Running since 1984, the German Socio-Economic Panel study (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing. The SOEP Survey Papers is comprised of the following series:

**Series A** – Survey Instruments (Erhebungsinstrumente)
**Series B** – Survey Reports (Methodenberichte)
**Series C** – Data Documentation (Datendokumentationen)
**Series D** – Variable Descriptions and Coding
**Series E** – SOEPmonitors
**Series F** – SOEP Newsletters
**Series G** – General Issues and Teaching Materials

The SOEP Survey Papers are available at [http://www.diw.de/soepsurveypapers](http://www.diw.de/soepsurveypapers)

**Editors:**
Dr. Jan Goebel, DIW Berlin
Prof. Dr. Martin Kroh, DIW Berlin and Humboldt Universität Berlin
Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin
Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:
The dataset FiDv4.0 is released with doi: 10.5684/soep.fid.v4.0. The data of the FiD-Study is now integrated completely in SOEP-Core, Version 31ff. (10.5684/soep.v31). The households will continue to be monitored as samples L1-L3 and interviewed with the SOEP-Core questionnaire.
Preface

This documentation consists of various parts and is meant to guide the user through the latest release of the “Familien in Deutschland” FiD data collection. This short introduction provides an overview of files and concepts, while the specific documentation files give some insights on the data generation process.

In case of questions or comments, please feel free to contact any one from the FiD-Team at the SOEP division of the DIW Berlin. We thank all SOEP researchers for their support and guidance throughout FiD’s first three waves of data collection, especially Elisabeth Liebau for help during the questionnaire construction phase, Martin Kroh for help on the weighting procedures, and Juliana Werneburg for her help on the marital datasets. Special thanks are in order for our current and past student assistants Stefan Damerow, Moritz Mannschreck, Guido Putzke, Alexander Raith, Nina Scherner, Linda Wittbrodt and Malisa Zobel for their excellent assistance throughout the data generation and documentation. We are also grateful to the SOEP-FiD group at TNS-Infratest, who provided valuable input to the documents. This work would not have been possible without the financial support of the Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (BMFSFJ) and the Federal Ministry of Finance (BMF). Last, but by no means least, we especially thank the FiD respondents for entrusting their information to us.

Berlin, February 2014

Mathis Fräßdorf
Jürgen Schupp jschupp@diw.de 030/89789-238
Rainer Siegers rsiegers@diw.de 030/89789-239
C. Katharina Spieß kspiess@diw.de 030/89789-254
Contents

General Documentation ..................................................................................................... 1
Documentation ppfad ...................................................................................................... 21
Documentation hpfad .................................................................................................. 32
Documentation $pgen ................................................................................................... 39
Documentation $hgen .................................................................................................. 82
Documentation bioage01-10 Files ............................................................................ 108
Documentation biobirth ............................................................................................... 147
Documentation biomarsy and biocouply ..................................................................... 156
Documentation $kind .................................................................................................. 175
Documentation biojob .................................................................................................. 185
Documentation hhrf and phrf ..................................................................................... 197
Documentation mipinc and mihinc .............................................................................. 220
Documentation pbiospe ............................................................................................... 254
Documentation pbrutto ............................................................................................... 260
Documentation hbrutto ............................................................................................... 275
Documentation paradatal ............................................................................................. 288
Documentation bioage17 ............................................................................................ 297
Documentation $bioparen ............................................................................................ 308
Änderungen in den Versionen 1.1-4.0 der FiD Datenweitergaben ............................... 324
General Documentation

Basics, Structure, and Overview of Datasets

Mathis Fräsdorf (geb. Schröder)
Contents

General information ....................................................................................................................... 3
The very basics of “Familien in Deutschland” .................................................................................. 4
Data structure: types of data files, names and data organization..................................................... 5
Combining data files ..................................................................................................................... 7
FiD data files .................................................................................................................................. 9
  Basic Data Files .......................................................................................................................... 9
    ppfad .......................................................................................................................................... 9
    hpfad ........................................................................................................................................ 9
    Shbrutto ................................................................................................................................. 9
    Spbrutto .................................................................................................................................. 9
    hbrutt10_fid ........................................................................................................................... 10
    hbrutt11_fid ........................................................................................................................... 10
  Original Data Files ..................................................................................................................... 11
    Sh ............................................................................................................................................. 11
    Sp ............................................................................................................................................. 11
    Sjela ......................................................................................................................................... 11
    Spkal ....................................................................................................................................... 12
    Sjugend ................................................................................................................................... 12
    Seltern1 .................................................................................................................................... 12
    Seltern2 .................................................................................................................................... 12
    Seltern3 .................................................................................................................................... 12
    Seltern4 .................................................................................................................................... 13
    Seltern5 .................................................................................................................................... 13
    Seltern6 .................................................................................................................................... 13
    Skind ......................................................................................................................................... 13
    Chuecke ..................................................................................................................................... 13
  Generated Data Files .................................................................................................................. 15
    Spgen ........................................................................................................................................ 15
    Shgen ....................................................................................................................................... 15
    artkalen .................................................................................................................................... 15
    pbiospe .................................................................................................................................... 15
    phrf and phrf_fidsoep.................................................................................................................. 16
    hhrf and hhrf_fidsoep.................................................................................................................. 16
    Smipinc ..................................................................................................................................... 16
    Smihinc ...................................................................................................................................... 16
    bioage01 .................................................................................................................................... 17
    bioage02 .................................................................................................................................... 17
    bioage03 .................................................................................................................................... 17
    bioage06 .................................................................................................................................... 17
    bioage08p1 and bioage08p2 ......................................................................................................... 17
    bioage10p1 and bioage10p2 ......................................................................................................... 18
    bioagel ....................................................................................................................................... 18
    biobirth ..................................................................................................................................... 18
    biomarsy ................................................................................................................................... 18
    biocouply .................................................................................................................................... 19
    biorresid .................................................................................................................................... 19
    bioparen .................................................................................................................................... 19
    bioage17 .................................................................................................................................... 19
    paradata ..................................................................................................................................... 20
Datasets not in FiD.......................................................................................................................... 20
General information

The project “Familien in Deutschland” – “Families in Germany” – is a longitudinal panel study financed by the German Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (BMFSFJ) and the German Federal Ministry of Finance (BMF). Its main purpose is to provide researchers with new and better data on specific groups in the German population: low-income families, families with more than two children, single parent families as well as families with young children. The data are the backbone of the first large scale evaluation of family policy measures in Germany on behalf of the two involved ministries.

The purpose of this documentation is to provide the user with the essentials of the data from “Familien in Deutschland” (FiD). We briefly describe the samples, the data structure, the naming logic, and give some basics on combining different data sets. We then turn to the different files included in the current data distribution (version 4.0), and for each provide a quick overview of the most important facts. More information is available from an Article to be published in “Schmollers Jahrbuch” (2013, 133, 4) by Schröder, Siegers & Spieß or in the SOEP-Survey paper 556¹. Either publication should be cited when using the FiD data.

This documentation is the start to the analysis with the FiD data – it somewhat resembles the Desk Top Companion known from the SOEP data collection, although it has been specifically constructed to suit the FiD data. There are several file-specific documentation files which are included as separate files in the data distribution. Users should refer to them, as they dive into the details of each data set. The questionnaire files, which are available as separate pdf-files for each questionnaire in the field, complement the different documentations. As FiD has a large overlap with the SOEP in terms of questions and variables, a further very useful documentation tool is provided in the correspondence file (FiD-SOEP_Correspondence_v4.0.xlsx), where variable correspondences are given within FiD (i.e. across waves) and across FiD and SOEP (i.e. variable correspondence from 2010 – “ba”, 2011 – “bb”, and 2012 – “bc”). This allows to ‘translate’ from one wave or dataset to the other. In addition, a future online documentation for the SOEP (formerly known as SOEPinfo) will include FiD and other related studies as well.

¹ see Schröder, Siegers & Spieß, 2013, included in the data distribution and also available online at: http://www.diw.de/documents/publikationen/73/diw_01.c.421623.de/diw_sp0556.pdf
The very basics of “Familien in Deutschland”

The role model for this project is the well-known “Socio-economic Panel” (SOEP), a representative study of the German population which runs annually since 1984. FiD is shaped after the SOEP in terms of structure (i.e. a household sample with individual interviews) and questionnaire content. However, the sample was drawn and weighted to represent the following four groups of the German population:

1. Families in "critical income brackets"
2. Single parents
3. Families with more than two children

The first three samples were selected through a screening process, which defined eligibility for households through a quick telephone interview before the actual interview. These three samples are thus often referred to as the “Screening Sample 2010”. The fourth group was selected through random sampling from German registries, and is referred to as the “Cohort Sample”.

All interviews were conducted in face-to-face mode, i.e. an interviewer was present during the interview. The general mode of the interviews was CAPI (Computer Assisted Personal Interview), however, some questionnaires allowed for a PAPI (Pen and Paper Interview) mode as well.

In 2010, wave 1 started containing overall 4.574 interviewed households with a total of 17.002 individuals. There are a total of 17.352 different interviews available, which include household interviews, person interviews and proxy interviews for children living in the household. Especially the latter provide valuable information for the evaluation purposes. The longitudinal dimension started with wave 2 in 2011. Here an additional subsample 924 households of both single parents and families with more than two children was added to the study, called the “Screening Sample 2011”.

2 For more details on the sampling procedures and definitions, on response rates, and on the number of observations, please see the methodological reports by TNS Infratest at: http://www.diw.de/de/diw_01.c.405871.de/fid_dokumentation.html.
**Data structure: types of data files, names and data organization**

“Familien in Deutschland” (FiD) is an annually conducted panel study, providing the collected information in various different data files. Overall, there are 100 files included in the data distribution of February 2014 (version 4.0), which can come on one of three different levels: household, adults, and children. It is always possible to match data from one level with the other levels via identifiers common to all data files (see the next section for details). The different levels are one reason for storing information in separate files rather than in one large data file. However, we do not combine data within one information level, either, because such a combination would make it very cumbersome to deal with the data – not only from a computational point of view, but keeping an overview of content and datasets with several hundreds of variables might not be easy.

In general, FiD publishes two different types of datasets: original files (i.e. datasets following the structure of questions and not including any generated variables) and generated files (datasets produced by using information from one or more of the original files).\(^3\) Original files are cross-sectional, while generated files may be cross-sectional or longitudinal. It is easy to identify cross-sectional datasets in FiD: they all have a wave specific label, which in documentation files, we generally denote by the letter “$”. The “$” is a substitute for a prefix, e.g. “f10” corresponds to the data collection in 2010. Cross-sectional datasets usually have only one observation for each key unit (person or household), while longitudinal datasets may have multiple observations at different time points for each key unit. Longitudinal generated data files can be either annually based (i.e. each year is an observation) or spell based. We provide more detail on the structure of each dataset below (see “FiD Data Files”).

