FDZ Datenreport
Documentation of labour market data

No. 5/2008 (EN)

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

## Volume I: Introduction and Overview

## Wave 1 (2006/2007)

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

Bundesagentur für Arbeit

## 1. The Panel Study 'Labour Market and Social Security': Introduction and Overview

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 political consulting. This Datenreport provides an overview of the first survey wave, for which 18,954 persons were interviewed in 12,794 households between December 2006 and July 2007.

The study is carried out as part of the IAB’s research into the German Social Code Book II (SGB II). The IAB is charged by law with studying the effects of benefits under SGB II for integration into the labour market and subsistence benefits. However, due to the complex sample design, it also enables researchers to answer questions far beyond these issues. Five core questions influenced the development of the new study, which are detailed in Achatz et al. (2007):

1. What options exist to regain independence from Unemployment Benefit II?
2. In which ways does the social situation of a household change when it receives benefits?
3. How do persons concerned cope with their situation? Will attitudes of the respondents that are constitutive for their actions change over time?
4. In which form do contacts between benefit recipients and institutions providing basic social security actually take place? What are the institutional procedures applied in practice?
5. Which employment career patterns or household dynamics lead to receipt of Unemployment Benefit II?

The following brief overview describes the motivation for carrying out the survey, its contents and the study design.

a) Additions to Existing Data

German labour market, poverty and welfare state research already has access to various micro datasets. In particular, there are a number of longitudinal datasets available covering relatively long survey periods. A particularly important source in the field of survey data is the German Socio-economic Panel Study (SOEP) (Wagner et al. 2007), which provides annual data on the individual and household level dating back to 1984. In addition, administrative data from the Federal Employment Agency (BA) is processed at the IAB and provided for research use by the Research Data Centre (FDZ) of the BA, for example in the form of the Integrated Employment Biographies (IEBS), the IAB Employment Samples (IABS) or the Linked Employer-Employee Dataset (LIAB).

The spectrum of questions and the design of PASS are intended to close gaps in the existing data. PASS has three main characteristics that extend analysis potential in addition to the Federal Employment Agency's administrative data:

1. The panel takes the household context into account – including the situation before and after receipt of Unemployment Benefit II.
2. The panel is complete, i.e. it does not exclude particular groups of persons or people with specific employment biographies, as it not only covers persons in dependent employment, unemployed persons and persons in need of assistance but all relevant groups. The dataset thus also provides information on the status during phases of unemployment, self-employment or employment as official civil servants.
3. The panel collects additional or significantly more detailed data on relevant characteristics such as attitudes, employment potential or job-seeking behaviour.

Compared to the existing surveys of individuals or households, PASS particularly aims to improve the data situation with regard to the following points:

1. The high case numbers of Unemployment Benefit II recipients (in wave 1: 10,197 interviewed persons in 7,350 households receiving such benefits) enable more detailed analyses – for example on the effect of SGB II on certain target groups – and more precise estimations of statistics and model coefficients than datasets in which benefit recipients are only included in proportion to their share of the population.

2. Gathering additional characteristics such as intensity and type of contacts to institutions providing basic social security or participation in labour market programmes enables analyses of the significance of institutional assistance for the population below the poverty line.
3. Linking the survey data with Federal Employment Agency process data enables both a validation of the characteristics surveyed, and secondly analyses in which the higher measurement precision of the process data can be combined with further variables and the household context from the survey.

b) Instruments and Core Topics of the Survey

An initial household interview was carried out with every head of household. Subsequently, the interviewers attempted to carry out personal interviews with every person in the household from the age of 15 and over. A briefer version of the personal questionnaire was used for persons aged 65 and over (referred to as a pensioner's questionnaire).

Figure 1 provides an overview of the subject spectrum of the survey. The socio-demographic characteristics are generally gathered using standard items. Please note the detailed record of migration background into the third generation (questionnaire p191-p202, variable code PMI) and the specific recording of employment status, which allows a better notation of parallel status, particularly relevant for benefit receipt, than standard demographic procedures (see items p24-p55 in the questionnaire, variable code PET).

The question block on the material situation includes detailed information on income and assets (although the first wave does not yet include all income components), along with a deprivation index (questionnaire: hh7, hh8, p141, variable code: HLS), which records the household's ownership of (consumer) goods.

The question complex on the social situation deals with non-material aspects of the living situation, which are generally not covered in the BA process data. In particular, these include the modules on health (p152-p164, PG) and social integration (p143-p146, PSK), on which the third wave of PASS will focus.

In addition, the panel records attitudes and orientations (p3, p23, p69-p70, p129, p190, PEO) as well as activities (e.g. job-seeking p116-p138, PAS) that are also not included in the process data.

Finally, PASS includes separate modules on receipt of insurance and social benefits (hh42-hh55, p71-p75), contact to the institutions providing these benefits (p76-p90, PTK), plus participation in labour market programmes (p92-p115, ALM).

**Figure 1: Core Topics of the Survey**

| Area / Level | Demography | Material situation | Social situation | Orientations & activities | Benefit receipt |
|---|---|---|---|---|---|
| **House-hold** | *List of household members (incl. age and gender)* | *Living standard of the household (deprivation index)* | *Housing / surrounding area* | | *Receipt of Unemployment Benefit II (spell data)* |
| | *Family relationships* | *Net household income* | | | |
| | *Children outside the household* | *Assets* | | | |
| | | *Debts and loans* | | | |
| **Indivi-dual** | *School and vocational education* | *Employment income, gross and net* | *Health (special focus topic Wave 3)* | *Gender-role orientation* | *Receipt of Unemployment Benefit I (spell data)* |
| | *Employment status* | *Other income components (e.g. special payments)* | *Social integration / networks, (special focus topic Wave 3)* | *Employment orientations* | *Institutional integration (e.g. contact to institutions, offers)* |
| | *Occupation* | | *Subjective evaluations (e.g. social positioning)* | *Job-seeking activities (search activities, reservation wage; special focus topic W1)* | *Participation in labour market programmes (e.g. type, length; spell data)* |
| | *Social origin (parents' education (W1) & occupation (W2))* | | *ISCO-based measures* | *Life satisfaction* | *Other social benefits (e.g. pensions)* |

## c) Survey Design

The many different requirements can only be achieved through a complex study design. Key considerations on this subject are detailed in an expertise by Schnell (2007), on the basis of which the final study design was developed at the IAB. The following points are of particular importance (see Rudolph and Trappmann 2007):

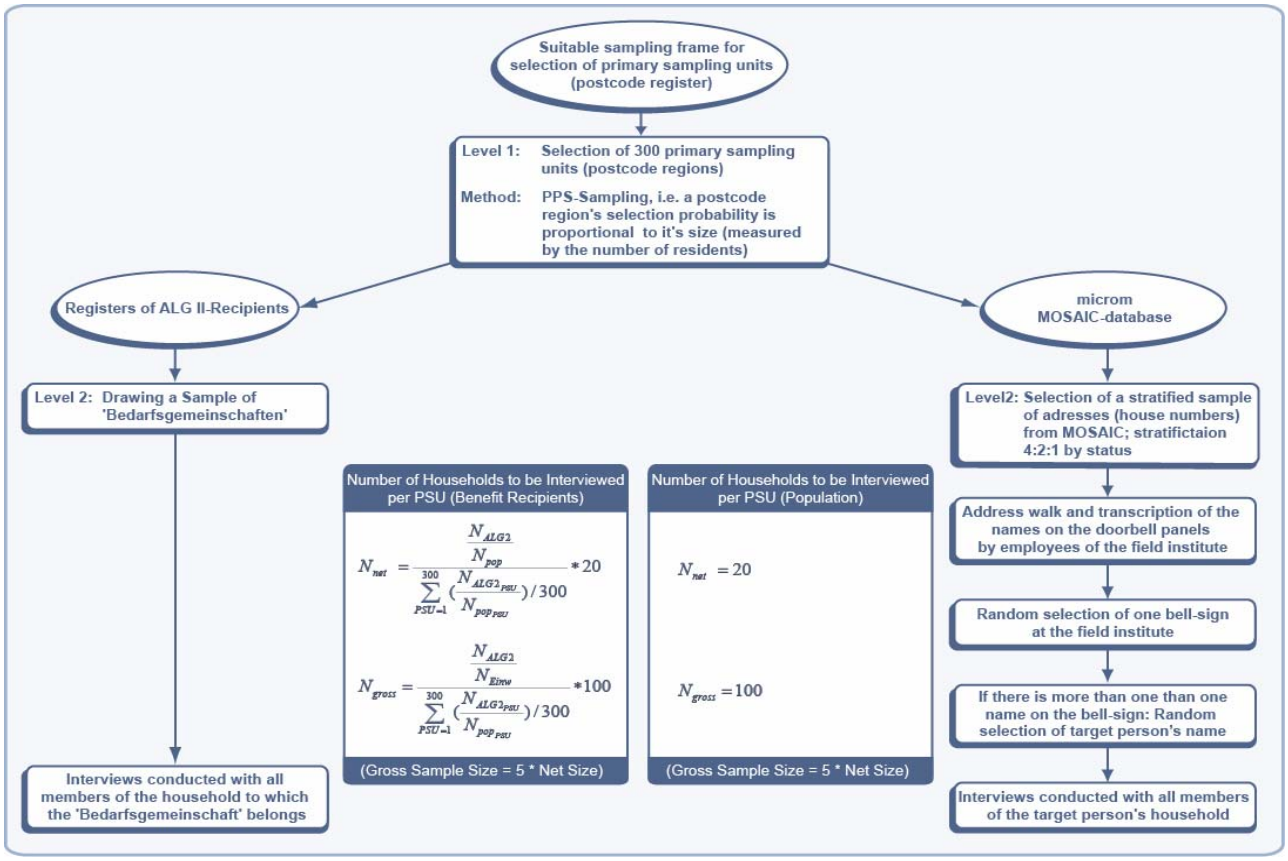
1. In order to avoid problems caused by low landline-phone-coverage of persons in the low-income bracket or transfer payment recipients, and at the same time carry out the survey on a cost-efficient basis, a CATI/CAPI mix was selected.[1]

[1] A recent mixed mode survey (CATI and CAPI) carried out by the IAB (Meßmann et al. 2008) found that more than 20 % of the interviewed benefit recipients (Unemployment Benefit II and social assistance recipients) do not have a fixed telephone line. These persons proved to be more deprived than other benefit recipients.

2. As one can assume that a considerable proportion of the target population does not have sufficient knowledge of German, the survey instruments were translated into the most common native languages of migrants in Germany – Turkish and Russian – and additionally into English.

3. For all persons who provided consent, the survey data were linked to the Federal Employment Agency's process data. This data merger, however, which succeeded for 72.0 % of the persons interviewed between the ages of 15 and 64, is not part of the scientific use file, but can only be used on site at the Research Data Centre of the BA in the IAB, for data protection reasons.

**Figure 2: The Sample Design**



4. As inflows into benefit receipt are also to be studied, and comparisons of SGB II benefit recipients with the general population are necessary or useful for many issues, a population sample was interviewed alongside a sample of households in which at least one person was receiving SGB II benefits as of July 2006. The population sample, based on a directory of residential addresses from the private company Microm, was also disproportionately stratified according to the provider's status index, so that persons with low social status and thus greater risk of entry into benefit receipt have a higher probability of inclusion (see Figure 2)[2]. Due to the complex panel design, we strongly advise against using the data without the resulting weightings, the generation and use of which are described in Chapter 7.

Figure 2 presents the key elements of the sample design. The target population of the population sample is all private households in Germany. The target population of the process data sample is all households in which at least one community in receipt of joint benefits (and thus at least one person) receives benefits in accordance with SGB II.

[2] The procedure is described in detail in Rudolph and Trappmann (2007). The results of the disproportionate stratification are detailed in Trappmann et al. (2007).

## 2. Central Statistics of the Study

This chapter provides a brief overview of central statistics of the study, such as the sample size and response rates. These are generally given separately for the two sub-samples described above (see Figure 2) as well as for the study as a whole.

- subsample 1 (BA sample) refers to the sample of benefit recipients from the Federal Employment Agency's process data.
- subsample 2 (Microm sample) refers to the stratified population sample.

a) Sample Size

The entire PASS sample includes a total of 12,794 households. Table 1 presents the distribution of these households across the two subsamples and the two modes of data collection.

**Table 1: Sample Size on Household Level**

| n | CATI | CAPI | Total |
|---|---|---|---|
| Subsample 1 (BA) | 5,139 | 1,665 | 6,804 |
| Subsample 2 (Microm) | 3,316 | 2,674 | 5,990 |
| Total | 8,455 | 4,339 | 12,794 |

On the individual level, these 12,794 households consist of 18,954 respondents. Table 2 presents the corresponding distribution of these persons across the samples and the two modes of data collection.

**Table 2: Sample Size on the Individual Level**

| n | CATI | CAPI | Total |
|---|---|---|---|
| Subsample 1 (BA) | 7,079 | 2,307 | 9,386 |
| Subsample 2 (Microm) | 5,335 | 4,233 | 9,568 |
| Total | 12,414 | 6,540 | 18,954 |

As described in Chapter 1, respondents without sufficient knowledge of German had the option of an interview in another language. The alternative interview languages were Turkish, Russian and English. Table 3 states how many households and persons were interviewed in the respective languages.

**Table 3: Sample Size of Alternative Language Interviews**

| n | Households | Persons |
|---|---|---|
| Russian | 276 | 426 |
| Turkish | 163 | 285 |
| English | 9 | 12 |
| Total | 448 | 723 |

b) Response Rates

In a household survey, one can distinguish between the response rate on the household level and the response rate within the households. “Response rate on the household level” refers to the proportion of completed households (in this case: complete household interview and at least one complete personal interview) to the number of eligible households.. “Response rate within households” refers to the mean proportion of persons aged 15 and over within households who took part in the survey if their household was taking part.

*Response rate on the household level*

TNS Infratest Sozialforschung, the institute commissioned to carry out the field work, reports the following response rates on the household level[3] :

| | |
|---|---|
| Response rate subsample 1 (BA): | 35.1 % |
| Response rate subsample 2 (Microm): | 26.6 % |
| Total response rate: | 30.5 % |

[3] As there are unfortunately no established standards for calculating response rates in German social research, we have provided these statistics as received. In comparison to other recent mixed-mode surveys of comparable populations, namely LSS 2005 (Meßmann et al. 2008, Infas 2006) and the benefit-recipient survey conducted as part of the evaluation of the experimentation clause (ZEW et al. 2007), a stricter definition of “eligibility” and “non-response” is used here. Despite comparable sampling bases, the proportion of non-eligible cases in PASS is 7, respectively 18 percent lower. Using comparable criteria, all three studies achieve an almost equal response rate in the population of benefit recipients. Using the stricter criteria of the American Association for Public Opinion Research’s RR1 (AAPOR 2006), the response rates in the benefit recipient samples of all three studies are around 30 %. A breakdown of the final disposition codes is contained in the TNS Infratest method and field report (Hartmann et al. 2008, Tables A2.1 and A2.2).

