (S. 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*, S. 1-56.

Gebhardt, D., Müller, G., Bethmann, A., Trappmann, M., Christoph, B., Gayer, C., Müller, B., Tisch, A., Siflinger, B., Kiesl, H., Huyer-May, B., Achatz, J., Wenzig, C., Rudolph, H.,

Graf, T. & Biedermann, A. (2009). Codebuch und Dokumentation des „Panel Arbeitsmarkt und soziale Sicherung“ (PASS). Datenreport Welle 2 (2007/2008). *FDZ Datenreport 06/2009*. Nürnberg.

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*, S. 111-129.

Harrison, E. & Rose, R. (2006). ESeC User Guide, Appendix 6 (SPSS-Syntax: Esec Full) [http://www.iser.essex.ac.uk/esec/guide/docs/ Appendix6.sps (31.11.2007)]

Hartmann, J., Brink, K., Jäckle, R. & Tschersich, N. (2008). IAB-Haushaltspanel im Niedrigeinkommensbereich. Methoden- und Feldbericht. *FDZ Methodenreport 07/2008*. Nürnberg.

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

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

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

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

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

Jesske, B. & Schulz, S. (2012). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 5. Erhebungswelle – 2011 (Haupterhebung), *FDZ Methodenreport 11/2012*. Nürnberg.

Jesske, B. & Schulz, S. (2013). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 6. Erhebungswelle – 2012 (Haupterhebung), *FDZ Methodenreport 10/2013*. Nürnberg.

Jesske, B. & Schulz, S. (2014). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 7. Erhebungswelle – 2013 (Haupterhebung), *FDZ Methodenreport 11/2014*. Nürnberg.

Jesske, B. & Schulz, S. (2015). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 8. Erhebungswelle – 2014 (Haupterhebung), *FDZ Methodenreport 11/2015*. Nürnberg.

Jesske, B., Knerr, P. & Schulz, S. (2016). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 9. Erhebungswelle – 2015 (Haupterhebung), *FDZ Methodenreport im Erscheinen*. Nürnberg.

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

Lechert, Y., Schroedter, J. Lüttinger, P. (2006). Die Umsetzung der Bildungsklassifikation CASMIN für die Volkszählung 1970, die Mikrozensus- Zusatzerhebung 1971 und die 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, S. 281-300.

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

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

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

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

Organisation for Economic Co-Operation and Development [OECD] (Hrsg.) (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*, S. 164-175.

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

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

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

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

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

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.

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*. Nürnberg.

Trappmann, M. (2013a). Weighting. In: Bethmann, A., Fuchs, B. & Wurdack, A. (Hrsg.)(2013). User Guide "Panel Labour Market and Social Security" (PASS). Wave 6. *FDZ Datenreport 07/2013* (S. 56-66). Nürnberg.

Trappmann, M. (2014b). Weights. In: Bethmann, A., Fuchs, B. & Wurdack, A. (Hrsg.)(2013). User Guide "Panel Labour Market and Social Security" (PASS). Wave 6. *FDZ Datenreport 07/2013* (S. 81-99). Nürnberg.

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

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

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

Imprint

FDZ-Datenreport 7/2016 (EN)

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

Editorial staff
Dana Müller, Dagmar Theune

Technical production
Dagmar Theune

Download
http://doku.iab.de/fdz/reporte/2016/DR_07-16_EN.pdf

Internet
http://fdz.iab.de/

Corresponding author:
Forschungsdatenzentrum (FDZ) der
Bundesagentur für Arbeit im Institut für
Arbeitsmarkt- und Berufsforschung (IAB),
Regensburger Str. 100, 90478 Nürnberg,
Email: iab.fdz@iab.de

Jonas Beste, Institut für Arbeitsmarkt- und
Berufsforschung (IAB),
Regensburger Str. 104, 90478 Nürnberg,
Tel.: +49 (0) 911/179-2279
Email: Jonas.Beste@iab.de