FDZ

FDZ·Datenreport

Documentation of labour market data

06/2010

EN

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

Datenreport Wave 3

Marco Berg,
Ralph Cramer,
Christian Dickmann,
Daniel Gebhardt,
Reiner Gilberg,
Birgit Jesske,
Karen Marwinski,
Claudia Wenzig,
Martin Wetzel

Bundesagentur für Arbeit

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

Datenreport Wave 3

Marco Berg, Ralph Cramer, Christian Dickmann, Reiner Gilberg, Birgit Jesske, Karen Marwinski (all from the infas Institute of Applied Social Sciences GmbH – infas Institut für angewandte Sozialwissenschaft GmbH), Daniel Gebhardt, Claudia Wenzig, Martin Wetzel (all from the Institute for Employment Research – Institut für Arbeitsmarkt- und Berufsforschung, IAB)

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, the people 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 Datenreport documents the data preparation of the third PASS wave and is based upon the second wave's Datenreport: Gebhardt, Daniel; Müller, Gerrit; Bethmann, Arne; Trappmann, Mark; Christoph, Bernhard; Gayer, Christine; Müller, Bettina; Tisch, Anita; Siflinger, Bettina; Kiesl, Hans; Huyer-May, Bernadette; Achatz, Juliane; Wenzig, Claudia; Rudolph, Helmut; Graf, Tobias; Biedermann, Anika (2009): Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS) Welle 2 (2007/2008), (FDZ Datenreport, 06/2009 (de), Nuremberg, 1097 pages. Sections whose procedures remain the same were adopted without any alterations (this applies to Chapters 1.1, 1.2). Other sections were modified (1.3, 2, 3, 4, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 6). In addition, there are also completely new chapters (5.7, 5.8).

Table of Contents

List of Tables and Figures

Data availability

The dataset described in this document is available for use by professional researchers. Further information can be found at http://fdz.iab.de/.

1   Introduction

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

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

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

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

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

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

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

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

This Datenreport provides an overview of the third survey wave, for which 13,439 individuals were interviewed in 9,535 households[2] between December 2008 and August 2009. 11,300 individuals and 8,207 households were interviewed again in the context of PASS.

For the first time, starting with this third wave's Datenreport, the report was divided into two components. The following is thus relevant for the documentation of the third wave:

1. the wave-specific Datenreport (including codebook) and

2. the cross-wave user guide[3].

---

[1]   Social Code Book II – Basic Social Security for Jobseekers (Sozialgesetzbuch (SGB) Zweites Buch (II) - Grundsicherung für Arbeitsuchende)

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

[3]   Up until the publication of the user guide it is possible to draw on the second wave's Datenreport which also contains cross-wave information, e.g. handling of data.

The cross-wave user guide is created under the responsibility of the PASS project team at the IAB. The documentation of the wave-specific third wave's Datenreport was created by infas. It is based on the second wave's Datenreport.

This wave-specific Datenreport aims to document the wave-related aspects of the study[4]. Following a short overview of the innovations and characteristics of the third wave (Chapter 1.3.), the key figures on samples and response rates of the third wave are reported (Chapter 2). Moreover, the steps of data preparation and the decisions made as part of this process are described (Chapter 5) and an overview of the variables generated is presented (Chapter 4). Additionally, the weighing procedure is presented (Chapter 6). The separate table reports list the frequencies of all variables included in the scientific use file that were recorded in wave 3, divided into their respective datasets (Volume II to Volume V).

1.2   Instruments and interview programme

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

The survey instruments and interview programme of the 3rd wave are based on those used in the 2nd wave of PASS. However, individual questions and modules have been revised or redeveloped (see cross-wave user guide or Chapter 1.3. for an overview).

---

[4]   In contrast, the cross-wave user guide aims to document the study as a whole. It describes in detail the objectives and the design of PASS and presents the content and instruments of the survey. Moreover, the structure of the scientific use file and the concept of the variable types and their names are described. Finally, it describes the utilisation of the various datasets based on examples.

[5]   The target person for the household interview should know as much as possible about general issues regarding the household. In re-interviewed households this was the same person who had completed the questionnaire in the previous wave. If this person was not available during the entire fieldwork period or was no longer a member of the household, then another adult who knew a lot about the household was selected. In the refreshment sample, which was drawn from the BA data, the person registered with the BA as the applicant of UB II should answer the questions about the household. In the case of split-off households a person who used to be a member of the original household and is at least 15 years old should be selected as the target person. If the person with whom the household interview was conducted in the original household in the previous wave now lived in the split-off part of the household, then this person should be selected as the target person of the household interview in the split-off household. Whenever a particular target person who was already known by name was not available during the fieldwork period, the interviewers tried to conduct the household interview with a person aged over 15 who knew as much as possible with regard to general household issues.

Also in the 3rd wave the instruments permit both initial interviews[6] and repeat interviews with households and individuals who had already taken part in one of the previous waves. In order to avoid seam effects[7] in the repeat interviews and to increase data quality, dependent interviewing has been used for certain questions since the second wave to update information that the respondent had provided in the last interview. Furthermore, information about constant characteristics was not gathered again. Owing to the complex updating of the household structure, at the household level, similar to the second wave, a separate questionnaire is available for re-interviewed households (HHalt) and for households participating in the survey for the first time (HHneu).

The individual instruments and the interview programme are described in detail in the cross-wave user guide. The following section provides an overview of the characteristics and innovations of the third wave.

1.3 Characteristics and innovations of the 3rd wave

At this point we would like to provide a brief outline of the characteristics of the 3rd wave of PASS for users who have already worked with the data from the first two panel waves.

The characteristics and innovations in wave 3 affect the set of questions[8] (updating the employment history information collected in wave 2 for the first time, utilisation of special focus modules in the areas of "networking", "health" and "old age provision" and discontinuation of existing modules), the sample, the preparation of data and the documentation.

As part of the third survey wave the employment history information collected in wave 2 is updated for the first time using so-called dependent interviewing. Besides the employment spells[9] also information on the further history of unemployment periods[10] which were ongoing in the previous wave and the receipt of Unemployment Benefit I[11] is collected. Other ongoing

---

[6]   The households interviewed for the first time in the third wave include: (1) households in the refreshment sample of the third wave and (2) households which have split off from households that were involved in the first or second wave of the survey (split-off households). Furthermore, two types of individuals are interviewed for the first time: (1) individuals who are members of a PASS household for the first time in the third wave and (2) individuals who were already members of a PASS household in the first or second wave but for whom no interview from one of the previous waves is available.

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

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

[9]   Questions E 38_X to E63_X in the personal questionnaire.

[10]   Questions A106 to A111 and A117 in the personal questionnaire.

[11]   Questions A112a to A116 in the personal questionnaire.

activities (e.g. vocational training, house wife/house husband, retired person) at the time of the last survey are not explicitly updated[12].

Repeatedly interviewed individuals who indicated at the time of the last interview that they were employed (with an income of more than EUR 400) are now asked if they are still working in the same job or until what point in time they were working in that job. For this updated employment the following information is collected again: (1) occupational status[13], (2) working hours[14] and (3), whether previously fixed-term employments were converted to permanent employments and additionally (4), how the employment was terminated (only employments that were terminated before the interview date of the 3rd wave).

According to a similar logic also the ongoing unemployment spell and the receipt of Unemployment Benefit I at the time of the last interview are updated. In this way, using dependent interviewing, it can be established for repeatedly interviewed individuals in wave 3, up to which point in time the registered unemployment indicated back then lasted as well as the reasons of termination, if any. In these cases it is also established whether and for how long Unemployment Benefit I was received in the period since the last survey. In the case of respondents who are in receipt of Unemployment Benefit I at the time of the 3rd wave interview, also information on the benefit amount is collected.

After updating the employment and unemployment spells specified in wave 2, additional employments (above EUR 400), unemployment periods and all other activities within so-called gaps in the employment history (a gap is defined as a period of more than three months in which neither employment nor unemployment is reported) since the last interview date are surveyed, if applicable, as well as the current (un)employment status at the time of the 3rd wave interview[15]. Information on additional employments, unemployment periods, periods in which Unemployment Benefit I was received and other activities is gathered by means of the known set of questions from wave 2.

The employment history information from newly interviewed individuals is also surveyed analogously to wave 2, only the start time is different. The employment history has been collected for these individuals since January 2006 (in wave 2 the start time for new participants was January 2005).

---

[12] In wave 2 and 3 the gap module in which these activities are collected solely serves to collect activities in periods of more than three months in which no employment or unemployment is indicated.

[13] In the case of terminated updated employments, the occupational status at the end of employment is listed, and in the case of ongoing updated employments the current occupational status is collected.

[14] In the case of terminated updated employments the working hours at the end of the employment are listed, and in the case of ongoing updated employments the current working hours are collected.

[15] Questions P126 to P132 in the personal questionnaire.

The scientific use file comprises as part of the relevant spell datasets[16] the information on periods of employment, unemployment and economic inactivity collected in both wave 2 and wave 3. The integration of all periods in the respective spell datasets follows specific rules (see Chapter 5.6, 5.7, 5.8). If periods are updated across multiple waves, a spell may also include several wave-specific pieces of information (e.g. working hours at the time of the interview in wave 2 and 3). It is stored in wave-specific variables. Wave-specific variables referring to wave 2 end with the digit "0", those referring to wave 3 end with the digit "1" etc. (see cross-wave user guide).

The participation in employment and training measures is surveyed in the 3rd wave by means of the known concept from the 2nd wave. However, the specified measures are not updated, because for each measure the actual or planned end or duration is already known from wave 2. Only the start time is different for repeatedly and newly interviewed individuals. In the case of newly interviewed individuals all measures since January 2007 are relevant for the survey (in wave 2 it was January 2006); repeatedly interviewed individuals are asked to state all measures and funded programmes, which they have participated in since the last interview date. The measure spell dataset (*mn_spells*) in the scientific use file contains all specified measures and funded programmes that were surveyed in the second and third wave.

Moreover, in the third wave additional questions are asked about three areas. Focal points are the areas of "networking", "health" and "old age provision".

In addition to the standard set of questions of the "networking" module, detailed information on the three most important friends is collected in the third wave (gender, school qualification, employment status, type of friendship, condition of friendship). Furthermore, it is surveyed to what extent the respondents have social resources. For this purpose, ten possible situations are presented, in which people commonly ask other people for their support, and additionally the respondent's private contact to certain groups of persons (e.g. entrepreneurs, criminal offenders) is surveyed[17].

In addition to the existing set of questions, wave 3 includes additional questions for all respondents aged 15 and over regarding their health.[18] Therefore, health-related quality of life using the so-called SF-12v2[19], limitations of employment, health behaviour (sports, alcohol and tobacco consumption) as well as obesity are surveyed for the first time in PASS.

There are plans to include both core topics in one of the future waves as well in order to be able to observe long-time changes.

The special focus topic "old age provision" is surveyed in both the household questionnaire and the personal questionnaire (for all individuals between 40 and 64 years as well as their

---

[16] Employment spells: et_spells; unemployment spells: al_spells; spells during periods of economic inactivity: lu_spells.

[17] Questions N1 to N17 in the personal and senior citizens' questionnaire.

[18] Questions G1 to G18 in the personal and senior citizens' questionnaire.

[19] The version used in PASS is not the original version (Ware et al., 2002) but by leave of the DIW the version developed for the SOEP (see Andersen et al., 2007a).

partners in the household, regardless of their age.[20] As part of the household interview, the head of the household is asked to answer detailed questions regarding their property ownership used by themselves (type, size, value). In the personal interview, the respondents are asked to indicate their state and private pension schemes (also supplementary schemes of public service, company pension schemes, "Riester" pension scheme, life insurance). They are asked to specify the duration of payments to date, the type of payout, the estimated amount of the future pension(s) and the age, at which they will have the pension at their disposal, if applicable. Furthermore, they are asked about a (premature) termination or exemption from contributions to private pension schemes (date of termination, reasons for termination, payout amount). Finally, they are asked about their satisfaction with their standard of living (current and at the time of retirement or before retirement). As part of the special focus module, the respondents are also asked for their permission to merge data from the German statutory pension insurance (Deutsche Rentenversicherung). The questions regarding pension schemes are stored in separate datasets at the household and individual level (*HAVDAT, PAVDAT*) and are not integrated in the regular household or individual dataset.

Additionally, three question modules at the individual level are not surveyed in the 3rd wave. This affects the respondent's attitude towards family and occupation, partnership and role relationships as well as the subject area of religion. However, these subject areas will only be skipped for a short period of time. The modules will be included again in future waves according to a rotation schedule.

Furthermore, also in the third wave a so-called refreshment sample was drawn for the BA subsample[21]. The aim is to guarantee the representativeness of the BA sample in the cross-section, and to be able to observe sufficient new transitions into receipt of Unemployment Benefit II over time. For the refreshment sample, benefit communities are drawn which were in receipt of Unemployment Benefit II on 01 July 2008 but not on the sampling date of the 1st or 2nd wave (see Chapter 2.1 and, on the concept of the refreshment sample, Trappmann et al 2009: 11 ff.). These households, which were surveyed for the first time in the third wave, can be identified via the sample indicator (*sample*).

The data preparation in the third wave is conducted for the first time by the survey institute infas Institute of Applied Social Sciences (infas Institut für angewandte Sozialwissenschaft) in Bonn. The first two waves were corrected and edited by the IAB itself. To ensure an analogous preparation with the previous waves, the IAB defined the steps to be taken in advance and provided infas with the necessary materials and preparation do-files. The actual data preparation was performed in close cooperation with the IAB. Basic procedures, e.g. for updating datasets and correcting problems in the household structures, were discussed during the preparation process and decided on by the IAB.

---

[20] Questions HV 1 to HV 9 in the household questionnaire and questions V1 to V99 in the personal questionnaire.

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

Another innovation in the 3rd wave refers to the documentation and the working tools that are available for users when they begin working with PASS. While questionnaires as well as the methods report, which describes the field work of the surveying institute, are still available unchanged, the concept of the Datenreport was revised.

In waves 1 and 2 the Datenreport included both *wave-specific* and *cross-wave* information of a rather general character. Instead of summarising this information in one document again, it will now be split. Also in the future, a Datenreport will be published with each wave. However, this Datenreport will have a much clearer focus and concentrate on *wave-specific* information. This means the Datenreport contains information on the content of a wave, its data preparation and the counts of variables collected in the wave from the various datasets. Additionally, the Datenreport contains information on key figures and the procedure to create weights for the respective wave.

From wave 3 onwards, there will be another document besides the Datenreport: the user guide. While the Datenreport concentrates on wave-specific information and is available as a document for each published wave, the user guide contains *general, cross-wave* information on PASS and the scientific use file. This new part of the documentation answers general questions about PASS, for example about study and sample design, the set of questions across waves, the data structure and the weighting concept. Moreover, the user guide contains examples for the utilisation of the datasets, for example for linking datasets or using the weights. The user guide will be adjusted and revised in the future as well. However, the old version will always be replaced by the new version. In contrast, one wave's Datenreport will not be replaced by the Datenreport of the subsequent wave, because the respective documents refer to different waves.

Due to the change in concept in the 3rd wave, there will be a transition period in which the Datenreport will be available according to the new concept, but the user guide will not yet be finished. For the time being, it is possible to draw on the second wave's Datenreport, which still contains general information on the survey and sample design, the survey instruments, the variable concept as well as practical examples for using PASS according to the old concept.

2   Key figures

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

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

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

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

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

2.1   Sample size

The sample size in a panel starts with the interviewed households from the first survey. In PASS the gross panel sample contains the interviewed households from the 1st wave but also the households from the refreshment samples of waves 2 and 3 that were interviewed for the first time. It must be taken into account that only those households interviewed for the first time are available for repeat interviews that are willing to participate in the panel[22]. The agreement to participate in the panel is only recorded in the first interview. A new confirmation of willingness for these households in the subsequent waves is not required. Besides the confirmation of willingness, access to the panel is already induced during the first interview by the general willingness to participate, that is, by realising an interview. Measures to ensure a best possible selection-free access to the panel as part of PASS are described in detail in the method and field report of waves 1 to 3[23].

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

The panel concept in PASS assumes that new households or split-off households emerge due to move-outs of individuals from panel households, which are counted as separate households as soon as a household interview was conducted. This results in an increasing number of households compared to the original sample. Detailed information on the procedures of the panel concept in PASS can be found under "Split-off households". Besides the expansion of the panel, there may also be a loss of households due to panel mortality. Households in which all respondents passed away or have moved abroad will be removed from the panel gross in the subsequent waves. Moreover, panel losses may occur if no

---

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

[23]   See Hartmann et al. (2008); Büngeler et al. (2009); Büngeler et al. (about to be published).

household interview could be conducted for one household for a period of two consecutive waves. This situation may arise for the first time at the end of the third wave and will then affect the panel gross in the fourth wave.

The case numbers for the gross sample of the respective surveys and subsamples are reported in the following table. 8,349 households of the 11,982 panel households were interviewed at least once in the 3rd wave. In addition to that, there are 1,186 interviewed households from the BA refreshment sample, 1,145 of which agreed to participate in the panel.

Table 1: **Panel sample on the household level by waves and subsamples**

n	BA	Microm	BA inflow 1	BA inflow 2	Total
			Sample		
Wave 1 HH interview conducted	6,804	5,990			12,794
of this: HH willing to participate	6,452	5,548			12,000
Wave 2 Panel HH gross	6,520	5,611			12,131
HH interview conducted	3,491	3,897	1,041		8,429
of this: HH willing to participate	3,360	3,766	1,003		8,129
Wave 3 Panel HH gross	5,833	5,141	1,008		11,982
HH interview conducted	3,754	3,901	694	1,186	9,535
of this: HH willing to participate	3,576	3,777	669	1,145	9,167

Source: HH-Register and PENDDAT; Scientific Use File IAB

The 9,535 household interviews conducted in the third wave correspond to 13,439 personal interviews. The following table lists the distribution of the respondents across the subsamples and the respective surveys.

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

Personal interview conducted		BA	Microm	BA inflow 1	BA inflow 2	Total
				Sample		
Wave 1	abs.	9,386	9,568			18,954
Wave 2	abs.	4,753	6,392	1,342		12,487
Wave 3	abs.	4,913	6,207	898	1,421	13,439

Source: P_Register; Scientific Use File IAB

Respondents without sufficient knowledge of the German language had the option of being interviewed in another language. The alternative interview languages offered were Turkish and Russian. Table 3 shows how many households or individuals were interviewed in the two interview languages.

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

		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

Source: PENDDAT; Scientific Use File IAB

For the overall data pool of the collected panel sample the following outline can be drawn regarding households and individuals over the three survey waves.

Figure 1:   Collected panel sample from households and individuals by survey waves
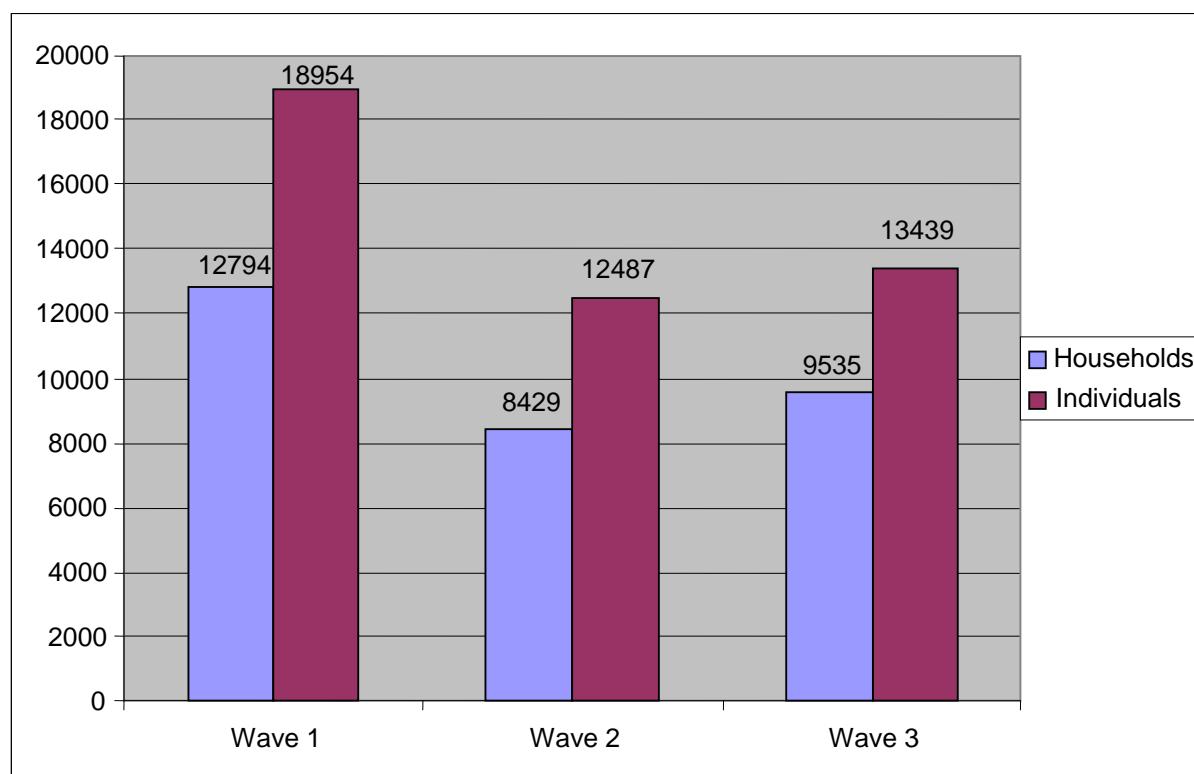
2.2 Response rates

The response rate is calculated in accordance with AAPOR standards (AAPOR 2006). The response rate RR1 is reported, which also includes all cases of unknown eligibility in the denominator and which therefore assumes the lowest value of all response rates[24].

The response rate on the household level is calculated from the share of usable household interviews as a proportion of the total of all usable household interviews and non-neutral non-responses. Only households in which all members have passed away and households which have moved abroad permanently are regarded as cases of neutral non-response. Households are considered usable if at least one complete household interview was conducted. New households are only considered usable if not only the household interview but also at least one complete personal interview is available.

The following response rates were obtained at the household level for the 3rd wave:

Table 4: **Response rate of the 3rd wave on the household level by subsamples**

Wave 3		BA	Microm	BA inflow 1	BA inflow 2	Total
				Sample		
HH gross	abs.	5,833	5,141	1,008	3,801	15,783
	%	100.0	100.0	100.0	100.0	100.0
neutral non-response	abs.	16	37	2	12	67
	%	0.3	0.7	0.2	0.3	0.4
HH gross filtered*		5,817	5,104	1,006	3,789	15,716
	%	100.0	100.0	100.0	100.0	100.0
HH interview conducted	abs.	3,754	3,901	694	1,186	9,535
	%	64.5	76.4	69.0	31.3	60.7
of this: HH willing to participate	abs.				1,145	
	%				30.2	

* HH gros - neutral non-responses

Source: HH-Register; Scientific Use File IAB - für BA-Zugang 2: Bruttodatensatz Welle 3 IAB

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

---

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

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

On average, the following response rates arise from within the interviewed households:

Table 5: **Average response rate within the interviewed households by waves and subsamples**

		BA	Microm	BA inflow1	BA inflow 2	Total
			Sample			
Wave 1	%	85.6	84.2			84.9
Wave 2	%	85.5	85.1	86.2		85.4
Wave 3	%	83.1	83.6	84.3	84.2	83.5

Source: P_Register; Scientific Use File IAB

In addition to the response rates at the household level and within the households, the following table shows the repeat interview rate at the individual level. This reports 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 wave 2 and 3 with respondents willing to participate in the panel from the previous wave by subsamples**

			BA	Sample Microm	BA inflow 2	Total
Wave 2	individuals willing to participate W1	abs.	8,925	8,938		17,863
	re-interviewed individuals in W2	abs.	4,274	5,829		10,103
	Proportion	%	*47.9*	*65.2*		*56.6*
Wave 3	individuals willing to participate W2	abs.	4,686	6,292	1,298	12,276
	re-interviewed individuals in W3	abs.	3,365	4,956	820	9,141
	Proportion	%	*71.8*	*78.8*	*63.2*	*74.5*

Source: PENDDAT; Scientific Use File IAB

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

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

Agreement to panel participation is described in detail in Chapter 2.1 in connection with the sample size. The agreement to participate for households that are interviewed for the first time in a wave[25] in PASS can be illustrated as follows:

Table 7: **Willingness to participate of households interviewed for the first time by waves**

	HH interviews conducted with HHs interviewed for the first time	HH interviews conducted with HHs interviewed for the first time willing to participate	Proportion willing to participate
	abs.	abs.	%
Wave 1	12,794	12,000	93.8
Wave 2	1,086	1,048	96.5
Wave 3	1,327	1,285	96.8

*HH interviewed for the first time from refreshment and split
Source: PENDDAT; Scientific Use File IAB

In the case of households interviewed for the first time in wave 3 the agreement to participate was recorded after the first individual interview. The information of this person was then transferred to the household. If the person agreed to participate, also the household was counted as willing to participate. If the person did not agree to participate, also the household was counted as not willing to participate[26].

In contrast to the agreement to participation, the permission to merge process data of the Federal Employment Agency with the survey data was obtained for each respondent who was interviewed using the personal questionnaire. This question does not apply to individuals

---

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

[26] Information regarding the agreement to participation is thus given by one person for the entire household. The available information on the household level was integrated in the individual dataset (*PENDDAT)* during data preparation. The individual persons interviewed in a household adopted the corresponding information available for the household. The same procedure was applied in wave 2. In wave 1, however, the agreement to participation was recorded after each individual and senior citizens' interview specifically for each person – therefore varying data within a household is possible. Households in which at least one individual with agreement to participation was living were counted as households willing to participate.
As part of the updating of address information after the first personal interview in re-interviewed households, it was explained that an interview would be conducted in the following year. If the respondent did not explicitly object to this announcement, the household was counted as still willing to participate, and the panel variable in the individual dataset (*PENDDAT*) was updated accordingly.

aged 65 and over, because it is not contained in the senior citizens' questionnaire. Agreement to merging of data is not obtained again in each new wave[27].

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

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

	Conducted personal interviews of the wave in which the question on merging of data was asked	Conducted personal interviews of the wave in which merging of data was agreed to	Proportion agreeing to merging of data
	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**

Basis: Individuals 15 to 64 years
Source: PENDDAT; Scientific Use File IAB

For 1,769 (83.1 percent) of the 2,128 individuals represented in Table 8, who gave their agreement to merging of data in wave 3, process data could be linked.

The agreement to merging of data remains valid in the future unless the respondent revokes it[28]. If the agreements from wave 1 and 2 which have not been revoked are taken into account, for 11,098 of a total of 12,104 conducted personal interviews in wave 3 (15 to under 65-year-olds) an agreement is available, which is a share of 91.7%. 10,436 (94.0 percent) of the personal interviews in wave 3 with agreement to merging of data since wave 1 could actually be linked to the process data. In total, 86.2 percent of the personal interviews (15 to under 65-year-olds) from wave 3 are linked to the process data of the Federal Employment Agency.

2.4   Split-off households

PASS is designed as a dynamic panel. People who move into or are born into sample households are also interviewed as long as they are aged 15 or over. People who move out of sample households or do not live in the household for one year or longer should continue to be interviewed, however. These individuals' new households are seen as split-offs from the original sample households. These split-off parts of the household (or split-off households) themselves become sample households of PASS. All of the individuals aged 15 or over living in these households become target persons for personal interviews. Should it occur in one of the subsequent waves that part of this split-off household in turn splits off,

---

[27]   Due to changes in filtering, it could occur that the question regarding agreement to merging of data was raised again in wave 2 and 3 if the person interviewed had not yet given his/her agreement in one of the previous waves.

[28]   Respondents who agreed to the linking of their data to process data of the Federal Employment Agency in the past can of course revoke this consent at any time.

then this new split-off household, too, becomes a PASS sample household, irrespective of whether there is still anyone from one of the original samples living there ("infinite degree contagion model", Rendtel and Harms 2009, 267). Individuals who have moved abroad, on the other hand, cease to be included in the survey as they no longer belong to the population and because the research questions specific to SGB II no longer apply. People who do not live in the household for less than one year continue to be counted as household members and do not constitute a new PASS household.

Between the survey dates of the 1st and 2nd wave a total of 344 households split off from the households already included in the first wave of the survey. It was possible to interview 46 of these split-off households during the fieldwork period of the 2nd wave. The split-off households that were not surveyed will be contacted again in the 3rd wave as long as they have not definitely refused to participate. In wave 3 there are 358 split-off households, 142 of which could be interviewed.

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 take over from their original household the sample indicator (*sample*), the information as to the sampling year (*jahrsamp*), the primary sampling unit (*psu*) and its stratification (*strpsu*).

3 Dataset structure

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

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

information is assigned to the same variable in each case in wave-specific rows of the observation units.

Preparing the data in long format has both advantages and disadvantages. The decisive advantage of this variant is that the data are already available in the structure required for many longitudinal analyses (such as duration history analyses). It is no longer necessary to invest additional time and effort in creating a cross-wave file. The switch from long format to wide format is also comparatively easy to perform. STATA for example provides a possibility to switch between the two formats with little effort using the "reshape" command. Until a few years ago, the central argument against using this type of dataset structure was the significantly larger memory space required, which mainly results from the fact that even variables recorded in only one or a small number of survey waves always require a complete column across all waves in the dataset. In addition, the long files become relatively large with increasing duration of the panel, simply as a result of all annual waves being appended to one another, which significantly increases the storage space required and the time to perform individual operations using the data. The wide availability of fast processors and large storage capacities in even simple desktop PCs makes this argument seem insignificant in the meantime. Another disadvantage is the merging with further information. Unlike the datasets prepared in wide format, an additional key variable is now required in order to be able to identify an observation clearly. This may be a wave identifier in the household or individual datasets, or alternatively the spell number in the spell datasets, which are also available in long format. Furthermore, it is not apparent at first sight which variables were surveyed for which waves, as all of the variables ever surveyed are present in the dataset. These variables are given a special code (-9) for waves in which they were not surveyed.

When the advantages and disadvantages of long format for the user are weighed up, the advantages clearly outweigh the disadvantages in our opinion. Accordingly, the household and individual datasets of PASS (HHENDDAT; PENDDAT) and the corresponding weighting data (hweights; pweights) were prepared in long format. The information collected as part of the special module regarding pension schemes at the household and individual level was outsourced into separate datasets (HAVDAT; PAVDAT), which can be merged with the household or individual dataset via the corresponding key variables.[29]

At the household level, the scientific use file contains the data on the household's receipt of Unemployment Benefit II processed in spell form (alg2_spells). At the individual level there are four spell datasets. These are (1) data about employment spells (et_spells), (2) periods of unemployment (al_spells) and (3) periods of economic inactivity (lu_spells), since January 2005 in each case, and (4) spell data on participation in employment and training measures (mn_spells) since January 2006. The household and the individual registers (hh_register; p_register) are available in wide format.

---

[29] The datasets each contain all households or individuals interviewed in wave 3. If in one case no information regarding pension schemes was collected, the variables with regard to contents are given the code "-3" (not applicable, filter).

Figure 2: **Dataset structure of PASS in wave 3**
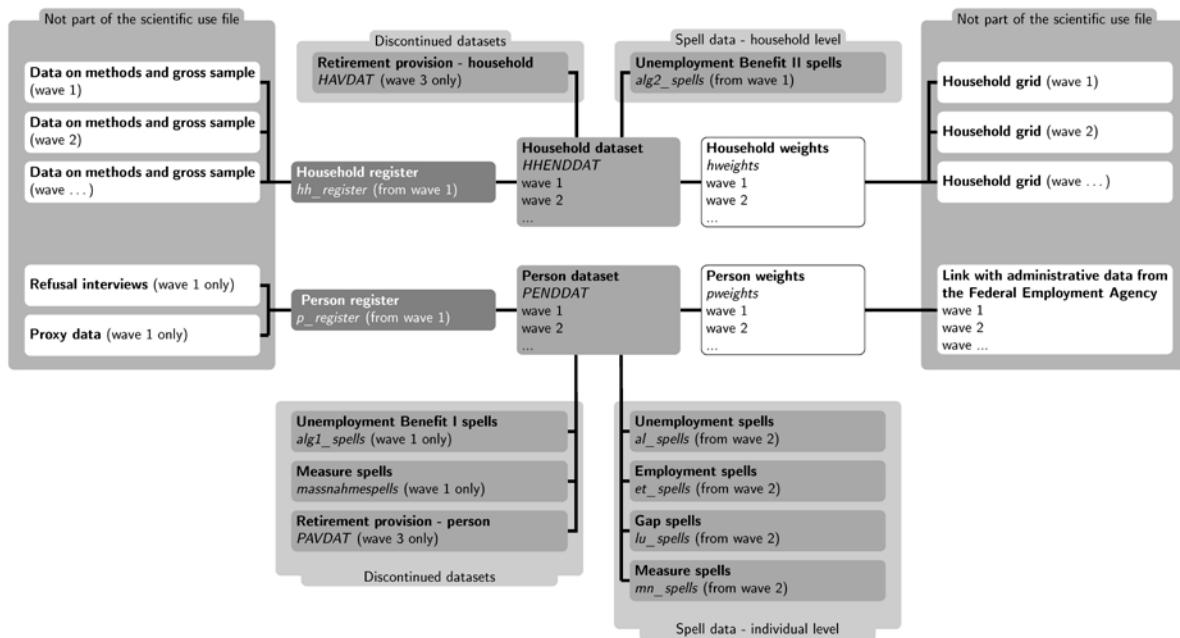
4 Generated variables

4.1 Coding of responses to open-ended survey questions

Some items of the survey were gathered as closed items with an open residual category or as open-ended items. In such cases, additional variables were usually generated[30] which differed from the original variable only insofar as the information from the open-ended responses was coded to the corresponding categories where possible. Moreover, in some cases new categories were created based on the information from open-ended questions. The name of these additional variables differs from that of the original variable in the last digit only, where the "0" was replaced by a "1". The items on country of birth, nationality, and the parents'/grandparents' country of residence before migration were also anonymised and given eloquent variable names[31]. Table 9 gives an overview of the open-ended survey questions which were coded in the third wave[32].

---

[30] Other information from open-ended questions was not coded, for example the name of the institution providing basic social security (P138).

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

[32] Variables for which information was gathered and coded via open-ended questions in the 1st wave but not in the subsequent waves are not listed (with the exception of the spell dataset for Unemployment Benefit II). For the observations as of the 2nd wave these variables are given the code -9 (item not surveyed in wave) and are documented in the Datenreport of the 1st wave.

Table 9: **Coding of responses to open-ended survey questions at the household level in wave 3**

Questionnaire number Re-interviewed HH	New sample HH	Coded to variable	Dataset	Description
HH64	n. in Q. vers.	*HW0881a-j*	*HHENDDAT*	other reason for moving out, not listed
HH85	HH42	*HD0601*	*HHENDDAT*	Language spoken in HH: other language, not listed
HH87	HH44	*HD0801*	*HHENDDAT*	Language spoken in HH after follow-up question about other languages: other language, not listed
HH88	HH45	*HD0901*	*HHENDDAT*	Language spoken in HH, equal use of two languages: first language is another language, not listed
HH89	HH46	*HD1001*	*HHENDDAT*	Language spoken in HH, equal use of two languages: second language is another language, not listed
HH99	HH56	*AL21301a-f* *AL21401a-f* *AL21501a-f* *AL21601a-f* *AL21701a-f* *AL21801a-f* *AL21851a-f* *AL21901 a-f* *AL22001a-f* *AL22101a-f* *AL22102a-f* *AL22103a-f*	*alg2_spells*	other reason for benefit cut, not listed
Z1	Z1	*AL22201a-f*	*alg2_spells*	other reason for discontinuation of receipt of UB II, not listed

Table 10: **Coding of responses to open-ended survey questions at the individual level in wave 3**

Questionnaire number Individuals	Senior cit's	Coded to variable	Dataset	Description
P8_6	n. in Q vers.	*PB0231*	*PENDDAT*	other German school qualification, not listed (update)
P8_7	n. in Q vers.	*PB0231*	*PENDDAT*	other foreign school qualification, not listed (update)
P10_9	P5_9	*PB0401*	*PENDDAT*	other German school qualification, not listed (first survey or not reported in previous wave)
P10_10	P5_10	*PB0401*	*PENDDAT*	other foreign school qualification, not listed (first survey or not reported in previous wave)
P11	n. in Q vers.	*PB1001*	*PENNDAT*	other foreign school qualification, not listed (first survey or not reported in previous wave)
P26_9	P7_9	*PB1301a-j*	*PENDDAT*	other German vocational qualification, not listed (update or first survey)
P26_10	P7_10	*PB1301a-j*	*PENDDAT*	other foreign vocational qualification, not listed (update or first survey)
P28	n. in Q vers.	*PB1601*	*PENDDAT*	other qualification, not listed, to which the foreign qualification corresponds
P111	n. in Q vers.	*AL0601*	*al_spells*	other reason, not listed, for no longer being registered as unemployed
P129	n. in Q vers.	*LU0101*	*lu_spells*	other gap status, not listed
P143	n. in Q vers.	*PTK0321a-g*	*PENDDAT*	other reason, not listed, for not having to seek employment
P162	n. in Q vers.	*MN0201a-h MN0202h*	*mn_spells*	other component of measure, not listed
P167	n. in Q vers.	*MN1001a-e*	*mn_spells*	other reason, not listed, why the measure was ended prematurely
Z2	n. in Q vers.	*ET2401*	*et_spells*	other way of getting to know of the employment, not listed
P184	n. in Q vers.	*PAS0901a-i*	*PENDDAT*	other places, not listed, where target pers. obtained information about job vacancies
P219	P51	*PG0901a-g*	*PENDDAT*	other health problems, not listed
P223	P54	*PG1301*	*PENDDAT*	other health insurance, not listed
P264	P73	*ogebland*	*PENDDAT*	other country of birth, not listed
P267	P76	*ostaatan*	*PENDDAT*	other nationality, not listed

Table 10: Coding of responses to open-ended survey questions at the individual level in wave 3: **Coding of responses to open-ended survey questions at the individual level in wave 3 (continued 1)**

Questionnaire number Individuals	Senior cit's	Coded to variable	Dataset	Description
P274	P80	*ozulanda-f*	*PENDDAT*	other country, not listed, from which parent/grandparent migrated
P275	P81	*PMI1111*	*PENDDAT*	Language spoken in circle of friends: other language, not listed
P276	P82	*PMI1121*	*PENDDAT*	Language spoken in circle of friends, equal use of two languages: first language is another language, not listed
P277	P83	*PMI1131*	*PENDDAT*	Language spoken in circle of friends, equal use of two languages: second language is another language, not listed
P278_9	n. in Q vers.	*PSH0201*	*PENDDAT*	other German school qualification of mother, not listed
P278_10	n. in Q vers.	*PSH0201*	*PENDDAT*	other foreign school qualification of mother, not listed
P279_7	n. in Q vers.	*PSH0301a-i*	*PENDDAT*	other German vocational qualification of mother, not listed
P279_8	n. in Q vers.	*PSH0301a-i*	*PENDDAT*	other foreign vocational qualification of mother, not listed
P289_9	n. in Q vers.	*PSH0501*	*PENDDAT*	other German school qualification of father, not listed
P289_10	n. in Q vers.	*PSH0501*	*PENDDAT*	other foreign school qualification of father, not listed
P290_7	n. in Q vers.	*PSH0601a-i*	*PENDDAT*	other German vocational qualification of father, not listed
P290_8	n. in Q vers.	*PSH0601a-i*	*PENDDAT*	other foreign vocational qualification of father, not listed

4.2 Harmonisation

For some variables, there were changes in the survey instruments across the waves. Above all, the integration of the employment biography module in wave 2 resulted in the fact that critical information on employment status, current main profession, economic inactivity status and receipt of Unemployment Benefit I was collected differently than in the first wave. Since then, information has been collected not only with regard to the date of the interview but also in spell form for certain periods of time.