*Response rate within the households*

Within the households, the following response rates are reported:

| | |
|---|---|
| Response rate subsample 1 (BA): | 85.6% |
| Response rate subsample 2 (Microm): | 84.3% |
| Total response rate: | 85.0% |

The relatively low response rate on the household level in the subsample for the general population should be considered reflective of the subjective importance of the subject matter. Schnell (1997) argues that the "middle class bias" to be found in many studies (i.e. persons from the middle social strata are frequently overrepresented in population surveys) is in actual fact a bias caused on one hand by the poor reachability of and high opportunity costs for the upper social strata, and on the other hand by the in many cases low importance of the subject matter for the lower social strata (Schnell 1997, 201 ff.). Our data appear to confirm this argument. In a survey dealing with the subject of social security, however, there is a shift in person groups for whom the subject is of high importance towards those directly affected. In contrast, it proved impossible to sufficiently convince the general population, serving as a control group in the survey design, of the importance of the study.

Another reason for the poor response rate is that the field work produced a very large number of households that could not be contacted, particularly in the BA sample, where the figure is larger than 35 %. There are various reasons for this. Similarly high numbers in the two reference studies (see footnote 3) provide evidence of problems with addresses – particularly in the area of the municipalities opting to be the sole agency administering the basic income for jobseekers (opting local authorities), for which the BA does not receive regular address updates[4]. The design of PASS intended to address these anticipated difficulties with an additional CAPI field and by putting a lot of effort into address-search. However, the combination of a short field period, an extremely short preparation period for the survey institute and errors in field control ultimately meant that the cases with initially incorrect addresses and/or telephone numbers could be dealt with only in part and for a in some cases very brief field period.

[4] Surprisingly, the selectivity models do, however, show a significant positive effect of data origin from opting local authorities on reachability.

Nevertheless, PASS can at least partly compensate for the relatively high non-response rate by means of an extremely positive data situation, as not only the usual regional and demographic variables used to correct non-response in surveys are available. In addition to these, individual information on unemployment, highest general educational school-leaving certificate, benefit community size and type can be used in the case of the process data sample. In both samples, moreover, micro-geographical information – i.e. information aggregated for a small geographical area, for example on types of buildings, age structure, status, proportion of non-German nationals, volume of moves, etc. – provided by Microm was merged and used for non-response analyses. The chapter on weighting deals with these models (Chapter 5)[5]. Only if the non-response process conditioned on all these variables is not "random", non-response bias will occur. Additionally, reference values are available for some of the variables in the two target populations. These reference values are used for calibrating the panel (see Chapter 5). This removes distortions in relation to these variables with known distributions in the respective population.

c) Agreement to Panel Participation and Mergers

The respondents' consent is required separately to save addresses for the purpose of repeat interviews in the next wave and for merging the survey data with process data from the Federal Employment Agency. In both cases, high consent rates were achieved:

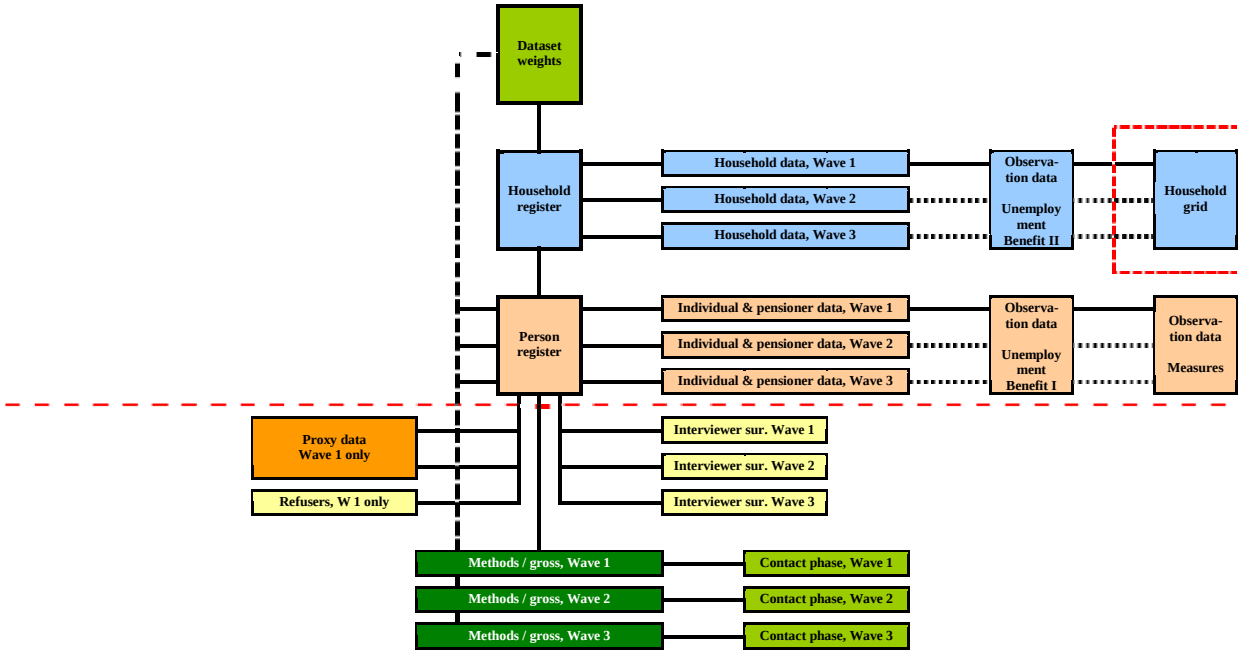
| | |
|---|---|
| Willingness to participate in panel: | 93.8% |
| Consent to merging of process data: | 79.8% |

[5] The model quality achieved, which is already very high– with a maximum McFadden $R^2$=0.1432 for modelling the reachability in the population sample – can be significantly enhanced by taking the variable 'tranche' into account (to 0.29 for reachability in the population sample and 0.069 for reachability in the process data sample). This indicates that particularly in the later tranches, the short field period had a strong influence on reachability. As the tranche ought not to be linked to any relevant variable because it represents an additional partition of the sample, this variable was not, however, used in the non-response analyses. Nevertheless, this result proves that a large part of the non-response process can be explained on the basis of the existing variables.

## 3. The Data Structure in Brief

The usual structure for editing a panel dataset, as used for example in surveys such as the German Socio-economic Panel (GSOEP) or the British Household Panel Survey (BHPS), is to store information on individuals and households in annual, individual datasets. If required, these can be supplemented with special-topic datasets, which might have a cross-wave data Structure. Examples are register or spell data. Figure 3 presents a possible such data structure for PASS.

**Figure 3: Standard Structure for Processing Panel Data**

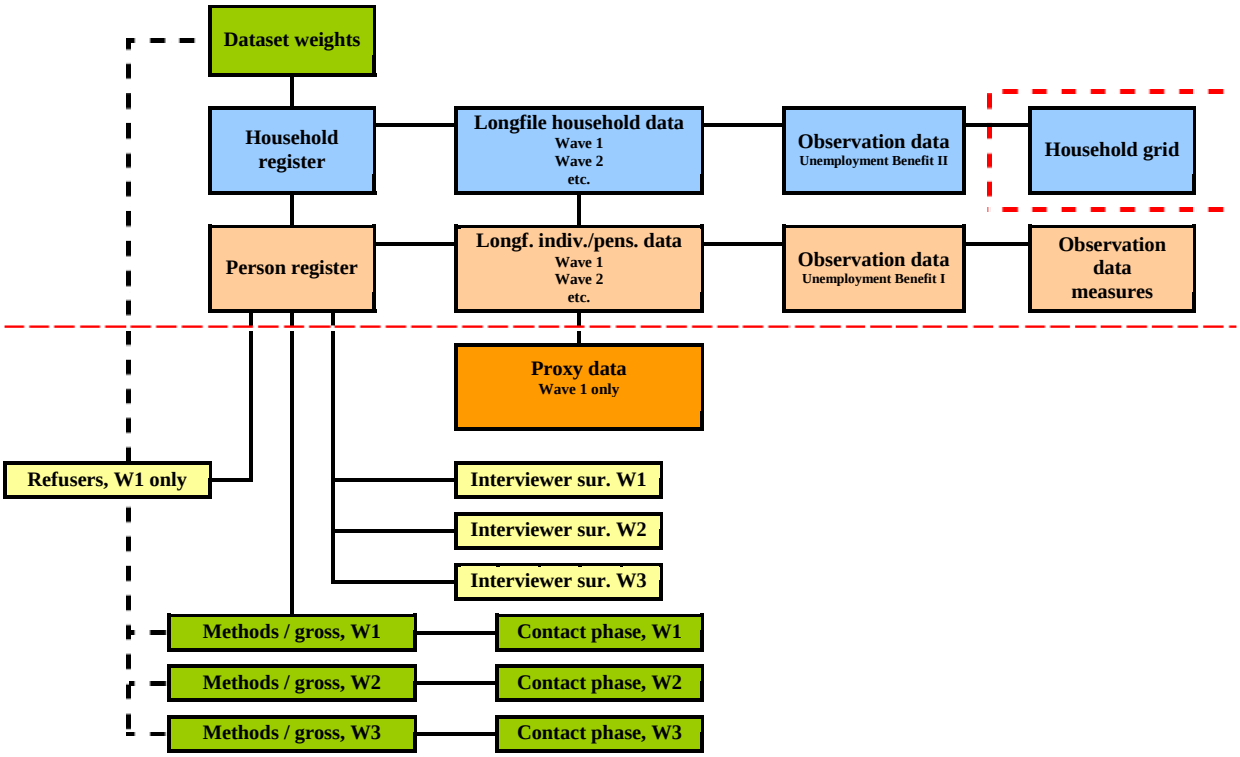



Following a long discussion process and consultations with various experts,[6] however, it was decided to process the data in an alternative structure. The central datasets (household and individual data) are not processed as annual rectangular files, but as cross-wave long files, in which all survey waves allocated to a person/household are written consecutively and identified via a variable for the year (see Figure 4). The advantage of this variant is that the data are already available in the form necessary for typical longitudinal applications such as event history analysis and thus do not require

6 We are particularly grateful to the DIW SOEP Group, especially Prof. Dr. Schupp, Dr. Krause and Dr. Goebel, and Stefan Bender from the FDZ of the BA at the IAB.

complicated processing before starting the analyses. Standard reformatting is also relatively easy to carry out.[7]

**Figure 4: Dataset Structure of the Panel Study Labour Market and Social Security (PASS)**




Until a few years ago, the central argument against applying this type of dataset structure was the significantly higher memory space required, which mainly results from the fact that even variables recorded only in one or a small number of survey waves will always require a complete column across all waves in the dataset. In addition, the long files become relatively large with increasing duration of the panel, simply by appending all annual waves to one-another, which significantly increases the storage space required and the length of individual operations performed using the data. For this reason, applying such a data structure would have been out of the question just a few years ago. In view of the developments in IT and the processors and memory sizes now available on even simple desktop PCs, we are of the opinion that this argument is no longer relevant.

[7] Reformatting to the comparably common wide format, in which the information allocated to one unit is written into different columns of the dataset, can be done automatically using statistics programs such as Stata (using the 'reshape' command). Preparing rectangular files for individual years is also possible via a simple filter instruction.

The only remaining disadvantage is thus the fact that allocating information from the household and individual dataset is made slightly more complicated, as it requires a wave identifier in addition to the actual classification code, for example in the form of a year variable. In comparison to the advantages of the long format, however, this disadvantage appears relatively minor.

Alongside the main datasets, the scientific use file contains the processed data on receipt of Unemployment Benefit I and II and participation in active labour market policy measures in observation form, register datasets on the household and individual level and weighting datasets. In addition, there are also datasets with detailed information on the family relationships in the household (household grid) and several datasets containing technical information (interviewer follow-up, gross datasets, etc.), which are not included in the scientific use file due to their purely technical content or for data protection reasons (these datasets are separated from the others in the diagram by a red dotted line).[8]

[8] Should research projects require access to these partial datasets, we kindly request that researchers contact the research data centre in order to find a suitable access possibility. The specific form of access depends on the nature of the project and the required data.

## 4. Variable Naming Conventions

a) General Issues

There are two main alternatives for naming variables from which we had to choose. The first option is naming the in accordance with their respective order in the questionnaire, as practiced by the GSOEP, for example. The advantage of this type of naming convention is that the items corresponding to the variables are easy to find in the questionnaire, which significantly enhances the value of the questionnaire as a documentation instrument. The central disadvantage of this approach is that identical items are given different names due to changes in the questionnaire, requiring considerable preparation for compiling and renaming the required variables even for simple trend analyses, as more and more panel waves become available.

The second main alternative is allocating independent variable names, which are kept constant across waves (apart from a wave indicator if necessary). The advantages and disadvantages of this strategy are opposite to those of the first alternative: identifying the variables corresponding to an item across waves is simple, whereas using the questionnaire as a documentation instrument becomes more difficult, as it is no longer possible to derive the position of an item in the questionnaire from its variable name.

In our opinion, the advantages of fixed variable names clearly outweigh the disadvantages in a long-term panel study. Moreover, the decision in favour of organising the data in long format as described above requires the use of standard variable names.

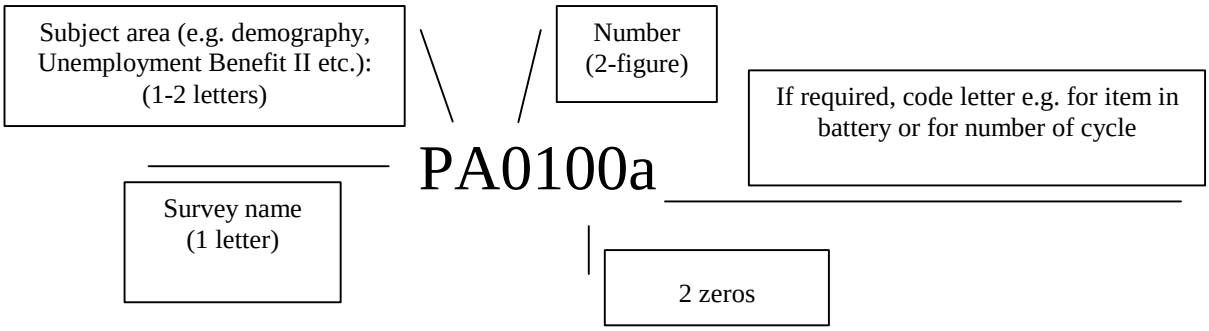
b) Variable Types

The codebook distinguishes between three different types of variables:

1. System variables: system variables are variables created in the course of the survey process. They can be used, firstly, to comprehend the filters documented in the questionnaire. At least some of the system variables can also be of interest from a content or methodological point of view, for example the interview mode or the number of children in a certain age group living in the household. System variables have been allocated individual names, for which lower-case letters and figures can be combined. The system variables also include the weights.

