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

FDZ-Datenreport

Documentation of labour market data

08/2011
EN

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

## Datenreport Wave 4

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

Bundesagentur für Arbeit

# Codebook and Documentation of the Panel Study ‘Labour Market and Social Security’ (PASS)

## Datenreport Wave 4

Marco Berg, Ralph Cramer, Christian Dickmann, Reiner Gilberg, Birgit Jesske, Martin Kleudgen, infas Institut für angewandte Sozialwissenschaft GmbH - Arne Bethmann, Benjamin Fuchs, Daniel Gebhardt, Institut für Arbeitsmarkt- und Berufsforschung (Institute for Employment Research – 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, users of the reports can ascertain whether the data offered is suitable for their research task, on the other hand, the data can be used to prepare evaluations. This Datenreport documents the data preparation of the fourth wave of PASS and is based upon the third wave’s Datenreport: Marco Berg, Ralph Cramer, Christian Dickmann, Reiner Gilberg, Birgit Jesske, Karen Marwinski (all from the 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): Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS) Welle 3 (2008/2009), FDZ Datenreport, 06/2010 (de), Nuremberg.

## Table of Contents

## List of Tables and Figures

## Data availability

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

# 1 Introduction

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

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

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

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

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

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

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

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

This Datenreport provides an overview of the fourth survey wave, for which 11,768 individuals were interviewed in 7,848 households[2] between February 2010 and September 2010. This included 10,049 individuals and 6,945 households that had already been interviewed repeatedly in the context of PASS.[3]

The present wave-specific Datenreport[4] of wave 4 documents the wave-related aspects of the study. Following a short overview of the innovations and characteristics of the fourth

---

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

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

3 The surveying institute changed after the third wave. infas Institut für angewandte Sozialwissenschaften, Bonn, conducted the survey interviews and prepared the survey data.

4 For the first time, starting with the third wave's documentation, the report was divided into two components: a wave-specific Datenreport (including codebook) and a cross-wave user guide. The PASS project team at the IAB is responsible for creating the cross-wave user guide. As of the third wave, infas has been creating the documentation of the wave-specific Datenreport. It is based on the second wave's Datenreport. 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

wave (Chapter 1.3.), the Datenreport reports the key figures on samples and response rates of the fourth wave (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 4, divided into their respective datasets (Volume II to Volume V).

## 1.2 Instruments and interview programme

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

The survey instruments and interview programme of the fourth wave are based on those used in the third wave of PASS. However, individual questions and modules have been revised or redeveloped (see User Guide PASS Wave 3, FDZ-Datenreport 04/2011 and Chapter 1.3 for an overview)[6].

The PASS survey instruments are designed in such a way that they allow repeated interviews of individuals and households that already participated in a previous wave but also first-time interviews[7]. In order to avoid seam effects[8] in the repeat interviews and to increase data quality, dependent interviewing has been used for certain questions since the third wave to update information that the respondent had provided in the previous interview. Further-

---

contents and instruments of the survey. Moreover, it describes the structure of the scientific use file and the concept of the variable types and their names. 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. The selection was based on certain rules and is documented in detail in the methods report (Jesske & Quandt, 2011).

6 The survey instruments had to be reprogrammed due to the change of the survey institute. This included optimising individual questions and modules with unchanged contents regarding the question sequence. In the contact phase this was in particular the intro, the household and relationship matrix for the inclusion of all members of a household and in the personal questionnaire the survey loops in the earnings module.

7 First-time interviewed households include: (1) Households from the refreshment sample of the current wave (2) households which split off from households interviewed in the previous waves (split-off households) (for further explanations, please the see methods report of the fourth wave (Jesske & Quandt, 2011)).

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

more, to a great extent, information about constant characteristics was not gathered again. Unlike in the past, there is an integrated questionnaire at the household level for repeatedly interviewed households (HHalt) and for first-time interviewed households (HHneu) as of wave 4[9].

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

## 1.3 Characteristics and innovations of the fourth wave

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

The characteristics and innovations in wave 4 concern the set of questions[10] (contact phase, update of employment history information surveyed in the previous wave, chronologic survey of activities and spells in the employment module, use of a new special focus module within the field of "childcare" and cuts in existing modules), the sample, data preparation and the documentation.

### 1.3.1 Personal questionnaire

The personal questionnaire updates the employment history information surveyed in wave 2 and wave 3[11]. The old, modular survey concept by employment spells, spells of unemployment and other ongoing activities (e.g. vocational training, homemaker, retired individual) was replaced by a chronologic retrospective survey. The respondents were to report all activities and spells chronologically, starting with the oldest activity. Each spell included the start and end date. The start of the chronology was the date of the last interview for re-interviewed individuals; for new respondents this was January 2008. A change of spells was tied to:

a) a change of employment status

b) a change of employment

c) an employment interruption

d) a change of employer.

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

[10] Minor changes in the set of questions (adding, modifying or deleting individual questions) are not listed here, as well as all changes and improvements in the contact phase (cf. method report of wave 4 (Jesske & Quandt, 2011)).

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

The respondents were repeatedly notified of the demarcation definition of their activities. The repeated survey of further spells was made up to the activity which lasted until the current interview date.

The new concept explicitly allowed parallel activities and spells. The set of questions was provided with a control mechanism which showed gaps of more than one month between reported spells as error message. In case of an error message, the interviewer was prompted to clarify and correct this implausibility immediately during the interview. Furthermore, the interviewer was provided with a chronological overview of all recorded activities at the end of an employment module which the responded was asked to confirm.

The individual questions within the spells survey remained and were surveyed following the update logic as in wave 3. This means that the following pieces of information for updated employment spells are surveyed again:

(1) occupational status[12] and

(2) working hours[13] and

(3) whether previously fixed-term employments were converted to permanent employments and

(4) how the employment relationship was terminated.

Collecting information on the monthly income was new in wave 4. Until wave 3, there was only summed up information on gross and net income across all current activities at the time of the interview. In addition to the summed up survey of current employments, also spell-related information on gross income *for* each finished activity has been surveyed as of wave 4.

In case of unemployment spells, respondents were asked whether this spell still continued and, if applicable, the reasons for the end of this spell. Another piece of information surveyed for the period since the last survey was whether, and if so, how long, respondents received Unemployment Benefit I. In the case of respondents who are in receipt of Unemployment Benefit I at the time of the fourth wave interview, also information on the benefit amount is collected.

The newly integrated chronological question concept also affected the scientific use file. As of wave 4, the spell datasets on employment, unemployment and gap will be combined to one spell dataset (*bio_spell*). This comprehensive spell dataset includes the information on spells of employment, unemployment and economic inactivity collected in wave 2, wave 3 and wave 4. The integration of all spells in the respective spell datasets follows specific rules (see Chapter 5.6, 5.7, 5.8). If spells 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 inter-

[12] In case of terminated updated activities, the occupational status at the end of the activity is listed; and in case of ongoing updated activities, the current occupational status is surveyed.

[13] In case of terminated updated activities, the working hours at the end of the activity are listed; in case of ongoing updated activities, the current working hours are collected.

view in wave 2 and 3). They are 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 labour market policy measures is surveyed in the fourth wave by means of a set of questions which has been completely revised in terms of content. It only relates to one-euro jobs (*ee_spells*). The start time for these measures was January 2009 for all respondents. Updates of the previous measure spell dataset (*mn_spells*) will no longer be made as of wave 4.

Furthermore, the personal questionnaire of wave 4 did not survey the additional questions from wave 3 in the subject areas "networks" and "health". However, this will only be skipped for a short period of time. Following a rotation plan, the questions will be reintegrated in later waves.

Furthermore, the following changes were made in the personal questionnaire:

- individual questions in the "education" module for pupils[14], trainees and students[15] were deleted;
- the question for personal experiences and opinions on institutions for basic benefits for job-seekers[16] was included again in the "contact to social security institutions" module, and the relevant group of respondents for this module was limited to current recipients of Unemployment Benefit II;
- the questions PEO0500, PEO0600 and PEO0700 were deleted in the group of questions on "parental education aspiration";
- questions on religious affiliation and religiousness were included again (PD0200-PD0400);
- two new questions on migration[17] were included to assess the German language skills of respondents with a migration background.

The focus module on "old age provision", which was especially designed for wave 3, was also deleted from the set of questions of the personal questionnaire.

### 1.3.2 Household questionnaire

A focus module on the topic of childcare, which was already used in wave 3, was included. The questions on childcare were revised for use in wave 4. This concerned all-day care of children under 7 years of age who are not yet attending school[18]. The questions on childcare were supplemented by the item "childcare through private individuals"[19].

14 Questions: PB0500, PB0600, PB0700, PB0800, PB0900.
15 Question PB1900
16 Question PTK1600
17 Questions PMI1200 and PMI1300
18 Questions: HKI0250b-o
19 Questions: HKI 0205 and HKI0405

Other significant changes in the household questionnaire of wave 4 concerned the Unemployment Benefit II module. In order to decrease complexity, the former update concept of the original household's target person was abandoned. In wave 4, the start date for the update was set to the period since the last interview date. The questions for possible reasons of receipt were re-developed within the Unemployment Benefit II module. The questions for reasons (AL20550) are posed to all new respondents in wave 4 who receive Unemployment Benefit II at the interview date and to re-interviewed individuals whose receipt was recorded for the first time in wave 4, i.e. the start date of the receipt is after the interview date of wave 3, or whose receipt is ongoing and will thus be updated. The questions for reasons are not posed if re-interviewed individuals currently do not receive Unemployment Benefit II, i.e. the receipt ends before the interview date of wave 4.

Furthermore, the following changes were made in the Unemployment Benefit II module:

- The question on possible reasons at the end of receipt was modified (AL22200),
- questions on possible delays of granting of Unemployment Benefit II were newly included (AL20560 and AL20570) and
- the question on which household member's Unemployment Benefit II was cut (AL22150a-e) was deleted.

In the overall household questionnaire, the following questions were abandoned:

- the questions on own living space in hostels (HW0910 and HW0920)
- the questions on the languages in the household (HD0600 – HD1000)
- the final question on the household's current standard of living

### 1.3.3 Sample and data preparation

In wave 4, like in the previous waves, a so-called refreshment sample was drawn for the BA subsample[20]. 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 in July 2009 but not on the sampling date of the first, second or third wave (see Chapter 2.1 and, on the concept of the refreshment sample, Trappmann et al 2009: 11 et seq.). These households, which were surveyed for the first time in wave 4, can be identified via the sample indicator (*sample*).

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

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

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

# 2 Key figures

This chapter provides a brief overview of important key figures of the study, such as sample sizes (gross and net) and response rates. For the panel sample, they are represented over the course of the previous 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.
- Refreshment sample 3 (BA sample) is the name of the sample drawn from the SGB II inflow between wave 3 and wave 4.

## 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 first wave but also the first-time interviewed households from the refreshment samples of waves 2, 3 and 4. 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[21]. 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 possi-

[21] The willingness to participate in the panel is only surveyed 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.

ble selection-free access to the panel as part of PASS are described in detail in the method and field report of waves 1 to 4[22].

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 moved abroad will be removed from the panel gross in the subsequent waves. Moreover, panel losses may occur if no household interview could be conducted for one household for a period of two consecutive waves. This situation could arise for the first time at the end of the third wave and affects the panel gross in the fourth wave.

The change of the survey institute is another factor influencing the panel gross in the fourth wave. Transferring the addresses of the panel participants from the IAB to infas required the target person's permission for circulation. For detailed explanations on this procedure and the results, please refer to the methods report of wave 4 (Jesske & Quandt, 2011). The gross sample used for wave 4 comprised a total of 9,497 panel households.

The case numbers for the gross sample size of the respective survey waves and subsamples are reported in the following table. In wave 4, at least one interview could be conducted in 7,848 households of the 9,465 panel households. In addition to that, there are 748 interviewed households from the BA refreshment sample, 723 of which agreed to participate in the panel.

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

**Table 1: Panel sample on the household level by waves and subsamples[23]**

| | n | Sample BA | Microm | BA inflow 1 | BA inflow 2 | BA inflow 3 | Total |
|---|---|---|---|---|---|---|---|
| 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 | 5,851 | 5,150 | 1,010 | | | 12,011 |
| | 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 |
| Wave 4* | Panel HH gross | 3,917 | 3,618 | 861 | 1,069 | | 9,465 |
| | HH interview conducted | 2,815 | 2,977 | 563 | 745 | 748 | 7,848 |
| | of this: HH willing to participate | 2,754 | 2,933 | 554 | 727 | 723 | 7,691 |

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

The 7,848 household interviews conducted in the fourth wave equal 11,768 personal interviews. The following table lists the distribution of the respondents across the subsamples and the respective survey waves.

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

| Personal interview conducted | | Sample BA | Microm | BA inflow 1 | BA inflow 2 | BA inflow 3 | Total |
|---|---|---|---|---|---|---|---|
| **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 |
| **Wave 4*** | abs. | 3,958 | 5,016 | 786 | 983 | 1,025 | 11,768 |

Source: P_Register; Scientific Use File IAB
*Reduction of the gross sample due to objection procedures

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 these two interview languages.

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

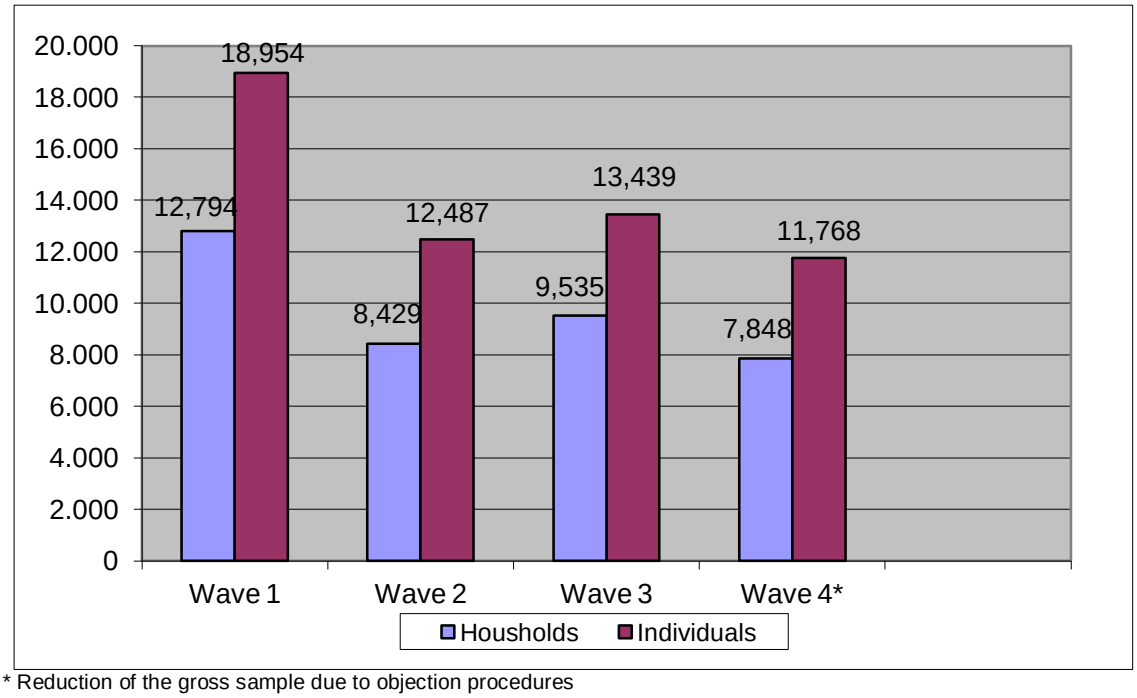
**Table 3: Panel sample size of foreign-language interviews by waves**

| | | 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 |
| Wave 4* | Households | 179 | 42 |
| | Individuals | 285 | 78 |

Source: PENDDAT; Scientific Use File IAB
*Reduction of the gross sample due to objection procedures

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

**Figure 1: Realised panel sample from households and individuals by survey waves**




* Reduction of the gross sample due to objection procedures

## 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 de-

nominator 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 passed away and households in which all members moved abroad permanently are regarded as cases of neutral non-response. Households are considered usable if at least one complete household interview is available. New households are only considered usable if not only the household interview but also at least one complete personal interview is available.

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

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

| Wave 4 | | sample | | | | | |
|---|---|---|---|---|---|---|---|
| | | **BA** | **Microm** | **BA inflow 1** | **BA inflow 2** | **BA inflow 3** | **Total** |
| HH gross | abs. | 3,917 | 3,618 | 861 | 1,096 | 2,440 | 11,932 |
| | *%* | *100.0* | *100.0* | *100.0* | *100.0* | *100.0* | *100.0* |
| neutral non-response | abs. | 12 | 7 | 5 | 3 | 19 | 46 |
| | *%* | *0.3* | *0.2* | *0.6* | *0.3* | *0.8* | *0.4* |
| HH gross filtered* | | 3,905 | 3,611 | 856 | 1,093 | 2,421 | 11,886 |
| | *%* | *100.0* | *100.0* | *100.0* | *100.0* | *100.0* | *100.0* |
| HH interview conducted | abs. | 2,815 | 2,977 | 563 | 745 | 748 | 7,100 |
| | *%* | *72.1* | *82.4* | *65.8* | *68.2* | *30.9* | *59.7* |
| of this: HH willing to participate | abs. | | | | | 723 | |
| | *%* | | | | | *29.9* | |

* HH gross - neutral non-respnses
Source: HH-Register; Scientific Use File IAB - für BA-Zugang 3: Methodendatensatz infas

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

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

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

[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.

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

| | | Sample | | | | | |
|---|---|---|---|---|---|---|---|
| | | **BA** | **Microm** | **BA inflow 1** | **BA inflow 2** | **BA inflow 2** | **Total** |
| 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 |
| Wave 4 | % | 88.4 | 88 | 90.2 | 88.3 | 89.6 | 88.5 |

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 is the proportion of individuals willing to participate in the panel with whom an interview could be conducted in the subsequent wave.

**Table 6: Proportion of personal interviews in waves 2, 3 and 4 with respondents from the previous wave willing to participate in the panel by subsamples**

| | | | Sample | | | | |
|---|---|---|---|---|---|---|---|
| | | | **BA** | **Microm** | **BA inflow 1** | **BA inflow 2** | **Total** |
| **Wave 2** | individuals willing to participate W1 | abs. | 8,925 | 8,938 | | | 17,863 |
| | re-invented 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-invented individuals in W3 | abs. | 3,365 | 4,956 | 820 | | 9,141 |
| | *Proportion* | *%* | *71.8* | *78.8* | *63.2* | | *74.5* |
| **Wave 4*** | individuals willing to participate W3 | abs. | 4,844 | 6,100 | 894 | 1,380 | 13,218 |
| | re-invented individuals in W4 | abs. | 3,287 | 4,347 | 626 | 854 | 9,114 |
| | *Proportion* | *%* | *67.9* | *71.3* | *70.0* | *61.9* | *69.0* |

Source: PENDDAT; Scientific Use File IAB
*Reduction of the gross sample due to objection procedures

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

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

Agreement to panel participation was explained in detail in Chapter 2.1 within the scope of the sample size. The agreement to participate in the panel for first-time interviewed households[25] in a wave in PASS can be illustrated as follows:

**Table 7: Agreement to panel participation of first-time interviewed households** 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 |
| **Wave 4*** | 903 | 866 | 95.9 |

* Reduction of the gross sample due to objection procedures
** HH interviewed for the first time from refreshment and splitt
Source: PENDDAT and HH_register; Scientific Use File IAB

The agreement to participate in the panel of first-time interviewed households in wave 3 was recorded following the first personal interview. The information given by this individual was then assumed for the household. If the individual agreed to participate in the panel, the household was considered willing to participate in the panel. If the individual did not agree to participate in the panel, the household was considered unwilling to participate in the panel[26].

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

[26] Hence, one individual provides the information on willingness to participate in the panel for the whole household. The information available on the household level was integrated in the individual dataset (*PENDDAT*) during data preparation. The individual respondents in the household 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 individual – therefore varying data within a household are possible. Households with at least one individual willing to participate in the panel were considered willing to participate in the panel.
As part of the updating of address information after the first personal interview in re-interviewed households, it was explained that an interview would be conducted again in the following year. If the respondent did not explicitly object to this notification, the household was considered as still agreeing to participate in the panel, and the panel variable in the individual dataset (*PENDDAT*) was updated accordingly.

In contrast to the agreement to participation, the permission to merge process data of the Federal Employment Agency with the survey data was obtained for each respondent who was interviewed using the personal questionnaire. This question does not apply to individuals aged 65 and over, because it is not included in the senior citizens' questionnaire. Agreement to merging of data is not obtained again in each new wave[27].

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

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

| | 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 |
| **Wave 4*** | 2,032 | 1,774 | 87.3 |

* Reduction of the gross sample due to objection procedures
Basis: Individials 15 to 64 years
Source: PENDDAT and HH_register; Scientific Use File IAB

## 2.4 Split-off households

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

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

From the survey date of the first to the fourth wave, a total of 343 households split off which could be interviewed subsequently in the following waves. In wave 4, 459 new split-off households were identified, 126 of which could be interviewed. The split-off households that were not surveyed will be contacted again in the fifth wave as long as they have not definitely refused to participate.

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

## 3 Dataset structure

The usual structure for 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 by specific datasets, which might have a cross-wave data structure, for example for register or spell data.

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

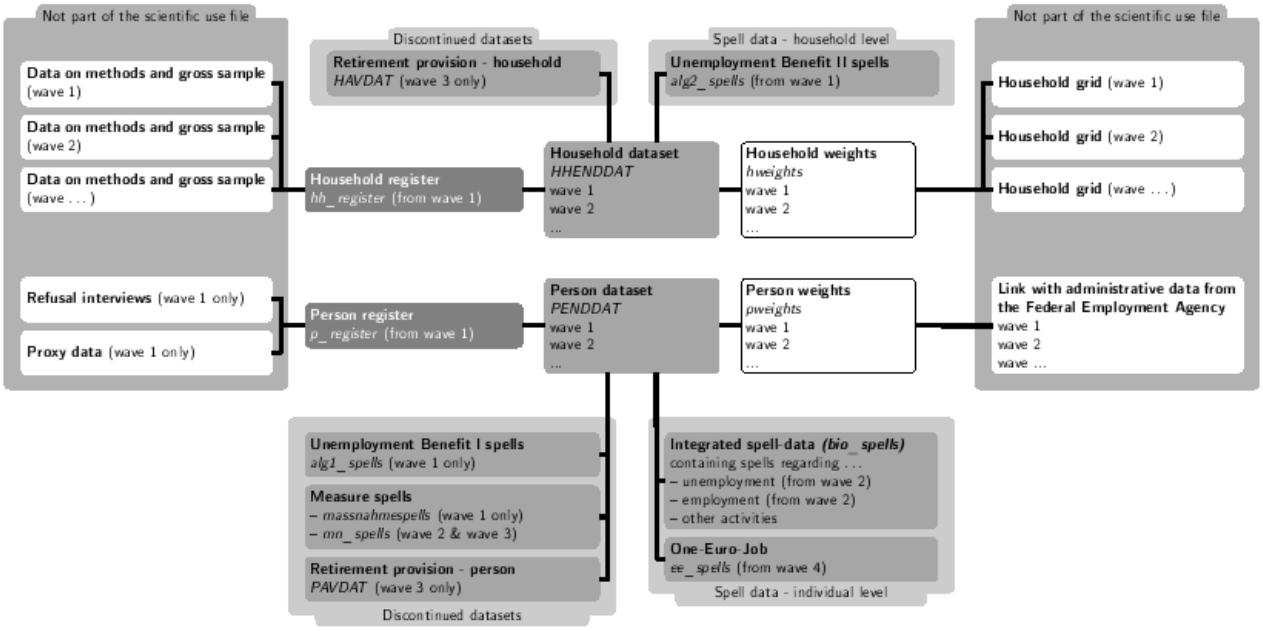
Preparing the data in long format has both advantages and disadvantages. The decisive advantage of this variant is that the data are already available in the structure required for many longitudinal analyses (such as event history analyses). It is no longer necessary to invest additional time and effort for creating a cross-wave file. The switch from long format to wide format is also quite easy to perform. STATA, for example, provides an option to switch between the two formats with little effort using the "reshape" command. Until a few years ago, the central argument against using this type of dataset structure was the significantly larger storage space required, which mainly results from the fact that even variables recorded in

only one or a small number of survey waves always require a complete column across all waves in the dataset. In addition, the long files become quite large with increasing duration of the panel, simply as a result of all annual waves being appended to one another, which significantly increases the storage space required and the time to perform individual operations using the data. The wide availability of fast processors and large storage capacities even on simple desktop PCs makes this objection seem insignificant in the meantime. Another disadvantage is the merging with additional information. Unlike the datasets prepared in wide format, an additional key variable is now required in order to be able to identify an observation clearly. This may be a wave identifier in the household or individual datasets, or alternatively the spell number in the spell datasets, which are also available in long format. Furthermore, it is not apparent at first sight which variables were surveyed for which waves, as all of the variables ever surveyed are 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, in our opinion the advantages clearly outweigh the disadvantages. Accordingly, the household and individual datasets of PASS (*HHENDDAT; PENDDAT*) and the corresponding weighting data (*hweights; pweights*) were prepared in long format.

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

**Figure 2: Dataset structure of PASS in wave 4**



# 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[28] which differed from the original variable only insofar as the information from the open-ended responses was coded to the corresponding categories where possible. Moreover, in some cases new categories were created based on the information from open-ended questions. The name of these additional variables frequently differs from that of the original variable in the last digit only, where "0" was replaced by "1". The items on country of birth, nationality and the parents'/grandparents' country of residence before migration were also anonymised and given meaningful variable names[29]. Table 9 and table 10 give an overview of the open-ended survey questions which were coded in the fourth wave[30].

**Table 9: Coding of responses to open-ended survey questions at the household level in wave 4**

| Regular variable name | Coded to variable | Dataset | Name |
|---|---|---|---|
| HD1100a-o | *HD1101a-o* | *HHENDDAT* | Employment status of HH members, proxy information, if necessary |
| HW0880 | *HW0881a-j* | *HHENDDAT* | Other reason for moving out, not listed |
| AL21300a-h – AL22100a-h | *AL21301a-h* *AL21401a-h* *AL21501a-h* *AL21601a-h* *AL21701a-h* *AL21801a-h* *AL21851a-h* *AL21901a-h* *AL22001a-h* *AL22101a-h* *AL22102a-h* *AL22103a-h* | *alg2_spells* | Other reason for benefit cut, not listed |
| AL22200a – AL22200h | *AL22201a-h* | *alg2_spells* | Other reason for discontinuation of receipt of UB II, not listed |
| AL20550a-h | *AL20551a-h* | *alg2_spells* | Other reason for why receipt of UB II started, not listed |

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

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

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

**Table 10: Coding of responses to open-ended survey questions at the individual level in wave 4**

| Regular variable name | Coded to variable | Dataset | Name |
|---|---|---|---|
| *PB0230 (code 6)* | *PB0231* | *PENDDAT* | Other German school qualification, not listed (update) |
| *PB0230 (code 7)* | *PB0231* | *PENDDAT* | Other foreign school qualification, not listed (update) |
| *PB0400 (code 9)* | *PB0401* | *PENDDAT* | Other German school qualification, not listed (first survey or not reported in previous wave) |
| *PB0400 (code 10)* | *PB0401* | *PENDDAT* | Other foreign school qualification, not listed (first survey or not reported in previous wave) |
| *PB1000* | *PB1001* | *PENNDAT* | Other foreign school qualification, not listed (first survey or not reported in previous wave) |
| *PB1300a-j (code 9)* | *PB1301a-j* | *PENDDAT* | Other German vocational qualification, not listed (update or first survey) |
| *PB1300a-j (code 10)* | *PB1301a-j* | *PENDDAT* | Other foreign vocational qualification, not listed (update or first survey) |

**Table** 10**: Coding of responses to open-ended survey questions at the individual level in wave 4 (continued)**

| | | | |
|---|---|---|---|
| *PB1600* | *PB1601* | *PENDDAT* | Other qualification to which the foreign qualification corresponds, not listed |
| AL0600 | *AL0601* | *bio_spells* | Other reason for no longer being registered as unemployed, not listed |
| BIO0100 | *BIO0101* | *bio_spells* | Other type of activity, not listed |
| *EE0300a-h* | *EE0301a-h* | *ee_spells* | Other reason for not participating in a one-euro job |
| *EE1000a-e* | *EE1001a-e* | *ee_spells* | Other reason why one-euro job was terminated prematurely |
| *PTK0320a-g* | *PTK0321a-g* | *PENDDAT* | Other reason for not having to seek employment, not listed |
| PEE0200a-d | *PEE0201a-e* | *PENDDAT* | Other source of information of one-euro jobs |
| PAS0900a-g | *PAS0901a-i* | *PENDDAT* | Other places where target pers. obtained information about job vacancies, not listed |
| PG0900a-f | *PG0901a-g* | *PENDDAT* | Other health problems, not listed |
| PG1300 | *PG1301* | *PENDDAT* | Other health insurance, not listed |
| PP1300a-e | *PP1301a-e* | *PENDDAT* | Other private caretaking activities |
| PMI0200 | *ogebland* | *PENDDAT* | Other country of birth, not listed |
| PMI0500 | *ostaatan* | *PENDDAT* | Other nationality, not listed |
| PMI1000a-f | *ozulanda-f* | *PENDDAT* | Other country from which parent/grandparent |

| | | | |
|---|---|---|---|
| | | | migrated, not listed |
| *PMI1110* | *PMI1111* | *PENDDAT* | Language spoken in circle of friends: other language, not listed |
| *PMI1120* | *PMI1121* | *PENDDAT* | Language spoken in circle of friends, equal use of two languages: first language is another language, not listed |
| *PMI1130* | *PMI1131* | *PENDDAT* | Language spoken in circle of friends, equal use of two languages: second language is another language, not listed |
| *PSH0200* | *PSH0201* | *PENDDAT* | Other German school qualification of mother, not listed |
| *PSH0200* | *PSH0201* | *PENDDAT* | Other foreign school qualification of mother, not listed |
| *PSH0300a-i (code 7)* | *PSH0301a-i* | *PENDDAT* | Other German vocational qualification of mother, not listed |
| *PSH0300a-i (code 8)* | *PSH0301a-i* | *PENDDAT* | Other foreign vocational qualification of mother, not listed |
| *PSH0500* | *PSH0501* | *PENDDAT* | Other German school qualification of father, not listed |
| *PSH0500* | *PSH0501* | *PENDDAT* | Other foreign school qualification of father, not listed |
| *PSH0600a-i (code 7)* | *PSH0601a-i* | *PENDDAT* | Other German vocational qualification of father, not listed |
| *PSH0600a-i (code 8)* | *PSH0601a-i* | *PENDDAT* | Other foreign vocational qualification of father, not listed |

## 4.2 Harmonisation

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

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

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

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

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

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

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

Therefore, it should be checked before working with the cross-wave but not harmonised variables, whether differences in the interviewed groups could cause problems for the respective evaluations and whether standardisation might be necessary[31].

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

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

| Variable | Subject area | Name |
|---|---|---|
| *isco88* | Employment | ISCO 88 (ZUMA coding), current employment, generated |
| *kldb* | Employment | Classification of occupations 1992, current employment |
| *azhpt2* | Employment | Current actual working hrs. main employment (without marginal employment, incl. cat. info.), gen. |
| *azges2* | Employment | Current total actual working hrs. (without marginal employment, incl. cat. info.), gen. |
| *befrist* | Employment | Current employment: fixed-term contract? Generated (all waves) |
| *mps* | Employment | Magnitude Prestige Scale, current employment, generated |
| *siops* | Employment | Standard International Occupational Prestige Scale, current |

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

| | | |
|---|---|---|
| | | employment, generated |
| *isei* | Employment | International Socio-Economic Index, current employment, generated |
| *egp* | Employment | Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), current occupation, generated |
| *esec* | Employment | European Socio-economic Classification (ESeC), current occupation, generated |
| *stib* | Employment | Occupational status, code number, current employment, generated |
| *netges* | Employment | Current total net income (without marginal employment, incl. cat. info.), gen. |
| *alg1abez* | Benefit receipt | Current receipt of UB I, generated |
| *aktmassn* | Participation in measures | Current participation in a programme funded/promoted by the employment agency, 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. Firstly, in some places only changes since the previous wave were to be recorded, partly depending on information on a certain set of questions already being available in the previous wave[32]. At these points, information from previous waves was used for controlling the filter. Secondly, the respondent should receive content information. In the places where changes since the previous wave were to be collected, the interview date of the previous wave was included in the question text to define the reporting period more clearly[33]. In other places, in particular where spell information was updated[34], also replies the respondent gave in the previous wave were integrated in the question texts. This was used to remind the respondent of his/her replies in the previous wave. This was to prevent that changes in status were reported which did not take place in reality but are an artefact of the open-ended survey arising from wrong memories or imprecise information.

If information from a single wave in the dataset is reviewed, only incomplete information is available for some respondents due to dependent interviewing, which only represents the changes between two survey dates. For respondents who are interviewed for the first time

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

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

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

about a certain topic there might be information available which is complete regarding this wave[35].

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 surveyed changes.

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

**Table 13a: Updated information from the previous wave in wave 4, household questionnaire**

| **Household questionnaire for re-interviewed households (HHalt)** | | | |
|---|---|---|---|
| **Construct** | **Q. no.** | **Note** | **Update in variable** |
| Housing situation | | Form of accommodation, type of tenancy and type of hostel/home/hall of residence updated during the interview | HHENDDAT: *HW0200 to HW0400* |
| Household structure | | Household size updated during the interview | HHENDDAT: *HA0100* |
| | | Sex of the individuals in the household corrected during the interview, if necessary | HHENDDAT: *HD0100a* to *HD0100o* |
| | | Age of the individuals in the household updated during the interview | HHENDDAT: *HD0200a* to *HD0200o* |
| | | Family relationships updated during the interview | not provided in the SUF |
| Year of move into current dwelling | HW0900 | Updated in generated variable | HHENDDAT: *einzugj* |
| Size of dwelling in sqm | HW1000 | Updated in generated variable | HHENDDAT: *wohnfl* |
| Receipt of Unemployment Benefit II | "Unemployment Benefit | Updated in Unemployment Benefit II spell dataset | alg2_spells: Variables of the Unem- |

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

| | | |
|---|---|---|
| 2" module | | ployment Benefit II spell dataset |
| | Information on the HH's current receipt of Unemployment Benefit II | HHENDDAT: alg2abez<br>PENDDAT: hhalg2 |
| | Information on the benefit community's Unemployment Benefit II receipt | p_register: bgbezs4; bgbezb4 |

**Table 13b Updated information from the previous wave in wave 4, personal questionnaire**

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

A distinction has to be drawn between these characteristics, where information collected in the past is updated with information on changes between the survey dates, and the so-called "constant characteristics". They are expected not to change over time. Therefore, these characteristics are recorded only once in PASS, although later corrections may be possible in some cases. Since information on these characteristics is usually only available in the surveyed variables at the date of the first interview, they are afterwards provided in the form of generated variables (see Chapter 4.4, User Guide PASS Wave 3).