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

Changes between the waves can affect the entire survey concept, categories and the interviewed groups. Therefore, harmonised variables consider different source variables that result from changed survey concepts, changes to categories as well as interviewed groups, and try to standardise them as far as possible across waves before generation is performed based on the variables.

So far, harmonisations have been performed for the employment status (erwerb2) and the simple classification of the occupational status (stibkz). However, the number of necessary harmonisations can be expected to increase with the duration of the panel.

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

Variable	Subject area	Description
erwerb2	Occupation	Employment status, generated (all waves)
stibkz	Occupation	Current occupational status, simple classification, harmonised (anonymised)

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

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

For this reason, before working with the cross-wave but not harmonised variables it should be verified whether differences in the interviewed groups might be problematic for the respective evaluations and whether a standardisation may be necessary[33].

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

---

[33]   For example, the groups of respondents which were asked about their occupation varied in wave 1 and the subsequent waves. Accordingly, also the respective groups which provided information on occupational status, occupational activities, working hours, fixed-term employment etc. varied.

Table 12: **Cross-wave but not completely harmonised variables in the individual dataset (*PENDDAT*)**

Variable	Subject area	Description
nichterw	occupation	employment status, generated (all waves)
nichtew2	occupation	current occupational status, simple classification, harmonised (anonymised)
isco88	occupation	ISCO 88 (ZUMA coding), current job, generated
isco88it	occupation	ISCO 88 (Infratest coding), current job, generated
kldb_it	occupation	classification of occupations 1992 (Infratest coding), current job,
arbzeit	occupation	weekly hours of work incl. details in the case of irregular working hours, generated
befrist	occupation	current job: fixed-term contract? Gen. (all waves)
mps	occupation	Magnitude-Prestige-Scale, current job, gen.
siops	occupation	Standard International Occupational Prestige Scale, current job, generated
isei	occupation	International Socio-Economic Index, current job, generated
egp	occupation	class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), current occupation, generated
esec	occupation	European Socio-economic Classification (ESeC), current occupation, generated
stib	occupation	occ. status, code number, current job, generated
alg1abez	receipt of benefits	current receipt of UB I, generated
aktmassn	participation in measures	current participation in a measure funded/promoted by the employment agency, generated

Furthermore, there are variables in the dataset which were generated for all waves, but are not evaluable in the longitudinal section. These are the generated income variables at the individual level. In these cases, the differences in survey concepts between wave 1 and the subsequent waves were not taken into account when generating variables. For wave 1 the relevant variables contain income data which refer to the main profession if the respondent worked at least one hour per week. As of wave 2, the data have included not only information on the main profession, but also on all employments with an income of more than EUR 400 which were permanent at the time of the interview – hence, the variables contain cumulated information in these cases. The introduction of the employment biography module as of wave 2 was one of the reasons for this.

An evaluation of these variables in the longitudinal section would cause errors, because the information contained is based on different survey concepts, includes constructs with different contents and is available for different groups of respondents.

A revision and standardisation in the form of cross-wave or harmonised variables is being planned and will be published with the scientific use file of a future wave.

Table 13: **Variables in the individual dataset (*PENDDAT*) generated for all waves, which are, however, not evaluable in the longitudinal section**

Variable	Subject area	Description
brutto	income	gross income, incl. categorised info., generated
bruttokat	income	categorised gross income, generated
netto	income	net income, incl. categorised info., generated
nettokat	income	categorised net income, generated

4.3 Dependent interviewing

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

There were mainly two goals that were pursued by utilising information from previous waves. First, at certain points only changes since the previous wave were supposed to be recorded, partly depending on whether there was information on specific questions available from the previous wave[34]. At these points, information from previous waves was used for controlling the filter. Secondly, the respondent should receive content information. Therefore, at those points where changes since the previous wave were supposed to be recorded, the interview date from the previous wave was included in the question text in order to be able to define the reporting period more clearly[35]. At other points, especially when updating spell information[36], also responses by the respondent from the previous wave were integrated in the question texts, which served as a reminder of the respondents' answers from the previous wave. This was to prevent that changes in status were reported which did not take place in reality but are an artefact of the open-ended survey arising from wrong memories or unprecise information.

If information from a single wave in the dataset is reviewed, only incomplete information is available for some respondents due to dependent interviewing, which only represents the changes between two survey dates. For respondents who were questioned on a subject for the first time, information may be available which is complete when considering this particular wave[37].

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

---

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

[35] For example, if only new school qualifications since the last interview were supposed to be recorded, the following question was raised first: "Have you obtained a general school qualification since our last interview on [display of interview date in previous wave]?"

[36] Examples are the updating of Unemployment Benefit II receipt from the previous wave in the household interview of the respective current wave or the updating of employment or unemployment spells in the personal interview.

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

Table 14, Table 15 and Table 16 below provide a brief overview of all of the relevant points in the questionnaires and show in which variables the updated information can be found. The cases where generated variables were updated or continued are additionally listed in Chapter 4.4 of this Datenreport.

Table 14: **Updated information from the previous wave in wave 3, re-interviewed households**

Household questionnaire for re-interviewed households (HHalt)			
Construct	**Q. no's.**	**Remarks**	**Update in variable**
Housing situation		Form of accommodation, type of rental contract and type of hostel/home/hall of residence etc., updated by Infratest during the interview	HHENDDAT: *HW0200 to HW0400*
Household structure	HH1-HH60	Household size, updated by Infratest during the interview	HHENDDAT: *HA0100*
		Gender of the individuals in the household, corrected, if necessary, by Infratest during the interview	HHENDDAT: *HD0100a to HD0100o*
		Age of the individuals in the household, updated by Infratest during the interview	HHENDDAT: *HD0200a to HD0200o*
		Family relationships, updated by Infratest during the interview	not provided in the SUF
Year of move into current dwelling	HH66	Updated in generated variable	HHENDDAT: *einzugj*
Size of dwelling in sqm	HH65/HH69	Updated in generated variable	HHENDDAT: *wohnfl*
Receipt of Unemployment Benefit II	HH91-HH104	Updated in spell dataset for Unemployment Benefit II	alg2_spells: Variables of the spell dataset for Unemployment Benefit II
		Information on the current Unemployment Benefit II receipt of the household	HHENDDAT: alg2abez PENDDAT: hhalg2
		Information on the Unemployment Benefit II receipt of the benefit community	p_register: bgbezs3; bgbezb3

Table 15: Updated information from the previous wave in wave 3, new sample households

Household questionnaire for new sample households (HHneu)			
Construct	Q. no's.	Remarks	Update in variable
Receipt of Unemployment Benefit II	HH48-HH60	Updated in spell dataset for Unemployment Benefit II	alg2_spells: *Variables of the spell dataset for Unemployment Benefit II*
		Information on the current Unemployment Benefit II receipt of the household	HHENDDAT: *alg2abez* PENDDAT: *hhalg2*
		Information on the Unemployment Benefit II receipt of the benefit community	p_register: *bgbezs3; bgbezb3*

Table 16: Updated information from the previous wave in wave 3, personal questionnaire

Personal questionnaire			
Construct	Q. no's.	Remarks	Update in variable
Highest general school qualification	P7-P12	Updated in generated variable	PENDDAT: *schul1* (without open-ended questions) *schul2* (with open-ended questions)
Year in which highest general school qual. was gained	P13	Updated in generated variable	PENDDAT: *schulabj*
Vocational qualification	P23-P29	Highest vocational qualification, updated in generated variable	PENDDAT: *beruf1* (without open-ended questions) *beruf2* (with open-ended questions)
Year in which vocational qual. was gained	P30	Updated in generated variable	*berabj*
Periods of employment with an income of more than EUR 400	E38-E63, P38-P63	Updated in spell dataset on employment;	et_spells: *Variables of the spell dataset on employment*
		Information on current employment, updated in generated variables	PENDDAT: *isco88; isco88_it; kldb_it;stib; stibkz; arbzeit; befrist; mps; siops; isei; egp; esec*
		Information on current employment/economic inactivity, updated in generated variables	PENDDAT: *erwerb2; nichterw; nichtew2*
Periods of registered unemployment incl. UB I receipt	A106-A117 P106-P117	Updated in spell dataset on unemployment	al_spells: *Variables of the spell dataset on unemployment*

Table 16: Updated information from the previous wave in wave 3, personal questionnaire **Updated information from the previous wave in wave 3, personal questionnaire (continued)**

Personal questionnaire			
Construct	**Q. no's.**	**Remarks**	**Update in variable**
		Information on current employment/economic inactivity, updated in generated variables	PENDDAT: *erwerb2; nichterw; nichtew2*
		Information on current Unemployment Benefit I receipt	al_spells: *Variables of Unemployment Benefit I receipt in the spell dataset on unemployment* PENDDAT: *alg1abez*

A distinction has to be drawn between these characteristics, where information collected in the past is updated with information on changes between the survey dates, and the so-called "constant characteristics". They are expected not to change over time. Therefore, these characteristics are recorded only once in PASS, although later corrections may be possible in some cases. Because information on these characteristics is usually available in the recorded variables of the first interview date only, it is subsequently provided in the form of generated variables (see Chapter 4.4, the PASS user guide, or until its publication Chapter 10.3 in the PASS Datenreport of wave 2).

4.4 Simple generated variables

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

For households or individuals who are interviewed on a subject for the first time the simple generated variables can always be created based on the information collected in the current wave. For households or individuals who answered questions on a subject already in a previous wave, however, they can be distinguished in the cross-section datasets (*HHENDDAT; PENDDAT*) with regard to the origin of the individual variables required for generating those variables. The three different types of simple generated variables are listed in

Table 17.

Table 17: **Types of simple generated variables in the cross-section datasets (*HHENDDAT; PENDDAT*) for households or individuals who answered questions on specific subjects already in a previous wave**

Type	Generated variable based on source data from		Description
	Wave in which the HH/person was interviewed for the first time on the subject	Current wave	
unveränderlich (uv)	yes	no	The information recorded in the first interview is usually adopted in the subsequent wave – unless input errors were corrected in the current wave. Example: *zpsex* (gender)
fortgeschrieben (fs)	yes	yes	Information that was current in the previous wave is combined with the information of the current wave and updated, if necessary. Example: *schul1* (highest school qualification)
unabhängig neu (neu)	no	yes	The variable is newly generated from the data of the current wave in each wave, regardless of the information from the previous wave. Example: *hhincome* (net income of household)

The simple generated variables are shown in the dataset-specific Table 18 to Table 25. Each variable has a short description. Additionally, the source variables necessary for generating the variable in wave 3 are listed[38]. Moreover, for the cross-section datasets (*HHENDDAT; PENDDAT*) the type of simple generation shown in Table 16 is indicated (uv; fs; neu). For the spell datasets, this subdivision does not make sense, since in these cases no wave-specific observations are available. Instead, the generated variables are newly generated on the spell level if the spell was newly created in the current wave or was updated with information collected in the current wave. Also the register datasets follow a different logic so that also in these cases a further differentiation was abandoned.

---

[38] Information on how the variables were generated in the cross-section datasets (*HHENDDAT; PENDDAT*) for observations in wave 1 or wave 2 can be found in the respective Datenreport. The documentation of the individual waves also describes the generation of the wave-specific variables in the register datasets. The generated variables in the spell datasets were always generated in the datasets that were already updated. If a spell was not updated, the corresponding generated variables remained unchanged (with an exception, if necessary, that a special code was set in the censoring indicator if the spell could not be continued for technical reasons). If a spell was updated, always the most current information was used, i.e. the variables containing the information from the current wave or the section variables in the spells relevant for the current wave.

Table 18: Simple generated variables for wave 3 in the household dataset (*HHENDDAT*) (in alphabetical order)

Variable	Label and description	Source var. for gen. in wave 3
alg2abez	*Current receipt of Unemployment Benefit II of the HH, generated* Indicator for the household's current receipt of Unemployment Benefit II (neu)	*zensiert; AL20300; AL20400; AL20500 (alg2_spells);* information on further receipts of Unemployment Benefit II (HHalt: HH104; HHneu: HH61); *hintjahr (HHENDDAT)*
bik	*BIK region size classes (GKBIK10), generated* *The information on the region size class was generated by TNS Infratest by converting the postcode available from the address data to GKBIK10 (neu).*	supplied by survey institute
blneualt	*Western German States or Eastern German States, generated* Aggregation of German federal states into the Western German States of the former FRG (without Berlin) and the Eastern German States of the former GDR (with Berlin). The federal state was identified by TNS Infratest based on the postcodes available from the address data (neu)	information generated and supplied by the survey institute on the federal state in which the household is resident at survey date
einzugj	*Year of move into current dwelling, generated* Information as to the year in which the household moved into the current dwelling. In the case of re-interviewed households, as of wave 2, the year of the move into the current dwelling was only asked if the household had been living in a residential home or if it had moved house since the previous wave (fs)	<u>In case of first-time interview:</u> *HW0900 (HHENDDAT)* <u>In case of repeat interview:</u> *einzugj from previous wave; HW0900, HW0200; umzug (HHENDAT)*
hhinckat	*categorised household income per month (in EUR), generated* *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, generated* *Generation of an integrated variable from categorised and open-ended survey questions on the net household income (neu)*	*HEK0600; HEK0700; HEK0800; HEK0900; HEK1000; HEK1100 (HHENDDAT)*
hintdat	*Date of household interview* *Generated variable with the date on which the household interview was conducted in the form YMMDD (neu)*	*hintjahr, hintmon, hinttag (HHENDDAT)*
kindu4	*Control variable: Child under age of 4 in the HH* *The variable indicates that at least one person 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 person aged four is actually the child of another person living in the household (neu)*	*HD0200a - HD0200o (HHENDDAT)*

Table 18: Simple generated variables for wave 3 in the household dataset (*HHENDDAT*) (in alphabetical order) (continued)

Variable	Label and description	Source var. for gen. in wave 3
kindu13	*Control variable: child under age of 13 in the HH* The variable indicates that at least one person in the household is below 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 person aged 13 is actually the child of another person living in the household (neu)	*HD0200a - HD0200o* (*HHENDDAT*)
kindu15	*Control variable: child under age of 15 in the HH* The variable indicates that at least one person in the household is below the age of 15 in the wave. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this person aged 15 is also actually the child of another person 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 included to generate the variable (neu)	*HD0200a - HD0200o*; categorical follow-up question about age group (in cases of no response in *HD0200*) (*HHENDDAT*)
wohnfl	*Living space in sqm, generated* Information on the size of the living space in the household's current dwelling. In the case of re-interviewed households, as of wave 2, the size of the living space was only asked if the household had moved house or if the house/apartment had changed since the previous wave (fs)	_In case of first-time interview:_ *HW1000 (HHENDDAT)* In case of repeat interview: *wohnfl from previous wave; HW1000; HW0910; HW0920* (*HHENDDAT*)

Table 19: Simple generated variables for wave 3 in the individual dataset (*PENDDAT*) (in alphabetical order)

Variable	Label and description	Source var. for gen. in wave 3
aktmassn	*Current participation in a measure funded/promoted by the employment agency, generated* Indicator: respondent is participating in a measure of active labour market policy at interview date (neu)	*MN0500; zensiert (mn_spells); PA0711b-f; PA0721a-f* (*PENDDAT*); information from follow-up validation question P 178_X_Prüf (personal questionnaire)
alg1abez	*current receipt of UB I, generated* Indicator: respondent is in receipt of Unemployment Benefit I at interview date. In the third wave the periods since January 2006 during which the respondent was registered as unemployed were surveyed. For each spell additional questions were asked as to whether the respondent received UB I and if so, during which period. This information was combined with a follow-up question for respondents who were aged 58 or over and were therefore entitled to Unemployment Benefit I without being registered as unemployed (neu)	*AL0700; AL1000; AL1100; AL1200; alg1bj; alg1ej (al_spells); PA0405* (*PENDDAT*); information as to whether there is a further spell of unemployment (P117/A117)

Variable	Label and description	Source var. for gen. in wave 3
alg1s05	*Indicator: Receipt of Unemployment Benefit I since Jan. 2005? Gen. (all waves)* Indicator: Respondent received UB I at some time since January 2005. In the 3rd wave the periods since January 2006 during which the respondent was registered as unemployed were surveyed. For each spell additional questions were asked as to whether the respondent received UB I and if so, during which period. This information was combined with a follow-up question for respondents who were aged 58 or over and were therefore entitled to UB I without being registered as unemployed (fs)	*In case of first-time interview:* AL0700; AL1200; alg1bj; alg1ej (*al_spells*); PET0911; PA0405 (*PENDDAT*); info. as to whether there is a further spell of unemployment (P117/A117) *In case of repeat interview:* AL0700; AL1200; alg1bj; alg1ej (*al_spells*); *alg1s05* from prev. wave; PET0911; PA0405 (*PENDDAT*); info. as to whether there is a further spell of unemployment (P117/A117)
apartner	*Control variable: cohabitee in the household* Indicator: respondent has a cohabitee or a partner whose status is not specified in the HH (neu)	Info. on relationships between HH members (household grid); *PD0500 - PD0900 (PENDDAT)*
arbzeit	*Weekly hours of work incl. details in the case of irregular working hours, gen.* Weekly hours of work in the job held by the respondent on the interview date, generated from responses to open-ended questions on working hours and categorical follow-up question in the case of irregular working hours (neu)	ET2101; ET2201 (*et_spells*); PET0510; PET0700 (*PENDDAT*)
befrist	*Current job: fixed-term contract? Gen. (all waves)* Indicator: The job held by the respondent on interview date is on a fixed-term contract (neu)	PET2510a; PET2510b (*PENDDAT*)
begjeewt	*Year of first employment, generated* Year in which the respondent first worked in a regular job. To generate the variable, information about the first regular employment was combined with information from the employment spells if the respondent had already reported his/her first regular employment during the questions on employment spells since January 2006 (uv)	*In case of first-time interview:* bjahr (*et_spells*); PET0150; PET0151; PET3200b (*PENDDAT*) <u>After first-time interview:</u> *begjeewt* from previous wave (*PENDDAT*)
begmeewt	*Month in which first job taken up, generated* Month in which the respondent first worked in a regular job (generation: see *begjeewt*) (uv)	*In case of first-time interview* bmonat (*et_spells*); PET0150; PET0151; PET3200a (*PENDDAT*) <u>After first-time interview:</u> *begmeewt* from previous wave (*PENDDAT*)
berabj	*Year in which highest vocational qual. gained* Year in which the respondent gained his/her highest vocational qualification at the time of the interview (fs) <u>Note:</u> The years in which the vocational qualifications reported in the first wave were gained were surveyed in the second wave.	<u>In case of first-time interview:</u> PB1300a-j; PB1310am-km; PB1310aj-kj (*PENDDAT*) <u>In case of repeat interview:</u> *berabj* from previous wave; PB1300a-j; PB1310am-km; PB1310aj-kj (*PENDDAT*)

Variable	Label and description	Source var. for gen. in wave 3
beruf1	*Highest vocational qual., excl. foreign qual's and open info. generated* Identification of the highest vocational qualification at the time of the interview by hierarchising the vocational qualifications indicated by the respondents, excl. information from open-ended questions (fs)	In case of first-time interview: *PB0100; PB0200; PB0300; PB1200b; PB1200c; PB1300a-j; (PENDDAT)* In case of repeat interview: *beruf1* from previous wave; *PB0100; PB0200; PB1200a; PB1300a-j (PENDDAT)*
beruf2	*Highest vocational qual., incl. foreign qual's and open info. generated* Like *beruf1* with the following differences: 1. Inclusion of responses to open-ended questions; 2. inclusion of information on foreign qualifications; 3. degrees not distinguished by type of institution (e.g. university or other institution of higher education) but by the qualification level (Bachelor's degree; Master's degree; Ph.D.). (fs)	In case of first-time interview: *PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601 (PENDDAT)* In case of repeat interview: *beruf2* from previous wave; *PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601 (PENDDAT)*
brutto	*Gross income incl. categorised information, generated* Generation of an integrated variable from categorised and open-ended survey questions on gross income (neu) Note: The variable was generated for all waves but is currently not evaluable in the longitudinal section (see Chapter 4.2).	*PEK0100b; PEK0200; PEK0300; PEK0400; PEK0500; PEK0600 (PENDDAT)*
bruttokat	*Categorised gross income, generated* Aggregation of the categorised information on gross income, combined from several items on income categories (neu) Note: The variable was generated for all waves but is currently not evaluable in the longitudinal section (see Chapter 4.2).	*PEK0200; PEK0300; PEK0400; PEK0500; PEK0600 (PENDDAT)*
ejhrlewt	*Time when last job ended (year)* Year in which the respondent was last in employment. To generate this variable, information from the employment spells was combined with information on the last job if the respondent had been out of work since Jan. 2005 (fs)	*In case of first-time interview:* *PET1200b (PENDDAT); ejahr; emonat (et_spells)* In case of repeat interview: *ejhrlewt* from prev. wave (*PENDDAT*); *ejahr; emonat (et_spells)*
ekin1517	*Control variable: own child aged between 15 and 17 in the household* This variable indicates that the resp. has a natural child, a stepchild/adopted child or a child of non-specified status between 15 and 17 in the HH (neu)	Information on relationships between household members (household grid)

Table 19: **Simple generated variables for wave 3 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 3)**

Variable	Label and description	Source var. for gen. in wave 3
ekind	*Control variable: Own child in HH* This variable indicates that the respondent has a natural child, a stepchild/adopted child or a child of non-specified status of any age in the household (neu)	Information on relationships between household members (household grid)
ekin614	*Control variable: own child aged between 6 and 14 in the household* This variable indicates that the respondent has a natural child, a stepchild/adopted child or a child of non-specified status aged between 6 and 14 in the household (neu)	Information on relationships between household members (household grid)
ekinu15	*Control variable: own child under age of 15 in household* This variable indicates that the respondent has a natural child, a stepchild/adopted child or a child of non-specified status under the age of 15 in the household (neu)	Information on relationships between household members (household grid)
ekinu18	*Control variable: own child under age of 18 in household* This variable indicates that the respondent has a natural child, a stepchild/adopted child or a child of non-specified status under the age of 18 in the household (neu)	Information on relationships between household members (household grid)
emonlewt	*Time when last employment ended (month)* Month in which the respondent was last in employment (generation: see *ejhrlewt*) (fs)	*In case of first-time interview:* PET1200a (*PENDDAT*); *emonat2* (*et_spells*) *In case of repeat interview:* *emonlewt* from previous wave (*PENDDAT*); *emonat* (*et_spells*)
epartner	*Control variable: spouse or registered partner in HH* This variable indicates that the respondent has a spouse or a same-sex registered partner in the household (neu)	Information on relationships between household members (household grid)

Table 19: **Simple generated variables for wave 3 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 4)**

Variable	Label and description	Source var. for gen. in wave 3
erwerb2	*Employment status, generated (all waves)* Integrated employment status variable, harmonised for the first wave. The *erwerb* variable created in the first wave could not be continued due to the changeover to employment biographies as of the second wave. A new status variable was therefore created which, for the 1st wave, is based on the previous employment status variable *erwerb* and, as of the 2nd wave, was generated based on the economic inactivity status (including responses to open-ended questions), the status of school pupil/student/trainee, the current working hours and the spell-related information on currently held jobs.	*PB0100; arbzeit; nichtew2 (PENDDAT) ET0601 (et_spells)*
	The basis for generating the variable is the information from the relevant spell dataset of the respective wave as to whether a certain type of spell is currently ongoing. In the case of a currently ongoing spell of economic inactivity from the gap dataset, the type of inactivity is identified via the *LU0101* variable (i.e. incl. information from open-ended survey questions). In the context of harmonisation, categories 2 ("unemployed") and 3 ("job-creation measure, one-Euro-job" etc.) of the 1st wave are combined to a joint category 2 ("unemployed"). The previous categories 8 ("apprenticeship/training/further training/retraining) and 10 ("student") were also merged into one category for the purpose of integration with the data as of wave 2 (neu)	
famstand	*Marital status, generated* Generation of an integrated marital status variable from the personal questionnaire and the *epartner* control variable generated from the household dataset (neu)	*epartner; PD0500; PD0700 (PENDDAT)*
gebhalbj	*Half-year of birth, generated* This variable indicates whether the date of birth is in the first or second half of the year of birth (neu)	Information on month of birth
hhalg2	*Control variable: current receipt of UB II* This variable indicates that the household is receiving Unemployment Benefit II at the time of the HH interview (neu)	*HA0250b (HHENDDAT) AL20400; AL20500 (alg2_spells)*
kindzges	*Total number of own children (living in and outside the HH), generated* Total number of 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)*

Table 19: **Simple generated variables for wave 3 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 5)**

Variable	Label and description	Source var. for gen. in wave 3
kindzihh	*Number of own children in the household, generated* Variable generated based on the responses in the household questionnaire concerning the number of children that a person in the household has (total number of persons in the household (half) matrix who count as children of the respondent plus the number of persons 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)
mberuf1	*highest vocational qualification attained by the mother, incl. mother in the household, excl. open info., gen.* In the first wave, the question on the mother's vocational qualification was only asked if the mother was not living in the survey household. If she was living in the household, the information on her vocational qualification was taken from her personal interview. As of the second wave the question on the mother's vocational qualification was asked of all newly interviewed individuals, irrespective of whether the mother was living in the household or not. For people taking part in a repeat interview as of the second wave the values were taken over from the generated variable *mschul1* from the previous wave (uv)	In case of first-time interview: *PSH0300a-i* (*PENDDAT*) After first-time interview: *mberuf1* from previous wave (*PENDDAT*)
mberuf2	*highest vocational qualification attained by the mother, incl. mother in the household, incl. open info., gen.* Like *mberuf1* apart from the fact that responses to open-ended survey questions were also taken into account for the generation of *mberuf2* (uv)	In case of first-time interview: *PSH0301a-i* (*PENDDAT*) After first-time interview: *mberuf2* from previous wave (*PENDDAT*)
mhh	*Control variable: mother living in HH* Variable indicating that the respondent's natural mother, stepmother, adoptive mother or mother of non-specified status is living in the household (neu)	Information on relationships between household members (household grid)

Table 19: **Simple generated variables for wave 3 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 6)**

Variable	Label and description	Source var. for gen. in wave 3
migration	*Respondent's migration background, generated* Generated variable for four categories of migration background: no migration background; personal migration (first generation); migration of at least one parent but no personal migration of the respondent (second generation); migration of at least one grandparent but no personal migration of the respondent or of either parent (third generation) (uv) Note: The concept for generating this variable was revised as of wave 2. To generate the variable in earlier waves, only the information on whether the respondent was born in Germany and on which generation/members of the family moved to Germany was used; now the information on whether a parent/grandparent was born outside Germany and, if applicable, which parent/grandparent, is also used. In order to guarantee a consistent logic across the waves, the variable for the 1st wave was also re-generated.	*In case of first-time interview:* *PMI0100; PMI0700; PMI0800a-f; PMI0900a-f (PENDDAT)* After first-time interview: *migration* from previous wave (*PENDDAT*)
mschul1	*highest general school qualification attained by the mother, incl. mother in HH, excl. info. from open-ended questions, generated* In the first wave, the question on the mother's highest school qualification was only asked if the mother was not living in the survey household. If she was living in the household the information on her highest school qualification was taken from her personal interview (uv) As of the second wave, the question on the mother's highest school qualification was asked of all newly interviewed individuals, irrespective of whether their mother was living in the survey household or not.	In case of first-time interview: PSH0200 (*PENDDAT*) After first-time interview: *mschul1* from previous wave (*PENDDAT*)
mschul2	*highest general school qualification attained by the mother, incl. mother in HH, incl. open info., gen.* Like *mschul1* apart from the fact that responses to open-ended survey questions were also taken into account for the generation of *mberuf2* (uv)	In case of first-time interview: PSH0201 (*PENDDAT*) After first-time interview: *mschul2* from previous wave (*PENDDAT*)
mstib	*Mother's occupational status, code number, generated* Detailed occupational status of mother, generated from the individual variables (uv)	In case of first-time interview: PSH0320; PSH0330; PSH0340; PSH0360; PSH0370; PSH0380 (*PENDDAT*) After first-time interview: *mstib (PENDDAT)*

Variable	Label and description	Source var. for gen. in wave 3
netto	*Net income incl. categorised information, generated* Generation of an integrated variable from categorised and open-ended survey questions on net income (neu) Note: The variable was generated for all waves but is currently not evaluable in the longitudinal section (see Chapter 4.2).	*PEK0700b; PEK0800; PEK0900; PEK1000; PEK1100; PEK1200* (*PENDDAT*)
nettokat	*Categorised net income, generated* Aggregation of the categorised information on net income, combined from several items on income categories (neu) Note: The variable was generated for all waves but is currently not evaluable in the longitudinal section (see Chapter 4.2).	*PEK0800; PEK0900; PEK1000; PEK1100; PEK1200* (*PENDDAT*)
nichterw	*Status: economic inactivity, generated (all waves)* Integrated variable for the respondent's status of economic inactivity. Generated from the *PET0800* variable for wave 1. As of *wave 2* generated from information regarding the type of the current economic inactivity from the gap module (*LU0100*, i.e. not taking into account the responses to open-ended survey questions) and information from the unemployment module regarding ongoing unemployment (neu)	*LU0100* (*lu_spells*); *censored* (al_*spells*); *PET0151; PET0911* (*PENDDAT*); indicator of cases for which no gap status was surveyed mistakenly
nichtew2	*Status: economic inactivity, generated, incl. information from open-ended survey questions (all waves)* Integrated variable for the respondent's status of economic inactivity. The responses to open-ended questions were also taken into account when generating *nichtew2*. Generated from the *PET0801* variable for wave 1. As of *wave 2* generated from information regarding the type of the current economic inactivity from the gap module (*LU0101*, i.e. taking into account the responses to open-ended survey questions) and information from the unemployment module regarding ongoing unemployment (neu)	*LU0101* (*lu_spells*); *censored* (al_*spells*); *PET0151; PET0911* (*PENDDAT*); Indicator of cases for which no gap status was surveyed mistakenly
palter	*Age (from p1), generated* Respondent's age, generated based on the date of birth and the date of the personal interview in the current wave (neu)	*p1; pintjahr, pintmon, pinttag* (*PENDDAT*)
panel	*Willingness to participate in panel* (neu)	Information supplied by the survey institute regarding the households' willingness to participate in the panel

Table 19: **Simple generated variables for wave 3 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 8)**

Variable	Label and description	Source var. for gen. in wave 3
pintdat	*Date of personal interview* *Generated variable with the date on which the personal interview was conducted in the form YMMDD (neu)*	*pintjahr, pintmon, pinttag* (*PENDDAT*)
schul1	*Highest school qual., excl. foreign qual's and open info.* Variable for the highest general school qualification; equivalent eastern and western German qualifications were combined (e.g. EOS and Abitur); excl. information from open-ended questions (fs)	In case of first-time interview: *PB0200; PB0220; PB0230; PB0300; PB0400* (*PENDDAT*) In case of repeat interview: *schul1 from previous wave*; *PB0200; PB0220; PB0230; PB0300; PB0400* (*PENDDAT*)
schul2	*Highest general school qual., incl. foreign qual's and open info.* Like *schul1* with the following differences: 1. Inclusion of responses to open-ended questions; 2. Inclusion of information on foreign qualifications (fs)	In case of first-time interview: *PB0200; PB0220; PB0231; PB0300; PB0401* (*PENDDAT*) In case of repeat interview: *schul2 from previous wave;* *PB0200; PB0220; PB0231; PB0300; PB0401* (*PENDDAT*)
schulabj	*Year in which highest school qual. was gained* Year in which respondent gained his/her highest school qualification (fs) Note: Re-interviewed respondents for whom information on the highest school qual. was already available from a previous wave were not asked in the current wave about the year when this qualification was gained if they had gained a new qualification since the previous wave. In this case the year in which the qualification was gained was estimated depending on the month and year of the interview. If the third wave interview was conducted before May 2009, it was assumed that the qualification was gained in 2008, if the interview was conducted later than May, the qualification was assumed to have been gained in 2009.	In case of first-time interview: *PB0220; PB0230; PB0400; PB0410; ;pintjahr; pintmon* (*PENDDAT*) In case of repeat interview: *schulabj* from previous wave; *PB0220; PB0230; PB0400; PB0410; pintjahr; pintmon* (*PENDDAT*)
stib	*occupational status, code number, generated* Generation of the detailed code number for occupational status from the individual variables. Generation of the variable using information from the employment module (*ET0601-ET1201*). 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 number of hours, the one that began first was selected (neu)	*ET0500; ET0601 ET0701; ET0801; ET0901; ET1001; ET1101; ET1201* (*et_spells*)

Variable	Label and description	Source var. for gen. in wave 3
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 about the first regular employment was combined with information from the employment spells if the respondent had already reported his/her first regular employment during the questions on employment spells since January 2006 (uv)	*In case of first-time interview:* PET3300b; PET3000; PET3400; PET3500; PET3600; PET3700; PET3800; PET3900 (*PENDDAT*) ET0601; ET0701; ET0801; ET0901; ET1001; ET1101; ET1201 (*et_spells*) After first-time interview: *stibeewt* from previous wave (*PENDDAT*)
stibkz	*Current occupational status, simple classification, harmonised (anonymised)* Gen. of the simple code number for occupational status from the individual variables (neu)	PET1510 (*PENDDAT*)
stiblewt	*Occupational status, last employment, code number, generated* Detailed code number of the occupational status in the respondent's last employment. To generate this variable, information from the employment spells was combined with information on the last job if the respondent had been out of work since Jan. 2006 (fs)	*In case of first-time interview:* PET1210b; PET1210; PET1220; PET1230; PET1240; PET1250; PET1260; PET1270 (*PENDDAT*) ET0601; ET0701; ET0801; ET0901; ET1001; ET1101; ET1200 (*et_spells*) In case of repeat interview: *stiblewt* from previous wave (*PENDDAT*); ET0601; ET0701; ET0801; ET0901; ET1001; ET1101; ET1200 (*et_spells*)
vberuf1	*highest vocational qualification attained by the father, incl. father in the household, excl. open info., gen.* Generation of variable for father's highest vocational qualification analogous to *mberuf1* (uv)	In case of first-time interview: PSH0600a-i (*PENDDAT*) After first-time interview: *vberuf1* from previous wave (*PENDDAT*)
vberuf2	*highest vocational qualification attained by the father, incl. father in the household, incl. open info., gen.* Generation of variable for father's highest vocational qualification (incl. information from open-ended survey questions) analogous to *mberuf2* (uv)	In case of first-time interview: PSH0601a-i (*PENDDAT*) After first-time interview: *vberuf2* 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 father, incl. father in household, excl. open info., gen.* Generation of variable for father's highest school qualification analogous to *mschul1* (uv)	In case of first-time interview: PSH0500 (*PENDDAT*) After first-time interview: *vschul1* from previous wave (*PENDDAT*)

Variable	Label and description	Source var. for gen. in wave 3
vschul2	*Highest school qualification attained by father, incl. father in household, incl. open info., gen.* Generation of variable for father's highest general school qualification (incl. information from open-ended survey questions) analogous to *mschul2* (uv)	In case of first-time interview: PSH0501 (*PENDDAT*)

After first-time interview: *vschul2* from previous wave (*PENDDAT*) |
| *vstib* | *Father's occupational status, code number, generated* Detailed occupational status of father, generated from the individual variables (uv) | In case of first-time interview: PSH0620; PSH0630; PSH0640; PSH0660; PSH0670; PSH0680 (*PENDDAT*)

After first-time interview: *vstib* from previous wave (*PENDDAT*) |

Table 20: **Simple generated variables for wave 3 in the spell dataset for Unemployment Benefit II (*alg2_spells*) (in the same order as in the dataset)**

Variable	Label and description	Source var. for gen. in wave 3
bmonat	*Spell of UB II: starting month, generated* Month in which the spell of Unemployment Benefit II began. To generate the variable, if information was only available on the season when a spell started, it was converted into a definite month.