2. Surveyed variables: surveyed variables are variables that were collected as part of the questionnaire. These variables have been allocated entirely new, abstract variable names. The concept behind this naming process is illustrated in Figure 5, using an example.

**Figure 5: The Variable Naming Scheme**




- The first letter of the variable name indicates the questionnaire level, i.e. household or individual dataset, by means of the letter H or P (upper-case).
- The following one or two upper-case letters, indicate the subject area the variable refers to (see Table 4 for a complete list).
- In the spell-datasets, there is no introductory P or H. Instead, the variables in these datasets are given a standard subject-based name consisting of two or three letters or two letters and one number.
- The introductory letter combination is then followed by two consecutively allocated numbers, which indicate the number of the question within the subject area.
- These two numbers are followed by two zeros, which are intended to enable the addition of further variables in later waves. To date, this option has only been used in cases where a second variant including coded information from an open-ended survey question or response category has been made available in addition to the original version of the variable. The final zero has been adjusted to a '1' for these variables (e.g. PA0101a rather than PA0100a).

**Table 4: List of Subject-Related Indicators Used in Variable Names**

| Individual Level | | Household Level | |
|---|---|---|---|
| **Code** | **Subject Area** | **Code** | **Subject Area** |
| PA | General | HA | General |
| PAS | Job-search | HD | Demography |
| PB | Education | HEK | Income |
| PD | Demography | HKI | Childcare |
| PEO | Attitudes and orientations | HLS | Standard of living |
| PEK | Income | HW | Housing |
| PET | Employment | | |
| PG | Health | | |
| PLS | Standard of living | | |
| PMI | Migration | | |
| PP | Care | | |
| PSH | Social origin | | |
| PSK | Social relations | | |
| PTK | Contact to social security institutions | | |
| AL | Receipt of Unemployment Benefit I (ALG 1) (spell data, individual level) | AL2 | Receipt of Unemployment Benefit II (ALG 2) (spell data, household level) |
| ALM | Active labour market policy measures (spell data, individual level) | | |

- In the case of variables for items from multi-item batteries or in a looped sequence of questions, a further lower-case letter may be added for identifying the item or the current cycle within the loop.

3. Generated variables: the generated variables are sub-divided into two further groups. The generated variables in a strict sense are aggregated from various other variables, e.g. from open and categorical income measures, or they are even more complex constructs such as equivalized household income or classifications for education (such as ISCED or Casmin) or status (e.g. EGP, ESEC). Generated variables in this strict sense are allocated individual names that are as clear and memorable as possible, in lower-case letters. For an overview of the generated variables, see Chapter 6.

The second group of generated variables includes those in which information from open ended survey questions or response categories were added to another (closed) variable. Although these variables are, strictly speaking, also generated variables and are classified as such in the frequency tables of the codebook, they have not been given individual names. Instead their names are based on those of the original variable, however with a ‘1’ as the final number rather than a ‘0’.

## 5. Documentation of Data Cleansing

Data cleansing took place in two stages. The first step was a detailed check of the filter questions (applying corrections, if necessary), Moreover, setting specific codes for missing values were defined. Then, selected items were checked for plausibility of the information provided by the respondents. Clearly implausible or contradictory responses were indicated as such by a specific missing code. However, such, corrections of the data were carried out on a very restrictive basis. In addition, problems with one of the foreign language versions of the survey instrument (Russian) made it necessary to repeat parts of the interview for this population. The integration of these data was also carried out during the data correction process.

a) Filter Checks

During the filter checks, the correct operation of the filter questions in the instrument was checked in the two datasets supplied by the field institute (one set each for the data from the personal and household questionnaires) using a statistics program. In cases where certain items had been filled in although the value of the relevant filter variable would have required a different sequence of questions (for example, if detailed information was recorded on vocational training although the respondent had stated that he or she did not have any vocational qualification), these variables were set to the missing code '-3' (not applicable), which they would also have received through correct use of the filters.[9] Moreover, there were incidents of items not being surveyed, although they ought to have been according to the relevant filter variable (e.g. if no further information was recorded on vocational education although the respondent had stated that he or she had undergone such training). In these cases, the specific missing code '-4' (question mistakenly not asked) was allocated.

In addition, the missing codes and system missings allocated by the field institute were replaced with standard values for all variables. Table 5 presents an overview of the allocated values. '-1' and '-2' are the standard recoding for the values 'Don’t know' and 'Refused' recorded during the survey. '-3' is the general 'Not applicable' code for questions not asked due to filters. '-4' is described above. '-5' to '-7' are question-specific

codes. These can be both specific missing codes (e.g. "Not applicable, not available for the labour market"), or special categories for valid values (e.g. a category for income above € 99,999 in the open question on income). These codes were only allocated as required.

**Table 5: Overview of Missing Codes**

| Code | Explanation |
|---|---|
| -1 | "Don't know" |
| -2 | "Refused" |
| -3 | "Not applicable (filter)" (question not asked due to filter) |
| -4 | "Question mistakenly not asked" (question should have been asked) |
| -5 | Question-specific code 1, allocated as required |
| -6 | Question-specific code 2, allocated as required |
| -7 | Question-specific code 3, allocated as required |
| -8 | "Implausible value" |
| -9 | "Item not administered in wave" |
| -10 | "Item not administered in questionnaire version" |

The value '-8' is a specific missing code allocated during plausibility checks; for a more detailed explanation, see the section below. '-9' is a missing code that is only required from the second wave on. It is always allocated if a certain item is not surveyed in a specific wave. Due to the long format of the dataset described above, in the data roll-out for the second wave, items that were not administered anymore in that year's survey would receive the value '-9' for wave 2. In just the same manner, new items would be retroactively coded '-9' for wave 1. The code '-10' can be used to take account of differences between questionnaire versions. In general, this code is allocated to cases where the respondents were interviewed using the senior citizens' questionnaire, and the respective question was included in the standard questionnaire but not in the shorter senior citizens' version.

[9] The correction took place in the standard questionnaire order, considering the (filter-)items asked first to be correct and correcting those items asked at a later point.

b) Plausibility Check

For the plausibility check, a comprehensive list of possible contradictions in the respondents' statements was checked. Information on the household structure was also checked for plausibility and the spell data were cleansed. 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 strong and/or the alteration could be defined as relatively minor. The latter applied, for example, if only a low number of cases were affected or one missing code (e.g. '-3') was simply replaced by another ('-8'). Two strategies were used to adjust implausible statements: direct correction of the implausible responses or allocation of a specific missing code.

   - Implausible responses were only corrected if it was extremely likely that the interviewer had entered information incorrectly. An example is a statement of a monthly gross income of € 22,000 with a simultaneous net income (plausible on the basis of occupation) of € 1,750 per month. In this case, for example, the last '0' of the gross income was deleted. Similar cases were income statements of € 99,999 or € 99,998. Here, the data correction process assumed that the six-figure missing codes '999998' (DK) or 999999 (Refused) had been wrongly entered, and the data were corrected correspondingly.

   - However, it was rather seldom the case during the checks on implausibilities that a value could be identified as an error with sufficient certainty. In most cases, it was only possible to establish a contradiction between two statements, but not to identify specific errors that had led to the implausibility. In these cases, no corrections were made, and the specific missing value code '-8' was allocated instead. Which of the variables involved in the contradiction the code was allocated to was decided on an individual basis.

2. Plausibility check of the household structure: this check was carried out on the basis of the information recorded in the household interview on the family relationships among household members, and the information on age, gender and first names. Before carrying out the check, partnership information was supplemented by information on partners living in the household reported during the personal interview. In the first part of the checks, the relationship information and the demographic data

on the individual household members were combined to identify implausible household structures. For the households identified as implausible in these checks, individual case decisions were made, on the basis of the entire household structure and other interview information (e.g. on marital status in the personal interview). Implausible relationships were marked as such ('-8') or corrected using additional information on the household context, if it was plausible that an error had occurred. One example: in the case of two persons of the same gender who were both natural parents of a third person in the household, the gender was corrected on the basis of the first name. If the forenames also indicated that both persons were of the same gender, and if no further relevant information was available, the family relationship was identified as implausible on the basis of the household structure. Additional checks were carried out, comparing sets of three family relationships with one another for plausibility. One example of a relationship classified as implausible in this check would be: person A is the spouse of person B. Person B is the natural child of person C. Person A is the natural child of person C. Cases identified as such or similar combinations of relationships during this check were marked as implausible responses ('-8').

3. During the check of the spell data, the correct chronological order of the spells was verified. If they had been mistakenly recorded in incorrect order during the survey, they were put back into chronological order. In some cases it occurred, that specific spells had been recorded twice. If two completely identical spells were available for a single person or household, one of these two was deleted.

c) Re-surveying the Russian CATI Field

The Russian version of the CATI programming had to be programmed anew using special software, as the Cyrillic characters were not compatible with the software used for the German, Turkish and English instrument. Thus, in contrast to the aforementioned other foreign-language instruments, it was not possible to retain the German version of the CATI program as a basis and simply translate and exchange the question texts.

In the course of this new programming process, errors in the filters or the generation of control variables occurred that affected a total of 294 of the 432 persons interviewed with the Russian instrument. This necessitated subsequent re-surveying of several modules of the personal questionnaire, particularly the questions on contacts with so-

cial security institutions (P76-P90), participation in labour-market policy measures (P93-P115) and job-search (P116-P138).

Out of these 294 cases, a total of 202 persons were available for subsequent interviews on the aforementioned modules. Each person was only interviewed on the module or modules that were mistakenly not included in the first, regular interview. These cases are identified with the variable "nachbef" in the personal dataset, which also states which (combination of) modules were subsequently administered and integrated into the dataset. In addition, those variables for which questions P93 to P115 were administered anew can be identified in the spell data for measures by the variable "nachbmas".

## 6. Generated Variables

a) Variables Including Responses to Open-Ended Survey Questions

Some items of the survey were gathered as closed items with an open residual category. In such cases, additional variables were generated which differed from the original variable only insofar as the information from the open-ended responses was coded to the respective category where possible. Moreover, in some cases new categories were created on the basis of the open-ended information. The naming of these additional variables differs from the original in the last digit only, where the '0' was replaced by a '1'. The following variables belong to this type:

| | |
|---|---|
| *Individual level:* | PB0401; PB1001; PB1001a – PB1301j; PB1601; PG0901a – PG0901g; PSH0201; PSH0301a – PSH0301i; PSH0501; PSH0601a – PSH0601i; ozulanda – ozulandf |
| *Household level:* | AL21301a-AL22203a, AL21301b-AL22203b, AL21301c-AL22203c, AL21301d-AL22203d, AL21301e-AL22203e |

b) Simple Re-codings or Aggregations

Variables of this type either contain information aggregated from particular items of a single construct (such as highest educational qualification) which had been collected separately for technical reasons, or for which additional information was merged from other partial datasets (e.g. indicators for ongoing receipt of Unemployment Benefit I or II) in completion of the individual and household data already included. These variables are shown in Tables 6 and 7.

**Table 6: Simple Re-codings or Aggregations on the Individual Level**

| Variable | Variable Label and Description | Source Variables |
|---|---|---|
| aktmassn | *Current participation in a measure funded/ promoted by the employment agency, generated*<br>Indicator for current participation in a measure of active labour market policy. | ALM0400/censored (massnahmespells) (p96) |
| alg1abez | *Ongoing receipt of Unemployment Benefit I, generated*<br>Indicator for ongoing receipt of Unemployment Benefit I. | censored (ALG1_Spells) (p72; p73) |
| arbzeit | *Weekly hours of work incl. indications on irregular working hours, generated*<br>Generation of an integrated variable from open-ended and categorized information on working hours. | PET1300/PET1400 (p36/p37) |
| beruf1 | *Highest educational qualification, excl. foreign qualifications and information from open ended survey questions, generated*<br>Identification of the highest educational attainment by hierarchising occupational qualifications attained by the respondents; excluding *information from open ended survey questions*. | PB1200; PB13000a-PB1300l (p15; p16_01-p16ka) |
| beruf2 | *Highest educational qualification, incl. foreign qualifications and information from open ended survey questions, generated*<br>Similar to 'beruf1', but additionally:<br>1. Inclusion of *information from open ended survey questions*;<br>2. Inclusion of information on foreign qualifications;<br>3. Degrees are not allocated to the school type (e.g. university or other institution of higher education) but to the qualification level (Bachelor's degree; Master's degree; Ph.D.). | PB1200; PB13000a-PB1300l; P1400; P1500a-P1500e (p15; p16_01- p16ka; p17; p18_01- p18ka) |
| brutto | *Gross income incl. categorised information, generated*<br>Generation of a variable integrating *information from* categorised and *open ended survey questions* on the household gross income. | PEK0200-PEK0600 (p56-p61) |
| bruttokat | *Categorised gross income, generated*<br>Aggregation of the categorised information on gross income, combined from several items on income categories. | PEK0100-PEK0600 (p57-p61) |
| erwerb | *Employment status, generated*<br>Generation of an integrated employment status variable including indications on employment and status of unemployment. | PET0100; PET0300; PET0400; PET0800 (p24, p26, p27, p28, p31) |
| famstand | *Marital status, generated*<br>Generation of an integrated marital status variable combining information from individual and household questionnaires. | PD0500; epartner (=variable taken over from the household questionnaire; 1=spouse lives in the same household)<br>(p177; epartner) |
| gebhalbj | *1st/ 2nd half-year of birth, generated*<br>Indicator: Date of birth is in the 1st or 2nd half of the year | PD0100 (p1) |