Wave identifiers (“$”) are generated in a simple way: Each wave is abbreviated by using the lower-case letter “f” and the third and fourth digit of the respective year in which the survey took place. For example, the first wave is called “f10”, the second wave is called “f11”, and so on. Wave identifiers are always used as a prefix to a dataset, and should not be confused with the “$$” used as suffix for variables in some datasets (mainly $pgen$ and $hgen$). In these files, the double dollar sign “$$” indicates a wave specific variable – it is a substitute for the last two digits of the respective year (e.g. SIZE$$ would be found as SIZE10 in $f10hgen$).\(^4\)

---

\(^3\) There is one exception to this rule with the dataset $kind$, see below.

\(^4\) As a matter of convenience to increase readability, throughout the documentation we refer to variable names in CAPITAL letters, even though in the datasets, data are usually in lower case letters. Datasets are in general referred to in **bold italic** letters.
For each generated data file, FiD provides an additional detailed documentation on the specifics of the generating process. With regard to the original data files, no additional documentation exists, as variables in the original files are named according to the sequence of the questionnaire. Principally, the following rule has been used for naming variables in original data files: the wave specific label is followed by the first letter of the name of the data file, followed by three digits indicating the number of a specific question in the questionnaire. Imagine, for example, that you are interested in question 10 of the “Household questionnaire” (i.e. “equipment of apartment”) and you want to find the corresponding variable with information on that question in wave 1 (2010).

This is how you would proceed:

- Current wave: “f10” +
- Household Questionnaire: “h” +
- 3-Digit Nr.: “010”

Table 1: Overview of Original Data Files and Corresponding Questionnaires

<table>
<thead>
<tr>
<th>Information on</th>
<th>Dataset</th>
<th>Questionnaire Name (pdf-file)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person Level (Adult)</td>
<td>$p</td>
<td>$Person_Qn (except Questions with prefix “L”)</td>
</tr>
<tr>
<td></td>
<td>$lela</td>
<td>$Person_Qn (all Questions with prefix “L”)</td>
</tr>
<tr>
<td></td>
<td>$spkal</td>
<td>$Person_Qn</td>
</tr>
<tr>
<td></td>
<td>$jugend</td>
<td>$Youth_Qn</td>
</tr>
<tr>
<td></td>
<td>$eltern1</td>
<td>$Parent_Qn1</td>
</tr>
<tr>
<td></td>
<td>$eltern2</td>
<td>$Parent_Qn2</td>
</tr>
<tr>
<td></td>
<td>$eltern3</td>
<td>$Parent_Qn3</td>
</tr>
<tr>
<td></td>
<td>$eltern4</td>
<td>$Parent_Qn4</td>
</tr>
<tr>
<td></td>
<td>$eltern5</td>
<td>$Parent_Qn5</td>
</tr>
<tr>
<td></td>
<td>$eltern6</td>
<td>$Parent_Qn6</td>
</tr>
<tr>
<td>Household Level</td>
<td>$h</td>
<td>$Household_Qn</td>
</tr>
</tbody>
</table>

The variable you are searching for is variable F10H010. Looking for it in the dataset $f10h$, you find that there are 13 variables called F10H010, each having a suffix. Going back to the questionnaire you can see that there are 13 answer possibilities (so-called items) for the “equipment of apartment”-question. Variables that correspond to questions of this type (i.e.
more than one answer is possible) generally receive an “A” for the first item, a “B” for the second and so forth. Hence, by taking a look into the respective questionnaire it is easily possible to match each variable with the corresponding question and vice versa. There is a slight exception to the above construction rule: in case of the “Parent-Questionnaires”, the 3-digit Number refers to the questionnaire number (i.e. “1” to “6”) in the first digit, while the last two identify the question number. Thus, a variable called f10e104 would indicate to look for Parent-Questionnaire 1 and the fourth question (i.e. you should read: f10-e1-04).  

**Combining data files**

Sometimes it is necessary to combine information from various datasets. For example, if someone working with the generated *bioage* data files (children specific information) wants to have more information on mothers, she needs to take that information from another file (e.g. the p-files). This can be easily done by merging the respective datasets. The variables of the using dataset are then added to the already existing observations in the master dataset. When merging data files it is crucial to know which variable uniquely identifies observations in the master and using file. Such a key variable is called an identifier and varies from dataset to dataset.

Usually the best way to merge is to use datasets which use the same logical observation to organize the data. For example, if you were interested in household income (e.g. *Shgen* dataset), and wanted to add information on whether a household has an immigrant background or not, you would need to merge the *Shgen* dataset with the *Shbrutto* dataset. The identifier to match the two would be the original household identification number (HHNR) or the current household identification number (HHNRAKT). But even if the logical observation organizing the data is a person and you are interested in some information based on household, you are able to match each person with the corresponding household by specifying the household identifier as the key variable.

It should be stressed that if your information is based on the household level and you want to keep it that way, you should refrain from merging it with information based on the person level (since a household includes various respondents and observations thus cannot be uniquely identified). Nevertheless, if you want to keep the household as your logical

---

5 Some known exceptions should be noted: In the biography section of the person questionnaire, the questions L62 to L62d have been coded as $L062A to$L062E. In the household questionnaire for 2010, question 42b has been coded as F10H043, and question 43 has been named F10H044. For the household questionnaires starting with wave 2 (2011) exceptions are questions W1, W2, W3, W4, and W5, which are named $H001A, $H001B, $H001C, $H001D, and $H001E. Additionally, question 44b is coded as $H044Z.
observation, but still want use information from a dataset based on person level, you can do so by first collapsing the person-based information to the household level.

**Table 2 Overview of identifiers**

<table>
<thead>
<tr>
<th>Household Identifiers</th>
<th>Person Identifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHNRAKT</td>
<td>PERSNR</td>
</tr>
<tr>
<td>$HHNR</td>
<td>PERSNRM, PERSNRRRESP</td>
</tr>
<tr>
<td>HHNR</td>
<td>COUPID</td>
</tr>
<tr>
<td></td>
<td>PARTNO$$</td>
</tr>
<tr>
<td></td>
<td>FATHNO$$/MOTHNO$$</td>
</tr>
</tbody>
</table>

As a further complication, datasets on the person level can either have adults or children as the logical observation organizing the data, with the identifiers PERSNR and PERSNRM or PERSNRRRESP. It is important to keep this in mind when merging datasets. Going back to the example above, consider augmenting the `bioage01` data files with more specific information on the mothers of these children. The `bioage01` data files are organized based on children, thus PERSNR refers to a specific child. Parents, or in our example, mothers can be identified using the variable PERSNRRRESP. Hence a merge of `bioage01` to `f10p` using PERSNR would lead to no matching observations, as children are not included in `f10p`. To achieve the merge, PERSNR in either `f10p` or `bioage01` would have to be renamed (e.g. rename PERSNR in `bioage01` to PERSNRK, then rename PERSNRRRESP to PERSNR, and merge `f10p` via PERSNR.) Similarly, children can be matched to their parents by using either FATHNO$ (PERSNR of the father in `$kind`) or MOTHNO$ (PERSNR of the mother in `$kind`) and partners can be matched by using PARTNO$ (PERSNR of the partner in `$pgen`).
**FiD data files**

**Basic Data Files**

Basic data files are those which allow the user to retrieve basic information on the unit of observation. *ppfad* and *hpfad* are datasets with which the user can monitor the development of each person and household through the panel life, along with some generated information. The *Spbrutto* and *Shbrutto* files provide similarly important information about the interviewing process for each wave. *hbrutt10_fid* and *hbrutt11_fid* contain the gross sample with which the survey started, i.e. the Cohort and Screening Gross Sample in 2010 and the Screening Gross Sample in 2011.

**ppfad**

Sample: all individuals in FiD  
Unit of Observation: Person  
Identifier: PERSNR (HHNR, $HHNR)  
Description: Generated person level path variables (basic demographic variables) follow a person’s path throughout her panel life.  
Documentation: Documentation *ppfad*

**hpfad**

Sample: all participating households in FiD  
Unit of Observation: Household  
Identifier: HHNR ($HHNR)  
Description: Generated household level path variables, follows a household’s path throughout its panel life.  
Documentation: Documentation *hpfad*

**Shbrutto**

Sample: all participating households in FiD  
Unit of Observation: Household  
Identifier: HHNR (HHNRAKT, $HHNR)  
Description: Information about all participating households in the respective wave  
Documentation: Documentation *shbrutto*

**Spbrutto**

Sample: all individuals in FiD  
Unit of Observation: Person  
Identifier: PERSNR (HHNR, HHNRAKT, $HHNR)  
Description: Information about all individuals in the respective wave  
Documentation: Documentation *spbrutto*
### hbrutt10_fid

<table>
<thead>
<tr>
<th>Sample</th>
<th>all households selected to participate in FiD in 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit of Observation</td>
<td>Household</td>
</tr>
<tr>
<td>Identifier</td>
<td>HHNR (HHNRAKT, $HHNR)</td>
</tr>
<tr>
<td>Description</td>
<td>Information about all households selected to participate in FiD in 2010 for the first time, i.e. the screening population of low income families, single parents, and families with more than two children as well as the cohort sample. This gross sample file contains information for non-responding households as much as possible.</td>
</tr>
<tr>
<td>Documentation</td>
<td>There is no specific documentation, but the documentation on <em>hbrutto</em> will help the user.</td>
</tr>
</tbody>
</table>

### hbrutt11_fid

<table>
<thead>
<tr>
<th>Sample</th>
<th>all households selected to participate in FiD in 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit of Observation</td>
<td>Household</td>
</tr>
<tr>
<td>Identifier</td>
<td>HHNR (HHNRAKT, $HHNR)</td>
</tr>
<tr>
<td>Description</td>
<td>Information about all households selected to participate in FiD in 2011 for the first time, i.e. the screening population of single parents and families with more than two children. This gross sample file contains information for non-responding households as much as possible.</td>
</tr>
<tr>
<td>Documentation</td>
<td>There is no specific documentation, but the documentation on <em>hbrutto</em> will help the user.</td>
</tr>
</tbody>
</table>
### Original Data Files

Original data files are those, which are left unchanged and do not include any generated variables (see section 1 for a more specific explanation). All generated data files are built up from the original data files and can be reproduced in that way. Nonetheless, we do not distribute the answers from open-ended questions in the original data files for data protection reasons. As a matter of fact, no string variables are included in the data, except for the data in $pkal$. However, string data are coded into categories if possible and are then included as numerical values in the generated files.