Note: The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent have been included in the source variables since the 2nd wave. 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 | AL20100 (*alg2_spells*) |
| *bjahr* | *Spell of UB II: starting year, generated* Year in which the spell of UB II receipt started.

Note: see *bmonat* | AL20200 (*alg2_spells*) |
| *emonat* | *Spell of UB II: ending month, generated* Month in which the spell of UB II receipt ended. To generate the variable info. the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the household was interviewed) the interview month was entered.

Note: see *bmonat* | AL20300 (*alg2_spells*) *hintmon* (*HHENDDAT*) |

Table 20: **Simple generated variables for wave 3 in the spell dataset for Unemployment Benefit II (*alg2_spells*) (in the same order as in the dataset) (continued 1)**

Variable	Label and description	Source var. for gen. in wave 3
ejahr	*Spell of UB II: ending year, generated* Year in which the spell of Unemployment Benefit II receipt ended. In the case of right-censored spells (i.e. spells that were still ongoing when the household was interviewed) the interview year was entered. Note: see *bmonat*	*AL20400* (*alg2_spells*) *hintjahr* (*HHENDDAT*)
alg2kbma - alg2kbmf	*UB II: 1st benefit cut: starting month, generated* Month in which the reduction of Unemployment Benefit II began. To generate the variable information on the season was converted into a definite month. Note: The UB II cuts are embedded in the spells of UB II receipt. The information on the individual benefit-cut spells can be distinguished via the indicator at the end of the respective variable (a-f). 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 the second wave.	1st benefit cut: *AL21000a* (*alg2_spells*) to 6th benefit cut: *AL21000f* (*alg2_spells*)
alg2kbja – alg2kbjf	*UB II: 1st benefit cut: starting year, generated* Year when Unemployment Benefit II cut began. Note: see *alg2kma - alg2kbmf*	1st benefit cut: *AL21100a* (*alg2_spells*) to 6th benefit cut: *AL21100f* (*alg2_spells*)
alg2kema - alg2kemf	*UB II: 1st benefit cut: ending month, generated* Month in which the Unemployment Benefit II cut ended. To generate the variable information the season was converted into a definite month. If the respondent reported a duration for the benefit cut, this was used to calculate the end date of the benefit cut based on the generated start date. Note: see *alg2kma - alg2kbmf*	1st benefit cut: *alg2kbma; alg2kbmja; AL21200a; AL21201a; AL21202a* (*alg2_spells*) to 6th benefit cut: *alg2kbmf; alg2kbmjf; AL21200f; AL21201f; AL21202f* (*alg2_spells*)
alg2keja - alg2kejf	*UB II: 1st benefit cut: ending year, generated* Year when Unemployment Benefit II cut ended. If the respondent reported a duration for the benefit cut, this was used to calculate the end date of the benefit cut based on the generated start date. Note: see *alg2kma - alg2kbmf*	1st benefit cut: *alg2kbma; alg2kbmja; AL21200a; AL21201a; AL21202a* (*alg2_spells*) to 6th benefit cut: *alg2kbmf; alg2kbmjf; AL21200f; AL21201f; AL21202f* (*alg2_spells*)

Table 20: **Simple generated variables for wave 3 in the spell dataset for Unemployment Benefit II (*alg2_spells*) (in the same order as in the dataset) (continued 2)**

Variable	Label and description	Source var. for gen. in wave 3
AL22150a to AL22150f	*UB II: benefit cut: which HH member's benefit was cut, gen.* This variable contains coded information about which HH members' Unemployment Benefit II was cut. It is a string variable with 15 positions. Starting from the left, each position of this variable stands for the position of one person in the household grid. The first position of the variable, for example, indicates whether the benefit was cut for the first person in the HH in the particular benefit cut spell, the second position shows whether the second person's benefit was cut and so on. Since the source information for the generation was only collected from the 2nd wave onwards, all 15 positions of the question variable are given the code "I" (item not surveyed in wave) for all benefit cuts reported in the first wave (see below). Each of the 15 positions of the variable, which stands for one of a maximum of 15 individuals in the household structure, is given one of the following codes indicating that person's benefit-cut status. Codes: 1 - the household member's UB II was cut 2 - the household member's UB II was not cut W - don't know K - not specified T - not applicable (filter) F - question mistakenly not asked U - implausible value I - item not recorded in wave	Information about which household member's benefit was cut in the particular benefit cut spell (HH102 in the household questionnaire for re-interviewed households; HH53 in the household questionnaire for split-off households and new sample households).

Table 20: **Simple generated variables for wave 3 in the spell dataset for Unemployment Benefit II (*alg2_spells*) (in the same order as in the dataset) (continued 3)**

Variable	Label and description	Source var. for gen. in wave 3
zensiert	*Spell of UB II: spell ongoing at time of last HH interview (right-censored.), generated* The censoring indicator shows whether a spell was still ongoing at the time of the last household interview. Note: A spell is regarded as censored if one of the following conditions is met: (a) It is a censored spell of a household from one of the previous waves which was not re-interviewed in the subsequent waves up to the current wave. (b) A household surveyed in wave 3 reports in H91/H93 (HHalt) / H48/H50 (HHneu) that a spell of UB II is still ongoing at the time of the interview in wave 3. Or in H91/H93 (HHalt) / H48/H50 (HHneu) an end date is reported which is identical to the interview date in wave 3, and it is confirmed in the follow-up question in H94 (HHalt) / HH51 (HHneu) that the benefit receipt is still currently ongoing. Code -5 was given if the household reference person of the previous wave was no longer living in the HH in wave 3 and was not interviewed in wave 3.	*AL20100; AL20500 (alg2_spells)*

Table 21: **Simple generated variables for wave 3 in the employment spell dataset (*et_spells*) (in the same order as in the dataset)**

Variable	Label and description	Source var. for gen. in wave 3
bmonat	*Occupation: starting month, generated* Month in which the employment spell began. To generate the variable information the season was converted into a definite month. Note: The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent are included in the source variables. Details regarding the season in which the spell 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 the year → December	*ET0100 (et_spells)*
bjahr	*Occupation: starting year, generated* Year in which the employment spell began. Note: see *bmonat*	*ET0200 (et_spells)*

Table 21: **Simple generated variables for wave 3 in the employment spell dataset (*et_spells*) (in the same order as in the dataset) (continued)**

Variable	Label and description	Source var. for gen. in wave 3
emonat	*Occupation: ending month, generated* Month in which the employment spell ended. To generate the variable information the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the person was interviewed) the interview month was entered. Note: see *bmonat*	ET0300; ET0500 (*et_spells*) *pintmon* (*PENDDAT*)
ejahr	*Occupation: ending year, generated* Year in which the employment spell ended. For right-censored spells (i.e. spells that were still ongoing when the HH was interviewed) the interview year was entered. Note: see *bmonat*	ET0400; ET0500 (*et_spells*) *pintjahr* (*PENDDAT*)
zensiert	*Occupation: spell still ongoing (right censoring)* The censoring indicator shows whether a spell was still ongoing at the time of the personal interview in the last wave, i.e. whether it is a right-censored spell. Note: A spell is regarded as censored if one of the two following conditions is met: The person reports in question P42 concerning the end date of the employment spell that the employment is still ongoing on the interview date (P42 end = 0). Or in P42 an end date is reported which is identical to the interview date, and it is confirmed in the follow-up question P43 that the employment spell is still currently ongoing. Additional employment spells reported in the gap module and corrected dates were taken into account before generating the variable.	ET0300; ET0400; ET0500 (*et_spells*)
stib	*occupational status, code number, generated* Generation of the detailed code number for occupational status from the individual variables.	collection of spell information in wave 3 ET0601; ET0701; ET0801; ET0901; ET1001; ET1101; ET1201 (*et_spells*) Otherwise, the value of the previous wave remains in place
arbzeit	*weekly hours of work incl. details in the case of irregular working hours, gen.* Integrated variable on weekly hours of work in the job held by the respondent, combining responses to open-ended questions on working hours and the categorical follow-up question. For the closed categories of the follow-up question the mean values for the categories were used, for the open-ended category (40 or more hours) the median of the weekly working hours reported in the open-ended questions was used.	*collection of spell information in wave 3* ET2101; ET2201 (*et_spells*) Otherwise, the value of the previous wave remains in place

Table 22: **Simple generated variables for wave 3 in the unemployment spell dataset (*al_spells*) (in the same order as in the dataset)**

Variable	Label and description	Source var. for gen. in wave 3
bmonat	*Registered unemployment: starting month, generated* Month in which the spell of registered unemployment began. To generate the variable information the season was converted into a definite month. Note: The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent are included in the source variables. Details regarding the season in which the spell 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 the year → December	*AL0100* (*al_spells*)
bjahr	*Registered unemployment: starting year, generated* Year in which the spell of registered unemployment began. Note: see *bmonat*	*AL0200* (*al_spells*)
emonat	*Registered unemployment: ending month, generated* Month in which the spell of registered unemployment ended. To generate the variable information the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the person was interviewed) the interview month was entered. Note: see *bmonat*	*AL0300; AL0500* (*al_spells*)
ejahr	*Registered unemployment: ending year, generated* Year in which the spell of registered unemployment ended. For right-censored spells (i.e. spells that were still ongoing when the HH was interviewed) the interview year was entered. Note: see *bmonat*	*AL0400; AL0500* (*al_spells*)
alg1bm	*Receipt of UB I: starting month, generated* Month in which the spell of UB I receipt began. To generate the variable information the season was converted into a definite month. Note: Periods of receipt of UB I are embedded in the spells of registered unemployment. A maximum of one period of UB I receipt is available per period of registered unemployment. The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent are included in the source variables. Conversion of the month details, see *bmonat*.	*AL0800* (*al_spells*)

Table 22: **Simple generated variables for wave 3 in the unemployment spell dataset (*al_spells*) (in the same order as in the dataset) (continued 1)**

Variable	Label and description	Source var. for gen. in wave 3
alg1bj	*Receipt of UB I: starting year, generated* Year in which the spell of Unemployment Benefit I receipt began. Note: see *alg1bm*	*AL0900 (al_spells)*
alg1em	*Receipt of UB I: ending month, generated* Month in which the spell of Unemployment Benefit I receipt ended. To generate the variable information the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the person was interviewed) the interview date was entered. Note: see *alg2kma - alg2kbme*	*AL1000; AL1200 (al_spells)* *pintmon (PENDDAT)*
alg1ej	*Receipt of UB I: ending year, generated* Year in which the spell of Unemployment Benefit I receipt ended. In the case of right-censored spells (i.e. spells that were still ongoing when the person was interviewed) the interview date was entered. Note: see *alg2kma - alg2kbme*	*AL1100; AL1200 (al_spells)* *pintjahr (PENDDAT)*
alg1akt	*Receipt of UB I: spell still ongoing (right censoring)* The censoring indicator shows whether the spell of Unemployment Benefit I receipt was still ongoing at the time of the personal interview in the last wave, i.e. whether it is a right-censored spell. Note: A spell is regarded as censored if one of the two following conditions is met: The person reports in question P114 concerning the end date of the spell of Unemployment Benefit I receipt that the benefit receipt is still ongoing on the interview date (P114 end = 0). Or he/she reports in P114 an end date, which is identical to the interview date, and it is confirmed in the follow-up question P115 that benefit receipt is still currently ongoing. The variable is generated based on the generated date variables, which are checked for plausibility.	*emonat, ejahr, AL1000; AL1100; AL1200 (al_spells)*

Table 22: **Simple generated variables for wave 3 in the unemployment spell dataset (*al_spells*) (in the same order as in the dataset) (continued 2)**

Variable	Label and description	Source var. for gen. in wave 3
zensiert	*Registered unemployment: spell still ongoing (right censoring)* The censoring indicator shows whether a spell was still ongoing at the time of the personal interview in the last wave, i.e. whether it is a right-censored spell. Note: A spell is regarded as censored if one of the two following conditions is met: The person reports in question P109 concerning the end date of the spell of registered unemployment that he/she is still registered as unemployed on the interview date (P109 end = 0). Or he/she reports in P109 an end date, which is identical to the interview date, and it is confirmed in the follow-up question P110 that the spell of registered unemployment is still ongoing. Additional employment spells reported in the gap module and corrected dates were taken into account before generating the variable.	*AL0300; AL0400; AL0500* (*al_spells*)

Table 23: **Simple generated variables for wave 3 in the gap spell dataset (*lu_spells*) (in the same order as in the dataset)**

Variable	Label and description	Source var. for gen. in wave 3
bmonat	*Spell: starting month, generated* Month in which the spell of economic inactivity began. To generate the variable information the season was converted into a definite month. Note: The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent are included in the source variables. Details regarding the season in which the spell 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 the year → December	*LU0200* (*lu_spells*)
bjahr	*Spell: starting year, generated* Year in which the spell of economic inactivity began. Note: see *bmonat*	*LU0300* (*lu_spells*)

Table 23: **Simple generated variables for wave 3 in the gap spell dataset (*lu_spells*) (in the same order as in the dataset) (continued)**

Variable	Label and description	Source var. for gen. in wave 3
emonat	*Spell: ending month, generated* Month in which the spell of economic inactivity ended. To generate the variable information the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the person was interviewed) the interview date was entered. Note: see *bmonat*	*LU0400; LU0600* (*lu_spells*) *pintjahr* (*PENDDAT*)
ejahr	*Spell: ending year, generated* Year in which the spell of economic inactivity ended. To generate the variable information the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the person was interviewed) the interview date was entered. Note: see *bmonat*	*LU0500; LU0600* (*lu_spells*) *pintjahr* (*PENDDAT*)
zensiert	*Spell: spell still ongoing (right censoring)* The censoring indicator shows whether a spell was still ongoing at the time of the personal interview in the last wave, i.e. whether it is a right-censored spell. Note: A spell is regarded as censored if one of the two following conditions is met: The person reports in question P130 concerning the end date that he/she is still economically inactive at the date of the interview (P130 end = 0). Or he/she reports in P130 an end date, which is identical to the interview date, and it is confirmed in the follow-up question P131 that the status of economic inactivity is still ongoing.	*LU0400; LU0500; LU0600* (*lu_spells*)

Table 24: **Simple generated variables for wave 3 in the employment and training measure spell dataset (*mn_spells*) (in the same order as in the dataset)**

Variable	Label and description	Source var. for gen. in wave 3
bmonat	*Measure: starting month, generated* Month in which the measure of active labour market policy spell began. To generate the variable information the season was converted into a definite month. Note: The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent (except for values identified as implausible when the range of values was checked) are included in the source variables. Details regarding the season in which the spell 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 the year → December	*MN0300 (mn_spells)*
bjahr	*Measure: starting year, generated* Year in which the measure spell began. Note: see *bmonat*	*MN0400 (mn_spells)*
emonat	*Measure: ending month, generated* Month in which the measure spell ended. To generate the variable information the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the person was interviewed) the interview date was entered. If the duration of the measure was reported instead of an end date, then the end date was calculated from the start date and the duration. Note: see *bmonat*	*MN0300; MN0400; MN0500; MN0600; MN0700; MN1100; MN1200 (mn_spells)* *pintjahr (PENDDAT)*
ejahr	*Measure: ending year, generated* Year in which the measure of active labour market policy spell ended. For right-censored spells (i.e. spells that were still ongoing when the person was interviewed) the interview date was entered. If the duration of the measure was reported instead of an end date, then the end date was calculated from the start date and the duration. Note: see *bmonat*	*MN0300; MN0400; MN0500; MN0600; MN0800; MN1100; MN1300 (mn_spells)* *pintjahr (PENDDAT)*
zensiert	*Measure: spell still ongoing (right censoring)* The censoring indicator shows whether a spell was still ongoing at the time of the personal interview in the last wave, i.e. whether it is a right-censored spell. Note: A spell is regarded as censored if the person reports in question P164 that he/she is currently still participating in a measure. (P164=1)	*MN0500 (mn_spells)*

Table 25: **Simple generated variables for wave 3 in the person register dataset (*p_register*) (in alphabetical order)**

Variable	Label and description	Source var. for gen. in wave 3
alter3	*Age of person in wave 3 (2008/2009)* Variable contains the "best" available information regarding a person's age. This is either (a) the age calculated from the date of birth reported in wave 3 or (b) if no date of birth is available from wave 3, then the age reported in the household interview. The information from *alter3* was also taken over into the household dataset and corresponds to the information in *HD0200a* to *HD0200o*. This procedure is consistent with that followed by Infratest. Already during the fieldwork, the age variable in the database was populated with the respective "best" information. During fieldwork, a variable in the database is first populated with the age information according to the household interview. If a personal interview is conducted, this variable in the database is overwritten with the age calculated based on the details given in the personal interview (date of birth, date of personal interview). Both the age details provided in the household dataset and those in the individual dataset are based on this variable of the database. The "best" information regarding the age of a person contained in the household dataset of wave 3 was taken into account in the plausibility check and for the generation of the types of benefit communities and households.	*p1, pintjahr, pintmon, pinttag* (*PENDDAT*) *HD0200a* to *HD0200o* (*HHENDDAT*)
korrsex	*Info. on gender was corrected between survey waves* For individuals who belonged to a sample HH in more than one wave this variable indicates whether the gender was corrected in the household interview.	*HD0100a* to *HD0100o* in all waves (*HHENDDAT*)
lastint	*Survey wave of last interview at individual level* This variable indicates the wave in which the last interview at the individual level was conducted with the person (personal interview or senior citizen's interview).	Personal interviews in all waves (*PENDDAT*)
neuj3	*Year in which person joined current HH, reported in wave 3 (2008/2009)* This variable indicates the year the person joined the household of which he/she is a member in the third wave. Note: Information on the date comes from the wave 3 interview with the re-interviewed household into which the person was born or has moved since the previous wave.	Information on the date at which a person moved into a household. Reported in the household questionnaire for re-interviewed households (HH18, HH37)

Variable	Label and description	Source var. for gen. in wave 3
neum3	*Month in which person joined current HH, reported in wave 3 (2008/2009)* This variable indicates the month the person joined the household of which he/she is a member in the third wave. Note: see *neuj3*	Information on the date at which a person moved into a household. Reported in the household questionnaire for re-interviewed households (HH18, HH37)
wegj3	*Year since which person is no longer living in previous HH, reported in wave 3 (2008/20009)* This variable indicates the year in which the person ceased to be a member of the household of the previous wave. Note: Information on the date comes from the wave 3 interview with the household in which the person was living in the previous wave.	Information on the date at which a person moved out of a household. Reported in the household questionnaire for re-interviewed households (HH8, HH28)
wegm3	*Month since which person no longer living in previous HH, reported in wave 3 (2008/20009)* This variable indicates the month in which the person ceased to be a member of the household of the previous wave. Note: see *wegj3*	Information on the date at which a person moved out of a household. Reported in the household questionnaire for re-interviewed households (HH8, HH28)
zmhh3	*Indicator: personal ID number of target person's mother in HH in wave 3 (2008/2009)* Contains the personal identification number of the mother if she is living in the household. Natural mothers, stepmothers, adoptive or foster mothers, or mothers whose status is not specified are counted as the mother.	Information on relationships between household members in the third wave (household grid)
zparthh3	*Indicator: personal ID number of target person's partner in HH in wave 3 (2008/2009)* Contains the personal identification number of a partner living in the household. Spouses, same-sex registered partners, cohabitees and partners whose status is not specified are counted as a partner.	Information on relationships between household members in the third wave (household grid)
zupanel	*Survey wave in which person joined panel* This variable indicates the wave in which the person was a member of a sample household for the first time.	Information on the people living in the household in all waves (household grid)
zvhh3	*Indicator: personal ID number of target person's father in HH in wave 3 (2008/2009)* Contains the personal identification number of the father if he is living in the household. Natural fathers, stepfathers, adoptive or foster fathers, or fathers whose status is not specified are counted as the father.	Information on relationships between household members in the third wave (household grid)

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

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

	PENDDAT					ET spells	AL spells	MN spells
	Current status	Employment history		Social origin		Employment biography	Unemployment biography	Participation in measures
		last ET	first ET	Mother	Father			
Education / training	berabj, beruf1, beruf2, schulabj, schul1, schul2			mberuf1, mberuf2, mschul1, mschul2	vberuf1, vberuf2, vschul1, vschul2			
Educational classifications	casmin, isced97, bilzeit			mcasmin, misced97, mbilzeit	vcasmin, visced97, vbilzeit			
Information on current status	aktmassn, erwerb2, nichtew2							
Socio-economic position	egp, esec, isei, mps, siops	egplewt, eseclewt, iseilewt, mpslewt, siopslewt	egpeewt, eseceewt, iseieewt, mpseewt, siopseewt	megp, mesec, misei, mmps, msiops	vegp, vesec, visei, vmps, vsiops	egp, esec, isei, mps, siops		
Occupational status	stib, stibkz	stiblewt	stibeewt	mstib	vstib	stib		
Dates of employment		emonlewt, ejhrlewt	begmeewt, begjeewt			bmonat, bjahr, emonat, ejahr		bmonat, bjahr, emonat, ejahr
Information on employment	arbzeit, befrist					arbzeit		
Occupation	isco88, isco88it, kldb_it	iscolewt, iscolewt_it, kldblewt	iscoeewt, iscoeewt_it, kldbeewt	misco, misco_it, mkldb	visco, visco_it, vkldb	isco88, isco88it, kldb_it		
Sector employed in	branche					branche		mnbranche

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

	PENDDAT					ET spells	AL spells	MN spells
	Current status	Employment history		Social origin		Employment biography	Unemployment biography	Participation in measures
		last ET	first ET	Mother	Father			
Income	brutto bruttokat netto nettokat							
Benefit receipt	alg1abez alg1s05 hhalg2 halg2s05 halg2s06						alg1akt	
Household context and marital 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 the individual	gebhalbj palter zpalthh zpsex							
Health	pcs mcs							
General	altbefr fb_vers panel pintdat RegP0100 sample							

4.5 Theory-based constructed variables

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

4.5.1 Individual level

Education in years

<u>Variable name</u>	*bilzeit*
<u>Variable label</u>	Duration of school education and vocational training in years, generated
<u>Source variables</u>	*schul2; beruf2*
<u>Category / dataset</u>	Education / individual-level data
<u>Prepared by</u>	Bernhard Christoph
<u>Explanation</u>	For many statistical models, using a linear variable for education is more appropriate than using a categorical one. For schooling levels, it is fairly easy to convert the categorical information into linear information. The linear value simply corresponds to the time spent at school until attainment of the final school leaving qualification. Care must be taken here, however, to ensure that equivalent qualifications are always allocated identical durations. An upper secondary school leaving certificate, for example, should always be labelled with the same duration, irrespective of whether it was attained after twelve or thirteen years of education. Secondary school qualifications were allocated the following education durations for this variable:
	Lower secondary school leaving certificate; lower secondary school leaving certificate from the former GDR (POS) after completion of grade 8; other lower secondary school leaving certificate: 9 years
	Intermediate secondary school leaving certificate; intermediate secondary school leaving certificate from the former GDR (POS) after completion of grade 10: 10 years
	Entrance qualification for University of Applied Sciences: 12 years
	General qualification for university entrance or subject-specific higher education entrance qualification (incl. EOS – comparable qualification in the former GDR) 13 years
	The situation is different for vocational qualifications. Due to the numerous different ways to gain a vocational qualification and the related potentially large differences in income even for qualifications with comparable training durations, the training duration may not be subjected to a simple one-to-one conversion process. This problem can be avoided by attempting to operationalize the growth in human capital related to a certain vocational qualification (see e.g. Helberger 1988).
	This study uses a similar approach. For the conversion process, only the respondent's highest vocational qualification was considered and the years estimated to represent the human capital growth resulting from this qualification were added to the years of school education.
	Training as a semi-skilled worker: +1 year
	Apprenticeship, vocational school, school for health care occupations: +1.5 years
	Master craftsman's certificate: +3 years
	College of advanced vocational studies: +3 years
	University of Applied Sciences/Bachelor: +3 years
	University/Master's degree: +5 years
	PhD.: +8 years
	Other German qualification: +1.5 years
	Other foreign qualification: +1.5 years
<u>Literature:</u>	Helberger (1988)

Education in years, mother

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

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

Training as a semi-skilled worker:	+1 year
Apprenticeship, vocational school, school for health care occupations:	+1.5 years
Master craftsman's certificate:	+3 years
College of advanced vocational studies:	+3 years
University of Applied Sciences:	+3 years
University:	+5 years
Other German qualification:	+1.5 years
Other foreign qualification:	+1.5 years

Literature:	Helberger (1988)

Education in years, father

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

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

Training as a semi-skilled worker:	+1 year
Apprenticeship, vocational school, school for health care occupations:	+1.5 years
Master craftsman's certificate:	+3 years
College of advanced vocational studies:	+3 years
University of Applied Sciences:	+3 years
University:	+5 years
Other German qualification:	+1.5 years
Other foreign qualification:	+1.5 years

Literature:	Helberger (1988)

CASMIN

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

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

occup. \ school	not surv.	pupil	not asked	not applic.	no details	don't know	no qual.	special needs school	lower sec. school	interm. sec. school	entrance qual. for univ. of app. sci.	upper sec. leaving cert.	other Ger. qual.	other foreign qual.
not surv.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-
implaus. value	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
pupil	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
not asked	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
not applic.	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
no details	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
don't know	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
no qualif.	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
semi-skilled	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
apprent-iceship	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
f-t voc. school	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
health occ. sch.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
master craftsm.	-	-	-	-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
UAS/ bachelor	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
univ./ masters	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
PhD	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
oth. Ger. qual.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
oth. for.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c

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

MCASMIN

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

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

school \ occup.	not surv.	pers. int. missing	parent un-known	not asked	not applic.	no details	don't know	no qual.	special needs schook	lower sec. school	interm. sec. school	entrance qual. for univ. of app. sci.	upper sec. leaving cert.	other Ger. qual.	other for. qual.
not surv.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
implaus. value	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
pers. int. missing	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
parent un-known	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
not asked.	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
not applic.	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
no details	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
don't know	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
no qual.	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
semi-skilled	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
apprent-iceship	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
master craftsm.	-	-	-	-	-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
univ. of appl. sci.	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
univ.	-	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
oth. Ger. qual.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
oth. for. qual.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c

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

VCASMIN

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

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

school / occup.	not surv.	pers. int. missing	parent un-known	not asked	not applic.	no details	don't know	no qual.	special needs schook	lower sec. school	interm. sec. school	entrance qual. for univ. of app. sci.	upper sec. leaving cert.	other Ger. qual.	other for. qual.
not surv.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
implaus. value	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
pers. int. missing	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
parent un-known	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
not asked.	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
not applic.	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
no details	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
don't know	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
no qual.	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
semi-skilled	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
apprent-iceship	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
master craftsm.	-	-	-	-	-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
univ. of appl. sci.	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
univ.	-	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
oth. Ger. qual.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
oth. for. qual.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c

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

ISCED 97

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

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

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

school occup.	not surveyed	pupil	not asked	not applic.	no details	don't know	no qual.	special needs school	lower sec. school	interm. sec. school	entrance qual. for univ. of app. sci.	upper sec. leaving cert.	other German qual.	other foreign qual.
not surveyed	-10	-	-	-	-	-	-	-	-	-	-	-	-	-
implaus. value	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
pupil	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
not asked	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
not applic.	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
no details	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
don't know	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
no qual.	-	-	-	-3	-2	-1	1	1	2	2	3a	3a	2	2
semi-skilled	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
apprenticeship	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
full-time voc. sch.	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
health occ. sch.	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
master craftsm.	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
BA	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
UAS/ bachelor	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
univ./ masters	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Ph.D.	-	-	-	6	6	6	6	6	6	6	6	6	6	6
oth. Ger. qual.	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
other foreign	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2

Literature:	BMBF (2003); OECD (1999)

MISCED 97

Variable name	*misced97*
Variable label	Education of mother classified acc. to isced97, updated version, generated
Source variables	*mschul2; mberuf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	For the theoretical background and generation details, see ISCED-97. In contrast to the ISCED-97 coding applied to data on the respondents' education, it is not possible generate ISCED level 6 for data on their parents. This is so, since data on the corresponding qualifications (i.e. PhD or equivalent) were not collected for the parents. Therefore, only ISCED levels 2 to 5 are covered in the coding applied in this dataset. The following table shows the coding details.

school occup.	not surv.	pers. int. missing	parent un-known	not asked	not applic.	no details	don't know	no qual.	special needs school	lower sec. school	interm. sec. school	entrance qual. for univ. of app. sci.	upper sec. leaving cert.	other German qual.	other foreign qual.
not surv.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
implaus. value	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
pers. int. missing	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
parent un-known	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
not asked	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
not applic.	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
no details	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
don't know	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
no qualif.	-	-	-	-	-3	-2	-1	1	1	2	2	3a	3a	2	2
semi-skilled	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
apprent-iceship	-	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
master craftsm.	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
BA	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
univ. of appl. sci.	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
univ.	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
oth. Ger. qual.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
oth. for. qual.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2

Literature:	BMBF (2003); OECD (1999)

VISCED 97

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

school occup.	not surv.	pers. int. missing	parent un-known	not asked	not applic.	no details	don't know	no qual.	special needs school	lower sec. school	interm. sec. school	entrance qual. for univ. of app. sci.	upper sec. leaving cert.	other German qual.	other foreign qual.
not surv.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
implaus. value	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
pers. int. missing	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
parent un-known	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
not asked	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
not applic.	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
no details	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
don't know	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
no qualif.	-	-	-	-	-3	-2	-1	1	1	2	2	3a	3a	2	2
semi-skilled	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
apprent-iceship	-	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
master craftsm.	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
BA	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
univ. of appl. sci.	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
univ.	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
oth. Ger. qual.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
oth. for. qual.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2

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

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

Generated	Occupation	Variable name	Source variables
	current	*isco88*	W1: P46; W2 onwards: P40_X
	Spell data *(et_spells)*	*isco88*	W2 onwards: P40_X
	first	*iscoeewt*	W2 onwards: P40_X, P91, P100
	last	*iscolewt*	W2 onwards: P40_X, P91
	of father	*visco*	W2 onwards: P299
	of mother	*misco*	W2 onwards: P288

Variable label	Current occup.: ISCO 88 (ZUMA coding), generated
	Spell data *(et_spells):* ISCO 88 (ZUMA coding), generated
	first occup.: ISCO 88 (ZUMA coding), first job, generated
	last occup.: ISCO 88 (ZUMA coding), last job, generated
	Father: ISCO 88 (ZUMA coding) of the father, generated
	Mother: ISCO 88 (ZUMA coding) of the mother, generated
Category / dataset	Occupation / individual-level data
Contact person	Bernhard Christoph
Explanation	The International Standard Classification of Occupations (ISCO) was developed by the International Labour Organization (ILO) as an internationally comparative classification. The special feature of the ISCO-88 is that in addition to the job performed, the qualification level generally necessary to perform the job is taken into account when assigning an occupation to a particular occupational code. This constitutes a major difference to the Classification of Occupations provided by the German Federal Statistical Office (KldB), which is also provided in this dataset.
	The actual coding was carried out by the Leibniz Institute for Social Sciences (GESIS, formerly ZUMA). In contrast to the coding variant used by Infratest, this coding of ISCO-88 constitutes an original coding of ISCO-88. It forms the basis for generating the ISCO-based measures of occupational status and prestige.
Literature:	ILO (1990)

International Standard Classification of Occupations 1988 (ISCO88); Infratest coding

Generated	Occupation	Variable name	Source variables
	current	*isco88it*	W1: P46; W2 onwards: P40_X
	Spell data *(et_spells)*	*isco88it*	W2 onwards: P40_X
	first	*Iscoeewt_it*	W2 onwards: P40_X, P91, P100
	last	*iscolewt_it*	W2 onwards: P40_X, P91
	of father	*visco_it*	W2 onwards: P299
	of mother	*misco_it*	W2 onwards: P288

Variable label	Current occup.: ISCO 88 (Infratest coding),generated
	Spell data *(et_spells):* ISCO 88 (Infratest coding),generated
	first occup.: ISCO 88 (Infratest coding), first job, generated
	last occup.: ISCO 88 (Infratest coding), last job, generated
	Father: ISCO 88 (Infratest coding) of the father, generated
	Mother: ISCO 88 (Infratest coding) of the mother, generated
Category / dataset	Occupation / individual-level data
Contact person	Bernhard Christoph
Explanation	The International Standard Classification of Occupations (ISCO) was developed by the International Labour Organization (ILO) as an internationally comparative classification. The special feature of the ISCO-88 is that in addition to the job performed, the qualification level generally necessary to perform the job is taken into account when assigning an occupation to a particular occupational code. This constitutes a major difference to the Classification of Occupations provided by the German Federal Statistical Office (KldB), which is also provided in this dataset.
	The coding was carried out by Infratest, the field institute of PASS for waves 1-3, using a procedure for deriving ISCO-88 codes from the German Federal Statistical Office's Classification of Occupations.
Literature:	ILO (1990)

Classification of Occupations 1992 (KldB92); Infratest Coding

Generated	Occupation	Variable name	Source variables
	current	*kldb_it*	W1: P46; W2 onwards: P40_X
	Spell data *(et_spells)*	*kldb_it*	W2 onwards: P40_X
	first	*kldbeewt*	W2 onwards: P40_X, P91, P100
	last	*kldblewt*	W2 onwards: P40_X, P91
	of father	*vkldb*	W2 onwards: P299
	of mother	*mkldb*	W2 onwards: P288

Variable label	Current occup.: KldB 92 (Infratest coding), generated
	Spell data *(et_spells):* KldB 92 (Infratest coding), generated
	first occup.: KldB 92 (Infratest coding), first occupation, generated
	last occup.: KldB 92 (Infratest coding), last occupation, generated
	Father: KldB 92 (Infratest coding) of the father, generated
	Mother: KldB 92 (Infratest coding) of the mother, generated
Category / dataset	Occupation / individual-level data
Contact person	Bernhard Christoph
Explanation	The KldB92 is the current version of the Classification of Occupations published by the German Federal Statistical Office. It is a classification system that was specifically constructed to match the particularities of the German occupational structure. It is based solely on job descriptions.
	The coding was carried out by Infratest, the field institute of PASS for waves 1-3.
Literature:	StBA (1992).