## 4.4 Simple generated variables

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

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

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

| Type | Generation based on source data from | | Description |
|---|---|---|---|
| | wave of the first survey of the topic for HH/individual | current wave | |
| *unveränderlich (uv)* | yes | no | Information gathered in the first survey is generally adopted in the subsequent wave – unless input errors were corrected in the current wave.<br>Example:<br>*zpsex* (sex) |
| *fortgeschrieben (fs)* | yes | yes | Information that was current in the previous wave is combined with information of the current wave and updated, if necessary.<br>Example:<br>*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.<br>Example:<br>*hhincome* (net income of household) |

More detailed explanations must be provided on the type "unveränderlich (uv)" regarding simple generation for *PENDDAT*. A first-time survey of a topic with an individual does not always have to take place only in the first wave in which the individual gives a personal/senior citizens' interview. Two groups of individuals are again treated as first-time interviewed respondents even if they give a repeat personal/senior citizens' interview.

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

On the other hand, only an individual-related preload for dependent interviewing is created for an individual (see Chapter 4.3) if he/she gave a personal/senior citizens' interview in one of the two directly preceding waves. The background is that there shall be a distinction up to which point in time an individual should remember the results surveyed in spell form. The reference date for individuals who gave their personal/senior citizens' interview for the last time in the third preceding wave or earlier were before the relevant date for first-time interviewed respondents. In order to limit stress on the target person and assuming that the validity of the surveyed information is too severely threatened beyond this limit, individuals whose reference date for information on spell results is before the date relevant for first-time interviewed individuals are treated like first-time interviewed respondents.[36] This situation has been occurring since wave 4 as this is the first time that a previous personal interview may be more than two waves past.

For these two groups of individuals the information on which the "constant" generations are based are collected again (e.g. in the module "social origin") since these individuals are again treated as first-time interviewed individuals. Data preparation treats this surveyed information just like the information from individuals who are actually interviewed for the first time within the framework of PASS. These generations, e.g. for the status information of mother and father, are thus based on the current wave. No transfer of information from the previous wave takes place and no data is made plausible with previous information. It can basically be assumed that the information of the target persons, which are processed to become "constant" generations, is consistent with previous information in case of a repeated survey. Inconsistencies and thus deviations as compared to information from the previous waves cannot, however, be generally excluded. Individuals included in one of the two groups described can be indentified in *PENDDAT* by them being flagged in more than one wave with the code variable *altbefr* as first-time respondent (code "0" or code "-9" for wave 1).

The simple generated variables are shown in the dataset-specific Table 15 to 20. They include short descriptions of the individual variables. Furthermore, the source variables neces-

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

sary for the generation of the variable in wave 4 are indicated[37]. For the cross-section datasets (*HHENDDAT; PENDDAT*) there is additional information on which type of simple generated variables shown in Table 16 they are (uv; fs; neu). This division does not make sense for spell datasets since there are no wave-specific observations. Instead, the generated variables are newly generated at spell level if the spell was newly included in the current wave or was updated with information surveyed in the current wave. Also register datasets follow a different logic so that no further differentiation was made here.

**Table 15: Simple generated variables for wave 4 in the household dataset (*HHENDDAT*) (in alphabetical order)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *alg2abez* | *Current receipt of UB II of the HH, generated*<br>Indicator for the household's current receipt of Unemployment Benefit II (neu) | *zensiert; AL20300; AL20400; AL20500* (*alg2_spells*); information on further receipts of Unemployment Benefit II (AL22700); *hintjahr* (*HHENDDAT*) |
| *bik* | *BIK region size classes (GKBIK10), generated*<br>The information on region size class was generated by infas by converting the post code available in the address data to *GKBIK10* (neu). | Supplied by survey institute |
| *blneualt* | *Western German States or Eastern German States, generated*<br>Aggregation of German federal states into the Western German States of the former FRG (without Berlin) and the Eastern German States of the former GDR (with Berlin). Infas determined the federal states based on the post codes available from the address data (neu). | Information generated and supplied by the survey institute on the federal state in which the household is resident at the survey date. |

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

**Table 15: Simple generated variables for wave 4 in the household dataset (HHENDDAT) (in alphabetical order) (continued)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *einzugj* | *Year of move into current dwelling, generated*<br>Information as to the year in which the household moved into the current dwelling. In the case of re-interviewed households, the year of the move into the current dwelling was only asked as of the second wave if the household was living in a hostel or if it had moved house since the previous wave (fs). | For first survey:<br>*HW0900 (HHENDDAT)*<br><br>For repeated survey:<br>*einzugj* from previous wave*;*<br>*HW0900, umzug* (*HHENDAT*) |
| *hhinckat* | *Categorised household income per month (in euros), generated*<br>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 euros) incl. categorised information, generated*<br>Generation of a variable integrating information from categorised and open-ended survey questions on net household income (neu). | *HEK0600; HEK0700; HEK0800; HEK0900; HEK1000; HEK1100* (*HHENDDAT*) |
| *hintdat* | *Date of household interview*<br>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*<br>The variable indicates that at least one individual in the household is under the age of four in the wave. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this individual aged four is actually the child of another individual living in the household (neu). | *HD0200a - HD0200o (HHENDDAT)* |
| *kindu13* | *Control variable: child under age of 13 in the HH*<br>The variable indicates that at least one individual 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 individual aged 13 is actually the child of another individual living in the household (neu). | *HD0200a - HD0200o* (*HHENDDAT*) |
| *kindu15* | *Control variable: child under age of 15 in the HH*<br>The variable indicates that at least one individual in the household is under the age of 15 in the wave. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this individual aged 15 is actually the child of another individual living in the household. If the response to the open-ended question on age was missing, the categorical follow-up question about the age groups was also used to generate the variable (neu). | *HD0200a - HD0200o*; categorical follow-up question about age group (in cases of no response in *HD0200*) (*HHENDDAT*) |

**Table 15: Simple generated variables for wave 4 in the household dataset (HHENDDAT) (in alphabetical order) (continued)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *wohnfl* | *Living space in sqm, generated*<br>Information on the size of the living space in the household's current dwelling. In the case of re-interviewed households, the size of the living space was only asked as of the second wave if the household had moved house or if the house/apartment had changed since the previous wave (fs). | For first survey:<br>*HW1000 (HHENDDAT)*<br><br>For repeated survey:<br>*wohnfl* from previous wave*;*<br>*HW1000;* (*HHENDDAT*) |

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

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *akt1euro* | *Current part. in one-euro job, generated*<br>Indicator: respondent participates in a one-euro job at the interview date (neu). | *zensiert* (ee*_spells*); |
| *alakt* | *Currently reported as unemployed, generated (as of wave 2)*<br>Indicates that the TP was reported unemployed at the date of the personal interview of the respective wave (neu).<br><br>Note: In wave 4, the variable was newly created; it was also created retrospectively for waves 2 and 3 (see Chapter 5.4). | zensiert, spintegr, BIO0101 (bio_spells) |
| *alg1abez* | *Current receipt of UB I, generated*<br>Indicator: respondent is in receipt of Unemployment Benefit I as of the interview date. In the fourth wave, the periods since January 2008 during which the respondent was registered as unemployed were surveyed. For each spell additional questions were asked as to whether the respondent received UB I and if so, during which period. (neu) | *AL0700, AL1000, AL1100, AL1200*<br>(bio_*spells*) |
| *apartner* | *Control variable: cohabitee in the HH*<br>Indicator: respondent has a cohabitee or a partner whose status is not specified in the household (neu). | Information on relationships between household members (household grid); *PD0500 - PD0900* (*PENDDAT*) |
| *azhpt1* | *Current contractual working hrs. main employment (without marginal employment), gen.*<br>Weekly contractual working hours in the main employment the respondent holds at the time of the interview, generated from open-ended questions on working hours (neu).<br><br>Note: In wave 4, the variable was newly created; it was also created retrospectively for waves 2 and 3 (see Chapter 5.4). | ET2002<br>(bio_spells) |

**Table 16: Simple generated variables for wave 4 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 1)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *azhpt2* | *Current actual working hrs. main employment (without marginal employment, incl. cat. info.), gen.*<br>Weekly actual working hours in the main employment held by the respondent at the interview date, generated from responses to open-ended questions on working hours and categorical follow-up question in the case of irregular working hours (neu).<br><br>Note: In wave 4, the variable was newly created; it was also created retrospectively as of wave 1 (see Chapter 5.4). | ET2102, ET2202 (bio_spells) |
| *azges1* | *Current total contractual working hrs. (without marginal employment), gen.*<br>Weekly contractual working hours in all employments the respondent holds at the time of the interview, generated from open-ended questions on working hours (neu).<br><br>Note: In wave 4, the variable was newly created; it was also created retrospectively for waves 2 and 3 (see Chapter 5.4). | ET2002 (bio_spells) |
| *azges2* | *Current total actual working hrs. (without marginal employment, incl. cat. info.), gen.*<br>Weekly actual working hours in all employments held by the respondent at the interview date, generated from responses to open-ended questions on working hours and categorical follow-up question in the case of irregular working hours (neu).<br><br>Note: In wave 4, the variable was newly created; it was also created retrospectively as of wave 1 (see Chapter 5.4). | ET2102, ET2202 (bio_spells) |
| *befrist* | *Current employment: fixed-term contract? Gen. (all waves)*<br>Indicator: The employment held by the respondent at the interview date is on a fixed-term contract (neu). | *PET2510a; PET2510b* (*PENDDAT*) |
| *begjeewt* | *Year in which first job taken up, generated*<br>Year in which the respondent first worked in a regular employment. To generate the variable, information about the first regular employment was combined with information from the employment spells if the respondent had already reported his/her first regular employment during the questions on employment spells since January 2008 (uv). | For first survey:<br>*bjahr* (bio_*spells*); *PET3200b* (*PENDDAT*)<br><br>After first survey::<br>*begjeewt* from previous wave (*PENDDAT*) |

**Table 16: Simple generated variables for wave 4 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 2)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *begmeewt* | *Starting month of first employment, generated*<br>Month in which the respondent first worked in a regular employment (generation: see *begjeewt*) (uv). | For first survey<br>*bmonat* (*et_spells*); *PET3200a* (*PENDDAT*)<br><br>For first survey:<br>*begmeewt* from previous wave *(PENDDAT)* |
| *berabj* | *Year of the highest vocational qualification*<br>Year in which the respondent gained his/her highest vocational qualification at the interview date (fs).<br><br>Note: The years in which the vocational qualifications reported in the first wave were gained were surveyed in the second wave. | For first survey:<br>*PB1310aj-kj* (*PENDDAT*)<br><br>For repeated survey:<br>*berabj* from previous wave; *PB1310aj-kj* (*PENDDAT*) |
| *beruf1* | *Highest vocational qual., excl. foreign qual's and open info., generated*<br>Identification of the highest vocational qualification at the interview date by hierarchising the vocational qualifications cited by the respondents, excl. information from open-ended questions (fs). | For first survey:<br>*PB0100; PB0200; PB0300; PB1200b; PB1200c; PB1300a-j;* (*PENDDAT*)<br><br>For repeated survey:<br>*beruf1* from previous wave; *PB0100; PB0200; PB1200a; PB1300a-j* (*PENDDAT*) |
| *beruf2* | *Highest vocational qual., incl. foreign qual's and open info., generated*<br>As *beruf1* with the following differences:<br>1. Inclusion of responses to open-ended questions;<br>2. inclusion of information on foreign qualifications;<br>3. degrees are not distinguished by type of institution (e.g. university or other institution of higher education) but by the qualification level (Bachelor's degree; Master's degree; Ph.D.). (fs) | For first survey:<br>*PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601* (*PENDDAT*)<br><br>For repeated survey:<br>*beruf2* from previous wave; *PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601* (*PENDDAT*) |
| *brges* | *Current total gross income (without marginal employment, incl. cat. info.), gen.*<br>Contains the cumulated information on gross income from all employments (>EUR 400). Generated from answers to open-ended questions on gross income and categorical follow-up question in case of "don't know" or "details refused" answers to open-ended questions (neu)<br><br>Note: In wave 4, the variable was newly created; it was also created retrospectively for waves 2 and 3 (see Chapter 5.4). | PEK0100b; PEK0200; PEK0300; PEK0400; PEK0500; PEK0600 |

**Table 16: Simple generated variables for wave 4 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 3)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *brutto* | *Gross income incl. categorised information, gen.*<br>Generation of a variable integrating information from categorised and open-ended survey questions on gross income (neu).<br><br>Note: The variable was generally revised also retrospectively for the previous waves (see Chapter 5.4). | *PEK0100b; PEK0200; PEK0300; PEK0400; PEK0500; PEK0600;* (*PENDDAT*) |
| *bruttokat* | *Categorised gross income, generated*<br>Aggregation of the categorised information on gross income for a specific variable, combined from several items on income categories (neu).<br><br>Note: The variable was generally revised also retrospectively for the previous waves (see Chapter 5.4). | *PEK0200; PEK0300; PEK0400; PEK0500; PEK0600* (*PENDDAT*) |
| *ejhrlewt* | *Time when last employment ended (year)*<br>Last year in which the respondent was in employment. To generate this variable, information from the employment spells was combined with information on the last employment if the respondent had been out of work since January 2008 (fs). | For first survey:<br>*PET1200b* (*PENDDAT*); *ejahr; emonat* (bio_*spells*)<br><br>For repeated survey:<br>*ejhrlewt* from previous wave (*PENDDAT*); *ejahr; emonat* (bio_*spells*) |
| *ekin1517* | *Control variable: own child aged between 15 and 17 in the household*<br>This variable indicates that the respondent has a natural child, a stepchild/adopted child or a child of non-specified status aged between 15 and 17 in the household (neu). | Information on relationships between household members (household grid) |
| *ekind* | *Control variable: own child in HH*<br>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*<br>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 HH*<br>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 HH*<br>This variable indicates that the respondent has a natural child, a stepchild/adopted child or a child of non-specified status under the age of 18 in the household (neu). | Information on relationships between household members (household grid) |

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

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *emonlewt* | *Time when last employment ended (month)*<br>Last month in which the respondent was in employment. (Generation: see *ejhrlewt*) (fs). | For first survey:<br>*PET1200a* (*PENDDAT*); *emonat2* (bio_*spells*)<br><br>For repeated survey:<br>*emonlewt* from previous wave (*PENDDAT*); *emonat* (*et_spells*) |
| *epartner* | *Control variable: spouse or registered partner in HH*<br>This variable indicates that the respondent has a spouse or a same-sex registered partner in the household (neu). | Information on relationships between household members (household grid) |
| *etakt* | *Currently employed (>EUR 400 per month), generated (as of wave 2)*<br>This variable indicates that the TP had an ongoing spell of employment at the time of the personal interview of the respective wave (i.e. employment earning > EUR 400) (neu).<br><br>Note: In wave 4, the variable was newly created; it was also created retrospectively for waves 2 and 3 (see Chapter 5.4). | zensiert, spintegr, BIO0101 (bio_spells) |
| *famstand* | *Marital status, generated*<br>Generation of a marital status variable integrating information from the personal questionnaire and the control variable *epartner* generated from the household dataset (neu). | *epartner; PD0500; PD0700* (*PENDDAT*) |
| *gebhalbj* | *Half-year of birth, generated*<br>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*<br>This variable indicates that the household is receiving Unemployment Benefit II at the time of HH interview (neu). | *HA250b* (*HHENDDAT*)<br>*AL20400; AL20500* (*alg2_spells*) |
| *kindzges* | *Total number of own children (living in and outside the household), generated*<br>Total number of the respondent's children including the children living in his/her household and the children living outside the household (neu). | Information on relationships between household members (household grid);<br>*PD0900; PD1000; PD1100* (*PENDDAT*) |
| *kindzihh* | *Number of own children in the household, generated*<br>Variable generated on the basis of the responses in the household questionnaire concerning the number of children that an individual in the household has (total number of individuals in the household (half) matrix who count as children of the respondent plus the number of individuals in the household (half) matrix for whom the respondent is classified as being a parent) (neu). | Information on relationships between household members (household grid) |

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

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| | *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.* | |
| *mberuf1* | *Highest vocational qualification attained by the mother, incl. mother in the HH, excl. information from open-ended survey questions, gen.*<br>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 had been asked of all newly interviewed individuals, irrespective of whether the mother was living in the household or not.<br>For people taking part in a repeat interview as of the second wave, the values were taken over from the generated variable *mberuf1* from the previous wave (uv). | For first survey:<br>*PSH0300a-i* (*PENDDAT*)<br><br>After first survey:<br>*mberuf1* from previous wave (*PENDDAT*) |
| *mberuf2* | *Highest vocational qualification attained by the mother, incl. mother in the household, incl. information from open-ended survey questions, generated*<br>Like *mberuf1* apart from the fact that responses to open-ended questions were also taken into account for the generation of *mberuf2* (uv). | For first survey:<br>*PSH0301a-i* (*PENDDAT*)<br><br>After first survey:<br>*mberuf2* from previous wave (*PENDDAT*) |
| *mhh* | *Control variable: mother living in HH*<br>Variable indicating that the respondent's natural mother, stepmother, adoptive mother or mother of non-specified status is living in the household (neu). | Information on relationships between household members (household grid) |
| *migration* | *Respondent's migration background, generated*<br>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 respondent or of either parent (third generation) (uv). | For first survey:<br>*PMI0100; PMI0700; PMI0800a-f; PMI0900a-f* (*PENDDAT*)<br><br>After first survey:<br>*migration* from previous wave (*PENDDAT*) |

**Table 16: Simple generated variables for wave 4 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 6)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| | Note: *The concept for generating this variable had been revised as of wave 2. To generate the variable in earlier waves, only the information on whether the respondent was born in Germany and on which 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 first wave was also re-generated.* | |
| *mschul1* | *Highest general school qualification attained by the mother, incl. mother in HH, excl. information from open-ended questions, generated*<br>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).<br><br>As of the second wave, the question on the mother's highest school qualification had been asked of all newly interviewed individuals, regardless of whether the mother was living in the survey household or not. | For first survey:<br>*PSH0200* (*PENDDAT*)<br><br>After first survey:<br>*mschul1* from previous wave (*PENDDAT*) |
| *mschul2* | *Highest general school qualification attained by the mother, incl. mother in HH, incl. information from open-ended questions, gen.*<br>Like *mschul1* apart from the fact that responses to open-ended survey questions were also taken into account for the generation of *mberuf2* (uv). | For first survey:<br>*PSH0201* (*PENDDAT*)<br><br>After first survey:<br>*mschul2* from previous wave (*PENDDAT*) |
| *mstib* | *Mother's occupational status, code number, gen.*<br>Detailed occupational status of mother, generated from the individual variables (uv). | For first survey:<br>*PSH0320; PSH0330; PSH0340; PSH0360; PSH0370; PSH0380* (*PENDDAT*)<br><br>After first survey:<br>*mstib (PENDDAT)* |
| *netges* | *Current total net income (without marginal employment, incl. cat. info.), gen.*<br>Contains the cumulated information on net income from all employments (>EUR 400). Generated from answers to open-ended questions on net income and categorical follow-up question as of wave 2 in case of "don't know" or "details refused" answers to open-ended questions (neu).<br><br>Note: In wave 4, the variable was newly created; it was also created retrospectively as of wave 1 (see Chapter 5.4). | PEK0700b; PEK0800; PEK0900; PEK1000; PEK1100; PEK1200 |

**Table 16: Simple generated variables for wave 4 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 7)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *netto* | *Net income incl. categorised information, generated*<br>Generation of an integrated variable from categorised and open-ended survey questions on net income (neu).<br><br>Note: The variable was generally revised also retrospectively for the previous waves (see Chapter 5.4). | *PEK0700b; PEK0800; PEK0900; PEK1000; PEK1100; PEK1200* (*PENDDAT*) |
| *nettokat* | *Categorised net income, generated*<br>Aggregation of the categorised information on net income for a specific variable, combined from several items on income categories (neu).<br><br>Note: The variable was generally revised also retrospectively for the previous waves (see Chapter 5.4). | *PEK0800; PEK0900; PEK1000; PEK1100; PEK1200* (*PENDDAT*) |
| *palter* | *Age (from PD010), generated*<br>Respondent's age, generated based on the date of birth and the date of the personal interview in the current wave (neu). | *PD0100; pintjahr, pintmon, pinttag* (*PENDDAT*) |
| *panel* | *Willingness to participate in panel*<br>(neu) | Information supplied by the survey institute regarding the households' willingness to participate in the panel. |
| *pintdat* | *Date of personal interview*<br>Generated variable with the date on which the personal interview was conducted in the form YYMMDD (neu). | *pintjahr, pintmon, pinttag* (*PENDDAT*) |
| *schul1* | *Highest school qualification, excl. foreign qualifications and information from open-ended survey questions*<br>Variable for the highest school qualification; equivalent eastern and western German qualifications were combined (e.g. EOS and Abitur); excl. information from open-ended questions (fs). | For first survey:<br>*PB0200; PB0220; PB0230; PB0300; PB0400* (*PENDDAT*)<br><br>For repeated survey:<br>*schul1* from previous wave; *PB0200; PB0220; PB0230; PB0300; PB0400* (*PENDDAT*) |
| *schul2* | *Highest school qualification, incl. foreign qualifications and information from open-ended survey questions*<br>Like schul1 with the following differences:<br>1. inclusion of responses to open-ended questions;<br>2. inclusion of information on foreign qualifications (fs). | For first survey:<br>*PB0200; PB0220; PB0231; PB0300; PB0401* (*PENDDAT*)<br><br>For repeated survey:<br>*schul2* from previous wave; *PB0200; PB0220; PB0231; PB0300; PB0401* (*PENDDAT*) |

**Table 16: Simple generated variables for wave 4 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 8)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *schulabj* | *Year in which highest school qual. was gained*<br>Year in which the respondent gained his/her highest general school qualification (fs).<br><br>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 fourth wave interview was conducted before May 2010, it was assumed that the qualification was gained in 2009, if the interview was conducted later than May, the qualification was assumed to have been gained in 2010. | For first survey:<br>*PB0220; PB0230; PB0410; pintjahr; pintmon* (*PENDDAT*)<br><br>For repeated survey:<br>*schulabj* from previous wave;<br>*PB0220; PB0230; PB0410; pintjahr; pintmon* (*PENDDAT*) |
| *statakt* | *Current main status, generated (as of wave 2)*<br>Indicates which main status the TP had at the date of the personal interview of the respective wave (neu).<br><br>Note: In wave 4, the variable was newly created; it was also created retrospectively for waves 2 and 3 (see Chapter 5.4). | zensiert, spintegr, BIO0101, az2ges (bio_spells) |
| *stib* | *Occupational status, code number, generated*<br>Generation of the detailed code number for occupational status on from the individual variables.<br><br>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 amounts of hours, the one that began first was selected (neu). | *ET0500; ET0601 ET0701; ET0801; ET0901; ET1001; ET1101; ET1201* (*et_spells*) |
| *stibeewt* | *Occupational status, first employment, code number, generated*<br>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 2008 (uv). | For first survey:<br>*PET3300b; PET3000; PET3400; PET3500; PET3600; PET3700; PET3800; PET3900 (PENDDAT)*<br>*ET0601; ET0701; ET0801; ET0901; ET1001; ET1101; ET1201* (*et_spells*)<br><br>After first survey:<br>*stibeewt* from previous wave (*PENDDAT*) |

**Table 16: Simple generated variables for wave 4 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 9)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *stibkz* | *Current occupational status, simple classification, harmonised (anonymised)*<br>Generation of the simple code number for occupational status from the individual variables (neu). | *PET1510* (*PENDDAT*) |
| *stiblewt* | *Occupational status, last employment, code number, generated*<br>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 employment if the respondent had been out of work since January 2008 (fs). | For first survey:<br>*PET1210b; PET1210; PET1220; PET1230; PET1240; PET1250; PET1260; PET1270 (PENDDAT) ET0601; ET0701; ET0801; ET0901; ET1001; ET1101; ET1200* (*et_spells*)<br><br>For repeated survey:<br>*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 HH, excl. open info., gen.*<br>Generation of variable for father's highest vocational qualification analogous to *mberuf1* (uv). | For first survey:<br>*PSH0600a-i* (*PENDDAT*)<br><br>After first survey:<br>*vberuf1* from previous wave (*PENDDAT*) |
| *vberuf2* | *Highest vocational qualification attained by the father, incl. father in the HH, incl. open info., gen.*<br>Generation of variable for father's highest vocational qualification (incl. information from open-ended survey questions) analogous to *mberuf2* (uv). | For first survey:<br>*PSH0601a-i* (*PENDDAT*)<br><br>After first survey:<br>*vberuf2* from previous wave (*PENDDAT*) |
| *vhh* | *Control variable: father living in HH*<br>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. information open info., gen.*<br>Generation of variable for father's highest general school qualification analogous to *mschul1* (uv). | For first survey:<br>*PSH0500* (*PENDDAT*)<br><br>After first survey:<br>*vschul1* from previous wave (*PENDDAT*) |
| *vschul2* | *Highest general school qualification attained by the father, incl. father in household, incl. open info., gen.*<br>Generation of variable for father's highest general school qualification (incl. information from open-ended survey questions) analogous to *mschul2* (uv). | For first survey:<br>*PSH0501* (*PENDDAT*)<br><br>After first survey:<br>*vschu2* from previous wave (*PENDDAT*) |

**Table 16: Simple generated variables for wave 4 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued 10)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *vstib* | *Father's occupational status, code number, gen.*<br>Detailed occupational status of father, generated from the individual variables (uv). | For first survey:<br>*PSH0620; PSH0630; PSH0640; PSH0660; PSH0670; PSH0680* (*PENDDAT*)<br><br>After first survey:<br>*vstib* from previous wave *(PENDDAT)* |

**Table 17: Simple generated variables for wave 4 in the spell dataset for Unemployment Benefit II (*alg2_spells*) (in the same order as in the dataset)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *bmonat* | *Spell of UB II: starting month, generated*<br>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.<br><br>Note: The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent have been included in the source variables as of the second wave.<br>Details regarding the season in which the spell began were recoded into month values as follows:<br>21 beginning of year/winter → January<br>24 spring/Easter → April<br>27 middle of year/summer → July<br>30 autumn → October<br>32 end of year → December | *AL20100* (*alg2_spells*) |
| *bjahr* | *Spell of UB II: starting year, generated*<br>Year in which the spell of Unemployment Benefit II ended.<br><br>Note: see *bmonat* | *AL20200* (*alg2_spells*) |
| *emonat* | *Spell of UB II: ending month, generated*<br>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.<br><br>Note: see *bmonat* | *AL20300* (*alg2_spells*)<br>hintmon (HHENDDAT) |
| *ejahr* | *Spell of UB II: ending year, generated*<br>Year in which the spell of Unemployment Benefit II ended. In the case of right-censored spells (i.e. | *AL20400* (*alg2_spells*)<br>hintjahr (HHENDDAT) |

spells that were still ongoing when the household was interviewed) the interview year was entered.

Note: see *bmonat*

**Table 17: Simple generated variables for wave 4 in the spell dataset for Unemployment Benefit II (alg2_spells) (in the same order as in the dataset) (continued 1)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *alg2kbma - alg2kbmh* | *UB II: 1. benefit cut: starting month, generated*<br>Month in which the reduction of Unemployment Benefit II began. To generate the variable information on the season was converted into a definite month.<br><br>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 - h). 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 as of wave 2. | 1. benefit cut: *AL21000a* (*alg2_spells*)<br><br>to 8th benefit cut: *AL21000h* (*alg2_spells*) |
| *alg2kbja – alg2kbjh* | *UB II: 1. benefit cut: starting year, generated*<br>Year in which the Unemployment Benefit II cut began.<br>Note: see *alg2kma - alg2kbmf* | 1. benefit cut: *AL21100a* (*alg2_spells*)<br>to<br>8th benefit cut:*AL21100h* (*alg2_spells*) |
| *alg2kema - alg2kemh* | *UB II: 1. benefit cut: ending month, generated*<br>Month in which the Unemployment Benefit II cut ended. To generate the variable information on the season was converted into a definite month. If the respondent reported a duration for the benefit cut, this was used to calculate the end date of the benefit cut based on the generated start date.<br>Note: see *alg2kma - alg2kbmf* | 1. benefit cut: *alg2kbma; alg2kbja; AL21200a; AL21201a; AL21202a* (*alg2_spells*)<br><br>to 8th benefit cut:*alg2kbmh; alg2kbjh; AL21200h; AL21201h; AL21202h* (*alg2_spells*) |
| *alg2keja - alg2kejf* | *UB II: 1. benefit cut: ending year, generated*<br>Year in which the Unemployment Benefit II cut ended. If the respondent reported a duration for the benefit cut, this was used to calculate the end date of the benefit cut based on the generated start date.<br>Note: see *alg2kma - alg2kbmf* | 1. benefit cut: *alg2kbma; alg2kbja; AL21200a; AL21201a; AL21202a* (*alg2_spells*)<br><br>to 8th benefit cut: *alg2kbmh; alg2kbjh; AL21200f; AL21201f; AL21202f* (*alg2_spells*) |
| *AL22150a to AL22150h* | *UB II: benefit cut: which HH member's benefit was cut, gen.*<br>This variable contains coded information about which HH members' Unemployment Benefit II was cut. It is a string variable with 15 positions. Starting from the left, each position of this variable stands for the position of one individual in the household grid. The first position of the variable, for example, indicates whether the benefit was cut for the first individual in the HH in the particular benefit cut spell, the second position shows whether the second individual's benefit was cut and so on. As the source information for the generation was only collected from the second wave onwards, all 15 positions of the question are given the code "I" | 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 17: Simple generated variables for wave 4 in the spell dataset for Unemployment Benefit II (alg2_spells) (in the same order as in the dataset) (continued 2)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| | (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 the individual's benefit-cut status.<br>Codes:<br>1 – the household member's UB II was cut<br>2 - the household member's UB II was not cut<br>W – don't know<br>K – not specified<br>T – not applicable (filter)<br>F – question mistakenly not asked<br>U – implausible value<br>I – item not recorded in wave. | |
| *zensiert* | *Spell of UB II: spell ongoing at time of last HH interview (right-censored.), generated*<br>The censoring indicator shows whether a spell was still ongoing at the time of the last household interview.<br><br>Note: A spell is regarded as censored if one of the following conditions is met:<br>(a) It is a censored spell of a household from one of the previous waves which had not been re-interviewed in the subsequent waves up to the current wave.<br>(b) A household surveyed in wave 3 reports that a spell of UB II is still ongoing on the interview date in wave 4. Or an end date is reported which is identical with the interview date in wave 4 and it is confirmed in the follow-up question that the benefit receipt is still currently ongoing.<br><br>Code -5 was given if the household reference person of the previous wave was no longer living in the household in wave 4 and was not interviewed in wave 4. | *AL20300; AL20400, AL20500* (*alg2_spells*) |

**Table 18: Simple generated variables for wave 4 in the BIO spell dataset (bio_*spells*) (in the same order as in the dataset)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *bmonat* | *Employment: starting month, generated*<br>Month in which the employment spell began. To generate the variable information on the season was converted into a definite month.<br><br>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.<br>Details regarding the season in which the spell began were recoded into months as follows:<br>21 beginning of year/winter → January<br>24 spring/Easter → April<br>27 middle of year/summer → July<br>30 autumn → October<br>32 end of year → December | *BIO0200* (bio_spells) |
| *bjahr* | *Employment: starting year, generated*<br>Year in which the employment spell began.<br><br>Note: see *bmonat* | *BIO0300* (bio_*spells*) |
| *emonat* | *Employment: ending month, generated*<br>Month in which the employment spell ended. To generate the variable information on the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the individual was interviewed) the interview month was entered.<br><br>Note: see *bmonat* | *BIO0400; BIO0600* (Bio_*spells*)<br>*pintmon* (*PENDDAT*) |
| *ejahr* | *Employment: ending year, generated*<br>Year in which the employment spell ended. For right-censored spells (i.e. spells that were still ongoing when the individual was interviewed) the interview month was entered.<br><br>Note: see *bmonat* | *BIO0500; BIO0600* (*et_spells*)<br>*pintjahr* (*PENDDAT*) |
| *zensiert* | *Employment: spell still currently ongoing (right censoring)*<br>The censoring indicator shows whether a spell was still ongoing at the time of the personal interview in the previous wave, i.e. whether it is a right-censored spell.<br><br>Note: A spell is regarded as censored if one of the two following conditions is met: The individual reports with regard to the end date of the BIO spell that the employment is still ongoing on the interview date. Or an end date is reported which is identical with the interview date and it is confirmed in the follow-up question that the activity is still currently ongoing. | *BIO0400; BIO0500; BIO0600* (*et_spells*) |

**Table 18: Simple generated variables for wave 4 in the BIO spell dataset (bio_spells) (in the same order as in the dataset) (continued 1)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *stib* | *Occupational status, code number, generated*<br>Generation of the detailed code number for occupational status on from the individual variables. | Collection of spell information in wave 4<br>*ET0602; ET0702; ET0802; ET0902; ET1002; ET1102; ET1202* (bio_*spells*)<br><br>Otherwise, the value from the previous wave remains |
| *az1* | *Weekly contractual working hours*. | Collection of spell information in wave 4<br>*ET2002* (bio_*spells*)<br><br>Otherwise, the value from the previous wave remains |
| *az2* | *Weekly working hours incl. details in the case of irregular working hours, gen.*<br>Integrated variable on weekly hours of work in the employment held by the respondent, combining responses to open-ended questions on working hours and the categorical follow-up question. For the closed categories of the follow-up question the mean values for the categories were used, for the open-ended category (40 hours or more) the median of the weekly working hours reported in the open-ended questions was used. | Collection of spell information in wave 4<br>*ET2102; ET2202* (bio_*spells*)<br><br>Otherwise, the value from the previous wave remains |
| *alg1bm* | *Receipt of UB I: starting month, generated*<br>Month in which the spell of Unemployment Benefit I receipt began. To generate the variable information on the season was converted into a definite month.<br><br>Note: Periods of receipt of Unemployment Benefit I are embedded in the spells of registered unemployment. A maximum of one period of UB I receipt is available per period of registered unemployment. The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent are included in the source variables.<br><br>Conversion of the month details, see *bmonat*. | *AL0800* (bio_*spells*) |
| *alg1bj* | *Receipt of UB I: starting year, generated*<br>Year in which the spell of Unemployment Benefit I receipt began.<br><br>Note: see *alg1bm* | *AL0900* (bio_*spells*) |