#### $Sh$

**Sample**

all households participating in the survey

**Unit of Observation**

Household

**Identifier**

HHNR, HHNRAKT, $HHNR$

**Description**

Annual questions concerning the household as a whole

**Documentation**

$Household\_Qn$

#### $Sp$

**Sample**

adults turning 18 in the current survey year or older

**Unit of Observation**

Person

**Identifier**

PERSNR (HHNR, HHNRAKT, $HHNR$)

**Description**

Annual individual questions

**Documentation**

$Person\_Qn$ (without L-questions)

#### $lela$

**Sample**

adults turning 18 in survey year or older, answering for the first time

**Unit of Observation**

Person

**Identifier**

PERSNR (HHNR, HHNRAKT, $HHNR$)

**Description**

Biographical person questions that are part of the Personal Questionnaire, questions are named by L#. In FiD 2010 there was one biographical part in questions L1 to L67. In 2011, a second part was added with questions L70 to L100, with the goal that persons participating for the first time in 2011, receive part 1, revisited respondents receive part 2, so that within two waves, a complete biography is gathered, that resembles the SOEP questionnaire. However, due to a technical error, all sample members of the samples drawn in 2010 received part 2, even if they entered the study only in 2011. The variable $LELTYP$ specifies which part of the biography questionnaire was filled out.

**Documentation**

$Person\_Qn$ (questions labeled by L#)
**Spkal**

Sample: adults turning 18 in survey year or older  
Unit of Observation: Person  
Identifier: PERSNR (HHNR, HHNRAKT, $HHNR)  
Description: Provides the calendar information from the personal questionnaire on what the respondent did in each month last year. Information is recorded in string variables.  
Documentation: $Person_Qn, activity calendar

**Sjugend**

Sample: all persons turning 17 in the survey year  
Unit of Observation: Person  
Identifier: PERSNR (HHNR, HHNRAKT, $HHNR)  
Description: The youth questionnaire is the first questionnaire a FiD member will ever fill out herself. Note that if a youth questionnaire has been conducted, no biography part will be conducted with this person in the future.  
Documentation: $Youth_Qn

**Seltern1**

Sample: newborn children or turning 1 in the survey year  
Unit of Observation: Person (child)  
Identifier: PERSNR, PERSNRM (HHNR, HHNRAKT, $HHNR)  
Description: Proxy parent questionnaire about children living in the household. Answered by the mother or single parent father living in the household.  
Documentation: $Parent_Qn1

**Seltern2**

Sample: children turning 2 in the survey year  
Unit of Observation: Person (child)  
Identifier: PERSNR, PERSNRM (HHNR, HHNRAKT, $HHNR)  
Description: Proxy parent questionnaire about children living in the household. Answered by the mother or father living in the household.  
Documentation: $Parent_Qn2

**Seltern3**

Sample: children turning 3 in the survey year  
Unit of Observation: Person (child)  
Identifier: PERSNR, PERSNRM (HHNR, HHNRAKT, $HHNR)  
Description: Proxy parent questionnaire about children living in the household. Answered by the mother or father living in the household.  
Documentation: $Parent_Qn3
**Seltern4**

Sample: children turning 6 in the survey year  
Unit of Observation: Person (child)  
Identifier: PERSNR, PERSNRM (HHNR, HHNRAKT, $HHNR)  
Description: Proxy parent questionnaire about children living in the household. Answered by the mother or father living in the household.  
Documentation: $Parent_Qn4

**Seltern5**

Sample: children turning 8 in the survey year  
Unit of Observation: Person (child)  
Identifier: PERSNR, PERSNRM (HHNR, HHNRAKT, $HHNR)  
Description: Proxy parent questionnaire about children living in the household. Answered by both the mother AND the father if living in the household. Note that this leads to multiple observations per child. PERSNR and PERSNRM can be used to identify a single observation.  
Documentation: $Parent_Qn5

**Seltern6**

Sample: children turning 10 in the survey year  
Unit of Observation: Person (child)  
Identifier: PERSNR, PERSNRM (HHNR, HHNRAKT, $HHNR)  
Description: Proxy parent questionnaire about children living in the household. Answered by both the mother AND the father if living in the household. Note that this leads to multiple observations per child. PERSNR and PERSNRM can be used to identify a single observation.  
Documentation: $Parent_Qn6

**Kind**

Sample: all children turning at most 16 during the survey year  
Unit of Observation: Person (child)  
Identifier: PERSNR, FATHNO10, MOTHNO10 (HHNR, HHNRAKT, $HHNR)  
Description: This file contains mainly original data from the so-called “children’s matrix” in the household questionnaire. The household head answers questions about all children living in the household. The data – originally on the household level – are then converted into children level information. Added to this data are the identifiers for mother and father as well as the type of their relationship (i.e. biological, social, or adoptive).  
Documentation: $Household_Qn, Documentation Kind

**Sluecke**

Sample: all eligible persons of the previous year who did not participate
<table>
<thead>
<tr>
<th><strong>Unit of Observation</strong></th>
<th>Person</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identifier</strong></td>
<td>PERSNR (HHNR, HHNRAKT, $HHNR)</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>If former FiD respondents decide not to participate in a given year, the following time the household is visited they are again asked to participate in the study. In such a case, they fill out the regular person questionnaire, but are also interviewed about their year of absence with a short questionnaire containing the main information. The information of this short questionnaire is stored in <em>$Luecke</em> (from German “Lücke”, gap), containing the exact same variable names of the previous year’s p-questionnaire. The data from this file belong to the previous wave and are hence stored in that year’s folder. (E.g. the data gathered in 2012 from respondents not participating in 2011 is called <em>f11luecke</em>, and this file is stored in the folder ~/2011.) Also, in <em>ppfad</em> the $NETTO code of the previous wave is changed accordingly (i.e. from 130 to 131).</td>
</tr>
<tr>
<td><strong>Documentation</strong></td>
<td>$Luecke_Qn</td>
</tr>
</tbody>
</table>
Generated Data Files

Generated data files are created for the user to ease the handling of the data. For the most part, they are constructed from original data which are also included in the data distribution. However, some files use information which for data protection reasons cannot be given to the user directly, for example detailed neighborhood information, which is used in the construction of the weights.

$pgen$
Sample: all adults turning 17 in survey year or older
Unit of Observation Person
Identifier: PERSNR (HHNR, HHNRAKT)
Description: Generated person-level variables
Documentation: Documentation on $pgen$

$shgen$
Sample: all households participating in the survey
Unit of Observation Household
Identifier: HHN, HHNRAKT
Description: Generated household-level variables
Documentation: Documentation on $shgen$

$artkalen$
Sample: all adults turning 18 in survey year or older.
Unit of Observation Person
Identifier: PERNSNR, SPELLNR (HHNR)
Description: The file contains monthly spells for events starting in January 2009, which are generated from the activity calendar $spkal$. This is in contrast to $pbiospe$ (see below), where spells are in yearly durations, and events prior to the first wave are included. $artkalen$ is longitudinal, as it combines information from all activity calendars. Observations in the data are uniquely identified through PERSNR and SPELLNR.
Documentation: There is no specific documentation; the activity calendar in $Person_Qn provides some help

$pbiospe$
Sample: all adults turning 18 in survey year or older.
Unit of Observation Person
Identifier: PERNSNR, SPELLNR (HHNR)
Description: The file contains yearly spells for events starting at the age of 15 for each respondent, generated from the biographical activity calendar
collected in the second part of the biography ($LELTYP=2 or 3). While the biography is recorded only once, $pbiospe is updated every year with spell information from $artkalen. $pbiospe is longitudinal, as it combines information from biography and all activity calendars. Observations in the data are uniquely identified through PERSNR and SPELLNR.

Documentation

Documentation on $pbiospe

$phrf and $phrf_fidsoep

Sample: all participating persons
Unit of Observation: Person
Identifier: PERSNR (HHNR, $HHNR)
Description: Person weighting factors for each wave, also includes the inverse probabilities of staying in the sample. $phrf_fidsoep provides this information for the integrated SOEP-FiD data.

Documentation: Documentation on $phrf and $hhrf

$hhrf and $hhrf_fidsoep

Sample: all participating households
Unit of Observation: Household
Identifier: HHNR, $HHNR
Description: Household weighting factors for each year, also includes the inverse probabilities of staying in the sample. $phrf_fidsoep provides this information for the integrated SOEP-FiD data.

Documentation: Documentation on $phrf and $hhrf

$smipinc

Sample: all persons participating in the survey
Unit of Observation: Person
Identifier: PERSNR, _MJ (HHNR, HHNRAKT)
Description: Contains multiply imputed data on personal incomes. In this specific dataset observations are uniquely identified by the combination of PERSNR with the identifier for each imputed implicate, _MJ.

Documentation: Documentation on $mihinc and $mipinc

$smihinc

Sample: all households participating in the survey
Unit of Observation: Household
Identifier: HHNR, _MJ, HHNRAKT
Description: Contains multiple imputed data on household incomes and other payments. In this specific dataset observations are uniquely identified by the combination of HHNR or HHNRAKT with the identifier for each imputed implicate, _MJ.
General information on the FiD data distribution

Documentation

Documentation on **mihinc** and **mipinc**

**bioage01**
Sample: newborn children (aged 0-1 during survey year)
Unit of Observation: Person (child)
Identifier: PERSNR, PERSNRRESP (HHNRAKT)
Description: Contains information about children when they were 0-1 years old and living in the household. This file is not longitudinal, but contains multiple cross sections as children are born into this age group.
Documentation: Documentation on **bioage** files

**bioage02**
Sample: children aged 1-2 during survey year
Unit of Observation: Person (child)
Identifier: PERSNR, PERSNRRESP (HHNRAKT)
Description: Contains information about children when they were 1-2 years old and living in the household. This file is not longitudinal, but contains multiple cross sections as children grow into this age group.
Documentation: Documentation on **bioage** files

**bioage03**
Sample: children aged 2-3 during survey year
Unit of Observation: Person (child)
Identifier: PERSNR, PERSNRRESP (HHNRAKT)
Description: Contains information about children when they were 2-3 years old and living in the household. This file is not longitudinal, but contains multiple cross sections as children grow into this age group.
Documentation: Documentation on **bioage** files

**bioage06**
Sample: children aged 5-6 during survey year
Unit of Observation: Person (child)
Identifier: PERSNR, PERSNRRESP (HHNRAKT)
Description: Contains information about children when they were 5-6 years old and living in the household. This file is not longitudinal, but contains multiple cross sections as children grow into this age group.
Documentation: Documentation on **bioage** files

**bioage08p1** and **bioage08p2**
Sample: children aged 7-8 during survey year
Unit of Observation: Person (child)
Identifier: PERSNR, PERSNRRESP (HHNRAKT)
**General information on the FiD data distribution**

**FiD-Documentation**

SOEP Survey Paper 332

Description

Contains information about children when they were 7-8 years old and living in the household. This file is not longitudinal, but contains multiple cross sections as children grow into this age group. Note that bioage08p1 contains all children for which a questionnaire was filled out. bioage08p2 adds information from an additional parent who has filled out the questionnaire.