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

Generated	Occupation	Variable name	Source variables
	current	*egp*	*isco88, stib*
	Spell data *(et_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 occup.: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), current occupation, generated
	Spell data *(et_spells):* Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), generated
	first occup.: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), first occupation, generated
	last occup.: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), last occupation, generated
	Father: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), occupation of father, generated
	Mother: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), occupation of mother, generated
Category / dataset	socio-economic position / individual-level data
Prepared by	Bernhard Christoph
Explanation	The class scheme developed by Erikson, Goldthorpe and Portocarrero (Erikson et al. 1979, 1982; Erikson & Goldthorpe 1992) is one of the most common instruments for operationalising class position.
	For this variable, data are coded exclusively based on the ISCO-88 occupational classification and the occupational status. The coding procedure is based on an earlier approach elaborated by Christoph et al. (2005), where a detailed description of the procedure can be found. In contrast to the procedure described by Christoph et al., here unpaid family workers were not coded as self-employed persons but as persons in dependent employment in accordance with the coding applied in the European Socio-Economic Classification (ESeC), which is described in the next section.
	One difference between the EGP codings applied here and the ESeC codings is that in the EGP coding procedure cases were set to "missing" (-7) where the occupational activity seemed to be incompatible with the occupational status (e.g. "directors and chief executives" [ISCO=1210] who reported that they were "employees performing simple duties" [StiB=51]). For reasons of compatibility with the strongly standardised coding procedure that we adopted for this instrument, we did not apply a comparable revision procedure when using EseC codings.
Literature:	Christoph et al. (2005); Erikson & Goldthorpe (1992); Erikson et al. (1982); Erikson et al. (1979):

European Socio-economic Classification (ESeC)

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

Variable label	current occup.: Magnitude-Prestige-Scale, current occupation, generated
	Spell data *(et_spells):* Magnitude-Prestige-Scale, generated
	first occup.: Magnitude-Prestige-Scale, first occupation, generated
	last occup.: Magnitude-Prestige-Scale, last occupation, generated
	Father: Magnitude-Prestige-Scale, occupation of father, generated
	Mother: Magnitude-Prestige-Scale, occupation of mother, generated
Category / dataset	socio-economic position / individual-level data
Contact person	Bernhard Christoph
Explanation	The Magnitude-Prestige-Scale [MPS] (Wegener 1985, 1988) is the only specifically German instrument available so far to operationalise social prestige based on detailed occupation information. It was originally developed for the older 1968 version of the International Standard Classification of Occupations (ISCO-68). Since occupation coding in the study at hand was conducted based on the more recent ISCO-88 classification and the Classification of Occupations (KldB) developed by the Federal Statistical Office, a variant of the scale transferred to ISCO-88 was used (Christoph 2005). The data were merged by the Centre for Survey Research and Methodology (GESIS-ZUMA) as part of the occupational coding procedure.
Literature:	Christoph (2005); Wegener (1985, 1988)

Standard International Occupational Prestige Scale (SIOPS/Treiman Scale)

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

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

International Socio-Economic Index (ISEI)

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

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

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

Generated	Occupation	Variable name	Source variables
	current	*branche*	P61_X
	Spell data *(et_spells)*	*branche*	P61_X
	Spell data *(mn_spells)*	*mnbranche*	P176_X

Variable label	current occup.: Current job: economic sector/industry (WZ2003)
	Spell data *(et_spells):* economic sector/industry (WZ2003), generated
	Spell data *(mn_spells):* measure: economic sector/industry (WZ2003)
Category / dataset	socio-economic position / individual-level data
Contact person	Bernhard Christoph
Explanation	The information from the open-ended survey question about the sector/ industry in which the respondent works was coded based on the 2-digit code in the Classification of Economic Activities of the Federal Statistical Office (WZ2003). At the two-digit level, this classification largely corresponds to the European "Nomenclature générale des Activités économiques dans les Communautés Européennes (NACE)" in revision 1.1.
	The coding was carried out by Infratest, the field institute of PASS for waves 1-3.
Literature:	StaBA (2002); EG (2002)

Physical scale of SF12v2 (SOEP version, NBS)

Variable name	*pcs*
Variable label	physical 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 a short questionnaire derived from SF36 to determine the health-related quality of life. Since 2002, the SOEP surveys the internationally recognised and utilised SF12-indicators (version 2 – SF12v2). The SOEP version, however, deviates in some parts from the original SF12v2 in terms of phrasing, order of questions and layout. For PASS, the SF12 indicators were surveyed analogously to the SOEP.
	The generation of pcs in PASS is based on the SPSS syntax as described in Nübling et al. (2006).
Literature:	Nübling et al. (2006); Andersen et al. (2007)

Psychological scale of SF12v2 (SOEP version, NBS)

Variable name	*mcs*
Variable label	Psychological 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 a short questionnaire derived from SF36 to determine the health-related quality of life. Since 2002, the SOEP surveys the internationally recognised and utilised SF12-indicators (version 2 – SF12v2). The SOEP version, however, deviates in some parts from the original SF12v2 in terms of phrasing, order of questions and layout. For PASS, the SF12 indicators were surveyed analogously to the SOEP.
	The generation of mcs in PASS is based on the SPSS syntax as described in Nübling et al. (2006).
Literature:	Nübling et al. (2006); Andersen et al. (2007)

4.5.2 Variables at the level of the household or benefit community

Equivalised household income, old OECD scale

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

Equivalised household income, modified OECD scale

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

Deprivation Index, unweighted

Variable name	*depindug*
Variable label	Deprivation index, unweighted (items missing for financial reasons; total of unweighted items: 26)
Source variables	*HLS0100a-HLS2600a; HLS0100b-HLS2600b*
Category / dataset	material situation / household-level data
Prepared by	Bernhard Christoph
Explanation	Following a proposal by Ringen (1988), a distinction is usually made in poverty research between a direct and an indirect measurement of poverty. Indirect measurement focuses on the resources available to attain a certain standard of living, in particular the (equivalised household) income. For this reason, this is also referred to as the resource-based approach to measuring poverty.
	In contrast, direct measurement attempts to record the households' actual ownership of goods and tries to determine the extent to which the households cannot afford certain goods or activities which are considered to be relevant, for financial reasons. This is also referred to as the deprivation approach (see e.g. Halleröd 1995).
	According to the general tenor of previous scientific research, the population classified as poor by the resource-based approach is not always identical to that defined by the deprivation approach. In order to define exactly who is to be considered poor in the narrow sense, it has therefore often been suggested to combine the measures of income-related poverty and deprivation and to count only those who are classified as poor by both approaches as belonging to the population living in poverty in the narrow sense (see Halleröd 1995; Nolan & Whelan 1996; Andreß and Lipsmeier 2001).
	The index is based on a list of 26 goods or activities. The households surveyed are asked to indicate whether they possess these goods or participate in the activities mentioned. The unweighted index calculated on this basis simply adds up the number of items which the respondents indicated that they do not possess or do not participate. However, only items which are missing for financial reasons are counted, in order to avoid certain consumer preferences (e.g. a household deliberately doing without a car or a television) being misinterpreted as a reduction in the standard of living.
	Additionally, an item was only accepted as missing for financial reasons if the answers to both questions explicitly confirmed this. "Don't know" or "details refused" answers were evaluated either as if the particular good was available in the household or as if it was missing for a reason other than financial reasons. This assumption is certainly not applicable to all cases. Alternatively, it would have been possible not to calculate an index value for households that failed to answer a question for (at least) one particular good ("istwise deletion"). With respect to the total of 26 goods and activities surveyed, however, this method could quickly have led to a large number of missing index values. For this reason, the first method described was selected. Nevertheless, compared to the listwise deletion procedure, there is a risk of the number of goods missing being underestimated with this method.
Literature:	Andreß & Lipsmeier (2001); Halleröd (1995); Nolan & Whelan (1996); Ringen (1988)

Deprivation Index, weighted

<u>Variable name</u>	*depindg*
<u>Variable label</u>	Deprivation index, weighted (items missing for financial reasons; total of weighted items: 12,8)
<u>Source variables</u>	*HLS0100a-HLS2600a; HLS0100b-HLS2600b; PLS0100-PLS2600*
<u>Category / dataset</u>	material situation / household-level data (weighted at the individual level)
<u>Prepared by</u>	Bernhard Christoph
<u>Explanation</u>	For a general description, see deprivation index, unweighted.

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

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

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

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

Household typology

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

Definition of relationships for generating the household type:

- **Couples:** married couples; registered partnerships; non-married partnerships and partnerships whose status is not further specified (missing value for the follow-up question about the type of partnership

- **Child of a person:** natural child; stepchild; adopted or foster child; child whose status is not further specified (missing value for the follow-up question about type of relationship to the child).

- **Parent of a person:** natural parent: step-parent; adoptive or foster parent: parent whose status is not further specified (missing value in follow-up question about type of parentship).

Definition of household types:

- **One-person household:** Household consisting of only one person

- **Couple without children:** Household consists of two individuals living together as a couple

- **One-parent household:** Household consists solely of one parent and his/her children. No restrictions are made with respect to the children's ages.

- **Couple with children under the age of 16:** Household consists solely of two individuals living as a couple and their respective and/or mutual children. All of the children are under the age of 16.

- **Couple with children aged 16 or over:** Household consists solely of two individuals living as a couple and their respective and/or mutual children. All of the children are aged 16 or over.

- **Couple with children under the age of 16 and children aged 16 or over:** Household consists solely of two individuals living as a couple and their respective and/or mutual children. There are both children under the age of 16 and children aged 16 or over living in the household.

- **Multi-generation household:** Household consists of members of at least three generations in linear succession. The core of the household is multi-generational, i.e. at least one person in the household is both a child and a parent of another member of the household. The other people living in the household are parents, children, siblings, partners of the central member(s) and partners' siblings.

- **Other household type:** Household which could not be assigned to one of the other defined household types.

- **Type generation not possible (missing values):** Basically all households with at least one missing value (-1,-2,-4) or implausible value (-8) in the main category of a relationship variable or the age variable (Exception: For households with three or less members in unambiguous relationship constellations, the household type was generated even if age details were missing.).

<u>Literature:</u>	Beckmann & Trometer (1991); Frick et al. (n.d.); Lengerer et al. (2005); Porst (1984)

Benefit community ID, wave 3

<u>Variable name</u>	*bgnr3*
<u>Variable label</u>	Benefit community ID in wave 3
<u>Source variables</u>	Household information on age and relationships between household members
<u>Category / dataset</u>	Benefit community (Bedarfsgemeinschaft) / person register
<u>Prepared by</u>	Gerrit Müller
<u>Explanation</u>	The *bgnr3* variable is created at the individual level. It assigns an identification number to each household member indicating the person's affiliation to a particular benefit community. Consequently, household members with the same ID constitute a benefit community together. The *bgnr3* variable is composed of the known household number and a two-digit indicator to identify the benefit community within the household.
	The identification of a household member's affiliation to a benefit community is based solely on the information on the relationships between the different household members from the household grid table as well as on the members' ages according to the household interview. The benefit communities identified in this way are, therefore, to be regarded as "synthetic" benefit communities. The identification process does not take into consideration information on actual benefit receipt or on the individual members' ability to work and qualification status. It is more a case of identifying groups of persons in the same household who are or would be regarded as household communities in joint receipt of benefits according to the provisions of the Social Code Book II in the event that they required benefits. This artificial allocation procedure is necessary, since information on the existence of a benefit community and the identification of individuals affiliated to this community cannot be collected directly in the context of an interview.
	With regard to content, the allocation of a person to a benefit community is based on the latest version of the German Social Code Book II, Section 7, Sub-section 3 (last amended on 26 March 2007). According to this, each individual aged between 25 and 64 constitutes a separate benefit community unless this person is living in a partnership and/or has a child/ children aged under 25 who has/have no own partner/children. In the latter case, the benefit community comprises the person, his/her partner and the child(ren). If two individuals live in the same household with a joint child, but do not indicate in the household grid table that they are living in a partnership, a partnership is nevertheless assumed to exist in terms of Section 7, Sub-section (3a), and the corresponding individuals and their child(ren) are assigned to the same benefit community. Individuals aged between 15 and 25 are in principle assigned to their parents unless they are already living together with a partner (or a child of their own) in a joint household. Individuals aged between 15 and 25 who live without their parents (or partner / children) constitute a separate benefit community.
	Persons aged 65 and over are not covered by the Social Code Book II and are therefore not counted as members of a benefit community (code 0) unless they live together with a partner who is aged under 65 (or a child aged under 25) in the same household. Likewise, children under the age of 15 who live in a household without their parents are not counted as members of a benefit community (code 0). They are covered by the provisions of the Social Code Book XII. Allocations to benefit communities were not made for households with missing information on relationships and/or the age of certain household members; instead, all members of these households were assigned code 99. By approximation, such households may be interpreted as households consisting of one benefit community only.
<u>Literature:</u>	German Social Code Book II – basic benefits for job-seekers (Sozialgesetzbuch, Zweites Buch - Grundsicherung für Arbeitssuchende (SGB II))

Benefit community typology, wave 3

Variable name	*bgtyp3*
Variable label	Type of benefit community in wave 3
Source variables	Household information on age and relationships between household members
Category / dataset	Benefit community / person register
Prepared by	Gerrit Müller
Explanation	The benefit community typology is based on the same concept of the synthetic benefit community as was used for variable *bgnr3*. Up to the age of 25, children are counted as members of the benefit community of their parents unless they themselves have a partner or children of their own. This is handled differently from the BA statistics, where typologies are often still established based on majority (18th birthday). As an example: households in which the youngest child is aged between 18 and 24 and which are classified as one-parent benefit communities according to our typology, are counted as single households in the BA statistics. This difference must be borne in mind when comparing PASS data with figures from the official statistics.
	Code 0 (no benefit community) was assigned to households in which one or more member(s) are not covered by the Social Code Book II (see also code 0 for variable bgnr3). Code -5 (generation impossible due to missing values) was allocated to households with missing information on relationships and/or the age of individual household members (see code 99 for bgnr2).
Literature:	–

Benefit community in receipt of unemployment benefit II as of the sampling date, wave 3

Variable name	**bgbezs3**
Variable label	Benefit community in receipt of UB II as of the sampling date in wave 3 (2007/2008)
Source variables	New sample households: *HH49, HH50, HH52, HH53, HH62, sample, hnr, bgnr2, hhgr*
	Re-interviewed households: *HH91, HH92, HH93, HH95, sample, hnr, bgnr2, hhgr*
Category / dataset	Benefit community / person register
Prepared by	Mark Trappmann
Explanation	For each benefit community that was identified in accordance with the procedure described for variable *bgnr3* this variable indicates whether the benefit community was in fact receiving Unemployment Benefit II as of the sampling date of wave 3 or not.
Literature:	–

Benefit Community in Receipt of Unemployment Benefit II as of the Survey Date, wave 3

Variable name	*bgbezb3*
Variable label	Benefit community in receipt of UB II as of the survey date in wave 3 (2007/2008)
Source variables	*AL20601, AL20701, zensiert* (*alg2_spells*), *sample, hhgr, bgnr3*
Category / dataset	Benefit community / person register
Prepared by	Daniel Gebhardt
Explanation	For each benefit community that was identified in accordance with the procedure described for variable *bgnr3* this variable indicates whether the benefit community was in fact receiving Unemployment Benefit II at the survey date of wave 3 or not.
Literature:	–

Number of benefit communities within the household

Variable name	*anzbg*
Variable label	Number of synthetic benefit communities in HH, generated
Source variables	*bgnr3, hnr*
Category / dataset	Benefit community / household dataset
Prepared by	Daniel Gebhardt
Explanation	This variable indicates the number of benefit communities existing in the household. The benefit communities were identified in accordance with the procedure described for the generation of variable *bgnr3*.
Literature:	–

Number of benefit communities in the household actually receiving benefits as of the sampling date

Variable name	*nbgbezug*
Variable label	No. of benefit communities receiving benefits in HH as of sampling date
Source variables	***bgbezs3, bgnr3**, hnr*
Category / dataset	Benefit community / household dataset
Prepared by	Daniel Gebhardt
Explanation	This variable indicates the number of benefit communities within the household which were in receipt of benefits in accordance with the Social Code Book II at the sampling date. The value was calculated by aggregating via the household number the benefit communities within each household which were actually receiving benefits according to the variable *bgbezs3* from the person register.
Literature:	–

5   Data preparation

In wave 3 for the first time not the IAB but infas was responsible for preparing the data[39]. In order to still guarantee the consistency of data preparation in the longitudinal section, infas was provided with the relevant syntax files of the data preparation in wave 2 together with the necessary source and intermediary data sets and a documentation of the individual operations. Important decisions, such as on the correction of structural problems in the participating households or on the integration of spell datasets, were made together with the IAB. The IAB was also available for questions beyond that during the period of data preparation.

The information gathered in the interviews of the 3rd wave is initially available at TNS Infratest in the form of ASCII data. In a first step, TNS Infratest created the following standardised datasets from these raw datasets (see Büngeler et al 2009:71ff.):

- Household dataset for re-interviewed households

- Household dataset for new sample households and split-off households

- Individual dataset (respondents aged 15 up to and including 64 years)

- Gap dataset (information on gaps in the employment biographies of more than three months duration since January 2005)

- Senior citizens' dataset (respondents aged 65 and over)

TNS Infratest conducted a basic check of the operation of the filter questions in these datasets. Questions that were not asked although they should have been were marked with a code. After the datasets had been prepared in this way, they were delivered to infas via the IAB. There the datasets were subjected to the second step of editing comprising further more detailed, formal and content-related checks and were then prepared as the scientific use file. In addition to this, TNS Infratest supplied datasets with information from open-ended survey questions (e.g. on the type of occupational activity), a gross dataset and other special datasets which are not obtained directly from the actual survey instruments.

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

---

[39]   As of wave 4, infas will also take over the field work for PASS. Data preparation as of wave 3 was part of a new invitation to tender, which became necessary due to the contract with TNS Infratest, currently limited to three waves.

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

Table 26: Overview of the steps involved in preparing the data of the 3rd wave of PASS

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

5.1 Structure checks and interviews removed from the dataset

Before the filter checks were carried out in the 3rd wave, a structure check was conducted. Here interviews which are regarded as not successfully surveyed in the sense of PASS were to be identified and were, if necessary, removed from the datasets for this reason. In addition, the structure of the re-interviewed households was compared with the structure reported in the previous wave in order to identify and, if necessary, correct implausible or problematic changes in the household composition and errors in the allocation of the personal interviews to their respective position in the household. For observing the households in the longitudinal section it is essential that the individuals are assigned consistently to their position in the household and that the respondents can be identified clearly across the waves. A definite personal identification number must not be allocated to different individuals in different waves. If the correct household composition was unclear, all of the interviews conducted with this household in the 3rd wave were removed from the dataset. If one of the personal interviews was conducted with the wrong person but without any further problems emerging in the household composition, then just the personal interview was removed.

Different checks were carried out to identify problematic cases:

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

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

- In order to identify households which are regarded as not successfully surveyed in the sense of PASS, the datasets at the household and the individual level were merged. Personal interviews without a full household interview were marked, as were household interviews for which no interview at the individual level was available[40].

- Also moves into and out of the household are another important factor. Panel household for which moves out of the household were reported were inspected regarding their household context and correlated with the realised split households. Evaluations were made as to whether the remaining household context of the panel household is self-evidently plausible. Interviews from panel households in which all household members leave the household, except individual children under 15 years of age, were discarded with regard to the panel household as well as with regard to split-off households. If more than one person moves out, it was checked whether these persons form a joint split-off household or several different ones, and whether this is plausible. Such cases were considered implausible, for instance, where one partner leaves the panel household together with young children, but the persons moving out form several different split-off

---

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

households according to field information, i.e. young children allegedly forming individual households. In case of the non-realisation of the split-off household, the moving out was considered as plausible, but all individuals that moved out were retroactively merged into one joint split-off household.

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

- In panel households that reported a move-out as of the 2nd wave, there can also be moves back in of members formerly belonging to the household as of the 3rd wave. The requirement of recognising these individuals as moving back in and placing them to their former household position instead of assigning them a new household position is a component of the household grid. It was evaluated subsequently whether these requirements were met in the field in all cases. For individuals that were subsequently identified as moving back in based on a comparison of first name, age and gender with the members moved out of the households, the household structure had to be changed. This led to retroactive changes of the personal identification number of the individual to be positioned and also an adjustment in the individual-related information in the household interview, e.g. on childcare or the reasons for a cut in Unemployment Benefit II to the position defined as correct within the framework of the structural check.

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

Individual case decisions were also made to deal with the cases which proved to be problematic during the structure checks. What was of significance here was how serious the

particular problem was considered to be. In cases where the correct household composition in the 3rd wave was unclear, all of the interviews from the 3rd wave were removed. In the 4th wave these households will be treated as households that did not participate in the 3rd wave. If in retroactively removed household interviews moves-out were reported, also the split-off households were discarded. This concerned both the interviews conducted in the current wave in these split-off households and also the sample of the subsequent wave. Split-off households that developed from a discarded interview of a panel household are retroactively classified as not having been conducted and do not count to the panel sample of the subsequent wave. If there was merely a problem in assigning individuals to their respective position in the household, i.e. if it was suspected that a personal interview had been conducted with the wrong person in the 3rd wave, then only the personal or senior citizens' interview concerned was removed. If the problem was a structural problem that had no serious consequences and could be solved, for example, by removing a personal interview, additional corrections of the first name, age and gender were made at the household level. The incorrect information concerned was then put back to the last valid value from the previous wave or in the case of age to the value from the previous wave + the number of years since the last valid interview in this household.

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

Furthermore, TNS Infratest reported with the gross dataset references to households whose interviews were not conducted correctly. This concerned on the one hand household structure problems as described above that became known already before transferring the raw data to the IAB, and on the other hand interviews with technical problems. In these cases, all of the interviews were removed.

The net variables (*hnettok3*, *hnettod3*, *pnettok3*, *pnettod3*) in the household register datasets and person register datasets provide an indication that interviews have been removed. Via the corresponding variables in the household register, it is possible to trace the re-interviewed households whose household interviews were removed later. By means of the net variables in the person register it is possible to trace the cases where only single individual-level interviews or all of the interviews of the household were deleted. In the case of households from the refreshment sample of the 3rd wave without at least one valid household and personal interview it is not possible to trace deleted interviews in the register datasets, as these households were not included in the datasets.

5.2 Filter checks

During the filter checks, the correct operation of the filter questions in the instruments was checked using a statistics program. If certain questions were asked although the value of the relevant filter variable would have required something else (for example, if detailed information was requested on vocational training although the respondent had stated that

he/she did not have a vocational qualification), these variables were set to the missing code "-3" (not applicable), which they would also have received through correct use of the filters.[41] Moreover, some items were not surveyed in individual cases although would have been necessary according to the relevant filter variable (e.g. if no further information was recorded on vocational training although the respondent had stated that he/she had undergone such training). In these cases, the specific missing code "-4" (question mistakenly not asked) was assigned. An assignment of the code "-4" can also be based on the household structure evaluation as described in Chapter 5.1. If the move-out of a person is retroactively discarded as implausible and the person is retroactively classified as still belonging to the former household then this also means that individual-related information on these individuals in the household interview must be coded retroactively as mistakenly surveyed or not surveyed. Thus, the code "-4" does not always refer to a problem in the survey instrument. If the code "-4" is assigned to a question that is relevant for filtering subsequent questions, then the subsequent questions are also coded with "-4" in case these subsequent questions were actually not surveyed. If subsequent questions were, however, surveyed, because, for instance several filter questions link to this subsequent question and another filter question triggered the subsequent question correctly, the value surveyed there remains.

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

 provides an overview of the assigned values. "-1" and "-2" are the standard recoding for the values "don't know" and "details refused" recorded during the survey. "-3" is the general "not

---

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

applicable" code for questions not asked due to filters. As described above, the code "-4" was assigned if a question was not asked as a result of a filter error. Codes "-5" to "-7" are question-specific codes. These can either be 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 over EUR 99,999 in the open question on income). These codes were only assigned as required.

Table 27:   Overview of the missing codes used

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

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

5.3   Plausibility checks

For the plausibility checks an extensive list of theoretically possible contradictions in the respondents' statements was checked. For this the list of checks conducted in the previous wave was adapted and extended for the current wave. In addition, the household structure was checked for plausibility. Furthermore, also the spell data were checked for plausibility – in particular with regard to inadmissible overlaps within the individual spell types. Here in principle only the data gathered in the cross-section of the 3rd wave were checked. No checks were carried out on the longitudinal section, in other words comparing the information provided in the current wave with that given in the previous wave.

In detail, the following steps were carried out:

1. Contradiction check: In general, contradictions were only corrected if either the implausibility could be defined as particularly serious and/or if the alteration was regarded as comparatively minor. The latter applied, for example, if only a small number of cases were affected or if one missing code (e.g. "-3") was simply replaced by another one (e.g. "-8"). Two strategies were used to filter implausible statements: Either the implausible responses were corrected directly or they were allocated a specific missing code.

- Implausible responses were only corrected when it was highly probable that the interviewer had entered information incorrectly. An example of this is a statement of a monthly total rent of € 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 had triggered a filter in subsequent questions, as is the case for the categorial question of income, then the categorial questions were retroactively set to code "-4" (question mistakenly not asked).

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

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

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

   - In a second step checks were carried out comparing sets of three family relationships with one another for plausibility. An example of a relationship structure that would be classified as implausible in this check is: person A is person B's spouse. Person A is the natural parent of person C. Person C is a sibling of person B. If such a combination or another similarly implausible combination of relationships was identified during the plausibility checks, then here, too, an attempt was made to make the relationship plausible based on

the household context. In the case described, the relationship data was corrected by person C being coded as a child of person B whose status was not further specified. The aim is to correct as many of the implausibilities identified as possible in terms of content, since a plausible and complete constellation of relationships is the necessary requirement for generating the benefit community.

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

5.4 Retroactive changes of the 1st and 2nd wave

During the data preparation process for the scientific use file of the 3rd wave, some changes were also made to the waves of PASS, which had already been delivered. These alterations included corrections of errors that were detected after the completion of the scientific use file of the 2nd wave. Table 28 to Table 33:   Overview of retrospective alterations in the weighting datasets (hweights; pweights) give an overview on the retroactive changes in the already delivered waves of PASS.[42]

---

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

Table 28: Overview of retroactive changes in the household dataset (*HHENDDAT*)

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
HW1900	*HHENDDAT*	2	Correction	Receiving housing assistance and Unemployment Benefit II at the same time is not possible. During the plausibility checks, a respective check for households with only one synthetic benefit community was conducted. If a household received both benefits at the same time, both the indicator for the receipt of housing assistance (*HW1800*) and the information on the monthly amount (*HW1900*) should be set to -8 (implausible value). In the data processing of the second wave only the variable *HW1800* was set to -8 in these cases. The information on the amount in *HW1900* remains. This mistake was corrected – for all cases for which the receipt of housing assistance is implausible (*HW1800*) also the monthly amount (*HW1900*) was set to "implausible".
HW0880a-i *HW0890* *HW0900* *HW0910* *HW0920* *HW1000* *HW1100* *HW1200* *HW1300* *HW1400* *HW1500* *HW1600* *HW1700* *HW1800* *HW1900* *HW2000* *HW2100* *einzugj* *umzug*	*HHENDDAT*	2	Correction	H64 (*HW0880a-i*), item I, surveyed the reason for a move (open-ended question). These responses were coded (*HW0881a-j*). If it became apparent during coding that there was no move and the question for the reasons of the move was asked mistakenly, the control variable *umzug was corrected.* The variables depending directly or indirectly from this variable were then again filter checked based on the corrected variable. Due to a mistake, these corrections of the control variable *umzug* and of H64 to H80 (HHalt) were not included in the dataset. This mistake was corrected. The control variable *umzug* corrected based on the open-ended responses and the other corrected variables depending on it in the entry filter are now included in the household dataset.

Retroactive changes of the 1st and 2nd wave

: **Overview of retrospective alterations in the household dataset (*HHENDDAT*) (continued)**

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
HEK1600	*HHENDDAT*	2	Correction	During the plausibility check, the number of children for whom the household receives child benefit should be checked against the number of children living in the household + the number of children living outside of the household. Implausible information was coded with the value "-8" during data preparation. Since not all necessary information for the checks is surveyed anymore in the current wave, a decision was made not to make the checks analogously to wave 1 as of wave 2. As of wave 2, only cases in *HEK1600* (number of children for whom the household receives child benefit) are set to -8 that indicate in the filter question *HEK1500* that they receive child benefit but then indicate in *HEK1600* the receipt for"0" children.
HEK1810	*HHENDDAT*	2	Correction	*HEK1810* surveyed the amount of advance child maintenance payment, which a household can receive for children under the age of 15. Due to a mistake in the plausibility check of this filter question, cases that indicated in the filter question as to whether the household even received advance child maintenance payment (*HEK1800*) that they did not receive such payment, or who were not asked this question according to the filter were mistakenly set to -8 (implausible value) in *HEK1810*. This mistake was corrected – cases which indicate in *HEK1800* that they do not receive advance child maintenance payment, or which were not asked this question were set to -3 (not applicable) in the question about the monthly amount *(HEK1810)*.
depindg	*HHENDDAT*	1,2	Correction	Besides an unweighted deprivation index, the dataset also provides a preference weighted version of the deprivation index. The preferences for weighting were used for personal interviews in wave 1 and then for both wave 1 and wave 2. To obtain representative weights for the overall population, it does, however, not suffice only to consider the preferences surveyed in the sample. This mistake was corrected – the preferences surveyed were initially projected to the overall population. These preference weights representative for the overall population were then used to weight the deprivation index for wave 1 and wave 2 again.

Table 29:   Overview of retrospective alterations in the individual dataset (*PENDDAT*)

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
PET0700	*PENDDAT*	2	Correction	The senior citizens' questionnaire surveyed whether the current employment is marginal or not (*PET0500*). The working hours of senior citizens in marginal employment and senior citizens with a different type of employment were assigned to two different variables (*PET0700* for marginal employment, *PET1300* for other types of employment). In the coding of *PET0700,* a mistake was made in assigning the missing codes. Senior citizens without marginal employment were mistakenly coded to -10. This mistake was corrected. The senior citizens who were not in marginal employment at the survey date of wave 2 are now coded in *PET0700* with -3.
PEK0100 *PEK0100a* *PEK0100b* *PEK0700* *PEK0700a* *PEK0700b* *PEK1300* *PEK1360a* *PEK1360b* *PEK1415* *PEK1425* *PEK1435* *PEK1445* *PEK1455* *PEK1500* *PEK1700* *PEK1900* *PEK2100*	*PENDDAT*	2	Correction	The open-ended responses to income in the individual dataset partly contain a special code for income exceeding a certain amount (top coding). Depending on which income variable is concerned, top coding is contained or not. If top coding is contained, its amount varies (e.g. "more than EUR 99,999" or "more than EUR 9,999"). For some income variables, wrong variable labels were assigned regarding the top coding. This mistake was corrected. New value labels were created and the variables were assigned with the respectively correct label.

Table 29: Overview of retrospective alterations in the individual dataset (*PENDDAT*) (continued 1)

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
PEK1600 *PEK1700* *PEK1800* *PEK1900* *PEK2000* *(neu)* *PEK2100* *(neu)*	*PENDDAT*	2	Correction	In wave 1, the receipt of education/parenting benefit was surveyed in the personal interview. *PEK1600* contains the response to the filter question whether the individual receives these benefits. *PEK1700* contains the information on the monthly amount of receipt. As of wave 2, the receipt of education/parenting benefit is no longer surveyed on the individual level but within the household interview (*HEK1610*, *HEK1620*). Furthermore, the spectrum of income components surveyed was expanded significantly as of wave 2. As of wave 2, the personal interviews for example include the question whether there is a receipt of BaföG/training allowance/student grant and if so, what the monthly amount is. This information was mistakenly not assigned to new variables but stored in *PEK1600* and/or *PEK1700* in wave 2. This mistake was corrected – *PEK1600* and *PEK1700* were filled for the observation of wave 2 with -9 (item not recorded in wave). To observe the correct order of questions, initially two variables had to be renamed which were also only surveyed in wave 2: *PEK1800* became *PEK2000* (indicator for receipt of government payments for employed persons), *PEK1900* became *PEK2100* (amount of monthly government payments for employed persons). The information whether training allowance (or a similar benefit) is received was assigned to *PEK1800*. The monthly amount of training allowance payments was stored in *PEK1900*.
PSK0400a *PSK0400b* *PSK0400c* *PSK0400d* *PSK0400e*	*PENDDAT*	2	Correction	In wave 1 and 2, *PSK0400a-e* was a multiple-choice question. The values of "don't know" and "details refused" given globally for all items are included in all individual items together with the special code "no, not actively". For wave 2, the information from the special code "no, not actively" was mistakenly not transferred to the individual items. This mistake was corrected – if the respondent indicated that he/she was not active, all individual items were set to the respective special code (-5) as in wave 1.

Table 29: **Overview of retrospective alterations in the individual dataset (*PENDDAT*) (continued 2)**

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
PAS0900a-f *PAS0901a-f* *PAS1000a-d* *PAS1110* *PAS1200* *PAS1300* *PAS1400a-f* *PAS1600* *PAS1800* *PAS1900* *PAS2000* *PAS2100* *PAS2200* *PAS2300*	*PENDDAT*	2	Correction	In *PAS0100,* employed persons are interviewed on job-seeking. Besides the three read out categories (1-3), the question has an additional category (4 "sought both additional and other employment") which, however, was not read out (since it is "below the line"). Since this additional category is a special code, the temporary code "-5" was assigned in the filter checks. *PAS0100* is a major filter variable for job-seeking of employed persons, for which reason *PAS0100* is relevant for controlling the following questions. In a check of entry filters in the following questions, this recoding was, however, mistakenly not considered. This led to mistakes in the filter checks of the subsequent questions. Cases which indicated in *PAS0100* that they were seeking both additional and other employment were mistakenly set to -3 in the indicated variables. This mistake in the filter checks was corrected. The responses by the respondents are now included.

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
PTK0200 *PTK0500* *PTK0900a* *PTK1000a* *PTK1100a*	*PENDDAT*	2	Correction	Wave 1 recorded in the contact to social security institutions module for some items (for the first time for *PTK0200*) the special code "not applicable, not available for the labour market". If the respondent provided a corresponding response, it was adapted in the data preparation in all following items of the contact to social security institutions module. This special code was no longer recorded as of wave 2. Instead, the new items *PTK0310* and *PTK0320** were included, which recorded whether the respondent was seeking work and for what reason he/she might not have to seek work. In the assignment of special codes as of wave 2, it had to be considered that the code -5 cannot be assigned – in wave 1, it was populated with the "not applicable" code. Furthermore, it had to be guaranteed that the special codes are coded analogous to wave 1. Here, mistakes were made in wave 2. The special codes recorded in wave 2 in the variables *PTK0200*, *PTK0500*, *PTK0900a*, *PTK1000a* and *PTK1100a* were not coded beginning with -6 but the special code -5 was used, too. In *PTK0200*, also the special code 997 (never) in wave 2 was not coded to "0" analogous to wave 1. These mistakes were corrected. The coding of the special codes for wave 2 now considers that the code -5 cannot be assigned since wave 1 already determined a meaning for it. Furthermore, the value labels were provided with information that shows which special codes were recorded in which wave.
PTK0321f *PTK0321g*	*PENDDAT*	2	Correction	The 2nd wave recorded reasons why a respondent does not have to seek work. The open responses thus recorded were coded in *PTK0321a-f*. Two new categories, *PTK0321f* and *PTK0321g,* were created here. For these variables a mistake was made in the data preparation in the variable labels – they were exchanged between the two variables. This mistake was corrected. *PTK0321f* (not job-seeking because in training) and *PTK0321g* (not job-seeking because employed) are now labelled correctly.

Table 29: Overview of retrospective alterations in the individual dataset (*PENDDAT*) (continued 4)

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
PTK0321f *PTK0321g*	*PENDDAT*	2	Correction	*PTK0321f* and *PTK0321g* are two new categories which were altered to a multiple choice item during coding of the open responses. Here a mistake was made in the data preparation. The "don't know" or "details refused" responses were not transferred correctly to these newly created variables. This mistake was corrected. If no further details were given in the multiple choice item, the "don't know" or "details refused" responses were also transferred to the variables created during coding of the open responses for the new categories.
PEO0600*	*PENDDAT*	1, 2	Correction	The variables *PEO0600a-o* contain the school qualifications the respondent expects for his/her children. A child's expected school qualification is always stored in the variable corresponding to the position of the child in the household structure of the respective wave. For instance, the information on the first child of the target person is stored in *PE0600c* if the first child comes in the third position of the household structure. The position corresponds to the zplfd (serial number of the individual within the household structure in the respective wave). These variable labels were ambiguous, since they suggested that the information on the first child could be found in the first variable (*PEO0600a*), whereas it is actually stored in the variable corresponding to the child's position in the household structure (in the example *PEO0600c* since the child comes in third position). The variable labels were corrected to give a more unambiguous reference to the assignment of the variables.
alg1s05	*PENDDAT*	2	Correction	The variable *alg1s05* was generated for wave 2. An analogous, harmonised variable on the basis of *PA0400* was created for wave 1. When generating the variable in wave 1, the category "no" was not coded as "0" as in wave 2, but incorrectly as "2". This mistake was corrected. The category "no" is now coded with "0" for wave 1 and 2.

Table 29: **Overview of retrospective alterations in the individual dataset (*PENDDAT*) (continued 5)**

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
arbzeit	*PENDDAT*	2	Correction	The open and categorial responses on working hours were included in the generated variable *arbzeit*. In case of irregular working hours (*ET2100* = -5) and the categorial response of working hours of 40 h and more (*ET2200* = 5), the median of open responses was imputed in *arbzeit*. However, mistakenly the median of all valid values was determined and imputed. This mistake was corrected – now, the median of valid open information of 40 h and more is imputed.
erwerb, erwerb2	*PENDDAT*	1	Correction	When generating the employment status variable for wave 1 (*erwerb*), implausible combinations are set to -8. Thus, the generated variable *erwerb* is set to -8 for persons who responded in *PB0100* that they were pupils, students or trainees while at the same time being employed in publically assisted employment (*PET0400*=1), since this is an implausibility that cannot be solved. Due to a mistake, however, not the pupils (*PB0100*=1) who are at the same time publicly employed are set to *erwerb*= -8 but these cases in which the target persons did not provide information about their status as pupil, student or trainee (*PB0100*=-2). This mistake was corrected. The pupils who are at the same time publically employed are now set to -8, those who responded with "details refused" when asked for their status (*PB0100*=-2) remain at *erwerb*=3 (publically assisted employment) since there is no implausibility. The harmonised variable *erwerb2,* based on *erwerb,* consequently changes analogously.
siops isei mps	*PENDDAT*	1	Correction	In wave 1, the MV codes from the ISCO88 coding made by GESIS were directly adopted in the variables *siops*, *isei* and *mps* based on this variable. As of wave 2, the MV codes -1, -2, -5, -6 and -8 from the ISCO coding are no longer differentiated in *siops*, *isei* and *mps* but consistently set to -5 (cannot be coded). Now, the MV codes are no longer differentiated also for wave 1 but set consistently to -5 in *siops*, *isei* and *mps*, too.