| Variable | Variable Label and Description | Source Variables |
|---|---|---|
| | of birth. | |
| kindzges | *Total number of own children (living in and outside the household), generated*<br>This variable indicates the total number of children of the respondent including children living in his/her own household as well as children living in another household. | HD0100x (interviewed household member); G0600xy / G0800xy (other household members); PD0900-PD1100<br>(hh28, hh33, hh35, p183-p185) |
| kindzihh | *Number of children living in the household, generated*<br>Variable generated on the basis of the responses in the household questionnaire containing the number of children of a person who live in the person's household (Total number of persons in the household (half-) matrix who are counted as children of the respondent plus the number of persons in the household (half-) matrix for whom the respondent is classified to be a parent).<br>N.B.: This variable is unusual in it relates to each individual person, which must be borne in mind when using it. This is to say that a child who lives together with his/her parents in the same household is counted as 'child in household' both for the father and the mother. Aggregating this variable over the household members will therefore not produce any meaningful results. | HD0100x (interviewed household member); G0600xy / G0800xy (other household members)<br>(hh28, hh33, hh35) |
| massntyp | *Measure: type of measure incl. 'one-euro-jobs', generated*<br>This variable is a combination of the question as to whether a measure is considered a 'one-euro-job'(p102) and the re-coding of open-ended information on the type of labour market policy measure (except for 'one-euro-jobs') the respondent participated in. | ALM1000, ALM 1500 (p102, p107) |
| mberuf1 | *Highest vocational qualification of the mother, incl. mothers in the household, excl. information from open ended survey questions, gen.*<br>The question on vocational qualification was administered only for mothers who were not living in the survey household. In case she was living in the household, her vocational qualification had to be adapted from the information she provided in her personal interview. | PSH0300a-l; beruf1 of mother (p205_01-p205_99; beruf1/mother) |
| mberuf2 | *Highest vocational qualification of the mother, incl. mothers in the household, incl. information from open ended survey questions, gen.*<br>The vocational qualification was administered only for mothers who were not living in the survey household. In case she was living in the household, her vocational qualification had to be adapted from the information she provided in her personal interview. | PSH0300a-l; beruf2 of mother (p205_01-p205_99; beruf2/ mother) |
| migration | *Respondents' background of migration, generated*<br>Generated variable to create four categories of migration background: no migration background; personal migration (1st generation); migration of at least one parent with no personal migration of the respondent (2nd generation); migration of at least one grandparent with no personal migration of neither the respondent | PMI0100, PMI0700-PMI0900f<br>(p191; p198-p200f) |

| Variable | Variable Label and Description | Source Variables |
|---|---|---|
| | nor one of his/her parents (3rd generation). | |
| mschul1 | *Highest general education school leaving certificate of the mother, incl. mothers in the household, excl. information from open ended survey questions, gen.*<br>The question on the school leaving certificate was administered only for mothers who were not living in the survey household. In the case she was living in the household, information on her school leaving certificate had to be adapted from the information she provided in her personal interview. | PSH0200; schul1 of mother<br>(p204; schul1/mother) |
| mschul2 | *Highest general education school leaving certificate of the mother, incl. mothers in the household, incl. indications in plain language, gen.*<br>The question on the school leaving certificate was administered only for mothers who were not living in the survey household. In case she was living in the household, information on her school leaving certificate had to be adapted from the information she provided in her personal interview. | PSH0200; schul2 of mother<br>(p204; schul2/mother) |
| netto | *Net income incl. categorised* information*, generated*<br>Generation of a variable integrating *information from* categorised and *open ended survey questions* on the respondent's net income. | PEK0700-PEK1200<br>p62-p67 |
| nettokat | *Categorised net income, generated*<br>Aggregation of the categorised information on the respondent's net income, combined from several items on income categories. | PEK0800-PEK1200<br>p63-p67 |
| ogebland | *Native country, incl. information from open ended survey questions, categories (anonymized)*<br>Variable combining the items 'native country = Germany' and *information from the open ended survey question* on the native country (other than Germany) | PMI0100; PMI0200<br>(p191; p192) |
| ostaatan | *Nationality, incl. information from open ended survey questions, categories (anonymized)*<br>Variable combining the items 'nationality = German' and *information from the open ended survey question* on the nationality (other than Germany). | PMI0400; PMI0500<br>(p195; p196) |
| palter | *Age (from p1), generated*<br>Age variable generated from the date of birth. | PD0100 (p1) |
| pintjahr | *Date of personal interview – year, generated*<br>Year of interview. | pintdat |
| pintmon | *Date of personal interview – month, generated*<br>Month of interview. | pintdat |
| pinttag | *Date of personal interview – day, generated*<br>Day of interview. | pintdat |
| schul1 | *Highest general education school leaving certificate, excl. foreign qualifications and information from open ended survey questions*<br>This variable indicates the highest school-leaving qualification; equivalent qualifications in East and West | PB0300 - PB0500<br>(p5-p7) |

| Variable | Variable Label and Description | Source Variables |
|---|---|---|
| | Germany were combined (e.g. 'EOS' and 'Abitur' ); excl. *information from open ended survey questions*. | |
| schul2 | *Highest general education school leaving certificate, incl. foreign qualifications and information from open ended survey questions*<br>Corresponds to 'schul1' with additional information on the educational qualification *from open ended survey questions*. | PB0300 - PB0500; PB1000; PB1100 (p5-p7; p13, p14) |
| stib | *Occupational status, detailed coding, generated*<br>Generation of a detailed code number of occupational status generated on the basis of various items gathering detailed information on occupational status. | PET1600-PET2200 (p39-p45) |
| vberuf1 | *Highest vocational qualification of the father, incl. fathers in the household, excl. information from open ended survey questions, gen.*<br>The question on vocational qualification was administered only for fathers who were not living in the survey household. In case he was living in the household, his vocational qualification had to be adapted from the information he provided in his personal interview. | PSH0600a-l; beruf1 of father<br>(p208_01-p208_99; beruf1/ father) |
| vberuf2 | *Highest vocational qualification of the father, incl. fathers in the household, incl. information from open ended survey questions, gen.*<br>The question on vocational qualification was administered only for fathers who were not living in the survey household. In case he was living in the household, his vocational qualification had to be adapted from the information he provided in his personal interview. | PSH0600a-l; beruf2 of father<br>(p208_01-p208_99; beruf2/ father) |
| vschul1 | *Highest general education school leaving certificate of the father, incl. fathers in the household, excl. information from open ended survey questions, gen.*<br>The question on the school leaving certificate was administered only for fathers who were not living in the survey household. In case he was living in the household, information on his vocational qualification had to be adapted from the information he had provided in his personal interview. | PSH0500; schul1 of father<br>(p207; schul1/ father) |
| vschul2 | *Highest general education school leaving certificate of the father, incl. fathers in the household, information from open ended survey questions, gen.*<br>The question on the school leaving certificate was administered only for fathers who were not living in the survey household. In case he was living in the household, information on his vocational qualification had to be adapted from information the indications he had provided in his personal interview. | PSH0500; schul2 of father<br>(p207; schul2/ father) |
| zmhh | *Indicator: Person-ID of the respondent's mother*<br>Indicates the individual *Person*-ID of a respondent's mother, if she lives in the same household. Mothers comprise: natural mothers, stepmothers, adoptive or foster mothers and mothers with a status not further specified. | HD0200a-HD0200o; household grid table<br>(HH26, HH28-HH32; HH33-HH37) |

| Variable | Variable Label and Description | Source Variables |
|---|---|---|
| zparthh | *Indicator: Person-ID of the respondent's partner in the household in wave 1 (2006/2007)*<br>Indicates the *Person*-ID of a respondent's partner living in the same household. Partners comprise: spouses, partner in a civil union, unmarried partners and partners with a status not further specified. | HD0200a-HD0200o; household grid table (HH26, HH28-HH32; HH33-HH37) |
| zvhh | *Indicator: individual ID of the respondent's father*<br>Indicates the individual ID of a respondent's father, if he lives in the same household. Fathers comprise: natural fathers, stepfathers, adoptive or foster fathers and fathers with a status not further specified. | household grid table (HH28-HH32; HH33-HH37) |

**Table 7: Simple Re-codings or Aggregations on the Household Level**

| Variable | Variable Label and Description | Source variables |
|---|---|---|
| alg2abez | *Current receipt of Unemployment Benefit II, generated*<br>The variable indicates that a household is currently receiving Unemployment Benefit II. | censored (ALG2_Spells) (hh43; hh44) |
| blneualt | *West German States or East German States, generated*<br>Aggregation of German states into the West German States of the former FRG (as it existed until the German Reunification in 1990, without Berlin) and the East German States of the former GDR (including Berlin). | bundesld |
| hhinckat | *Categorised household income, generated*<br>Categorised information on the household's net income combined from several items on income categories. | HEK0700-HEK1100 (hh61-hh65) |
| hhincome | *Household income incl. categorised information, generated*<br>Generation of a variable integrating information from categorised and open-ended survey questions on the household's net income. | HEK0600-HEK1100 (hh60-hh65) |
| hintjahr | *Date of household interview – year, generated*<br>Year of interview. | hintdat |
| hintmon | *Date of household interview – month, generated*<br>Month of interview. | hintdat |
| hinttag | *Date of household interview – day, generated*<br>Day of interview. | hintdat |

**Table 8: Simple Re-codings or Aggregations of Other Available Datasets**

| Variable | Dataset | Variable label and description | Source variables |
|---|---|---|---|
| bmonat<br>bjahr<br>emonat<br>ejahr | alg1_spells<br>alg2_spells<br>observations<br>on measures | Converted start and end dates of individual spells in the spell data. Where only information on the season during which the spell started or ended was available, these have been replaced with estimated values for the month. In case of current receipt of benefits / participation in measures, censoring information has been replaced with the interview-date. In the spell- data for participation in measures of active labour market policies the end date was calculated on the basis of the start date and information on the duration of receipt of benefits. | AL10100 - AL10400<br>AL20100 - AL20400<br>ALM0200,<br>ALM0300,<br>ALM0400,<br>ALM0500, ALM0800 |
| zensiert | alg1_spells<br>alg2_spells<br>observations<br>on measures | This variable indicates whether a spell was still ongoing at the interview-date of the latest wave, i.e. the observation is right-censored.<br>In the datasets on receipt of Unemployment Benefit I and II, this variable was coded as one if respondents either answered that they still received benefits when asked about the end-date of the spell or, when indicating the current month as the spell's end date, answered in a second question, that they still received benefits.<br>In the spell data on labour-market measures this variable includes recoded answers to the question as to whether a person is currently participating in any such programme, with missing values being interpreted as non-participations. | AL10300-AL10500<br>AL20300-AL20500<br>ALM0400 |

## c) Theory-Based Constructed Variables

Theory-based constructed variables are variables which require extensive re-coding and/or coding to be generated. In most cases these variables have been empirically tested and have a foundation in theoretical concepts. Moreover, at least some of these are standardised instruments used in social sciences or economics. Examples for such standardised instruments are the European Socio-economic Classification (ESeC), the International Standard Classification of Education (ISCED) or equivalised household income. This chapter provides a detailed description of all theory-based constructed variables provided in the PASS-data and a short overview on their theoretical background and the most important references.

### Variables on the Individual Level

*Education in years*

| | |
|---|---|
| Variable name | bilzeit |
| Variable label | Duration of school education and vocational training in years, generated |
| Source variable | schul2 beruf2 |
| Category / dataset | Education / person data |
| Prepared by | Bernhard Christoph |
| Description | For many statistical models using a linear variable for education is more appropriate than using a categorical one. For schooling levels, converting the categorical information to a linear one is fairly simple. The linear value corresponds to the time spent in school until attainment of the final school leaving qualification. The only particularity to be considered is that equivalent qualifications should be assigned to identical periods. An upper secondary school leaving certificate, for example, should always be labelled with the same duration, independent of whether it had been attained after twelve or thirteen years of education. For constructing this variable, secondary school leaving certificate were assigned to the following education periods for this variable: |

| | |
|---|---|
| Lower secondary school leaving certificate;<br>Lower secondary school leaving certificate from the former GDR (POS) after completion of grade 8;<br>Other lower secondary school leaving certificate: | 9 years |
| Intermediate secondary school leaving certificate;<br>Intermediate secondary school leaving certificate from the former GDR (POS) after completion of grade 10: | 10 years |
| University of Applied Sciences entrance qualification: | 12 years |
| General qualification for university entrance or subject-specific higher education entrance qualification (incl. EOS - comparable qualification in the former GDR) | 13 years |

This is, however, different for vocational qualifications. Due to the numerous possibilities to obtain a vocational degree and the potentially huge differences in income even degrees with a comparable period of education might result in, the required training period may not be subject to a simple one-to-one conversion process as in the case of school leaving certificates. To avoid this problem, it may be attempted to estimate the growth in human capital related to a certain vocational qualification. (cf. e.g. Helberger 1988).

This study uses such an approach. For the conversion process, only the highest vocational qualification of each respondent was considered and the years estimated to represent human capital growth entailed by this qualification were added to the years of secondary school education.

| | |
|---|---|
| Training as a semi-skilled worker: | +1 years |
| Apprenticeship, vocational school, school for health care | |

| | |
|---|---|
| | professionals: +1.5 years<br>Master craftsmen's degree: +3 years<br>Professional college: +3 years<br>University of Applied Sciences/Bachelor: +3 years<br>University/Master: +5 years<br>Ph.D.: +8 years<br>Other German qualification: +1.5 years<br>Other foreign qualification: +1.5 years |
| Reference | 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 / person data |
| Prepared by | Bernhard Christoph |
| Description | General description: see 'education in years' |
| | When generating the variable parents' years of education the values added for vocational degrees differ from those used when constructing the corresponding variable for the respondents. This is so, since information vocational education was collected in less detail for the parents (especially as far as tertiary education is concerned). The values corresponding to particular courses of education are the following: |
| | Training as a semi-skilled worker: +1 years<br>Apprenticeship, vocational school, school for health care professionals: +1.5 years<br>Master craftsmen's degree: +3 years<br>Vocational academy: +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 |
| Reference | Helberger (1988) |

*Education in years, father*

| | |
|---|---|
| Variable name | vbilzeit |
| Variable label | Duration of school education and vocational training in years, generated |
| Category/ dataset | Education / person data |
| Category | Qualification |
| Prepared by | Bernhard Christoph |
| Description | When generating the variable parents' years of education the values added for vocational degrees differ from those used when constructing the corresponding variable for the respondents. This is so, since information vocational education was collected in less detail for the parents (especially as far as tertiary education is concerned). The values corresponding to particular courses of education are the following: |
| | Training as a semi-skilled worker: +1 years<br>Apprenticeship, vocational school, school for health care professionals: +1.5 years<br>Master craftsmen's degree: +3 years<br>Vocational academy: +3 years |

| | | |
|---|---|---|
| | University of Applied Sciences: | +3 years |
| | University: | +5 years |
| | Other German qualification: | +1.5 years |
| | Other foreign qualification: | +1.5 years |
| Reference | Helberger (1988) | |

## *CASMIN*

| | |
|---|---|
| Variable name | casmin |
| Variable label | Education classified acc. to CASMIN, updated version, generated |
| Source variable | schul2 beruf2 |
| Category/ dataset | Education / person data |
| Prepared by | Bernhard Christoph |
| Description | 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 available (Brauns & Steinmann 1999).<br>The procedures for re-coding qualifications acc. to CASMIN applied in the panel follow the descriptions of Lechert et al. (2006) and Granato (2000), especially for problematic cases, not neglecting, of course, the slightly differing category-values of the education variable in this dataset. Details can be found in the table below. Cells containing valid combinations according to CASMIN are highlighted in light grey, such containing defined missing values are dark grey. |