Documentation

Documentation on bioage files

**bioage10p1 and bioage10p2**

Sample

children aged 9-10 during survey year

Unit of Observation

Person (child)

Identifier

PERSNR, PERSNRRESP (HHNRAKT)

Description

Contains information about children when they were 9-10 years old and living in the household. This file is not longitudinal, but contains multiple cross sections as children grow into this age group. Note that bioage10p1 contains all children for which a questionnaire was filled out. bioage10p2 adds information from an additional parent who has filled out the questionnaire.

Documentation

Documentation on bioage files

**bioage1**

Sample

all children in bioage01-bioage10

Unit of Observation

Person (child)

Identifier

PERSNR, BIOAGE, PERSNRRESP (HHNRAKT)

Description

Contains the combined information about children who were covered in the bioage files in a longitudinal format, i.e. multiple observations per child are possible.

Documentation

Documentation on bioage files

**Biobirth**

Sample

all adults at least 17 years old

Unit of Observation

Person

Identifier

PERSNR (HHNR)

Description

Contains information about a respondent’s biological children and the situation surrounding the birth. This information is mainly based on the biography part of the person questionnaire.

Documentation

Documentation on biobirth

**biomarsy**

Sample

all adults at least 17 years old

Unit of Observation

Person

Identifier

PERSNR SPELLNR (HHNR)

Description

Contains yearly spell information on the respondent’s martial history. This information is mainly based on the biography part of the person
questionnaire. Unique observations are identified through PERSNR and SPELLNR.

Documentation

Documentation on *biomarsy* and *biocouply*

### biocouply

**Sample**
all adults at least 17 years old

**Unit of Observation**
Person

**Identifier**
PERSNR SPELLNR (HHNR)

**Description**
Contains yearly spell information on the respondent’s partner history, including information on marriages. This information is mainly based on the biography part of the person questionnaire. Unique observations are identified through PERSNR and SPELLNR.

Documentation
Documentation on *biomarsy* and *biocouply*

### bioresid

**Sample**
all adults at least 17 years old

**Unit of Observation**
Person

**Identifier**
PERSNR (HHNR, HHNRAKT)

**Description**
Contains information on when a person moved into her current accommodation, and whether secondary accommodations exist. The information so far stems from the biography questionnaire and information on moves during the panel.

Documentation
No documentation yet, please refer to the SOEP documentation for this dataset.

### bioparen

**Sample**
all adults at least 17 years old

**Unit of Observation**
Person

**Identifier**
PERSNR (HHNR, HHNRAKT)

**Description**
Contains information on a person’s parents, mainly information from the second part of the biography questionnaires. This dataset contains the same variable names as the corresponding SOEP data, and can thus be used easily in combination with these data.

Documentation
Documentation on *bioparen*

### bioage17

**Sample**
youths turning 17 years during the survey year

**Unit of Observation**
Person

**Identifier**
PERSNR (HHNR, HHNRAKT)

**Description**
Contains information from the youth questionnaires (*Sjugend*) in a combined format, i.e. all persons who ever answered this questionnaire are included in this dataset (similar to the *bioage01 - bioage06* files).

Documentation
Documentation on *bioage17*. 
paradatal

- **Sample**: all respondents
- **Unit of Observation**: Questionnaire
- **Identifier**: SYEAR PERSNR PERSNRK, $QSTNR (HHNR, HHNRAKT, $HHNR)
- **Description**: Provides background information about the interview (interviewer ID, date of interview, mode, etc.) for all survey years. It is based on the questionnaire level, hence unique observations are identified only through the combination of SYEAR, PERSNR, PERSNRK and $QSTNR, the identifier of the questionnaire answered. (Note that prior to FiDv3.1, these data were distributed in their yearly format in the datasets [paradata](#).)
- **Documentation**: Documentation on [paradatal](#)

Datasets not in FiD, which are known from the SOEP

There are a couple of datasets which users from the SOEP may miss when skimming through the FiD datasets. Some of the data sets simply could not be produced, because the data are not available, i.e. the information asked for does not include the information needed to create the data sets. This pertains to the datasets [Shv](#), [bioimmig](#), and [abroad](#). Other datasets could not be provided simply due to the amount of work necessary to produce them (prominently [Spequiv](#)). These will be provided as FiD is integrated into the SOEP.
Documentation \textit{ppfad}

Person related meta-dataset

This documentation is based on the comparable SOEP documentation on \textit{ppfad} and has benefited from previous work. For readability reasons, we do not specifically cite and specify text that has been used directly from the SOEP document.
General information

This file is designed to support longitudinal analysis when linking personal information from various waves. Each person documented in at least one Spbrutto file is also in ppfad. The only sorting key is the never-changing person ID number PERSNR. The $HHN R variables contain the current household number of the particular household in which the person lived and was interviewed in at the time of the survey, either as an adult household member or as a child. The $NETTO variables give the wave-specific interview status, and tell which dataset contains a person’s record in a particular wave (Sp, Skind, etc.) or whether the person has permanently left FiD.

In addition, ppfad contains basic demographic variables, which provide longitudinally consistent information over all waves:

- SEX (sex)
- GEBJAHR (4-digit birth year)
- GEBMONAT (2-digit birth month)
- TODJAHR (4-digit year of death, if applicable)
- IMMIYEAR (year of first immigration to Germany)
- GERMBORN (born in Germany)
- CORIGIN (country of origin)
- MIGBACK (migration background)
- MIGINFO (information source for MIGBACK)

It is recommended that these tested variables be used for cross-sectional and longitudinal analyses. These variables are adjusted on a wave-by-wave basis in the framework of demographic testing.

Further methodological variables (year, four-digit) are included to provide more information on the sample composition:

- EINTRITT: year a person joined FiD, e.g. the year a person was first included in Spbrutto
- AUSTRITT: year a person ultimately left FiD
- ERSTBEFR: year of a person’s first interview
- LETZTBEFR: year of a person’s most recent interview
### List of variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHNR</td>
<td>24</td>
</tr>
<tr>
<td>PERSNR</td>
<td>24</td>
</tr>
<tr>
<td>SHHNR</td>
<td>24</td>
</tr>
<tr>
<td>PSAMPLE</td>
<td>24</td>
</tr>
<tr>
<td>SEX</td>
<td>25</td>
</tr>
<tr>
<td>GEBJAHR</td>
<td>25</td>
</tr>
<tr>
<td>GEBMONAT</td>
<td>25</td>
</tr>
<tr>
<td>GEBMOVAL</td>
<td>25</td>
</tr>
<tr>
<td>TODJAHR</td>
<td>26</td>
</tr>
<tr>
<td>TODINFO</td>
<td>26</td>
</tr>
<tr>
<td>LOC1989</td>
<td>26</td>
</tr>
<tr>
<td>SNETTO</td>
<td>26</td>
</tr>
<tr>
<td>SNETOLD</td>
<td>27</td>
</tr>
<tr>
<td>SCASEMAT</td>
<td>28</td>
</tr>
<tr>
<td>SPOP</td>
<td>28</td>
</tr>
<tr>
<td>SSAMPREG</td>
<td>28</td>
</tr>
<tr>
<td>GERMBOREN</td>
<td>29</td>
</tr>
<tr>
<td>IMMIYEAR</td>
<td>29</td>
</tr>
<tr>
<td>CORIGIN</td>
<td>29</td>
</tr>
<tr>
<td>MIGBACK</td>
<td>30</td>
</tr>
<tr>
<td>MIGINFO</td>
<td>31</td>
</tr>
</tbody>
</table>
HHNR
Variable label  “Original household number”
Variable format  7-digit integer
Comment  Identifier for the household in which a person lived at the time she was captured in $pbrutto$ for the first time. Note that this does not mean that a person had an interview in this household – they could be children or non-responding household members for other reasons as well. People entering a household at a later time receive the original household number as their HHNR, even if the household is a split household.

PERSNR
Variable label  “Never changing Person ID”
Variable format  8-digit integer
Comment  Identifies each person uniquely and for all waves and datasets. Note that some statistical software programs have issues with the length of the variable. E.g. in Stata™, note that when generating a variable identical to PERSNR, you will have to specify that the new variable is in “long” format, i.e. “gen long varx=persnr”.

$HHNR$
Variable label  “Current household number $$$$”
Variable format  7-digit integer
S - Wave  F10, F11, F12, F13
Comment  Identifier for the household in which a person lived at the time of the current wave interview. $HHNR$ is set to “-2 does not apply” if a person was not in a FiD-household in the respective wave.

PSAMPLE
Variable label  “Subsample”
Value label  PSAMPLE
(61) FiD 2007 Birth Cohort
(62) FiD 2008 Birth Cohort
(63) FiD 2009 Birth Cohort
(64) FiD 2010 Birth Cohort
(65) FiD Screening (sampled 2010)
(66) FiD Screening (sampled 2011)
Variable format  2-digit integer
Comment  Note that this variable is included in all datasets, and provides information whether the household or person originates from the cohort or one of the screening samples in FiD. In $kind$ it is also named PSAMPLE, in $hpfad$ it is named HSAMPLE; in all other datasets it is called SAMPLE1, which is analogous to the SOEP notation.
**SEX**
Variable label  “Sex”
Value label  SEX
(1) male
(2) female
Variable format  1-digit integer

**GEBJAHR**
Variable label  “Year of birth”
Variable format  4-digit integer
Comment  Respondent’s year of birth, checked for consistency across survey years. Note that this information may vary across years due to misinformation, but always provides the best information available. Hence GEBJAHR should be used for any analysis involving the age of respondent (or child).

**GEBMONAT**
Variable label  “Month of birth”
Value label  GEBMONAT
(1) January
(2) February
(3) March
(4) April
(5) May
(6) June
(7) July
(8) August
(9) September
(10) October
(11) November
(12) December
Variable format  2-digit integer
Comment  Respondent’s month of birth, checked for consistency across survey years. Note that this information may vary across years due to misinformation, but always provides the best information available.

**GEBMOVAL**
Variable label  “Month of birth, data source”
Value label  GEBMOVAL
(1) Generated from GEBMONAT (parents)
(2) Info from ppfad
(3) Info from $kind
(4) Info from $p
(5) Info from $lela
(6) Info from $bioage
(7) Info from Sjugend

Comment GEBMOVAL provides information from which data source the month of birth was taken from. In principle, GEBMONAT can come from four different sources:
1) it can be taken from a parent-questionnaire;
2) it can come from a person’s biography information;
3) it can be provided in a regular person interview;
4) for children, it can come from the household interview and the children’s matrix; or
5) it can be generated through other means (e.g. biographical information from other persons).
The hierarchy is exactly in the above order, meaning that the first source is preferred to the second and so on, which also clears conflicts between the sources.

TODJAHR
Variable label “Year of death, four-digit”
Variable format 4-digit integer

TODINFO
Variable label “Year of death, information source”
Value label TODINFO
   (1) from annual survey (pbr_exit)
Variable format 2-digit integer
Comment TODINFO provides information from which data source information on the death of respondent was taken. In FiD, there is only one source at the moment, which is the annual survey.