Table 29: Overview of retrospective alterations in the individual dataset (*PENDDAT*) (continued 6)

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
isei *siops* *iseilewt* *siopslewt* *iseieewt* *siopseewt*	*PENDDAT*	2	Correction	ZUMA delivers the variables *isco88, isei, siops* and *mps*. For isco88=110 (special code for soldiers) no *siops* and *isei* assignment is created. This was now added retroactively. The correction for current soldiers also affects prestige scales of the first and last employment. The prestige scales of the occupations of the fathers and mothers also based on the ISCO88 coding are not affected by this since their assignment was correct.
migration	*PENDDAT*	2	Correction	The generated variable *migration* contains information on the migration background of the respondent. In wave 1, the variable could not be generated for senior citizens' interviews since only information on the own migration background was available but information on the migration background of parents and grandparents was not recorded. As of wave 2, this information is also recorded within the framework of senior citizens' interviews. The generation of *migration* is thus possible for senior citizens' interviews as of wave 2. However, these cases were mistakenly set to -10 (item not recorded in wave) for the senior citizens newly interviewed in wave 2. Furthermore, the variable for re-interviewed senior citizens' in wave 2 could be generated but was mistakenly generated with -10, too.
vegp *megp* *egplewt* *egpeewt*	*PENDDAT*	2	Correction	A generation mistake happened when generating the EGP values for the first and last employment of the target person as well as of the mother and the father of the target person. These variables should (among others) carry the value "-5" (cannot be generated) if the respective corresponding information on occupational status was "-5" (e.g. cannot be generated). Instead of the respective corresponding occupational status, the information was mistakenly taken from the current employment in all cases. This mistake was corrected.

Table 29:   Overview of retrospective alterations in the individual dataset (*PENDDAT*) (continued 7)

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
vkldb *visco_it*	*PENDDAT*	2	Correction	As of wave 2, also the description of the job of the mother and the father is recorded at the time the respondent was 15 years old. The information recorded in wave w was initially coded according to KldB 92 (*vkldb*, *mkldb*) by TNS Infratest. A second step brought the transition from KldB 92 to ISCO 88 (*visco_it*, *misco_it*). Besides that, GESIS conducted a direct coding (*visco*, *misco*). In the information coded by Infratest there is one case with an invalid code for *vkldb* (*vkldb*=7670). From this code there also was a transition to an ISCO code (*visco_it*=3449). Essentially, the open responses do not suffice for the coding. Correspondingly, the case also was not coded by GESIS (*visco*=-5). The code 7670 is not included in the classification of the Federal Statistical Office. Furthermore, no KldB code (and as a consequence also no ISCO code) should have been assigned since the responses did not suffice for coding. The values in *vkldb* and *visco_it* were assigned "-5" (cannot be coded) in this case.

Table 30:   Overview of retrospective alterations in the spell data at the household level (*alg2_spells*)

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
AL21201a-e *AL21202a-e* *AL21850a-e* *AL21851a-e* *AL21900a-e* *AL21901a-e* *AL22150a-e* *AL22170a-e**	*alg2_spells*	1,2	Correction	Within the spells of UB II it was recorded whether there was a cut of UB II. This included among others also the reasons for the cut and whose household members' benefit was cut. Not all information was, however, recorded in all waves, which can lead to "-9" values (item not recorded in wave) in filled cut spells. "-9" was, however, also assigned if periods of cuts were not filled. This mistake was corrected. The variables of unfilled periods of cuts are now consistently set to "-3" (not applicable (filter)), instead of individual "-9" values.

Table 31: **Overview of retrospective alterations in the spell data at the individual level (*et_spells; al_spells; lu_spells; mn_spells*)**

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
arbzeit	*et_spells*	2	Correction	The open and categorial responses on working hours were included in the generated variable *arbzeit*. In case of irregular working hours (*ET2100* = -5) and the categorial response of working hours of 40 h and more (*ET2200* = 5), the median of open responses was imputed in *arbzeit*. However, mistakenly the median of all valid values was determined and imputed. This mistake was corrected – now, the median of valid open information of 40 h and more is imputed.
isei siops	*et_spells*	2	Correction	ZUMA delivers the variables *isco88, isei, siops* and *mps*. For isco88=110 (special code for soldiers) no *siops* and *isei* assignment is created. This was now added retroactively.
MN0200f MN0201f	*mn_spells*	2	Correction	The variable label of *MN200f* was incorrect ("part of the prog.: employment in transition company"). This mistake was corrected, the correct label is now "part of the prog.: employment in training company".

Table 32: **Overview of retrospective alterations in the register datasets (*hh_register; p_register*)**

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
pnettok2 pnettod2	*p_register*	2	Correction	Within the framework of the objection procedure it became apparent that one household was included twice in the sample, namely as hnr=12002409 and as hnr=21006023. The household hnr=21006023 is a split-off household of hnr=12002409 in wave 2, which is obviously incorrect. The split-off household was not surveyed in wave 2. The split-off household was removed since the split formation was incorrect. Since no retroactive alterations are made in household structures, the original household remains in the form as recorded in wave 2. The individual who is incorrectly marked as moved-out is assigned a special code in the person register.

Table 33: **Overview of retrospective alterations in the weighting datasets (hweights; pweights)**

Altered variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
prop_t0	hweights	2	Correction	In the entry weights Infratest had delivered in the variable gew_ges the product from 1/design weight and participation propensity. This product was mistakenly included in the variable *prop_t0* (participation propensity). In the correction, a division by the design is carried out before.

5.5 Anonymisation

All data gathered by the IAB as a department of the Federal Employment Agency (BA) are social data, which places high demands on data protection. It was therefore necessary to include some of the variables in the scientific use file in a simplified form. These variables are generally identified as "anonymised" in the variable label. For the same reason it was also necessary to exclude available regional information, with the exception of the German federal states and information on East/West Germany derived from this. For reasons of data protection, neither the data on family relationships in the household nor the first names of the household members are part of the scientific use file. References to the household structure are provided, however, by generated variables, for example on the household and benefit community type (*hhtyp*[43], *bgtyp*[44]), indicator variables on partners in the household (*apartner*; *epartner*[45]), pointer variables for parents and partners in the household (*zmhh*; *zvhh*; *zparthh*[46]) and various indicator variables which show whether parents (*mhh*; *vhh*[47]) or children of the target person (e. g *ekind*[48]) are living in the household. **Fehler! Ungültiger Eigenverweis auf Textmarke.** gives an overview of the variables concerned and the process of anonymisation [49] in the individual dataset. Table 35: Overview of the anonymised variables in the employment spell dataset (*et_spells*) in wave 3 shows the anonymised variables of the employment spell dataset.

---

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

Table 34:   Overview of the anonymised variables in the individual dataset (*PENDDAT*)

Varname	Question number		Variable label	Procedure
	Standard quest.	**Sen. cit's quest.**		
PD0100	P1	P1	Year of birth (date of birth, anonymised)	The precise date of birth was shortened to year of birth.
gebhalbj	generated	generated	Half-year of birth, generated	The precise date of birth was shortened to an indicator for the 1st or 2nd half of the year.
PET1210	P84	n. in Q vers.	Last occupational status, simple classification (before January 2005) (anon.)	For technical reasons, professional and regular soldiers were recorded separately in the survey. Due to the small amount of case numbers and as this group is not usually asked about occupational status anyway, this group was merged with that of civil servants and judges.
PET1250	P87, P88	n. in Q vers.	Last occup. status civil servant: detailed information, incl. soldiers (before January 2005)(anon.)	This variable contains additional cases. The professional and regular soldiers from P87 were added to the corresponding civil servant category. The variable for professional and regular soldiers (P87) is not supplied.
PET1211	generated	n. in Q vers.	Last occup. status, simple class. (incl. spell info.) (anon.), gen.	Procedure as for *PET1210*.
PET1251	generated	n. in Q vers.	Last occup. status civil servant: detailed information, incl. soldiers (incl. spell info.)(anon.), gen.	Procedure as for *PET1250*. The variable for professional and regular soldiers (P87) is not supplied.
stiblewt	generated	n. in Q vers.	Occupational status, last job, code number, generated	When generating the occupational status variable, professional and regular soldiers are assigned to the corresponding civil servant category.
PET1510	generated	P12	Current occup. status, simple classification, surv'd from W2 (anon.)	Procedure as for *PET1210*.

Table 34: Overview of the anonymised variables in the individual dataset (*PENDDAT*) in wave 3 (continued 1)

Varname	Question number		Variable label	Procedure
	Standard quest.	**Sen. cit's quest.**		
PET1900	generated	P15, P16	Current occup. status civil servant: detailed information, incl. soldiers (anon.)	Procedure as for *PET1250*. The variable for professional and regular soldiers surveyed in the senior citizens' interviews (P15) is not supplied. As regards the personal interviews, no generated variable for prof. and regular soldiers is incorporated into the individual dataset from the employment spells (P47).
stibkz	generated	generated	Current occupational status, simple classification, harmonised (anonymised)	When generating the occupational status variable, professional and regular soldiers are assigned to the corresponding civil servant category.
stib	generated	generated	Occ. status, code number, generated	Procedure as for *stiblewt*.
PET3300	P93	n. in Q vers.	First occup. status, simple classification (anon.)	Procedure as for *PET1210*.
PET3700	P96, P97	n. in Q vers.	First occup. status civil servant: detailed info., incl. soldiers	Procedure as for *PET1250*. The variable for professional and regular soldiers (P96) is not supplied.
PET3301	generated	n. in Q vers.	First occup. status, simple class. (merged, incl. spell info.) (anon.), gen.	Procedure as for *PET1210*.
PET3701	generated	n. in Q vers.	First occup. status civil servant: detailed info., incl. soldiers, (merged, incl. spell info.) (anon.), generated.	Procedure as for *PET1250*. The variable for professional and regular soldiers (P96) is not supplied.
stibeewt	generated	n. in Q vers.	Occupational status, first job, code number, generated	Procedure as for *stiblewt*.

Table 34: Overview of the anonymised variables in the individual dataset (*PENDDAT*) in wave 3 (continued 2)

Varname	Question number		Variable label	Procedure
	Standard quest.	**Sen. cit's quest.**		
PSH0320	P281	n. in Q vers.	Mother's occup. status at that time, simple classification (anon.)	Procedure as for *PET1210*.
PSH0360	P284, P285	n. in Q vers.	Mother's occup. status at time civil servant, incl. soldiers: detailed info (anon.)	Procedure as for *PET1250*. The variable for professional and regular soldiers (P284) is not supplied.
mstib	generated	n. in Q vers.	Mother's occupational status, code number, generated	Procedure as for *stiblewt*.
PSH0620	P292	n. in Q vers.	Father's occup. status at that time, simple classification (anon.)	Procedure as for *PET1210*.
PSH0660	P295, P296	n. in Q vers.	Father's occup. status at that time, incl. soldiers: detailed info (anon.)	Procedure as for *PET1250*. The variable for professional and regular soldiers (P284) is not supplied.
vstib	generated	n. in Q vers.	Father's occupational status, code number, generated	Procedure as for *stiblewt*.
PMI0200	P264	P73	Not born in Germany: country of birth	Countries with low case numbers were grouped into larger categories.
ogebland	generated	generated	Country of birth, incl. open info., categories (anonymised)	Procedure as for *PMI0200*.
PMI0500	P267	P76	No German nationality: which nationality? (anonymised)	Nationalities of countries with low case numbers were grouped into larger categories.
ostaatan	generated	generated	Nationality, incl. open info., categories (anonymised)	Procedure as for *PMI0500*.

Table 34: Overview of the anonymised variables in the individual dataset (*PENDDAT*) in wave 3 (continued 3)

Varname	Question number		Variable label	Procedure
	Standard quest.	Sen. cit's quest.		
PMI1000a	P274a	P80a	Father: country of residence before migration (anonymised)	Countries of residence before migration with low case numbers were grouped into larger categories.
PMI1000b	P274b	P80b	Mother: country of residence before migration (anonymised)	Procedure as for *PMI1000a*
PMI1000c	P274c	P80c	Father's father: country of residence before migration (anonymised)	Procedure as for *PMI1000a*
PMI1000d	P274d	P80d	Father's mother: country of residence before migration (anonymised)	Procedure as for *PMI1000a*
PMI1000e	P274e	P80e	Mother's father: country of residence before migration (anonymised)	Procedure as for *PMI1000a*
PMI1000f	P274f	P80f	Mother's mother: country of residence before migration (anonymised)	Procedure as for *PMI1000a*
ozulanda	generated	generated	Father: country of residence before migration, incl. info. from open-ended questions, categories (anonymised)	Procedure as for *PMI1000a*
ozulandb	generated	generated	Mother: country of residence before migration, incl. info. from open-ended questions, categories (anonymised)	Procedure as for *PMI1000a*
ozulandc	generated	generated	Father's father: country of residence before migration, incl. info. from open-ended questions, categories (anonymised)	Procedure as for *PMI1000a*
ozulandd	generated	generated	Father's mother: country of residence before migration, incl. info. from open-ended questions, categories (anonymised)	Procedure as for *PMI1000a*

Table 34: Overview of the anonymised variables in the individual dataset (*PENDDAT*) in wave 3 (continued 4)

Varname	Question number		Variable label	Procedure
	Standard quest.	**Sen. cit's quest.**		
ozulande	generated	generated	Mother's father: country of residence before migration, incl. info. from open-ended questions, categories (anonymised)	Procedure as for *PMI1000a*
ozulandf	generated	generated	Mother's mother: country of residence before migration, incl. info. from open-ended questions, categories (anonymised)	Procedure as for *PMI1000a*

Table 35: Overview of the anonymised variables in the employment spell dataset (*et_spells*) in wave 3

Varname	Question number		Variable label	Procedure
	Standard quest.	**Sen. cit's quest.**		
ET0601	P44		Occup. status, simple classification (anon.)	Procedure as for *PET1210*.
ET1001	P47, P48		Occ. status civil servant: detailed info. (anon.)	Procedure as for *PET1250*. The variable for professional and regular soldiers (P47) is not supplied.
stib	generated		Occ. status, code number, generated	Procedure as for *stiblewt*.

5.6 Receipt of Unemployment Benefit II

Receipt of Unemployment Benefit II at the household level was already recorded in spell form in the 1st and 2nd wave. This concept was continued in wave 3 but with a slightly revised set of questions. Besides changes in phrasing, the question for reasons for the end of the receipt of Unemployment Benefit II was newly included (Z 1 in both versions of the household questionnaire; variables *AL22200a* to *AL22200f* and with coded open responses *AL22201a* to *AL22201f*).

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

In order to update the spells of Unemployment Benefit II receipt which were still ongoing in the previous wave and were therefore right-censored in the spell dataset, dependent interviewing questions are included in both versions of the household questionnaire (HH91 in the household questionnaire for re-interviewed households; HH48 in the household questionnaire for split-off households and new sample households). In cases where the household interviewed in the previous wave had split up, the censored spells of Unemployment Benefit II receipt were updated via the part of the household in which the person with whom the household interview was conducted in the previous wave is living (termed hereafter as "HRP"[50] for short).

If the HRP is a member of the household which is first reached at the old address / under the old telephone number, the spell is updated via the responses in the household questionnaire for re-interviewed households ("HHalt" for short). The procedure is different, however, if the part of the household reached at the old address / under the old telephone number gives the information that the HRP has moved out / has not been present for a year or longer / or remained at the place of residence of the previous wave. In these cases, the part of the household that split off from the original household is regarded as a separate survey household and is interviewed using the questionnaire for new sample households ("HHneu" for short). If the HRP of the previous wave belongs to this split-off household, the spell of Unemployment Benefit II receipt of the original household that was still ongoing in the previous wave is updated using the details provided by the HRP in the split-off part of the household.

There are also differences between re-interviewed and split-off households with regard to the period for which information is collected about receipt of Unemployment Benefit II in the 3rd wave. Here, too, it is of importance whether the HRP of the previous wave is living in the household. If the HRP of the previous wave is living in the household, then spells of Unemployment Benefit II receipt since the interview date of the previous wave are recorded. If the HRP is not living in the household, then only spells of Unemployment Benefit II receipt since the date when the HRP moved out or the date when the respondent moved out of the joint household with the HRP are recorded.

The households of the refreshment sample which were interviewed for the first time in wave 3 were asked about their receipt of Unemployment Benefit II during the period since the last change in the household composition. If this was before January 2007 or if no information was provided about changes in the household, then the household's receipt of Unemployment Benefit II from January 2007 onwards was recorded.

5.6.2 Structure of the spell dataset on Unemployment Benefit II

The structure and the contents of the spell dataset on Unemployment Benefit II change due to the integration of the spells of Unemployment Benefit II receipt reported in wave 3. Here it

---

[50]   HRP stands for "household reference person".

is necessary to distinguish between (1) new variables that refer to a particular wave, (2) new variables that do not refer to a particular wave and (3) variables that are no longer surveyed in wave 3.

1. Also in wave 3 new wave-specific cross-sectional variables were included in the Unemployment Benefit II spell dataset. These were: *AL20602, AL20702a to AL20702o, AL20802 and AL20902*. These variables refer to the interview date of the 3rd wave. Cross-sectional variables also exist for the interview dates of the previous waves which contain the analogous information referring to the respective wave. Structure of the spell dataset on Unemployment Benefit II

2. gives an overview of the cross-sectional information contained in the Unemployment Benefit II spell dataset.

Table 36:   Cross-sectional variables in the UB II spell dataset (*alg2_spells*)

	Cross-sectional variable with information referring to …		
	Wave 1:	**Wave 2:**	**Wave 3:**
Does the HH receive UB II for all HH members?	*AL20600*	*AL20601*	*AL20602*
Does the HH receive UB II for the individuals 1 to 15?	*AL20700a* to *AL20700o*	*AL20700a* to *AL20701o*	*AL20700a* to *AL20702o*
Amount of monthly UB II receipt?	*AL20800*	*AL20801*	*AL20802*
Has a cut of UB II begun?	*AL20900*	*AL20901*	*AL20902*

3. Embedded in the spells of receipt of Unemployment Benefit II is information on times of benefit cuts. Up to wave 2, there were up to five cuts during a period of receipt of benefits. Within the framework of an update of the unemployment benefit II receipt that was censored in the previous wave, also information on the newly begun cuts are recorded. The new cuts are transferred to the existing spells of Unemployment Benefit II that are to be updated. Since the existing maximum number of cuts per period of receipt did no longer suffice due to the renewed update, an additional, sixth cut was introduced, which carries the abbreviation "f" [51]. Furthermore, the data structure corresponds to that from wave 2.

4. The reason for the cut, *AL21900a* to *AL21900e*, was also not recorded in wave 3. Accordingly, no responses to open-ended questions were coded to the variables *AL21901a* to *AL21901f* any longer.

---

[51]   The variables indicating a cut can be recognised from a letter at the end of the variable. Cut variables relating to the first cut end with an "a", those relating to a second cut with a "b" etc.

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

As was done in wave 1 and 2, the information on receipt of Unemployment Benefit II was also subjected to a number of plausibility checks in wave 3. Inadmissible overlaps and datings of spells of Unemployment Benefit II receipt or of benefit cuts were corrected if necessary. In principle, changes were only made to the generated date variables (*bmonat; bjahr; emonat; ejahr*) of the spell of Unemployment Benefit II receipt, the spells of benefit cuts (*alg2kbm; alg2kbj; alg2kem; alg2kej*) and in the censoring indicator of the spell of Unemployment Benefit II receipt (zensiert). If it was not possible to remove implausibilities by correcting the date variables, then in a small number of cases spells of Unemployment Benefit II receipt were merged or spells of Unemployment Benefit II receipt or benefit cuts were deleted entirely.

5.6.4 Updating the spell dataset on Unemployment Benefit II receipt

After the spells of Unemployment Benefit II receipt that were reported in wave 3 had been converted into spell format and following the plausibility checks and corrections, where inadmissible overlaps and spells with implausible dates were corrected, the spells of Unemployment Benefit II receipt which were still ongoing at the time of the interview in the previous wave were updated using the information gathered in wave 3. Three variants are to be distinguished here. In the first two, (1) and (2), only the censoring indicator zensiert is changed. The third variant (3) is an update of the spell which was censored in the previous wave using information gathered in wave 3 in the narrow sense. Here the censoring indicator is integrated into the spell of Unemployment Benefit II receipt which was still ongoing in the previous wave, as are the generated and surveyed end dates, the wave-specific cross-sectional information (see above) and information about new spells of benefit cuts. In addition to updating spells which were censored in the previous wave, new spells that were reported in wave 3 are merged with the spell dataset (4). These four variants are outlined briefly below:

1. *Cases in which the HRP of the previous wave no longer lives in the household and is also no member of a split-off household interviewed in the current wave.*

   In order to prevent the censored Unemployment Benefit II spells of the original household that were recorded in the previous wave continuing to be evaluated as current benefit receipt of this household, the censoring indicator was set to *zensiert* "-5" (HRP of the previous wave not in the household and not interviewed) in these cases. The indicator *zensiert* was also set to "-5" in cases where the HRP of the previous wave had died. The reported and generated variables for the end date of the spell (*AL20300, AL20400* and *emonat, ejahr*) as well as the question whether a spell still continues (*AL20500*) remain unchanged.[52]

---

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

2. *Cases in which the household in wave 3 contradicts an ongoing spell of Unemployment Benefit II receipt as of the interview date in the previous wave.*

If the household contradicted the information that there was an ongoing spell of Unemployment Benefit II receipt at the time of the previous wave, either explicitly or implicitly (by reporting an end date that preceded the interview date in the previous wave) in the update question (HH91 in HHalt; HH48 in HHneu), then *zensiert* was set to "2" (no). The information provided in the interview of the previous wave is presumed to have been correct. As it is not possible to make any reliable statements about the continued duration of the benefit receipt beyond the date of the interview in the previous wave, it is assumed that the benefit receipt ended in the month of the interview in the previous wave. The reported and generated variables for end date of the spell (*AL20300, AL20400* and *emonat, ejahr*) as well as the question whether a spell still continues (*AL20500*) remain unchanged.[53] The generated end date of the Unemployment Benefit II spell (*emonat; ejahr*) was already in the previous wave set to the interview date of the previous wave.

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

If information about the end date of a spell of Unemployment Benefit II receipt that was censored in the previous wave is available in wave 3, then the spell which was censored in the previous wave was updated using the current information. First, the surveyed end date (*AL20300; AL20400*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the receipt of Unemployment Benefit II is still ongoing (*AL20500*) and the censoring indicator (*zensiert*) were overwritten with the information gathered in the previous wave. Furthermore, the spells of benefit cuts reported in the 3rd wave and the cross-sectional data referring to wave 3 (*AL20602; AL20702a to AL20702o, AL20802, AL20902*) were included.

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

Spells reported for the first time in wave 3 were added to the Unemployment Benefit II spell dataset. Then the spell counter was generated anew in order to create a variable without gaps *spellnr*.

5.7 Employment biographies

Employment, unemployment and gap periods at the individual level were recorded in spell form already in the 2nd wave. This concept was continued in wave 3. In addition to the adjustments necessary for updating the employment spells (ET-Spells) and unemployment spells (AL spells) that were still ongoing at the time of the interview of the 2nd wave, the contents were expanded and minor corrections and expansions were made in response to

---

[53]   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.

the experiences made in the 2nd wave. For individuals that were asked for their employment biography for the first time in wave 3, the reference date for the start of the retrospective interval was adjusted. In wave 3, all spells of employment and unemployment since January 2006 are to be reported here (in wave 2: January 2005). Individuals who were interviewed on their employment biography already in the previous wave, however, should report all new spells since the date of the last interview.

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

ET and AL spells lasting from wave 2 were updated in the 3rd wave. Not updated were gap spells (LU spells) since they project the gaps at the respective interview date and are thus not set up cross-wave in the conception of the questionnaire. In order to update the ET and AL spells which were still ongoing in the previous wave and were therefore right-censored in the spell dataset, dependent interviewing questions are included in the personal questionnaires (E38 for ET spells and A106 for AL spells). Up to two ET spells and one AL spell from the previous wave could be updated. For respondents with more than two ongoing ET spells at the interview date, in each case the employment with the largest amount of working hours was updated.

5.7.2 Structure of the spell datasets

The LU spell dataset remains unaltered as compared to wave 2 concerning its structure and the variables contained. Due to the integration of the spells of employment and unemployment reported in wave 3 into the spell data sets of the previous wave, the ET and AL spell dataset is expanded by new variables. Here it is necessary to distinguish between (1) new variables that refer to a particular wave and (2) new variables that do not refer to a particular wave.

1. The variables *ET0600* to *ET2200* from the ET spell dataset are seen as wave-specific, cross-sectional information referring to wave 2. For the cross-sectional information gathered in wave 3, analogously the new variables *ET0601 to ET2201* were included in the ET spell dataset. Table 37 gives an overview of the cross-sectional information contained in the ET spell dataset.

Table 37: Cross-sectional variables in the ET spell dataset (*et_spells*)

	Cross-sectional variable with information referring to …		
	Wave 1:	**Wave 2:**	**Wave 3:**
Occupational status (simple and detailed classification)	*(no ET spells)*	*ET0600* *ET0700* *ET0800* *ET1000* *ET1100* *ET1200*	*ET0601* *ET0701* *ET0801* *ET1001* *ET1101* *ET1201*
Supervisory function; number of employees supervised	*(no ET spells)*	*ET1300* *ET1400*	*ET1301* *ET1401*
Cancellation of limitation of an initially limited employment	*(no ET spells)*	*ET1700*	*ET1701*
Working hours (contracted; actual; average for irregular working hours)	*(no ET spells)*	*ET2000* *ET2100* *ET2200*	*ET2001* *ET2101* *ET2201*

The variable *AL1300* from the AL spell dataset are seen as wave-specific cross-sectional information referring to wave 2. For the cross-sectional information gathered in wave 3, the new variable *AL1301* was included analogously in the AL spell dataset. **Fehler! Ungültiger Eigenverweis auf Textmarke.** gives an overview of the cross-sectional information contained in the spell dataset.

Table 38: Cross-sectional variables in the AL spell dataset (*al_spells*)

	Cross-sectional variable with information referring to …		
	Wave 1:	**Wave 2:**	**Wave 3:**
Amount of monthly UB I receipt?	*(no spells)*	*AL1300*	*AL1301*

2. The non wave-specific variable *ET2400* (How did the person first get to know about the new position?) and the appropriate variable including coding *ET2401* were first recorded in wave 3 and integrated in the ET spell dataset.

5.7.3 Plausibility checks and corrections of the spell datasets

In the gap module respondents could make different kinds of responses to close gaps of more than three months in the employment biography. The dates of already recorded spells could either be corrected or the respondent could add spells that had been forgotten before (spells of employment or unemployment) or were only recorded within the framework of the gap module (spells of economic inactivity).

The date corrections reported in the gap module were initially applied to the respective spell dates. Subsequently, ET and AL spells reported in the gap module were allocated to the ET or AL spell dataset and further processed there. Besides the ET and AL spells reported in the gap module directly as such (categorial response), further gap spells were identified as ET or AL spells in the coding of open-ended responses and assigned to the respective spell dataset.

At the individual level the spell datasets on employment and unemployment spells and the gap dataset were checked for plausibility and corrected, if necessary. Checks were only made within one type of spell. Cross-dataset checks were not carried out. As with the spell data on Unemployment Benefit II receipt, corrections and recodings were only carried out in the generated date variables. Here, too, seasons were recoded into months, "-8" values were allocated for implausible responses and date information was replaced or rendered plausible. As only the generated date variables were edited, the original information gathered in the survey is available to the user in the date variables *ET0100-ET0400, AL0100-AL0400* and *AL0800-AL1100,* and *LU0200-LU0500*, thus permitting the user to conduct his/her own checks and corrections.

In addition, it seemed necessary to delete entire spells in some cases. Most of these deletions can be attributed to faults in the gap module. For example in the gap module further spells were recorded for a person although the entire retrospective period was already covered. Or, as a result of mistakes in operating the gap module, interviewers recorded virtually identical spells more than once instead of using the available correction function. Spells that are completely outside the period surveyed but for which data were nonetheless collected were also deleted.

5.7.4 Update of ET and AL spell datasets

After the spells that were reported in wave 3 had been converted into spell format and following the plausibility checks and corrections where inadmissible overlaps and spells with implausible dates were corrected, the ET and AL spells which were still ongoing at the time of the interview in the previous wave were updated using the information recorded in wave 3.

Three variants are to be distinguished here. In the first (1), only the censoring indicator *zensiert* is changed. The second variant (2) is an update of the spell which was censored in the previous wave using information gathered in wave 3 in the narrow sense. Here, the censoring indicator is integrated into the spell which was still ongoing in the previous wave, as are the generated and recorded end dates and wave-specific cross-sectional information (see above).

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

1. *Cases in which the individual in wave 3 contradicts an ongoing spell at the interview date in the 2nd wave.*

   If the individual contradicted the information that there was an ongoing ET or AL spell at the time of the previous wave, either explicitly or implicitly (by reporting an end date that preceded the interview date in wave 2) in the update question (E38 for ET spells and A106 for AL spells), then the censoring indicator *zensiert* was set to "2" (no). The information provided in the interview of the previous wave is presumed to have been correct. As it is not possible to make any reliable statements about the continued duration of the spell beyond the date of the interview in wave 2, it is assumed that the benefit receipt ended in the month of the interview in wave 2. The reported and generated end date of the spell (*ET0300, ET0400* or *AL0300, AL0400* and *emonat, ejahr)* as well as the question whether a spell still continues (*ET0500 or AL0500*) remain unchanged [54]. The generated end date of the spell (*emonat; ejahr*) was set to the interview date of the 2nd wave already in the previous wave.

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

   If information about the end date of an ET or AL spell that had been censored in wave 2 is available in wave 3, then the spell which had been censored in wave 2 was updated using the current information. For ET spells the recorded end date (*ET0300; ET0400*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the spell was still ongoing (*ET0500*), the reason for the cancellation of a work contract (*ET2300*), the generated variables on occupational status and weekly working hours (*stib, arbzeit*) and the censoring indicator (zensiert) are overwritten with the information gathered in wave 3. Furthermore, the cross-sectional data referring to wave 3 (*ET0601 to ET2201*) were included.

---

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

For AL spells the recorded end date (*AL0300; AL0400*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the spell was still ongoing (*AL0500*), the reason for the end of unemployment (*AL0600, AL0601*) and the censoring indicator (*zensiert*) are overwritten with the information gathered in wave 3. Furthermore, the cross-sectional data referring to wave 3 (*AL1301*) was included. AL spell data, moreover, feature the exception that the spell of Unemployment Benefit I (receipt of UB I) is recorded within an AL spell. Which information is updated depends on whether there already was a receipt of UB I in this spell of unemployment and whether this receipt was ongoing in the previous wave:

- If in the previous wave there also was a continuous receipt of receipt of UB I in the AL spell to be updated, the end of the receipt was recorded[55]. In this case the recorded end date of the receipt (*AL1000, AL1100*), the indicator as to whether the spell is ongoing (*AL1200*), the generated end date of receipt (*alg1em, alg1ej*) and the censoring indicator of the receipt (*alg1akt*) were overwritten with the information collected in wave 3.

- If the respondent never received UB I in this AL spell to updated, he/she was asked whether there has been a UB I receipt since the interview date of the previous wave. If there was, all information on UB I receipt was overwritten with the information obtained in wave 3. Besides the indicator as to whether UB I was received in the AL spell (*AL0700*), the corrected beginning 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 which, however, ended already in the previous wave, or if the individual provided no information on UB I receipt in the previous wave, no question was asked on the receipt of UB I within this AL reporting spell. Consequently, no changes were made in these spells.

3. *Spells of ET or AL reported for the first time in wave 3 which do not update any spells that were censored in wave 2.*

Spells reported for the first time in wave 3 were added to the respective spell dataset. Then the spell counter was generated anew in order to create a variable without gaps *spellnr*.

Updating the spell datasets does not affect the spell number of the spells on wave 2 already contained in the SUF. These spells keep their spell number. The new spells from wave 3 are added to the respective dataset and the spell number is updated.

5.7.5 Update of the LU spell dataset

When updating the LU spells, other than for the ET and AL spells no attached spells from wave 2 were updated in wave 3. The integration of the gap spells reported in wave 3 into the

---

[55] If the respondent confirmed the receipt in the previous wave (see A 112a in the personal questionnaire wave 3).

gap spells from wave 2 was therefore modified. If there was a censored episode in the LU spell dataset of wave 2, then it was checked whether in wave 3 there was a spell of the same type in which falls the interview date of the previous wave. If this was the case, the spells were merged, i.e. the spell from wave 2 took over the generated end date (*emonat und ejahr*) and the reported end date (*LU0400 und LU0500*) and the censoring indicator *zensiert* of the spell from wave 3. For censored gap spells from wave 2 for which there was only a senior citizen's interview in wave 3, the censoring variable *zensiert* was set to the special code -5 ("censoring in data preparation cancelled"). Gap spells reported for the first time in wave 3 were added to the respective spell dataset. Then the spell counter was generated anew in order to create a variable without gaps *spellnr*.

5.8 Participation in measures

In the 2nd wave, the concept for surveying participation in employment and training measures was thoroughly reworked. This concept was continued in wave 3. The reference date to record the measure for newly participating individuals was January 2007 and for individuals that already participated the interview date of the last interview.

The measure (MN) spell dataset has several exceptions as compared to the other spell datasets of the PASS arising from the survey concept. These will be presented briefly in the following.

No update of measure spells censored in the previous wave is planned for the participation in labour market policy measures which are stored in the MN spells. The measure module of the personal questionnaire was not only supposed to record programmes of a comparatively long duration (e.g. 1-Euro-Jobs) in wave 2 and 3; the respondents were also asked to report significantly shorter measures (such as application trainings etc.). The update of censored spells in the following wave would have been difficult since it must be expected that people have a bad memory of shorter spells.

Therefore, each new MN spell recorded besides the beginning date also information on duration and/or end date. For measures that were already finished, the actual duration or the actual end date was recorded. For measures still ongoing at the interview date, respondents should report the planned duration or the planned end date.

Difficulties for the update of the MN spell dataset arise because no update is planned for censored MN spells from the previous wave and because information of different "quality" is available for the end of the measure (actual versus planned end). Instead of closing the censoring of spells from wave 2 following a rule (e.g. on the basis of the planned end of the measure) when integrating it into wave 3, a different approach was taken.

Wave 3 included a wave ID (*spwelle)* in the MN spell dataset with which the wave can be identified in which a spell was stored. The censoring indicator of the MN spells (*zensiert*) was not edited for spells from the previous waves. This was also not done if the interview date of wave 3 was after the planned end of a measure reported in the previous wave. Table 39 gives an overview, which cases can be contained in the MN spell dataset of wave 3 and which information is available in these cases on the end of the spell.

Table 39: **Overview on the information on end date in the integrated MN spell dataset of wave 2 and 3 (*mn_spells*)**

	Value of the variable in MN spells wave 3		Reported duration/ end date		Generated end date
	spwelle	zensiert	Type	Variables	
Spell reported in W2 and finished back then	*2*	*2*	*actual*	MN0600	*emonat*
				MN0700	*ejahr*
				MN0800	
Spell reported in W2 and ongoing back then	*2*	*1*	*planned*	MN1100	*emonat*
				MN1200	*ejahr*
				MN1300	
Spell reported in W3 and finished back then	*3*	*2*	*actual*	MN0600	*emonat*
				MN0700	*ejahr*
				MN0800	
Spell reported in W3 and ongoing back then	*3*	*1*	*planned*	MN1100	*emonat*
				MN1200	*ejahr*
				MN1300	

This means that the censoring indicator (*zensiert*) – in contrast to the other spell data of the PASS – does not show a censoring of the spell at the time of the last interview in the MN spell dataset but the censoring of the spell in which it was reported. The wave ID (*spwelle*) shows which wave the censoring refers to.

In wave 3 there was thus no adjustment of the information gathered in wave 2, i.e. censored spells from wave 2 remain censored even after the integration of wave 3.

5.8.1 Structure of the MN spell dataset

The structure of the MN spell dataset remains almost the same as compared to wave 2. Only the variable "name of the programme" (MN1500), which in wave 2 was filled with missing values anyway, was removed from the dataset.

5.8.2 Plausibility checks and corrections in the MN spell dataset

The MN spell dataset on the participation in measures of labour market policy was checked for plausibility and corrected. Only the generated date variables were corrected and re-coded. Seasons were recoded into months, "-8" values were allocated for implausible responses and date information was replaced or rendered plausible.

New measure spells reported in wave 3 were added to the MN spell dataset of wave 2. Then the spell counter was generated anew in order to generate a variable without gaps *spellnr*.

6 Weighting wave 3

The construction of the weights for the 3rd wave was generally made similar to the 2nd wave (see Gebhardt et al. 2009, Chapter 9; it also contains a schematic overview on the weighting concept of PASS). Differences arise from the integration of temporary non-responses that became necessary for the first time, i.e. households which participated in the wave two years ago (wave 1) but not in the previous wave (wave 2), and from a different calibration procedure. Instead of GREG, a raking procedure (IPF) was used, which however has no major importance for the weights[56].The starting point for the weighting procedure for the third wave and for the longitudinal section from wave 2 to wave 3 were the cross-sectional weights from wave 2 for households and individuals. The two weights of each household and the two weights of each individual were updated again.