**Table 18: Simple generated variables for wave 4 in the BIO spell dataset (bio_spells) (in the same order as in the dataset) (continued 2)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *alg1em* | *Receipt of UB I: ending month, generated*<br>Month in which the spell of Unemployment Benefit I receipt ended. To generate the variable information on the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the individual was interviewed) the interview date was entered.<br><br>Note: see *alg2kma - alg2kbme* | *AL1000; AL1200* (bio_*spells*)<br>*pintmon* (*PENDDAT*) |
| *alg1ej* | *Receipt of UB I: ending year, generated*<br>Year in which the spell of Unemployment Benefit I receipt ended. In the case of right-censored spells (i.e. spells that were still ongoing when the individual was interviewed) the interview date was entered.<br><br>Note: see *alg2kma - alg2kbme* | *AL1100; AL1200* (bio_*spells*)<br>*pintjahr* (*PENDDAT*) |
| *alg1akt* | *Receipt of UB I: spell still currently ongoing (right censoring)*<br>The censoring indicator shows whether the spell of Unemployment Benefit I receipt was still ongoing at the time of the personal interview in the previous wave, i.e. whether it is a right-censored spell.<br><br>Note: A spell is regarded as censored if one of the two following conditions is met: The individual reports with regard to the end date of the spell of Unemployment Benefit I receipt that the benefit receipt is still ongoing on the interview date. Or an end date is reported which is identical with the interview date and it is confirmed in the follow-up question that benefit receipt is still currently ongoing. The variable is generated based on the generated date variables, which are checked for plausibility. | *emonat, ejahr, AL1000; AL1100; AL1200* (bio_*spells*) |

**Table 19: Simple generated variables for wave 4 in the one-euro spell dataset (ee_*spells*) (in the same order as in the dataset)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *bmonat* | *Measure: starting month, generated*<br>Month in which the measure of active labour market policy spell began. To generate the variable information on the season was converted into a definite month.<br><br>Note: The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent (apart from values identified as implausible when the range of values was checked) are included in the source variables.<br>Details regarding the season in which the spell began were recoded into months values as follows:<br>21 beginning of year/winter → January<br>24 spring/Easter → April<br>27 middle of year/summer → July<br>30 autumn → October<br>32 end of year → December | *EE0600a* (ee_*spells*) |
| *bjahr* | *Measure: starting year, generated*<br>Year in which the measure of active labour market policy spell began.<br><br>Note: see *bmonat* | *EE0600b* (*mn_spells*) |
| *emonat* | *Measure: ending month, generated*<br>Month in which the measure of active labour market policy ended. To generate the variable information on the season was converted into a definite month and for right-censored spells (i.e. spells that were still ongoing when the individual was interviewed) the interview date was entered.<br><br>Note: see *bmonat* | *EE0600a, EE0600b, EE0700, EE0800a, EE0800b* (ee_*spells*)<br>*pintmon, pintjahr* (*PENDDAT*) |
| *ejahr* | *Measure: ending year, generated*<br>Year in which the measure of active labour market policy spell ended. For right-censored spells (i.e. spells that were still ongoing when the individual was interviewed) the interview date was entered.<br><br>Note: see *bmonat* | *EE0600a, EE0600b, EE0700, EE0800a, EE0800b* (ee_*spells*)<br>*pintjahr, pintjahr* (*PENDDAT*) |
| *zensiert* | *Measure: spell still currently ongoing (right censoring)*<br>The censoring indicator shows whether a spell was still ongoing at the time of the personal interview in the previous wave, i.e. whether it is a right-censored spell. | *EE0700* (ee_*spells*) |

**Table 20: Simple generated variables for wave 4 in the person register dataset (*p_register*) (in alphabetical order)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *alter4* | *Age of individual in wave 4 (2010)*<br>Variable contains the "best" available information regarding an individual's age. This is either (a) the age calculated from the date of birth reported in wave 4 or (b) if no date of birth is available from wave 4, then the age reported in the household interview. The information from alter4 was also taken over into the household dataset and corresponds to the information in *HD0200a* to *HD0200o*. This procedure is consistent with that followed in the field. Already during the fieldwork, the age variable in the database was populated with the respective "best" information. During fieldwork, a variable in the database is first populated with the age information according to the household interview. If a personal interview is conducted, this variable in the database is overwritten with the age calculated based on the details given in the personal interview (date of birth, date of personal interview). Both the age 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 an individual contained in the household dataset of wave 4 was taken into account in the plausibility check and for the generation of the types of benefit communities and households. | *PD0100, pintjahr, pintmon, pinttag* (*PENDDAT*)<br>*HD0200a* to *HD0200o* (*HHENDDAT*) |
| *erwprox4* | *Employment status according to HH interview in wave 4 (2010)*<br>Variable is an unchanged transfer of HD1101* from the current wave from HHENDDAT. | *HD1101** |
| *korrsex* | *Info. on sex was corrected between survey waves*<br>For individuals who belonged to a sample HH in more than one wave this variable indicates whether the sex was corrected in the household interview. | *HD0100a* to *HD0100o* of all waves (*HHENDDAT*) |
| *lastint* | *Survey wave of last interview at individual level*<br>This variable indicates the wave in which the last interview at the individual level was conducted with the individual (personal interview or senior citizen's interview). | Personal interviews from all waves (*PENDDAT*) |
| *neuj4* | *Year in which individual joined current HH, reported in wave 4 (2010)*<br>This variable indicates the year the individual joined the household of which he/she is a member in the fourth wave.<br>Note: Information on the date comes from the wave 4 interview with the re-interviewed household into which the individual has moved or was born since the previous wave. | Information on the date at which an individual moved into a household. Surveyed in the household grid |

**Table20: Simple generated variables for wave 4 in the person register dataset (*p_register*) (in alphabetical order) (continued 1)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *neum4* | *Month in which individual joined current HH, reported in wave 4 (2010)*<br>This variable indicates the month the individual joined the household of which he/she is a member in the fourth wave.<br><br>Note: see *neuj4* | Information on the date at which an individual moved into a household. Surveyed in the household grid |
| *wegj4* | *Year since which individual has no longer been living in previous HH, reported in wave 4 (2010)*<br>This variable indicates the year the individual ceased to be a member of the household of the previous wave.<br><br>Note: Information on the date comes from the wave 4 interview with the household in which the individual was living in the previous wave. | Information on the date at which an individual moved out of a household. Surveyed in the household grid |
| *wegm4* | *Month since which individual has no longer been living in previous HH, reported in wave 4 (2010)*<br>This variable indicates the month the individual ceased to be a member of the household of the previous wave.<br><br>Note: see *wegj4* | Information on the date at which an individual moved out of a household. Surveyed in the household grid |
| *zdub4* | *Pointer: Personal identification no. of the individual doubled by the TP in wave 4 (2010)*<br>Indicates that an individual from an original HH currently lives in a split-off HH without the original HH having reported the move of this individual.<br><br>Note: Chapter 5.4 provides a detailed explanation on the reasons for the introduction of this variable. | Information on all household members of an original household and all of its split-off households in the household grid of the current and the previous wave |
| *zmhh4* | *Pointer: personal ID number of target person's mother in HH in wave 4 (2010)*<br>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 wave 4 (household grid) |
| *zparthh4* | *Pointer: personal ID number of target person's partner in HH in wave 4 (2010)*<br>Contains the personal identification number of a partner living in the household. Spouses, registered partners, cohabitees and partners whose status is not specified are counted as a partner. | Information on relationships between household members in wave 4 (household grid) |
| *zupanel* | *Survey wave in which individual joined panel*<br>This variable indicates the wave in which the individual was a member of a sample household for the first time. | Information on the individuals living in a household in all waves (household grid) |

**Table20: Simple generated variables for wave 4 in the person register dataset (*p_register*) (in alphabetical order) (continued 2)**

| Variable | Variable label and description | Source var. for generated var. in wave 4 |
|---|---|---|
| *zvhh4* | *Pointer: personal ID number of target person's father in HH in wave 4 (2010)*<br>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 wave 4 (household grid) |

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

**Figure 3: Overview of generated variables at the individual level in wave 4**

| | PENDDAT | | | | | BIO-Spells | EE_Spells |
|---|---|---|---|---|---|---|---|
| | Aktueller Status | Erwerbshistorie | | Soziale Herkunft | | Erwerbs- und Arbeitslosigkeits-biografie | Ein-Euro-Job-Teilnahme |
| | | letzte ET | erste ET | Mutter | Vater | | |
| **Bildung** | berabj | | | | | | |
| | beruf1 | | | mberuf1 | vberuf1 | | |
| | beruf2 | | | mberuf2 | vberuf2 | | |
| | schulabj | | | | | | |
| | schul1 | | | mschul1 | vschul1 | | |
| | schul2 | | | mschul2 | vschul2 | | |
| **Bildungsklasifikation** | casmin | | | mcasmin | vcasmin | | |
| | isced97 | | | misced97 | visced97 | | |
| | bilzeit | | | mbilzeit | vbilzeit | | |
| **Informationen zum aktuellen Status** | akt1euro | | | | | | |
| | alakt | | | | | | |
| | etakt | | | | | | |
| | statakt | | | | | spelltyp | |
| **Sozio-ökonomische Position** | egp | egplewt | egpeewt | megp | vegp | egp | |
| | esec | eseclewt | eseceewt | mesec | vesec | esec | |
| | isei | iseilewt | iseieewt | misei | visei | isei | |
| | mps | mpslewt | mpseewt | mmps | vmps | mps | |
| | siops | siopslewt | siopseewt | msiops | vsiops | siops | |
| **Stellung im Beruf** | stib | stiblewt | stibeewt | mstib | vstib | stib | |
| | stibkz | | | | | | |
| **Datierung der Erwerbstätigkeit** | | | begmeewt | | | bmonat | bmonat |
| | | | begjeewt | | | bjahr | bjahr |
| | | emonlewt | | | | emonat | emonat |
| | | ejhrlewt | | | | ejahr | ejahr |
| **Datierung der Arbeitslosigkeit** | | | | | | alg1bm | |
| | | | | | | alg1bj | |
| | | | | | | alg1em | |
| | | | | | | alg1ej | |
| **Informationen zur Erwerbstätigkeit** | befrist | | | | | | |
| | azhpt1 | | | | | az1 | |
| | azhpt2 | | | | | az2 | |
| | azges1 | | | | | | |
| | azges2 | | | | | | |
| **Berufliche Tätigkeit** | isco88 | iscolewt | iscoeewt | misco | visco | isco88 | |
| | kldb | kldblewt | kldbeewt | mkldb | vkldb | kldb | |
| **Tätig in Branche** | branche | | | | | branche | |

**Figure 3: Overview of generated variables at the individual level in wave 4 (continued)**

| | PENDDAT | | | | | BIO-Spells | EE_Spells |
|---|---|---|---|---|---|---|---|
| | Aktueller Status | Erwerbshistorie | | Soziale Herkunft | | Erwerbs- und Arbeitslosigkeits-biografie | Ein-Euro-Job-Teilnahme |
| | | letzte ET | erste ET | Mutter | Vater | | |
| **Einkommen** | netges | | | | | | |
| | brges | | | | | | |
| **Leistungsbezug** | alg1abez | | | | | alg1akt | |
| | hhalg2 | | | | | | |
| **Haushaltskontext und Familienstand** | hhgr | | | | | | |
| | famstand | | | | | | |
| | vhh | | | | | | |
| | mhh | | | | | | |
| | apartner | | | | | | |
| | epartner | | | | | | |
| | ekind | | | | | | |
| | ekin614 | | | | | | |
| | ekinu15 | | | | | | |
| | ekinu18 | | | | | | |
| | ekin1517 | | | | | | |
| | kindzges | | | | | | |
| | kindzihh | | | | | | |
| **Migrationshintergrund** | ogebland | | | | | | |
| | ostaatan | | | | | | |
| | ozulanda | | | | | | |
| | ozulandb | | | | | | |
| | ozulandc | | | | | | |
| | ozulandd | | | | | | |
| | ozulande | | | | | | |
| | ozulandf | | | | | | |
| | migration | | | | | | |
| **Informationen zur Person** | gebhalbj | | | | | | |
| | palter | | | | | | |
| | zpalthh | | | | | | |
| | zpsex | | | | | | |
| **Allgemein** | altbefr | | | | | | |
| | fb_vers | | | | | | |
| | panel | | | | | | |
| | pintdat | | | | | | |
| | RegP0100 | | | | | | |
| | sample | | | | | | |

## 4.5 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

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

## Education in years, mother

| | |
|---|---|
| Variable name | *mbilzeit* |
| Variable label | Duration of school education and vocational training in years, generated |
| Source variables | *mschul2; mberuf2* |
| Category / dataset | Education / individual-level data |
| Prepared by | Bernhard Christoph |
| Explanation | General description: see "Education in years"<br>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:<br>Training as a semi-skilled worker: +1 year<br>Apprenticeship, vocational school, school for health care occupations: +1.5 years<br>Master craftsman's certificate +3 years<br>College of advanced vocational studies: +3 years<br>University of Applied Sciences: +3 years<br>University: +5 years<br>Other German qualification: +1.5 years<br>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"<br>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:<br>Training as a semi-skilled worker: +1 year<br>Apprenticeship, vocational school, school for health care occupations: +1.5 years<br>Master craftsman's certificate +3 years<br>College of advanced vocational studies: +3 years<br>University of Applied Sciences: +3 years<br>University: +5 years<br>Other German qualification: +1.5 years<br>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).<br>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. |

| school / occup. | 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

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

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

| school occup. | not surv. | 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

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

| school occup. | not surv. | 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)

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

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

## Classification of Occupations 1992 (KldB92)

| | | | |
|---|---|---|---|
| Generated | Employment | Variable name | Source variables |
| | current | *kldb_it* | ET2500 |
| | Spell data (bio_*spells*) | *kldb* | ET2500 |
| | first | *kldbeewt* | ET2500, PET1280, PET3950 |
| | last | *kldblewt* | ET2500, PET1280 |
| | of father | *vkldb* | PSH0800 |
| | of mother | *mkldb* | PSH0700 |
| Variable label | Current empl.: Classification of Occupations 1992, current employment | | |
| | Spell data (*et_spells*): Classification of Occupations 1992, generated | | |
| | First empl.: Classification of Occup. 1992, first empl., gen. | | |
| | Last empl.: Classification of Occupations 1992, last empl., gen. | | |
| | Father: Classification of Occupations 1992 of father, generated | | |
| | Mother: Classification of Occupations 1992 of mother, generated | | |
| Category / dataset | Occupation / individual-level data | | |
| Contact person | Bernhard Christoph | | |
| Explanation | The KldB92 is the current version of the Classification of Occupations published by the German Federal Statistical Office. 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. | | |
| Literature: | StBA (1992) | | |

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

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

## European Socio-economic Classification (ESeC)

| Generated | Employment | Variable name | Source variables |
|---|---|---|---|
| | current | *esec* | *isco88, stib, PET2000, PET2700* |
| | Spell data (bio_*spells*) | *esec* | *isco88, stib, ET1100, ET1101, ET1102, ET1300, ET1301, ET1302,* |
| | first | *eseceewt* | *iscoeewt, stibeewt, PET1261* |
| | last | *eseclewt* | *iscolewt, stiblewt, PET3801* |
| | of father | *vesec* | *visco, vstib, PSH0670* |
| | of mother | *mesec* | *misco, mstib, PSH0370* |
| Variable label | Current empl.: European Socio-economic Classification (ESeC), current occupation, gen. | | |
| | Spell data (bio_*spells*): European Socio-economic Classification (ESeC), gen. | | |
| | First empl.: European Socio-economic Classification (ESeC), first employment, gen. | | |
| | Last empl.: European Socio-economic Classification (ESeC), last employment, gen. | | |
| | Father: European Socio-economic Classification (ESeC), occupation of father, gen. | | |
| | Mother: European Socio-economic Classification (ESeC), occupation of mother, gen. | | |
| Category / dataset | socio-economic position / individual-level data | | |
| Prepared by | Bernhard Christoph | | |
| Explanation | With regard to its theoretical conception, the European Socio-economic Classification is largely based on the EGP class scheme. In contrast to the latter, however, great importance was attached to international comparability of operationalisation procedures and comprehensive validation of the classification scheme (for a general description, see: Rose & Harrison 2007, and Müller et al. 2006, 2007 for Germany). | | |
| | The Stata do-file required to generate the ESeC was kindly provided by Heike Wirth from GESIS-ZUMA (Fischer & Wirth 2007). We simply adjusted it to the requirements of this study. This do-file, originally written in standard SPSS syntax by Harrison & 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 | Employment | Variable name | Source variables |
| | current | *mps* | *isco88* |
| | Spell data (bio_*spells*) | *mps* | *isco88* |
| | first | *mpseewt* | *iscoeewt* |
| | last | *mpslewt* | *iscolewt* |
| | of father | *vmps* | *visco* |
| | of mother | *mmps* | *misco* |
| Variable label | Current empl.: Magnitude Prestige Scale, current occupation, gen. | | |
| | Spell data (bio_*spells*): Magnitude Prestige Scale, generated | | |
| | First empl.: Magnitude Prestige Scale, first employment, gen. | | |
| | Last empl.: Magnitude Prestige Scale, last employment, gen. | | |
| | Father: Magnitude Prestige Scale, occupation of father, gen. | | |
| | Mother: Magnitude Prestige Scale, occupation of mother, gen. | | |
| Category / dataset | socio-economic position / individual-level data | | |
| Contact person | Bernhard Christoph | | |
| Explanation | The Magnitude Prestige Scale [MPS] (Wegener 1985, 1988) is the only specifically German instrument available so far to operationalise social prestige based on detailed occupation information. It was originally developed for the older 1968 version of the International Standard Classification of Occupations (ISCO-68). Since occupation coding in the study at hand was conducted based on the more recent ISCO-88 classification and the Classification of Occupations (KldB) developed by the Federal Statistical Office, a variant of the scale transferred to ISCO-88 was used (Christoph 2005). 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 | Employment | Variable name | Source variables |
| | current | *siops* | *isco88* |
| | Spell data (bio_*spells*) | *siops* | *isco88* |
| | first | *siopseewt* | *iscoeewt* |
| | last | *siopslewt* | *iscolewt* |
| | of father | *vsiops* | *visco* |
| | of mother | *msiops* | *misco* |
| Variable label | Current empl.: Standard International Occupational Prestige Scale, current occupation, gen. | | |
| | Spell data (bio_*spells*): Standard International Occupational Prestige Scale, generated | | |
| | First empl.: Standard International Occupational Prestige Scale, first employment, gen. | | |
| | Last empl.: Standard International Occupational Prestige Scale, last employment, gen. | | |
| | Father: Standard International Occupational Prestige Scale, occupation of father, gen. | | |
| | Mother: Standard International Occupational Prestige Scale, occupation of mother, gen. | | |
| Category / dataset | socio-economic position / individual-level data | | |
| Contact person | Bernhard Christoph | | |
| Explanation | The Treiman Prestige Scale, which was originally constructed by Treiman (1977) for 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 | Employment | Variable name | Source variables |
| | current | *isei* | *isco88* |
| | Spell data (bio_*spells*) | *isei* | *isco88* |
| | first | *iseieewt* | *iscoeewt* |
| | last | *iseilewt* | *iscolewt* |
| | of father | *visei* | *visco* |
| | of mother | *misei* | *misco* |
| Variable label | Current empl.: International Socio-Economic Index, current employment, gen. | | |
| | Spell data (bio_*spells*): International Socio-Economic Index, generated | | |
| | First empl.: International Socio-Economic Index, first employment, gen. | | |
| | Last empl.: International Socio-Economic Index, last employment, gen. | | |
| | Father: International Socio-Economic Index, occupation of father, gen. | | |
| | Mother: International Socio-Economic Index, occupation of mother, gen. | | |
| Category / dataset | socio-economic position / individual-level data | | |
| Contact person | Bernhard Christoph | | |
| Explanation | The International Socio-Economic Index is certainly one of the most common indices of its kind. This is due not least to the fact that, in contrast to most other SEIs, the ISEI is based on an original theoretical concept which sees the occupation and its socio-economic status as an "intervening variable" between education and income.<br>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).<br>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 | Employment | Variable name | Source variables |
| | current | *branche* | ET2600 |
| | Spell data (bio_*spells*) | *branche* | ET2600 |
| Variable label | Current empl.: Current activity: economic sector/industry (WZ2003) | | |
| | Spell data (bio_*spells*): economic sector/industry (WZ2003), generated | | |
| Category / dataset | socio-economic position / individual-level data | | |
| Contact person | Bernhard Christoph | | |
| Explanation | The information from the open-ended survey question about the sector / industry in which the respondent works was coded based on the 2-digit code in the Classification of Economic Activities of the Federal Statistical Office (WZ2003). At the two-digit level, this classification largely corresponds to the European "Nomenclature générale des Activités économiques dans les Communautés Européennes (NACE)" in revision 1.1. | | |
| Literature: | StaBA (2002); EG (2002) | | |

### 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-individual households as compared to single households. To do this, the per-capita income in multi-individual households is not calculated based on the actual number of individuals living in the household, but by using a divisor which is usually below this figure and is calculated based on the assumed needs of the household members (equivalised household size).<br>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.<br>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".<br>The modified OECD equivalence scale assumes a weighting factor of 1.0 only for the first household member (aged 15 or over). Further household members aged 15 or over are assigned a weighting factor of 0,5; children up to the age of 14 are counted with a weighting factor of 0.3 to calculate the equivalised household size.<br>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.<br>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).<br>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).<br>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.<br>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 ("listwise 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

| | |
|---|---|
| Variable name | *depindg* |
| Variable label | Deprivation index, weighted (items missing for financial reasons; total of weighted items: 12,8) |
| Source variables | *HLS0100a-HLS2600a; HLS0100b-HLS2600b; PLS0100-PLS2600* |
| Category / dataset | material situation / household-level data (weighted at the individual level) |
| Prepared by | Bernhard Christoph |
| Explanation | For a general description, see deprivation index, unweighted.<br>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).<br>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.<br>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. p. 28 pp.). |
| Literature: | Andreß & Lipsmeier (1995, 2001); Andreß et al. (1996); Halleröd (1995); Lipsmeier (1999); Nolan & Whelan (1996) |

## Household typology

| | |
|---|---|
| Variable name | *hhtyp* |
| Variable label | Household type, generated |
| Source variables | Household information on age and relationships between household members |
| Category / dataset | Household structure / household data |
| Prepared by | Daniel Gebhardt |
| Explanation | A number of variants and suggestions exist regarding the definition of household types (see e.g. Lengerer 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 generation relationships. Whereas the SOEP typology is merely based on the relationship of the household members to the head of the household, PASS uses information on interrelationships between all household members for the generation. In addition, the PASS typology includes the age of the household members as indicated in the household interview and the household size. |

**Definition of relationships for generating the household type:**

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

**Definition of household types:**

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

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

## Benefit community ID, wave 4

| | |
|---|---|
| Variable name | *bgnr4* |
| Variable label | Benefit community ID in wave 4 |
| Source variables | Household information on age and relationships between household members |
| Category / dataset | Benefit community / person register |
| Prepared by | Gerrit Müller |
| Explanation | The *bgnr4* variable is created at the individual level. It assigns an identification number to each household member indicating the individual's affiliation to a particular benefit community. Consequently, household members with the same ID constitute a benefit community together. The *bgnr4* variable is composed of the known household number and a two-digit indicator to identify the benefit community within the household.<br>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 individuals in the same household who are or would be regarded as benefit 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.<br>With regard to content, the allocation of an individual 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 individual is living in a partnership and/or has a child / children aged under 25 who has/have no own partner/children. In the latter case, the benefit community comprises the individual, his/her partner and the child(ren). If two individuals live in the same household with a mutual child, but do not indicate in the household grid table that they are living in a partnership, a partnership is nevertheless assumed to exist in terms of Section 7, Sub-section (3a), and the corresponding individuals and their child(ren) are assigned to the same benefit community. Individuals aged between 15 and 25 are generally assigned to their parents unless they are already living together with a partner (or a child of their own) in a joint household. Individuals aged between 15 and 25 who live without their parents (or partner / children) constitute a separate benefit community.<br>Individuals 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. |
| Literature: | German Social Code Book II – basic benefits for job-seekers (Sozialgesetzbuch, Zweites Buch - Grundsicherung für Arbeitssuchende (SGB II)) |

## Benefit community typology, wave 4

| | |
|---|---|
| Variable name | *bgtyp4* |
| Variable label | Type of benefit community in wave 4 |
| 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 *bgnr4*. 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.<br>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 bgnr4). Code -5, generation impossible (missing values), was allocated to households with missing information on relationships and/or the age of individual household members (see code 99 for bgnr4). |
| Literature: | – |

## Benefit community in receipt of unemployment benefit II on the sampling date, wave 4

| | |
|---|---|
| Variable name | *bgbezs4* |
| Variable label | Benefit community in receipt of UB II on the sampling date in wave 4 (2010) |
| Source variables | *HA0250*, HA0300, AL20100, AL20200, AL20300, AL20400, AL20603, AL20703*, HA0400, sample, hnr, bgnr4, 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 *bgnr4* this variable indicates whether the benefit community was in fact receiving Unemployment Benefit II on the sampling date of wave 4 or not. |
| Literature: | – |

## Benefit community in receipt of unemployment benefit II on the survey date, wave 4

| | |
|---|---|
| Variable name | *bgbezb4* |
| Variable label | Benefit community in receipt of UB II on the survey date in wave 4 (2010) |
| Source variables | *AL20603, AL20703, zensiert* (*alg2_spells*), *sample, hhgr, bgnr4* |
| 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 *bgnr4* this variable indicates whether the benefit community was in fact receiving Unemployment Benefit II on the survey date of wave 4 or not. |
| Literature: | – |

Number of benefit communities within the household

| | |
|---|---|
| Variable name | *anzbg* |
| Variable label | Number of synthetic benefit communities in the HH, generated |
| Source variables | *bgnr4, 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 *bgnr4*. |
| Literature: | – |

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

| | |
|---|---|
| Variable name | *nbgbezug* |
| Variable label | Number of benefit communities in the HH receiving benefits on the sampling date |
| Source variables | *bgbezs4, bgnr4, 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 on the sampling date. The value was calculated via the household number by aggregating the benefit communities within each household which were actually receiving benefits according to the variable *bgbezs4* from the person register. |
| Literature: | – |

# 5 Data preparation

Since wave 3, not the IAB but infas has been responsible for preparing the data. In order to 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 development of the *bio_spells* dataset, which was first developed in wave 4, were made together with the IAB. The IAB was also available for questions beyond that during the period of data preparation.

The information gathered in the interviews of the fourth wave is initially available at infas in the form of ASCII data. In a first step, infas prepared the following datasets from these raw data[38]:

- Household dataset for questions surveyed in the cross-section
- Household dataset for data surveyed in the longitudinal section (module "Unemployment Benefit II")
- Dataset on the update of the household composition (matrix)
- Dataset on the update of the family relationships in the household (relationship matrix)
- Personal/senior citizens' dataset for questions surveyed in the cross-section
- Individual dataset for data surveyed in the longitudinal section I (module "employment biography [spells]")
- Individual dataset for data surveyed in the longitudinal section II (module "measures")
- Dataset for open texts (across all household, personal and senior citizens' interviews)

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

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

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

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

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

| No. | Step of the procedure |
|---|---|
| 1 | Import of the surveyed raw data in working datasets |
| 2 | Check of the household structure (see Chapter 5.1) |
| 3 | Removal of problematic interviews (household and/or individual level) (see Chapter 5.1 ) |
| 4 | Integration of individual dataset and senior citizens' dataset |
| 5 | Correction of the household structure of re-interviewed households (see Chapter 5.1) |
| 6 | Filter checks at the household level (see Chapter 5.2) |
| 7 | Construction of a household grid dataset and plausibility checks on this (see Chapter 5.3) |
| 8 | Generation of the synthetic benefit 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 datasets, individual datasets, register datasets) |
| 17 | Anonymisation (see Chapter 5.5) |

## 5.1 Structure checks and interviews removed from the dataset

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

Different checks were carried out to identify problematic cases. The cases concerned were discussed in a formalised procedure between infas and the IAB. The final decision on how to proceed with these cases was made by the IAB. It should be considered that the following specifies the extent of the checks conducted. Not every check in every wave leads to the identification of problems. The result of a check is usually that a checked issue occurs in a

low case number or not at all. Furthermore, known error sources are absorbed already during the interview. The survey instrument thus intends that not all known target persons can move out of a panel household at the same time and that among the individuals remaining after the moves at least one must be over 15 years of age.

- By comparing the first names reported in the current and the previous wave, cases were identified in which changes in the household composition had not been recorded correctly. Instead of including moves into and out of the household in the relevant places in the household interview, it sometimes happened that interviewers renamed household members or changed their age or sex. All cases where a first name had been changed and this could not be put down to a correction of spelling and where the year of birth reported in the previous wave differed by more than one year from that reported in the current wave were subjected to individual case reviews. Here a decision was made as to whether the change in the data was simply a matter of correcting the first name, age or sex, or whether the interviewer had made an inadmissible change to the household structure.
- Furthermore, it was checked whether more than one individual with the same date of birth was living in the household. In the household context of the two waves, it was decided whether these were plausible or implausible cases. The remaining cases then underwent another check. For this, households were identified in which a date of birth was reported in the current and previous wave by individuals in different positions in the household structure. Here it seemed reasonable to suspect that a different individual from that in the previous wave conducted the particular personal interview in the current wave. In the context of the household and individual-level data of the current and previous wave, individual case decisions were made regarding the respective household and personal interviews.
- In general, the date of birth from the personal/senior citizens' interview of the current wave displaces all other age information on these individuals, e.g. from the household grid, and is the basis for all generations which are among others based on age. In a special constellation, the date of birth is, however, corrected in *PD0100*. If the year of birth of an individual changes significantly according to *PD0100*, the day and month, however, stay the same, the hitherto known date of birth has never changed according to *PD0100* and at least two pieces of information on the date of birth from *PD0100* are available from previous waves, then the year of birth is reset to the value known from the previous waves considering the whole household constellation. A theoretical example is an individual whose date of birth is known as 01 February 1972 from at least two previous waves and whose date of birth is now recorded as 01 February 1992, which would make this individual younger than the children living in the household. Without a correction, such a constellation would lead to an implausible relationship structure, which would consequently also lead to the fact that, for example, the synthetic benefit communities cannot be generated. Hence, the information from the example is being corrected to the value 01 February 1972 in the current wave.
- In order to identify households which are regarded as not successfully surveyed in the sense of PASS, the datasets at the household and the individual level were merged. Per-

sonal interviews without a full household interview were marked, as were household interviews for which no interview at the individual level was available[39].

- Also moves into and out of the household are another important factor. Panel households with reported move-outs of the household were generally inspected regarding their household context and correlated with the realised split-off households. Evaluations were made as to whether the remaining household context of the panel household is self-evidently plausible. Interviews from panel households in which all household members leave the household, except for individual children under 15 years of age, were discarded with regard to the panel household as well as with regard to split-off households. If more than one individual moved out, it was checked whether these individuals formed a joint split-off household or several different ones, and whether this is plausible. Such cases were considered implausible, for instance, where one partner left the panel household together with young children, but the individuals moving out formed several different split-off households according to field information, i.e. the young children allegedly forming individual households. In case of the non-realisation of the split-off household, the move-outs were considered as plausible, but all individuals that moved out were retroactively merged into one joint split-off household.
- Individual cases occurred in which, according to the interview in the panel household, individual persons formed a split-off household, however, all members of the panel household could be found in the split-off household. In an alternative situation not all members of the panel household live in the split-off household, but at least one member of the panel household who, in the interview there, was not reported as having moved out or having moved to another split-off household than the one observed. Here, too, differentiated decisions were made as to which reported 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 split-off households it is verified whether individuals who are not known from the panel household but join PASS through the split-off household might still originate from the panel household. Two constellations promote these cases. On the one hand, it occurs that a panel household reports in case of several individuals moving out that the split-off individuals formed more than one split-off household. In this case, a dynamical preload is created for the current filed for all the split-off households known through the panel household. If, however, individuals who live in various split-off households according to the panel household, are actually found in a shared split-off household, these individuals who are not assigned to this split-off household by the panel household but to another split-off household do not have a preload and are included as new individuals.
- On the other hand, it is possible that individuals move out of or into a household which was formed as split-off household in a previous wave and was already successfully surveyed back then. Thus, there is another move from the original panel household into this

[39] 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 first wave. In contrast, the household interviews of re-interviewed households and split-off households were retained.

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

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

- In panel households that reported a move-out as of wave 2, there can also be moves back in of members formerly belonging to the household as of the wave 3. The requirement of recognising these individuals as moving back in and 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 who were subsequently identified in the current wave as moving back in based on a comparison of first name, age and sex with the members who previously moved out of the households, the household structure had to be changed. This led to retroactive changes of the personal identification number of the individual to be positioned and also an adjustment in the individual-related information in the household interview, e.g. on childcare or the reasons for a cut in Unemployment Benefit II to the position defined as correct within the framework of the structural check. Conversely, it is also checked whether an individual who is marked in the field as moving back in really is the same individual who moved out in a previous wave. If not, this is a move-in of an individual who is new to PASS. The described changes in the household structure are also made in this case.
- In case of moves back in it is checked whether the split-off household in which the individual lived before he/she moved back into the panel household was successfully surveyed in the current wave and whether the split-off household considers the individual moving back in as having moved out. Also individuals who moved back into their panel household in a previous wave must continue to be checked regarding their status in the split-off household as long as the split-off household is part of the current panel sample. If an individual who moves back is still considered a current household member in his/her split-off household, a decision was made during data preparation for these cases whether this was plausible or whether the household structure of the panel or split-off household had to be corrected.

- Not only moves back can lead to individuals being considered as current household member of several households. It can also occur that an individual is considered a member of a split-off household although he/she was not recorded as having moved out of the panel household. Individual cases of this can be acknowledged as plausible after examination of the household structure of the respective households. Such a case resulted in the introduction of the variable *zdub4* in the person register. For further explanations, please refer to Chapter 4.4 and Chapter 5.4.1.4.
- There can be other issues regarding the relationship of a panel household and its split-off households. There is a possibility that individuals who joined PASS via a split-off household move to the panel household. Another possibility is that individuals move from one split-off household to another split-off household. Generally, all individuals in a panel household and all split-off households connected to it must be considered a network. The structure checks are designed in such a way that individual moves between the households of such a network are detected regardless of the direction in which an individual moves in the network.
- Household structure checks generally do not evaluate the structure of the household in terms of plausibility but they consider the changes between the waves. Therefore, the household structure of households interviewed for the first time can only be checked to a limited extent. For households interviewed for the first time a check is made based on information concerning first name, age and sex 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 at position 2 and 3 are identical, not only individual 3 is removed but also individual 4 is retroactively moved to position 3. As a rule, in a household interviewed for the first time with X household members, the positions 1 to X are to be filled without gaps. Just like for someone retroactively recognised as moving back in, a subsequent change in the personal identification number of the individual to be moved also requires moving the individual-related information in the household interview.
- Thanks to feedback by a field interviewer, a household was detected which was included twice in the panel sample. Household 10015439 has been in the sample as identical household 15044862 since W1. Both households were successfully surveyed in wave 1 and wave 3 and not surveyed in wave 2. In wave 4, household 10015439 was successfully surveyed. This duplicate was detected since "both" households were given to the CAPI interviewer of this point. The household composition in the two households remained the same across all waves. Household 15044862, which was not surveyed in wave 4, will be deleted from the sample for wave 5. There will be no retroactive removal of the duplicate from waves 1 to 3 since this would affect weighting. The duplicate household is marked with code 26 in the variable hnettod4 in hh_register which makes the reason for non-surveying transparent. All household members of the duplicate household are marked with code 56 in the variable *pnettod4* in *p_register*.