LOC1989
Variable label “Where did you live in 1989?”
Value label LOC1989
   (1) East Germany (GDR) incl. East Berlin
   (2) West Germany (FRG) incl. West Berlin
   (3) abroad (foreign country)
Variable format 2-digit integer
Comment This variable captures where the respondent lived at the time of the German reunification. This information is taken from the second part of the biography questionnaire in FiD, which was fielded in 2011 for the first time. Hence if respondents dropped out of the sample before answering the second part of the biography questionnaire, they will have a missing value in this variable.

$NETTO
Variable label “Current Wave Survey Status”
Value label $NETTO$

(110) FiD: Person questionnaire
(111) FiD: P-questionnaire & LeLa I
(112) FiD: P-questionnaire & LeLa II
(113) FiD: P-questionnaire & LeLa I+II
(117) FiD: Youth questionnaire, 17
(118) FiD: P-questionnaire and below 17
(119) FiD: P-questionnaire, no household
(120) FiD: Child in part. household (_KIND)
(121) FiD: Parent Questionnaire 1 (0-1)
(122) FiD: Parent Questionnaire 1 2 (1-2)
(123) FiD: Parent Questionnaire 1 3 (2-3)
(124) FiD: Parent Questionnaire 1 4 (5-6)
(125) FiD: Parent Questionnaire 1 5 (7-8)
(126) FiD: Parent Questionnaire 1 6 (9-10)
(130) FiD: in Gross-Sample-HH w/o P-Int.
(131) FiD: Gap Questionnaire (_LUECKE)
(132) FiD: Biography only, no household
(133) FiD: Youth questionnaire, no household
(135) FiD: Parent questionnaire, no household
(160) FiD: Only Questionnaire Without Indiv. And HH Interview
(161) FiD: Gap Interview without HH reference
(162) FiD: Gap Interview with drop out
(170) FiD: Only Participation In Tests, Experiments, etc.
(180) FiD: Individual Without Any Current Information
(181) FiD: Prior Interviewee Without Any Current Information
(188) FiD: Repatriate - (moved abroad before [191])
(189) FiD: Repatriate - (was drop out [190])
(190) FiD: Drop-outs
(191) FiD: Moved abroad
(199) FiD: Deceased

Variable format 3-digit integer
$ - Wave F10, F11, F12, F13$

Comment $NETTO$ follows the SOEP logic in principle, although some of the provided value labels are not relevant for FiD at the moment. The main difference is that FiD uses 3-digit numbers and features an integrated personal interview, capturing the biography information as well as the personal information (codes 111, 112, 113).

Note that code 131 (Gap-Questionnaire) is set retrospectively for the year of the gap, e.g. if a person missed the 2011 interview and fills out the gap questionnaire in 2012, the F11$NETTO$ code is changed accordingly from 130 (gross sample only) to 131.

$NETOLD$

Variable label “Current Wave Survey Status”
Value label $NETOLD$

(0) Person Gap (YPBRUTTO)
(1) Successful Interview (_P [,_JUGEND])
(2) Below Survey Age (_KIND)
(3) Did not participate (_PBRUTTO)
(4) Missing this wave (_PLUECKE)
(5) Interviewee without household interview

Variable format 1-digit integer
$ - Wave F10, F11, F12, F13

Comment This variable is a compressed version of the variable $NETTO.

$CASEMAT
Variable label “Case-match, combined panel households”
Variable format 6-digit integer
$ - Wave F10, F11, F12, F13

Comment This variable is built after the SOEP to capture the rare event of two FiD-households moving together. This event is extremely unlikely, so this variable is missing (“-2 Does not apply”) for all cases at the moment.

$POP
Variable label “Sample membership $$$$”
Value label $POP
(1) Private household, German HH-head
(2) Private household, foreign HH-head
(3) Institutional household, German HH-head
(4) Institutional household, foreign HH-head
(5) Not Completed Private HH, German HH-Head
(6) Not Completed Private HH, Foreign HH-Head
(7) Not Completed Institutional HH, German HH-Head
(8) Not Completed Institutional HH, Foreign HH-Head

Variable format 1-digit integer
$ - Wave F10, F11, F12, F13

Comment $POP was derived from $pbrutto, using the variables $pnat and $stell. Note that a value of “-1 No answer” is possible, if the nationality was not reported by the respondent.

$SAMPREG
Variable label “Current wave sample region $$$$”
Value label $SAMPREG
(1) West Germany
(2) East Germany

Variable format 1-digit integer
$ - Wave F10, F11, F12, F13
Comment $SSAMPREG$ specifies whether the household is located in the former Western or Eastern part of Germany, the “old” or “new” Länder, respectively. This information uses the old borders as they were defined at the time of the reunification, hence the former West Berlin is labeled as $SSAMPREG=1$. This information is time-dependent; thus for each year, there is the wave-specific information $SSAMPREG$.

**GERMBORN**
Variable label “Born in Germany or immigration prior to 1949”
Value label GERMBORN
(1) German born or immigrant prior to 1949
(2) immigrant after 1948
Variable format 1-digit integer
Comment GERMBORN is taken from the biography information the respondents provide. If this information is missing due to item or unit non-response, the variable is set to “-1 No answer”.

**IMMIYEAR**
Variable label “Year of immigration to Germany after 1948”
Variable format 4-digit integer
Comment GERMBORN is taken from the biography information the respondents provide. For all individuals born in Germany, this information is set to “-2 Does not apply”. If this information is missing due to item or unit non-response, the variable is set to “-1 No answer”.

**CORIGIN**
Variable label “Country of origin”
Value label CORIGIN (see below)
Variable format 3-digit integer

(001) Germany
(002) Turkey
(003) Ex-Yugoslavia
(004) Greece
(005) Italy
(006) Spain
(007) Ex-GDR
(008) Austria
(011) France
(012) Benelux
(013) Denmark
(014) Great Britain
(015) Sweden
(016) Norway
(017) Finland
(018) USA
(019) Switzerland
(020) Chile
(021) Rumania
(022) Poland
(023) Korea
(024) Iran
(025) Indonesia
(026) Hungary
(027) Bolivia
(028) Portugal
(029) Bulgaria
(030) Syria
(031) Czech Republic
(032) Russia
(033) Kurdistan
(034) Mexico
(035) Argentina
(036) Cape Verde Is.
(037) Benin
(038) Philippines
(039) Israel
(040) Japan
(041) Australia
(042) India
(043) Afghanistan
(044) Thailand
(045) Jamaica
(046) Saudi-Arabia
(047) Ethiopia
(048) Columbia
(049) Ghana
(050) Bangladesh
(051) Venezuela
(052) Tunisia
(053) Mauritius
(054) Nigeria
(055) Canada
(056) New Zealand
(057) Tanzania
(058) Cyprus
(059) Cuba
(060) Iraq
(061) Brazil
(062) Monaco
(063) Hong Kong
(064) Peru
(065) Sri Lanka
(066) Nepal
(067) Morocco
(068) China
(069) Liechtenstein
(070) Iceland
(071) Ireland
(072) St. Lucia
(073) Moldavia
(074) Kazakhstan
(075) Albania
(076) Lebanon
(077) Kyrgyzstan
(078) Ukraine
Comment  CORIGIN is generated with information from $lela$ and the youth questionnaire. If GERMBORN is true (=1), the country is Germany. If children were born before the immigration, the CORIGIN of the mother is transferred to them; if this was not available, the CORIGIN of the father was used.

MIGBACK

Variable label  “Migration Background”

Value label  MIGBACK

   (1) No migration background
   (2) Direct migration background
   (3) Indirect migration background
   (4) Migration background, not differentiated

Variable format  1-digit integer

Comment  Generally MIGBACK is defined as in the SOEP. Due to a different information basis there are some slight differences (information about respondent’s parents migration status is only available for the second wave respondents, except mother tongue). We use information on country of birth, citizenship, immigration, immigration group, mother tongue of parents and citizenship of parents where available. A direct migration background was assigned if the person migrated herself (i.e. was not born in Germany). Otherwise an indirect migration background was assigned, when one of the person’s parents migrated. For children (under the age of 17), migration background is set to be direct, if their biological parents moved after they were born. Consequently, if the parents had already been in Germany at the time of birth, the migration background becomes indirect. Note that there is a
difference to the SOEP here: as FiD has more detailed information on the parent-child relationship, the migration background is only derived if at least one of the child’s biological parents is known.

**MIGINFO**

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Information Source for MIGBACK”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value label</td>
<td>MIGINFO</td>
</tr>
<tr>
<td></td>
<td>(1) Direct info without parental info</td>
</tr>
<tr>
<td></td>
<td>(2) Proxy info without parental info</td>
</tr>
<tr>
<td></td>
<td>(3) Direct info with parental info</td>
</tr>
<tr>
<td></td>
<td>(4) Proxy info with parental info</td>
</tr>
</tbody>
</table>

Variable format 1-digit integer
Documentation *hpfad*

Household related meta-dataset

*Rainer Siegers*

This documentation is based on the comparable SOEP documentation on *hpfad* and has benefited from previous work. For readability reasons, we do not specifically cite and specify text that has been used directly from the SOEP document.
General information

This dataset is designed to assist linking data from different waves for the substantive evaluation of household level data. This dataset includes all households that ever had contact with FiD in any wave up to the present. Thus a household interview does not necessarily have to have taken place in the current wave; it is enough to simply have appeared once in any one of the “$hhbrutto” household datasets. The sorting keys are the current household numbers (HHNRAKT and $SHHN). While “old” households retain their old HHNRAKT in case of a move (e.g. F10HHNR=F11HHNR), “new” panel households (e.g. formed by members of an existing panel household that have moved into a new household) receive a different HHNRAKT upon their first interview in a particular wave.

The variables $SHHN also contain the current household number if the household entered the sample in the respective wave. In any years prior to this first interview, the variable $SHHN is “missing”, which then helps to prevent mistakes in linking household level data across time. The variables $SHNETTO give information on whether it was possible to carry out a HH interview and whether it would pay off to create a link to the corresponding data in that wave’s $H file.
List of variables:

- HHNR ................................................................. 35
- HHNRAKT ......................................................... 35
- HSAMPLE .......................................................... 35
- $HHNR .............................................................. 35
- $SAMPREG ........................................................ 36
- $HPOP ............................................................... 36
- $HNETTO .......................................................... 36
- SINGPA .............................................................. 37
- LRGFAM ............................................................. 37
- LOWINC ............................................................ 37
HHNR
Variable label “Original household number”
Variable format 7-digit integer

Comment HHNR denotes the original household number of the household and allows relating households that have split over the years. For a household in its first wave, HHNR always equals HHNRAKT and $HHNR. After that, a HHNR differing from the current household number indicates a break-off household that has successfully been followed.
Note that individuals moving into a break-off household receive the HHNR of the household, even though they never lived in it.