6.1 Design weights for the wave 2 households in the 3rd wave

New "household design weights" were generated for the 3rd wave from the cross-sectional weights for households of the second wave, taking into account people moving into households from within Germany. This was again done by using the weight share procedure as described in wave 2. Births, deaths or moves out of households have no influence on the weight; moves into households from within Germany, on the other hand, increase the inclusion probability of a household as the individuals who have moved into the household also had the chance of being included in the sample in wave 1 or wave 2 (only refreshment sample BA). The new design weight for subsample i $dw_i hh_3$ is therefore calculated from the old cross-sectional weight $wq_i hh_2$:

---

[56]   The weights in these procedures are generally very similar. No corrected standard errors can be estimated in the raking procedure as compared to GREG. However, considering the effects of the calibration on the standard error in Stata is not possible anyway as in most statistics programmes.

$1/ dw_i hh_3 = 1 / wq_i hh_2 + (n_{sample\ i} / n_{population\ i})$

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

6.2 Design weights for the wave 3 refreshment sample

In the third wave, the panel was only refreshed by sampling new households from the new inflows to benefit recipiency. All households that were in receipt of benefit in July 2008 but had had no probability of being selected for the register data sample in the same month 2007 and the same month 2006 had a chance of being drawn. This refreshment of the sample could be done by selecting only benefit communities in which no member was receiving benefits in July of the two previous years. The refreshment sample was drawn in the 300 points of the first wave. Analogous with the special pps procedure used to draw the first register data sample, which is described in Rudolph and Trappmann (2007), the sample size was proportional to the share of new benefit recipients in the population in the sampling point (at the time when the sampling points were selected). The calculation of the design weights is also described in the same article. For cases with *sample=4* 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 2, the probability of re-participation is estimated for each household that participated in the second wave on the basis of logit models for willingness to participate in a panel, availability and participation. Households that only participated in wave 1 but not in wave 2 (temporary non-responses) were not considered for the modelling. In addition to variables from the household interview and the personal interview with the head of the household in the previous wave, also other variables are included which are associated with the fieldwork, e.g. interview mode, number of contact attempts, relocation, household size, number of individuals willing to participate. 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 a sample of the first wave for the total possible period $[t_1; t_2; t_3]$ between all three waves can be obtained as product from the cross-sectional weight to t_1, *hpbleib* (wave 1 to wave 2) and *hpbleib* (wave 2 to wave 3).

Table 40: Variable overview, codes and reference categories for the logit models of the re-participating households

Variable code and reference category	Explanation
alter29	Household reference person (HRP) younger than 30 years
alter3039	HRP 30 – 39 years old
alter4049	HRP 40 – 49 years old
alter65	HRP older than 65 years
Reference category	HRP 50 – 64 years
Mann	HRP male
Reference category	HRP female
staatandere	HRP has nationality other than German
Reference category	HRP has German nationality or missing information
arbzeit2	HRP: Weekly working hours > 0 and < 40 hours
arbzeit3	HRP: Weekly working hours >= 40 hours
Reference category	HRP: Weekly working hours = 0 hours
Dschul1	School qualification HRP: still pupil, other German school qualification, foreign qualification, missing information
Dschul2	School qualification HRP: School finished without qualification, qualification from special school
Dschul3	School qualification HRP: Lower secondary school leaving certificate, lower secondary school leaving certificate from the former GDR (POS) after completion of grade 8
Dschul5	School qualification HRP: Entrance qualification for University of Applied Sciences, general or subject-specific university entrance qualification
Reference category	School qualification HRP: Intermediate secondary school leaving certificate
DhealthKAWN	Subjective evaluation of the health state of HRP: missing information
DhealthZu	Subjective evaluation of the health state of HRP: satisfactory
DhealthSchlecht	Subjective evaluation of the health state of HRP: not so good, bad
Reference category	Subjective evaluation of the health state of HRP: very good, good
DZufrKAWN	General life satisfaction HRP: missing information
DZufr04	General life satisfaction HRP: scale value 0 – 4
DZufr910	General life satisfaction HRP: scale value 9 – 10
Reference category	General life satisfaction HRP: scale value 5 – 8
eigentum	Type of residential property: proprietor
Reference category	Type of residential property: tenant, missing information
anz_0_3	Number of persons aged 0 – 3 years
anz_4_6	Number of persons aged 4 – 6 years
anz_7_14	Number of persons aged 7 – 14 years
anz_15_64	Number of persons aged 15 – 64 years
anz_65	Number of persons aged 65 years and older

Table 40: **Variable overview, codes and reference categories for the logit models of the re-participating households (continuation 1)**

Variable code and reference category	Explanation
DinvalidAge	age responses that cannot be evaluated: Yes
Reference category	age responses that cannot be evaluated: No
in_hh2	Number of personal interviews in the HH: 2
in_hh3	Number of personal interviews in the HH: 3 and more
Reference category	Number of personal interviews in the HH: 1
Dhhincom	Household income: missing or implausible response
hhincom1	Household income: up to EUR 871
hhincom3	Household income: EUR 1401 - 2200
hhincom4	Household income: more than EUR 2200
Reference category	Household income: EUR 872 - 1400
alg2abez	UB II receipt of the household: current receipt
Reference category	UB II receipt of the household: no current receipt
halg2st3	UB II receipt of the household in the three years prior to the interview: receipt
Dhalg2st3	UB II receipt of the household in the three years prior to the interview: missing information
Reference category	UB II receipt of the household in the three years prior to the interview: no receipt
Dhpwnka0	Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: none
Dhpwnka2	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
eastwest	Old and new federal states: new federal states
Reference category	Old and new federal states: old federal states
Dbundesl1	Federal state: Schleswig-Holstein
Dbundesl2	Federal state: Hamburg
Dbundesl3	Federal state: Lower Saxony
Dbundesl4	Federal state: Bremen
Dbundesl6	Federal state: Hesse
Dbundesl7	Federal state: Rhineland-Palatinate
Dbundesl8	Federal state: Baden-Württemberg
Dbundesl9	Federal state: Bavaria
Dbundesl10	Federal state: Saarland
Dbundesl11	Federal state: Berlin
Dbundesl12	Federal state: Brandenburg
Dbundesl13	Federal state: Mecklenburg-Vorpommern

Table 40: Variable overview, codes and reference categories for the logit models of the re-participating households (continuation 2)

Variable code and reference category	Explanation
Dbundesl14	Federal state: Saxony
Dbundesl15	Federal state: Saxony-Anhalt
Dbundesl16	Federal state: Thuringia
Reference category	Federal state: North Rhine-Westphalia
Dbik_2	BIK size class of municipality: 2000 – under 5000 inhabitants
Dbik_3	BIK size class of municipality: 5000 – under 20000 inhabitants
Dbik_4	BIK size class of municipality: 20000 – under 50000 inhabitants
Dbik_5	BIK size class of municipality: 50000 – under 100000 inhabitants surroundings
Dbik_6	BIK size class of municipality: 10000 – under 100000 inhabitants centre
Dbik_7	BIK size class of municipality: 100000 – under 500000 inhabitants surroundings
Dbik_8	BIK size class of municipality: 10000 – under 500000 inhabitants centre
Dbik_9	BIK size class of municipality: 500000 and more inhabitants surroundings
Dbik_10	BIK size class of municipality: 500000 and more inhabitants centre
Reference category	BIK size class of municipality: under 2000 inhabitants
kcati0	Contact attempts CATI: 0
kcati2	Contact attempts CATI: 5 – 10
kcati3	Contact attempts CATI: 11 – 27
kcati4	Contact attempts CATI: 28 and more
Reference category	Contact attempts CATI: 1 – 4
kcapi0	Contact attempts CAPI: 0 or missing information
kcapi2	Contact attempts CAPI: 3
kcapi3	Contact attempts CAPI: 4
kcapi4	Contact attempts CAPI: 5 and more
Reference category	Contact attempts CAPI: 1 – 2

Table 41: Logit models on re-participation for willingness to participate in a panel, availability and participation

	Willingness to participate in panel		Contact		Participation	
	Coef.	p	Coef.	p	Coef.	p
alter29	-.4179966	0.191	-.8311556	0.000	-.5715283	0.000
alter3039	-.6758633	0.017	-.6455897	0.000	-.2683763	0.045
alter4049	-.2119312	0.427	-.2243778	0.046	-.1361424	0.217
alter65	.5006297	0.313	.0803519	0.764	.1780036	0.465
Mann	.0951685	0.619	-.0361669	0.656	-.0940206	0.271
staatandere	-.7545756	0.009	-.3118517	0.024	-.2040702	0.232
arbzeit2	.227652	0.448	.2733963	0.044	-.0801597	0.520

Table 41: Logit models on re-participation for willingness to participate in a panel, availability and participation (continuation 1)

	Willingness to participate in panel		Contact		Participation	
	Coef.	p	Coef.	p	Coef.	p
arbzeit3	.3782709	0.211	.0284119	0.824	.0635809	0.624
Dschul1	-.0029973	0.995	-.0959908	0.644	.0328029	0.894
Dschul2	-.1588747	0.687	-.3502206	0.024	-.1364375	0.492
Dschul3	-.23506	0.291	-.1238202	0.199	-.1683959	0.092
Dschul5	-.0376717	0.885	.1295561	0.248	.0028406	0.980
DhealthKAWN	2.937874	0.024	-.8767879	0.196	-.942785	0.159
DhealthZu	.0708809	0.732	.1378718	0.134	.1871987	0.050
DhealthSchlecht	.260788	0.301	-.1212947	0.237	-.1009442	0.353
DZufrKAWN	-1.844352	0.006	.6504756	0.348	.1695471	0.795
DZufr04	-.4867122	0.032	-.0084285	0.934	-.072271	0.526
DZufr910	.0658411	0.834	.2003547	0.152	.0495304	0.702
eigentum	-.2927264	0.227	.2691123	0.028	-.0191214	0.859
anz_0_3	.0684786	0.837	-.0999625	0.423	.1360783	0.388
anz_4_6	.3885585	0.266	-.1451466	0.192	-.1446253	0.252
anz_7_14	-.0008379	0.996	.0708032	0.317	.0613971	0.407
anz_15_64	-.0186249	0.884	-.2013297	0.000	-.477008	0.000
anz_65	-.5701717	0.040	-.0753248	0.645	-.3791891	0.007
DinvalidAge	-1.200804	0.160	-.7285887	0.266	-1.648609	0.006
in_hh2	1.369934	0.000	.4284563	0.000	.5352592	0.000
in_hh3	2.13238	0.004	.7714587	0.000	1.161619	0.000
Dhhincom	-1.175575	0.004	.4580934	0.234	-.4479153	0.101
hhincom1	.021228	0.936	-.193863	0.073	.2231407	0.072
hhincom3	-.4689878	0.070	.0460383	0.687	.163656	0.167
hhincom4	.3694011	0.339	.3583555	0.020	.3563863	0.016
alg2abez	.3348003	0.226	.1252726	0.276	-.0393806	0.763
halg2st3	.1390143	0.624	-.0921112	0.460	.2290695	0.084
Dhalg2st3			-1.477168	0.003	.466441	0.581
Dhpwnka0	.5728716	0.010	.2676548	0.001	.1530721	0.071
Dhpwnka2	-1.084382	0.000	.0414075	0.785	.001363	0.994
eastwest	-.0897827	0.664				
Dbundesl1			-.2056056	0.272	.1275223	0.556
Dbundesl2			-.7086271	0.009	-.2949669	0.402
Dbundesl3			-.0139394	0.922	-.01159	**0.938**
Dbundesl4			.7596985	0.156	.0246193	**0.958**
Dbundesl6			.5715236	0.005	.1964544	**0.279**
Dbundesl7			-.2165509	0.287	.011608	**0.958**

Table 41: Logit models on re-participation for willingness to participate in a panel, availability and participation (continuation 2)

	Willingness to participate in panel		Contact		Participation	
	Coef.	p	Coef.	p	Coef.	p
Dbundesl8			.173082	0.290	-.0377617	0.806
Dbundesl9			.1098784	0.437	.0428446	0.766
Dbundesl10			-.3480398	0.276	.5340396	0.278
Dbundesl11			-.2323356	0.147	-.1222737	0.532
Dbundesl12			.0672585	0.747	.1917281	0.371
Dbundesl13			.1826805	0.499	.3863675	0.183
Dbundesl14			.173065	0.345	.4142374	0.031
Dbundesl15			.2912595	0.159	.21407	0.302
Dbundesl16			.3309171	0.190	.6795506	0.015
Dbik_2			.4852294	0.014	-.0985176	0.587
Dbik_3			.1629304	0.185	-.0916656	0.513
Dbik_4			.2484895	0.088	-.0364019	0.812
Dbik_5			.5679636	0.032	-.3075491	0.225
Dbik_6			.0964047	0.564	-.0833082	0.651
Dbik_7			.2507438	0.127	-.1531106	0.350
Dbik_8			.3224649	0.070	-.2950502	0.090
Dbik_9			.2409303	0.326	-.3722877	0.113
Dbik_10			.8106134	0.067	-.8256749	0.004
kcati0					2.595051	0.000
kcati2					.4144359	0.001
kcati3					.4314559	0.000
kcati4					.4958591	0.000
kcapi0					.9932981	0.000
kcapi2					.1224835	0.552
kcapi3					-.1604692	0.473
kcapi4					-.4160623	0.017
cons	4.099317	0.000	2.325814	0.000	1.277427	0.000
n	**8440**		**8304**		**7487**	
Log likelihood	**-624.93564**		**-2489.6614**		**-2263.2766**	
Pseudo R^2	**0.1025**		**0.0675**		**0.1000**	

6.4 Propensity to participate – first interviewed split-off households

This step calculated the propensities to participate for new split-off households, i.e. households that are included in the panel due to the relocation of one individual of the panel sample in a new household. Here, only split-off households were considered that had not

been interviewed in the first two waves. Thus, the propensities to participate of first interviewed split-off households were modelled. The probability of re-participation was estimated via logit models for availability and participation. Missing time stable information on the household reference person (HRP) was added from the previous wave, if necessary. The estimated propensities of the two models were multiplied. The reciprocal value of the product for the split-off households can also be found in the variable *hpbleib*.

Table 42: **Variable overview, codes and reference categories for the logit models of the split-off households participating for the first time**

Variable code and reference category	Explanation
alter3039	HRP 30 – 39 years old
alter4049	HRP 40 – 49 years old
alter5064	HRP 50 – 64 years
alter65	HRP older than 65 years
Reference category	HRP younger than 30 years
Mann	HRP male
Reference category	HRP female
staatandere	HRP has nationality other than German
staatsysmis	HRP has missing response for nationality
Reference category	HRP has German nationality
Dschul2	School qualification HRP: school finished without qual., qual. from special school
Dschul3	School qualification HRP: lower secondary school leaving certificate, lower sec. school leaving certificate from the former GDR (POS) after completion of grade 8
Dschul4	School qualification HRP: Intermediate secondary school leaving certificate
Dschul5	School qualification HRP: Entrance qualification for University of Applied Sciences, general or subject-specific university entrance qualification
Reference category	School qualification HRP: still pupil, other German school qualification, foreign qualification, missing information
sample_BA	From BA sample wave 1
sample_Auffrischer	From BA refreshment sample wave 2
Reference category	From Microm sample
kcati0	Contact attempts CATI: 0
kcati1	Contact attempts CATI: 1 – 7
kcati2	Contact attempts CATI: 8 – 19
kcati4	Contact attempts CATI: 49 and more
Reference category	Contact attempts CATI: 20 – 48
kcapi0	Contact attempts CAPI: 0
kcapi2	Contact attempts CAPI: 3
kcapi3	Contact attempts CAPI: 4
kcapi4	Contact attempts CAPI: 5 and more
kon_capi_sysmis	Contact attempts CAPI: missing information
Reference category	Contact attempts CAPI: 1 – 2

Table 43: Logit models on the first participation of split-off households for availability and participation

	Contact		Participation	
	Coef.	p	Coef.	p
alter3039	-.0457938	0.882	-.4974757	0.232
alter4049	.1541013	0.685	.435613	0.414
alter5064	-.2011846	0.584	-.0742023	0.890
alter65	1.03412	0.099	-.6649824	0.332
Mann	.0829381	0.680	.2489844	0.360
staatandere	-.6301316	0.258	-1.874751	0.030
staatsysmis	-1.014355	0.010	-.7183244	0.128
Dschul2	-.5599909	0.348	.1770058	0.830
Dschul3	-.8941441	0.048	-.3839034	0.496
Dschul4	-.5311538	0.205	-.2538346	0.608
Dschul5	-.750205	0.089	-.2666337	0.624
sample_BA	.0285283	0.889	-.1333173	0.625
sample_Auffrischer	-1.319607	0.017	-.8828261	0.365
kcati0	.678541	0.101		
kcati1	1.313406	0.002		
kcati2	2.137653	0.000		
kcati4	-.0279659	0.955		
kon_capi_sysmis	-.4791278	0.134		
kcapi0	.0868798	0.827		
kcapi2	-.0524601	0.882		
kcapi3	.1001297	0.821		
kcapi4	-.233398	0.487		
cons	.2971993	0.591	.897202	0.045
n	**483**		**262**	
Log likelihood	**-305.3388**		**-168.78392**	
Pseudo R^2	**0.0832**		**0.0410**	

6.5 Non-response weighting for households from the wave 3 refreshment sample

For the households in the refreshment sample, non-response was again modelled in a two-step procedure (availability and participation) as was done for the second wave. The participation probability derived from this can be found in variable *prop_t0*.

Table 44:   **Variable overview, codes and reference categories for the logit models of the refreshment sample wave 3**

Variable code and reference category	Explanation
alter3039	HRP 30 – 39 years old
alter4049	HRP 40 – 49 years old
alter5059	HRP 50 – 59 years old
alter60	HRP 60 years and older
Reference category	HRP younger than 30 years
sex_w	HRP female
Reference category	HRP male
staat_rge	HRP nationality: European (not German), Russian, former CIS countries
staat_son	HRP nationality: Turkish, Asian, Australian, African, American
Reference category	HRP has German nationality or missing information
Schul_Kein	School qualification HRP: no qualification
Schul_HS	School qualification HRP: Lower secondary school leaving certificate
Schul_MR	School qualification HRP: Intermediate secondary school leaving certificate
Schul_FA	School qualification HRP: Entrance qualification for University of Applied Sciences
Reference category	School qualification HRP: University entrance qualification
typ_alleinerz	BC type: single parent
typ_paarokind	BC type: couple without children
typ_paarmkind	BC type: couple with children
typ_sonst	BC type: other
Reference category	BC type: single
erw_NEF	HRP not capable of work
erw_KA	HRP without determination of capability of work
Reference category	HRP capable of work
catifeld	Household was (originally) in the CATI field
Reference category	Household was (originally) in the CAPI field
tranche2	Tranche: 2
tranche3	Tranche: 3
tranche4	Tranche: 4

Table 44: Variable overview, codes and reference categories for the logit models of the refreshment sample wave 3 (continuation 1)

Variable code and reference category	Explanation
tranche5	Tranche: 5
tranche6	Tranche: 6
Reference category	Tranche: 1
anz_verwf	Number of persons capable of work in BC
kontcati	Number of CATI contacts
kontcati_ka	Missing information on the number of CATI contacts: yes
Reference category	Missing information on the number of CATI contacts: no
kontcapi	Number of CAPI contacts
kontcapi_ka	Missing information on the number of CAPI contacts: yes
Reference category	Missing information on the number of CAPI contacts: no
bula1	Federal state: Schleswig-Holstein
bula2	Federal state: Hamburg
bula3	Federal state: Lower Saxony
bula4	Federal state: Bremen
bula6	Federal state: Hesse
bula7	Federal state: Rhineland-Palatinate
bula8	Federal state: Baden-Württemberg
bula9	Federal state: Bavaria
bula10	Federal state: Saarland
bula11	Federal state: Berlin
bula12	Federal state: Brandenburg
bula13	Federal state: Mecklenburg-Vorpommern
bula14	Federal state: Saxony
bula15	Federal state: Saxony-Anhalt
bula16	Federal state: Thuringia
Reference category	Federal state: North Rhine-Westphalia
bik_1	BIK size class of municipality: under 2000 inhabitants
bik_2	BIK size class of municipality: 2000 – under 5000 inhabitants
bik_3	BIK size class of municipality: 5000 – under 20000 inhabitants
bik_4	BIK size class of municipality: 20000 – under 50000 inhabitants
bik_5	BIK size class of municipality: 50000 – under 100000 inhabitants surroundings
bik_6	BIK size class of municipality: 10000 – under 100000 inhabitants centre
bik_7	BIK size class of municipality: 100000 – under 500000 inhabitants surroundings
bik_8	BIK size class of municipality: 10000 – under 500000 inhabitants centre
bik_9	BIK size class of municipality: 500000 and more inhabitants surroundings
Reference category	BIK size class of municipality: 500000 and more inhabitants centre

Table 45:   Logit models on first participation for availability and participation

	Contact		Participation	
	Coef.	**p**	**Coef.**	**p**
alter3039	.3090823	0.003	-.1314074	0.249
alter4049	.3885464	0.000	-.1144351	0.314
alter5059	.3887638	0.001	-.2671515	0.035
alter60	.4672857	0.035	-.3456237	0.138
sex_w	.023271	0.777	-.012965	0.884
Schul_Kein	-.5021477	0.000	-.2747929	0.114
Schul_HS	-.0455601	0.651	-.1582463	0.144
Schul_MR	.0568066	0.635	-.0463256	0.703
Schul_FA	.2362025	0.094	.1239275	0.389
staat_rge	-.5830485	0.000	.13027	0.445
staat_son	-.8213009	0.000	-.4579772	0.013
typ_alleinerz	.473528	0.005	.3056752	0.094
typ_paarokind	.4321966	0.002	-.0737591	0.658
typ_paarmkind	.6248467	0.001	.1028116	0.665
typ_sonst	.3926251	0.022	.179789	0.488
erw_NEF	-.0766154	0.614	-.0265557	0.867
erw_KA	-.7834184	0.005	.2289636	0.527
catifeld	.8830739	0.000		
anz_verwf			-.1363106	0.183
kontcati			.0017048	0.326
kontcati_ka			.2885493	0.213
kontcapi			.0815515	0.003
kontcapi_ka			-3.742772	0.000
tranche2	.1046926	0.428	.0937907	0.498
tranche3	.17847	0.179	.0056628	0.967
tranche4	-.0669418	0.599	-.1135456	0.421
tranche5	-.0269351	0.833	-.3703891	0.010
tranche6	.1719382	0.196	-.2572793	0.066

Table 45:   Logit models on first participation for availability and participation (continued)

	Contact		Participation	
	Coef.	p	Coef.	p
bik_1	.2860343	0.114	-.0870541	0.653
bik_2	.1278021	0.262	.0802627	0.533
bik_3	.4756256	0.002	-.1189641	0.446
bik_4	.4653795	0.065	.1358901	0.580
bik_5	.2334757	0.180	.1972223	0.290
bik_6	.3774596	0.016	-.0071909	0.964
bik_7	.3402342	0.068	-.0200336	0.918
bik_8	.2427334	0.398	-.5795322	0.052
bik_9	-.2359266	0.602	.2676086	0.601
bula1	.0358522	0.864	-.0866758	0.686
bula2	-.0826558	0.751	-.0283367	0.925
bula3	-.0031247	0.983	-.0724963	0.640
bula4	-.4717033	0.164	-.1457921	0.763
bula6	-.1771432	0.280	.373867	0.043
bula7	-.2555386	0.205	.0989945	0.658
bula8	-.2784098	0.047	.0020443	0.990
bula9	.1582747	0.288	.1193608	0.439
bula10	-.1490653	0.641	.2069501	0.540
bula11	.0950723	0.540	-.2387027	0.177
bula12	.0297586	0.895	-.5428247	0.023
bula13	-.157555	0.559	.409411	0.139
bula14	-.0000678	1.000	.000055	1.000
bula15	.0489525	0.823	.0454987	0.831
bula16	.3197248	0.243	.0673526	0.773
cons	-.3548499	0.054	.0960608	0.605
n	**3785**		**2704**	
Log likelihood	**-2138.9356**		**-1754.3689**	
Pseudo R^2	**0.0553**		**0.0543**	

6.6   Propensity to participate again – individuals

The decisive longitudinal weight is not the one at the household level but the one at the individual level, as the units here are stable over time. As in wave 2, propensities to participate again for individuals were estimated including additional personal characteristics via logit models for availability and participation. The dependence of the personal sample conveyed via the household context and correction of the estimation of standard errors made necessary by it was considered in these models by clustering the missing terms at the household level. The predicted propensities of the models were again multiplied. The reciprocal value of this product can be found in variable *ppbleib*. The longitudinal weight for

an individual for the period $[t_1; t_2; t_3]$ across all three waves can be obtained as product of the cross-sectional weight to t_1, *ppbleib* (wave 1 to wave 2) and *ppbleib* (wave 2 to wave 3).

Table 46:   **Variable overview, codes and reference categories for the logit models of re-participating individuals**

Variable code and reference category	Explanation
alter29	Individual younger than 30 years
alter3039	Individual 30 – 39 years old
alter4049	Individual 40 – 49 years old
alter65	Individual older than 65 years
Reference category	Individual 50 – 64 years
Mann	Individual male
Reference category	Individual female
staatandere	Individual has nationality other than German (or missing information)
staatsysmis	Individual has missing response for nationality
Reference category	HRP has German nationality (or missing information, if staatsymis not in model)
arbzeit1	Weekly working hours <= 32 hours
arbzeit2	Weekly working hours > 32 and <= 40 hours
arbzeit3	Weekly working hours > 40 hours
Reference category	HRP: Weekly working hours = 0 hours
Dschul1	School qualification: still pupil, other German school qualification, foreign qualification, missing information
Dschul2	School qualification: School finished without qualification, qualification from special school
Dschul4	School qualification: Intermediate secondary school leaving certificate
Dschul5	School qualification: Entrance qualification for University of Applied Sciences, general or subject-specific university entrance qualification
Reference category	School qualification: Lower secondary school leaving certificate, lower secondary school leaving certificate from the former GDR (POS) after completion of grade 8
DhealthKAWN	Subjective evaluation of the health state of HRP: missing information

Table 46:   **Variable overview, codes and reference categories for the logit models of re-participating individuals (continuation 1)**

Variable code and reference category	Explanation
DhealthZu	Subjective evaluation of the health state of HRP: satisfactory
DhealthSchlecht	Subjective evaluation of the health state of HRP: not so good, bad
Reference category	Subjective evaluation of the health state of HRP: very good, good
DZufrKAWN	General life satisfaction HRP: missing information
DZufr04	General life satisfaction HRP: scale value 0 – 4
DZufr910	General life satisfaction HRP: scale value 9 – 10
Reference category	General life satisfaction HRP: scale value 5 – 8
Dpwnka0	Number of "don't know" and "details refused" responses in household and personal interview: 0
Dpwnka2	Number of "don't know" and "details refused" responses in household and personal interview: more than 10
Reference category	Number of "don't know" and "details refused" responses in household and personal interview: 1 - 10
Dhsprache	Main language: German
Reference category	Main language: not German
eigentum	Type of residential property: proprietor
Reference category	Type of residential property: tenant, missing information
anz_0_3	Number of persons aged 0 – 3 years
anz_4_6	Number of persons aged 4 – 6 years
anz_7_14	Number of persons aged 7 – 14 years
anz_65	Number of persons aged 65 years and older
Danz_15_64_1	Number of persons aged 15 – 64 years: 1
Danz_15_64_2	Number of persons aged 15 – 64 years: 2
Danz_15_64_3	Number of persons aged 15 – 64 years: 3
Danz_15_64_4	Number of persons aged 15 – 64 years: 4 and more
Reference category	Number of persons aged 15 – 64 years: none
DinvalidAge	Age responses that cannot be evaluated: yes
Reference category	Age responses that cannot be evaluated: no
in_hh2	Number of personal interviews in the HH: 2
in_hh3	Number of personal interviews in the HH: 3 and more
Reference category	Number of personal interviews in the HH: 1
Dhhincom	Household income: missing or implausible response
hhincom1	Household income: up to EUR 986
hhincom3	Household income: EUR 1501 - 2500
hhincom4	Household income: more than EUR 2500
Reference category	Household income: EUR 987 - 1500
alg2abez	UB II receipt of the household: current receipt

Table 46:   **Variable overview, codes and reference categories for the logit models of re-participating individuals (continuation 2)**

Variable code and reference category	Explanation
Reference category	UB II receipt of the household: no current receipt
halg2st3	UB II receipt of the household in the three years prior to the interview: receipt
Reference category	UB II receipt of the household in the three years prior to the interview: no receipt
sample_BA	BA sample
Reference category	Microm sample
Dbundesl1	Federal state: Schleswig-Holstein
Dbundesl2	Federal state: Hamburg
Dbundesl3	Federal state: Lower Saxony
Dbundesl4	Federal state: Bremen
Dbundesl6	Federal state: Hesse
Dbundesl7	Federal state: Rhineland-Palatinate
Dbundesl8	Federal state: Baden-Württemberg
Dbundesl9	Federal state: Bavaria
Dbundesl10	Federal state: Saarland
Dbundesl11	Federal state: Berlin
Dbundesl12	Federal state: Brandenburg
Dbundesl13	Federal state: Mecklenburg-Vorpommern
Dbundesl14	Federal state: Saxony
Dbundesl15	Federal state: Saxony-Anhalt
Dbundesl16	Federal state: Thuringia
Reference category	Federal state: North Rhine-Westphalia
Dbik_2	BIK size class of municipality: 2000 – under 5000 inhabitants
Dbik_3	BIK size class of municipality: 5000 – under 20000 inhabitants
Dbik_4	BIK size class of municipality: 20000 – under 50000 inhabitants
Dbik_5	BIK size class of municipality: 50000 – under 100000 inhabitants surroundings
Dbik_6	BIK size class of municipality: 10000 – under 100000 inhabitants centre
Dbik_7	BIK size class of municipality: 100000 – under 500000 inhabitants surroundings
Dbik_8	BIK size class of municipality: 10000 – under 500000 inhabitants centre
Dbik_9	BIK size class of municipality: 500000 and more inhabitants surroundings
Dbik_10	BIK size class of municipality: 500000 and more inhabitants centre
Reference category	BIK size class of municipality: under 2000 inhabitants
kcati0	Contact attempts CATI: 0
kcati2	Contact attempts CATI: 6 – 14
kcati3	Contact attempts CATI: 14 – 42 or 14 and more
kcati4	Contact attempts CATI: 43 and more
Reference category	Contact attempts CATI: 1 – 5

Table 46: Variable overview, codes and reference categories for the logit models of re-participating individuals (continuation 3)

Variable code and reference category	Explanation
kcapi0	Contact attempts CAPI: 0 or missing information
kcapi2	Contact attempts CAPI: 3
kcapi3	Contact attempts CAPI: 4
kcapi4	Contact attempts CAPI: 5 and more
Reference category	Contact attempts CAPI: 1 – 2

Table 47: Logit models on re-participation for availability and participation

	Contact		Participation	
	Coef.	p	Coef.	p
alter29	-.3194031	0.007	-.9574916	0.000
alter3039	-.1641501	0.208	-.4442976	0.000
alter4049	.0162689	0.896	-.0822184	0.329
alter65	.2934335	0.379	-.3358986	0.047
Mann	-.04211	0.533	-.2058298	0.000
staatandere	-.3855001	0.022	-.3378695	0.005
staatsysmis	-.0434743	0.972		
arbzeit1	.213965	0.169	.0032377	0.975
arbzeit2	.1397969	0.272	-.0777185	0.369
arbzeit3	.0705923	0.606	.040118	0.651
Dschul1	.0264277	0.870	.3279346	0.003
Dschul2	-.5820174	0.000	-.0636041	0.633
Dschul4	.0247343	0.807	.1842116	0.008
Dschul5	.0261321	0.828	.2079732	0.009
DhealthKAWN	-.2165862	0.772	-1.963675	0.001
DhealthZu	.1195256	0.179	.1115784	0.075
DhealthSchlecht	-.0363983	0.717	-.0250116	0.746
DZufrKAWN	.9939802	0.146	-.5768915	0.352
DZufr04	.041639	0.698	.0073365	0.932
DZufr910	.176946	0.173	.0829721	0.279

Table 47: **Logit models on re-participation for availability and participation (continuation 1)**

	Contact		Participation	
	Coef.	**p**	**Coef.**	**p**
Dpwnka0	.2501217	0.002	.1773289	0.002
Dpwnka2	.0653905	0.696	-.1595383	0.185
Dhsprache	.0024693	0.993	-.1833932	0.289
eigentum	.3327693	0.026	.0403194	0.616
anz_0_3	-.2668426	0.080	.303347	0.014
anz_4_6	-.1438577	0.299	-.0555842	0.539
anz_7_14	.0557589	0.546	.0848012	0.117
Danz_15_64_1	.1226313	0.782	-.5429105	0.000
Danz_15_64_2	-.2948062	0.527	-2.556655	0.000
Danz_15_64_3	-.337792	0.502	-.1068527	0.000
Danz_15_64_4	-.3430471	0.516	.2323787	0.000
anz_65	-.3113932	0.089	-.383435	0.000
DinvalidAge	-.2810615	0.793	-.2419069	0.002
in_hh2	.3757126	0.002	-.0890333	0.192
in_hh3	.6023466	0.011	.1173713	0.062
Dhhincom	.2755263	0.493	.0533518	0.115
hhincom1	-.4698099	0.007	.0801471	0.040
hhincom2	-.2008494	0.168	.05335	0.331
hhincom4	.2859489	0.151	2.084369	0.213
alg2abez	.0990292	0.476	.5554232	0.583
halg2st3	-.2049675	0.237	.271659	0.461
sample_BA	.1647116	0.180	-.4007131	0.538
kcati0	2.156775	0.000	.6827928	0.000
kcati2	.3595989	0.075	-.1429823	0.000
kcati3	-1.001132	0.000	-.3206767	0.007
kcati4			-.5170584	0.000
kcapi0	1.811629	0.000	-.1676071	0.000
kcapi2	.5007148	0.013	-.6044384	0.297

Table 47:   Logit models on re-participation for availability and participation (continuation 2)

	Contact		Participation	
	Coef.	p	Coef.	p
kcapi3	.0638927	0.764	.0435586	0.042
kcapi4	-.3761208	0.015	.4082219	0.000
Dbundesl1	-.2458053	0.255	.0656108	0.283
Dbundesl2	-.7702539	0.025	-.0115265	0.045
Dbundesl3	-.0309503	0.866	-.1124558	0.710
Dbundesl4	.3374801	0.590	.1536802	0.365
Dbundesl6	.3380317	0.147	.2855574	0.645
Dbundesl7	-.5305397	0.037	-.3208774	0.949
Dbundesl8	.1868234	0.343	-.0324926	0.343
Dbundesl9	.0400736	0.818	-.1450802	0.178
Dbundesl10	-.4245764	0.308	.3532369	0.358
Dbundesl11	.0253143	0.903	-.0259127	0.046
Dbundesl12	-.0930031	0.740	.5016664	0.826
Dbundesl13	.2009278	0.491	-.0585923	0.448
Dbundesl14	.1927373	0.388	-.2191701	0.021
Dbundesl15	.6268441	0.008	.1047016	0.849
Dbundesl16	.286547	0.311	-.0202333	0.004
Dbik_2	.7123127	0.002	-.2138616	0.683
Dbik_3	.0246498	0.872	-.1596903	0.040
Dbik_4	.2873936	0.118	-.3368483	0.369
Dbik_5	.759438	0.019	-.2437077	0.921
Dbik_6	.1190219	0.559	-.375665	0.106
Dbik_7	.406349	0.044	2.400581	0.195
Dbik_8	.4490884	0.058	-.5429105	0.014
Dbik_9	.3872928	0.197	-2.556655	0.166
Dbik_10	.7727135	0.171	-.1068527	0.130
cons	.9859379	0.096	.2323787	0.000
n	**12259**		**11203**	
Log pseudolikelihood	**-2754.1834**		**-4782.647**	
Pseudo R^2	**0.2346**		**0.1114**	

Note: The correction of standard errors was made by means of an estimation clustered by 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 refreshment and panel household samples (including the refreshments from wave 2), which have been modified by the non-response modelling. The double selection probability of a newly sampled benefit recipient who was living in the same household as benefit recipients in one of the two previous years but without being a member of the benefit unit him/herself was ignored again. The new design weights of the benefit recipient sample project in the cross-section to all individuals who were living in a household containing at least one benefit community in either 7/2006, 7/2007 or 7/2008. It is only when calculating new weights for the total sample that it becomes necessary to adjust the weights for all households in receipt of benefits in 7/2008. For this adjustment the inclusion probability in the respective other sample was estimated for cases from the Microm sample (wave 1) and the refreshment sample (wave 3). For cases from the refreshment sample, the mean wave 1 selection probability in the Microm sample in the respective postcode sector and the average participation probability (for W1, W2 and W3) in that sample were assumed. For cases from the Microm sample, if they are (according to survey data) new recipients of Unemployment Benefit II who first received the benefit between the last two sampling dates (W2; W3), the mean selection probability of a household in the refreshment sample in the respective postcode sector 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 non-responses), could participate again in wave 3 for the first time. 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 is no longer contacted. In order to calculate mutual cross-sectional weights including the temporary non-responses, there was a convex combination of the modified household weights of the temporary non-responses and the modified household weights of the panel household sample (not of the refreshment sample) before calibration. The convex combination of the household weights was hence made before calibration; the calibration was then made with the new combined household weights.