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

| | |
|---|---|
| References | Brauns et al. (1999); Granato (2000); König et al. (1987); Lechert et al. (2006) |

*MCASMIN*

| | |
|---|---|
| Variable name | mcasmin |
| Variable label | Education of the mother classified acc. to CASMIN, updated version, generated |
| Source variable | mschul2 mberuf2 |
| Category/ dataset | Education / person data |
| Prepared by | Bernhard Christoph |
| Description | General description: see CASMIN |
| | Since the education variable has different category-values for respondents and their parents, the coding pattern of 'mcasmin' and 'vcasmin' slightly differs from the 'casmin' pattern. The following table shows the differences in detail. |

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

| | |
|---|---|
| References | Brauns et al. (1999); Granato (2000); König et al. (1987); Lechert et al. (2006) |

## *VCASMIN*

| | |
|---|---|
| Variable name | vcasmin |
| Variable label | Education of the father classified acc. to CASMIN, updated version, generated |
| Source variable | vschul2 vberuf2 |
| Category/ dataset | Education / person data |
| Prepared by | Bernhard Christoph |
| Description | General description: see CASMIN |
| | Since the education variable has different category-values for respondents and their parents, the coding pattern of 'mcasmin' and 'vcasmin' slightly differs from the 'casmin' pattern. The following table shows the differences in detail. |

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

| | |
|---|---|
| References | Brauns et al. (1999); Granato (2000); König et al. (1987); Lechert et al. (2006) |

## *ISCED 97*

| | |
|---|---|
| Variable name | isced |
| Variable label | Education classified acc. to ISCED-97, updated version, generated |
| Source variables | mschul2 mberuf2 |
| Category/ dataset | Education / person data |
| Prepared by | Bernhard Christoph |
| Description | ISCED-97 (International Standard Classification of Education) developed by the OECD (OECD 1999, for an outline, see also BMBF 2003) is an education qualification which can be used alternatively to CASMIN.<br>What is particular about the ISCED-97 classification is that it embodies 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 of age. Instead a separate group of persons with an education below ISCED level 2 (ISCED 2 = lower or intermediate secondary school leaving certificate) was generated. Therefore, only ISCED levels 2 to 6 are considered in the coding procedure applied in this dataset.<br>Coding details are shown in the table below. Cells containing valid combinations according to ISCED are highlighted in light grey, such containing defined missing values are dark grey. |

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

| | |
|---|---|
| References | BMBF (2003); OECD (1999) |

## *MISCED 97*

| | |
|---|---|
| Variable name | misced |
| Variable label | Education of the mother classified acc. to ISCED-97, updated version, generated |
| Source variables | mschul2 mberuf2 |
| Typ / Datensatz | Education / person data |
| Prepared by | Bernhard Christoph |
| Description | For the theoretical background and details please refer to ISCED-97.<br>In contrast to the re-coding applied to data respondents' education in terms of ISCED-97,, it is not possible generate ISCED level 6 for data on their parents. This is so, since data on the respective qualification (i.e. PhD or equivalent) had not been collected for the parents. Therefore only ISCED levels 2 to 5 are considered in the procedure applied in this dataset. The following table shows the coding details. |

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

| | |
|---|---|
| References | BMBF (2003); OECD (1999) |

## *VISCED 97*

| | |
|---|---|
| Variable name | visced |
| Variable label | Education of the father classified acc. to ISCED-97, updated version, generated |
| Source variable | vschul2 vberuf2 |
| Category/ dataset | Education / person data |
| Prepared by | Bernhard Christoph |
| Description | For the theoretical background and details please refer to ISCED-97.<br>In contrast to the re-coding applied to data respondents' education in terms of ISCED-97,, it is not possible generate ISCED level 6 for data on their parents. This is so, since data on the respective qualification (i.e. PhD or equivalent) had not been collected for the parents. Therefore only ISCED levels 2 to 5 are considered in the procedure applied in this dataset. The following table shows the coding details. |

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

| | |
|---|---|
| References | BMBF (2003); OECD (1999) |

## *International Standard Classification of Occupations 1988 (ISCO88); ZUMA coding*

| | |
|---|---|
| Variable name | isco88 |
| Variable label | ISCO 88 (ZUMA coding), generated |
| Source variable | P46 |
| Category/ dataset | Occupation / person data |
| Prepared by | Bernhard Christoph |
| Description | The International Standard Classification of Occupations (ISCO) was developed by the International Labour Organization (ILO), allowing to classify occupations in an internationally comparable manner. The ISCO-88's most notable particularity is that in addition to the job performed, the qualification level necessary to do so is essential for assigning an occupation to a particular occupational code. This is also a mayor difference between the ISCO-88 and the Classification of Occupations provided by the German Federal Statistical Office (KldB), which is also provided in this dataset.<br>The actual coding was carried out by the Centre for Survey Research and Methodology (Gesis-ZUMA). In contrast to the coding variant applied by TNS Infratest, ZUMA's ISCO-88 is generated using a coding procedure using the original information from the open ended job-descriptions and is not derived from the KldB codes. The ISCO-based measures of occupational status and prestige provided in this dataset are based on the ZUMA coding.<br>A list of ISCO 88 occupations and the respective values of occupational status and prestige scales are given in Table A 3 on page 623 pp. |
| Reference | ILO (1990) |

## *International Standard Classification of Occupations 1988 (ISCO88); TNS Infratest Coding*

| | |
|---|---|
| Variable name | isco88it |
| Variable label | ISCO 88 (TNS Infratest coding), generated |
| Source variable | P46 |
| Category/ dataset | Occupation / person data |
| Prepared by | Bernhard Christoph |
| Description | The International Standard Classification of Occupations (ISCO) was developed by the International Labour Organization (ILO), allowing to classify occupations in an internationally comparable manner. The ISCO-88's most notable particularity is that in addition to the job performed, the qualification level necessary to do so is essential for assigning an occupation to a particular occupational code. This is also a mayor difference between the ISCO-88 and the Classification of Occupations provided by the German Federal Statistical Office (KldB), which is also provided in this dataset.<br>Coding of this variable was carried out by TNS Infratest, the PASS's field institute, using a procedure for deriving ISCO88 codes from the German Federal Statistical Office's Classification of Occupations.<br>A list of ISCO 88 occupations and the respective values of occupational status and prestige scales are given in Table A 3 on page 623 pp. |
| Reference | ILO (1990) |

## *Classification of Occupations 1992 (KldB92); TNS Infratest Coding*

| | |
|---|---|
| Variable name | kldb_it |
| Variable label | Classification of Occupations 1992 (TNS Infratest coding), generated |
| Source variable | P46 |
| Category/ dataset | Occupation / person data |
| Prepared by | Bernhard Christoph |
| Description | The KldB92 is the latest version of the Classification of Occupations published by the German Federal Statistical Office. It is a coding system that was specifically constructed to match the particularities of the German occupational structure. It is only based on job descriptions.<br>Coding of this variable was carried out by TNS Infratest Sozialforschung, the PASS's field institute. |
| Reference | StBA (1992). |

## *Coding of Occupations, Problem Indicator (TNS Infratest Coding)*

| | |
|---|---|
| Variable name | berpr_it |
| Variable label | Coding of occupations, problem indicator (TNS Infratest coding), generated |
| Source variable | P46 |
| Category/ dataset | Occupation / person data |
| Prepared by | - |
| Description | Indicator developed by TNS Infratest to assess the coding of occupations according to KldB 1992. |
| Reference | |

*Class scheme according to Erikson, Goldthorpe and Portocarero (EGP Class Scheme)*

| | |
|---|---|
| Variable name | egp |
| Variable label | Class scheme acc. to Erikson, Goldthorpe & Portocarero (EGP class scheme), current occupation, generated |
| Source variables | isco88, stib |
| Category/ dataset | Socio-economic position / person data |
| Prepared by | Bernhard Christoph |
| Description | The class scheme according to Erikson, Goldthorpe and Portocarero (Erikson et al. 1979, 1982; Erikson & Goldthorpe 1992) is one of the most common instruments for operationalising class position.<br>For this variable, data are coded exclusively on the basis of the ISCO-88 occupational classification and the detailed coding of occupational status. The coding procedure is essentially based on an earlier approach elaborated by Christoph et al. (2005), where detailed description is provided. In contrast to the procedure described by Christoph et al., unpaid family workers were not coded as self-employed persons but as persons in dependent employment in accordance with the coding applied in the European Socio-Economic Classification (ESeC), which is described in the next section.<br>Other than with the ESeC coding, cases were set to 'missing' (-7) in the EGP coding procedure, where the occupational activity seemed to be incompatible with the occupational status (e.g. 'directors and chief executives' [ISCO=1210], who indicated to be 'employees performing simple duties' [StiB=51]). For reasons of compatibility with the strongly standardised ESeC coding procedure we have adopted, we did not apply a comparable revision procedure when generating ESeC. |
| References | Christoph et al. (2005); Erikson & Goldthorpe (1992); Erikson et al. (1982); Erikson et al. (1979): |

*European Socio-economic Classification (ESeC)*

| | |
|---|---|
| Variable name | esec |
| Variable label | European Socio-economic Classification (ESeC), current occupation, gen. |
| Source variables | isco88, stib, PET2000, PET2700 |
| Category/ dataset | Socio-economic position / person data |
| Prepared by | Bernhard Christoph |
| Description | With regard to its theoretical conception, the European Socio-economic Classification is to a large extent 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).<br>The Stata do-file required to generate the ESeC was kindly provided by Heike Wirth from Gesis ZUMA (Fischer & Wirth 2007) and merely had to be adjusted to the requirements of this study. Originally, this do-file was written in SPSS by Harrison & Rose (2006) as a standard program for the generation of the EseC, which was converted into a Stata datafile. |
| References | Fischer & Wirth (2007); Harrison & Rose (2006); Müller et al. (2006, 2007); Rose & Harrison (2007) |

*Magnitude-Prestige-Scale (MPS)*

| | |
|---|---|
| Variable name | mps |
| Variable label | Magnitude-Prestige-Scale, current occupation, generated |
| Source variable | isco88 |
| Category/ dataset | Socio-economic position / person data |
| Prepared by | Bernhard Christoph |
| Description | The Magnitude-Prestige-Scale [MPS] (Wegener 1985, 1988) is the only instrument available so far to operationalise social prestige on the basis of detailed occupation codes, which was specifically developed for use on the national level. It was originally developed for the older 1968 version of the International Standard Classification of Occupations (ISCO-68). Since occupation coding in the PASS was subjected to the more recent ISCO-88 classification as well as to the Classification of Occupations (KldB) by the Federal Statistical Office, a variant of the scale transferred to ISCO-88 was used (Christoph 2005). The variable was generated by the Centre for Survey Research and Methodology (Gesis-ZUMA) as part of occupational coding procedures. |
| References | Christoph (2005); Wegener (1985, 1988) |

*Standard International Occupational Prestige Scale (SIOPS/Treiman-Skala)*

| | |
|---|---|
| Variable name | siops |
| Variable label | Standard International Occupational Prestige Scale, current occupation, generated |
| Source variable | isco88 |
| Category/ dataset | Socio-eonomic position / person data |
| Prepared by | Bernhard Christoph |
| Description | The scale originally constructed by Treiman (1977) for the ISCO-68 is the first and only prestige scale available so far which might be used in internationally comparative researchl. 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 variable was generated by the Centre for Survey Research and Methodology (Gesis-ZUMA) as part of occupational coding procedures. |
| References | Ganzeboom & Treiman (1996, 2003); Treiman (1977) |

*International Socio-Economic Index (ISEI)*

| | |
|---|---|
| Variable name | isei |
| Variable label | International Socio-Economic Index, current occupation, generated |
| Source variable | isco88 |
| Category/ dataset | Socio-eonomic position / person data |
| Prepared by | Bernhard Christoph |
| Description | Certainly, the International Socio-Economic Index is one of the most common indices for operationalising social position. This is not least due to the fact that, in contrast to most other SEIs, the ISEI is based on a well defined theoretical conception conceiving occupations and their social position as an 'intervening variable' between education and income.<br>Initially, the ISEI was developed for the ISCO-68 (Ganzeboom et al. 1992) and was adapted to the ISCO-88 later (Ganzeboom & Treiman 1996, 2003).<br>The variable was generated by the Centre for Survey Research and Methodology (Gesis-ZUMA) as part of occupational coding procedures. |
| References | Ganzeboom et al. (1992); Ganzeboom & Treiman (1996, 2003) |

## Variables at the Household Level and Household Communities in Joint Receipt of Benefit

### *Equivalized Household Income, Old OECD Scale*

| | |
|---|---|
| Variable name | oecdinca |
| Variable label | *Equivalized household income*, old OECD scale (rounded) |
| Source variables | HD0200a-HD0200o; HA0100; hhincome |
| | (hh27_01-hh27_15; hh1; hhincome) |
| Category/ dataset | Socio-eonomic position / household data |
| Prepared by | Bernhard Christoph |
| Description | With what is called 'e*quivalized household income*', statisticians try to reflect the savings achievable by common housekeeping in multi-person households compared to single households. To do so, the per-capita income in multi-person households is not calculated on the basis of the real number of persons living in the household, but by using a divisor which is usually below this figure and implies the assumed needs of the household (needs-adjusted household size). For more information on the old OECD scale, see OECD (1982); an overview on the topic is provided by Hauser (1996).<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 needs-adjusted household size. |
| References | Hauser (1996); OECD (1982) |

### *Equivalized Household Income, Modified OECD Scale*

| | |
|---|---|
| Variable name | oecdincn |
| Variable label | *Equivalized household income*,, modified OECD scale (rounded) |
| Source variables | HD0200a-HD0200o; HA0100; hhincome |
| | (hh27_01-hh27_15; hh1; hhincome) |
| Category/ dataset | Socio-eonomic position / household data |
| Prepared by | Bernhard Christoph |
| Description | General description: cf. *Equivalized household income*,, old OECD scale. For the Modified OECD scale, see Hagenaars et al. (1994).<br>The Modified OECD equivalence scale assumes a weighting factor of 1.0 only for the first household member (aged 15 or over). Any further household members aged 15 or over are assigned a weighting factor of 0.5; children up to the age of 14 are counted with a weighting factor of 0.3 to calculate the needs-adjusted household size. |
| Reference | Hagenaars et al. (1994) |