HHNRAKT
Variable label “Current household number”
Variable format 7-digit integer

Comment HHNRAKT is the current household number. It is identical to the wave specific household number, $HHNR. HHNRAKT is not updated if a household is no longer in the sample, hence the variable is never missing (i.e. “-2 does not apply”). This is a key difference to $HHNR, which can be missing (either if a household is not yet in the panel in a specific wave or has dropped out).

HSAMPLE
Variable label “Subsample”
Value label HSAMPLE
(61) FiD 2007 Birth Cohort
(62) FiD 2008 Birth Cohort
(63) FiD 2009 Birth Cohort
(64) FiD 2010 Birth Cohort
(65) FiD Screening (sampled 2010)
(66) FiD Screening (sampled 2011)
Variable format 2-digit integer

Comment HSAMPLE is fixed across waves and households and denotes the sample membership at first contact with FiD (at the point of selection of the original household into the sample).

$HHNR
Variable label “Household number $$$$”
$ - Wave F10, F11, F12, F13
Variable format 7-digit integer

Comment $HHNR denotes the household’s current identifier. It is set to missing (“-2 does not apply”) in case the household is not yet or no longer in the
sample. E.g., all households in the Screening-Sample of 2011 (HSAMPLE=66) have a F10HHNR of “-2”, as they entered the sample in 2011 only.

$\text{SAMPREG}$
Variable label “Current wave sample reason $$$$$$
Value label $\text{SAMPREG}$
   (1) West Germany (former FRG)
   (2) East Germany (former GDR)
$ - \text{Wave} $ F10, F11, F12, F13
Variable format 1-digit integer
Comment $\text{SAMPREG}$ specifies whether the household is located in the former Western or Eastern part of Germany, the “old” or “new” Länder, respectively. This information uses the old borders as they were defined at the time of the reunification, hence the former West Berlin is labeled as $\text{SAMPREG}=1$. This information is time-dependent; thus for each year, there is the wave-specific information $\text{SAMPREG}$.

$\text{HPOP}$
Variable label “Sample membership $$$$$
Value label $\text{HPOP}$
   (1) private household, German head
   (2) private household, foreign head
   (3) institutional household, German head
   (4) institutional household, foreign head
$ - \text{Wave} $ F10, F11, F12, F13
Variable format 1-digit integer
Comment $\text{HPOP}$ is derived from $\text{WUM2}$ (in $\text{Shbrutto}$ differentiating private from institutional households) as well as $\text{SPNAT}$ and $\text{SSTELL}$ (nationality and relationship to household head in $\text{Shbrutto}$). Missing values are imputed taking into account respondent’s history. Thus the only admissible missing value is “-2 does not apply”.

$\text{HNETTO}$
Variable label “Wave survey status $$$$$$
Value label $\text{HNETTO}$
   (1) successful household interview
   (2) household in gross sample only
Variable format 1-digit integer
$ - \text{Wave} $ F10, F11, F12, F13
Comment The wave-specific survey status sets $\text{HNETTO}$ to 1 if there is an interview in the respective year, i.e. an entry in $\text{Sh}$ or in the $\text{Sp}$ (note
that for very few cases there is no household interview, but one or more person interviews). For $HNETTO=2$ there is only an entry in the file $Shbrutto$, but the household was not interviewed (the reason for which is given by the variable $HERGS$ in the file $Shbrutto$).

**SINGPA**

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Single-parent-household at first contact”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value label</td>
<td>SINGPA</td>
</tr>
<tr>
<td></td>
<td>(1) Yes</td>
</tr>
<tr>
<td></td>
<td>(2) No</td>
</tr>
<tr>
<td>Variable format</td>
<td>1-digit integer</td>
</tr>
</tbody>
</table>

**Comment**

SINGPA is a time-fixed indicator providing information whether a household was defined as “single-parent-household” the first time it was interviewed. This information belongs to the original household and is thus associated with the variable HHNR. It is carried over to all split households from the original household. In the Screening-Samples 2010 and 2011, this corresponds to the sampling-design-definition (exactly one adult person with at least one child younger than 18) at the point of selection of the original household into the sample. For the Cohort-Sample, this information was not used during sampling, but during the process of generating the integrated weights.

**LRGFAM**

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Large-family-household at first contact”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value label</td>
<td>LRGFAM</td>
</tr>
<tr>
<td></td>
<td>(1) Yes</td>
</tr>
<tr>
<td></td>
<td>(2) No</td>
</tr>
<tr>
<td>Variable format</td>
<td>1-digit integer</td>
</tr>
</tbody>
</table>

**Comment**

LRGFAM is a time-fixed indicator providing information whether a household was defined as “large-family-household” the first time it was interviewed. This information belongs to the original household and is thus associated with the variable HHNR. It is carried over to all split households from the original household. In the Screening-Samples 2010 and 2011, this corresponds to the sampling-design-definition (at least three children younger than 18 years in household) at the point of selection of the original household into the sample. For the Cohort-Sample, this information was not used during sampling, but during the process of generating integrated weights.

**LOWINC**

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Low-income-household at first contact”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value label</td>
<td>LOWINC</td>
</tr>
<tr>
<td></td>
<td>(1) Yes</td>
</tr>
<tr>
<td></td>
<td>(2) No</td>
</tr>
<tr>
<td></td>
<td>(3) Unknown</td>
</tr>
<tr>
<td>Variable format</td>
<td>1-digit integer</td>
</tr>
</tbody>
</table>
Comment

LOWINC is a time-fixed indicator providing information whether a household was defined as “low-income-household” the first time it was interviewed. This information belongs to the original household and is thus associated with the variable HHNR. It is carried over to all split households from the original household. In the Screening-Sample 2010, this corresponds to the sampling-design-definition at the point of selection of the original household into the sample. For the cohort sample, this information was not used during sampling, but during the process of generating integrated weights. LOWINC is set to “-2 Does not apply” for all households in the Screening-Sample 2011, as this information was not used for sampling or weighting purposes.

The value “(3) Unknown” indicates that no low-income-decision could be made due to missing values on the net household income in 2010 (F10H045A). Given a valid value on net household income, a low-income-household is defined as:

a) a household with one adult and children and a maximum net income of 1500€,

b) a household with more than one adult and one child and a maximum net income of 2000€ or

c) a household with more than one adult and more than one child and a maximum net income of 2500€ a month.

If possible, the information of F10H045B is also used.
Documentation $pgen$

Person Related Status Variables and Generated Variables

Stefan Damerow, Mathis Fräßdorf (geb. Schröder) and Juliana Werneburg

This documentation is based on the comparable SOEP documentation on PGEN and has benefited from previous work of Silke Anger, Joachim Frick, Markus Grabka, Jan Goebel, Peter Krause, Henning Lohmann, Olaf Groh-Samberg, and Pia Schober. For readability reasons we do not specifically cite and specify text that has been used directly from the SOEP document.

General information

The $pgen$ datasets provide individual level information for each survey wave. Similar to the data included in $shgen$ for the household level, the $pgen$ files give the researcher an easier access to individual information over the waves. Variable names and formats are consistent over all years included, such that a combination of these files facilitates a comparison of different years. In addition, some variables are completely generated or imputed, and hence provide information which is not included in the regular questionnaire. The following documentation lists all variables in $pgen$ and provides information on how they were generated. For all variables, the information provided looks as follows:

Variable label    Provides the label of the variable as it is given in the dataset. Variables are given in CAPTIAL letters, even though they might appear in small letters in the dataset. This is simply for readability.
**Spgen: Person related status variables and generated variables**

**Value labels**: $LBLNME$

In case $VARNME$ is categorical, $LBLNME$ specifies the labels for each category, and the value labels are listed here. Note that the standard missing value labels (-1: No answer; -2: Does not apply; -3: Not valid) are not listed, but apply to all variables in this dataset.

**Variable format**: Specifies the format for each variable, e.g. “1-digit integer” or “string”.

**$SS$ - Survey Years**: Specifies the years for which the variable is provided. This is provided for 2000+, such that “10” refers to 2010, etc.

**Comment**: Provides more detailed information on the generating process, also on the population the variable is specified for, if necessary. Here, variables used, changes between waves, or any other anomalies are mentioned and their relevance explained.

Some variables will be written forward from previous waves, as they are only collected once and do not necessarily change every year. These variables include some employment variables, the education variables, marital status, and partner variables.

If you have questions regarding **Spgen** data for the FiD-distribution, unless noted otherwise, please contact Mathis Schröder at +49 (0)30 / 89789 - 222.
List of variables:

EMPLST$$ ............................................................................................................................... 43
LFSS$$ ....................................................................................................................................... 43
JOBCH$$ .................................................................................................................................. 44
EXPFT$$ .................................................................................................................................. 45
EXPPT$$ .................................................................................................................................. 46
EXPUE$$ .................................................................................................................................. 46
TENURE$$ .................................................................................................................................. 47
JOBTRN$$ .................................................................................................................................. 48
REQUTR$$ .................................................................................................................................. 48
COSIZE$$ .................................................................................................................................. 49
CRSIZE$$ .................................................................................................................................. 49
CIVIL$$ ...................................................................................................................................... 50
OCCPOS$$ .................................................................................................................................. 50
CLASS$$ ...................................................................................................................................... 51
IS88$$ ...................................................................................................................................... 53
NACE$$ ...................................................................................................................................... 54
ISEI$$ ...................................................................................................................................... 56
EGP$$ ...................................................................................................................................... 56
SIOP$$ ...................................................................................................................................... 58
MP$$ ...................................................................................................................................... 58
AUTON$$ .................................................................................................................................. 59
AGRHR$$ .................................................................................................................................. 60
ACTHR$$ .................................................................................................................................. 60
OVRHR$$ .................................................................................................................................. 60
PART$$ ...................................................................................................................................... 61
PARTNO$$ .................................................................................................................................. 62
COUPST$$ .................................................................................................................................. 62
COUPID$$ .................................................................................................................................. 63
MARRST$$ .................................................................................................................................. 63
Educational Variables in FiD ................................................................................................... 66
SCEDU$$ ................................................................................................................................... 67
Spgen: Person related status variables and generated variables

- SCEDUE$$ ................................................................. 67
- SCEDUASS$$ .............................................................................................................................. 67
- VCDEG$$ ............................................................................................................................. 68
- VCDEGE$$ .............................................................................................................................. 68
- VCDEGA$$ ............................................................................................................................. 68
- VCNONE$$ ................................................................. 68
- COLLEGS$$ ................................................................. 69
- TIMEDU$$ ................................................................. 69
- ISCED$$ ............................................................................................................................. 70
- CASMIN$$ ............................................................................................................................. 72
- LABGRO$$ ............................................................................................................................. 73
- IMPGRO$$ ............................................................................................................................. 74
- LABNET$$ ............................................................................................................................. 74
- IMPNET$$ ............................................................................................................................. 74
- FIELD$$ ............................................................................................................................. 74
- DEGREE$$ ............................................................................................................................. 76
- TRAINA$$ ............................................................................................................................. 78
- TRAINB$$ ............................................................................................................................. 80
- TRAIND$$ ............................................................................................................................. 80
- FDT_F$$ ............................................................................................................................. 81
**EMPLST$$**

**Variable label**  “Employment status”

**Value label**

- (1) full-time employment
- (2) regular part-time employment
- (3) vocational training
- (4) marginal, irregular part-time employment
- (5) not employed
- (6) sheltered workshop

**Variable format**  1-digit integer

**$$ - Survey Years**  $$=10, 11, 12, 13

**Comment**

This variable is generated from the annual question on current labor market participation, which has a central filter function in the questionnaire to separate employed from non-employed respondents for further questions. It is designed to provide consistent longitudinal data on employment status across all waves.