- Individual case decisions were also made to deal with the cases which proved to be problematic during the structure checks. What was of significance here was how serious the particular problem was considered to be. In cases where the correct household composition in the fourth wave was unclear, all of the interviews from the fourth wave were removed. In the fifth wave these households will be treated as households that did not participate in the fourth 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 individual in the third wave, then only the respective personal or senior citizens' interview was removed. If it was a structural problem that had no serious consequences and could be solved, for example, by removing a personal interview, additional corrections of the first name, age and sex were made at the household level. The incorrect information concerned was then set back to the last valid value from the previous wave or in the case of age to the value from the previous wave + the number of years since the last valid interview in this household.

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

The *Netto* variables (*hnettok4*, *hnettod4*, *pnettok4*, *pnettod4*) in the household and person register datasets indicate removed interviews. Via the corresponding variables in the household register it is possible to trace the re-interviewed households whose household interviews were removed later. By means of 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 fourth 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 respective 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.[40] 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 an individual is retroactively discarded as implausible and the individual is retroactively classified as still belonging to the former household then this also means that individual-related information on these individuals in the household interview must be coded retroactively as mistakenly not surveyed. Thus, the code "-4" does not always refer to a problem in the survey instrument. If the code "-4" is assigned to a question that is relevant for filtering subsequent questions, then the subsequent questions are also coded with "-4" in case these subsequent questions were actually not surveyed. If subsequent questions were, however, surveyed, because, for instance several filter questions 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. Table 22 provides an overview of the assigned values. "-1" and "-2" are the standard recoding for the values "don't know" and "details refused" recorded during the survey. "-3" is the general "not applicable" code for questions not asked due to filters. As described above, the code "-4" was assigned if a question was not asked as a result of a filter error. Codes "-5" to "-7" are question-specific codes. These can 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.

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

**Table 22: Overview of the missing codes used**

| Code | Explanation |
|---|---|
| -1 | "don't know" |
| -2 | "details refused" |
| -3 | "not applicable (filter)" (question not asked due to filter) |
| -4 | "question mistakenly not asked" (question should, however, have been asked) |
| -5 | question-specific code no. 1, only assigned as required |
| -6 | question-specific code no. 2, only assigned as required |
| -7 | question-specific code no. 3, only assigned as required |
| -8 | "implausible value" |
| -9 | "item not surveyed in wave" |
| -10 | "item not surveyed in questionnaire version"[41] |

The value "-8" is a specific missing code assigned during the plausibility checks (see Chapter 5.3 on plausibility checks). The missing code "-9" became necessary for the first time in 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 have not been surveyed in any version of the questionnaire as of the second wave are given the value "-9" for the observations in this wave. The same is done for observations from the first wave. Variables that were surveyed for the first time after the first wave are retroactively coded "-9" for observations of waves in which they were not surveyed. Code "-10" can be used to consider differences between the questionnaire versions, in other words between the personal questionnaire and the senior citizens' questionnaire or between the two versions of the household questionnaire until wave 3.

## 5.3 Plausibility checks

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

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

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

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

2. Plausibility check of the household structure: This check was carried out based on the information collected in the household interview on the family relationships between the household members, and the information on age, sex and first 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. An example: In the case of two people of the same sex who were both natural parents of a third member of the household, the sex was corrected based on the first name. If the first names also indicated that the two people were of the same sex, and if there was no other relevant

information available, then the relationship was marked as implausible based on the household structure.

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

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

## 5.4 Retroactive changes of waves 1 to 3

### 5.4.1 Conceptional revisions

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

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

- W1: panel concept, i.e. only survey of latest available data
- W2/3: modular survey of ET/AL spells[42] + filling of gaps of > 3 months and of latest available data
- W4: integrated survey of ET/AL/LU spells

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

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

#### 5.4.1.1 Variables on current participation in measures in PENDDAT

**akt1euro** (*Current part. in one-euro job, generated*)

W1[43] to W3 surveyed periods at the individual level during which the target person participated in labour market policy measures. The interest in results was similar despite the revision of the module after W1: Measures of different types, ranging from one-day training measures to longer-term measures, such as job creation measures or one-euro jobs, were to be reported. The goal was to have the clearest possible picture of the measure history of the TP. This is why a very comprehensive range of measures was surveyed.

Based on the spell data surveyed in that way (and the spell datasets *massnahmespells* and *mn_spells* generated from that) the variable *aktmassn* was generated for the individual dataset. This variable indicates whether the TP was in a currently ongoing measure at the interview date of the respective wave.

As of W4, the result interest is limited to one-euro jobs in which the TP participated in a certain period. The variable *aktmassn* which was previously generated for the individual dataset cannot be continued due to the reduced range of surveyed measures (as of W4: code -9).

Instead, a new variable is created which indicates whether the TP currently participated in a one-euro job on the survey date. This variable is generated for all waves - respectively based on the spell dataset relevant for the wave, i.e. based on *massnahmespells* for W1, *mn_spells* for W2/W3 and *ee_spells* for W4.

Categories of the variable to be generated:

| | |
|---|---|
| -10 | Item not relevant for questionnaire version |
| -5 | No generation poss. (missing values) |
| -3 | Not applicable (filter) |
| 1 | Currently participates in one-euro job |
| 2 | Currently does not participate in one-euro job |

[43] Here and in the following: W1 = wave 1, W2 = wave 2 etc.

**Table 23: Generation bases of akt1euro**

| Wave | Dataset | Determination that measure is one-euro job | Determination that one-euro job continues |
|---|---|---|---|
| 1 | massnahmespells | massntyp=1 | zensiert=1 |
| 2 | mn_spells | MN0100 | spwelle=2 & zensiert=1 |
| 3 | mn_spells | MN0100 | spwelle=3 & zensiert=1 |
| 4 | ee_spells | - | zensiert=1 |

#### 5.4.1.2 Variables on the employment / economic inactivity status in PENDDAT

The concept to survey employments has been revised several times across the waves:

- W1: panel concept, i.e. only survey of latest available data
- W2/3: modular survey of ET/AL spells + filling of gaps of > 3 months and of latest available data
- W4: integrated survey of ET/AL/LU spells

The information available for the individual waves varies regarding the following due to changes in the survey concept:

- form of the information available (panel vs. spells)
- degree of details of the available information (main status vs. parallel conditions)
- consistency of the available parallels (gap filling vs. complete survey of parallel conditions)

The previous concept of generated variables on the employment / economic inactivity status is oriented heavily on the survey logic of wave 1. This logic is described in the following (simplified):

- Is there ET of at least 1/h per week?
- If ET: one or more ETs?
- If ET (information for main ET): successive determination whether ET is marginal employment, one-euro job or similar, or part of vocational training
- If no ET (or main ET= marginal employment): determination of the status of economic inactivity (AL or other status)

The concept of the generated variables (*erwerb*, *erwerb2*, *nichterw*, *nichtew2*) mainly follows this survey logic from W1. While, due to the question logic, no competing statuses were possible in W1 (TPs with ET which is no marginal employment, were not asked for other activities), the necessity for decision arose in W2 if there was more than one ongoing spell. The following logic was used for the generation of variables on the employment / economic inactivity status in W1 and W2:

**Table 24: Generation logic *erwerb*, *erwerb2*, *nichterw*, *nichtew2***

| Variable | Generation logic wave 1 | Generation logic wave 2 |
|---|---|---|
| erwerb | *(1) Differentiation status main ET*<br>- No main ET<br>- Main ET: not vocational training/job creation measure/marginal employment<br>- Main ET: part of vocational training<br>- Main ET: job creation measure etc.<br>- Main ET: marginal employment<br><br>*(2) Differentiation status main ET is basis for further generation*<br>- Main ET: not vocational training/job creation measure/marginal employment→ ET as employment status<br>(exceptions:<br>trainees (from PB0100) with working hours < 21 -→ trainees;<br>pupils (from PB0100) with working hours >0 & working hours < 24 → pupils;<br>students (from PB0100) with working hours >0 & working hours < 21 → students;<br>employees with working hours > 0 & working hours < 16 → other)<br><br>- no main ET or main ET: marginal employment → use employment status from PET0801 (i.e. using status of economic inactivity)<br><br>- no main ET + pupils/students according to PB0100→ use employment status from PB0100<br><br>- Main ET: job creation measure etc. → transfer as employment status (job creation measure, one-euro job etc.)<br><br>*(3) Deciding on contradictory information*<br>- erwerb: job creation measure etc. + PB0100: pupil/student/trainee→ -8<br>- erwerb: pupil + PB0100: student → -8<br>- erwerb: retired individual + PB0100: trainee→ -8<br>- erwerb: pupil + PB0100: trainee → transfer of status from PB0100<br>- erwerb: other + PB0100: pupil/student/trainee → transfer of employment status from PB0100 | not generated (-9) |

**Table 24: Generation logic *erwerb*, *erwerb2*, *nichterw*, *nichtew2* (continued)**

| Variable | Generation logic wave 1 | Generation logic wave 2 |
|---|---|---|
| erwerb2 | *(1) Recode of erwerb*<br><br>- Merging of categories:<br><br>- Unemployed + job creation measure/one-euro job/similar → unemployed<br>- Vocational training/training/further education/<br>retraining + student → vocational training/training/studies | *(1) Recode of nichtew2*<br><br>*(2) Integrating ETs*<br>- Replacing values if current ET (>EUR 400 from ET spells) is available<br><br>*(3) Making adjustments*<br>- erwerb2: ET + PB0100: Student + working hours <= 20h → student<br>- erwerb2: AL + PB0100: Student → student<br>- erwerb2: pupil + PB0100: Student → status not clear |
| nichterw | *(1) Recode of PET0800* | *(1) Recode of LU0100 (gap status without answers to open-ended questions) + current AL from AL spells*<br><br>- Merging of categories:<br><br>- Reported as unemployed + not reported → unemployed<br>- Vocational training/training/studies + other → other<br><br>- Determination MV from PET0151/PET0911 + indicator for cases which were mistakenly not filtered into the gap module |
| nichtew2 | *(1) Recode of PET0801*<br><br>- Merging of categories:<br><br>- Unemployed + job creation measure/one-euro job/similar → unemployed<br>- Vocational training/training/further education/<br>retraining + student → vocational training/training/studies | *(1) Recode of LU0101 (gap status with answers to open-ended questions)*<br>- Merging of categories:<br><br>- Reported as unemployed + not reported → unemployed<br>- Something else/main status not clear → other/main status unclear<br><br>*(2) Considering pupils/students/trainees from PB0100*<br>- If no valid status available until now → Transfer of information from PB0100 |

The generated variables thus continue the logic of the survey concept from W1 which can also be found as basic logic in the generated variable:

→ ET generally overrides all other conditions (excluding exceptions); AL overrides all conditions except for ET (excluding exceptions)

Another logic (e.g. AL overrides ET) could not have been implemented in W1 since the survey logic prioritised the ET of the TP and other conditions were only surveyed as replacement. The process for variable generation thus follows the survey process.

However, this process is not really useful for the determination of an individual's main status and furthermore ignores basic considerations which can be found, for example, in the definition of unemployment (Section 16, 119 SGB III; also applies to SGB II in accordance with Section 53a SGB II).

Unemployment requires the following (according to the definition in SGB III):

- economic inactivity (i.e. no paid employment or employment only to a limit of 15h/week; variations are possible) (Section 119 SGB III)
- availability (i.e. is available for placement efforts by the BA; search for employment and readiness to be employed >= 15h/week; can comply with integration proposals timely; readiness to participate in professional integration measures) (Section 119 SGB III)
- own efforts (i.e. efforts to end economic inactivity) (Section 119 SGB III)
- report (i.e. personal report as unemployed with the BA) (Section 16 SGB III)
- no current participation in measures (Section 16 SGB III)

The previous logic in which AL was overridden by ET regardless of the amount of hours is thus more driven by the survey logic of W1 than by a content-wise consideration of what should actually be considered as main status.

Further criticism on the employment / economic inactivity status variables concerns the basic aim of these variables. What are they supposed to indicate? An individual's main status? The employment status (if so, what exactly is that)? When considered in detail, the aim does not seem consistent since two concepts are mixed:

- The description of the main status of the TP (i.e. in cases of competing conditions a decision is made which status "overrides" another under which conditions).
- The description whether the TP is currently in a certain condition (also if this is possibly "overridden" in the main status by another condition).

Generally, there are two options to generate employment / economic inactivity status variables as of W4:

- Continuing the previous generation logic with the new database.
- Revising the generation logic with the following aim:
  - conceptional streamlining (what exactly do the variables describe?)
  - improvement of decisions which were made based on the available data in the past, but are suboptimal in terms of content (i.e. not only continuation with a new database, but content-wise use of the more detailed database)
  - purification (i.e. removal of variables with very limited additional benefits)

A fundamental revision of the generation logic was decided. The former variables are used as follows:

**Table 25: Revision *erwerb*, *erwerb2*, *nichterw*, *nichtew2***

| Variable | Decision | Reason |
|---|---|---|
| erwerb | maintained<br>(W1: generated content-wise)<br><br>(W2 et seq.: -9) | Variable describes the survey concept of W1 ideally. The focus is on ET (to put it in simple terms, these override unemployment which in turn overrides anything else). Arguments against continuing are concerns relating to content which, thanks to the more detailed database, can be met with a reconception.<br><br>The variable is maintained for W1 since it is tailored to the survey concept. The characteristics (no parallels; focus on ET; no differentiation of unemployment in reported and not reported) remain limited to W1. |
| erwerb2 | omitted from SUF | This variable continues the logic of the survey concept in W1 in harmonised form.<br><br>Different problems occur here:<br><br>(1) There is a change as to which ETs are surveyed (W1: 1h/week vs. W2 et seq.: >EUR 400)<br>(2) Change of focus (W1: If ET [not marginal employment] exists → no survey of parallel AL/LU conditions; W2 et seq.: ET/AL/(in part also LU) in parallel possible)<br>(3) By continuing the logic of W1, the possibilities of the new database cannot be used sensibly (for example, to make more sensible decisions in terms of content)<br><br>Conclusion: A harmonised variable with a focus on ET (like previously in erwerb2) is the only way for a cross-wave harmonised variable. Generating it would be possible, but only based on content-wise/conceptual decisions which are not sensible. Since the concept of W1 is considered problematic, the provision of a harmonised variable is waived. |
| nichterw | omitted from SUF | The previous division into employment and economic inactivity status is omitted and replaced by main status + indicator for current employment (subject to social insurance) + indicator for current report of unemployment<br><br>W1: variable does not offer additional information as compared to the new main status variable<br>W2et seq.: only little additional information as compared to the new main status variable<br>Conclusion: Generally rather additional complexity with very limited use (e.g. students > 20h/working hours per week). For evaluations probably own determination of secondary conditions more sensible than previously provided variables. |
| nichtew2 | omitted from SUF | (see nichterw) |

The following variables have been generated as of wave 4:

| | |
|---|---|
| *etakt:* | *Currently employed (>EUR 400 per month), generated (as of wave 2)* |
| *alakt:* | *Currently reported as unemployed, generated (as of wave 2)* |
| *statakt:* | *Current main status, generated (as of wave 2)* |

The revision has the following aims:

- separation of the information on main status (*statakt*) from the information on currently ongoing spell types (*etakt*, *alakt*)
- clear documentation of rules for creating the main status
- differentiation between reported and not reported AL (where possible)

**etakt** (Currently employed (>EUR 400 per month), generated (as of wave 2))

This variable indicates that the TP had an ongoing spell of employment at the time of the personal interview of the respective wave (i.e. employment earning > EUR 400). The variable cannot be generated for W1 since the survey concept between W1 and the following waves differs (W1: min 1h/week; W2 et seq. > EUR 400/month).

A current ET exists if a censored ET spell is available in the spell dataset of the respective wave.

Categories of the variable to be generated:

| | |
|---|---|
| -10 | Item not relevant for questionnaire version |
| -5 | No generation poss. (missing values) |
| -3 | Not applicable (filter) |
| 1 | TP curr. employed (> EUR 400) |
| 2 | TP curr. economically inactive (> EUR 400) |

**alakt** (Currently reported as unemployed, generated (as of wave 2))

Indicates that the TP was reported unemployed at the date of the personal interview of the respective wave. This variable cannot be generated for wave 1 since the survey concept changed between W1 and the following waves (W1: AL only surveyed if there is no ET; W1: unemployed; W2 et seq.: reported unemployed).

A current reported AL exists if a censored AL spell (reported AL) is available in the spell dataset of the respective wave.

Categories of the variable to be generated:

| | |
|---|---|
| -10 | Item not relevant for questionnaire version |
| -5 | No generation poss. (missing values) |
| -3 | Not applicable (filter) |
| 1 | TP curr. reported unemployed |
| 2 | TP curr. not reported unemployed |

**statakt** (Current main status, generated (as of wave 2))

Indicates which main status the TP had at the date of the personal interview of the respective wave.

Basis for the generation are the spell datasets (W2/3: *ET/AL/LU spells*; W4f.: *BIO spells*) and the status as pupil/student/trainee in PB0100.

If in the respective wave a spell of a certain type is currently ongoing, then the respective condition exists for the individual. The spell type is determined in W2/W3 via the respective spell dataset (*ET-/AL-Spells*) or the gap condition (*LU0101* in *LU spells*). As of W4, the variable spelltyp can be used. In all waves only ongoing spells at the interview date (i.e. censored=1 in SUF of the respective wave) are considered. The current status as pupil or students/trainee from PB0100 is considered as if there were a currently ongoing spell in the respective wave.

Categories of the variable to be generated:

| | |
|---|---|
| -10 | Item not relevant for questionnaire version |
| -5 | No generation poss. (missing values) |
| -3 | Not applicable (filter) |
| 1 | Employed with income > EUR 400/month |
| 2 | Registered as unemployed |
| 3 | Pupil (until W3 in spells surv. acc. to cat.) |
| 4 | Vocational training/training/studies |
| 5 | Military service/civil service/etc. |
| 6 | Homemaker |
| 7 | Maternity protection/extended maternity leave/parental leave |
| 8 | Senior citizen/pensioner/early retirement |
| 9 | Other/main status unclear |
| 10 | Unempl., n. reported (as of W4 from o-e. q.) |
| 11 | Sick/unfit f. w./incapacit./disab.(from o-e. q.) |
| 12 | Self-e./helping fam. memb. (from o-e. q.) |

The allocation of codes should take place in stages:

**Table186: Rough allocation –spell of higher priority overrides spell of lower priority**

| Priority of a current spell (or analogous condition from PB0100) | Code in statakt (analogous to variable spelltyp) | Meaning |
|---|---|---|
| 1 | 2 | Reported unemployed/part. in measure |
| 2 | 1 | Employed with income > EUR 400/month |
| 3 | 8 | Senior citizen/pensioner/early retirement |
| 4 | 7 | Maternity protection/extended maternity leave/parental leave |
| 5 | 5 | Military service/civil service/etc. |
| 6 | 4 | Vocational training/training/studies |
| 7 | 3 | Pupil (until W3 from o-e. q.) |
| 8 | 12 | Self-e./helping fam. memb. (from o-e. q.) |
| 9 | 11 | Sick/unfit f. w./incapacit./disab. (from o-e. q.) |
| 10 | 10 | unempl., n. reported (as of W4 from o-e. q.) |
| 11 | 6 | Homemaker |
| 12 | 9 | Other/main status unclear |

The rough allocation remains in place if there are no valid values for additional information (see Table7).

**Table 27: Fine allocation for special cases**

| Rough allocation | Additional information | Decision |
|---|---|---|
| Registered as unemployed | Employed with income > EUR 400/month<br>+<br>Working hours (az2ges; actual working hours, sum via censored ETs) >= 15h | Employed with income > EUR 400/month |
| Employed with income > EUR 400/month | Vocational training/training/studies<br>+<br>Working hours (az2ges; actual working hours, sum via censored ETs) <= 20h | Vocational training/training/studies |

A current reported AL exists if a censored AL spell (reported AL) is available in the spell dataset of the respective wave (W2/3: *AL spells*; W4f.: *BIO spells*).

#### 5.4.1.3 Income variables and working hours in PENDDAT and BIO spells

The variables on current employment refer to the main employment in W1-W4[44]. Excluded from that is information on gross/net income in W2-W4 – this information refers to all currently ongoing employments > EUR 400 (imprecision regarding marginal employment wages). Spell-specific information is not available and will only be surveyed as of W5. The information is only surveyed as total value across all employments. This leads to two partial problems:

I. The generated variables on working hours and gross/net wage have referred to different employments (main ET or all ETs) as of W2. If hourly wages are calculated on this basis, this leads to errors for TPs with several ETs.

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

The generated variables on income and working hours will thus be revised accordingly in W4.

**Income variables**

The survey concept of income variables changed significantly between wave 1 and 2 without this leading to the formation of new variables: *brutto* (*bruttokat*) and *netto* (*nettokat*) reflect the income from the main employment in W1; as of W2, the income from all employments which are not marginal. This is inconsistent and potentially leads to errors in the evaluation. The revision is to correct this problem:

**Table 28: Revision of income variables[45]**

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

[44] W2-W3; this is the censored ET in the ET spell dataset. In case of several censored spells the spell with the highest amount of hours was selected. In case of several spells with the same amount of hours the longest lasting spell was selected. Only one employment was surveyed for senior citizens.

[45] In W1, there is only a categorical follow-up question for the main employment's net wage but not for other activities. This is accepted when generating *netges*. If the information (MV) on net income from other activities is missing, the variable *netges* cannot be generated.

**Revised variables (in W1-W3 already in the dataset):**

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

These variables refer to the respective main ET in W1. As of W2, they have been, however, filled with the cumulated information for all ETs (>EUR 400) since only this information is surveyed. The variable labels have been adjusted respectively as of W4. For W2-W4, the variables were filled with -9 since generation analogous to W1 is not possible.

**New variables in W4:**

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

This variable contains the cumulated information on gross income from all ETs (>EUR 400). This variable cannot be generated in this form for W1 since only the gross income for the main ET was surveyed. For W2 and W3, the variable is identical in terms of content with the variable *brutto*, which was included in the SUF W3 (i.e. prior to the revision as explained above). In W2-W4, only the cumulated gross income was surveyed – the source variables used in W2/W3 thus already include the respective information on total income from ETs > EUR 400. For W4, the variable will be generated analogous to W2/W3. As of W5, it will be generated based on spell-specific income information.

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

This variable contains the cumulated information on net income from all ETs (>EUR 400). The variable can be generated for W1 by combining the open-ended and categorical information on net income of the main employment with the information for other activities (however, the categorical follow-up question is missing here). For W2 and W3, the variable is identical with the variable *netto*, which was provided in the SUF W3. In W2-W4, only the cumulated net income was surveyed – the source variables used in W2/W3 thus already include the respective information on total income from ETs > EUR 400. For W4, the variable was formed analogous to W2/W3. As of W5, it will be generated based on spell-specific income information.

**Working hours**

By correcting the variables on (gross/net) income (see same section above), it has become impossible to generate hourly wages in the individual dataset since the only working hours available are the actual working hours from the main ET (variable *arbzeit* in *PENDDAT* of the SUF W3).

Analogous to the revision of the income variables, also a revision of the working hours variables in *PENDDAT* and in the *BIO spells* is necessary.

**Table 29: Revision of working hours variables**

| Variable | Content | Dataset | Generated for | | | Basis | | Note |
|---|---|---|---|---|---|---|---|---|
| | | | W1 | W2 | W3 | OFFA | KatA | |
| az1 | Spell ET, contract. | Bio spells | 0 | 1 | 1 | 1 | 0 | Cat W2f. |
| azhpt1 | Main ET, contract. | PENDDAT | 0 | 1 | 1 | 1 | 0 | Cat W2f. |
| azges1 | Total ET, contract. | PENDDAT | 0 | 1 | 1 | 1 | 0 | Cat W2f. |
| az2 | Spell ET, contract. | Bio spells | 0 | 1 | 1 | 1 | 1 | equivalent to former variable arbzeit (BIO spells); cat W2f.; ET with max(az2) = main ET (if two identical: ET which starts earliest |
| azhpt2 | Main ET, actual | PENDDAT | 1 | 1 | 1 | 1 | 1 | equivalent to former variable arbzeit (PENDDAT); cat. W1 != cat W2f. |
| azges2 | Total ET, actual | PENDDAT | 1 | 1 | 1 | 1 | 1* | cat W1!= cat W2f.; in W1 no cat. for secondary ET |

**Revised variables (in W1-W3 already in the dataset):**

arbzeit *(Weekly working hours incl. details in the case of irregular working hours, gen.)*

Variable omitted from *PENDDAT* and *BIO spells*. It is replaced in terms of content by *azhpt2* (*PENDDAT*) and *az2* (*BIO spells*).

**New variables in W4:**

az1 *(Contract. working hours, gen.)*
The variable is generated for all episodes in *BIO spells*. It contains the most current information, respectively, for the spell on contractual working hours (ET > EUR 400). The basis for the generation are the respective cross-section variables for which information was last surveyed in the respective spell.

Example:

- Spell created in W2, ended in W2: cross-section variables W2
- Spell created in W2, updated in W3 and W4: cross-section variables W4

**azhpt1** *(Current contractual working hrs. main employment (without marginal employment), gen.)*

The variable is generated for *PENDDAT*. It contains the contractual working hours of the main ET currently ongoing in the respective wave from the spell data (ET > EUR 400). The variable cannot be generated for W1 (-9) since the respective information has only been surveyed as of W2. As of W2, the generated variable on contractual working hours of the main ET from the respective spell data (*az1*) has been adopted into *PENDDAT*. Which ongoing spell is the main ET is also determined based on the actual working hours (generated variable *az2* in the spell data; analogous to the proceedings in W2/W3 which used variable *arbzeit* to determine this).

**azges1** *(Total current contractual working hrs. (without marginal employment), gen.)*

The variable is generated for *PENDDAT*. It contains the cumulated contractual working hours of all currently ongoing employments in the respective wave from the spell data (ET > EUR 400). The variable cannot be generated for W1 (-9) since the respective information has only been surveyed as of W2.

As of W2, this variable has been generated based on the generated variable for contractual working hours (*az1*) from the spell data. During generation, the information in the generated variable on contractual working hours (*az1*) is cumulated across all spells which are currently ongoing at the time of the interview. This information is not adopted into *PENDDAT*.

**az2** *(Actual working hours incl. details in case of irregular working hours, gen.)*

The variable is generated for all episodes in *BIO spells*. It contains the respective most current information on actual working hours for the spell and thus also integrates categorical information in case of irregular working hours. The basis for the generation are the respective cross-section variables for which information was last surveyed in the respective spell.

Example:

- Spell created in W2, ended in W2: cross-section variables W2
- Spell created in W2, updated in W3 and W4: cross-section variables W4

This variable replaces the formerly generated variable *arbzeit* in the ET spells (which is thus omitted). Generation is made analogous to the generation of *arbzeit* during data preparation W2/W3.

The variable is used to determine the main employment (main ET) in a wave, for which different information are adopted into *PENDDAT*. The main ET is the ongoing ET in a respective wave with the largest amount of hours. If there are several employments with the same amount of hours then the one which started first is selected.

**azhpt2** *(Actual current working hours main empl. (without marginal employment, incl. cat. info.), gen.)*

The variable is generated for *PENDDAT*. It contains the actual working hours of the currently ongoing main employment and thus also integrates the categorical follow-up question for irregular working hours. The variable replaces the variable *arbzeit* which was omitted in *PENDDAT*. The generation is made for W1-W4 analogous to the generation of the omitted variable.

The variable is generated based on the cross-section data in W1. It thus combines both the open-ended information on actual working hours and the categorical follow-up question. One-euro jobs, job creation measures, marginal employment and activities which are part of an apprenticeship are not considered here – for these cases the variable cannot be generated (-3) since no analogous information was surveyed in W2-W4.

As of W2, the generated variable on actual working hours of the main ET from the respective spell data (*az2*) has been adopted into *PENDDAT*. Which ongoing spell is the main ET is also determined based on the actual working hours (generated variable *az2* in the spell data; analogous to the proceedings in W2/W3 which used variable *arbzeit* to determine this).

The categorical follow-up question for irregular working hours differs between W1 and W2 et seq. Still, an integration across the waves is made.

**azges2** *(Actual total current working hours (without marginal employment, incl. cat. info.), gen.)*

The variable is generated for *PENDDAT*. It contains the cumulated actual working hours of all employments currently ongoing in the respective wave.

Therefore, W1 combines the hours of the main ET (after integration of the categorical information for irregular working hours) with the information on actual working hours of the other ETs. One-euro jobs, job creation measures, marginal employment and activities which are part of an apprenticeship are not considered here – for these cases the variable cannot be generated (-3) since no analogous information was surveyed in W2-W4.

As of W2, this variable has been generated on the basis of the generated variable for actual working hours (*az2*) from the spell data. During generation, the information in the generated variable on actual working hours (*az2*) is cumulated across all spells which are currently ongoing at the time of the interview. This information is not adopted into *PENDDAT*.

#### 5.4.1.4 Duplicate pointer in p_register

**zdub4** *(Pointer: personal identification no. of the individual doubled by the TP in wave 4 (2009/2010)*

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

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

A wave-specific pointer variable (zdub4) marking these cases is created in the person register to achieve this.

Two different types of problems must be differentiated:

1. (Real) duplicates

(Real) duplicates are individuals who de facto belong to two households in a wave[46]. The households concerned were interviewed and the individual is included in the respective household structures.

Since both individuals cannot have the same personal identification number in the SUF, the individual which was newly created in the current wave (=duplicate in duplicate HH) receives

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

a new permanent personal identification number following the known rules. The individual who already was part of the panel in a previous wave (i.e. the individual is already known with a permanent number in the person register of the previous wave) keeps the former personal identification number (=original in original HH).

If there were individual-level interviews with duplicate and original in the current wave, then the interview of the duplicate is removed and will not be used for the preload generation in the next wave either. Analogous to the other personal interviews deleted during data preparation, marking occurs in *pnetto*4*. The duplicate is treated as new respondent for whom no previous information is available in the next wave. Weighting thus only uses one of the two observations of the individual in the current wave. Special treatment of these cases is thus not necessary for weighting.

If there is only one individual-level interview for either duplicate or original in the current wave, then this interview is not removed, i.e. if there is no competing information from the interview with the original, the duplicate interview remains in the SUF. In this case, the information from the personal interview is also used for the preload generation of the next wave.

Wave 4 is the first to integrate a wave-specific pointer variable in the person register which points from a duplicate to the original (irrespective of the interview status of duplicate and original on the individual level). For the observation of the duplicate in the person register, this pointer variable thus contains the permanent personal identification number of the original, i.e. it can only be filled with a personal identification number for individuals who are duplicates. If an observation is no duplicate, it is filled with the variable "0" (analogous to the proceedings with other pointer variables) or with "-6" if the individual's household was not surveyed in the current wave or the individual is no longer part of a survey household (analogous to the allocation of code -6 in the other variables of the person register).

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

## 2. Potential duplicates

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

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

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

The following is done for the preload of the following wave:

- Original HH of the individual

The potential duplicate is not removed from the household structure, so the move-out can be reported. The personal interviews conducted in the duplicate HH are not used for the preload generation of the potential duplicate in the original HH. Individual and household are thus treated like regular skips.

- Duplicate HH of the individual

The preload is generally generated based on the respective last information, i.e. the individual is included in the household structure of the duplicate HH and the personal interviews conducted in the duplicate HH are considered for the generation.

The following wave can determine whether this is a real duplicate (see 1). In this case, the personal identification number of the individual in the duplicate HH will be corrected retrospectively, meaning that the individual will receive a new personal identification number retrospectively as described in (1) (in all relevant datasets). The duplicate pointer, too, will be filled retrospectively.

Categories of the variable to be generated:

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

#### 5.4.1.5 Spell data on the employment biography in BIO spells

**spellnr** *(all activities: spell number)*

This is the chronological cross-spell type spell counter. (Generated based on the same source variables as for chronological sorting in W2/W3)

**spellnret** *(employments: spell number)*

This is the spell type internal, chronological spell counter for the ET spells (ET > EUR 400). (Generated based on the same source variables as for chronological sorting in W2/W3)

**spellnral** *(unemployments: spell number)*

This is the spell type internal, chronological spell counter for the AL spells (reported AL). (Generated based on the same source variables as for chronological sorting in W2/W3)

**spellnrlu** *(other activities: spell number)*

This is the spell type internal, chronological spell counter for the LU spells (i.e. all other activities which are not ET or reported AL). (Generated based on the same source variables as for chronological sorting in W2/W3)

**spelltyp** *(all activities: type of activity, generated)*

W2/W3 first successively included all relevant ET spells and after that all AL spells of the TP. Following that, gaps of > 3 months and gaps in latest available data were included (LU). This led to three spell datasets which only contained one spell type (ET, AL or LU) for the SUF based on this information.

Starting in W4, the ET/AL spells will not be surveyed successively in a modular way and then completed by a gap survey, but the survey is made in an integrated way whereby also LU spells (i.e. spells which are not ET or reported AL) are surveyed directly. This results in an integrated biography spell dataset. The spells surveyed type-specifically in W2/W3 are integrated in the integrated dataset of W4.

The generated variable *spelltyp* is formed to allow easy distinction of the individual spell types across the individual waves.