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

For temporary non-responses the probability of non-participation in wave 2 in case of participation in wave 1 (non-participation propensities W2) and the probability of participation in wave 3 in case of a non-participation in wave 2 (participation propensities W3) was determined. The probability of non-participation in wave 2 is calculated from 1– participation probability in wave 2. Since the probability of re-participation in wave 2 was already

estimated via logit models for willingness to participate in a panel, availability and participation, the results of these models could be used. The participation propensities for the participation in wave 3 in case of non-participation in wave 2, however, had to be estimated again. Logit models for availability and participation with only a few variables (reason for non-response, method of contact attempt, sample and label whether it is a split-off household from the first wave) were used here. The models reproduce in the result the average probability of participation of the temporary non-responses with a comparatively low variance of the propensities.

The product of the projected probabilities of both models was multiplied by the probability of non-participation in wave 2. The modified household weight of the temporary non-responses is then calculated by multiplying the calibrated household weights of the first wave by the reciprocal value of this product.

Table 48:   **Variable overview, codes and reference categories for the logit models of the temporary non-responses**

Variable code and reference category	Explanation
end_ne	Reason for non-response in wave 2: HH could not be contacted
end_nidl	Reason for non-response in wave 2: HH was not fit to be interviewed
end_aw	Reason for non-response in wave 2: HH moved, invalid telephone number
Reference category	Reason for non-response in wave 2: refusal
cati1	Contact attempt via CATI or CAPI: first via CATI
cati2	Contact attempt via CATI or CAPI: again via CATI
Reference category	Contact attempt via CATI or CAPI: first via CAPI or research
sample_microm	Subsample Microm
Reference category	Subsample BA
split	Split-off household
Reference category	No split-off household
uv_imp	Values of independent variables imputed
Reference category	Values of independent variables not imputed

Table 49: **Logit models on re-participation in wave 3 in case of non-participation in wave 2 for availability and participation**

	Contact		Participation	
	Coef.	p	Coef.	p
end_ne	-.2280403	0.034	.2571192	0.056
end_nidl	-.4508862	0.024	-.0650028	0.805
end_aw	-.0852084	0.370	.3998799	0.001
cati1	.564227	0.000	-.2624226	0.015
cati2	-.3230414	0.128	.3277073	0.271
sample_microm	.4433349	0.000	-.1261313	0.178
split	-.3645094	0.009	.1304105	0.514
uv_imp	-.6368384	0.309	-.535854	0.567
cons	.2587525	0.008	.6734514	0.000
n	**3773**		**2453**	
Log likelihood	**-2394.156**		**-1541.6122**	
Pseudo R^2	**0.0198**		**0.0056**	

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

$dw_i hh_{temp.non-response}$ * ($n_{temp.non-response\ i}$ / ($n_{temp.non-response\ i}$ + $n_{panel\ household\ sample\ i}$)) for temporary non-responses and

$dw_i hh_{panel\ household\ sample}$ * ($n_{panel\ household\ sample\ i}$ / ($n_{temp.non-response\ i}$ + $n_{panel\ household\ sample\ i}$)) for the panel household sample.

6.9 Calibration to the household weight, 3rd wave, cross-section

Following that came another calibration of the modified design weights including the non-response weighting at the household level by raking to the benchmark values of the Federal Statistical Office for 2008. For households in receipt of benefits the weights are adjusted to the statistics of the Federal Employment Agency for July 2008. As in the previous year, also the increase in Unemployment Benefit II receipt since the previous year at the level of benefit communities (477,034) was also included as an additional benchmark value in the total sample. Those cases in the previous samples from wave 1 and wave 2 which, according to wave 3 of the survey, were receiving Unemployment Benefit II in July 2008 will be calibrated

to the benchmark statistics of the Federal Employment Agency on receipt of Unemployment Benefit II.

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

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

- Also the interval of weights was limited during calibration, i.e. a maximum and a minimum limit for weights was determined.

- Here also the total width of the weights was determined; the range of the pure calibration weights can be calculated from the relation of original weights to the trimmed input weight.

- It had to be observed here that narrower limits for the weights result in less distribution and thus less variance of the estimations; too narrow limits can, however, make the calibration of all benchmark values impossible.

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

6.9.1   Calibration of the BA sample

The population of the BA sample of the first three waves consists of all households in Germany with at least one benefit community receiving benefits in accordance with SGB II at one of the, up to now, three drawing dates (in July 2006, July 2007 or July 2008). In wave 3 only the benchmark values of BA statistics from July 2008 are calibrated. The calibration thus only influences the weights of the households from the BA sample in which at least one benefit community receiving benefits in accordance with SGB II was living in July 2008.

Starting point for the calibration were modified design weights including the non-response weighting. The modified design weights were trimmed at the 5% percentile and the 95% percentile of their distribution and after that rescaled in such a way that their total resulted in the total of the untrimmed calibrated design weights. The projection factors reach from 260.62 to 1887.11 (weighting factors from 0.32 to 2.34). The interval of the total projection factors was limited downwards to 82.78 and upwards to 3311.18, which equals a limitation of the total weighting factors to the area from 0.01 to 4.0.

A calibration was made for the following characteristics:

Benefit communities basis BA statistics:

- Increase in Unemployment Benefit II recipients
- Number BCs receiving benefits in accordance with SGB II by federal states
- Number BCs receiving benefits in accordance with SGB II by number of individuals under 65 years of age in the benefit community, by west/east
- Number BCs receiving benefits in accordance with SGB II by number of children under 15 years of age in the benefit community, by west/east
- Number BCs receiving benefits in accordance with SGB II consisting of one single parent with children, by west/east

As in the previous year, an additional benchmark was included; this is the increase in Unemployment Benefit II recipients since the previous year at the level of benefit communities (477,034).

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 most frequent value of the respective variable.

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

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

Benchmark figure	Value benchmark figure	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	181	123,366	123,366
	Number BCs Hamburg	70	107,745	107,746
	Number BCs Lower-Saxony	436	332,147	332,147
	Number BCs Bremen	49	50,367	50,367
	Number BCs North Rhine-Westphalia	931	812,279	812,279
	Number BCs Hesse	271	216,808	216,808

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics	Distribution with calibrated weights
	Number BCs Rhineland Palatine	153	119,932	119,932

Table 50: Nominal distributions and distributions after calibration (BA sample, households) (continuation 1)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics	Distribution with calibrated weights
	Number BCs Baden-Württemberg	289	234,615	234,615
	Number BCs Bavaria	372	259,709	259,709
	Number BCs Saarland	65	43,632	43,632
	Number BCs Berlin	372	331,568	331,555
	Number BCs Brandenburg	234	179,084	179,081
	Number BCs Mecklenburg-Vorpommern	148	138,195	138,195
	Number BCs Saxony	326	294,155	294,149
	Number BCs Saxony-Anhalt	307	196,771	196,781
	Number BCs Thuringia	172	136,945	136,949
Number BCs receiving benefit in accordance with SGB by number of individuals under 65 years of age in the benefit community (1, 2, 3, 4, and „5 or more") and by west/east (10 categories)	Number BCs with 1 individual under 65 (west)	1,147	1,176,271	1,176,263
	Number BCs with 2 individuals under 65 (west)	698	486,072	486,070
	Number BCs with 3 individuals under 65 (west)	491	309,659	309,662
	Number BCs with 4 individuals under 65 (west)	271	193,824	193,828
	Number BCs with 5 or more individuals under 65 (west)	210	134,775	134,779
	Number BCs with 1 individual under 65 (east)	678	702,562	702,548
	Number BCs with 2 individuals under 65 (east)	388	291,769	291,769
	Number BCs with 3 individuals under 65 (east)	273	156,596	156,589
	Number BCs with 4 individuals under 65 (east)	137	83,122	83,118
	Number BCs with 5 or more individuals under 65 (east)	83	42,669	42,686
Number BCs receiving benefit in accordance with SGB II by number of individuals under 15 years of age in the benefit community (0, 1, 2, 3, "4 or more") and by west/east (10 categories)	Number BCs without children under 15 years of age (west)	1,821	1,520,533	1,520,536
	Number BCs with 1 child under 15 years of age (west)	542	420,221	420,220
	Number BCs with 2 children under 15 years of age (west)	312	242,250	242,250
	Number BCs with 3 children under 15 years of age (west)	99	83,664	83,663

Table 50: Nominal distributions and distributions after calibration (BA sample, households) (continuation 2)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics	Distribution with calibrated weights
	Number BCs with 4 or more children under 15 y. of age (west)	43	33,933	33,933
	Number BCs without children under 15 years of age (east)	1,125	933,101	933,111
	Number BCs with 1 child under 15 years of age (east)	273	209,901	209,891
	Number BCs with 2 children under 15 years of age (east)	124	96,928	96,921
	Number BCs with 3 children under 15 years of age (east)	25	26,552	26,552
	Number BCs with 4 or more children under 15 y. of age (east)	12	10,236	10,236
Number BCs receiving benefits in accordance with SGB II consisting of a single parent with children by west/east (2 categories)	Number BCs with a single parent (west)	720	459,676	459,676
	Rest BCs without a single parent (west)	2,097	1,840,924	1,840,926
	Number BCs with a single parent (east)	302	202,694	202,688
	Rest BCs without a single parent (east)	1,257	1,074,024	1,074,022

Table 51: Parameters of distribution of weights

1% percentile	129.9741
5% percentile	222.0341
10% percentile	301.1921
25% percentile	411.9451
50% percentile	630.5933
75% percentile	1163.411
90% percentile	1657.794
95% percentile	1963.006
99% percentile	2336.944
Average value	827.7941
Standard deviation	557.7468
Minimum	82.78
Maximum	3311.18
Case number	4231
Efficiency measure	**68.8%**

6.9.2 Microm sample

All private households in Germany form the population. Starting point for the calibration were modified design weights including the non-response weighting. The modified design weights were trimmed at the 5% percentile and the 95% percentile of their distribution and after that rescaled in such a way that their total resulted in the total of the untrimmed calibrated design weights. The projection factors reach from 1794.10 to 26840.67 (weighting factors from 0.18 to 2.66). The interval of the total projection factors was limited downwards to 101.09 and upwards to 73799.05, which equals a limitation of the total weighting factors to the area from 0.01 to 7.3.

A calibration was made for the following characteristics:

Benefit communities: Basis BA statistics:

- Number BCs receiving benefits in accordance with SGB II by federal states
- Number of BCs receiving benefits in accordance with SGB II by number of individuals under 65 years of age in the benefit community, by west/east
- Number of BCs receiving benefits in accordance with SGB II by number of children under 15 years of age in the benefit community, by west/east
- Number of BCs receiving benefits in accordance with SGB II consisting of a single parent with children, by west/east

Households: Basis Microcensus 2008:

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

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

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

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

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
Number BCs receiving benefits in accordance with SGB II by federal states (16 categories)	Number BCs Schleswig-Holstein	16	123,366	123,695
	Number BCs Hamburg	2	107,745	107,745
	Number BCs Lower-Saxony	47	332,147	332,417
	Number BCs Bremen	7	50,367	50,367
	Number BCs North Rhine-Westphalia	86	812,279	811,044
	Number BCs Hesse	17	216,808	216,808
	Number BCs Rhineland-Palatinate	9	119,932	119,931
	Number BCs Baden-Württemberg	15	234,615	234,615
	Number BCs Bavaria	46	259,709	259,857
	Number BCs Saarland	13	43,632	43,632
	Number BCs Berlin	13	331,568	331,576
	Number BCs Brandenburg	25	179,084	178,620
	Number BCs Mecklenburg-Vorpommern	6	138,195	138,180
	Number BCs Saxony	19	294,155	294,282
	Number BCs Saxony-Anhalt	27	196,771	196,759
	Number BCs Thuringia	12	136,945	136,941
Number BCs receiving benefits in accordance with SGB II by number of individuals under 65 years of age in the benefit community (1, 2, 3, 4, and "5 or more") and by west/east (10 categories)	Number BCs with 1 individual under 65 (west)	73	1,176,271	1,176,265
	Number BCs with 2 individuals under 65 (west)	60	486,072	484,356
	Number BCs with 3 individuals under 65 (west)	55	309,659	310,563
	Number BCs with 4 individuals under 65 (west)	41	193,824	194,156
	Number BCs with 5 or more individuals under 65 (west)	29	134,775	134,772
	Number BCs with 1 individual under 65 (east)	25	702,562	702,376
	Number BCs with 2 individuals under 65 (east)	33	291,769	291,724
	Number BCs with 3 individuals under 65 (east)	29	156,596	156,308
	Number BCs with 4 individuals under 65 (east)	7	83,122	83,215

Table 52: Nominal distributions and distributions after calibration (Microm sample, households) (continuation 1)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number BCs with 5 or more individuals under 65 (east)	8	42,669	42,735
Number BCs receiving benefits in accordance with SGB II by number of individuals under 15 years of age in the benefit community (0,1, 2, 3, "4 or more") and by west/east (10 categories)	Number BCs without children under 15 years of age (west)	139	1,520,533	1,520,053
	Number BCs with 1 child under 15 years of age (west)	47	420,221	420,220
	Number BCs with 2 children under 15 years of age (west)	48	242,250	242,243
	Number BCs with 3 children under 15 years of age (west)	13	83,664	83,664
	Number BCs with 4 or more children under 15 y. of age (west)	11	33,933	33,933
	Number BCs without children under 15 years of age (east)	70	933,101	932,784
	Number BCs with 1 child under 15 years of age (east)	16	209,901	209,862
	Number BCs with 2 children under 15 years of age (east)	8	96,928	96,924
	Number BCs with 3 children under 15 years of age (east)	7	26,552	26,550
	Number BCs with 4 or more children under 15 y. of age (east)	1	10,236	10,237
Number BCs receiving benefits in accordance with SGB II consisting of a single parent with children by west/east (2 categories)	Number BCs with a single parent (west)	68	459,676	459,676
	Rest BCs without a single parent (west)	190	1,840,924	1,840,436
	Number BCs with a single parent (east)	22	202,694	202,690
	Rest BCs without a single parent (east)	80	1,074,024	1,073,667
Number of households by federal state and BIK type (spelling: "Federal state.BIK size category")	1.1 to 1.4	21	316,000	315,870
	1.5	22	81,000	80,841
	1.6	15	66,000	65,856
	1.7	26	139,000	139,204
	1.8	19	374,000	374,233
	1.9	38	170,000	169,963
	1.10	27	208,000	208,214

Table 52: Nominal distributions and distributions after calibration (Microm sample, households) (continuation 2)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	2.10	42	957,000	957,075
	3.1 to 3.2	13	151,000	151,034
	3.3	52	318,000	318,236
	3.4	30	458,000	458,229
	3.5	31	406,000	406,175
	3.7	126	906,000	906,204
	3.8	71	541,000	541,041
	3.9	71	640,000	640,022
	3.10	41	380,000	380,148
	4.8 to 4.10	30	352,000	352,150
	5.2 to 5.3	43	355,000	354,910
	5.4	69	1,037,000	1,037,127
	5.5	80	643,000	643,176
	5.6	31	304,000	304,040
	5.7	101	867,000	867,164
	5.8	205	2,551,000	2,551,411
	5.9	49	318,000	318,236
	5.10	237	2,446,000	2,446,515
	6.1 to 6.2	30	67,000	67,039
	6.3	45	330,000	330,067
	6.4	32	237,000	237,001
	6.5 to 6.7	64	627,000	627,008
	6.8	32	462,000	462,172
	6.9	59	363,000	363,192
	6.10	55	776,000	776,071
	7.1	17	208,000	208,214
	7.2	10	97,000	97,009
	7.3	28	189,000	188,891
	7.4	19	150,000	149,851
	7.5	18	166,000	166,019
	7.6	14	79,000	78,869
	7.7	36	399,000	399,077

Table 52:   Nominal distributions and distributions after calibration (Microm sample, households) (continuation 3)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	7.8	18	326,000	326,123
	7.9 to 7.10	12	257,000	257,113
	8.1 to 8.3	39	645,000	645,148
	8.4	71	516,000	516,197
	8.5 to 8.6	43	466,000	466,116
	8.7	77	1,058,000	1,058,027
	8.8	74	786,000	785,929
	8.9	22	355,000	354,910
	8.10	102	1,064,000	1,064,337
	9.1	10	93,000	93,065
	9.2	15	263,000	263,028
	9.3	70	500,000	500,029
	9.4	89	668,000	668,020
	9.5 to 9.6	43	455,000	455,074
	9.7	117	1,007,000	1,007,157
	9.8	64	631,000	630,952
	9.9	117	721,000	721,257
	9.10	157	1,478,000	1,478,399
	10.3 to 10.5	20	154,000	154,189
	10.7 to 10.8	26	333,000	333,221
	11.10	112	1,949,000	1,949,458
	12.1 to 12.3	18	262,000	261,874
	12.4	20	250,000	250,042
	12.5 to 12.6	18	138,000	138,036
	12.7	21	122,000	121,866
	12.8	20	142,000	141,979
	12.9 to 12.10	30	323,000	323,005
	13.1 to 13.3	9	192,000	192,067
	13.4	12	179,000	179,054
	13.5 to 13.6	14	156,000	156,178
	13.7	19	104,000	104,119
	13.8	14	217,000	216,913

Table 52: Nominal distributions and distributions after calibration (Microm sample, households) (continuation 4)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	14.1	24	19,000	18,931
	14.2	46	137,000	136,853
	14.3	16	183,000	182,998
	14.4	10	259,000	259,112
	14.5	13	172,000	171,953
	14.6	17	132,000	132,120
	14.7 to 14.8	10	373,000	373,090
	14.9	7	230,000	229,928
	14.10	48	663,000	663,360
	15.1 to 15.3	28	205,000	205,082
	15.4	28	161,000	160,910
	15.5 to 15.6	17	313,000	313,145
	15.7	34	236,000	236,238
	15.8	35	281,000	281,198
	16.1 to 16.3	16	233,000	233,084
	16.4	35	264,000	264,241
	16.5 to 16.6	40	307,000	307,229
	16.7 to 16.8	35	317,000	317,091
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)	764	11,753,010	11,753,015
	Number households with 2 individuals (west)	1,115	10,484,510	10,484,721
	Number households with 3 individuals (west)	525	4,043,850	4,043,870
	Number households with 4 individuals (west)	503	3,354,560	3,354,511
	Number households with 5 or more individuals (west)	228	1,279,700	1,279,730
	Number households with 1 individual (east)	214	3,566,830	3,566,699
	Number households with 2 individuals (east)	304	3,023,290	3,023,250
	Number households with 3 individuals (east)	135	1,178,940	1,178,760

Table 52: Nominal distributions and distributions after calibration (Microm sample, households)(continuation 5)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number households with 4 individuals (east)	81	598,880	599,038
	Number households with 5 and more individuals (east)	32	153,350	153,406
Number of households by "children under 15 years of age in the household yes/no" and west/east	Number households with children under 15 (west)	878	5,799,000	5,799,211
	Number households without children under 15 (west)	2,257	25,116,000	25,116,637
	Number households with children under 15 (east)	142	1,215,000	1,215,054
	Number households without children under 15 (east)	624	7,306,000	7,306,099

Table 53: Parameters of distribution of weights

1% percentile	867.5521
5% percentile	1756.647
10% percentile	2241.188
25% percentile	4117.769
50% percentile	7659.663
75% percentile	13893.96
90% percentile	22045.54
95% percentile	25891.3
99% percentile	34531.22
Average value	10109.46
Standard deviation	8008.667
Minimum	106.1988
Maximum	73080.57
Case number	3901
Efficiency measure	**61.4%**

6.9.3 Total sample

All private households in Germany form the population. Starting point for the calibration were modified design weights including the non-response weighting. The modified design weights were trimmed at the 5% percentile and the 95% percentile of their distribution and after that

rescaled in such a way that their total resulted in the total of the untrimmed calibrated design weights. The projection factors reach from 211.64 to 19028.16 (weighting factors from 0.05 to 4.60). The interval of the total projection factors was limited downwards to 41.36 and upwards to 35156.21, which equals a limitation of the total weighting factors to the area from 0.01 to 8.5.

A calibration was made for the following characteristics:

Benefit communities basis BA statistics:

- Number BCs receiving benefits in accordance with SGB II by federal states
- Number of BCs receiving benefits in accordance with SGB II by number of individuals under 65 years of age in the benefit community, by west/east
- Number of BCs receiving benefits in accordance with SGB II by number of children under 15 years of age in the benefit community, by west/east
- Number of BCs receiving benefits in accordance with SGB II consisting of a single parent with children, by west/east

Households basis Microcensus 2008:

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

Besides that also the increase in Unemployment Benefit II recipients since the previous year at the level of benefit communities (477,034) 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 most frequent value of the respective variable.

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

Table 54: Nominal distributions and distributions after calibration (total sample, households)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
Number BCs receiving benefits in accordance with SGB II by federal states (16 categories)	Number BCs Schleswig-Holstein	197	123,366	123,970
	Number BCs Hamburg	72	107,745	105,131
	Number BCs Lower-Saxony	483	332,147	331,190
	Number BCs Bremen	56	50,367	47,804
	Number BCs North Rhine-Westphalia	1,017	812,279	813,795
	Number BCs Hesse	288	216,808	216,588
	Number BCs Rhineland-Palatinate	162	119,932	118,016
	Number BCs Baden-Württemberg	304	234,615	234,657
	Number BCs Bavaria	418	259,709	258,818
	Number BCs Saarland	78	43,632	40,953
	Number BCs Berlin	385	331,568	338,139
	Number BCs Brandenburg	259	179,084	179,032
	Number BCs Mecklenburg-Vorpommern	154	138,195	133,426
	Number BCs Saxony	345	294,155	299,359
	Number BCs Saxony-Anhalt	334	196,771	198,354
	Number BCs Thuringia	184	136,945	129,345
Number BCs receiving benefits in accordance with SGB by number of individuals under 65 years of age in the benefit community (1, 2, 3, 4, and "5 or more") and by west/east (10 categories)	Number BCs with 1 individual under 65 (west)	1,220	1,176,271	1,173,311
	Number BCs with 2 individuals under 65 (west)	758	486,072	487,197
	Number BCs with 3 individuals under 65 (west)	546	309,659	303,253
	Number BCs with 4 individuals under 65 (west)	312	193,824	190,407
	Number BCs with 5 or more individuals under 65 (west)	239	134,775	136,754
	Number BCs with 1 individual under 65 (east)	703	702,562	702,518
	Number BCs with 2 individuals under 65 (east)	421	291,769	293,264
	Number BCs with 3 individuals under 65 (east)	302	156,596	153,588
	Number BCs with 4 individuals under 65 (east)	144	83,122	87,070

Table 54: Nominal distributions and distributions after calibration (total sample, households) (continuation 1)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number BCs with 5 or more individuals under 65 (east)	91	42,669	41,215
Number BCs receiving benefit in accordance with SGB II by number of individuals under 15 years of age in the benefit community (0, 1, 2, 3, "4 or more") and by west/east (10 categories)	Number BCs without children under 15 years of age (west)	1,960	1,520,533	1,514,073
	Number BCs with 1 child under 15 years of age (west)	589	420,221	414,400
	Number BCs with 2 children under 15 years of age (west)	360	242,250	244,125
	Number BCs with 3 children under 15 years of age (west)	112	83,664	83,664
	Number BCs with 4 or more children under 15 y. of age (west)	54	33,933	34,660
	Number BCs without children under 15 years of age (east)	1,195	933,101	931,962
	Number BCs with 1 child under 15 years of age (east)	289	209,901	211,236
	Number BCs with 2 children under 15 years of age (east)	132	96,928	97,669
	Number BCs with 3 children under 15 years of age (east)	32	26,552	26,552
	Number BCs with 4 or more children under 15 y. of age (east)	13	10,236	10,236
Number BCs receiving benefits in accordance with SGB II consisting of a single parent with children by west/east (2 categories)	Number BCs with a single parent (west)	788	459,676	455,311
	Rest BCs without a single parent (west)	2,287	1,840,924	1,835,611
	Number BCs with a single parent (east)	324	202,694	202,410
	Rest BCs without a single parent (east)	1,337	1,074,024	1,075,245
Number of households by federal state and BIK type (spelling: "Federal state.BIK size category")	1.1 to 1.4	77	316,000	316,059
	1.5	33	81,000	81,015
	1.6	37	66,000	66,012
	1.7	45	139,000	139,026
	1.8	100	374,000	374,070
	1.9	66	170,000	170,032
	1.10	68	208,000	208,039

Table 54: Nominal distributions and distributions after calibration (total sample, households) (continuation 2)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	2.10	126	957,000	957,179
	3.1 to 3.2	20	151,000	151,028
	3.3	111	318,000	318,060
	3.4	74	458,000	458,086
	3.5	48	406,000	406,076
	3.7	299	906,000	906,170
	3.8	211	541,000	541,101
	3.9	127	640,000	640,120
	3.10	105	380,000	380,071
	4.8 to 4.10	91	352,000	352,066
	5.2 to 5.3	103	355,000	355,066
	5.4	135	1,037,000	1,037,194
	5.5	180	643,000	643,120
	5.6	63	304,000	304,057
	5.7	220	867,000	867,162
	5.8	557	2,551,000	2,551,477
	5.9	94	318,000	318,060
	5.10	653	2,446,000	2,446,458
	6.1 to 6.2	47	67,000	67,013
	6.3	96	330,000	330,062
	6.4	68	237,000	237,044
	6.5 to 6.7	118	627,000	627,117
	6.8	99	462,000	462,086
	6.9	95	363,000	363,068
	6.10	142	776,000	776,145
	7.1	23	208,000	208,039
	7.2	17	97,000	97,018
	7.3	40	189,000	189,035
	7.4	31	150,000	150,028
	7.5	45	166,000	166,031
	7.6	57	79,000	79,015
	7.7	61	399,000	399,075

Table 54: Nominal distributions and distributions after calibration (total sample, households) (continuation 3)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	7.8	48	326,000	326,061
	7.9 to 7.10	45	257,000	257,048
	8.1 to 8.3	70	645,000	645,121
	8.4	108	516,000	516,097
	8.5 to 8.6	54	466,000	466,087
	8.7	137	1,058,000	1,058,198
	8.8	160	786,000	786,147
	8.9	37	355,000	355,066
	8.10	233	1,064,000	1,064,199
	9.1	12	93,000	93,017
	9.2	28	263,000	263,049
	9.3	106	500,000	500,094
	9.4	133	668,000	668,125
	9.5 to 9.6	66	455,000	455,085
	9.7	164	1,007,000	1,007,188
	9.8	188	631,000	631,118
	9.9	150	721,000	721,135
	9.10	325	1,478,000	1,478,277
	10.3 to 10.5	49	154,000	154,029
	10.7 to 10.8	72	333,000	333,062
	11.10	572	1,949,000	1,949,506
	12.1 to 12.3	117	262,000	262,068
	12.4	47	250,000	250,065
	12.5 to 12.6	87	138,000	138,036
	12.7	60	122,000	122,032
	12.8	35	142,000	142,037
	12.9 to 12.10	78	323,000	323,084
	13.1 to 13.3	49	192,000	192,050
	13.4	55	179,000	179,047
	13.5 to 13.6	60	156,000	156,041
	13.7	54	104,000	104,027
	13.8	45	217,000	217,056

Table 54: **Nominal distributions and distributions after calibration (total sample, households) (continuation 4)**

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	14.1	36	19,000	19,005
	14.2	119	137,000	137,036
	14.3	29	183,000	183,048
	14.4	53	259,000	259,067
	14.5	42	172,000	172,045
	14.6	109	132,000	132,034
	14.7 to 14.8	54	373,000	373,097
	14.9	23	230,000	230,060
	14.10	155	663,000	663,173
	15.1 to 15.3	76	205,000	205,053
	15.4	68	161,000	161,042
	15.5 to 15.6	67	313,000	313,082
	15.7	161	236,000	236,061
	15.8	158	281,000	281,073
	16.1 to 16.3	41	233,000	233,061
	16.4	113	264,000	264,069
	16.5 to 16.6	101	307,000	307,080
	16.7 to 16.8	104	317,000	317,083
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)	2,158	11,753,010	11,753,070
	Number households with 2 individuals (west)	2,053	10,484,510	10,484,563
	Number households with 3 individuals (west)	1,183	4,043,850	4,043,869
	Number households with 4 individuals (west)	888	3,354,560	3,354,575
	Number households with 5 or more individuals (west)	485	1,279,700	1,279,706
	Number households with 1 individual (east)	1,053	3,566,830	3,566,802
	Number households with 2 individuals (east)	836	3,023,290	3,023,265
	Number households with 3 individuals (east)	494	1,178,940	1,178,928

Table 54: Nominal distributions and distributions after calibration (total sample, households) (continuation 5)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number households with 4 individuals (east)	264	598,880	598,873
	Number households with 5 and more individuals (east)	121	153,350	153,348
Number of households by "children under 15 years of age in the household yes/no" and west/east	Number households with children under 15 (west)	2,161	5,799,000	5,799,147
	Number households without children under 15 (west)	4,606	25,116,000	25,116,637
	Number households with children under 15 (east)	683	1,215,000	1,215,031
	Number households without children under 15 (east)	2,085	7,306,000	7,306,185

Table 55: Parameters of distribution of weights

1% percentile	68.68383
5% percentile	120.0191
10% percentile	169.9333
25% percentile	353.8263
50% percentile	801.0488
75% percentile	5528.988
90% percentile	14697.51
95% percentile	19157.05
99% percentile	23630.28
Average value	4136.025
Standard deviation	6211.77
Minimum	41.3601
Maximum	35155.65
Case number	9535
Efficiency measure	**30.7%**

6.10 Calibration to the person weight, 3rd wave, cross-section

As in the two previous waves, the person weights were calibrated under the restriction that they differ as little as possible from the calibrated household weights. The calibration is therefore not based directly on the person weights of the first wave. The calibrated

household weights were instead to some extent bequeathed to the individual household members. Following this, these input weights were calibrated at the personal level.

As in the previous year, also the increase in Unemployment Benefit II recipients since the previous year at the level of individuals between 15 and 64 years (648,988) was also included as an additional benchmark value in the total sample. Again, those cases in the two samples from wave 1 and wave 2 which, according to wave 3 of the survey, were receiving Unemployment Benefit II in July 2008 will be calibrated to the benchmark statistics of the Federal Employment Agency on receipt of Unemployment Benefit II.

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

6.10.1 BA sample

The population of the cumulated BA sample of the first three waves consists of all individuals aged 15 and over who are living in a household in which there was at least one benefit community receiving benefits in accordance with SGB II at one of the three drawing dates (in July 2006, July 2007 or July 2008). Only those individuals aged 15 and over who are living in a benefit community receiving benefits in accordance with SGB II were considered for the calibration. Individuals living in a household that does not receive benefits and individuals living in a household with at least on benefit community in accordance with SGB II but are no part of a benefit community themselves were removed from the dataset for the calibration. The weighting of these individuals was calculated in a different way (see below).

The starting point for the calibration is the calibrated household weights of the BA sample. They were trimmed at the 5% percentile and the 95% percentile of their distribution and after that rescaled in such a way that their total resulted in the total of the untrimmed calibrated household weights. The trimmed projection factors extend from 187.22 to 1927.14 (weighting factors from 0.24 to 2.51). The interval of the total projection factors was limited downwards to 144.25 and upwards to 5289.12, which equals a limitation of the total weighting factors to the area from 0.15 to 5.5.

A calibration was made for the following characteristics:

Benefit recipients basis BA statistics:

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

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

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

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

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

As in the previous year, also the increase in Unemployment Benefit II recipients since the previous year at the level of individuals between 15 and 64 years (648,988) was also included as an additional benchmark value in the total sample.

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

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

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

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics	Distribution with calibrated weights
Number of persons aged 15 and over in benefit communities receiving benefits in accordance with SGB II by federal states (16 categories)	Number of individuals in BCs Schleswig-Holstein	223	175,064	175,064
	Number individuals in BCs Hamburg	72	147,471	147,471
	Number individuals in BCs Lower Saxony	528	481,832	481,831
	Number individuals in BCs Bremen	61	70,213	70,213
	Number individuals in BCs North Rhine-Westphalia	1,141	1,184,500	1,184,502
	Number individuals in BCs Hesse	323	316,240	316,239
	Number individuals in BCs Rhineland-Palatinate	179	174,363	174,363
	Number individuals in BCs Baden-Württemberg	336	330,928	330,928
	Number individuals in BCs Bavaria	452	354,691	354,691
	Number individuals in BCs Saarland	78	61,668	61,668

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

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics	Distribution with calibrated weights
	Number individuals in BCs Berlin	454	454,535	454,535
	Number individuals in BCs Brandenburg	288	254,167	254,167
	Number individuals in BCs Mecklenburg-Vorpommern	170	195,559	195,559
	Number individuals in BCs Saxony	405	414,878	414,878
	Number individuals in BCs Saxony-Anhalt	384	282,916	282,916
	Number individuals in BCs Thuringia	201	192,959	192,959
Number of individuals in benefit communities receiving benefits in accordance with SGB II by age (15-24 and 25-64; 2 categories)	Number individuals in BCs aged 15-24	874	1,004,739	1,004,740
	Number individuals in BCs aged 25-64	4,421	4,087,245	4,087,244
Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by gender and west/east (4 categories)	Number men in BCs (west)	1,491	1,565,168	1,565,174
	Number women in BCs (west)	1,902	1,731,802	1,731,797
	Number men in BCs (east)	896	903,032	903,033
	Number women in BCs (east)	1,006	891,982	891,981
Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by "single parent yes/no" gender and west/east (8 categories)	Number single parents in BCs (west)	690	459,676	459,676
	Number single parents in BCs (east)	290	202,694	202,694
	Number non single parents in BCs (west)	2,703	2,837,294	2,837,295
	Number non single parents in BCs (east)	1,612	1,592,320	1,592,319
Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by nationality (German/not German) and west/east	Number of German individuals in BCs	4,736	4,116,358	4,127,427
	Number non-German individuals in BCs	559	961,971	964,557

Table 57:   Parameters of distribution of weights

1% percentile	164.9536
5% percentile	230.2621
10% percentile	308.6981
25% percentile	404.0884
50% percentile	735.0017
75% percentile	1280.256
90% percentile	1943.159
95% percentile	2435.064
99% percentile	3829.328
Average value	961.6589
Standard deviation	762.4775
Minimum	144.25
Maximum	5289.12
Case number	5295
Efficiency measure	**61.4%**

6.10.2 Microm sample

All individuals over 14 years of age in private households in Germany form the population. Starting point for the calibration were calibrated household weights of the Microm sample. They were trimmed at the 10% percentile and the 90% percentile of their distribution and after that rescaled in such a way that their total resulted in the total of the untrimmed calibrated household weights. The trimmed projection factors extend from 2768.07 to 26258.62 (weighting factors from 0.24 to 2.31). The interval of the total projection factors was limited downwards to 567.55 and upwards to 147563.72, which equals a limitation of the total weighting factors to the area from 0.05 to 13.0.