*Deprivation Index, Unweighted*

| | |
|---|---|
| Variable name | depindug |
| Variable label | Deprivation index, unweighted (items not missing for financial reasons; total of unweighted items: 26) |
| Source variable | HLS0100a-HLS2600a; HLS0100b-HLS2600b (HH7a_01-HH7b_13; HH8a_01-HH8b_13) |
| Category/ dataset | Material situation / household data |
| Prepared by | Bernhard Christoph |
| Description | Following a proposal by Ringen (1988), in poverty research there is usually a distinction been made between direct and indirect measurement of poverty. Indirect measurement is focussed on the resources available to attain a certain standard of living, especially the (equivalised household) income. Therefore, it is also referred to as the resource-based approach to measuring poverty.<br><br>In contrast, direct measurement attempts to record the households' actual ownership of goods and tries to find out to which extent a household cannot afford certain goods or activities considered essential for living for financial reasons. This is also referred to as the deprivation approach (see e.g. Halleröd 1995).<br><br>According to the general tenor of previous scientific research, the population classified as poor by the resource-based approach is not entirely identical with that referred to by the deprivation approach. In order to determine exactly who is to be considered poor in the true sense, recommendations exist to combine the measures of income-related poverty and deprivation and to take into account only those who are classified as poor in the true sense by both approaches (see Halleröd 1995; Nolan & Whelan 1996; Andreß and Lipsmeier 2001).<br><br>The index is based on a list of 26 goods or activities for which 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 they do not possess or do not participate in. However, only items which are missing for financial reasons are counted, in order to avoid misinterpretation of certain consumer preferences as a deterioration in the standard of living (e.g. if a household goes without a car or TV set deliberately).<br><br>Additionally, a missing for financial reasons was only assumed if answers to both questions explicitly confirmed the assumption, i.e. 'don't know' or 'no response' answers were valued either as availability of the respective commodity in the household or missing for another reason. This assumption is certainly not applicable to all cases. As an alternative, the calculation of index items for households that failed to answer (at least) one question for a certain commodity could have been dropped ('listwise deletion'). With respect to the total of 26 goods and activities surveyed, however, this method could have caused a high number of missing values in the index very quickly, for which the first method described had been preferred. Nevertheless, other than in the case of listwise deletion this method bears the risk of underestimating the number of goods missing. |
| References | Andreß & Lipsmeier (2001); Halleröd (1995); Nolan & Whelan (1996); Ringen (1988) |

*Deprivation Index, Weighted*

| | |
|---|---|
| Variable name | depindg |
| Variable label | Deprivation index, weighted (items not missing for financial reasons; total of weighted items: 12.8) |
| Source variables | HLS0100a-HLS2600a; HLS0100b-HLS2600b; PLS0100-PLS2600 (HH7a_01-HH7b_13; HH8a_01-HH8b_13; P141a-P141z) |
| Category/ dataset | Material situation / household data (weighting on the individual level) |
| Prepared by | Bernhard Christoph |
| Description | For a general description, see deprivation index, unweighted.<br>With respect to unweighted indices, there is often criticism that the items included are apportioned identical weightings. When comparing two items, for example, such as the existence of an indoor toilet or a VCR/ DVD player in a household, it becomes obvious that the dimensions to which a household's standard of living would be restrained by the lack of one of these items would be totally different. Therefore, weighting index items seems to be reasonable, even if empirical research has proven that weighted and unweighted index variants do not deliver significantly different results in most cases (see Lipsmeier 1999).<br>For the present survey, we decided to weight items according to the share of survey households who considered a particular item to be indispensable. We have chosen this procedure not only for its conceptual conclusiveness and commonness (applied by Halleröd 1995, for example), but also because it could be implemented without unreasonable effort. As the deprivation weightings to be determined for the individual questionnaire items are considered highly stable over time, they these items have to be administered either only once or at comparably long intervals. Moreover, thanks to the large population of the PASS sample, we were able to split the population up into several randomly selected subsamples, which were presented with selections of the items.<br>Alternative weighting methods such as a restriction of the index to items which are considered indispensable by a certain minimum share of 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 easily be generated on the basis of the data provided by the user. A discussion summarising the different methods of index weighting can be found at Andreß & Lipsmeier (2001, esp. pp. 28 ff.) |
| References | Andreß & Lipsmeier (1995, 2001); Andreß et al. (1996); Halleröd (1995); Lipsmeier (1999); Nolan & Whelan (1996) |

## *Household Typology*

| | |
|---|---|
| Variable name | hhtyp |
| Variable label | Household typology, generated |
| Source variables | HH27-HH37 |
| Category/ dataset | Household structure / household data |
| Prepared by | Daniel Gebhardt |
| Description | 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.y.) for the SOEP). The household typology used in PASS follows the SOEP version. Decisive criteria of differentiation are existing partnerships, number and age of children and existing family relationships. Whereas the SOEP typology is merely based on the relationship of the household members to the head of household, PASS uses interrelationships of all household members. In addition, the PASS typology includes the age of the household members as indicated in the interview and the household size.<br>**Definition of relationships for the household type generation:**<br>• Couples: married couples; couple living in a civil union; non-married couples and partnerships with a status not further specified (missing value for indication of the kind of partnership).<br>• Child of a person: natural child; stepchild, adopted or foster child; child with a status not further specified (missing value for parent-child relationship).<br>• Parent of a person: natural parent; step-parent, adopted or foster parent; parents with a status not further specified (missing value for kind of parentship).<br>**Definition of household types:**<br>• One-person household: Household consists of one person only.<br>• Couple without children: Household consists of two adults living together as a couple.<br>• One-parent household: Household consists of one parent and his/her children. No restrictions are made with respect to the childrens' age.<br>• Couple with children up to the age of 15: Household consists of two adults living together as a couple and their individual or common children. All children are up to the age of 15.<br>• Couple with children aged 16 or over: Household consists of two adults living together as a couple and their individual or common children. All children are aged 16 or over.<br>• Couple with children under 16 plus children aged 16 or over: Household consists of two adults living together as a couple and their individual or common children. Children up to 15 as well as chlidren aged 16 or over are living in the household.<br>• Multi-generation household: Household comprises members of at least three generations in linear succession. The central member(s) of the household has/have a multi-generation background, i.e. at least one person is both child and parent of another person in the household. Other persons living in the household are parents, children, siblings and partners of the central member(s) as well as partners' siblings.<br>• Other type: household which could not be assigned to one of the defined household types.<br>• Type generation not possible (missing values): Basically all households with at least one missing (-1,-2,-4) or implausible value (-8) in the main category of a relationship variable or in the age variable. (Exception: For households with three or less members in unambiguous relationship constellations, the household type was generated even if indications of age were missing.) |
| References | Beckmann & Trometer (1991); Frick et al. (o.J.); Lengerer et al. (2005); Porst (1984) |

*Benefit Community ID, Wave 1*

| | |
|---|---|
| Variable name | bgnr1 |
| Variable label | Benefit community ID in wave 1 |
| Source variables | HH27-HH37 |
| Category/ dataset | Household communities in joint receipt of benefits / individual register |
| Prepared by | Gerrit Müller |
| Description | The bgnr1 variable is created on the person level. It assigns an ident number to every household member indicating the person's affiliation to a particular household community in joint receipt of benefits. Consequently, household members with identical IDs form a common household community. The bgnr1 variable is established from the household ID and a two-digit indicator to identify the benefit community within the household.<br>The identification of a household member's affiliation to a community in joint receipt of benefits is solely based on the information on the relationships between household members from the household grid table as well as on the members' age according to the household interview. The identified benefit communities should, therefore, be considered to be 'synthetic' benefit communities. The identification process does not consider indications on actual benefit receipt, ability to work and qualification status. It rather intends to groups of persons in the same household who are or would be regarded as household communities in joint receipt of benefits according to the provisions of the Social Code Book II in the event that they are needy. This artificial allocation procedure became necessary, since information on the existence of a benefit community and the identification of persons related to this community cannot directly be collected from an interview.<br>With regard to implementation, the allocation of a person to a benefit community is based on the latest version of the German Social Code Book II, Section 7, sub-section 3 (last amended on 26/03/2007). Accordingly, a person aged between 25 and 64 forms an individual benefit community, unless this person lives in a partnership and/or has a child/ children aged under 25; in this case the benefit community comprises the person, his/her partner and the child(ren). For two persons who live in the same household with a common child, but indicated that they do not live in a partnership in the household grid table, a partnership is nevertheless implied to exist in terms of Section 7, sub-section (3a), for which the respective persons and their child/ren are assigned to the same benefit community. Persons between 15 and 25 are basically assigned to their parents unless they already live together with a partner (or a child of their own) in a common household. Persons between 15 and 25 who live without their parents (or partner/ children) form an independent benefit community.<br>Persons aged 65 and over are not covered by the Social Code Book II and are, therefore, not counted as members of a benefit community (code 0) unless they live together with a partner who is aged under 65 (or a child aged under 25) in the same household. Likewise, children under 15 who live 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 were not made for households with missing information on relationships and/or age; their members were assigned to code 99. By approximation, such households may be interpreted as households consisting of one benefit community only. |
| Reference | German Social Code Book II |

*Benefit Community Typology, Wave 1*

| | |
|---|---|
| Variable name | bgtyp1 |
| Variable label | Type of benefit community in wave 1 |
| Source variables | HH27-HH37 |
| Category/ dataset | Household communities in joint receipt of benefits / individual register |
| Prepared by | Gerrit Müller |
| Description | The concept of variable 'bgnr1' of creating synthetic household communities in receipt of benefits is also applicable to the typology of benefit communities. Up to the age of 25, children are counted as members of the benefit community of their parents unless they have a partner or children of their own. This is handled differently from the BA statistics, where typologies are often established on the basis of majority (completion of the 18th year). As an example: households with the youngest child aged between 18 and 24 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 panel data to data from the official statistics.<br>Code 0 (household community not in joint receipt of benefits) was assigned to households with members who are not covered by the Social Code Book II (see also code 0 for variable 'gbnr1'). Code -5 (generation impossible due to missing values) was allocated to households with missing information on relationships or age (see code 99 for bgnr1). |
| Reference | – |

*Household Community in Receipt of Unemployment Benefit II as of the Sampling Date, Wave 1*

| | |
|---|---|
| Variable name | bgbezs1 |
| Variable label | Household community in receipt of Unemployment Benefit II as of the sampling date, wave 1 (2006/2007) |
| Source variables | HH43, HH45, HH46, HH54, sample, bgnr1 |
| Category/ dataset | Household communities in joint receipt of benefits / individual register |
| Prepared by | Mark Trappmann |
| Description | This variable indicates whether a ‘synthetic’ household community in joint receipt of benefits which had been identified according to the procedure described for variable 'bgnr1' was in fact receiving Unemployment Benefit II as of the sampling date of wave 1. |
| Reference | – |

*Survey Household Community in Receipt of Unemployment Benefit II, Wave 1*

| | |
|---|---|
| Variable name | bgbezb1 |
| Variable label | Household community in receipt of Unemployment Benefit II at the time the survey was conducted, wave 1 (2006/2007) |
| Source variables | HH43-HH46, sample, bgnr1 |
| Category/ dataset | Household communities in joint receipt of benefits / individual register |
| Prepared by | Mark Trappmann |
| Description | This variable indicates whether a benefit community identified according to the procedure described for variable 'bgnr1' was in receipt of Unemployment Benefit II at the time of survey wave 1. |
| Reference | – |

*Number of Receiving Communities Within the Household*

| | |
|---|---|
| Variable name | anzbg |
| Variable label | Number of synthetic benefit communities in a household, generated |
| Source variables | bgnr1, hnr |
| Category/ dataset | Household communities in joint receipt of benefits / household data |
| Prepared by | Mark Trappmann |
| Description | This variable identifies the number of communities in joint receipt of benefits existing in a household according to the procedure described for the generation of variable 'bgnr1'. |
| Reference | – |

*Number of Household Communities Receiving Benefits as of the Sampling Date*

| | |
|---|---|
| Variable name | nbgbezug |
| Variable label | Number of household communities receiving benefits as of the sampling date. |
| Source variables | bgbezs1, bgnr1, hnr |
| Category/ dataset | Household communities in joint receipt of benefits / household data |
| Prepared by | Mark Trappmann |
| Description | The variable indicates the number of benefit communities within a household, which were in receipt of benefits according to the Social Code Book II as of the sampling date. The variable value was calculated by aggregating the benefit communities within each household that according to the variable 'bgbezs1' were actually receiving benefits when the sample for wave 1 was drawn. |
| Reference | – |

## 7. Weighting and the Use of Weights

a) Documentation of the Weighting Process

The weighting process is based on a three-stage weighting concept:

1. In the first stage, design weights are produced for the gross sample used.
2. Subsequently, non-response is modelled in the second stage.
3. The third stage is a calibration of the weights.

*Stage 1: design weighting*

The design weights are reciprocal selection probabilities for the gross sample used. The generation procedure is described in detail in Rudolph and Trappmann (2007). The design weights are contained in the dataset *hweights*. These are:

| | |
|---|---|
| dw_ba_1 | Design weight of a household in the BA sample (target population: households, in which at least one community was in joint receipt of benefits in accordance with SGB II as of July 2006) |
| dw_mi_1 | Design weight of a household in the Microm sample (target population: households in Germany) |
| dw_1 | Design weight of a household in the total sample (target population: households in Germany) |

*Stage 2: non-response modelling*

With the aid of two Logit models, the participation probability was estimated for all households in the gross sample. The first Logit model explains the probability of a contact. The second Logit model explains the participation (at least household interview and one complete personal interview) in the case of successful contact. These Logit models were calculated separately for the two samples, as the contact procedures were based on two different processes. Only the micro-geographical variables supplied by Microm were used for modelling the Microm sample. In the case of the models for the BA sample, additional characteristics from the sampling frame (A2LL or XSozial) could be used. The models applied contain variables with significant effects only (Likelihood-Ratio test, two-sided, 10% level). The table below shows the variables used and the coefficients of the models. A detailed description of the non-response modelling is

also contained in Chapter 5.2 of the TNS Infratest method and field report (Hartmann et al. 2008).

The dataset *hweights* contains the variables *pr_ba_1* and *pr_mi_1*. These are the products of the predicted probabilities of the two models, in separate form for the BA and Microm samples.

| | |
|---|---|
| pr_ba_1 | Estimated participation probability for cases from the BA sample |
| pr_mi_1 | Estimated participation probability for cases from the Microm sample |

Dividing the design weights by the estimated participation probabilities results in the modified design weights, which formed the starting point for the third stage – calibration.

*Stage 3: calibration*

The calibration process is described in detail in Kiesl (2008). We will therefore merely sketch out the basic procedure here.

HOUSEHOLD LEVEL

In an initial step, the two samples and the total sample were calibrated on the household level using official statistics.

The total and BA weights for benefit recipients in the two samples were adjusted to statistical values from the Federal Employment Agency (reporting month July 2006). The total and Microm weights were additionally adjusted to statistical values on private households in Germany for 2007 from the Federal Statistical Office. The figures used are detailed in Kiesl (2008).