The category “not employed” comprises non-working individuals, those in military/community service, those on maternity leave, and employed persons in a phased retirement scheme (*Altersteilzeit*), whose current actual working hours are zero. The additional category “sheltered workshop” is included for disabled persons in sheltered employment.

EMPLST$$ is supplemented by the variable LF$$, which differentiates among persons who are not employed.

**LF$$**

**Variable label**  “Labor force status”

**Value label**

- (1) non-working without further information
- (2) non-working, and older than 65
- (3) non-working, in training program
- (4) non-working, on maternity leave
- (5) non-working, in military/community service
- (6) non-working, and registered unemployed
- (8) non-working, but sometimes second job
- (9) non-working, but working past 7 days
- (10) non-working, but regular second job
- (11) working
- (12) working, but non-working past 7 days

**Variable format**  2-digit integer

**$$ - Survey Years**  $$=10, 11, 12, 13

**Comment**

This variable is based on the annual question on current labor market participation, combined with additional information on activities of non-working individuals. The number of values assigned has been based on a large number of highly differentiated answer categories.

LF$$ provides a differentiation between “working” (Code 11-12) and “non-working” (Code 1-10). Non-employment is split further in order
to make it possible to efficiently apply different labor market concepts in studying the data. To construct this variable, the variables on employment status, age, maternity leave, second jobs, registration at the employment office, participation in paid work during the past 7 days and training status are used.

For respondents who have multiple status codes and different values for this variable, the following hierarchy was used to determine which of the values would play the determining role (increasing dominance)

- 11 - working
- 1 - non-working without further information
- 2 - non-working, and older than 65
- 3 - non-working, and currently in a training program
- 6 - non-working, and registered unemployed
- 4 - non-working, on maternity leave
- 5 - non-working, in military/community service
- 9 - non-working, but working past 7 days
- 10 - non-working, but regular second job
- 8 - non-working, but occasional second job
- 12 – working, but non-working past 7 days

LFSS$ supplants the variable EMPLST$$, which differentiates among persons who are employed.

**JOBCH$$**

**Variable label**

“Job change”

**Value label**

JOBCH$

(1) not employed
(2) employed, no change
(3) employed, no info if change
(4) employed, with change
(5) first time employed

**Variable format**

1-digit integer

**$\$ - Survey Years**

$$=10, 11, 12, 13$

**Comment**

This variable indicates a change of job since the last interview for respondents with a follow-up interview, whereas for first-time respondents, the information refers to a change of job since the beginning of the previous year. JOCH$$ is generated based on the central filter variable, which indicates whether a respondent has changed jobs since the previous year.

The variable is used to integrate the information for first-time respondents and follow-up respondents. The following hence only applies from the second wave (2011) onwards, where both types of respondents appear for the first time.

The variable also identifies respondents who have entered employment for the first time. The variable will provide consistent longitudinal information on job changes as well. The JOBCH$$ variable is
generated by correcting the original job change information in various ways:
1. We check whether the job changes stated by a respondent in two consecutives interviews refer to one and the same job change. The date of the job change and the interview month are used to correct double entries.
2. If the respondent indicates a job change with a date before the previous interview but did not state a job change in the previous interview, this is coded as a job change in the current interview.
3. If a respondent indicates no job change and was not employed in the previous interview, this is coded as "no job change" because there could have been short-term employment spells between the previous year's and this year's interview.
4. Respondents can be "first-time employed" only once. If a respondent states being "first-time employed" for a second time, this is coded as "employed, with change".

**EXPFT$$**

**Variable label**
"Working experience full-time employment"

**Variable format**
3-digit integer

**$$ - Survey Years**
$$=10, 11, 12, 13

**Comment**
Full-time working experience

This variable reflects the total length of full-time employment in the respondent’s career up to the point of the interview. The variable is created by combining monthly information from the calendar dataset `artkalen` (which provides monthly information on activity status since an individual entered FiD) and annual information from the biographical dataset `pbiospe` (which provides information on activity status over the individual’s life course). EXPFT$$ gives the length of time in months – different from the SOEP, where it is provided in years with months in decimal form.

If there is no monthly calendar data available in a given year of a respondent’s career, the annual data from `pbiospe` is used for that year. If the year in which a spell started and ended is the same, and if there is no monthly data, a spell of 6 months is assumed. Persons without annual data (not contained in `pbiospe`) are only assigned a non-missing value for this variable if they joined FiD by the age of 18 and if there is calendar data for them in `artkalen`.

Persons whose life course has been observed completely but with no spell of full-time employment are assigned the code (0). The code (“-1 No answer”) is assigned to all persons whose life course has not been observed completely. Persons with inconsistent information receive a (“-3 Answer implausible”).

Note that differently to the SOEP, respondents in FiD receive a value of “-2 Does not apply” if their biography data has not been collected yet.
For example, this is the case for individuals, who only participate in 2010. For a person to have valid values in EXPFT$$, the respective biography part has to be answered ($\text{LELTYP}=2$ or $3$).

Please also see EXPPT$$ and EXPUE$$.

### EXPPT$$

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Working experience part-time employment”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable format</td>
<td>3-digit integer</td>
</tr>
<tr>
<td>$\text{Survey Years}$</td>
<td>$10, 11, 12, 13$</td>
</tr>
<tr>
<td>Comment</td>
<td>Part-time working experience</td>
</tr>
</tbody>
</table>

This variable reflects the total length of part-time employment in the respondent’s career up to the point of the interview. The variable is created by combining monthly information from the calendar dataset artkalen (which provides monthly information on activity status since an individual entered FiD) and annual information from the biographical dataset pbiospe (which provides information on activity status over the individual’s life course). EXPPT$$ gives the length of time in months – different from the SOEP, where it is provided in years with months in decimal form.

If there is no monthly calendar data available in a given year of a respondent’s career, the annual data from pbiospe is used for that year. If the year in which a spell started and ended is the same, and if there is no monthly data, a spell of 6 months is assumed. Persons without annual data (not contained in pbiospe) are only assigned a non-missing value for this variable if they joined FiD by the age of 18 and if there is calendar data on them in artkalen.

Persons whose life course has been observed completely but with no spell of part-time employment are assigned the code (0). The code (“-1 No answer”) is assigned to all persons whose life course has not been observed completely. Persons with inconsistent information receive a (“-3 Answer implausible”).

Note that differently to the SOEP, respondents in FiD receive a value of “-2 Does not apply” if their biography data has not been collected yet. For example, this is the case for individuals, who only participate in 2010. For a person to have valid values in EXPPT$$, the respective biography part has to be answered ($\text{LELTYP}=2$ or $3$).

Please also see EXPFT$$ and EXPUE$$.

### EXPUE$$

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Unemployment experience”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable format</td>
<td>3-digit integer</td>
</tr>
</tbody>
</table>

SOEP Survey Paper 332
Comment

Unemployment experience
This variable reflects the total length of unemployment in the respondent’s career up to the point of the interview. The variable is created by combining monthly information from the calendar dataset artkalen (which provides monthly information on activity status since an individual entered FiD) and annual information from the biographical dataset pbiospe (which provides information on activity status over the individual’s life course). EXPFT$$ gives the length of time in months – different from the SOEP, where it is provided in years with months in decimal form.

If there is no monthly calendar data available in a given year of a respondent’s career, the annual data from pbiospe is used for that year. If the year in which a spell started and ended is the same, and if there is no monthly data, a spell of 6 months is assumed. Persons without annual data (not contained in pbiospe) are only assigned a non-missing value for this variable if they joined FiD by the age of 18 and if there is calendar data on them in artkalen.

Persons whose life course has been observed completely but with no spell of unemployment are assigned the code (0). The code (“-1 No answer”) is assigned to all persons whose life course has not been observed completely. Persons with inconsistent information receive a (“-3 Answer implausible”).

Note that differently to the SOEP, respondents in FiD receive a value of “-2 Does not apply” if their biography data has not been collected yet. For example, this is the case for individuals, who only participate in 2010. For a person to have valid values in EXPUE$$, the respective biography part has to be answered ($LELTYP=2 or 3).

Please also see EXPFT$$ and EXPPT$$.

**TENURE$$**
Variable label “Tenure in months”
Variable format 3-digit integer
$$_{SS}$$- Survey Years $$=10, 11, 12, 13$
Comment

The variable TENURE$$ is designed to offer data on the length of time with the firm at the point in time of the interview for all employed persons. This variable is generated from the respondent's start date with the current employer and the start date of the current position if there was a job change.

The variable provides consistent longitudinal information on the length of time with the same employer. Data that show longitudinal inconsistencies are corrected. In case of no job change, the information on the start date with the current employer given in the earliest
interview available is treated as dominant and carried forward to the subsequent years. In case of a job change, the information on the start of the current position is used and carried forward to the subsequent years. In the case that a respondent starts working again after a period of non-employment, he/she is assumed to have returned to the former employer if the start date with the current employer was before the previous interview date. In this case, the start date with the current employer given in the previous interview is treated as dominant. Otherwise, the present information on the start date with the current employer is used and carried forward to the subsequent years. For respondents who are assumed to have returned to their former employer, the full length of time with the firm is calculated. There is no deduction for the time during which the respondent was not employed. Both monthly and annual information is used in the variable, which is provided – differently to the SOEP – in months.

**JOBTRN$$**

Variable label: “Working in occupation trained for”

Value label:

- (1) yes
- (2) no
- (3) currently in training
- (4) has no job training

Variable format: 1-digit integer

$$ - Survey Years $$= 10, 11, 12, 13

Comment: This variable is designed to offer annual data on all employed persons, indicating whether they are working in the occupation they were trained for. This question is asked – similar to the SOEP – on a biannual basis for those who did not change their job. Hence, this information is written forward in following waves.

**REQUTR$$**

Variable label: “Required job training”

Value label:

- (1) no training
- (2) brief on-the-job training
- (3) extensive on-the-job training
- (4) attended courses
- (5) completed vocational training
- (8) Fachhochschule degree
- (9) University degree

Variable format: 1-digit integer

$$ - Survey Years $$= 10, 11, 12, 13

Comment: This variable is designed to provide annual data on required job training for all employed persons. The variable is generated using questions on required formal education and required on-the-job-training which are categorized into up to seven independent variables with 0/1 coding. Out
of these, the highest available level of required training is used for the generation of the status variable. This question is asked – similar to the SOEP – on a biannual basis for those who did not change their job. Hence, this information is written forward in following waves. The missing value (“-2 Does not apply”) was assigned to all non-employed persons and also includes persons in occupational training, in occupational retraining programs, and those doing an internship at the time of the survey.