Categories of the variable to be generated:

| | |
|---|---|
| 1 | Employed with income > EUR 400/month |
| 2 | Reported unemployed/part. in measure |
| 3 | Unempl., n. reported (as of W4 from o-e. q.) |
| 4 | Vocational training/training/studies |
| 5 | Military service/civil service/etc. |
| 6 | Homemaker |
| 7 | Maternity protection/extended maternity leave/parental leave |
| 8 | Senior citizen/pensioner/early retirement |
| 9 | Did something else |
| 10 | Pupil (until W3 from o-e. q.) |
| 11 | Main status not clear (from o-e. q.) |
| 12 | Sick/unfit f. w./incapacit./disab.(from o-e. q.) |
| 13 | Self-e./helping fam. memb. (from o-e. q.) |
| -1 | Don't know |
| -2 | Details refused |

**Table 30: Formation of spelltyp**

| Characteristics of spelltyp in terms of content | Characteristic is true for spell from W2/W3 if … | Characteristic is true for spell from W4 if … |
|---|---|---|
| Employed with income > EUR 400 | all ET spells | BIO0101=1 |
| Reported unemployed/part. in measure | all AL spells | BIO0101=2 |
| unempl., n. reported (as of W4 from o-e. q.) | LU spells + LU0101=3 | (corresp. code in BIO0101) |
| Vocational training/training/studies | LU spells + LU0101=4 | BIO0101=4 |
| Military service/civil service/voluntary social year | LU spells + LU0101=5 | BIO0101=5 |
| Homemaker | LU spells + LU0101=6 | BIO0101=6 |
| Maternity protection/extended maternity leave/parental leave | LU spells + LU0101=7 | BIO0101=7 |
| Senior citizen/pensioner/early retirement | LU spells + LU0101=8 | BIO0101=8 |
| Did something else | LU spells + LU0101=9 | BIO0101=9 |
| Pupil (until W3 from o-e. q.) | LU spells + LU0101=10 | BIO0101=3 |
| Main status not clear (from o-e. q.) | LU spells + LU0101=11 | (corresp. code in BIO0101) |
| Sick/incapacit./disab./unfit f. w. (from o-e. q.) | LU spells + LU0101=12 | (corresp. code in BIO0101) |
| Self-employed/helping family memb. (from o-e. q.) | LU spells + LU0101=13 | (corresp. code in BIO0101) |
| Don't know | LU spells + LU0101=-1 | BIO0101=-1 |
| Details refused | LU spells + LU0101=-2 | BIO0101=-2 |

**spintegr** *(all activities: concerned by spell integration W2/W3 and W4 et seq.?)*

In order to make it clear for the user with which survey logic a spell was surveyed in the integrated biography spell dataset, a generated variable is formed with which three types of spells can be differentiated: (1) Spells which only contain information reported in W2/W3, (2) spells which were created in W2/W3 but were updated with the new biography module in W4 and (3) spells only containing information reported in W4 et seq.

**Table 31: Formation of spintegr**

| Characteristics of spintegr in terms of content | Characteristic is true for spell from W2/W3 if … | Characteristic is true for spell from W4 if … |
|---|---|---|
| Information entirely from W2/W3 | ET/AL: created + ended<br>LU: created + ended | - |
| Information from W2/W3 and W4 (et seq.) | ET/AL: created + censored<br>LU: created + censored | ET/AL: Update previous W<br>LU: Update previous W |
| Information entirely from W4 (et seq.) | - | ET/AL/LU: created |

**Variables on working hours (az1, az2)**

The variables are described as part of the revision of the working hours variables (see above).

### 5.4.2 Error corrections

During the data preparation process for the scientific use file of the fourth wave, some changes were also made to the waves of PASS, which had already been delivered. These changes included corrections of errors that were detected after the completion of the scientific use file of the third wave. Tables 32 to 36 give an overview on the retroactive changes in the already delivered waves of PASS[47].

**Table 32: Overview of retroactive changes in the household dataset (*HHENDDAT*)**

| Altered variable | Dataset concerned | Altered wave | Type of alteration | Description of the alteration |
|---|---|---|---|---|
| *HEK1630* | *HHENDDAT* | 3 | Correction | HEK1630 has been surveyed since wave 3. The correct filter condition for HEK1630 is "HEK1600>0 & HEK1600<99 & HHALG2_an=0". During survey and preparation of the data from W3, the term "HHALG2_an=0" was not considered since it was not documented in the questionnaire template for wave 3. In terms of content, however, it is necessary to consider the control variable HHALG2_an since only households which do not receive UB II can receive Children's Allowance (Kinderzuschlag). Households with currently ongoing receipt of UB II in wave 3 were coded using code "-3" retroactively. This creates data consistency with the following waves. Additionally, the variable label was corrected. The name of the label must be "Kinderzuschlag" (children's allowance) and not "Kindergeldzuschlag" (supplementary child benefit). |
| *HD0300* | *HHENDDAT* | 2,3 | Correction | The value labels of codes 1 to 4 referred to the reference date 01/2005. This was only correct in the SUF of wave 1. The reference date moves back one year in each survey wave. The labels were adjusted: Instead of "01.2005" there is now a reference to "Ref.dat". |

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

**Table 32: Overview of retrospective alterations in the household dataset (HHENDDAT) (continued)**

| Altered variable | Dataset concerned | Altered wave | Type of alteration | Description of the alteration |
|---|---|---|---|---|
| *nextstra* | *HHENDDAT* | 1-3 | Omitted | The variable "nextstra" is no longer included in the PASS delivery of wave 4. The variable describes the next drawing layer with which a layer can be merged if there is only one single PSU left in one layer. This variable must, however, be updated across all previous and subsequent waves since otherwise strpsu (the layer code) is no longer constant within a PSU. Considering a large amount of layers to be merged in the refreshment sample UB II 2007 in wave 4 and the fact that further small subsamples are added every year, it must be expected that there will have to be a very rough merging of layers very quickly, which no longer corresponds to the original layers. The variable was therefore omitted. |
| *bik* | *HHENDDAT* | 1-3 | Correction | The variable bik was provided in an unusual coding in which code 0 stood for densely populated agglomerations and code 9 for the smallest and most weakly linked regions. It was recoded to the usual coding in which code 10 stands for densely populated agglomerations and code 1 for the smallest and most weakly linked regions. |

**Table 33: Overview of retrospective alterations in the individual dataset (*PENDDAT*)**

| Altered variable | Dataset concerned | Altered wave | Type of alteration | Description of the alteration |
|---|---|---|---|---|
| **isco_it*<br>*berpr_it* | *PENDDAT* | 1-3 | Omitted | The Infratest coding is no longer available as of wave 4. The ZUMA coding is the basis for the subject-related generations. |
| **kldb** | *PENDDAT* | 1-3 | Correction | No coding by Infratest anymore as of wave 4. Integration in existing variables possible - names and variable labels had to be adjusted for that. |
| *alg1s05* | *PENDDAT* | 1-3 | Omitted | According to the label, the variable refers to whether the respondent has received UB I since January 2005. In the newer subsamples this fact can no longer be determined since the retrospective interview period is too long past. Continuing this variable thus does not make sense. |
| *arbzeit* | *PENDDAT* | 1-3 | Omitted | See Chapter 5.4.1.3 |
| *brutto**<br>*netto** | *PENDDAT* | 2-3 | Correction | See Chapter 5.4.1.3 |
| *nichterw*<br>*nichtew2*<br>*erwerb2* | *PENDDAT* | 1-3 | Omitted | See Chapter 5.4.1.2 |
| *alakt*<br>*etakt*<br>*statakt* | *PENDDAT* | 2-3 | Added | See Chapter 5.4.1.2 |
| *brges* | *PENDDAT* | 2-3 | Added | See Chapter 5.4.1.3 |
| *netges* | *PENDDAT* | 1-3 | Added | See Chapter 5.4.1.3 |
| *azhpt1*<br>*azges1* | *PENDDAT* | 2-3 | Added | See Chapter 5.4.1.3 |
| *azhpt2*<br>*azges2* | *PENDDAT* | 1-3 | Added | See Chapter 5.4.1.3 |
| *mhh*<br>*vhh* | *PENDDAT* | 3 | Correction | The two variables must be filled with -10 for senior citizens' interviews (fb_vers=3). Instead, these two variables were filled with content in wave 3. |
| *beruf1*<br>*beruf2* | *PENDDAT* | 3 | Correction | The two variables are updated. Initially, the information from the current wave is considered. The value from the previous wave is adopted if there is no new information from the current wave. Minor errors occurred during updating. In *beruf1* corrections were made in 19 cases, in *beruf2* in 34 cases. |

**Table 34: Overview of retroactive corrections in spell datasets as basis for the integrated Bio spell dataset (bio_*spells*)**

| Altered variable | Dataset concerned | Altered wave | Type of alteration | Description of the alteration |
|---|---|---|---|---|
| *egp esec* | *et_spells* | 2-3 | Correction | As of W2, StiB (simple classification) has been queried with a changed scale as compared to W1 (W2 et seq.: PET1510 and ET060*; W1: PET1500). This was not considered in the et_spells when forming egp and esec. This led to wrong allocations since StiB is a source when generating these two indicators. |
| *zensiert* | *lu_spells* | 2-3 | Correction | Spells of target persons with a censored LU spell from W2 but without an overlapping spell of the same type in W3 who were interviewed in W3 were closed in the censoring variable with special code -5. |
| *zensiert* | *lu_spells, et_spells, al_spells* | 2-3 | Correction | Ongoing spells from W2 for which the TP was interviewed with the senior citizen's questionnaire in W3 were closed in the censoring variable with special code -5. |

**Table 35: 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 |
|---|---|---|---|---|
| *bgbezb3* | *p_register* | 3 | Correction | There were significant differences in the preparation logic between wave 3 on the one hand and the previous waves on the other hand. For example, there was no comparison with *alg2abez* anymore. Additionally, the codes -1, -2 and -4 from variables *AL20602* and *AL20702** were processed inconsistently. The preparation logic was retroactively adjusted to the proceedings from wave 3. |

**Table 36: Overview of retrospective alterations in the weighting datasets (hweights; pweights)**

| Altered variable | Dataset concerned | Altered wave | Type of alteration | Description of the alteration |
|---|---|---|---|---|
| *ppleib* | pweights | 1 | Correction | When creating the ppleib variable for the re-participation probabilities of participants from wave 1 to wave 2, the non-responses due to refusal of participation in the panel were not considered. This was corrected in the SUF of wave 4. Due to refusal of panel participation from wave 1 to wave 2 the medium failure rate (propensity model with the constant only) was used as the constant factor instead of a propensity model, due to the low failure rate for the correction of the ppbleib variable. The ppbleib variable was constantly multiplied by the factor 1.059 (18954 participants in the first wave divided by 17900 participants willing to participate in the panel in the first wave). |

## 5.5 Anonymisation

All data surveyed by the IAB as a special department of the Federal Employment Agency (BA) are social data, which places high demands on data protection. It was therefore necessary to include some of the variables in the scientific use file in simplified form. These variables are generally labelled with the flag "anonymised" in the variable label. For the same reason it was also necessary to exclude available regional information, with the exception of the German federal states and information on East/West Germany derived from this. For reasons of data protection, neither the data on family relationships in the household nor the first names of the household members are part of the scientific use file. References to the household structure are provided, however, by generated variables, for example on the household and benefit community type (*hhtyp*[48], *bgtyp*[49]), indicator variables on partners in the household (*apartner; epartner*[50]), pointer variables pointing to parents and partners in the household (*zmhh; zvhh; zparthh*[51]) and various indicator variables which show whether parents (*mhh; vhh*[52]) or children of the target person (e. g. *ekind*[53]) are living in the household.

Table37 provides an overview of the variables concerned and the process of anonymisation[54] in the individual dataset. Table38 shows the anonymised variables of the employment spell dataset.

48 Contained in the household dataset (*HHENDDAT*), see Chapter 4.5.2
49 Wave-specific variables contained in the person register (*p_register*), see Chapter 4.4.
50 Contained in the individual dataset (*PENDDAT*), see Chapter 4.4.
51 Wave-specific variables contained in the person register (*p_register*), see Chapter 4.4.
52 Contained in the individual dataset (*PENDDAT*), see Chapter 4.4.
53 Contained in the individual dataset (*PENDDAT*), see Chapter 4.4.
54 If non-anonymised versions of one or several variables are indispensable for your research, please contact the Forschungsdatenzentrum (Research Data Centre) to find a suitable possibility of obtaining access to the data. The form of this access will depend on the research project and the variables necessary for it.

**Table 37: Overview of the anonymised variables in the individual dataset (*PENDDAT*) in wave 4**

| Varname | Variable label | Procedure |
|---|---|---|
| *PD0100* | Year of birth (date of birth, anonymised) | The precise date of birth was shortened to year of birth. |
| *gebhalbj* | Half-year of birth, generated | The precise date of birth was shortened to an indicator for the first or second half of the year. |
| *PET1210* | Last occupational status, simple classification (before January 2005) (anon.) | For technical reasons, professional and regular soldiers were recorded separately in the survey. Due to the small amount of case numbers and as this group is not usually asked about occupational status anyway, this group was merged with that of civil servants and judges. |
| *PET1250* | Last occup. status civil servant: detailed information, incl. soldiers (before January 2005) (anon.) | This variable contains additional cases. The professional and regular soldiers from PET1240 were added to the corresponding civil servant category. The variable for professional and regular soldiers PET1240 is not supplied. |
| *PET1211* | Last occup. status, simple class. (incl. spell info.) (anon.), gen. | Procedure analogous to *PET1210*. |
| *PET1251* | Last occup. status civil servant: detailed information, incl. soldiers (incl. spell info.) (anon.), gen. | Procedure analogous to *PET1250*. The variable for professional and regular soldiers PET1240 is not supplied. |
| *stiblewt* | Occupational status, last employment, code number, generated | When generating the occupational status variable, professional and regular soldiers are assigned to the corresponding civil servant category. |
| *PET1510* | Current occup. status, simple classification, surv. as of W2 (anon.) | Procedure analogous to *PET1210*. |

**Table37: Overview of the anonymised variables in the individual dataset (PENDDAT) in wave 4 (continued 1)**

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

**Table37: Overview of the anonymised variables in the individual dataset (PENDDAT) in wave 4 (continued 2)**

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

**Table37: Overview of the anonymised variables in the individual dataset (PENDDAT) in wave 4 (continued 3)**

| Varname | Variable label | Procedure |
|---|---|---|
| *ozulandc* | Father's father: country of residence before migration, incl. open info., categories (anonymised) | Procedure analogous to *PMI1000a* |
| *ozulandd* | Father's mother: country of residence before migration, incl. open info., categories (anonymised) | Procedure analogous to *PMI1000a* |
| *ozulande* | Mother's father: country of residence before migration, incl. open info., categories (anonymised) | Procedure analogous to *PMI1000a* |
| *ozulandf* | Mother's mother: country of residence before migration, incl. open info., categories (anonymised) | Procedure analogous to *PMI1000a* |

**Table 38: Overview of the anonymised variables in the BIO spell dataset (bio_*spells*) in wave 4**

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

## 5.6 Receipt of Unemployment Benefit II

Receipt of Unemployment Benefit II at the household level was already recorded in spell form the first to the third waves. This concept was continued in wave 4 but with a slightly revised set of questions. Besides changes in phrasing, the question for reasons for the start of the receipt of Unemployment Benefit II was newly included (AL20550a to AL2055h or with coded answers to open-ended questions *AL20551a* to *AL20551h*).

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

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

The households from the refreshment sample which were interviewed for the first time in wave 4 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 2008 or if no information was provided about changes in the household, then the household‘s receipt of Unemployment Benefit II from January 2008 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 4. Here it is necessary to distinguish between (1) new variables that refer to a particular wave, (2) new variables that do not refer to a particular wave and (3) variables that are no longer surveyed in wave 3.

1. Also in wave 4, new wave-specific cross-sectional variables were included in the Unemployment Benefit II spell dataset. These were: *AL20603, AL20703a to AL20703o, AL20803 and AL20903*. These variables refer to the interview date of wave 4. Cross-sectional variables also exist for the interview dates of the previous waves which contain the analogous information referring to the respective wave. Table provides an overview of the cross-sectional information contained in the Unemployment Benefit II spell dataset.

**Table 39: Cross-sectional variables in the UB II spell dataset (*alg2_spells*)**

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

2. Embedded in the spells of receipt of Unemployment Benefit II is information on times of benefit cuts. Up to wave 3, there were up to six 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. Two additional cuts were introduced, which carry the abbreviations "g" and "h", since the existing maximum number of six cuts per period of receipt did no longer suffice due to the renewed update[55]. Furthermore, question 22200 features two new items (AL22200g and AL22200h; including open coding AL22201g and AL22201h). Furthermore, the data structure corresponds to that from wave 3.

3. The reason for the cut, *AL21900a to AL21900e*, was not surveyed in wave 4 either. Accordingly, no responses to open-ended questions were coded to the variables *AL21901a to AL21901h* any longer.

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

As was done in wave 1 to 3, the information on receipt of Unemployment Benefit II was also subjected to a number of plausibility checks in wave 3. Inadmissible overlaps and dates of spells of Unemployment Benefit II receipt or of benefit cuts were corrected if necessary. In principle, changes were only made to the generated date variables (*bmonat; bjahr; emonat; ejahr*) of the spell of Unemployment Benefit II receipt, the spells of benefit cuts (*alg2kbm; alg2kbj; alg2kem; alg2kej*) and 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 4 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 4. Two variants are to be distinguished here. In the first (1), only the censoring indicator zensiert is changed. The second variant (2) is an update of the spell which was censored in the previous wave using information gathered in wave 4 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 4 are merged with the spell dataset (3). These three variants are outlined briefly below:

1. *Cases in which the household in wave 4 contradicts an ongoing spell of Unemployment Benefit II receipt at the interview date in the previous wave.*

[55] 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 "a", those relating to the second cut with "b" etc.

If the household contradicted the information that there was an ongoing spell of Unemployment Benefit II receipt at the time of the previous wave, either explicitly or implicitly (by reporting an end date that preceded the interview date in the previous wave) in the update question, then *zensiert* was set to "2" (no). The information provided in the interview of the previous wave is assumed to be correct. As it is not possible to make any reliable statements about the continued duration of the benefit receipt beyond the date of the interview in the previous wave, it is assumed that the benefit receipt ended in the month of the interview in the previous wave. The reported and generated variables for the end date of the spell (*AL20300, AL20400* and *emonat, ejahr*) as well as the question whether a spell still continues (*AL20500*) remain unchanged.[56] The generated end date of the Unemployment Benefit II spell (*emonat; ejahr*) was already set to the interview date of the previous wave in the previous wave.

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

   If information about the end date of a spell of Unemployment Benefit II receipt that was censored in the previous wave is available in wave 4, then the spell which was censored in the previous wave was updated using the current information. First, the surveyed end date (*AL20300; AL20400*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the receipt of Unemployment Benefit II is still ongoing (*AL20500*) and the censoring indicator (zensiert) are overwritten with the information gathered in the previous wave. Furthermore, the spells of benefit cuts reported in the fourth wave and the cross-sectional data referring to wave 4 (*AL20603; AL20703a to AL20703o, AL20803, AL20903*) were included.

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

Spells reported for the first time in wave 4 were added to the Unemployment Benefit II spell dataset. Then the spell counter was generated anew in order to create a variable *spellnr* without gaps.

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

### 5.6.5 Data editing due to erroneous preload input

In wave 4, an error occurred during the transmission of the preload to the database used for questionnaire control. The original preload version imported to the database had to be updated again. This led to an error where variable alg2bez_h_vw (current receipt of UB II of the HH in the previous wave) was not updated or wrongly updated.

This resulted in a certain group (372 households) being confronted with a wrong preload in the wave 4 interviews. These cases did not currently receive UB II in the previous wave. Nevertheless, the UB II update question was posed mistakenly in wave 4.

This concerns cases which updated and ended an attached spell of the previous wave in wave 3 and which did not have an ongoing UB II spell at the interview date in wave 3.
Within this context, data preparation conventions for spells were made as follows, which can be determined via variable *fallgruppe* in the dataset:

**a.) Attached spell contradicted (case group 1)**

These spells were not included in the UB II spell dataset since there is a contradiction.

**b.) Attached spell updated and ongoing without interruptions until the interview date of wave 4 (case group 2)**

In this case, the wrong information from the preload was updated since there was no contradiction. For the time being, the most current information was trusted. This results in the spell end being considered as unproblematic in this case. The begin of the spell, however, is doubtful. The following variants existed for corrections:

a. Accepting the update of the wrong attached spell (i.e. spell becomes part of the UB II spell dataset)
b. Determination of the generated start date (bmonat, bjahr): interview date of the previous wave
c. Determination of the reported start date (AL20100, AL20200): -4

**c.) Attached spell updated and ended before the interview date of wave 4 with other UB II spells (case group 3) or without other UB II spells (case group 4)**

In these cases, it must be suspected that these are artificial spells which actually did not occur. In contrast, only an artificial start date is assumed for case group 2. In cases in which the mistakenly updated attached spell ended directly with the interview date of the previous wave or before, it was assumed that this was an incorrectly included contradiction. These spells were removed accordingly.
In cases in which the mistakenly updated attached spell ended after the interview date of the previous wave, it is unclear whether this is an artificial spell. Therefore, these spells remain in the dataset:

- Determination of the generated start date (bmonat, bjahr): interview date of the previous wave
- Determination of the reported start date (AL20100, AL20200): -4

**d.) Attached spell not updated because UB II block was not activated**

In these 17 spells, the censoring in the variable *zensiert* was closed with special code -6.

## 5.7 Employment biographies

Employment, unemployment and gap periods at the individual level were recorded in spell form already in the second and third waves. This concept of modular spell survey was changed to an integrated survey of the employment biography in wave 4. For individuals that were asked for their employment biography for the first time in wave 4, the reference date for the start of the retrospective interval was adjusted. In wave 4, all spells of employment and unemployment since January 2008 were to be reported here. Individuals who were interviewed on their employment biography already in the previous wave, however, should report all new spells since the date of the last interview.

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

Besides ET and AL spells lasting from the previous wave, also gap spells were updated in wave 4 for the first time. In order to update the spells which were still ongoing in the previous wave and were therefore right-censored in the spell dataset, dependent interviewing questions are included in the personal questionnaires. 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 BIO spell dataset

Regarding its structure, the BIO spell dataset orients itself on the modular ET, AL and LU spell datasets. The information surveyed in a modular way in waves 2 and 3 was merged with the information surveyed in an integrated way in wave 4. ET-specific variables kept their name in the BIO spell dataset as compared to the ET SUF of wave 3, analogously also AL- and LU-specific variables. Variables which are the same in ET, AL and LU were standardised (BIO0100, BIO0101, BIO0200, BIO0300, BIO0400, BIO0500, BIO0600) or were already standardised in the original datasets of the SUF wave 3 (bmonat, bjahr, emonat, ejahr, zensiert). Furthermore, a variable for the type of activity (spelltyp) and for spell integration (spintegr) was included, a comprehensive spell number was generated (spellnr) and the existing spell numbers were renamed (spellnret, spellnral, spellnrlu).

Due to the integration of the employment and unemployment spells reported in wave 4 into the BIO spell dataset, new ET- and AL-specific variables are added. Here it is necessary to distinguish between (1) new variables that refer to a particular wave and (2) new variables that do not refer to a particular wave.

1. The ET-specific variables in the BIO spell dataset *ET0600 to ET2200* are considered as wave-specific, cross-section information related to wave 2; variables ET0601 to ET2201 as related to wave 3 and the new variables *ET0552 to ET2202* as cross-section information related to wave 4. Table 40 provides an overview of the ET-specific cross-section information in the BIO spell dataset.

**Table 40: ET-specific cross-section variables in the BIO spell dataset (bio_spells)**

| | **Cross-sectional variable with information referring to …** | | |
|---|---|---|---|
| | **Wave 2:** | **Wave 3:** | **Wave 4:** |
| Occupational status (simple and detailed classification) | *ET0600*<br>*ET0700*<br>*ET0800*<br>*ET1000*<br>*ET1100*<br>*ET1200* | *ET0601*<br>*ET0701*<br>*ET0801*<br>*ET1001*<br>*ET1101*<br>*ET1201* | *ET0502*<br>*ET0602*<br>*ET0702*<br>*ET0802*<br>*ET1002*<br>*ET1102*<br>*ET1202* |
| Supervisory function; number of employees supervised | *ET1300*<br>*ET1400* | *ET1301*<br>*ET1401* | *ET1302*<br>*ET1402* |
| Cancellation of limitation of an initially limited employment | *ET1700* | *ET1701* | *ET1702* |
| Working hours (contracted; actual; average for irregular working hours) | *ET2000*<br>*ET2100*<br>*ET2200* | *ET2001*<br>*ET2101*<br>*ET2201* | *ET2002*<br>*ET2102*<br>*ET2202* |

The BIO spell dataset furthermore includes an AL-specific variable which is to be understood as wave-specific cross-section information (AL1300 for wave 2; AL1301 for wave 3 and AL1302 for wave 4). Table 41 gives an overview of the cross-sectional information contained in the spell dataset.

**Table 41: Cross-sectional variables in the AL spell dataset (*al_spells*)**

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

2. The non wave-specific ET variable *ET2700* (gross income for ended activity) was surveyed first in wave 4 and was integrated in the BIO spell dataset.

### 5.7.3 Plausibility checks and corrections of the spell datasets

At the individual level, the plausibility checks and corrections orient themselves on wave 2 and wave 3. Similar to wave 3, the checks were made only within one spell type. Cross-spell type checks were not carried out. As with the spell data on Unemployment Benefit II receipt, corrections and recodings were only carried out in the generated date variables. Here, too, details on seasons were recoded into definite months, "-8" values were set for implausible responses and date information was replaced or rendered plausible. As only the generated date variables were edited, the original information gathered in the survey is available to the user in the date variables BIO0200-BIO0500 and AL0800-AL1100, thus permitting the user to conduct his/her own checks and corrections.

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

### 5.7.4 Update of spell datasets

After the spells that were reported in wave 4 had been converted into spell format and following the plausibility checks and corrections where inadmissible overlaps and spells with implausible dates were corrected, the spells which were still ongoing at the time of the interview in the previous wave were updated using the information recorded in wave 4.

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 4 are merged with the spell dataset (3). These three variants are outlined briefly below:

1. *Cases in which the individual in wave 4 contradicts an ongoing spell at the interview date in the previous wave.*

   If the individual contradicted the information that there was an ongoing spell at the time of the previous wave, either explicitly or implicitly (by reporting an end date that preceded the interview date in the previous wave) in the update question, then the censoring indicator *zensiert* was set to "2" (no). The information provided in the interview of the previ-

ous wave is assumed to be correct. As it is not possible to make any reliable statements about the continued duration of the spell beyond the date of the interview in the previous wave, it is assumed that the spell ended in the month of the interview in the previous wave. The reported and generated variables on the end date of the spell (BIO0400*, BIO0500 and emonat, ejahr*) as well as the question whether a spell still continues (BIO0600) remain unchanged[57]. The generated end date of the spell (*emonat; ejahr*) was already set to the interview date of the previous wave in the previous wave.

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

   If information about the end date of a spell that was censored in the previous wave is available in wave 4, then the spell which was censored in the previous wave was updated using the current information. For ET spells the recorded end date (BIO0400*; BIO0500*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the spell was still ongoing (BIO0600), the reason for the cancellation of a work contract (*ET2300*), the generated variables on occupational status and weekly working hours (*stib, az1, az2*) and the censoring indicator (zensiert) were overwritten with the information gathered in wave 4. Furthermore, the cross-sectional data referring to wave 4 (*ET0552 to ET2202*) were included.

   For AL spells the recorded end date (BIO0400*; BIO0500*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the spell was still ongoing (BIO0600), the reason for the end of unemployment (*AL0600, AL0601*) and the censoring indicator (zensiert) were overwritten with the information gathered in wave 4. Furthermore, the cross-sectional data referring to wave 4 (*AL1302*) was included. AL spell data, moreover, feature the exception that the spell of Unemployment Benefit I (receipt of UB I) is recorded within an AL spell. Which information is updated depends on whether there already was a receipt of UB I in this spell of unemployment and whether this receipt was ongoing in the previous wave:

   - If, in the previous wave, there also was an ongoing receipt of UB I in the AL spell to be updated, then the surveyed end date of the receipt (*AL1000, AL1100*), the indicator as to whether the spell is ongoing (*AL1200*), the generated end date of the receipt (*alg1em, alg1ej*) and the censoring indicator of the receipt (*alg1akt*) were overwritten with the information surveyed in wave 4.

   - If no UB I was received in any previous waves in the AL spell to be updated, then the information on UB I receipt was overwritten with the information surveyed in wave 4. Besides the indicator as to whether UB I was received in the AL spell (*AL0700*), the reported start and end date (*AL0800, AL0900, AL1000, AL1100*), the indicator for on-

[57] 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.

going receipt (*AL1200*) and the respective generated variables (*alg1bm, alg1bj, alg1em, alg1ej, alg1akt*) were replaced with the newly recorded information.

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

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

   Spells reported for the first time in wave 4 were added to the BIO spell dataset. Then the spell counter was generated anew in order to create a variable *spellnr* without gaps.

Updating the spell datasets does not affect the module-specific spell numbers of the spells already contained in the SUF from wave 3 (spellnret, spellnral, spellnrlu). These spells keep their spell number. The new spells from wave 4 are added to the respective dataset and the spell number is updated.

## 5.8 One-euro job spell dataset (ee_spells)

In the fourth wave, the concept for surveying participation in employment and training measures was thoroughly revised. The MN spell dataset was replaced by the one-euro spell dataset (ee_spells) in wave 4. The reference date as of which to consider one-euro jobs was January 2009.

### 5.8.1 Structure of the EEJ spell dataset

For the EEJ (one-euro jobs) spell dataset it must be considered that there are also spells if the one-euro job was not performed, i.e. there was no participation.

### 5.8.2 Plausibility checks and corrections in the EEJ spell dataset

The EEJ spell dataset on the participation in one-euro jobs was checked for plausibility and corrected. The plausibility checks contained checks of dates, of the reference date (1/2009) and checks for logical inconsistencies in case of respondents with several one-euro job spells.

Only the generated date variables were corrected and recoded. Details on seasons were recoded into definite months, "-8" values were allocated for implausible responses and date information was replaced or rendered plausible. Following that, a spell counter *spellnr* was generated. The generation was performed analogously to the chronological counters in the BIO spell datasets. Non-participating spells were not included in the sorting and thus kept their original *spellnr*.

# 6 Weighting wave 4

The construction of the weights for the fourth wave was generally made similar to the third wave (see Berg et.al, FDZ Datenreport 06/2010). Other than in wave 3, the GREG procedure was again used for the calibration of wave 4. This, however, has no significant relevance for the weights[58] (wave 3 was calibrated using the raking procedure (IPF)).

The starting point for the weighting procedure for the fourth wave and for the longitudinal section from wave 3 to wave 4 were the cross-sectional weights from wave 3 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 panel households in the fourth wave

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

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

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

## 6.2 Design weights for the refreshment sample in the fourth wave

In the fourth wave, the panel was again refreshed by sampling new households from the new inflows to benefit recipiency. All households that were in receipt of benefits in July 2009 but had had no probability of being selected for the register data sample in the same month in 2008, 2007 and 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=5*, the design weight of the refreshment sample is included in the variable *dw_ba*.

[58] The weights in these procedures are generally very similar. No corrected standard mistakes can be estimated in the raking procedure as compared to GREG.

## 6.3 Propensity to participate again - households

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

**Table 42: Variable overview, codes and reference categories for the logit models of the re-participating households**

| Variable code and reference category | Explanation |
|---|---|
| alter_1 | Household reference person (HRP) younger than 30 years |
| alter_2 | HRP 30 – 39 years old |
| alter_4 | HRP 50 – 64 years old |
| alter_5 | HRP 65 years and older |
| Reference category | HRP 40 – 49 years old |
| sex_1 | HRP male |
| Reference category | HRP female |
| nichtdeutsch | HRP has nationality other than German |
| Reference category | HRP has German nationality or missing information |
| sprache_nichtdeutsch | Language HRP: primarily spoken language in the household is not German |
| Reference category | Language HRP: primarily spoken language in the household is German |
| schulbil_1 | School qualification HRP: no qualification |
| schulbil_2 | School qualification HRP: lower secondary school |
| schulbil_4 | School qualification HRP: college/university qualification |
| Reference category | School qualification HRP: intermediate secondary school/still pupil |
| gesundheit_3 | Subjective evaluation of the health state of the HRP: satisfactory |
| gesundheit_4 | Subjective evaluation of the health state of the HRP: not so good |
| gesundheit_5 | Subjective evaluation of the health state of the HRP: bad |
| Reference category | Subjective evaluation of the health state of the HRP: very good to good |
| zufrieden_1 | General life satisfaction HRP: scale value 0-2 |
| zufrieden_2 | General life satisfaction HRP: scale value 3-5 |
| zufrieden_3 | General life satisfaction HRP: scale value 6-8 |
| Reference category | General life satisfaction HRP: scale value 9-10 |
| anz_0_3 | Number of individuals in the household aged 0 – 3 years |
| anz_4_6 | Number of individuals in the household aged 4 – 6 years |
| anz_7_14 | Number of individuals in the household aged 7 – 14 years |
| anz_65 | Number of individuals in the household aged 65 years and older |
| DinvalidAge | Age responses that cannot be evaluated |
| Reference category | Number of individuals in the household aged 15 – 64 years |
| eigentum | Type of residential property: proprietor |
| Reference category | Type of residential property: tenant, missing information |
| wnka_1 | Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: none |
| wnka_3 | Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: 11 and more |
| Reference category | Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: 1-10 |
| hhincome_1 | Household income: up to EUR 870 |
| hhincome_2 | Household income: EUR 871-1.400 |
| hhincome_4 | Household income: more than EUR 2,200 |
| Reference category | Household income: EUR 1,401-2,200 |
| alg2_1 | UB II receipt of the household: current receipt of UB II |
| Reference category | UB II receipt of the household: no current receipt of UB II |
| stichprobe1 | BA sample |
| stichprobe3 | Refreshment sample (BA) wave 2 |
| stichprobe4 | Refreshment sample (BA) wave 3 |
| Reference category | Microm sample |
| anzkon_1 | Number of contact attempts CATI/CAPI: 1 contact attempt |
| anzkon_3 | Number of contact attempts CATI/CAPI: 4-9 contact attempts |
| anzkon_4 | Number of contact attempts CATI/CAPI: 10 and more contact attempts |
| Reference category | Number of contact attempts CATI/CAPI: 2-3 contact attempts |

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

| Variable code and reference category | Explanation |
|---|---|
| blneualt_2 | New federal states |
| Reference category | Old federal states |
| bundesld_1 | Federal state: Schleswig-Holstein |
| bundesld_2 | Federal state: Hamburg |
| bundesld_3 | Federal state: Lower-Saxony |
| bundesld_4 | Federal state: Bremen |
| bundesld_6 | Federal state: Hesse |
| bundesld_7 | Federal state: Rhineland-Palatinate |
| bundesld_8 | Federal state: Baden-Wuerttemberg |
| bundesld_9 | Federal state: Bavaria |
| bundesld_10 | Federal state: Saarland |
| bundesld_11 | Federal state: Berlin |
| bundesld_12 | Federal state: Brandenburg |
| bundesld_13 | Federal state: Mecklenburg-Vorpommern |
| bundesld_14 | Federal state: Saxony |
| bundesld_15 | Federal state: Saxony-Anhalt |
| bundesld_16 | Federal state: Thuringia |
| Reference category | Federal state: North Rhine-Westphalia |
| bik_1 | BIK size class of municipality: population of less than 2,000 |
| bik_2 | BIK size class of municipality: population of 2.000 to under 5.000 |
| bik_3 | BIK size class of municipality: population of 5,000 to under 20,000 |
| bik_4 | BIK size class of municipality: population of 20,000 to under 50,000 |
| bik_5 | BIK size class of municipality: population of 50,000 to under 100,000 STYP 2/ 3/ 4 |
| bik_6 | BIK size class of municipality: population of 50,000 to under 100,000 STYP 1 |
| bik_7 | BIK size class of municipality: population of 100,000 to under 500,000 STYP 2/ 3/ 4 |
| bik_8 | BIK size class of municipality: population of 100,000 to under 500,000 STYP 1 |
| bik_9 | BIK size class of municipality: population of 500,000 and more STYP 2/ 3/ 4 |
| Reference category | BIK size class of municipality: population of 500,000 and more STYP 1 |

**Table 43: 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 |
| alter_1 | .2449415 | 0.048 | -.563471 | 0.000 | -.5253855 | 0.000 |
| alter_2 | .2266293 | 0.038 | -.3007197 | 0.053 | -.2684374 | 0.002 |
| alter_4 | -.1421211 | 0.119 | .4448706 | 0.009 | .3424315 | 0.000 |
| alter_5 | -.0611345 | 0.760 | .6635783 | 0.196 | .1243719 | 0.605 |
| sex_1 | .1283306 | 0.059 | -.2378574 | 0.032 | -.1962515 | 0.002 |
| nichtdeutsch | -.0153975 | 0.926 | .0860101 | 0.714 | -.1611078 | 0.194 |
| sprache_nichtdeutsch | .1221601 | 0.421 | .02652 | 0.902 | -.1991273 | 0.078 |
| schulbil_1 | .0868956 | 0.621 | -.3618542 | 0.100 | -.440921 | 0.001 |
| schulbil_2 | .006146 | 0.940 | -.1451147 | 0.272 | -.077809 | 0.297 |
| schulbil_4 | -.1066143 | 0.220 | -.0391485 | 0.797 | .2277794 | 0.008 |
| gesundheit_3 | .051909 | 0.498 | -.0757783 | 0.547 | .0519512 | 0.467 |
| gesundheit_4 | -.0000444 | 1.000 | .1316258 | 0.433 | .1421355 | 0.129 |
| gesundheit_5 | .221958 | 0.160 | -.2521271 | 0.214 | -.0532613 | 0.674 |

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

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

## 6.4 Propensity to participate – first-time interviewed split-off households

This step calculated the propensities to participate for new split-off households, i.e. households that are included in the panel due to the relocation of one individual of the panel sample in a new household. Here, only split-off households were considered that had not been interviewed in the previous waves. This means that the participation propensities for first-time participating split-off households were modelled separately following the criterion of split-off households originated in wave 3 (split-off W3 households) and split-off households originated in wave 4 (split-off W4 households). The probability of re-participation was estimated via logit models for availability and participation. Missing time-stable information on the household reference person (HRP) was added from the previous wave if necessary. The estimated propensities of the two models were multiplied. The reciprocal value of the product for the split-off households can also be found in the variable *hpbleib*.