All weights are household weights. The BA statistics, however, are based on the level of household communities in joint receipt of benefits. The link is created using synthetic benefit communities, generated as described in Chapter 6 (variable *bgnr1* in the dataset *p_register*). Households are initially broken down into synthetic benefit communities. The characteristics used for the calibration process are then generated on the benefit community level. This also includes the characteristic of whether the community received Unemployment Benefit II as of the sampling date. After calibration, the characteristics of all benefit communities in receipt of benefits as of the sampling

date are multiplied by the weighting factors to produce the household statistics. Communities in receipt of benefits within the same households therefore always receive the same weighting factors.

It is not always possible to determine the benefit receipt of a household or even less of a benefit community. Therefore as many information as possible is provided in order to enable users to make independent decisions. Thus, the variable *alg2samp* on the household level is supplied in the dataset *hh_register*. This contains the benefit receipt as of the sampling date for all households in the categories: *0 no receipt, 1 receipt, 2 no receipt according to survey (but included in BA sample and thus receipt according to process data), 3 receipt unclear from survey (but included in BA sample and thus receipt according to process data), 4 receipt unclear from survey (Microm sample)*. In addition, every user can generate this variable on his own, using unemployment benefit spell data (dataset *alg2_spells*). Other useful variables are *AL20600* and *AL20700a-o* (for whom does the household receive benefits?) and the variable *HA0400* from *HHENDDAT*, which for households founded after July 2006 records whether at least one household member received benefits in July 2006. The variable *sample* in *hweights* states which sample each household is taken from.

To generate the weight, however, a clear decision is necessary on which benefit communities should be regarded recipients of Unemployment Benefit II at the time of sampling. Below is an explanation of the decisions on which the weighting is based.

On the household level, it was decided that:

1. All households from the BA sample (*sample=1*) received benefits as of the sampling date, even if they denied doing so, provided at least one person aged between 15 and 64 lives in the household.
2. Households from the Microm sample for which benefit receipt cannot be clearly established based on the survey data are regarded as households receiving Unemployment Benefit II for the purpose of weighting if they state that they have ever received Unemployment Benefit II (*HA0300=1*) and if the start or end date of at least one observation lies in 2006 (with an undetermined end or start).

Transferring from the household to the benefit community level is wrought with even greater insecurity. This is so because it is not possible to obtain reliable retrospective information on which parts of the household received benefits in July 2006 in an em-

pirical survey. In most cases, the entire household consists of only one benefit community, making the question redundant, as the benefit community receives benefits precisely when the household does so. In cases where the household consists of several benefit communities, the following approach was selected.

The information on the persons for whom the household is currently receiving benefits (AL20600 and AL20700a-o) is used. A benefit community is regarded as receiving benefits if at least one of its members is reported as a benefit recipient. In a household with several benefit recipients and without information on the persons for whom the household receives benefits (e.g. as the questionnaire responses state that no benefits are being claimed), all synthetic benefit communities were regarded as being in receipt of benefits. The result of this generation is contained in the variable *bgbezs1* in the dataset *p_register*.

The weights following calibration on the household level are also contained in the dataset *hweights*.

| | |
|---|---|
| wqbahh1 | Calibrated household weight of the BA sample |
| wqmihh1 | Calibrated household weight of the Microm sample |
| wqhh1 | Calibrated household weight of the total sample |

### INDIVIDUAL LEVEL

Following the calibration on the household level, the persons who gave a person or pensioner's interview were adjusted to the benchmark statistics on the person-level. The starting point for this step was the calibrated household weights.

The total and BA weights for benefit recipients in both samples were adjusted to key statistics from the Federal Employment Agency (reporting month July 2006). The total and Microm weights were additionally adjusted to statistics from the Federal Statistical Office on private households in Germany for 2007. The statistics used are detailed in Kiesl (2008).

Pensioners' interviews were calibrated to population statistics in the same way as personal interviews. The BA statistics, however, do not contain figures on the number of pensioners in households receiving benefits. Nor do they identify persons living in households receiving benefits who are not part of a benefit community. It was therefore impossible to obtain the BA person weights for these persons by means of calibration.

The participation probability of these persons, given their household takes part in the survey, was estimated through a Logit model with the covariates: number of persons aged 15 and over in the household, interview mode, age and gender. The modified design weight was subsequently divided by this value.

The calibrated person weights are contained in the dataset *pweights*.

| | |
|---|---|
| wqbap1 | Calibrated person weight of the BA sample |
| wqmip1 | Calibrated person weight of the Microm sample |
| wqgesp1 | Calibrated person weight of the total sample |

b) Recommendations for the use of surveyset in Stata

All weights described here are *probability weights*: the weight of a household or person is equivalent to the reciprocal value of its/his/her inclusion probability (corrected by non-response modelling and calibration). In Stata, starting with version 9, probability weights have to be set in the *surveyset command*. However, *surveyset* not only has the purpose of defining the weights to be used, but also of defining the aspects of the survey design that influence the standard error.

There are two different possibilities of doing so in Stata: by specifying the design or by using replication weights. In the first option, the aspects of the survey design that influence the standard error have to be entered in the command line. These are – alongside the weights – clusters, stratification characteristics and finite population corrections in the selection without replacement. The effect of calibration on the standard error and other factors such as pps-sampling cannot be taken into account. The second option, on the other hand, makes use of a set of replication weights, which are calculated for all units of the study using processes such as jackknifing, BRR or bootstrapping. This procedure also potentially enables the calibration to be taken into account.

There are no replication weights available for PASS to date, so that researchers will have to use the first variation for their *surveyset*. However, the complex PASS survey design cannot be used for variance estimation with the *surveyset* command in all details. We recommend the following approach:

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

*wqX* stands for the adequate weight for the intended analyses. An indicator for the *primary sampling units* (which are the same for both subsamples) is the variable *psu* in the household dataset HHENDDAT.dta. The strata for the selection of the *primary sampling units* are represented by the variable *strpsu* in the same dataset. Strata with less than two units in the sample have been collapsed. In the case of drawing with replacement, strata and clusters do not play a role in variance estimation from the second level on (Särndal et al. 1992, 144ff.). If the sampling rate is very low, variance estimation for drawing without replacement can be approximated very well using the formulae for drawing with replacement. This is the case for PASS (only approximately 3.6% of the postal codes in Germany were selected for the survey). There is therefore no need to give finite population corrections or further clusters (here: households). However, the recommended *surveyset* then takes neither calibration nor pps-sampling into account, nor the low finite population correction for drawing without replacement. The resulting standard errors will be rather too large and thus should be considered conservative estimates.

c) Use of the Weights

All weights add up to the respective population size. Dividing these weights by their mean value results in weights that add up to the sample size. The design weights (*dw_mi_1, dw_ba_1, dw_1*) and the estimated participation probabilities (*pr_ba_1, pr_mi_1*) are provided with the panel study; however, we recommend using the calibrated weights. Researchers who wish to do without calibration should bear in mind that although division of the household weights by the adequate participation probabilities estimated for the respective subsamples does produce modified household design weights, these do not take into account the fact that there were also non-responses within participating households. Use on the individual level thus initially requires an estimation of the person's participation probability, given the household takes part.

The following sections provide examples showing how to use the weights for various different research questions.

*Analyses of benefit recipients in July 2006*

The best way to obtain findings on the population of the BA sample (households in which there was at least one benefit community receiving benefits in accordance with SGB II as of July 2006; referred to below as “households receiving benefits as of July 2006”) is to use only the BA sample and the relevant weights. In this case, proceeding in this way is more efficient than using the total sample, as the weights in the BA sample have less variance.

### Analyses on the household level

To make analyses of households receiving benefits as of July 2006, researchers should use wqbahh1. The example below demonstrates its use in Stata (from version 9). It is intended to calculate the number or percentage of households receiving benefits in possession of a car (variable HLS0800a). To start with, the household weights have to be merged with the household dataset, then the *surveyset* has to be carried out, and then the projected value can be calculated:

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

Approximately 37.9% of households with at least one benefit recipient have a car, 62.1% have no car, and the percentage with no relevant information is extremely low.

### Analyses on the benefit community level

Researchers working on receipt of Unemployment Benefit II are often not interested in households but in benefit communities. If the above question on the percentage of households receiving benefits as of July 2006 in possession of a car is to be transferred to benefit communities, the PASS data can be used to answer the question of how many benefit communities live in a household that has a car (as the benefit communities were identified retrospectively, there are no questions in the questionnaire relating directly to benefit communities – it is therefore not possible to identify which benefit community owns the car in a household consisting of several benefit communities). This question is relatively easy to answer, using the variable *nbgbezug*, which states how many benefit communities in joint receipt of Unemployment Benefit II a

household contains as of the sampling date[10]. The fastest way to do so is to multiply the household weights by this value.

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

The percentages deviate slightly from the analysis presented above (37.9% of households receiving benefits, but 38.2% of the benefit communities receiving benefits have a car in their household). Above all, however, the absolute numbers are different: the sum of all households receiving benefits was 3,882,013, whereas the sum of all benefit communities receiving benefits is 4,011,889, and matches the BA statistics due to the calibration.

[10] For this variable, the decisions required when the statements do not clearly identify how many benefit communities receive Unemployment Benefit II in the household were made in the same way as for the calibration process. Every user is of course free to make his or her own decisions on the basis of the Unemployment Benefit II observation.

### Analyses on the Individual Level

Analyses on the individual level are similarly simple. Researchers should use the weight wqbap1 in this case. An interim step is necessary, as the variables psu and strpsu are only contained in the household dataset. The following example calculates the number of persons aged 15 years and above[11] in households receiving benefits with a background of migration (variable *migration*).

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

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

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

Fewer than 61.3% do not have a background of migration, 24.4% migrated to Germany themselves, at least one parent migrated to Germany for a further 7.6%, and at least one grandparent for another 1.9%. The code “Item not administered in questionnaire”[12] applies to 3.6%. This is due to the fact that the data from the short questionnaire for persons aged 65 and above is stored in the same dataset as data from the personal questionnaire. Persons aged 65 and above receive this code for questions not asked in the pensioners’ questionnaire. In order to run analyses excluding these persons, researchers can limit the frequencies to data from the person questionnaires (fb_vers=1).

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

The person weights of the BA sample project onto all persons in households receiving benefits. Some households, however, consist of several synthetic benefit communities,

[11] As younger persons cannot be interviewed in person, the PASS data can only be used to establish characteristics surveyed in the household questionnaires (e.g. age, gender). The household weights should be used in this case.

[12] For a further 1.2%, the characteristic cannot be formed due to missing information.

not all of which receive benefits. Researchers wanting to project only onto persons who are members of SGB II benefit communities have to exclude persons with *bgbezs1=0* (by merging the variable *bgbezs1* from the person register *p_register* with the individual database):

```
drop _m
sort pnr
merge pnr using p_register.dta
tab _m pnetto1
drop if _merge==2
svy, subpop(if bgbezs1==1): tab migration if fb_vers==1, count ///
cell format(%9.0g)
```

The percentage of persons who migrated to Germany themselves is therefore slightly higher among the persons who are members of a benefit community, at 25.5%, than among persons living in a household receiving benefits (25.3%).

*Analyses on the residential population of Germany*

Analyses on the residential population of Germany can be carried out both with the total weights and with the Microm weights, with results differing only minorly in most cases. The percentage of households with a car in the total population is calculated either using the following commands using the total weights:

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

Or alternatively with the Microm weights:

```
svyset psu [pw=wqmihh1], strata(strpsu)
svy: tab HLS0800a, cell ci format(%10.0g)
```

In the first case, the percentage of households with a car is 76.0% (95% confidence interval from 74.1% to 77.9%), and in the second case 75.8% (95% confidence interval from 73.7% to 77.8%). The confidence interval using the total weights is slightly smaller, as in this case the part of the population receiving benefits under SGB II is rep-

resented much more precisely, which is why we prefer its use. The same applies to the person weights.

*Analyses on benefit recipients using the latest available data*

Working with the BA sample and the appropriate weights, results refer to recipients in July 2006. For analyses of this population, this approach achieves the greatest statistical power, as the BA weights have a relatively low variance. However, researchers will prefer to carry out many analyses – especially on fast-changing characteristics – using the latest available data, to which many characteristics refer, such as employment status, income or employment volume. The survey date is between 6 and 13 months after the sampling date. Working on the latest available data exclusively with the BA sample, researchers can only make analyses of *'stayers'*, those who continued to receive benefits from the sampling date to the survey date. In view of a rather high turnover (37% of persons receiving benefits under SGB II in January 2005 were no longer doing so by December 2006 (Graf 2007)), this group may be significantly distinct from the current benefit recipients in its makeup.

ANALYSES ON BENEFIT RECIPIENTS USING THE LATEST AVAILABLE DATA ON THE HOUSEHOLD LEVEL

Representative results for current benefit recipients can therefore only be achieved for current benefit recipients using the total weights. The characteristic of whether the household is currently receiving benefits (*alg2abez*) is contained in the household dataset (HHENDDAT.dta). Calculations are therefore relatively simple on the household level. The example below shows this, again using the question of car ownership.

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

Out of households currently receiving benefits, 36.2% have a car. The value is slightly lower than for the households receiving benefits in July 2006.

If we were to estimate this using the BA weights and the BA sample,

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

the value would be calculated at 33.7%. However, as these are only *'stayers'* or households that received benefits both in July 2006 and as of the survey date, it is plausible that fewer of these households have cars than those that stopped or started receiving benefits during the intervening period.

One consequence of using the total weights rather than the BA weights is the significant increase in the confidence intervals. The variance of the total weights is significantly larger in both subsamples due to the very different sampling rates. The analyses on car ownership in households receiving Unemployment Benefit II as of the sampling date (see above) resulted in a 95% confidence interval from 36.0% to 39.7%. For the survey date, we receive an almost doubled 95% confidence interval from 33.5% to 39.1%.

### Analyses on benefit recipients using the latest available data on the benefit community level

In comparison to the analyses referring to the sampling date (in 5.3.1.2), an extra step has to be taken as there is no variable equivalent to *nbgbezug* for receipt of benefits as of the survey date. Therefore, in a first step, this variable has to be generated using the variable *bgbezb1* in *p_register*, which indicates for each benefit community whether this particular community was receiving Unemployment Benefit II as of the survey date[13].

[13] In the sample code, 'recode bgbezb1 (-5=0)' is used to treat all benefit communities for which current receipt of benefits is unclear on the basis of the survey data as non-recipients.

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

The estimated value of 36.2% does not differ from that obtained in the analysis on the household level. However, the value no longer relates to a sub-population of just under 3,489,000 households as in the section above, but to just below 3,519,000 benefit communities receiving benefits as of the survey date. During the survey period, the number of benefit communities varied between 3,724,000 (July 07) and 3,818,000 (March 07), according to the BA statistics. This reference value is thus not quite achieved. The under-reporting arises from the fact that, unlike in the figures referring to the sampling date, information on benefit receipt at the time of the survey is not available from process data for all respondents. Thus the under-reporting on benefit receipt[14] is not corrected for when using the latest available data.