A calibration was made for the following characteristics:

Benefit recipients basis BA statistics:

- Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by federal states
- Number of individuals in benefit communities receiving benefits in accordance with SGB II by age (15-24 and 25-64)
- Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by gender, by west/east
- Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by "single parent yes/no", by west/east

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

Population basis Microcensus 2008:

- 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, gender and west/east

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

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

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

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

Population basis BA statistics:

- Number of unemployed persons including participants in measures by west/east-
- Number of employees covered by social security by west/east

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

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

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

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

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by federal states (16 categories)	Number individuals in BCs Schleswig-Holstein	21	175,064	174,610
	Number individuals in BCs Hamburg	1	147,471	147,302
	Number individuals in BCs Lower-Saxony	58	481,832	480,589
	Number individuals in BCs Bremen	9	70,213	69,984

Table 58: Nominal distributions and distributions after calibration (Microm sample, individuals) (continuation 1)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number individuals in BCs North Rhine-Westphalia	119	1,184,500	1,181,446
	Number individuals in BCs Hesse	23	316,240	315,281
	Number individuals in BCs Rhineland-Palatinate	16	174,363	173,807
	Number individuals in BCs Baden-Württemberg	24	330,928	329,967
	Number individuals in BCs Bavaria	62	354,691	353,784
	Number individuals in BCs Saarland	17	61,668	61,467
	Number individuals in BCs Berlin	18	454,535	453,123
	Number individuals in BCs Brandenburg	34	254,167	253,690
	Number individuals in BCs Mecklenburg-Vorpommern	8	195,559	195,016
	Number individuals in BCs Saxony	25	414,878	413,743
	Number individuals in BCs Saxony-Anhalt	41	282,916	282,270
	Number individuals in BCs Thuringia	19	192,959	192,264
Number of individuals in benefit communities receiving benefits in accordance with SGB II by age (15-24 and 25-64; 2 categories)	Number individuals in BCs age 15-24	77	1,004,739	1,003,208
	Number individuals in BCs age 25-64	418	4,087,245	4,075,136
Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by gender and west/east (4 categories)	Number men in BCs (west)	151	1,565,168	1,561,331
	Number women in BCs (west)	199	1,731,802	1,726,907
	Number men in BCs (east)	59	903,032	900,280
	Number women in BCs (east)	86	891,982	889,826

Table 58: Nominal distributions and distributions after calibration (Microm sample, individuals) (continuation 2)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by "single parent yes/no" gender and west/east (8 categories)	Number single parents in BCs (west)	68	459,676	458,183
	Number single parents in BCs (east)	20	202,694	201,955
	Number non single parents in BCs (west)	282	2,837,294	2,830,054
	Number non single parents in BCs (east)	125	1,592,320	1,588,151
Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by nationality (German/not German) and west/east	Number German individuals in BCs	419	4,116,358	4,116,250
	Number non-German individuals in BCs	76	961,971	962,094
Number of individuals aged 15 and over in private households by federal state (16 categories)	Number individuals in private households Schleswig-Holstein	272	2,400,000	2,399,916
	Number individuals in private households Hamburg	51	1,543,000	1,542,600
	Number individuals in private households Lower-Saxony	695	6,743,000	6,742,834
	Number individuals in private households Bremen	41	573,000	572,983
	Number individuals in private households North Rhine-Westphalia	1,270	15,327,000	15,327,031
	Number individuals in private households Hesse	513	5,189,000	5,189,462
	Number individuals in private households Rhineland-Palatinate	286	3,441,000	3,441,713
	Number individuals in private households Baden-Württemberg	699	9,123,000	9,123,633
	Number individuals in private households Bavaria	1,136	10,611,000	10,610,649
	Number individuals in private households Saarland	71	889,000	889,149
	Number individuals in private households Berlin	155	3,007,000	3,006,483

Table 58: **Nominal distributions and distributions after calibration (Microm sample, individuals) (continuation 3)**

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number individuals in private households Brandenburg	195	2,241,000	2,240,645
	Number individuals in private households Mecklenburg-Vorpommern	101	1,483,000	1,483,095
	Number individuals in private households Saxony	304	3,731,000	3,730,605
	Number individuals in private households Saxony-Anhalt	208	2,132,000	2,132,164
	Number individuals in private households Thuringia	210	2,023,000	2,023,008
Number of individuals aged 15 and over in private households by age (in 5-year classes), gender and west/east (56 categories)	Number men in private households (west), 15-19 years	183	1,978,000	1,968,446
	Number men in private households (west), 20-24 years	129	1,933,000	1,930,497
	Number men in private households (west), 25-29 years	109	1,928,000	1,927,047
	Number men in private households (west), 30-34 years	136	1,872,000	1,872,007
	Number men in private households (west), 35-39 years	180	2,363,000	2,363,064
	Number men in private households (west), 40-44 years	270	2,942,000	2,942,498
	Number men in private households (west), 45-49 years	271	2,605,000	2,605,497
	Number men in private households (west), 50-54 years	215	2,255,000	2,255,853
	Number men in private households (west), 55-59 years	205	2,067,000	2,068,040
	Number men in private households (west), 60-64 years	170	1,697,000	1,698,207
	Number men in private households (west), 65-69 years	194	1,970,000	1,971,883
	Number men in private households (west), 70-74 years	142	1,611,000	1,612,766
	Number men in private households (west), 75-79 years	65	1,029,000	1,030,183

Table 58:   Nominal distributions and distributions after calibration (Microm sample, individuals) (continuation 4)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number men in private households (west), 80+ years	48	947,000	948,170
	Number women in private households (west), 15-19 years	173	1,836,000	1,825,491
	Number women in private households (west), 20-24 years	136	1,833,000	1,830,958
	Number women in private households (west), 25-29 years	138	1,959,000	1,958,785
	Number women in private households (west), 30-34 years	174	1,902,000	1,902,171
	Number women in private households (west), 35-39 years	262	2,334,000	2,334,396
	Number women in private households (west), 40-44 years	335	2,850,000	2,850,380
	Number women in private households (west), 45-49 years	335	2,578,000	2,578,703
	Number women in private households (west), 50-54 years	269	2,287,000	2,288,177
	Number women in private households (west), 55-59 years	239	2,115,000	2,116,401
	Number women in private households (west), 60-64 years	174	1,746,000	1,747,779
	Number women in private households (west), 65-69 years	205	2,172,000	2,174,532
	Number women in private households (west), 70-74 years	136	1,832,000	1,834,340
	Number women in private households (west), 75-79 years	81	1,341,000	1,342,971
	Number women in private households (west), 80+ years	60	1,858,000	1,860,730
	Number men in private households (east), 15-19 years	31	434,000	432,517
	Number men in private households (east), 20-24 years	25	568,000	567,605
	Number men in private households (east), 25-29 years	28	541,000	540,857
	Number men in private households (east), 30-34 years	26	475,000	474,789

Table 58: Nominal distributions and distributions after calibration (Microm sample, individuals) (continuation 5)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number men in private households (east), 35-39 years	30	579,000	578,898
	Number men in private households (east), 40-44 years	55	728,000	727,928
	Number men in private households (east), 45-49 years	70	691,000	691,109
	Number men in private households (east), 50-54 years	61	639,000	639,043
	Number men in private households (east), 55-59 years	46	580,000	580,077
	Number men in private households (east), 60-64 years	36	436,000	436,236
	Number men in private households (east), 65-69 years	46	574,000	574,399
	Number men in private households (east), 70-74 years	31	453,000	453,322
	Number men in private households (east), 75-79 years	20	266,000	266,247
	Number men in private households (east), 80+ years	6	198,000	198,158
	Number women in private households (east), 15-19 years	38	387,000	386,086
	Number women in private households (east), 20-24 years	35	519,000	518,663
	Number women in private households (east), 25-29 years	32	497,000	496,944
	Number women in private households (east), 30-34 years	28	419,000	418,910
	Number women in private households (east), 35-39 years	52	526,000	525,883
	Number women in private households (east), 40-44 years	62	688,000	687,960
	Number women in private households (east), 45-49 years	66	672,000	672,024
	Number women in private households (east), 50-54 years	81	632,000	632,081
	Number women in private households (east), 55-59 years	82	613,000	613,137

Table 58: **Nominal distributions and distributions after calibration (Microm sample, individuals) (continuation 6)**

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number women in private households (east), 60-64 years	49	461,000	461,254
	Number women in private households (east), 65-69 years	59	649,000	649,536
	Number women in private households (east), 70-74 years	38	552,000	552,496
	Number women in private households (east), 75-79 years	23	373,000	373,364
	Number women in private households (east), 80+ years	17	466,000	466,476
Number individuals aged 15 years and over in private households by household size (1, 2, 3, 4, "5 or more individuals") and west/east (10 categories)	Number individuals in private households with 1 individual (west)	756	11,753,000	11,752,588
	Number individuals in private households with 2 individuals (west)	1,751	20,499,000	20,509,147
	Number individuals in private households with 3 individuals (west)	974	9,987,000	9,985,175
	Number individuals in private households with 4 individuals (west)	999	9,335,000	9,331,354
	Number individuals in private HH with 5 or more individuals (west)	554	4,265,000	4,261,707
	Number individuals in private households with 1 individual (east)	211	3,567,000	3,566,706
	Number individuals in private households with 2 individuals (east)	476	5,867,000	5,869,110
	Number individuals in private households with 3 individuals (east)	255	2,978,000	2,976,941
	Number individuals in private households with 4 individuals (east)	155	1,707,000	1,706,347
	Number individuals in private HH with 5 or more individuals (east)	76	497,000	496,895
Number of individuals aged 15 years and over in private households by highest school qualification and west/east (12 categories)	Number individuals in private households with highest school qualification: still pupil (west)	222	2,350,000	2,379,775
	Number individuals in private households with highest school qualification: no qualification (west)	142	1,880,000	1,904,762
	Number individuals in private HH with highest school qualification: Lower secondary school (west)	1,705	23,781,000	24,094,794

Table 58: Nominal distributions and distributions after calibration (Microm sample, individuals) (continuation 7)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number individuals in private households with highest school qualification: Intermediate secondary school; intermediate secondary school in the former GDR (west)	1,426	13,221,000	13,392,817
	Number individuals in private households with highest school qualification: university (of applied sciences) qualification (west)	1,539	13,889,000	14,067,822
	Number individuals in private households with highest school qualification: still pupil (east)	37	442,000	445,272
	Number individuals in private households with highest school qualification: no qualification (east)	19	269,000	270,954
	Number individuals in private HH with highest school qualification: Lower secondary school (east)	288	3,842,000	3,871,263
	Number individuals in private households with highest school qualification: Intermediate secondary school; intermediate secondary school in the former GDR (east)	523	6,473,000	6,521,235
	Number individuals in private households with highest school qualification: university (of applied sciences) qualification (east)	306	3,481,000	3,507,276
Number of individuals aged 15 years and over in private households by marital status and west/east (10 categories)	Number individuals in private HH with marital status: single (west)	1,170	9,714,000	9,713,419
	Number individuals in private households with marital status: married, civil partnership (west)	3,163	36,549,000	36,548,855
	Number individuals in private HH with marital status: divorced (west)	405	4,729,000	4,728,653
	Number individuals in private households with marital status: widowed (west)	296	4,848,000	4,849,044
	Number individuals in private HH with marital status: single (east)	275	3,246,000	3,245,413
	Number individuals in private households with marital status: married, civil partnership (east)	687	8,578,000	8,577,406
	Number individuals in private households with marital status: divorced (east)	116	1,499,000	1,498,954

Table 58: Nominal distributions and distributions after calibration (Microm sample, individuals) (continuation 8)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number individuals in private households with marital status: widowed (east)	95	1,294,000	1,294,227
Number of individuals aged 15 years and over in private households by nationality and west/east	Number individuals in private households German	5,947	64,137,000	64,138,315
	Number individuals in private households non-German	260	6,318,000	6,317,656
Unemployed persons incl. participants in measures west/east	Unemployed persons incl. participants in measures (west)	322	3,213,295	3,213,970
	Unemployed persons incl. participants in measures (east)	142	1,618,732	1,619,007
Employees subject to social security contributions west/east	Employees subject to social security contributions (west)	2,042	22,205,091	22,205,091
	Employees subject to social security contributions (east)	480	5,178,108	5,178,108

Table 59: Parameters of distribution of weights

1% percentile	1269.838
5% percentile	2001.988
10% percentile	2597.447
25% percentile	4170.006
50% percentile	7965.598
75% percentile	14576.53
90% percentile	24197.6
95% percentile	30944.07
99% percentile	53203.17
Average value	11351.05
Standard deviation	11179.41
Minimum	567.55
Maximum	147563.7
Case number	6207
Efficiency measure	**50.8%**

6.10.3 Total sample

As for the Microm sample, all individuals of aged 15 and over in private households in Germany form the population. Starting point for the calibration were calibrated household weights of the total sample. They were trimmed at the 10% percentile and the 90% percentile of their distribution and after that rescaled in such a way that their total resulted in the total of the untrimmed calibrated household weights. The trimmed projection factors extend from 222.42 to 19648.46 (weighting factors from 0.04 to 3.75). The interval of the total projection factors was limited downwards to 52.43 and upwards to 52426.52, which equals a limitation of the total weighting factors to the area from 0.01 to 10.0.

A calibration was made for the following characteristics:

Benefit recipients basis BA statistics:

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

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

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

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

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

Population basis Mikrocensus 2008:

- 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, gender and west/east

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

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

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

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

Population basis BA statistics:

- Number of unemployed persons including participants in measures by west/east

- Number of employees covered by social security by west/east

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

Besides that, also the increase in Unemployment Benefit II recipients since the previous year at the level of individuals between 15 and 64 years (648,988) 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 most frequent value of the respective variable.

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

Table 60: Nominal distributions and distributions after calibration (total sample, individuals)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by federal states (16 categories)	Number individuals in BCs Schleswig-Holstein	244	175,064	174,668
	Number individuals in BCs Hamburg	73	147,471	147,243
	Number individuals in BCs Lower-Saxony	586	481,832	480,636
	Number individuals in BCs Bremen	70	70,213	70,004
	Number individuals in BCs North Rhine-Westphalia	1,260	1,184,500	1,181,743
	Number individuals in BCs Hesse	346	316,240	315,493
	Number individuals in BCs Rhineland-Palatinate	195	174,363	173,824
	Number individuals in BCs Baden-Württemberg	360	330,928	330,083
	Number individuals in BCs Bavaria	514	354,691	353,830
	Number individuals in BCs Saarland	95	61,668	61,507
	Number individuals in BCs Berlin	472	454,535	453,559

Table 60: Nominal distributions and distributions after calibration (total sample, individuals) (continuation 1)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number individuals in BCs Brandenburg	322	254,167	253,593
	Number individuals in BCs Mecklenburg-Vorpommern	178	195,559	195,000
	Number individuals in BCs Saxony	430	414,878	413,963
	Number individuals in BCs Saxony-Anhalt	425	282,916	282,237
	Number individuals in BCs Thuringia	220	192,959	192,510
Number of individuals in benefit communities receiving benefits in accordance with SGB II by age (15-24 and 25-64; 2 categories)	Number individuals in BCs aged 15-24	951	1,004,739	1,004,157
	Number individuals in BCs aged 25-64	4,839	4,087,245	4,075,739
Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by gender and west/east (4 categories)	Number men in BCs (west)	1,642	1,565,168	1,561,349
	Number women in BCs (west)	2,101	1,731,802	1,727,684
	Number men in BCs (east)	955	903,032	900,819
	Number women in BCs (east)	1,092	891,982	890,044
Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by "single parent yes/no" gender and west/east (8 categories)	Number single parents in BCs (west)	758	459,676	458,214
	Number single parents in BCs (east)	310	202,694	202,082
	Number non single parents in BCs (west)	2,985	2,837,294	2,830,818
	Number non single parents in BCs (east)	1,737	1,592,320	1,588,781
Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by nationality (German/not German) and west/east	Number German individuals in BCs	5,155	4,116,358	4,117,533
	Number non-German individuals in BCs	635	961,971	962,362

Table 60: Nominal distributions and distributions after calibration (total sample, individuals) (continuation 2)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
Number of individuals aged 15 and over in private households by federal state (16 categories)	Number individuals in private households Schleswig-Holstein	604	2,400,000	2,400,106
	Number individuals in private households Hamburg	143	1,543,000	1,542,235
	Number individuals in private households Lower-Saxony	1,416	6,743,000	6,743,028
	Number individuals in private households Bremen	122	573,000	572,991
	Number individuals in private households North Rhine-Westphalia	2,802	15,327,000	15,326,658
	Number individuals in private households Hesse	951	5,189,000	5,189,359
	Number individuals in private households Rhineland-Palatinate	530	3,441,000	3,441,734
	Number individuals in private households Baden-Württemberg	1,179	9,123,000	9,123,526
	Number individuals in private households Bavaria	1,795	10,611,000	10,611,113
	Number individuals in private households Saarland	166	889,000	889,149
	Number individuals in private households Berlin	722	3,007,000	3,006,510
	Number individuals in private households Brandenburg	580	2,241,000	2,240,660
	Number individuals in private households Mecklenburg-Vorpommern	341	1,483,000	1,482,956
	Number individuals in private households Saxony	870	3,731,000	3,730,596
	Number individuals in private households Saxony-Anhalt	712	2,132,000	2,132,291
	Number individuals in private households Thuringia	506	2,023,000	2,022,986

Table 60: Nominal distributions and distributions after calibration (total sample, individuals) (continuation 3)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
Number of individuals aged 15 and over in private households by age (in 5-year classes), gender and west/east (56 categories)	Number men in private households (west), 15-19 years	366	1,978,000	1,966,053
	Number men in private households (west), 20-24 years	308	1,933,000	1,928,487
	Number men in private households (west), 25-29 years	313	1,928,000	1,925,612
	Number men in private households (west), 30-34 years	320	1,872,000	1,872,046
	Number men in private households (west), 35-39 years	388	2,363,000	2,363,314
	Number men in private households (west), 40-44 years	526	2,942,000	2,942,956
	Number men in private households (west), 45-49 years	529	2,605,000	2,606,021
	Number men in private households (west), 50-54 years	413	2,255,000	2,256,273
	Number men in private households (west), 55-59 years	407	2,067,000	2,068,481
	Number men in private households (west), 60-64 years	302	1,697,000	1,698,692
	Number men in private households (west), 65-69 years	250	1,970,000	1,972,389
	Number men in private households (west), 70-74 years	153	1,611,000	1,613,207
	Number men in private households (west), 75-79 years	68	1,029,000	1,030,409
	Number men in private households (west), 80+ years	50	947,000	948,361
	Number women in private households (west), 15-19 years	381	1,836,000	1,823,687
	Number women in private households (west), 20-24 years	340	1,833,000	1,829,468
	Number women in private households (west), 25-29 years	427	1,959,000	1,958,518
	Number women in private households (west), 30-34 years	503	1,902,000	1,902,386
	Number women in private households (west), 35-39 years	566	2,334,000	2,334,731

Table 60: Nominal distributions and distributions after calibration (total sample, individuals) (continuation 4)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number women in private households (west), 40-44 years	672	2,850,000	2,850,792
	Number women in private households (west), 45-49 years	657	2,578,000	2,579,148
	Number women in private households (west), 50-54 years	509	2,287,000	2,288,701
	Number women in private households (west), 55-59 years	426	2,115,000	2,116,905
	Number women in private households (west), 60-64 years	304	1,746,000	1,748,322
	Number women in private households (west), 65-69 years	235	2,172,000	2,175,227
	Number women in private households (west), 70-74 years	148	1,832,000	1,834,938
	Number women in private households (west), 75-79 years	83	1,341,000	1,343,455
	Number women in private households (west), 80+ years	64	1,858,000	1,861,318
	Number men in private households (east), 15-19 years	113	434,000	432,594
	Number men in private households (east), 20-24 years	160	568,000	567,021
	Number men in private households (east), 25-29 years	169	541,000	540,223
	Number men in private households (east), 30-34 years	144	475,000	474,602
	Number men in private households (east), 35-39 years	129	579,000	578,804
	Number men in private households (east), 40-44 years	174	728,000	727,931
	Number men in private households (east), 45-49 years	238	691,000	691,170
	Number men in private households (east), 50-54 years	200	639,000	639,158
	Number men in private households (east), 55-59 years	191	580,000	580,274
	Number men in private households (east), 60-64 years	94	436,000	436,368

Table 60: Nominal distributions and distributions after calibration (total sample, individuals) (continuation 5)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number men in private households (east), 65-69 years	59	574,000	574,637
	Number men in private households (east), 70-74 years	38	453,000	453,501
	Number men in private households (east), 75-79 years	20	266,000	266,350
	Number men in private households (east), 80+ years	8	198,000	198,266
	Number women in private households (east), 15-19 years	120	387,000	385,876
	Number women in private households (east), 20-24 years	151	519,000	517,979
	Number women in private households (east), 25-29 years	193	497,000	496,651
	Number women in private households (east), 30-34 years	156	419,000	418,771
	Number women in private households (east), 35-39 years	171	526,000	525,879
	Number women in private households (east), 40-44 years	219	688,000	687,950
	Number women in private households (east), 45-49 years	258	672,000	672,202
	Number women in private households (east), 50-54 years	247	632,000	632,232
	Number women in private households (east), 55-59 years	216	613,000	613,344
	Number women in private households (east), 60-64 years	104	461,000	461,462
	Number women in private households (east), 65-69 years	72	649,000	649,834
	Number women in private households (east), 70-74 years	41	552,000	552,732
	Number women in private households (east), 75-79 years	24	373,000	373,517
	Number women in private households (east), 80+ years	22	466,000	466,672

Table 60: Nominal distributions and distributions after calibration (total sample, individuals) (continuation 6)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
Number of individuals aged 15 and over in private households by household size (1, 2, 3, 4, "5 or more individuals") and west/east (10 categories)	Number individuals in private HH with 1 individual (west)	2,113	11,753,000	11,748,501
	Number individuals in private HH with 2 individuals (west)	2,960	20,499,000	20,513,222
	Number individuals in private HH with 3 individuals (west)	1,930	9,987,000	9,985,555
	Number individuals in private HH with 4 individuals (west)	1,652	9,335,000	9,331,486
	Number individuals in private HH with 5 or more individuals (west)	1,053	4,265,000	4,261,133
	Number individuals in private HH with 1 individual (east)	1,030	3,567,000	3,566,369
	Number individuals in private HH with 2 individuals (east)	1,178	5,867,000	5,869,179
	Number individuals in private HH with 3 individuals (east)	809	2,978,000	2,977,100
	Number individuals in private HH with 4 individuals (east)	471	1,707,000	1,706,453
	Number individuals in private HH with 5 or more individuals (east)	243	497,000	496,898
Number of individuals aged 15 and over in private households by highest school qualification and west/east (12 categories)	Number individuals in private HH with highest school qualification: still pupil (west)	439	2,350,000	2,379,250
	Number individuals in private HH with highest school qualification: no qualification (west)	488	1,880,000	1,904,834
	Number individuals in private HH with highest school qualification: Lower secondary school (west)	3,588	23,781,000	24,095,400
	Number individuals in private HH with highest school qualification: Intermediate secondary school; intermediate secondary school in the former GDR (west)	2,779	13,221,000	13,392,814
	Number individuals in private HH with highest school qualification: university (of applied sciences) qualification (west)	2,414	13,889,000	14,067,601
	Number individuals in private HH with highest school qualification: still pupil (east)	117	442,000	445,187

Table 60:   Nominal distributions and distributions after calibration (total sample, individuals) (continuation 7)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
	Number individuals in private HH with highest school qualification: no qualification (east)	119	269,000	270,922
	Number individuals in private HH with highest school qualification: Lower secondary school (east)	939	3,842,000	3,872,147
	Number individuals in private HH with highest school qualification: Intermediate secondary school; intermediate secondary school in the former GDR (east)	1,850	6,473,000	6,520,965
	Number individuals in private HH with highest school qualification: university (of applied sciences) qualification (east)	706	3,481,000	3,506,779
Number of individuals aged 15 and over in private households by marital status and west/east (10 categories)	Number individuals in private HH with marital status: single (west)	3,100	9,714,000	9,713,491
	Number individuals in private HH with marital status: married, civil partnership (west)	4,773	36,549,000	36,548,722
	Number individuals in private HH with marital status: divorced (west)	1,420	4,729,000	4,728,660
	Number individuals in private HH with marital status: widowed (west)	415	4,848,000	4,849,026
	Number individuals in private HH with marital status: single (east)	1,470	3,246,000	3,245,419
	Number individuals in private HH with marital status: married, civil partnership (east)	1,506	8,578,000	8,577,426
	Number individuals in private HH with marital status: divorced (east)	585	1,499,000	1,498,907
	Number individuals in private HH with marital status: widowed (east)	170	1,294,000	1,294,248
Number of individuals aged 15 years and over in private households by nationality and west/east	Number individuals in private HH German	12,464	64,137,000	64,138,414
	Number individuals in private HH non-German	975	6,318,000	6,317,484

Table 60:   Nominal distributions and distributions after calibration (total sample, individuals) (continuation 8)

Benchmark figure	Value benchmark figure	Unweighted distribution	Nominal values from BA statistics and MZ 2008	Distribution with calibrated weights
Unemployed persons incl. participants in measures west/east	Unemployed persons incl. participants in measures (west)	2,435	3,213,295	3,213,962
	Unemployed persons incl. participants in measures (east)	1,418	1,618,732	1,619,109
Employees subject to social security contributions west/east	Employees subject to social security contributions (west)	3,389	22,205,091	22,205,091
	Employees subject to social security contributions (east)	1,285	5,178,108	5,178,108

Table 61:   Parameters of distribution of weights

1% percentile	61.5697
5% percentile	114.8264
10% percentile	165.983
25% percentile	324.2362
50% percentile	1287.922
75% percentile	7187.478
90% percentile	15795.81
95% percentile	21377.3
99% percentile	40675.11
Average value	5242.644
Standard deviation	8205.737
Minimum	52.43
Maximum	52426.52
Case number	13439
Efficiency measure	**29.0%**

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

Finally, also in wave 3 some households and individuals remained that could not be assigned a BA cross-sectional household weight or a BA cross-sectional person weight by means of calibration. The number of these households is now larger in wave 3 than in wave 2, since a larger part of the BA sample of wave 1 meanwhile withdrew from receiving benefits. These are the following three groups which did not receive benefits in July 2008 but which belong to the population of the BA sample (households with receipt of Unemployment Benefit II in 7/2006 or 7/2007 or 7/2008 and individuals in households with receipt of Unemployment Benefit II in 7/2006 or 7/2007 or 7/2008).

- From the refreshment sample: Individuals in the household who are not members of a benefit community: Here, the person weight was obtained from the BA household weight in wave 3 after calibration (*wqbahh*) by dividing it by the proportion of these individuals who gave a personal or senior citizens' interview – provided that their household was participating.

- Wave 2 households in which nobody was in receipt of Unemployment Benefit II any longer in July 2008: The household retains the BA weight before calibration. Individuals in these households with interviews in both waves were given a new BA person weight, which is obtained by multiplying their old BA person weight from the previous wave by the reciprocal re-participation probability *ppbleib*. Individuals in these households who did not provide a personal interview in wave 2 are given a new BA person weight calculated by dividing the BA household weight of their household for wave 3 by the proportion of such individuals who participate provided that their household is taking part.

- Individuals who are not members of a benefit community in wave 1 and 2 households that were still in receipt of Unemployment Benefit II in July 2008: Individuals in these households with interviews in both waves were given a new BA person weight which is obtained by multiplying their BA person weight from the previous wave by the reciprocal re-participation probability *ppbleib*.

7 Appendix: Brief description of the dataset

Content characteristics

Categories	Comments
Topics/characteristics categories	**Socio-demographic characteristics:** Artificial individual ID; gender; 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:** Status of employment/ economic inactivity; mini-job; working hours; occupational status (detailed); occupation (ISCO-88 and KldB-92); ISCO-based measures of occupational status and prestige (ISEI, SIOPS, MPS, EGP, ESeC); earned income (gross and net); employment biographies with employment/unemployment spells and periods of economic inactivity since 01/2005 (from wave 2 onwards); fixed-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 01/2005 (wave 1 only); information on periods of Unemployment Benefit I receipt in the context of registered unemployment since 01/2005 (from wave 2 onwards); amount of benefit; reason for end; Unemployment Benefit II: start and end dates of the spell of benefit receipt since 01/2005; reason for end; identification of household members receiving benefits; amount of benefits received; benefit cuts (start date, duration, reasons, which household members' benefit cut); Participation in measures: type of measure; start and end dates of measure; indicator of dropout; reasons for dropout; type of access to measure; assessment of measure; working hours in measure; comparison to regular employment; economic sector/industry; 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; gender role orientation; subjective social position (top-bottom scale); subjective assessment of health state

Categories	Comments
Data unit	Individuals and households in receipt of Unemployment Benefit II in 7/2006 (sample I) Individuals and households in the resident population of Germany (sample II) Individuals and households in receipt of Unemployment Benefit II in 7/2007 but without receipt in 7/2006 (sample III; refreshment sample 1) Individuals and households in receipt of Unemployment Benefit II in 7/2008 but without receipt in 7/2006 or 7/2007 (sample IV; refreshment sample 2) Note: individuals aged 65 and over are interviewed using a shorter version of the questionnaire
Number of cases	Wave 1: Sample I: 9,386 individuals (living in 6,804 households) Sample II: 9,568 individuals (living in 5,990 households) Wave 2: Sample I: 4,753 individuals (living in 3,491 households) Sample II: 6,392 individuals (living in 3,897 households) Sample III: 1,342 individuals (living in 1,041 households) Wave 3: Sample I: 4,913 individuals (living in 3,754 households) Sample II: 6,207 individuals (living in 3,901 households) Sample III: 898 individuals (living in 694 households) Sample IV: 1,421 individuals (living in 1,186 households)
Data collection mode	CATI and CAPI CAPI interviews were conducted when a sample household could not be reached by telephone or when a personal interview was desired. 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): 7776 individuals (5664 households) N (CAPI): 5663 individuals (3871 households)

Categories	Comments
Interview languages	Wave 1: German: 18,205 individuals (12,347 households) Russian: 432 individuals (275 households) Turkish: 305 individuals (163 households) English: 12 individuals (9 households) Wave 2: German: 12,237 individuals (8,234 households) Russian: 219 individuals (156 households) Turkish: 31 individuals (39 households) English: no longer offered in wave 2 due to the low case numbers in wave 1 Wave 3: German: 13,000 individuals (9,256 households) Russian: 330 individuals (210 households) Turkish: 109 individuals (69 households)
Response rates	Wave 1: Sample I: 35.1 % Sample II: 26.6 % Total: 30.5 % Wave 2: Sample II (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: 31.3 % Total: 60,7 %

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 % Total: 83,5 %
Fieldwork period:	Wave 1: December 2006-June 2007 Wave 2: December 2007-June 2008 Wave 3: December 2008-August 2009
Period covered	Wave 1: fieldwork period and retrospective spell data as of 01/2005 Wave 2: fieldwork period and retrospective spell data from 01/2005 or the respective reference period of the spell type Wave 3: fieldwork period and retrospective spell data from 01/2006 or the respective reference period of the spell type
Time reference	Repeat interview (household panel)
Regional structure	Federal state, east/west (Further regional information is available but is not contained in the scientific use file for data protection reasons. Detailed information available on request)
Territorial allocation	At the survey date

Methodological characteristics

Categories	Comments
Survey design	Original sample wave 1: two-stage random sample with two sub-populations
	1st stage: selection of 300 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).
	2nd stage, sample I: drawing of benefit communities 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 benefit recipients within the respective postcode sector (probability proportional to size/pps). The average size of the gross sample was N=100 per postcode sector.
	2nd stage, sample II: for sample II, first a sample of residential buildings was drawn from a commercial database (Microm mosaic). This was then stratified by a stratification index contained in the database at a ratio of 4:2:1 for households with a low, medium or high status respectively. Interviewers from the surveying institute visited the selected buildings. In the event that a building accommodated several households, this was noted and then one of the households was selected by the institute as the household to be interviewed. The gross sample comprised N=100 households per postcode sector.
	Refreshment sample 1 for sample I in wave 2 (sample III): In addition to continuing the samples I which were drawn for wave 1, in the 2nd wave a refreshment sample was drawn from the register data of the Federal Employment Agency. For this, benefit communities which were in receipt of Unemployment Benefit II in July 2007 but not in July 2006 were selected. These benefit communities thus depict the inflows to benefit receipt. The sample was drawn in the postcode sectors selected for wave 1 following the procedure used in the 1st wave.
	Refreshment sample 2 for sample I in wave 3 (sample IV): Also in wave 3 a refreshment sample for sample I was drawn from the register data of the Federal Employment Agency. For this, benefit communities which were in receipt of Unemployment Benefit II in July 2008 but not in July 2007 and July 2006 were selected. These benefit communities thus depict the inflows to benefit receipt. The sample was drawn in the postcode sectors selected for wave 1 following the procedure used in the 1st wave.
Institutions involved in survey	Institute for Employment Research (IAB); TNS Infratest Sozialforschung, infas Institut für angewandte Sozialwissenschaft GmbH (data preparation and documentation wave 3)
Frequency of data collection	Annually (panel)
File format and size	STATA, SPSS (several files)

Categories	Comments
File architecture	Household dataset: HHENDDAT.dta/.sav
	Individual dataset: PENDDAT.dta/.sav
	Spell data Unemployment Benefit I: alg1_spells.dta/.sav (wave 1 only)
	Spell data Unemployment Benefit II: alg2_spells.dta/.sav
	Spell data unemployment: al_spells.dta/.sav (from wave 2 onwards)
	Spell data employment: et_spells.dta/.sav (from wave 2 onwards)
	Spell data gaps: lu_spells.dta/.sav (from wave 2 onwards)
	Spell data measures: mn_spells.dta/.sav (from wave 2 onwards)
	Spell data participation in measures: massnahmespells.dta/.sav (wave 1 only)
	Register data on households: hh_register.dta/.sav
	Register data on individuals: p_register.dta/.sav
	Weighting data on households: hweights.dta/.sav
	Weighting data on individuals: pweights.dta/.sav
	Old-age provision household-level: HAVDAT.dta/.sav (only wave 3)
	Old-age provision individual level: PAVDAT.dta/.sav (only wave 3)

Data access

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

References

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

Achatz, Juliane; Hirseland, Andreas; Promberger, Markus (2007): IAB-Panelbefragung von Haushalten im Niedrigeinkommensbereich. Entwurf für ein Rahmenkonzept, pp. 11-32 in Promberger, Markus (Ed.) Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung 'Arbeitsmarkt und Soziale Sicherung', No. 12/2007 in IAB-Forschungsbericht, Nürnberg, p. 11-32.

Andersen, Hanfried H.; Mühlbacher, Axel; Nübling, Matthias (2007a): Die SOEP-Version des SF-12 als Instrument gesundheitsökonomischer Analysen. SOEPpapers 06/2007, Deutsches Institut für Wirtschaftsforschung, Berlin.

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

Andersen, H. H., Mühlbacher, A., Nübling, M., (2007c): Entwicklung eines Verfahrens zur Berechnung der körperlichen und psychischen Summenskalen auf Basis der SOEP – Version des SF 12 (Algorithmus). Data Documentation 16, Deutsches Institut für Wirtschaftsforschung, Berlin.

Andreß, Hans-Jürgen; Burkatzki, Eckhard; Lipsmeier, Gero; Salentin, Kurt; Schulte, Katja; Strengmann-Kuhn; Wolfgang (1996): Leben in Armut. Analysen der Verhaltensweisen armer Haushalte mit Umfragedaten. Final report of the DFG project "Versorgungsstrategien privater Haushalte im unteren Einkommensbereich (VuE)", Bielefeld.

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

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

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

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

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

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

Büngeler, Kathrin; Gensicke, Miriam; Hartmann, Josef; Jäckle, Robert; Tschersich, Nikolai (about to be published): IAB-Haushaltspanel im Niedrigeinkommensbereich Welle 3 (2008/2009). Methoden- und Feldbericht. FDZ-Methodenreport, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.

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

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

Erikson, Robert; Goldthorpe, John H. (1992): The Constant Flux. A Study of Class Mobility in Industrial Society. Oxford: Clarendon Press.

Erikson, Robert; Goldthorpe, John H.; Portocarero, Lucienne (1979): Intergenerational Class Mobility in Three Western Societies: England, France and Sweden. British Journal of Sociology, Vol. 30, No. 4, p. 415-441.

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

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

Frick, Joachim R.; Göbel, Jan; Krause, Peter (n.d.): $HGEN: Generated House-hold-Level Variables. Download: [http://www.diw.de/documents/dokumentenarchiv/17/60053/hgen.pdf (8.11.2007)].

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

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

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

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

Granato, Nadia (2000): CASMIN-Bildungsklassifikation. Eine Umsetzung mit dem Mikrozensus 1996. Technical Report, ZUMA, Mannheim.

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

Halleröd, Björn (1995): The Truly Poor: Direct and Indirect Consensual Measurement of Poverty in Sweden. Journal of European Social Policy, Vol. 5, No. 2, p. 111-129.

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

Hartmann, Josef; Brink, Kathrin; Jäckle, Robert; Tschersich, Nikolai (2008): IAB-Haushaltspanel im Niedrigeinkommensbereich. Technical Report, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg. Download: [http://ideas.repec.org/p/iab/iabfme/200807.html]

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

Helberger, Christof (1988): Eine Überprüfung der Linearitätsannahme der Humankapitaltheorie. In: Bodenhöfer, Hans-Joachim (Ed.): Bildung, Beruf, Arbeitsmarkt. Berlin: Dunker & Humblot, p. 151-170.

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

Jäckle, Annette (2008): The Causes of Seam Effects in Panel Surveys. ISEP Working Paper Series 14/2008, University of Essex.

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

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

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

Lipsmeier, Gero (1999): Die Bestimmung des notwendigen Lebensstandards – Einschätzungsunterschiede und Entscheidungsprobleme. In: Zeitschrift für Soziologie, Vol. 28, No. 4, p. 281-300.

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

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

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

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

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

Porst, Rolf (1984): Haushalt und Familien 1982. Zur Erfassung und Beschreibung von Haushalts- und Familienstrukturen mit Hilfe repräsentativer Bevölkerungsumfragen. In: Zeitschrift für Soziologie, Vol. 13, No. 2, p. 164-175.

Rendtel, Ulrich; Harms, Torsten (2009): Weighting and calibration for household panels. In: Lynn, Peter (Ed.) Methodology of Longitudinal Surveys, Chichester: Wiley, p. 265-286.

Ringen, Stein (1988): Direct and Indirect Measurement of Poverty. In: Journal of Social Policy, Vol. 17, No. 3, p. 351-365.

Rose, Richard; Harrison, Eric (2007): The European Socio-Economic Classification: A New Social Class Schema for Comparative European Research. In: European Societies, Vol. 9, No. 3, p. 459-490.

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

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

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

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

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

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

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

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

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

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

Imprint

Corresponding author:
Arne Bethmann, Institute for Employment
Research, Regensburger Str. 104,
D - 90478 Nürnberg; Tel.: +49 (0) 911/179-2307
Email: arne.bethmann@iab.de

Mark Trappmann, Institute for Employment
Research, Regensburger Str. 104,
D - 90478 Nürnberg; Tel.: +49 (0) 911/179-3096
Email: mark.trappmann@iab.de