**Table 44: Variable overview, codes and reference categories for the logit models of the split-off households participating for the first time (W3 and W4)**

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

**Table 45: Logit models on the first participation of split-off W3 households for availability and participation**

| | Contact | | Participation | |
|---|---|---|---|---|
| | **Coef.** | **p** | **Coef.** | **p** |
| alter_1 | -.2418126 | 0.716 | .5429442 | 0.340 |
| alter_2 | -.4146809 | 0.570 | -.1851265 | 0.791 |
| alter_4 | -.7330313 | 0.198 | -.0424319 | 0.937 |
| alter_5 | -1.770686 | 0.241 | 2.556612 | 0.055 |
| sex_1 | -.6319621 | 0.188 | .3655047 | 0.410 |
| nichtdeutsch | -.1078378 | 0.931 | 1.012818 | 0.281 |
| schulbil_1 | -.2794609 | 0.724 | -.3160496 | 0.716 |
| schulbil_2 | .47696 | 0.384 | -1.09537 | 0.060 |
| schulbil_4 | .5901841 | 0.384 | .7000542 | 0.189 |
| anzkon_1 | -1.314818 | 0.011 | | |
| anzkon_3 | 1.649364 | 0.145 | | |
| anzkon_4 | .1804791 | 0.838 | | |
| stichprobe1 | 1.034975 | 0.035 | -.1910267 | 0.702 |
| stichprobe3 | -.2036549 | 0.782 | -.6937177 | 0.457 |
| cons | 1.963171 | 0.004 | -1.323982 | 0.012 |
| **n** | 187 | | 158 | |
| **Log likelihood** | -67.599448 | | -75.077429 | |
| **Pseudo $R^2$** | 0.1621 | | 0.0875 | |

**Table 46: Logit models on the first participation of split-off W4 households for availability and participation**

| | Contact | | Participation | |
|---|---|---|---|---|
| | **Coef.** | **p** | **Coef.** | **p** |
| alter_1 | -.3103986 | 0.495 | .5612502 | 0.121 |
| alter_2 | -.0615905 | 0.911 | .2041636 | 0.602 |
| alter_4 | -.4011371 | 0.236 | .1618036 | 0.556 |
| alter_5 | 1.05171 | 0.220 | 1.097297 | 0.075 |
| sex_1 | -.3221642 | 0.275 | -.1990163 | 0.429 |
| nichtdeutsch | -.1466203 | 0.837 | .0084587 | 0.988 |
| schulbil_1 | -.5262951 | 0.463 | .9209869 | 0.072 |
| schulbil_2 | -.0384243 | 0.915 | .165165 | 0.548 |
| schulbil_4 | .2198667 | 0.553 | .3418422 | 0.283 |
| anzkon_1 | -.3701073 | 0.235 | | |
| anzkon_3 | 2.284895 | 0.000 | | |
| anzkon_4 | 2.754416 | 0.000 | | |
| stichprobe1 | 3.516057 | 0.000 | -.9831306 | 0.000 |
| stichprobe3 | .289362 | 0.500 | -.2668892 | 0.548 |
| stichprobe4 | .0596691 | 0.902 | -.3008521 | 0.563 |
| cons | .4192136 | 0.235 | -.347527 | 0.211 |
| **n** | 459 | | 368 | |
| **Log likelihood** | -153.64726 | | -226.55592 | |
| **Pseudo $R^2$** | 0.3278 | | 0.0590 | |

## 6.5 Non-response weighting for households from the wave 4 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 third wave. The participation probability derived from this can be found in variable *prop_t0*.

**Table 47: Variable overview, codes and reference categories for the logit models of the refreshment sample of wave 4**

| Variable code and reference category | Explanation |
|---|---|
| alter_1 | Household reference person (HRP) younger than 30 years |
| alter_2 | HRP 30 – 39 years old |
| alter_4 | HRP 50 – 64 years old |
| Reference category | HRP 40 – 49 years old |
| sex_1 | HRP male |
| Reference category | HRP female |
| nichtdeutsch | HRP has nationality other than German |
| Reference category | HRP has German nationality or missing information |
| anz_persBG_2 | Number of individuals in the benefit community: 2 individuals |
| anz_persBG_3 | Number of individuals in the benefit community: 3 and more individuals |
| Reference category | Number of individuals in the benefit community: 1 individual |
| anz_verwfBG_1 | Number of individuals capable of work in the benefit community: none |
| anz_verwfBG_3 | Number of individuals capable of work in the benefit community: 2 individuals |
| anz_verwfBG_4 | Number of individuals capable of work in the benefit community: 3 and more individuals |
| Reference category | Number of individuals capable of work in the benefit community: 1 individual |
| BG_typ_2 | Type of benefit community: single parent |
| BG_typ_3 | Type of benefit community: couple without children |
| BG_typ_4 | Type of benefit community: couple with children under the age of 18 |
| BG_typ_5 | Type of benefit community: other benefit community |
| Reference category | Type of benefit community: single |
| famstand_2 | Marital status: married |
| famstand_3 | Marital status: widowed |
| famstand_4 | Marital status: divorced |
| famstand_5 | Marital status: separated |
| famstand_6 | Marital status: relationship similar to a marriage or registered partnership |
| Reference category | Marital status: single |
| anzkon_1 | Number of contact attempts CATI/CAPI: 1 contact attempt |
| anzkon_3 | Number of contact attempts CATI/CAPI: 4-9 contact attempts |
| anzkon_4 | Number of contact attempts CATI/CAPI: 10 and more contact attempts |
| Reference category | Number of contact attempts CATI/CAPI: 2-3 contact attempts |
| blneualt_2 | New federal states |
| Reference category | Old federal states |

**Table 48: Logit models on the first participation for availability and participation of the wave 4 refreshment sample**

| | Contact | | Participation | |
|---|---|---|---|---|
| | **Coef.** | **p** | **Coef.** | **p** |
| alter_1 | -.2709841 | 0.450 | -.0748669 | 0.585 |
| alter_2 | .5520724 | 0.191 | .1400829 | 0.311 |
| alter_4 | .0117875 | 0.975 | .2277945 | 0.118 |
| sex_1 | -.1020647 | 0.701 | .0284883 | 0.776 |
| nichtdeutsch | -.0272298 | 0.936 | -.2228597 | 0.076 |
| anz_persBG_2 | -1.702769 | 0.063 | .3772027 | 0.264 |
| anz_persBG_3 | -1.524367 | 0.197 | .5694042 | 0.153 |
| anz_verwfBG_1 | -.9406263 | 0.436 | .2104376 | 0.678 |
| anz_verwfBG_3 | .1560236 | 0.821 | -.1769682 | 0.496 |
| anz_verwfBG_4 | -.0581056 | 0.958 | -.7962362 | 0.060 |
| BG_typ_2 | 1.785789 | 0.074 | -.3529742 | 0.328 |
| BG_typ_3 | 1.017784 | 0.247 | -.6112298 | 0.095 |
| BG_typ_4 | .8798209 | 0.436 | -.5367045 | 0.193 |
| BG_typ_5 | 1.407635 | 0.105 | -.0245079 | 0.931 |
| famstand_2 | .8228556 | 0.202 | .3322433 | 0.124 |
| famstand_3 | -1.126793 | 0.186 | .2749957 | 0.497 |
| famstand_4 | -.5357952 | 0.169 | -.0970361 | 0.538 |
| famstand_5 | .2135627 | 0.676 | -.0288342 | 0.863 |
| famstand_6 | -.3597392 | 0.578 | .5774257 | 0.045 |
| blneualt_2 | .4622306 | 0.120 | -.1443263 | 0.172 |
| anzkon_1 | -1.265884 | 0.008 | | |
| anzkon_3 | -.2950892 | 0.584 | | |
| anzkon_4 | -2.737432 | 0.000 | | |
| cons | 4.895443 | 0.000 | -.7310508 | 0.000 |
| **n** | 2440 | | 2359 | |
| **Log likelihood** | -297.79633 | | -1487.1327 | |
| **Pseudo $R^2$** | 0.1622 | | 0.0085 | |

## 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. The propensities to participate again for individuals in wave 4 were estimated including additional personal characteristics via logit models for willingness to participate in the panel, availability and participation. The dependence of the personal sample conveyed via the household context and correction of the estimation of standard mistakes 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 [$t_1$; $t_2$; $t_3$; $t_4$] across all four waves can be obtained as the product of the cross-sectional weight to $_1$, *ppbleib* (wave 1 to wave 2) and *ppbleib* (wave 2 to wave 3 etc.).

**Table 49: Variable overview, codes and reference categories for the logit models of re-participating individuals**

| Variable code and reference category | Explanation |
|---|---|
| alter_1 | Individual younger than 30 years |
| alter_2 | Individual 30 – 39 years old |
| alter_4 | Individual 50 – 64 years old |
| alter_5 | Individual 65 years and older |
| Reference category | Individual 40 – 49 years old |
| sex_1 | Individual male |
| Reference category | Individual female |
| nichtdeutsch | Individual has nationality other than German |
| Reference category | Individual has German nationality or missing information |
| sprache_nichtdeutsch | Language of individual: primarily spoken language in the household is not German |
| Reference category | Language of individual: primarily spoken language in the household is German |
| schulbil_1 | School qualification individual: no qualification |
| schulbil_2 | School qualification individual: lower secondary school |
| schulbil_4 | School qualification individual: college/university qualification |
| Reference category | School qualification individual: intermediate secondary school/still pupil |
| gesundheit_3 | Subjective evaluation of the health state of the individual: satisfactory |
| gesundheit_4 | Subjective evaluation of the health state of the individual: not so good |
| gesundheit_5 | Subjective evaluation of the health state of the individual: bad |
| Reference category | Subjective evaluation of the health state of the individual: very good to good |
| zufrieden_1 | General life satisfaction of the individual: scale value 0-2 |
| zufrieden_2 | General life satisfaction of the individual: scale value 3-5 |
| zufrieden_3 | General life satisfaction of the individual: scale value 6-8 |
| Reference category | General life satisfaction of the individual: scale value 9-10 |
| anz_0_3 | Number of individuals in the household aged 0 – 3 years |
| anz_4_6 | Number of individuals in the household aged 4 – 6 years |
| anz_7_14 | Number of individuals in the household aged 7 – 14 years |
| anz_65 | Number of individuals in the household aged 65 years and older |
| DinvalidAge | Age responses that cannot be evaluated |
| Reference category | Number of individuals in the household aged 15 – 64 years |
| eigentum | Type of residential property: proprietor |
| Reference category | Type of residential property: tenant, missing information |
| wnka_1 | Number of "don't know" and "details refused" responses in household and personal interviews of the individual: none |
| wnka_3 | Number of "don't know" and "details refused" responses in household and personal interviews of the individual: 11 and more |
| Reference category | Number of "don't know" and "details refused" responses in household and personal interviews of the individual: 1-10 |
| hhincome_1 | Household income: up to EUR 870 |
| hhincome_2 | Household income: EUR 871-1.400 |
| hhincome_4 | Household income: more than EUR 2,200 |
| Reference category | Household income: EUR 1,401-2,200 |
| alg2_1 | UB II receipt of the household: current receipt of UB II |
| Reference category | UB II receipt of the household: no current receipt of UB II |
| stichprobe1 | BA sample |
| stichprobe3 | Refreshment sample (BA) wave 2 |
| stichprobe4 | Refreshment sample (BA) wave 3 |
| Reference category | Microm sample |
| anzkon_1 | Number of contact attempts CATI/CAPI: 1 contact attempt |
| anzkon_3 | Number of contact attempts CATI/CAPI: 4-9 contact attempts |
| anzkon_4 | Number of contact attempts CATI/CAPI: 10 and more contact attempts |
| Reference category | Number of contact attempts CATI/CAPI: 2-3 contact attempts |

**Table 49: Variable overview, codes and reference categories for the logit models of the re-participants (continued)**

| Variable code and reference category | Explanation |
|---|---|
| blneualt_2 | New federal states |
| Reference category | Old federal states |
| bundesld_1 | Federal state: Schleswig-Holstein |
| bundesld_2 | Federal state: Hamburg |
| bundesld_3 | Federal state: Lower-Saxony |
| bundesld_4 | Federal state: Bremen |
| bundesld_6 | Federal state: Hesse |
| bundesld_7 | Federal state: Rhineland-Palatinate |
| bundesld_8 | Federal state: Baden-Wuerttemberg |
| bundesld_9 | Federal state: Bavaria |
| bundesld_10 | Federal state: Saarland |
| bundesld_11 | Federal state: Berlin |
| bundesld_12 | Federal state: Brandenburg |
| bundesld_13 | Federal state: Mecklenburg-Vorpommern |
| bundesld_14 | Federal state: Saxony |
| bundesld_15 | Federal state: Saxony-Anhalt |
| bundesld_16 | Federal state: Thuringia |
| Reference category | Federal state: North Rhine-Westphalia |
| bik_1 | BIK size class of municipality: population of less than 2,000 |
| bik_2 | BIK size class of municipality: population of 2,000 to under 5,000 |
| bik_3 | BIK size class of municipality: population of 5,000 to under 20,000 |
| bik_4 | BIK size class of municipality: population of 20.000 to under 50.000 |
| bik_5 | BIK size class of municipality: population of 50,000 to under 100,000 STYP 2/ 3/ 4 |
| bik_6 | BIK size class of municipality: population of 50,000 to under 100,000 STYP 1 |
| bik_7 | BIK size class of municipality: population of 100,000 to under 500,000 STYP 2/ 3/ 4 |
| bik_8 | BIK size class of municipality: population of 100,000 to under 500,000 STYP 1 |
| bik_9 | BIK size class of municipality: population of 500,000 and more STYP 2/ 3/ 4 |
| Reference category | BIK size class of municipality: population of 500,000 and more STYP 1 |

**Table 50: Logit models on re-participation for willingness to participate in a panel, availability and participation**

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

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

**Table 50: Logit models on re-participation for willingness to participate in a panel, availability and participation (continued)**

| | Willingness to participate in panel | | Contact | | Participation | |
|---|---|---|---|---|---|---|
| | Coef. | p | Coef. | p | Coef. | p |
| anz_0_3 | .0798781 | 0.530 | .0133438 | 0.930 | -.1653959 | 0.026 |
| anz_4_6 | -.0104995 | 0.939 | .3446568 | 0.120 | .0843735 | 0.362 |
| anz_7_14 | -.0078977 | 0.908 | -.0451142 | 0.655 | -.0053103 | 0.916 |
| anz_65 | -.246353 | 0.043 | .18392 | 0.531 | -.0071902 | 0.947 |
| DinvalidAge | -.1436989 | 0.557 | -.2284297 | 0.619 | -.2672323 | 0.323 |
| eigentum | -.6452986 | 0.000 | .8104465 | 0.000 | .131467 | 0.134 |
| wnka_1 | .1024658 | 0.152 | .0668878 | 0.585 | .1685842 | 0.005 |
| wnka_3 | -.1999418 | 0.021 | -.105062 | 0.420 | -.2232673 | 0.001 |
| hhincome_1 | -.0332209 | 0.807 | -.6980556 | 0.000 | .1603949 | 0.088 |
| hhincome_2 | -.0335545 | 0.758 | -.1987762 | 0.298 | .1456451 | 0.076 |
| hhincome_4 | .1431539 | 0.175 | .2071721 | 0.353 | .0296048 | 0.746 |
| alg2_1 | .1801189 | 0.107 | .1525633 | 0.283 | -.0604414 | 0.426 |
| stichprobe1 | .2394516 | 0.030 | .1340638 | 0.414 | -.2718389 | 0.002 |
| stichprobe3 | .7714802 | 0.000 | -.0910405 | 0.701 | -.1922442 | 0.145 |
| stichprobe4 | -.0358222 | 0.798 | -.1917936 | 0.316 | -.3642542 | 0.001 |
| blneualt_2 | -.1490614 | 0.078 | | | | |
| anzkon_1 | | | -.7274809 | 0.000 | | |
| anzkon_3 | | | .0244946 | 0.877 | | |
| anzkon_4 | | | -.580114 | 0.000 | | |
| bundesld_1 | | | -.4421672 | 0.142 | .0398692 | 0.802 |
| bundesld_2 | | | 1.253176 | 0.226 | .2258841 | 0.418 |
| bundesld_3 | | | -.4988578 | 0.013 | -.0247771 | 0.827 |
| bundesld_4 | | | .0021874 | 0.997 | -.313746 | 0.233 |
| bundesld_6 | | | .2739805 | 0.301 | -.4146181 | 0.001 |
| bundesld_7 | | | -.4197543 | 0.177 | -.238903 | 0.123 |
| bundesld_8 | | | .0759532 | 0.775 | .0657664 | 0.609 |
| bundesld_9 | | | -.2790896 | 0.185 | -.1254794 | 0.230 |
| bundesld_10 | | | .4008025 | 0.534 | -.5064959 | 0.022 |
| bundesld_11 | | | .4743736 | 0.141 | .0279104 | 0.849 |
| bundesld_12 | | | -.3197083 | 0.300 | .1375164 | 0.405 |
| bundesld_13 | | | .0319421 | 0.934 | .0353293 | 0.864 |
| bundesld_14 | | | -.1628645 | 0.532 | -.0117899 | 0.938 |
| bundesld_15 | | | -.0687428 | 0.807 | .05052 | 0.732 |
| bundesld_16 | | | .3727655 | 0.316 | .0765401 | 0.671 |
| bik_1 | | | .0821929 | 0.889 | .045135 | 0.879 |
| bik_2 | | | -.1168019 | 0.741 | .0455952 | 0.812 |
| bik_3 | | | .3013398 | 0.265 | -.018835 | 0.883 |
| bik_4 | | | -.0769723 | 0.752 | -.0342773 | 0.787 |
| bik_5 | | | -.0814823 | 0.744 | -.0441525 | 0.739 |
| bik_6 | | | 1.052581 | 0.014 | -.2114996 | 0.237 |
| bik_7 | | | .1470519 | 0.484 | -.2208845 | 0.045 |
| bik_8 | | | .036178 | 0.841 | -.0282934 | 0.776 |
| bik_9 | | | .6274816 | 0.046 | .0138276 | 0.917 |
| cons | 2.081648 | 0.000 | 3.541147 | 0.000 | 1.829531 | 0.000 |
| **n** | 13439 | | 11912 | | 11439 | |
| **Log likelihood** | -4573.7302 | | -1824.7061 | | -5479.8072 | |
| **Pseudo $R^2$** | 0.0387 | | 0.0828 | | 0.0512 | |

## 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 waves 2 and 3), which were modified by the non-response modelling. The multiple selection probability of a sampled benefit recipient who was living in the same household as benefit recipients in the previous years but without being a member of the benefit community himself/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, in 7/2007, in 7/2008 or in 7/2009. 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 7/2009. 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 4). 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, W3 and W4) in that sample were assumed. For cases from the Microm sample, if they were (according to survey data) new recipients of Unemployment Benefit II who first received the benefit between the last two sampling dates (W2; W3, W4), 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 4 as was possible in wave 3. No longitudinal weights are calculated for these households, i.e. (weighted) longitudinal evaluations can only be made with participants across all waves in question. Non-participation of a household can only occur in one wave; if a household skips two consecutive waves, it will no longer be contacted. In order to calculate mutual cross-sectional weights including the temporary non-responses, there was a convex combination of the modified household weights of the temporary non-responses and the modified household weights of the panel household sample (not of the refreshment sample) before calibration. The convex combination of the household weights was hence made before calibration; the calibration was then made with the new combined household weights.

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

For temporary non-responses the probability of non-participation in wave 3 in case of participation in wave 2 (non-participation propensities W3) and the probability of participation in wave 4 in case of a non-participation in wave 3 (participation propensities W4) was determined. The probability of non-participation in wave 3 is calculated from 1– participation probability in wave 3.

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

**Table 51: Variable overview, codes and reference categories for the logit models of the temporary non-responses**

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

**Table 51: Variable overview, codes and reference categories for the logit models of the temporary non-responses (continued)**

| Variable code and reference category | Explanation |
|---|---|
| wnka_1 | Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: none |
| wnka_3 | Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: 11 and more |
| Reference category | Number of "don't know" and "details refused" responses in household and personal interviews of the HRP: 1-10 |
| hhincome_1 | Household income: up to EUR 870 |
| hhincome_2 | Household income: EUR 871-1.400 |
| hhincome_4 | Household income: more than EUR 2,200 |
| Reference category | Household income: EUR 1,401-2,200 |
| alg2_1 | UB II receipt of the household: current receipt of UB II |
| Reference category | UB II receipt of the household: no current receipt of UB II |
| bundesld_1 | Federal state: Schleswig-Holstein |
| bundesld_2 | Federal state: Hamburg |
| bundesld_3 | Federal state: Lower-Saxony |
| bundesld_4 | Federal state: Bremen |
| bundesld_6 | Federal state: Hesse |
| bundesld_7 | Federal state: Rhineland-Palatinate |
| bundesld_8 | Federal state: Baden-Wuerttemberg |
| bundesld_9 | Federal state: Bavaria |
| bundesld_10 | Federal state: Saarland |
| bundesld_11 | Federal state: Berlin |
| bundesld_12 | Federal state: Brandenburg |
| bundesld_13 | Federal state: Mecklenburg-Vorpommern |
| bundesld_14 | Federal state: Saxony |
| bundesld_15 | Federal state: Saxony-Anhalt |
| bundesld_16 | Federal state: Thuringia |
| Reference category | Federal state: North Rhine-Westphalia |
| bik_1 | BIK size class of municipality: population of less than 2,000 |
| bik_2 | BIK size class of municipality: population of 2,000 to under 5,000 |
| bik_3 | BIK size class of municipality: population of 5,000 to under 20,000 |
| bik_4 | BIK size class of municipality: population of 20,000 to under 50,000 |
| bik_5 | BIK size class of municipality: population of 50,000 to under 100,000 STYP 2/ 3/ 4 |
| bik_6 | BIK size class of municipality: population of 50,000 to under 100,000 STYP 1 |
| bik_7 | BIK size class of municipality: population of 100,000 to under 500,000 STYP 2/ 3/ 4 |
| bik_8 | BIK size class of municipality: population of 100,000 to under 500,000 STYP 1 |
| bik_9 | BIK size class of municipality: population of 500,000 and more STYP 2/ 3/ 4 |
| Reference category | BIK size class of municipality: population of 500,000 and more STYP 1 |

**Table 52: Logit models of temporary non-responses**

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

**Table 52: Logit models of temporary non-responses (continued)**

| | Re-participation in wave 3 to determine the non-participation probability W3 (1-participation probability W3) | | Re-participation in wave 4 in case of non-participation in wave 3 | |
|---|---|---|---|---|
| | **Coef.** | **p** | **Coef.** | **p** |
| bik_1 | -.1718978 | 0.389 | -.314913 | 0.547 |
| bik_2 | -.1070126 | 0.456 | .2166599 | 0.529 |
| bik_3 | -.1862369 | 0.077 | -.6639829 | 0.029 |
| bik_4 | -.2260286 | 0.019 | -.9719687 | 0.001 |
| bik_5 | -.1548135 | 0.141 | -.3110231 | 0.242 |
| bik_6 | .0426827 | 0.785 | -.4441523 | 0.281 |
| bik_7 | .0438941 | 0.624 | -.3476504 | 0.138 |
| bik_8 | .0090798 | 0.911 | -.1945911 | 0.316 |
| bik_9 | -.0567797 | 0.601 | -.7660066 | 0.024 |
| cons | .3782731 | 0.003 | -1.92955 | 0.000 |
| **n** | 8429 | | 2933 | |
| **Log likelihood** | -5253.1301 | | -924.46904 | |
| **Pseudo $R^2$** | 0.0355 | | 0.0556 | |

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

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

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

## 6.9 Calibration to the household weight, fourth wave, cross-section

Following that came another calibration of the modified design weights including the non-response weighting at the household level using the GREG procedure to the benchmark values of the Federal Statistical Office for 2009. For households in receipt of benefits the weights were adjusted to the statistics of the Federal Employment Agency for July 2009. As in the previous year, also the increase in Unemployment Benefit II receipt since the previous year at the level of benefit communities (380,663) was also included as an additional benchmark value in the total sample. Those cases in the previous samples from wave 1, wave 2, and wave 3 which, according to wave 4 of the survey were receiving Unemployment Benefit II in July 2009, 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 to the weights if necessary. Within the framework of the calibration at hand, this was made in two points:

- The input weights for the calibration (the modified design weights after considering non-response analyses) were trimmed before calibration, i.e. they were replaced by new input weights. The maximum and minimum of the trimmed design weights was determined by using certain percentiles of the distribution depending on the distribution of the design weights.
- Also the interval of weights was limited during calibration, i.e. a maximum and a minimum limit for weights was determined. Here also the total width of the weights was determined; the range of the pure calibration weights can be calculated from the relation of original weights to the trimmed input weight. It had to be observed here that narrower limits for the weights result in less 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 cumulated BA sample of the first four 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, four drawing dates (in July 2006, July 2007, July 2008 or July 2009). In wave 4, only the benchmark values of the BA statistics from July 2009 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 2009. Starting point for the calibration were modified design weights including the non-response weighting. The modified design weights were trimmed at the 5 % percentile and the 95 % percentile of their distribution and after that rescaled in such a way that their total again resulted in the total of the untrimmed design weights. The projection factors of the trimmed design weights range from 313,86 to 2630,18. The relation between the total projection fac-

tors after calibration and the trimmed design weights was limited downwards to 0.3 and upwards to 2.0. Thus, the total projection factors after calibration lie between min. 94.15834 and max. 4165.227.

A calibration was made for the following characteristics:

Benefit communities basis BA statistics:

- Increase in BC Unemployment Benefit II recipients
- Number BCs receiving benefits in accordance with SGB II by federal states
- Number of BCs receiving benefits in accordance with SGB II by number of individuals under 65 years of age in the benefit 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

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 (380,663).

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

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

**Table 53: Nominal distributions and distributions after calibration (BA sample, households)**

| Benchmark figure | Characteristics benchmark figure from BA statistics | Unweighted distribution | Nominal values from BA statistics | Distribution with calibrated weights |
|---|---|---|---|---|
| Number BCs receiving benefits in accordance with SGB II by federal states (16 categories) | Number BCs Schleswig-Holstein | 135 | 121,765 | 121,765 |
| | Number BCs Hamburg | 79 | 107,774 | 107,774 |
| | Number BCs Lower-Saxony | 325 | 331,822 | 331,822 |
| | Number BCs Bremen | 36 | 50,549 | 50,549 |
| | Number BCs North Rhine-Westphalia | 837 | 823,814 | 823,814 |
| | Number BCs Hesse | 203 | 218,820 | 218,820 |
| | Number BCs Rhineland-Palatinate | 117 | 122,169 | 122,169 |
| | Number BCs Baden-Wuerttemberg | 272 | 246,337 | 246,337 |
| | Number BCs Bavaria | 293 | 265,707 | 265,707 |
| | Number BCs Saarland | 52 | 44,384 | 44,384 |
| | Number BCs Berlin | 271 | 330,193 | 330,193 |
| | Number BCs Brandenburg | 179 | 172,677 | 172,677 |
| | Number BCs Mecklenburg-Vorpommern | 105 | 131,213 | 131,213 |
| | Number BCs Saxony | 252 | 287,711 | 287,711 |
| | Number BCs Saxony-Anhalt | 245 | 191,791 | 191,791 |
| | Number BCs Thuringia | 128 | 134,303 | 134,303 |
| 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) | 981 | 1,240,872 | 1,240,872 |
| | Number BCs with 2 individuals under 65 (west) | 595 | 480,659 | 480,659 |
| | Number BCs with 3 individuals under 65 (west) | 427 | 300,506 | 300,506 |
| | Number BCs with 4 individuals under 65 (west) | 203 | 184,448 | 184,448 |
| | Number BCs with 5 or more individuals under 65 (west) | 143 | 126,656 | 126,656 |
| | Number BCs with 1 individual under 65 (east) | 554 | 714,495 | 714,495 |
| | Number BCs with 2 individuals under 65 (east) | 309 | 276,254 | 276,254 |
| | Number BCs with 3 individuals under 65 (east) | 180 | 143,684 | 143,684 |
| | Number BCs with 4 individuals under 65 (east) | 90 | 74,190 | 74,190 |
| | Number BCs with 5 or more individuals under 65 (east) | 47 | 39,265 | 39,265 |

**Table 53: Nominal distributions and distributions after calibration (BA sample, households) (continued)**

| Benchmark figure | Characteristics benchmark figure from BA statistics | Unweighted distribution | Nominal values from BA statistics | Distribution with calibrated weights |
|---|---|---|---|---|
| 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) | 1554 | 1,585,891 | 1,585,891 |
| | Number BCs with 1 child under 15 years of age (west) | 466 | 407,636 | 407,636 |
| | Number BCs with 2 children under 15 years of age (west) | 223 | 229,110 | 229,110 |
| | Number BCs with 3 children under 15 years of age (west) | 68 | 79,014 | 79,014 |
| | Number BCs with 4 or more children under 15 years of age (west) | 38 | 31,490 | 31,490 |
| | Number BCs without children under 15 years of age (east) | 886 | 926,414 | 926,414 |
| | Number BCs with 1 child under 15 years of age (east) | 187 | 195,946 | 195,946 |
| | Number BCs with 2 children under 15 years of age (east) | 84 | 90,169 | 90,169 |
| | Number BCs with 3 children under 15 years of age (east) | 15 | 25,468 | 25,468 |
| | Number BCs with 4 or more children under 15 years of age (east) | 8 | 9,891 | 9,891 |
| Number BCs receiving benefits in accordance with SGB II consisting of a single parent with children by west/east (4 categories) | Number BCs with a single parent (west) | 592 | 452,688 | 452,688 |
| | Rest BCs without a single parent (west) | 1757 | 1,880,453 | 1,880,453 |
| | Number BCs with a single parent (east) | 202 | 192,016 | 192,016 |
| | Rest BCs without a single parent (east) | 978 | 1,055,872 | 1,055,872 |

**Table 54: Parameters of distribution of weights**

| | |
|---|---|
| 1% percentile | 193.7486 |
| 5% percentile | 267.6218 |
| 10% percentile | 331.4976 |
| 25% percentile | 470.8878 |
| 50% percentile | 734.9039 |
| 75% percentile | 1433.385 |
| 90% percentile | 2257.767 |
| 95% percentile | 2667.289 |
| 99% percentile | 2921.039 |
| Average value | 1030.394 |
| Standard deviation | 736.9397 |
| Minimum | 94.15834 |
| Maximum | 4165.227 |
| Case number | 3404 |
| **Efficiency measure** | **65.65%** |

### 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 again resulted in the total of the untrimmed design weights. The projection factors of the trimmed design weights range from 2394,48 to 35018,87. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.1 and upwards to 3.0. Thus, the total projection factors after calibration lie between min. 239.4479 and max. 56098.93.

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 Mikrozensus 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 modal value of the respective variable.