[14] As receipt of Unemployment Benefit II is likely to be a socially undesirable characteristic, a certain amount of under-reporting is not surprising.

### Analyses on benefit recipients using the latest available data on the individual level

Analyses can be transferred to the individual level in much the same way as it was done when using data referring to the sampling date. To start with, researchers should merge the individual dataset with the person weights and the information for the *surveyset* again. For analyses on persons from households currently receiving benefits, the frequencies should be limited to persons with *alg2abez=1*. This characteristic has to be merged from the household dataset.

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

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

Of the persons in households currently receiving Unemployment Benefit II, 60.6% have no background of migration, 26.1% migrated to Germany themselves, 8.1% have at least one migrated parent and 2.0% one migrated grandparent.

In most cases, however, analyses will not be limited to persons in households receiving benefits, but to persons in benefit communities receiving benefits. This characteristic is contained in the person register. The following series of commands produces the percentage of migrants among persons in benefit communities aged between 15 and 64.

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

*Analyses on benefit recipients at later dates*

The biographical data on Unemployment Benefit II receipt on the household level will in principle allow to perform analyses referring to other points in time which are between sampling and the date the first wave of the survey was administered. However, variables such as *bgbezs1*, *bgbezb1* or *nbgbezug* are only provided for the two dates described above. Users who would like to run projections referring to other points in time therefore will have to generate analogous variables. Doing so, imprecision as well as the problem of under-reporting of benefit receipt will have to be taken into account.

*Comparison of benefit recipients with the general population*

The large variety of options for studying benefit recipients and their households results in an equally large variety of options for comparing benefit recipients with the general population. Table 9 provides an overview. The total weights should be used in all cases.

**Table 9: Variables and their uses for comparison of SGB II benefit recipients with the general population**

| Variable | Dataset | Values | Suitable for comparison of … |
|---|---|---|---|
| sample | PENDDAT<br>HHENDDAT | 1 BA sample<br>2 Microm sample | a) Households receiving Unemployment Benefit II as of July 2006 (sample=1) with households in the residential population (sample=2).<br>b) Persons in households receiving Unemployment Benefit II as of July 2006 (sample=1) with persons in households in the residential population (sample=2).<br>Households receiving Unemployment Benefit II are in this case defined via the subsample. |
| alg2samp | hh_register | 0 No benefit receipt<br>1 Benefit receipt<br>2 No benefit receipt according to survey (BA-SP)<br>3 Benefit receipt unclear according to survey (BA-SP)<br>4 Benefit receipt unclear according to survey (Microm-SP) | a) Households receiving Unemployment Benefit II as of July 2006 (alg2samp=1) with households not receiving Unemployment Benefit II as of July 2006 (alg2samp=0).<br>b) Persons in households receiving Unemployment Benefit II as of July 2006 (alg2samp=1) with persons in households not receiving Unemployment Benefit II as of July 2006 (alg2samp=0).<br>Users may choose how to treat cases that received Unemployment Benefit II according to the sample but not the survey. |
| bgbezs1 | p_register | 1 Unemployment Benefit II receipt on sampling date<br>0 No Unemployment Benefit II receipt on sampling date | Persons in benefit communities receiving Unemployment Benefit II as of July 2006 (bgbezs1=1) with persons in benefit communities not receiving Unemployment Benefit II as of July 2006 (bgbezs1=0)<br>As this variable was used for the weighting, a decision was taken for every ambiguous case. |
| alg2abez | HHENDDAT | 1 HH currently receiving Unemployment Benefit II<br>2 HH not currently receiving Unemployment Benefit II<br>-5 No generation possible (missing values) | a) Households receiving Unemployment Benefit II as of survey date (alg2abez=1) with households not receiving Unemployment Benefit II as of survey date (alg2abez=2).<br>b) Persons in households receiving Unemployment Benefit II as of survey date (alg2abez=1) with persons in households not receiving Unemployment Benefit II as of survey date (alg2abez=2). |
| bgbezb1 | p_register | 1 Unemployment Benefit II receipt in wave 1<br>0 No Unemployment Benefit II receipt in wave 1<br>-5 No generation possible (missing values) | Persons in benefit communities receiving Unemployment Benefit II as of survey date (bgbezb1=1) with persons in benefit communities not receiving Unemployment Benefit II as of survey date (bgbezb1=0) |

## 8. Anonymisation

All data gathered by the IAB as a department of the Federal Employment Agency (BA) are social data, which places high demands on data protection. It was therefore necessary to include several variables only in more or less strongly aggregated form in the dataset. These variables are generally identified as *anonymisiert* 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 these. Table 10 provides an overview of the variables in question and the steps taken to render them anonymous.[15]

**Table 10: Overview of the variables rendered anonymous**

| Question | Var name | Var label | Procedure |
|---|---|---|---|
| p1 | PD0100 | Year of birth | The precise date of birth was shortened to year of birth and an indicator for 1st/2nd half-year. |
| | gebhalbj | 1st/2nd half-year of birth | |
| p38 | PET1500 | Occupational status | For technical reasons, professional soldiers and judges were recorded separately in the survey. Due to the low case numbers and as these are not usually recorded as separate categories of occupational status, these two professions have been grouped with the civil servants. |
| p41 | PET1800 | Current occupational status: professional soldier | This variable is not supplied. The cases have been added to the relevant categories in PET1900. |
| p42 | PET1900 | Current occupational status: civil servant | This variable contains additional cases. The judges from P38 have been grouped with the higher-ranking civil servants, the professional soldiers from P41 with the relevant civil servant category. |
| p192 | PMI0200 | No native Germans: native country | Countries with low case numbers have been grouped into larger categories. |
| p196 | PMI0500 | Nationality not German: which nationality? | Nationalities with low case numbers have been grouped into larger categories. |
| p201a – p201f | PMI1000a -PMI1000f | Country parents and grandparents migrated from | Countries with low case numbers have been grouped into larger categories. |
| Generated variable | ogebland | Native country, incl. information from open ended survey questions | Countries with low case numbers have been grouped into larger categories. |
| Generated variable | ostaatan | Nationality, incl. information from open ended survey questions | Nationalities with low case numbers have been grouped into larger categories. |
| Generated variable | ozulanda - ozulandf | Country parents and grandparents migrated from, incl. information from open ended survey questions | Countries with low case numbers have been grouped into larger categories. |

[15] Should the use of the non-anonymous version of one or more of these variables be essential for your research project, please contact the Research Data Centre to find a suitable data access possibility. This will depend on the research project and the variables required.

## 9. References

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

Achatz J., Hirseland A. and Promberger M. (2007), “IAB-Panelbefragung von Haushalten im Niedrigeinkommensbereich. Entwurf für ein Rahmenkonzept“, pp. 11-32 in Promberger, Markus (ed.), Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung ‘Arbeitsmarkt und Soziale Sicherung’, IAB Forschungsbericht No.12/2007, Nuremberg.

Andreß H.-J., Burkatzki E., Lipsmeier G., Salentin K., Schulte K. and 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. and Lipsmeier G. (1995): "Was gehört zum notwendigen Lebensstandard und wer kann ihn sich leisten? Ein neues Konzept zur Armutsmessung." Aus Politik und Zeitgeschichte, issue B 31-32/1995: pp. 35-49.

Andreß H.-J. and Lipsmeier G. (2001): "Armut und Lebensstandard. Gutachten im Rahmen des Armuts- und Reichtumsberichts der Bundesregierung." Federal Ministry of Labour and Social Affairs (BMAS). Bonn.

Beckmann P. and Trometer R. (1991): "Neue Dienstleistungen des ALLBUS: Haushalts- und Familientypologien, Klassenschema nach Goldthorpe." ZUMA-Nachrichten No. 28: pp. 7-17.

Brauns H. and Steinmann S. (1999): "Educational Reform in France, West-Germany and the United Kingdom: Updating the Casmin Classification." ZUMA-Nachrichten No. 44: pp. 7-45.

BMBF [Federal Ministry of Education and Research] (2003):"Berufsausbildung sichtbar gemacht. Schaubildsammlung." 4th edition. BMBF, Bonn.

Christoph B. (2005): "Zur Messung des Berufsprestiges: Aktualisierung der Magnitude-Prestigeskala auf die Berufsklassifikation ISCO88." ZUMA-Nachrichten No. 57: pp. 79-127.

Christoph B., Sachweh P. and Müller T. (2005): "Umfrage zum Forschungsprojekt 'Die Akzeptanz des Wohlfahrtsstaats'. Codebuch und Dokumentation." Mannheim Centre for European Social Research (MZES), Mannheim.

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

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

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

Frick J. R., Göbel J. and Krause P. (n.d.): "$HGEN: Generated Household-Level Variables." Download: [http://www.diw.de/documents/dokumentenarchiv/17/60053/hgen.pdf (08/11/2007)].

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

Ganzeboom H.B.G., De Graaf P.M. and Treiman D.J. (1992). "A Standard International Socio-Economic Index of Occupational Status." Social Science Research, 21: pp. 1-56.

Graf T. (2007). "Bedarfsgemeinschaften 2005 und 2006: Die Hälfte war zwei Jahre lang durchgehend bedürftig." IAB-Kurzbericht No. 17/2007.

Granato N. (2000): "CASMIN-Bildungsklassifikation. Eine Umsetzung mit dem Mikrozensus 1996." ZUMA-Technischer Bericht No. 2000/12.

Hagenaars A., J. M. de Vos K. and Zaidi M. A. (1994): "Poverty Statistics in the Late 1980s: Research Based on Microdata." Office for Official Publications of the European Communities, Luxembourg.

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

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

Hartmann J., Brink K., Jäckle R. and Tschersich N. (2008). "IAB-Haushaltspanel im Niedrigeinkommensbereich: Methoden- und Feldbericht." Munich.

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

Helberger C. (1988):"Eine Überprüfung der Linearitätsannahme der Humankapitaltheorie." pp. 151-170 in: H.-J. Bodenhöfer (ed.): Bildung, Beruf, Arbeitsmarkt. Duncker & Humblot, Berlin.

Infas (2006), IAB project: "Zusammenlegung von Arbeitslosen- und Sozialhilfe mit Einführung des SGB II: Übergangsmuster, Strukturen und Folgen für die Betroffenen". Representative cross-sectional survey. Final report on the methodological implementation. Infas, Bonn.

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

Kiesl H. (2008): "Kalibrierte Hochrechnung für das Panel 'Arbeitsmarkt und soziale Sicherung' (Welle 1)". FDZ Methodenreport No. x/2008. Research Data Centre of the BA at the IAB, Nuremberg [forthcoming].

König W., Lüttinger P. and Müller W. (1987): "Eine vergleichende Analyse der Entwicklung und Struktur von Bildungssystemen. Methodologische Grundlagen und Konstruktion einer vergleichbaren Bildungsskala." CASMIN project. Working Paper No. 12.

Lechert Y., Schroedter J. and Lüttinger, P. (2006): "Die Umsetzung der Bildungsklassifikation CASMIN für die Volkszählung 1970, die Mikrozensus- Zusatzerhebung 1971 und die Mikrozensen 1976-2004." ZUMA Methodenbericht No. 2006/12.

Lengerer A., Bohr J. and Janßen A. (2005): "Haushalte, Familien und Lebensformen im Mikrozensus – Konzepte und Typisierungen." ZUMA Arbeitsbericht No. 2005/05. ZUMA. Mannheim.

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

Meßmann S.. Bender S., Rudolph H., Hirseland A., Bruckmeier K., Wübbeke C., Dundler A., Städele D. and Schels B. (2008): "Lebenssituation und Soziale Sicherung 2005 (LSS 2005) * IAB-Querschnittsbefragung SGB II. Handbuch-Version 1.0.0." (FDZ Datenreport No. 04/2008 (de)), Nuremberg.

Müller W., Wirth H., Bauer G., Pollak R. and Weiss F. (2007): "Entwicklung einer europäischen sozioökonomischen Klassifikation." Wirtschaft und Statistik No. 5/2007: pp. 527-530.

Müller W., Wirth H., Bauer G., Pollak R. and Weiss F. (2006): "ESeC – Kurzbericht zur Validierung und Operationalisierung einer europäischen sozioökonomischen Klassifikation." ZUMA-Nachrichten No. 59: pp. 111–119.

Nolan B. and Whelan C.T. (1996). "Measuring Poverty Using Income and Deprivation Indicators: Alternative Approaches." Journal of European Social Policy, Vol. 6, No. 3: pp. 225-240

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

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

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

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

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

Rudolph H. and Trappmann M. (2007): "Design und Stichprobe des Panels 'Arbeitsmarkt und Soziale Sicherung' (PASS)", pp. 60-101 in Promberger, Markus. (ed.), "Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung 'Arbeitsmarkt und Soziale Sicherung' ", IAB Forschungsbericht No. 12/2007, Nuremberg.

Särndal C.-E., Swensson B. and Wretman J. (1992): "Model Assisted Survey Sampling." Springer, New York.

Schnell R. (1997): "Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen", Leske+Budrich, Opladen.

Schnell R. (2007): "Alternative Verfahren zur Stichprobengewinnung für ein Haushaltspanelsurvey mit Schwerpunkt im Niedrigeinkommens- und Transferleistungsbezug", pp. 33-59, in Promberger, Markus. (ed.), "Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung 'Arbeitsmarkt und Soziale Sicherung' ", IAB Forschungsbericht No. 12/2007, Nuremberg.

Sozialgesetzbuch Zweites Buch [German Social Code Book II - SGB II]: "Grundsicherung für Arbeitssuchende." (Basic security for job seekers).

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

Trappmann M., Christoph B., Achatz J. and Wenzig C. (2007). " 'Labour Market and Social Security': A New Panel Study for Research on Long Term Unemployment. Paper Presented at the International Conference of the German Association of Po-

litical Economy", Nuremberg, October 11th, 2007. http://doku.iab.de/veranstaltungen/2007/cape_2007_trappmann.pdf

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

Wagner G.G., Frick J.R. and Schupp J. (2007): "The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements". Schmollers Jahrbuch 127 (1), pp.139-169.

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

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

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

*Imprint*

FDZ Datenreport
No. 05/2008 (EN)

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

**Editorial staff**
Stefan Bender, Dagmar Herrlinger

**Technical production**
Dagmar Herrlinger



**Download**
http://doku.iab.de/fdz/reporte/2008/DR_05-08_I_EN.pdf

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

**Corresponding authors**
Bernhard Christoph, Institute for Employment Research, Regensburger Str. 104, D-90478 Nuremberg;
Phone: +49-(0)911/179-3507;
e-mail: bernhard.christoph@iab.de

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