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

**Table 55: Nominal distributions and distributions after calibration (Microm sample, households)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number BCs receiving benefits in accordance with SGB II by federal states (16 categories) | Number BCs Schleswig-Holstein | 16 | 121,765 | 121,765 |
| | Number BCs Hamburg | 2 | 107,774 | 107,774 |
| | Number BCs Lower-Saxony | 43 | 331,822 | 331,778 |
| | Number BCs Bremen | 6 | 50,549 | 50,549 |
| | Number BCs North Rhine-Westphalia | 75 | 823,814 | 823,858 |
| | Number BCs Hesse | 18 | 218,820 | 218,820 |
| | Number BCs Rhineland-Palatinate | 13 | 122,169 | 122,169 |
| | Number BCs Baden-Wuerttemberg | 14 | 246,337 | 246,337 |
| | Number BCs Bavaria | 40 | 265,707 | 265,707 |
| | Number BCs Saarland | 10 | 44,384 | 44,384 |
| | Number BCs Berlin | 18 | 330,193 | 330,193 |
| | Number BCs Brandenburg | 17 | 172,677 | 172,677 |
| | Number BCs Mecklenburg-Vorpommern | 6 | 131,213 | 131,213 |
| | Number BCs Saxony | 23 | 287,711 | 287,711 |
| | Number BCs Saxony-Anhalt | 21 | 191,791 | 191,791 |
| | Number BCs Thuringia | 15 | 134,303 | 134,303 |

**Table 55: Nominal distributions and distributions after calibration (Microm sample, households) (continued 1)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| 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) | 82 | 1,240,872 | 1,240,872 |
| | Number BCs with 2 individuals under 65 (west) | 62 | 480,659 | 480,658 |
| | Number BCs with 3 individuals under 65 (west) | 32 | 300,506 | 300,506 |
| | Number BCs with 4 individuals under 65 (west) | 31 | 184,448 | 184,488 |
| | Number BCs with 5 or more individuals under 65 (west) | 30 | 126,656 | 126,617 |
| | Number BCs with 1 individual under 65 (east) | 31 | 714,495 | 714,495 |
| | Number BCs with 2 individuals under 65 (east) | 29 | 276,254 | 276,254 |
| | Number BCs with 3 individuals under 65 (east) | 24 | 143,684 | 143,684 |
| | Number BCs with 4 individuals under 65 (east) | 10 | 74,190 | 74,190 |
| | Number BCs with 5 or more individuals under 65 (east) | 6 | 39,265 | 39,265 |
| 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) | 151 | 1,585,891 | 1,585,891 |
| | Number BCs with 1 child under 15 years of age (west) | 29 | 407,636 | 407,636 |
| | Number BCs with 2 children under 15 years of age (west) | 38 | 229,110 | 229,110 |
| | Number BCs with 3 children under 15 years of age (west) | 10 | 79,014 | 79,014 |
| | Number BCs with 4 or more children under 15 years of age (west) | 9 | 31,490 | 31,490 |
| | Number BCs without children under 15 years of age (east) | 74 | 926,414 | 926,414 |
| | Number BCs with 1 child under 15 years of age (east) | 12 | 195,946 | 195,946 |
| | Number BCs with 2 children under 15 years of age (east) | 6 | 90,169 | 90,169 |
| | Number BCs with 3 children under 15 years of age (east) | 7 | 25,468 | 25,468 |
| | Number BCs with 4 or more children under 15 years of age (east) | 1 | 9,891 | 9,891 |

**Table 55: Nominal distributions and distributions after calibration (Microm sample, households) (continued 2)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number BCs receiving benefits in accordance with SGB II consisting of a single parent with children by west/east (4 categories) | Number BCs with a single parent (west) | 48 | 452,688 | 452,688 |
| | Rest BCs without a single parent (west) | 189 | 1,880,453 | 1,880,453 |
| | Number BCs with a single parent (east) | 16 | 192,016 | 192,016 |
| | Rest BCs without a single parent (east) | 84 | 1,055,872 | 1,055,872 |
| Number of households by federal state and BIK type (spelling: "Federal state.BIK type") | 1,1 to 1,4 | 17 | 316,000 | 316,000 |
| | 1,5 to 1,6 | 24 | 149,000 | 149,000 |
| | 1,7 to 1,8 | 28 | 517,000 | 517,000 |
| | 1.9 | 18 | 172,000 | 172,000 |
| | 1.10 | 29 | 211,000 | 211,000 |
| | 2.10 | 40 | 969,000 | 969,000 |
| | 3,2 to 3,3 | 39 | 489,000 | 489,000 |
| | 3.4 | 25 | 434,000 | 434,000 |
| | 3.5 | 25 | 451,000 | 451,000 |
| | 3.7 | 101 | 857,000 | 857,000 |
| | 3.8 | 59 | 548,000 | 548,000 |
| | 3.9 | 54 | 636,000 | 636,000 |
| | 3.10 | 34 | 376,000 | 376,000 |
| | 4.8 to 4.10 | 22 | 358,000 | 358,000 |
| | 5,2 to 5,3 | 38 | 349,000 | 349,000 |
| | 5.4 | 57 | 1,052,000 | 1,052,000 |
| | 5.5 | 71 | 631,000 | 631,000 |
| | 5.6 | 25 | 301,000 | 301,000 |
| | 5.7 | 72 | 836,000 | 836,000 |
| | 5.8 | 156 | 2,333,000 | 2,333,000 |
| | 5.9 | 37 | 363,000 | 363,000 |
| | 5.10 | 183 | 2,659,000 | 2,659,000 |
| | 6.2 | 16 | 64,000 | 64,000 |
| | 6.3 | 33 | 330,000 | 330,000 |
| | 6,4 to 6,5 | 20 | 304,000 | 304,000 |
| | 6.7 | 40 | 564,000 | 564,000 |
| | 6.8 | 25 | 465,000 | 465,000 |
| | 6.9 | 50 | 366,000 | 366,000 |
| | 6.10 | 46 | 784,000 | 784,000 |
| | 7,1 to 7,2 | 23 | 305,000 | 305,000 |
| | 7.3 | 20 | 177,000 | 177,000 |
| | 7.4 | 14 | 154,000 | 154,000 |
| | 7,5 to 7,6 | 24 | 240,000 | 240,000 |
| | 7.7 | 34 | 388,000 | 388,000 |
| | 7.8 | 13 | 329,000 | 329,000 |
| | 7.9 to 7.10 | 11 | 269,000 | 269,000 |
| | 8,2 to 8,3 | 37 | 639,000 | 639,000 |
| | 8.4 | 61 | 518,000 | 518,000 |
| | 8,5 to 8,6 | 18 | 464,000 | 464,000 |

**Table 55: Nominal distributions and distributions after calibration (Microm sample, households) (continued 3)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of households by federal state and BIK type (spelling: "Federal state.BIK type") | 8.7 | 51 | 1,034,000 | 1,034,000 |
| | 8.8 | 63 | 825,000 | 825,000 |
| | 8.9 | 17 | 369,000 | 369,000 |
| | 8.10 | 83 | 1,059,000 | 1,059,000 |
| | 9,1 to 9,2 | 12 | 364,000 | 364,000 |
| | 9.3 | 62 | 502,000 | 502,000 |
| | 9.4 | 62 | 693,000 | 693,000 |
| | 9.5 | 34 | 372,000 | 372,000 |
| | 9,6 to 9,7 | 82 | 1,080,000 | 1,080,000 |
| | 9.8 | 44 | 634,000 | 634,000 |
| | 9.9 | 83 | 720,000 | 720,000 |
| | 9.10 | 120 | 1,485,000 | 1,485,000 |
| | 10,3 to 10,4 | 17 | 155,000 | 155,000 |
| | 10,7 to 10,8 | 19 | 325,000 | 325,000 |
| | 11.10 | 95 | 1,971,000 | 1,971,000 |
| | 12,1 to 12,3 | 11 | 277,000 | 277,000 |
| | 12.4 | 21 | 223,000 | 223,000 |
| | 12,5 to 12,6 | 14 | 136,000 | 136,000 |
| | 12.7 | 15 | 126,000 | 126,000 |
| | 12.8 | 19 | 147,000 | 147,000 |
| | 12.9 to 12.10 | 21 | 327,000 | 327,000 |
| | 13,1 to 13,4 | 16 | 363,000 | 363,000 |
| | 13,5 to 13,6 | 12 | 164,000 | 164,000 |
| | 13.7 | 11 | 102,000 | 102,000 |
| | 13.8 | 14 | 221,000 | 221,000 |
| | 14.1 | 13 | 21,000 | 21,000 |
| | 14.2 | 34 | 143,000 | 143,000 |
| | 14,3 to 14,4 | 17 | 442,000 | 442,000 |
| | 14.5 | 12 | 171,000 | 171,000 |
| | 14.6 | 15 | 133,000 | 133,000 |
| | 14,7 to 14,8 | 9 | 378,000 | 378,000 |
| | 14.9 to 14.10 | 45 | 893,000 | 893,000 |
| | 15,1 to 15,2 | 16 | 138,000 | 138,000 |
| | 15,3 to 15,4 | 16 | 231,000 | 231,000 |
| | 15,5 to 15,6 | 14 | 309,000 | 309,000 |
| | 15.7 | 25 | 241,000 | 241,000 |
| | 15.8 | 31 | 276,000 | 276,000 |
| | 16,1 to 16,2 | 9 | 141,000 | 141,000 |
| | 16,3 to 16,4 | 25 | 371,000 | 371,000 |
| | 16.5 | 22 | 238,000 | 238,000 |
| | 16.6 | 14 | 56,000 | 56,000 |
| | 16.7 to 16.8 | 23 | 312,000 | 312,000 |

**Table 55: Nominal distributions and distributions after calibration (Microm sample, households) (continued 4)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of households by household size (1,2,3,4,“5 and more individuals“) and west/east (10 categories) | Number households with 1 individual (west) | 608 | 11,877,000 | 11,877,000 |
| | Number households with 2 individuals (west) | 868 | 10,551,000 | 10,551,000 |
| | Number households with 3 individuals (west) | 382 | 3,977,000 | 3,977,000 |
| | Number households with 4 individuals (west) | 357 | 3,305,000 | 3,305,000 |
| | Number households with 5 or more individuals (west) | 173 | 1,274,000 | 1,274,000 |
| | Number households with 1 individual (east) | 192 | 3,627,000 | 3,627,000 |
| | Number households with 2 individuals (east) | 223 | 3,064,000 | 3,064,000 |
| | Number households with 3 individuals (east) | 99 | 1,140,000 | 1,140,000 |
| | Number households with 4 individuals (east) | 56 | 571,000 | 571,000 |
| | Number households with 5 and more individuals (east) | 19 | 149,000 | 149,000 |
| Number of households by "children under 15 years of age in the household yes/no" and west/east | Number households with children under 15 (west) | 622 | 5,704,000 | 5,704,000 |
| | Number households without children under 15 (west) | 1766 | 25,280,000 | 25,280,000 |
| | Number households with children under 15 (east) | 105 | 1,217,000 | 1,217,000 |
| | Number households without children under 15 (east) | 484 | 7,334,000 | 7,334,000 |

**Table 56: Parameters of distribution of weights**

| | |
|---|---|
| 1% percentile | 1212.23 |
| 5% percentile | 2331.255 |
| 10% percentile | 3064.392 |
| 25% percentile | 5468.877 |
| 50% percentile | 10128.85 |
| 75% percentile | 18479.22 |
| 90% percentile | 28758.1 |
| 95% percentile | 34701.2 |
| 99% percentile | 44240.42 |
| Average value | 13280.15 |
| Standard deviation | 10250.35 |
| Minimum | 239.4479 |
| Maximum | 56098.93 |
| Case number | 2977 |
| **Efficiency measure** | **62.67%** |

### 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 again resulted in the total of the untrimmed design weights. The projection factors of the trimmed design weights range from 176.35 to 25479.46. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.1 and upwards to 5.0. Thus, the total projection factors after calibration lie between min. 18.79 and max. 43378.58.

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 Mikrozensus 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 (380,663) was included as an additional benchmark value in the total sample.

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

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

**Table 57: Nominal distributions and distributions after calibration (total sample, households)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number BCs receiving benefits in accordance with SGB II by federal states (16 categories) | Number BCs Schleswig-Holstein | 151 | 121,765 | 121,767 |
| | Number BCs Hamburg | 81 | 107,774 | 107,781 |
| | Number BCs Lower-Saxony | 368 | 331,822 | 331,808 |
| | Number BCs Bremen | 42 | 50,549 | 50,549 |
| | Number BCs North Rhine-Westphalia | 912 | 823,814 | 823,659 |
| | Number BCs Hesse | 221 | 218,820 | 218,743 |
| | Number BCs Rhineland-Palatinate | 130 | 122,169 | 122,156 |
| | Number BCs Baden-Wuerttemberg | 286 | 246,337 | 246,286 |
| | Number BCs Bavaria | 333 | 265,707 | 266,183 |
| | Number BCs Saarland | 62 | 44,384 | 44,384 |
| | Number BCs Berlin | 289 | 330,193 | 330,123 |
| | Number BCs Brandenburg | 196 | 172,677 | 172,671 |
| | Number BCs Mecklenburg-Vorpommern | 111 | 131,213 | 131,231 |
| | Number BCs Saxony | 275 | 287,711 | 287,601 |
| | Number BCs Saxony-Anhalt | 266 | 191,791 | 191,768 |
| | Number BCs Thuringia | 143 | 134,303 | 134,320 |

**Table 57: Nominal distributions and distributions after calibration (total sample, households) (continued 1)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| 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) | 1063 | 1,240,872 | 1,240,874 |
| | Number BCs with 2 individuals under 65 (west) | 657 | 480,659 | 480,666 |
| | Number BCs with 3 individuals under 65 (west) | 459 | 300,506 | 300,388 |
| | Number BCs with 4 individuals under 65 (west) | 234 | 184,448 | 184,409 |
| | Number BCs with 5 or more individuals under 65 (west) | 173 | 126,656 | 126,980 |
| | Number BCs with 1 individual under 65 (east) | 585 | 714,495 | 714,495 |
| | Number BCs with 2 individuals under 65 (east) | 338 | 276,254 | 276,239 |
| | Number BCs with 3 individuals under 65 (east) | 204 | 143,684 | 143,688 |
| | Number BCs with 4 individuals under 65 (east) | 100 | 74,190 | 74,114 |
| | Number BCs with 5 or more individuals under 65 (east) | 53 | 39,265 | 39,176 |
| 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) | 1705 | 1,585,891 | 1,586,151 |
| | Number BCs with 1 child under 15 years of age (west) | 495 | 407,636 | 407,605 |
| | Number BCs with 2 children under 15 years of age (west) | 261 | 229,110 | 229,119 |
| | Number BCs with 3 children under 15 years of age (west) | 78 | 79,014 | 79,014 |
| | Number BCs with 4 or more children under 15 years of age (west) | 47 | 31,490 | 31,428 |
| | Number BCs without children under 15 years of age (east) | 960 | 926,414 | 926,318 |
| | Number BCs with 1 child under 15 years of age (east) | 199 | 195,946 | 195,886 |
| | Number BCs with 2 children under 15 years of age (east) | 90 | 90,169 | 90,149 |
| | Number BCs with 3 children under 15 years of age (east) | 22 | 25,468 | 25,468 |
| | Number BCs with 4 or more children under 15 years of age (east) | 9 | 9,891 | 9,891 |

**Table 57: Nominal distributions and distributions after calibration (total sample, households) (continued 2)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number BCs receiving benefits in accordance with SGB II consisting of a single parent with children by west/east (4 categories) | Number BCs with a single parent (west) | 640 | 452,688 | 452,612 |
| | Rest BCs without a single parent (west) | 1946 | 1,880,453 | 1,880,705 |
| | Number BCs with a single parent (east) | 218 | 192,016 | 192,033 |
| | Rest BCs without a single parent (east) | 1062 | 1,055,872 | 1,055,679 |
| Number of households by federal state and BIK type (spelling: "Federal state.BIK type") | 1.1 to 1.4 | 66 | 316,000 | 316,000 |
| | 1.5 to 1.6 | 55 | 149,000 | 149,000 |
| | 1.7 to 1.8 | 108 | 517,000 | 517,000 |
| | 1.9 | 41 | 172,000 | 172,000 |
| | 1.10 | 63 | 211,000 | 211,000 |
| | 2.10 | 147 | 969,000 | 969,000 |
| | 3.2 to 3.3 | 82 | 489,000 | 489,000 |
| | 3.4 | 63 | 434,000 | 434,000 |
| | 3.5 | 42 | 451,000 | 451,000 |
| | 3.7 | 225 | 857,000 | 857,000 |
| | 3.8 | 186 | 548,000 | 548,000 |
| | 3.9 | 101 | 636,000 | 636,000 |
| | 3.10 | 87 | 376,000 | 376,000 |
| | 4.8 to 4.10 | 71 | 358,000 | 358,000 |
| | 5.2 to 5.3 | 89 | 349,000 | 349,000 |
| | 5.4 | 140 | 1,052,000 | 1,052,000 |
| | 5.5 | 183 | 631,000 | 631,000 |
| | 5.6 | 60 | 301,000 | 301,000 |
| | 5.7 | 180 | 836,000 | 836,000 |
| | 5.8 | 461 | 2,333,000 | 2,333,000 |
| | 5.9 | 74 | 363,000 | 363,000 |
| | 5.10 | 545 | 2,659,000 | 2,659,000 |
| | 6.2 | 26 | 64,000 | 64,000 |
| | 6.3 | 69 | 330,000 | 330,000 |
| | 6.4 to 6.5 | 52 | 304,000 | 304,000 |
| | 6.7 | 78 | 564,000 | 564,000 |
| | 6.8 | 87 | 465,000 | 465,000 |
| | 6.9 | 79 | 366,000 | 366,000 |
| | 6.10 | 114 | 784,000 | 784,000 |
| | 7.1 to 7.2 | 36 | 305,000 | 305,000 |
| | 7.3 | 28 | 177,000 | 177,000 |
| | 7.4 | 23 | 154,000 | 154,000 |
| | 7.5 to 7.6 | 78 | 240,000 | 240,000 |
| | 7.7 | 50 | 388,000 | 388,000 |
| | 7.8 | 40 | 329,000 | 329,000 |
| | 7.9 to 7.10 | 41 | 269,000 | 269,000 |
| | 8.2 to 8.3 | 64 | 639,000 | 639,000 |
| | 8.4 | 107 | 518,000 | 518,000 |
| | 8.5 to 8.6 | 27 | 464,000 | 464,000 |

**Table 57: Nominal distributions and distributions after calibration (total sample, households) (continued 3)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of households by federal state and BIK type (spelling: "Federal state.BIK type") | 8.7 | 103 | 1,034,000 | 1,034,000 |
| | 8.8 | 145 | 825,000 | 825,000 |
| | 8.9 | 42 | 369,000 | 369,000 |
| | 8.10 | 212 | 1,059,000 | 1,059,000 |
| | 9.1 to 9.2 | 24 | 364,000 | 364,000 |
| | 9.3 | 93 | 502,000 | 502,000 |
| | 9.4 | 95 | 693,000 | 693,000 |
| | 9.5 | 53 | 372,000 | 372,000 |
| | 9.6 to 9.7 | 137 | 1,080,000 | 1,080,000 |
| | 9.8 | 136 | 634,000 | 634,000 |
| | 9.9 | 113 | 720,000 | 720,000 |
| | 9.10 | 271 | 1,485,000 | 1,485,000 |
| | 10.3 to 10.4 | 42 | 155,000 | 155,000 |
| | 10.7 to 10.8 | 61 | 325,000 | 325,000 |
| | 11.10 | 477 | 1,971,000 | 1,971,000 |
| | 12.1 to 12.3 | 79 | 277,000 | 277,000 |
| | 12.4 | 61 | 223,000 | 223,000 |
| | 12.5 to 12.6 | 62 | 136,000 | 136,000 |
| | 12.7 | 49 | 126,000 | 126,000 |
| | 12.8 | 39 | 147,000 | 147,000 |
| | 12.9 to 12.10 | 56 | 327,000 | 327,000 |
| | 13.1 to 13.4 | 87 | 363,000 | 363,000 |
| | 13.5 to 13.6 | 48 | 164,000 | 164,000 |
| | 13.7 | 32 | 102,000 | 102,000 |
| | 13.8 | 42 | 221,000 | 221,000 |
| | 14.1 | 18 | 21,000 | 21,000 |
| | 14.2 | 104 | 143,000 | 143,000 |
| | 14.3 to 14.4 | 53 | 442,000 | 442,000 |
| | 14.5 | 40 | 171,000 | 171,000 |
| | 14.6 | 96 | 133,000 | 133,000 |
| | 14.7 to 14.8 | 37 | 378,000 | 378,000 |
| | 14.9 to 14.10 | 158 | 893,000 | 893,000 |
| | 15.1 to 15.2 | 60 | 138,000 | 138,000 |
| | 15.3 to 15.4 | 49 | 231,000 | 231,000 |
| | 15.5 to 15.6 | 59 | 309,000 | 309,000 |
| | 15.7 | 126 | 241,000 | 241,000 |
| | 15.8 | 127 | 276,000 | 276,000 |
| | 16.1 to 16.2 | 37 | 141,000 | 141,000 |
| | 16.3 to 16.4 | 80 | 371,000 | 371,000 |
| | 16.5 | 55 | 238,000 | 238,000 |
| | 16.6 | 46 | 56,000 | 56,000 |
| | 16.7 to 16.8 | 76 | 312,000 | 312,000 |

**Table 57: Nominal distributions and distributions after calibration (total sample, households) (continued 4)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of households by household size (1,2,3,4,“5 and more individuals“) and west/east (10 categories) | Number households with 1 individual (west) | 1853 | 11,877,000 | 11,877,000 |
| | Number households with 2 individuals (west) | 1741 | 10,551,000 | 10,551,000 |
| | Number households with 3 individuals (west) | 969 | 3,977,000 | 3,977,000 |
| | Number households with 4 individuals (west) | 662 | 3,305,000 | 3,305,000 |
| | Number households with 5 or more individuals (west) | 370 | 1,274,000 | 1,274,000 |
| | Number households with 1 individual (east) | 914 | 3,627,000 | 3,627,000 |
| | Number households with 2 individuals (east) | 687 | 3,064,000 | 3,064,000 |
| | Number households with 3 individuals (east) | 372 | 1,140,000 | 1,140,000 |
| | Number households with 4 individuals (east) | 198 | 571,000 | 571,000 |
| | Number households with 5 and more individuals (east) | 82 | 149,000 | 149,000 |
| Number of households by "children under 15 years of age in the household yes/no" and west/east | Number households with children under 15 (west) | 1691 | 5,704,000 | 5,704,000 |
| | Number households without children under 15 (west) | 3904 | 25,280,000 | 25,280,000 |
| | Number households with children under 15 (east) | 532 | 1,217,000 | 1,217,000 |
| | Number households without children under 15 (east) | 1721 | 7,334,000 | 7,334,000 |

**Table 58: Parameters of distribution of weights**

| | |
|---|---|
| 1% percentile | 52.711 |
| 5% percentile | 113.8978 |
| 10% percentile | 167.971 |
| 25% percentile | 368.6647 |
| 50% percentile | 959.6308 |
| 75% percentile | 6079.649 |
| 90% percentile | 18154 |
| 95% percentile | 24881.24 |
| 99% percentile | 30300.4 |
| Average value | 5037.589 |
| Standard deviation | 7894.869 |
| Minimum | 18.79417 |
| Maximum | 43378.58 |
| Case number | 7848 |
| **Efficiency measure** | **28.93%** |

## 6.10 Calibration to the person weight, fourth wave, cross-section

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

As in the previous year, also the increase in Unemployment Benefit II recipients since the previous year at the level of individuals between 15 and 64 years (499,434) was also included as an additional benchmark value in the total sample. Again, those cases in the previous samples from wave 1, wave 2, and wave 3 which, according to wave 4 of the survey were receiving Unemployment Benefit II in July 2009 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 all four 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, up to now, four drawing dates (in July 2006, July 2007, July 2008 or July 2009). 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 again resulted in the total of the untrimmed calibrated household weights. The trimmed projection factors range from 392,22 to 4011,3. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.5 and upwards to 2.0. Thus, the total projection factors after calibration lie between min. 196.11 and max. 7392.75.

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 sex, 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 (499,434) was also included as an additional benchmark value in the total sample.

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

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

**Table 59: Nominal distributions and distributions after calibration (BA sample, individuals)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Un-weighted distribution | Nominal values | Distribu-tion 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 | 176 | 170,449 | 170,449 |
| | Number individuals in BCs Hamburg | 93 | 145,876 | 145,876 |
| | Number individuals in BCs Lower-Saxony | 417 | 473,772 | 473,772 |
| | Number individuals in BCs Bremen | 51 | 69,246 | 69,246 |
| | Number individuals in BCs North Rhine-Westphalia | 1,114 | 1,188,016 | 1,188,016 |
| | Number individuals in BCs Hesse | 251 | 316,524 | 316,524 |
| | Number individuals in BCs Rhineland-Palatinate | 145 | 175,232 | 175,232 |
| | Number individuals in BCs Baden-Wuerttemberg | 318 | 343,728 | 343,728 |
| | Number individuals in BCs Bavaria | 373 | 359,156 | 359,156 |
| | Number individuals in BCs Saarland | 64 | 61,900 | 61,900 |
| | Number individuals in BCs Berlin | 344 | 447,970 | 447,970 |
| | Number individuals in BCs Brandenburg | 238 | 237,387 | 237,387 |
| | Number individuals in BCs Mecklenburg-Vorpommern | 127 | 180,515 | 180,515 |
| | Number individuals in BCs Saxony | 331 | 396,110 | 396,110 |
| | Number individuals in BCs Saxony-Anhalt | 331 | 268,792 | 268,792 |
| | Number individuals in BCs Thuringia | 163 | 184,513 | 184,513 |
| 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 | 763 | 971,951 | 971,951 |
| | Number individuals in BCs aged 25-64 | 3,773 | 4,047,235 | 4,047,235 |
| Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by sex and west/east (4 categories) | Number men in BCs (west) | 1,348 | 1,608,020 | 1,608,020 |
| | Number women in BCs (west) | 1,654 | 1,695,879 | 1,695,879 |
| | Number men in BCs (east) | 746 | 879,978 | 879,978 |
| | Number women in BCs (east) | 788 | 835,309 | 835,309 |
| Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by "single parent yes/no", sex and west/east (8 categories) | Number non single parents in BCs (west) | 2,416 | 2,851,211 | 2,851,211 |
| | Number single parents in BCs (west) | 586 | 452,688 | 452,688 |
| | Number non single parents in BCs (east) | 1,327 | 1,523,271 | 1,523,271 |
| | Number single parents in BCs (east) | 207 | 192,016 | 192,016 |
| Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by nationality (German/non-German) | Number non-German individuals in BCs | 516 | 968,390 | 968,390 |
| | Number German individuals in BCs | 4,020 | 4,050,796 | 4,050,796 |

**Table 60: Parameters of distribution of weights**

| | |
|---|---|
| 1% percentile | 205.9219 |
| 5% percentile | 258.8751 |
| 10% percentile | 307.8417 |
| 25% percentile | 465.0251 |
| 50% percentile | 771.7853 |
| 75% percentile | 1472.96 |
| 90% percentile | 2390.529 |
| 95% percentile | 2791.07 |
| 99% percentile | 4901.122 |
| Average value | 1107.988 |
| Standard deviation | 943.3924 |
| Minimum | 196.1115 |
| Maximum | 7392.752 |
| Case number | 4530 |
| **Efficiency measure** | **57.97%** |

### 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 5 % percentile and the 95 % percentile of their distribution and after that rescaled in such a way that their total again resulted in the total of the untrimmed calibrated household weights. The trimmed projection factors range from 2676,57 to 38818,75. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.2 and upwards to 3.5. Thus, the total projection factors after calibration lie between min. 535.32 and max. 135865.6.

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

Population basis BA statistics:

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

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

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

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

**Table 61: Nominal distributions and distributions after calibration (Microm sample, individuals)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | 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 | 25 | 170,449 | 170,449 |
| | Number individuals in BCs Hamburg | 2 | 145,876 | 145,876 |
| | Number individuals in BCs Lower-Saxony | 56 | 473,772 | 473,772 |
| | Number individuals in BCs Bremen | 8 | 69,246 | 69,246 |
| | Number individuals in BCs North Rhine-Westphalia | 109 | 1,188,016 | 1,188,016 |
| | Number individuals in BCs Hesse | 30 | 316,524 | 316,524 |
| | Number individuals in BCs Rhineland-Palatinate | 20 | 175,232 | 175,232 |
| | Number individuals in BCs Baden-Wuerttemberg | 20 | 343,728 | 343,728 |
| | Number individuals in BCs Bavaria | 46 | 359,156 | 359,156 |
| | Number individuals in BCs Saarland | 11 | 61,900 | 61,900 |
| | Number individuals in BCs Berlin | 25 | 447,970 | 447,970 |
| | Number individuals in BCs Brandenburg | 27 | 237,387 | 237,387 |
| | Number individuals in BCs Mecklenburg-Vorpommern | 11 | 180,515 | 180,515 |
| | Number individuals in BCs Saxony | 25 | 396,110 | 396,110 |
| | Number individuals in BCs Saxony-Anhalt | 35 | 268,792 | 268,792 |
| | Number individuals in BCs Thuringia | 26 | 184,513 | 184,513 |
| 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 | 102 | 971,951 | 971,951 |
| | Number individuals in BCs aged 25-64 | 374 | 4,047,235 | 4,047,235 |
| Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by sex and west/east (4 categories) | Number men in BCs (west) | 152 | 1,608,020 | 1,608,020 |
| | Number women in BCs (west) | 175 | 1,695,879 | 1,695,879 |
| | Number men in BCs (east) | 67 | 879,978 | 879,978 |
| | Number women in BCs (east) | 82 | 835,309 | 835,309 |
| Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by "single parent yes/no", sex and west/east (8 categories) | Number non single parents in BCs (west) | 279 | 2,851,211 | 2,851,211 |
| | Number single parents in BCs (west) | 48 | 452,688 | 452,688 |
| | Number non single parents in BCs (east) | 132 | 1,523,271 | 1,523,271 |
| | Number single parents in BCs (east) | 17 | 192,016 | 192,016 |
| Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by nationality (German/non-German) | Number non-German individuals in BCs | 62 | 968,390 | 968,390 |
| | Number German individuals in BCs | 414 | 4,050,796 | 4,050,796 |

**Table 61: Nominal distributions and distributions after calibration (Microm sample, individuals) (continued 1)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of individuals aged 15 and over in private households by federal state (16 categories) | Number individuals in private households Schleswig-Holstein | 204 | 2,397,000 | 2,397,000 |
| | Number individuals in private households Hamburg | 58 | 1,553,000 | 1,553,000 |
| | Number individuals in private households Lower-Saxony | 561 | 6,735,000 | 6,735,000 |
| | Number individuals in private households Bremen | 29 | 575,000 | 575,000 |
| | Number individuals in private households North Rhine-Westphalia | 1,050 | 15,301,000 | 15,301,000 |
| | Number individuals in private households Hesse | 407 | 5,192,000 | 5,192,000 |
| | Number individuals in private households Rhineland-Palatinate | 238 | 3,438,000 | 3,438,000 |
| | Number individuals in private households Baden-Wuerttemberg | 590 | 9,130,000 | 9,130,000 |
| | Number individuals in private households Bavaria | 866 | 10,627,000 | 10,627,000 |
| | Number individuals in private households Saarland | 60 | 878,000 | 878,000 |
| | Number individuals in private households Berlin | 140 | 3,011,000 | 3,011,000 |
| | Number individuals in private households Brandenburg | 170 | 2,223,000 | 2,223,000 |
| | Number individuals in private households Mecklenburg-Vorpommern | 90 | 1,470,000 | 1,470,000 |
| | Number individuals in private households Saxony | 223 | 3,692,000 | 3,692,000 |
| | Number individuals in private households Saxony-Anhalt | 169 | 2,101,000 | 2,101,000 |
| | Number individuals in private households Thuringia | 161 | 1,993,000 | 1,993,000 |
| Number of individuals aged 15 and over in private households by age (in 5-year classes), sex and west/east (56 categories) | Number men in private households (west), 15-19 years | 171 | 1,943,000 | 1,943,000 |
| | Number men in private households (west), 20-24 years | 131 | 1,938,000 | 1,938,000 |
| | Number men in private households (west), 25-29 years | 79 | 1,962,000 | 1,962,000 |
| | Number men in private households (west), 30-34 years | 88 | 1,876,000 | 1,876,000 |
| | Number men in private households (west), 35-39 years | 136 | 2,178,000 | 2,178,000 |
| | Number men in private households (west), 40-44 years | 207 | 2,916,000 | 2,916,000 |
| | Number men in private households (west), 45-49 years | 204 | 2,681,000 | 2,681,000 |
| | Number men in private households (west), 50-54 years | 203 | 2,304,000 | 2,304,000 |
| | Number men in private households (west), 55-59 years | 166 | 2,063,000 | 2,063,000 |

**Table 61: Nominal distributions and distributions after calibration (Microm sample, individuals) (continued 2)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of individuals aged 15 and over in private households by age (in 5-year classes), sex and west/east (56 categories) | Number men in private households (west), 60-64 years | 133 | 1,724,000 | 1,724,000 |
| | Number men in private households (west), 65-69 years | 146 | 1,896,000 | 1,896,000 |
| | Number men in private households (west), 70-74 years | 127 | 1,703,000 | 1,703,000 |
| | Number men in private households (west), 75-79 years | 56 | 1,037,000 | 1,037,000 |
| | Number men in private households (west), 80+ years | 44 | 997,000 | 997,000 |
| | Number women in private households (west), 15-19 years | 158 | 1,810,000 | 1,810,000 |
| | Number women in private households (west), 20-24 years | 122 | 1,878,000 | 1,878,000 |
| | Number women in private households (west), 25-29 years | 101 | 1,959,000 | 1,959,000 |
| | Number women in private households (west), 30-34 years | 119 | 1,889,000 | 1,889,000 |
| | Number women in private households (west), 35-39 years | 177 | 2,186,000 | 2,186,000 |
| | Number women in private households (west), 40-44 years | 238 | 2,797,000 | 2,797,000 |
| | Number women in private households (west), 45-49 years | 280 | 2,656,000 | 2,656,000 |
| | Number women in private households (west), 50-54 years | 230 | 2,338,000 | 2,338,000 |
| | Number women in private households (west), 55-59 years | 199 | 2,113,000 | 2,113,000 |
| | Number women in private households (west), 60-64 years | 151 | 1,766,000 | 1,766,000 |
| | Number women in private households (west), 65-69 years | 162 | 2,058,000 | 2,058,000 |
| | Number women in private households (west), 70-74 years | 125 | 1,941,000 | 1,941,000 |
| | Number women in private households (west), 75-79 years | 65 | 1,335,000 | 1,335,000 |
| | Number women in private households (west), 80+ years | 45 | 1,882,000 | 1,882,000 |
| | Number men in private households (east), 15-19 years | 28 | 365,000 | 365,000 |
| | Number men in private households (east), 20-24 years | 27 | 557,000 | 557,000 |
| | Number men in private households (east), 25-29 years | 31 | 563,000 | 563,000 |
| | Number men in private households (east), 30-34 years | 23 | 487,000 | 487,000 |
| | Number men in private households (east), 35-39 years | 22 | 540,000 | 540,000 |
| | Number men in private households (east), 40-44 years | 43 | 698,000 | 698,000 |
| | Number men in private households (east), 45-49 years | 55 | 698,000 | 698,000 |
| | Number men in private households (east), 50-54 years | 60 | 643,000 | 643,000 |

| | | | | |
|---|---|---|---|---|
| | Number men in private households (east), 55-59 years | 38 | 608,000 | 608,000 |
| | Number men in private households (east), 60-64 years | 33 | 417,000 | 417,000 |
| | Number men in private households (east), 65-69 years | 31 | 550,000 | 550,000 |
| | Number men in private households (east), 70-74 years | 24 | 487,000 | 487,000 |
| | Number men in private households (east), 75-79 years | 17 | 274,000 | 274,000 |
| | Number men in private households (east), 80+ years | 7 | 221,000 | 221,000 |

**Table 61: Nominal distributions and distributions after calibration (Microm sample, individuals) (continued 3)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of individuals aged 15 and over in private households by age (in 5-year classes), sex and west/east (56 categories) | Number women in private households (east), 15-19 years | 26 | 335,000 | 335,000 |
| | Number women in private households (east), 20-24 years | 32 | 513,000 | 513,000 |
| | Number women in private households (east), 25-29 years | 20 | 492,000 | 492,000 |
| | Number women in private households (east), 30-34 years | 22 | 448,000 | 448,000 |
| | Number women in private households (east), 35-39 years | 40 | 496,000 | 496,000 |
| | Number women in private households (east), 40-44 years | 40 | 648,000 | 648,000 |
| | Number women in private households (east), 45-49 years | 55 | 676,000 | 676,000 |
| | Number women in private households (east), 50-54 years | 57 | 641,000 | 641,000 |
| | Number women in private households (east), 55-59 years | 65 | 627,000 | 627,000 |
| | Number women in private households (east), 60-64 years | 45 | 439,000 | 439,000 |
| | Number women in private households (east), 65-69 years | 40 | 625,000 | 625,000 |
| | Number women in private households (east), 70-74 years | 35 | 586,000 | 586,000 |
| | Number women in private households (east), 75-79 years | 22 | 378,000 | 378,000 |
| | Number women in private households (east), 80+ years | 15 | 478,000 | 478,000 |
| Number of individuals aged 15 and over in private households by household size (1, 2, 3, 4, "5 or more individuals") and west/east (10 categories) | Number individuals in private households with 1 individual (west) | 598 | 11,877,000 | 11,877,000 |
| | Number individuals in private households with 2 individuals (west) | 1,445 | 20,639,000 | 20,639,000 |
| | Number individuals in private households with 3 individuals (west) | 763 | 9,826,000 | 9,826,000 |
| | Number individuals in private households with 4 individuals (west) | 787 | 9,229,000 | 9,229,000 |
| | Number individuals in private households with 5 or more individuals (west) | 470 | 4,255,000 | 4,255,000 |
| | Number individuals in private households with 1 individual (east) | 187 | 3,627,000 | 3,627,000 |
| | Number individuals in private households with 2 individuals (east) | 380 | 5,953,000 | 5,953,000 |
| | Number individuals in private households with 3 individuals (east) | 203 | 2,847,000 | 2,847,000 |
| | Number individuals in private households with 4 individuals (east) | 128 | 1,595,000 | 1,595,000 |
| | Number individuals in private households with 5 or more individuals (east) | 55 | 468,000 | 468,000 |

**Table 61: Nominal distributions and distributions after calibration (Microm sample, individuals) (continued 4)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of individuals aged 15 and over in private households by highest school qualification and west/east (12 categories) | Number individuals in private households with highest school qualification: still pupil (west) | 206 | 2,348,000 | 2,348,000 |
| | Number individuals in private households with highest school qualification: no qualification (west) | 104 | 1,887,000 | 1,887,000 |
| | Number individuals in private households with highest school qualification: lower secondary school (west) | 1,329 | 23,468,000 | 23,468,000 |
| | Number individuals in private households with highest school qualification: intermediate secondary school; intermediate secondary school in the former GDR (west) | 1,168 | 13,583,000 | 13,583,000 |
| | Number individuals in private households with highest school qualification: university (of applied sciences) qualification (west) | 1,256 | 14,540,000 | 14,540,000 |
| | Number individuals in private households with highest school qualification: still pupil (east) | 35 | 380,000 | 380,000 |
| | Number individuals in private households with highest school qualification: no qualification (east) | 20 | 275,000 | 275,000 |
| | Number individuals in private households with highest school qualification: lower secondary school (east) | 226 | 3,802,000 | 3,802,000 |
| | Number individuals in private households with highest school qualification: Intermediate secondary school; intermediate secondary school in the former GDR (east) | 425 | 6,449,000 | 6,449,000 |
| | Number individuals in private households with highest school qualification: university (of applied sciences) qualification (east) | 247 | 3,584,000 | 3,584,000 |
| Number of individuals aged 15 and over in private households by marital status and west/east (10 categories) | Number individuals in private households with marital status: single (west) | 992 | 9,886,000 | 9,886,000 |
| | Number individuals in private households with marital status: married, civil partnership (west) | 2,504 | 36,324,000 | 36,324,000 |
| | Number individuals in private households with marital status: divorced (west) | 342 | 4,849,000 | 4,849,000 |
| | Number individuals in private households with marital status: widowed (west) | 225 | 4,767,000 | 4,767,000 |
| | Number individuals in private households with marital status: single (east) | 242 | 3,325,000 | 3,325,000 |
| | Number individuals in private households with marital status: married, civil partnership (east) | 540 | 8,389,000 | 8,389,000 |

| | | | | |
|---|---|---|---|---|
| | Number individuals in private households with marital status: divorced (east) | 99 | 1,497,000 | 1,497,000 |
| | Number individuals in private households with marital status: widowed (east) | 72 | 1,279,000 | 1,279,000 |
| Number of individuals aged 15 and over in private households by nationality | Number individuals in private households non-German | 178 | 6,354,000 | 6,354,000 |
| | Number individuals in private households German | 4,838 | 63,962,000 | 63,962,000 |

**Table 61: Nominal distributions and distributions after calibration (Microm sample, individuals) (continued 5)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Unemployed individuals incl. participants in measures west/east | Not unemployed west | 3,828 | 52,539,412 | 52,539,412 |
| | Unemployed individuals incl. participants in measures west | 235 | 3,286,588 | 3,286,588 |
| | Not unemployed east | 845 | 12,998,476 | 12,998,476 |
| | Unemployed individuals incl. participants in measures east | 108 | 1,491,524 | 1,491,524 |
| Employees subject to social security contributions west/east | Employees not subject to social security contributions west | 2,400 | 33,504,848 | 33,504,848 |
| | Employees subject to social security contributions west | 1,663 | 22,321,152 | 22,321,152 |
| | Employees not subject to social security contributions east | 551 | 9,246,200 | 9,246,200 |
| | Employees subject to social security contributions east | 402 | 5,243,800 | 5,243,800 |

**Table 62: Parameters of distribution of weights**

| | |
|---|---|
| 1% percentile | 795.634 |
| 5% percentile | 1747.698 |
| 10% percentile | 2419.551 |
| 25% percentile | 4431.909 |
| 50% percentile | 9077.082 |
| 75% percentile | 17655.82 |
| 90% percentile | 31548.64 |
| 95% percentile | 43427.17 |
| 99% percentile | 73782.65 |
| Average value | 14018.34 |
| Standard deviation | 15070.33 |
| Minimum | 535.3148 |
| Maximum | 135865.6 |
| Case number | 5016 |
| **Efficiency measure** | **46.39%** |

### 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 5 % percentile and the 95 % percentile of their distribution and after that rescaled in such a way that their total again resulted in the total of the untrimmed calibrated household weights. The trimmed projection factors range from 141.75 to 29423.42. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.2 and upwards to 5.0. Thus, the total projection factors after calibration lie between min. 28.35 and max. 123298.8.

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

Population basis BA statistics:

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

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

Besides that, also the increase in Unemployment Benefit II recipients since the previous year at the level of individuals between 15 and 64 years (499,434) was included as an additional benchmark value in the total sample.

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

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

**Table 63: Nominal distributions and distributions after calibration (total sample, individuals)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | 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 | 201 | 170,449 | 170449 |
| | Number individuals in BCs Hamburg | 95 | 145,876 | 145876 |
| | Number individuals in BCs Lower-Saxony | 473 | 473,772 | 473772 |
| | Number individuals in BCs Bremen | 59 | 69,246 | 69246 |
| | Number individuals in BCs North Rhine-Westphalia | 1,223 | 1,188,016 | 1188016 |
| | Number individuals in BCs Hesse | 281 | 316,524 | 316524 |
| | Number individuals in BCs Rhineland-Palatinate | 165 | 175,232 | 175232 |
| | Number individuals in BCs Baden-Wuerttemberg | 338 | 343,728 | 343728 |
| | Number individuals in BCs Bavaria | 419 | 359,156 | 359156 |
| | Number individuals in BCs Saarland | 75 | 61,900 | 61900 |
| | Number individuals in BCs Berlin | 369 | 447,970 | 447970 |
| | Number individuals in BCs Brandenburg | 265 | 237,387 | 237387 |
| | Number individuals in BCs Mecklenburg-Vorpommern | 138 | 180,515 | 180515 |
| | Number individuals in BCs Saxony | 356 | 396,110 | 396110 |
| | Number individuals in BCs Saxony-Anhalt | 366 | 268,792 | 268792 |
| | Number individuals in BCs Thuringia | 189 | 184,513 | 184513 |
| 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 | 865 | 971,951 | 971951 |
| | Number individuals in BCs aged 25-64 | 4,147 | 4,047,235 | 4047235 |
| Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by sex and west/east (4 categories) | Number men in BCs (west) | 1,500 | 1,608,020 | 1608020 |
| | Number women in BCs (west) | 1,829 | 1,695,879 | 1695879 |
| | Number men in BCs (east) | 813 | 879,978 | 879978 |
| | Number women in BCs (east) | 870 | 835,309 | 835309 |
| Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by "single parent yes/no", sex and west/east (8 categories) | Number non single parents in BCs (west) | 2,695 | 2,851,211 | 2851211 |
| | Number single parents in BCs (west) | 634 | 452,688 | 452688 |
| | Number non single parents in BCs (east) | 1,459 | 1,523,271 | 1523271 |
| | Number single parents in BCs (east) | 224 | 192,016 | 192016 |
| Number of individuals aged 15 and over in benefit communities receiving benefits in accordance with SGB II by nationality (German/non-German) | Number non-German individuals in BCs | 578 | 968,390 | 968390 |
| | Number German individuals in BCs | 4,434 | 4,050,796 | 4050796 |

**Table 63: Nominal distributions and distributions after calibration (total sample, individuals) (continued 1)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of individuals aged 15 and over in private households by federal state (16 categories) | Number individuals in private households Schleswig-Holstein | 513 | 2,397,000 | 2,397,000 |
| | Number individuals in private households Hamburg | 192 | 1,553,000 | 1,553,000 |
| | Number individuals in private households Lower-Saxony | 1,178 | 6,735,000 | 6,735,000 |
| | Number individuals in private households Bremen | 103 | 575,000 | 575,000 |
| | Number individuals in private households North Rhine-Westphalia | 2,621 | 15,301,000 | 15,301,000 |
| | Number individuals in private households Hesse | 768 | 5,192,000 | 5,192,000 |
| | Number individuals in private households Rhineland-Palatinate | 447 | 3,438,000 | 3,438,000 |
| | Number individuals in private households Baden-Wuerttemberg | 1,080 | 9,130,000 | 9,130,000 |
| | Number individuals in private households Bavaria | 1,444 | 10,627,000 | 10,627,000 |
| | Number individuals in private households Saarland | 148 | 878,000 | 878,000 |
| | Number individuals in private households Berlin | 641 | 3,011,000 | 3,011,000 |
| | Number individuals in private households Brandenburg | 517 | 2,223,000 | 2,223,000 |
| | Number individuals in private households Mecklenburg-Vorpommern | 302 | 1,470,000 | 1,470,000 |
| | Number individuals in private households Saxony | 728 | 3,692,000 | 3,692,000 |
| | Number individuals in private households Saxony-Anhalt | 634 | 2,101,000 | 2,101,000 |
| | Number individuals in private households Thuringia | 452 | 1,993,000 | 1,993,000 |
| Number of individuals aged 15 and over in private households by age (in 5-year classes), sex and west/east (56 categories) | Number men in private households (west), 15-19 years | 345 | 1,943,000 | 1,943,000 |
| | Number men in private households (west), 20-24 years | 312 | 1,938,000 | 1,938,000 |
| | Number men in private households (west), 25-29 years | 289 | 1,962,000 | 1,962,000 |
| | Number men in private households (west), 30-34 years | 268 | 1,876,000 | 1,876,000 |
| | Number men in private households (west), 35-39 years | 333 | 2,178,000 | 2,178,000 |
| | Number men in private households (west), 40-44 years | 424 | 2,916,000 | 2,916,000 |
| | Number men in private households (west), 45-49 years | 438 | 2,681,000 | 2,681,000 |

**Table 63: Nominal distributions and distributions after calibration (total sample, individuals) (continued 2)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of individuals aged 15 and over in private households by age (in 5-year classes), sex and west/east (56 categories) | Number men in private households (west), 50-54 years | 404 | 2,304,000 | 2,304,000 |
| | Number men in private households (west), 55-59 years | 363 | 2,063,000 | 2,063,000 |
| | Number men in private households (west), 60-64 years | 275 | 1,724,000 | 1,724,000 |
| | Number men in private households (west), 65-69 years | 204 | 1,896,000 | 1,896,000 |
| | Number men in private households (west), 70-74 years | 143 | 1,703,000 | 1,703,000 |
| | Number men in private households (west), 75-79 years | 61 | 1,037,000 | 1,037,000 |
| | Number men in private households (west), 80+ years | 47 | 997,000 | 997,000 |
| | Number women in private households (west), 15-19 years | 322 | 1,810,000 | 1,810,000 |
| | Number women in private households (west), 20-24 years | 339 | 1,878,000 | 1,878,000 |
| | Number women in private households (west), 25-29 years | 355 | 1,959,000 | 1,959,000 |
| | Number women in private households (west), 30-34 years | 396 | 1,889,000 | 1,889,000 |
| | Number women in private households (west), 35-39 years | 432 | 2,186,000 | 2,186,000 |
| | Number women in private households (west), 40-44 years | 542 | 2,797,000 | 2,797,000 |
| | Number women in private households (west), 45-49 years | 600 | 2,656,000 | 2,656,000 |
| | Number women in private households (west), 50-54 years | 461 | 2,338,000 | 2,338,000 |
| | Number women in private households (west), 55-59 years | 405 | 2,113,000 | 2,113,000 |
| | Number women in private households (west), 60-64 years | 278 | 1,766,000 | 1,766,000 |
| | Number women in private households (west), 65-69 years | 203 | 2,058,000 | 2,058,000 |
| | Number women in private households (west), 70-74 years | 135 | 1,941,000 | 1,941,000 |
| | Number women in private households (west), 75-79 years | 71 | 1,335,000 | 1,335,000 |
| | Number women in private households (west), 80+ years | 49 | 1,882,000 | 1,882,000 |
| | Number men in private households (east), 15-19 years | 105 | 365,000 | 365,000 |
| | Number men in private households (east), 20-24 years | 127 | 557,000 | 557,000 |
| | Number men in private households | 173 | 563,000 | 563,000 |

| | | | | |
|---|---|---|---|---|
| | (east), 25-29 years | | | |
| | Number men in private households (east), 30-34 years | 137 | 487,000 | 487,000 |
| | Number men in private households (east), 35-39 years | 118 | 540,000 | 540,000 |
| | Number men in private households (east), 40-44 years | 142 | 698,000 | 698,000 |

**Table 63: Nominal distributions and distributions after calibration (total sample, individuals) (continued 3)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of individuals aged 15 and over in private households by age (in 5-year classes), sex and west/east (56 categories) | Number men in private households (east), 45-49 years | 190 | 698,000 | 698,000 |
| | Number men in private households (east), 50-54 years | 185 | 643,000 | 643,000 |
| | Number men in private households (east), 55-59 years | 166 | 608,000 | 608,000 |
| | Number men in private households (east), 60-64 years | 116 | 417,000 | 417,000 |
| | Number men in private households (east), 65-69 years | 45 | 550,000 | 550,000 |
| | Number men in private households (east), 70-74 years | 30 | 487,000 | 487,000 |
| | Number men in private households (east), 75-79 years | 17 | 274,000 | 274,000 |
| | Number men in private households (east), 80+ years | 7 | 221,000 | 221,000 |
| | Number women in private households (east), 15-19 years | 110 | 335,000 | 335,000 |
| | Number women in private households (east), 20-24 years | 133 | 513,000 | 513,000 |
| | Number women in private households (east), 25-29 years | 156 | 492,000 | 492,000 |
| | Number women in private households (east), 30-34 years | 152 | 448,000 | 448,000 |
| | Number women in private households (east), 35-39 years | 133 | 496,000 | 496,000 |
| | Number women in private households (east), 40-44 years | 155 | 648,000 | 648,000 |
| | Number women in private households (east), 45-49 years | 229 | 676,000 | 676,000 |
| | Number women in private households (east), 50-54 years | 221 | 641,000 | 641,000 |
| | Number women in private households (east), 55-59 years | 175 | 627,000 | 627,000 |
| | Number women in private households (east), 60-64 years | 113 | 439,000 | 439,000 |
| | Number women in private households (east), 65-69 years | 57 | 625,000 | 625,000 |
| | Number women in private households (east), 70-74 years | 41 | 586,000 | 586,000 |
| | Number women in private households (east), 75-79 years | 23 | 378,000 | 378,000 |
| | Number women in private households (east), 80+ years | 18 | 478,000 | 478,000 |
| Number of individuals aged 15 and over in private households by household size (1, 2, 3, 4, "5 or more individuals") and west/east (10 categories) | Number individuals in private households with 1 individual (west) | 1,824 | 11,877,000 | 11,877,000 |
| | Number individuals in private households with 2 individuals (west) | 2,667 | 20,639,000 | 20,639,000 |
| | Number individuals in private households with 3 individuals (west) | 1,758 | 9,826,000 | 9,826,000 |
| | Number individuals in private households with 4 individuals (west) | 1,355 | 9,229,000 | 9,229,000 |

| | | | | |
|---|---|---|---|---|
| | Number individuals in private households with 5 or more individuals (west) | 890 | 4,255,000 | 4,255,000 |
| | Number individuals in private households with 1 individual (east) | 904 | 3,627,000 | 3,627,000 |
| | Number individuals in private households with 2 individuals (east) | 1,086 | 5,953,000 | 5,953,000 |
| | Number individuals in private households with 3 individuals (east) | 678 | 2,847,000 | 2,847,000 |
| | Number individuals in private households with 4 individuals (east) | 403 | 1,595,000 | 1,595,000 |
| | Number individuals in private households with 5 or more individuals (east) | 203 | 468,000 | 468,000 |

**Table 63: Nominal distributions and distributions after calibration (total sample, individuals) (continued 4)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of individuals aged 15 and over in private households by highest school qualification and west/east (12 categories) | Number individuals in private households with highest school qualification: still pupil (west) | 408 | 2,348,000 | 2,348,000 |
| | Number individuals in private households with highest school qualification: no qualification (west) | 426 | 1,887,000 | 1,887,000 |
| | Number individuals in private households with highest school qualification: lower secondary school (west) | 3,065 | 23,468,000 | 23,468,000 |
| | Number individuals in private households with highest school qualification: intermediate secondary school; intermediate secondary school in the former GDR (west) | 2,432 | 13,583,000 | 13,583,000 |
| | Number individuals in private households with highest school qualification: university (of applied sciences) qualification (west) | 2,163 | 14,540,000 | 14,540,000 |
| | Number individuals in private households with highest school qualification: still pupil (east) | 118 | 380,000 | 380,000 |
| | Number individuals in private households with highest school qualification: no qualification (east) | 107 | 275,000 | 275,000 |
| | Number individuals in private households with highest school qualification: lower secondary school (east) | 772 | 3,802,000 | 3,802,000 |
| | Number individuals in private households with highest school qualification: Intermediate secondary school; intermediate secondary school in the former GDR (east) | 1,634 | 6,449,000 | 6,449,000 |
| | Number individuals in private households with highest school qualification: university (of applied sciences) qualification (east) | 643 | 3,584,000 | 3,584,000 |
| Number of individuals aged 15 and over in private households by marital status and west/east (10 categories) | Number individuals in private households with marital status: single (west) | 2,759 | 9,886,000 | 9,886,000 |
| | Number individuals in private households with marital status: married, civil partnership (west) | 4,120 | 36,324,000 | 36,324,000 |
| | Number individuals in private households with marital status: divorced (west) | 1,282 | 4,849,000 | 4,849,000 |
| | Number individuals in private households with marital status: widowed (west) | 333 | 4,767,000 | 4,767,000 |
| | Number individuals in private households with marital status: single (east) | 1,304 | 3,325,000 | 3,325,000 |

| | | | | |
|---|---|---|---|---|
| | Number individuals in private households with marital status: married, civil partnership (east) | 1,329 | 8,389,000 | 8,389,000 |
| | Number individuals in private households with marital status: divorced (east) | 514 | 1,497,000 | 1,497,000 |
| | Number individuals in private households with marital status: widowed (east) | 127 | 1,279,000 | 1,279,000 |

**Table 63: Nominal distributions and distributions after calibration (total sample, individuals) (continued 5)**

| Benchmark figure | Characteristics benchmark figure from BA statistics and Mikrozensus 2009 | Unweighted distribution | Nominal values | Distribution with calibrated weights |
|---|---|---|---|---|
| Number of individuals aged 15 and over in private households by nationality | Number individuals in private households non-German | 873 | 6,354,000 | 6,354,000 |
| | Number individuals in private households German | 10,895 | 63,962,000 | 63,962,000 |
| Unemployed individuals incl. participants in measures west/east | Not unemployed west | 6,499 | 52,539,412 | 52,539,412 |
| | Unemployed individuals incl. participants in measures west | 1,995 | 3,286,588 | 3,286,588 |
| | Not unemployed east | 2,170 | 12,998,476 | 12,998,476 |
| | Unemployed individuals incl. participants in measures east | 1,104 | 1,491,524 | 1,491,524 |
| Employees subject to social security contributions west/east | Employees not subject to social security contributions west | 5,466 | 33,504,848 | 33,504,848 |
| | Employees subject to social security contributions west | 3,028 | 22,321,152 | 22,321,152 |
| | Employees not subject to social security contributions east | 2,044 | 9,246,200 | 9,246,200 |
| | Employees subject to social security contributions east | 1,230 | 5,243,800 | 5,243,800 |

**Table 64: Parameters of distribution of weights**

| | |
|---|---|
| 1% percentile | 31.43041 |
| 5% percentile | 79.49212 |
| 10% percentile | 134.8719 |
| 25% percentile | 321.3353 |
| 50% percentile | 1247.177 |
| 75% percentile | 6821.727 |
| 90% percentile | 18490.31 |
| 95% percentile | 27502.24 |
| 99% percentile | 48845.01 |
| Average value | 5975.187 |
| Standard deviation | 10608.77 |
| Minimum | 28.3505 |
| Maximum | 123298.8 |
| Case number | 11768 |
| **Efficiency measure** | **24.08%** |

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

Finally, also in wave 4 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 4 than in wave 3, since a larger part of the BA sample of wave 1 and wave 2 has meanwhile withdrawn from receiving benefits. These are the following three groups which did not receive benefits in July 2009 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 or 7/2009 and individuals in households with receipt of Unemployment Benefit II in 7/2006 or 7/2007 or 7/2008 or 7/2009).

- 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 4 after calibration (*wqbahh*) by dividing it by the proportion of these individuals who gave a personal or senior citizens' interview – provided that their household was participating.
- Panel households in which nobody was in receipt of Unemployment Benefit II any longer in July 2009: 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 3 are given a new BA person weight calculated by dividing the BA household weight of their household for wave 4 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 panel households that were still in receipt of Unemployment Benefit II in July 2009: 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:**<br>artificial individual ID; sex; year of birth; age; marital status; number of children living in and outside the household; nationality; country of origin and migration background; school and vocational qualifications (incl. generated scales: CASMIN, ISCED-97, number of years of schooling and vocational training), parents' school and vocational qualifications; health indicators; religious denomination; social contacts; childcare and school attendance of children; household income (incl. individual components and equivalised household income); basic information on assets and liabilities; household equipment (deprivation index); housing and residential environment; detailed information on the topic of old age benefits (only wave 3);<br>**Employment-related characteristics:**<br>employment status / economic inactivity status; marginal employment; working hours; occupational status (detailed); employment (ISCO-88 and KldB-92); ISCO-based measures of occupational status and prestige (ISEI, SIOPS, MPS, EGP, ESeC); earned income (gross and net); employment biographies with employment/unemployment spells and periods of economic inactivity since 01/2005 (from wave 2 onwards); 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;<br>**Characteristics on receiving benefits:**<br>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;<br>Unemployment Benefit II: start and end dates of the spell(s) 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);<br>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; from wave 4 onwards only one-euro job;<br>Contacts with Unemployment Benefit II institutions: number and type of contacts; contents of discussion; offers; integration agreement; assessment of institution;<br>**Subjective indicators:**<br>satisfaction; fears and problems; employment orientation; education aspiration; sex role orientation; subjective social position (top-bottom scale); subjective assessment of health state |

| Categories | Comments |
|---|---|
| Data unit | Individuals and households in receipt of Unemployment Benefit II in 7/2006 (sample I)<br>Individuals and households in the resident population of Germany (sample II)<br>Individuals and households in receipt of Unemployment Benefit II in 7/2007 but without receipt in 7/2006 (sample III; refreshment sample 1)<br>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)<br>Individuals and households in receipt of Unemployment Benefit II in 7/2009 but without receipt in 7/2006, 7/2007 or 7/2008 (sample V; refreshment sample 3)<br>Note: individuals aged 65 and over are interviewed using a shorter version of the questionnaire |
| Number of cases | Wave 1:<br>Sample I: 9.386 individuals (living in 6.804 households)<br>Sample II: 9.568 individuals (living in 5.990 households)<br>Wave 2:<br>Sample I: 4.753 individuals (living in 3.491 households)<br>Sample II: 6.392 individuals (living in 3.897 households)<br>Sample III: 1.342 individuals (living in 1.041 households)<br>Wave 3:<br>Sample I: 4.913 individuals (living in 3.754 households)<br>Sample II: 6.207 individuals (living in 3.901 households)<br>Sample III: 898 individuals (living in 694 households)<br>Sample IV: 1.421 individuals (living in 1.186 households)<br>Wave 4:<br>Sample I: 3.958 individuals (living in 2.815 households)<br>Sample II: 5,016 individuals (living in 2,977 households)<br>Sample III: 786 individuals (living in 563 households)<br>Sample IV: 983 individuals (living in 745 households)<br>Sample V: 1,025 individuals (living in 748 households) |

| Categories | Comments |
|---|---|
| Data collection mode | CATI and CAPI<br>CAPI interviews were conducted when a sample household could not be reached by telephone or when a personal interview was requested.<br>Wave 1:<br>N (CATI): 12.414 individuals (8.445 households)<br>N (CAPI): 6.540 individuals (4.339 households)<br>Wave 2:<br>N (CATI): 7.888 individuals (5.378 households)<br>N (CAPI): 4.599 individuals (3.051 households)<br>Wave 3:<br>N (CATI): 7.776 individuals (5.664 households)<br>N (CAPI): 5.663 individuals (3.871 households)<br>Wave 4:<br>n (CATI): 6.913 individuals (4.669 households)<br>n (CAPI): 4.855 individuals (3.179 households) |
| Interview languages | Wave 1:<br>German: 18.205 individuals (12.347 households)<br>Russian: 432 individuals (275 households)<br>Turkish: 305 individuals (163 households)<br>English: 12 individuals (9 households)<br>Wave 2:<br>German: 12.237 individuals (8.234 households)<br>Russian: 219 individuals (156 households)<br>Turkish: 31 individuals (39 households)<br>English: no longer offered in wave 2 due to the low case numbers in wave 1<br>Wave 3:<br>German: 13.000 individuals (9.256 households)<br>Russian: 330 individuals (210 households)<br>Turkish: 109 individuals (69 households)<br><br>Wave 4:<br>German: 11,405 individuals (7,627 households)<br>Russian: 285 individuals (179 households)<br>Turkish: 78 individuals (42 households) |

| Categories | Comments |
|---|---|
| Response rates | Wave 1:<br>Sample I: 35.1 %<br>Sample II: 26.6 %<br>Total: 30.5 %<br>Wave 2:<br>Sample I (HHs agreeing to participate only): 51.1 %<br>Sample II (HHs agreeing to participate only): 64.7 %<br>Sample III: 26.3 %<br>Split-off households (from samples I and II): 13.4 %<br>Total: 45.0 %<br>Wave 3:<br>Sample I (HHs agreeing to participate only): 64.5 %<br>Sample II (HHs agreeing to participate only): 76.4 %<br>Sample II (HHs agreeing to participate only): 69.0 %<br>Sample IV: 31.2%<br>Total: 60.6 %<br>Wave 4:<br>Sample I (HHs agreeing to participate only): 72.1 %<br>Sample II (HHs agreeing to participate only): 82.4 %<br>Sample III (HHs agreeing to participate only): 65.6 %<br>Sample IV (HHs agreeing to participate only): 68.2 %<br>Sample V: :30.9 %<br>Total: 59.5 % |

| Categories | Comments |
|---|---|
| Response rates within households | Wave 1:<br>Sample I: 85.6 %<br>Sample II: 84.3 %<br>Total: 85.0 %<br>Wave 2:<br>Sample I (re-interviewed households only): 85.5 %<br>Sample II (re-interviewed households only): 85.1 %<br>Sample III: 86.2 %<br>Split-off households (from samples I and II): 88.3 %<br>Total: 85.4 %<br>Wave 3:<br>Sample I (re-interviewed households only): 83.1 %<br>Sample I (re-interviewed households only): 83.6 %<br>Sample III (re-interviewed households only): 84.3 %<br>Sample IV: 84.2 %<br>Split-off households (from samples I - II): 84.2 %<br>Total: 83.5 %<br>Wave 4:<br>Sample I (re-interviewed households only): 88.4 %<br>Sample I (re-interviewed households only): 88.0 %<br>Sample III (re-interviewed households only): 90.2 %<br>Sample IV (re-interviewed households only): 88.3 %<br>Sample V: 89.6 %<br>Split-off households (from samples I - IV): 86.4 %<br>Total: 88.5 % |
| Fieldwork period: | Wave 1: December 2006-June 2007<br>Wave 2: December 2007-July 2008<br>Wave 3: December 2008-August 2009<br>Wave 4: February 2010-September 2010 |
| Period covered | Wave 1: fieldwork period and retrospective spell data as of 01/2005<br>Wave 2: fieldwork period and retrospective spell data as of 01/2005 or the respective reference period of the spell type<br>Wave 3: fieldwork period and retrospective spell data as of 01/2006 or the respective reference period of the spell type<br>Wave 4: fieldwork period and retrospective spell data as of 01/2008 or the respective reference period of the spell type |
| Time reference | Repeat interview (household panel) |
| Regional structure | German federal state, east/west Germany<br>(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<br><br>Stage 1: 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).<br><br>Stage 2, 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.<br><br>Stage 2, sample II: for sample II, first a sample of residential buildings was drawn from a commercial database (Microm mosaic). This was then stratified by a stratification index contained in the database at a ratio of 4:2:1 for households with a low, medium or high status respectively. Interviewers from the surveying institute visited the selected buildings. In the event that a building accommodated several households, this was noted and then one of the households was selected by the institute as the household to be interviewed. The gross sample comprised N=100 households per postcode sector.<br><br>Refreshment sample 1 for sample I in wave 2:<br>In addition to continuing sample I which was drawn for wave 1, in the second wave a refreshment sample was drawn from the register data of the Federal Employment Agency. For this, benefit 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 first wave.<br><br>Refreshment sample 2 for sample I in wave 3:<br>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 first wave.<br><br>Refreshment sample 3 for sample I in wave 4:<br>Also in wave 4 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 2009 but not in July 2006, July 2007 or July 2008 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 proce- |

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

**Data access**

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

## Literature:


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

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

Andersen, H.H., Mühlbacher, A., & Nübling, M. (2007a). Die SOEP-Version des SF-12 als Instrument gesundheitsökonomischer Analysen. *SOEPpapers 6/2007*. Berlin.

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

Andersen, H.H., Mühlbacher, A., Nübling, M., Schupp, J. & Wagner, G.G. (2007): Computation of Standard Values for Physical and Mental Health Scale Scores Using the SOEP-Version of SF-12v2. *Schmollers Jahrbuch 127(1)*. pp. 171-182.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Granato, N. (2000). Mikrodaten-Tools: CASMIN-Bildungsklassifikation. Eine Umsetzung mit dem Mikrozensus 1996. *ZUMA-Technischer Bericht 2000/12.* Mannheim.

Hagenaars, A., de Vos, K. & Zaidi, M. (1994). *Poverty Statistics in the Late 1980s: Research Based on Micro-data.* Luxembourg: Office for Official Publications of the European Communities.

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

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

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

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

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

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

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

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

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

Lechert, Y., Schroedter, J. & Lüttinger, P. (2006). Die Umsetzung der Bildungsklassifikation CASMIN für die Volkszählung 1970, die Mikrozensus- Zusatzerhebung 1971 und die Mikrozensen 1976-2004. *ZUMA-Methodenbericht 2006/12.* Mannheim.

Lengerer, A., Bohr, J. & Janßen, A. (2005). Haushalte, Familien und Lebensformen im Mikrozensus – Konzepte und Typisierungen. *ZUMA-Arbeitsbericht 2005/05.* Mannheim.

Lipsmeier, G. (1999). Die Bestimmung des notwendigen Lebensstandards – Einschätzungsunterschiede und Entscheidungsprobleme. *Zeitschrift für Soziologie 28(4).* S. 281-300.

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

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

Nolan, B. & Whelan, C. (1996). Measuring Poverty Using Income and Deprivation Indicators: Alternative Approaches. *Journal of European Social Policy 6(3).* pp. 225-240.

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

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

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

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

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

Rose, R. & Harrison, E. (2007). The European Socio-Economic Classification: A New Social Class Schema for Comparative European Research. *European Societies 9(3).* pp. 459-490.

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

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. Wiesbaden: Statistisches Bundesamt.

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

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

Trappmann, M., Christoph, B., Achatz, J., Wenzig, C., Müller, G. & Gebhardt, D. (2009). Design and stratification of PASS. A New Panel Study for Research on Long Term Unemployment. *IAB-Discussion Paper 5/2009.* Nuremberg.

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

Ware J.E., Kosinski, M., Turner-Bowker, D.M. & Gandek, B. (2002). *How to score Version 2 of the SF-12 Health Survey*. Lincoln, RI: Qualitymetric.

Wegener, B. (1985): Gibt es Sozialprestige? *Zeitschrift für Soziologie 14(3).* S. 209-235.

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

# Imprint

**FDZ-Datenreport 08/2011**

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

**Editorial staff**
Stefan Bender, Dagmar Theune

**Technical production**
Dagmar Theune



**Download**
http://doku.iab.de/fdz/reporte/2011/DR_08-11_I_EN.pdf

**Internet**
http://fdz.iab.de/

**Corresponding author:**
Forschungsdatenzentrum (FDZ) der
Bundesagentur für Arbeit im Institut für
Arbeitsmarkt- und Berufsforschung (IAB),
Regensburger Str. 104, 90478 Nürnberg,
E-Mail: iab.fdz@iab.de

Arne Bethmann, Institut für
Arbeitsmarkt- und Berufsforschung (IAB),
Regensburger Str. 104, 90478 Nürnberg,
Tel.: +49 (0) 911/179-2307
E-Mail: arne.bethmann@iab.de