COSIZE$$

Variable label
“Size of company”
Value label
COSIZE$$
(1) less than 5
(2) 5 to 10
(3) 11 to 20
(4) [up to 1990 less than 20]
(5) [1991-2004 5 to 20]
(6) 20 to 100
(7) 100 to 200
(8) [up to 1998 20 to 200]
(9) 200 to 2000
(10) 2000 or more
(11) Self-employed without other employees

Variable format
2-digit integer
$\$ - Survey Years
$\$=10, 11, 12, 13

Comment
This variable is designed to offer annual data on company size for all employed persons. The codes 4, 5, and 8 were given in the data to make this variable comparable to the SOEP logic, although they will not be set in FiD. (These codes were necessary due to the differentiation of items for small and medium-sized companies over the years.) This question is asked – similar to the SOEP – on a biannual basis for those who did not change their job. Hence, this information is written forward in following waves. Please also see COSIZE$$ for a broader categorization of the firm size, again comparable to the SOEP logic. (Note that COSIZE$$ was namedFSIZE$$ in FiD v1.2. To avoid confusions with the identically named variable in Shgen, we renamed the variable in Spgen. Nothing else has changed.)

CRSIZE$$

Variable label
“Core size category of the company”
Value label
CRSIZE
(1) fewer than 20
(2) 20 to 200
(3) 200 to 2000
(4) 2000 or more
(5) Self-employed without other employees
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13
Comment This variable is designed to provide annual data on the core size category of the company for all employed persons compared to the SOEP logic.
This question is asked – similar to the SOEP – on a biannual basis for those who did not change their job. Hence, this information is written forward in following waves.
(Note that CRSIZE$$ was named CFSIZE$$ in FiD v1.2.)

CIVILS$$
Variable label “Civil Service”
Value label CIVILS$$
   (1) yes
   (2) no
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13
Comment This status variable is designed to provide annual data on employment in the civil service for all employed persons. This question is asked – similar to the SOEP – on a biannual basis for those who did not change their job. Hence, this information is written forward in following waves.

OCCPO$$
Variable label “Occupational Position”
Value label OCCPO$$ (see below)
Variable format 3-digit integer
$$ - Survey Years $$=10, 11, 12, 13
( -1) No Answer
( 10) Not Employed
( 11) In Education
( 12) Unemployed, Not Employer
( 13) Pensioner
( 15) Military, Community Service
(110) Apprentice
(120) Apprentice, Trainee Industry Technology
(130) Apprentice, Trainee Trade and Commerce
(140) Trainee, Intern
(150) Research assistant
(210) Untrained Worker
(220) Semi-Trained Worker
(230) Trained Worker
(240) Foreman, Team Leader
(250) Foreman
   (310) Agricultural Worker
   (320) Agricultural Specialist
   (330) Agricultural Foreman
   (340) Agricultural Manager
   (410) Self-Employed Farmer
   (411) Self-Employed Farmer, No Employees
   (412) Self-Employed Farmer LE 9 Employees
   (413) Self-Employed Farmer GT 9 Employees
   (420) Free-Lance Professional
   (421) Free-Lance Professional, No Employees
   (422) Free-Lance Professional, LE 9 Employees
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>423</td>
<td>Free-Lance Professional, GT 9 Employees</td>
<td>521</td>
<td>Untrained Employee with Simple Tasks</td>
</tr>
<tr>
<td>430</td>
<td>Other Self-Employed No Or LE 9 Employees</td>
<td>521</td>
<td>Untrained W-Collar Worker with Simple Tasks</td>
</tr>
<tr>
<td>431</td>
<td>Other Self-Employed No Employees</td>
<td>522</td>
<td>Trained Employee with Simple Tasks</td>
</tr>
<tr>
<td>432</td>
<td>Other Self-Employed LE 9 Employees</td>
<td>522</td>
<td>Trained W-Collar Worker with Simple Tasks</td>
</tr>
<tr>
<td>433</td>
<td>Other Self-Employed GT 9 Employees</td>
<td>530</td>
<td>Qualified Professional</td>
</tr>
<tr>
<td>440</td>
<td>Help In Family Business</td>
<td>540</td>
<td>H. Qualified Professional</td>
</tr>
<tr>
<td>510</td>
<td>Foreman</td>
<td>550</td>
<td>Managerial</td>
</tr>
<tr>
<td>520</td>
<td>Employee with Simple Tasks</td>
<td>610</td>
<td>Low-Level Civil Service</td>
</tr>
<tr>
<td>520</td>
<td>W-Collar Worker with Simple Tasks</td>
<td>620</td>
<td>Middle-Level Civil Service</td>
</tr>
<tr>
<td></td>
<td></td>
<td>630</td>
<td>High-Level Civil Service</td>
</tr>
<tr>
<td></td>
<td></td>
<td>640</td>
<td>Executive Civil Service</td>
</tr>
</tbody>
</table>

Comment: The variable represents a compilation of all relevant information on current occupational position. It is generated by combining information on “occupational group”, “unemployed (yes/no)”, “military/community service”, “in education (yes/no)” and “pensioner”. A hierarchical scheme is used to determine which data is given precedence when a variety of divergent information exists (increasing dominance):

10 – not employed
13 – pensioner
11 – currently in education
15 – military / community service
12 – registered unemployed
110-150 – apprentice
410-440 – self-employed
210-250 – manual laborer
510-550 – employee
610-640 – civil service

The categories (150) and (310) to (340) were only assigned in the SOEP to respondents in East Germany in 1990. In OCCPOS$$, non-working persons are only assigned to the category (13) "pensioner" if they are recipients of retirement pension or if they are recipients of widow’s pension AND are older than 60 years. Moreover, if there is missing information on pension receipt, additional information from artkalen (retrospective information from the activity calendar for the previous year) will be used in the subsequent waves in the generation process to determine whether a person was in retirement or early retirement (Vorruhestand) at the time of the interview.

CLASS$$
Variable label “StaBuA 1992 Job Classification”
Value label CLASS$$ (see below)
Variable format 4-digit integer
This variable is designed to provide annual data on job classification for all employed persons according to the classification of the German Federal Statistical Office (StaBuA). Respondents answer the question on their current occupational title in their own words, and this response is entered into a blank in the questionnaire. Due to data protection regulations, this information cannot be provided to data users and was therefore completely recoded by Infratest Sozialforschung. This recoding has been documented in Hartmann/Schütz 2002.

This question is asked – similar to the SOEP – on a biannual basis for those who did not change their job. Hence, this information is written forward in following waves.

The occupational classification of the German Federal Statistical Office differentiates among six main occupational types (see next page).

I KLAS-Codes 0100-0629  **Berufe in der Land-, Tier-, Forstwirtschaft und im Gartenbau**
II KLAS-Codes 0700-0809  **Bergleute, Mineralgewinner**
III KLAS-Codes 0900-1459  **Fertigungsberufe**
   IIIa KLAS-Codes 1000-1129  Berufe in der Steinbearbeitung und Baustoffherstellung
   IIIb KLAS-Codes 1200-1359  Keramik-, Glasberufe
   IIIc KLAS-Codes 1400-1539  Chemie-, Kunststoffberufe
   IIId KLAS-Codes 1600-1799  Berufe in der Papierherstellung, -verarbeitung und im Druck
   IIIe KLAS-Codes 1800-1859  Berufe in der Holzverarbeitung, Holz- und Flechtwarenherstellung
   IIIf KLAS-Codes 1900-2459  Berufe in der Metallerzeugung und –bearbeitung
   IIIg KLAS-Codes 2500-3099  Metall-, Maschinenbau- und verwandte Berufe
   IIIh KLAS-Codes 3100-3189  Elektroberufe
   IIIi KLAS-Codes 3200-3239  MontiererInnen und Metallberufe, a.n.g.
   IIIk KLAS-Codes 3300-3619  Textil- und Bekleidungsberufe
   IIIl KLAS-Codes 3700-3789  Berufe in der Lederherstellung, Leder- und Fellverarbeitung
   IIIm KLAS-Codes 3900-4359  Ernährungsberufe
   IIIn KLAS-Codes 4400-4729  Hoch-, Tiefbauberufe
   IIIo KLAS-Codes 4800-4929  Ausbauberufe, PolsterInnen
   IIIp KLAS-Codes 5000-5099  Berufe in der Holz- und Kunststoffverarbeitung
   IIIq KLAS-Codes 5100-5149  MalerInnen, LackiererInnen und verwandte Berufe
   IIIr KLAS-Codes 5200-5239  WarenprüferInnen, VersandfertigmacherInnen
   IIIs KLAS-Codes 5300-5319  HilfsarbeiterInnen ohne nähere Tätigkeitsangabe
   IIIt KLAS-Codes 5400-5509  MaschinistInnen und zugehörige Berufe
   IIIu KLAS-Codes 5600-5699  ZahnarztInnen und zugehörige Berufe
   IIIv KLAS-Codes 5700-5799  KellnerInnen und Kellner
   IIIw KLAS-Codes 5800-5899  Küche und Küchenhilfe
   IIIx KLAS-Codes 5900-5999  VerkäuferInnen, Verkäufer, Verkäuferinnen und -verkäufer
IIII KLAS-Codes 6000-6129  IngenieurInnen, ChemikerInnen, PhysikerInnen, MathematikerInnen
   IVa KLAS-Codes 6200-6529  TechnikerInnen, Technische Sonderfachkräfte
V KLAS-Codes 6600-6899  Warenkaufleute
   Va KLAS-Codes 6600-6899  Warenkaufleute
   Vb KLAS-Codes 6900-7069  Dienstleistungskaufleute und zugehörige Berufe
   Vc KLAS-Codes 7100-7449  Verkehrsberufe
   Vd KLAS-Codes 7500-7899  Organisations-, Verwaltungs-, Büroberufe
   Ve KLAS-Codes 7900-8149  Ordnungs- und Sicherheitsberufe
   Vf KLAS-Codes 8200-8399  Schriftwerkschaffende, -ordinende und künstlerische Berufe
   Vg KLAS-Codes 8400-8599  Gesundheitsdienstberufe
   Vh KLAS-Codes 8600-8949  Sozial- und Erziehungsberufe, anderweitig nicht genannte geistes- und sozialwissenschaftliche Berufe
   Vj KLAS-Codes 7900-7979  Sonstige Dienstleistungsberufe
VI KLAS-Codes 9000-9979  **Sonstige Arbeitskräfte**
Because of gaps in the answers provided by respondents, the following “new” codes were created:

- 9711 Mithelfende Familienangehörige außerhalb der Landwirtschaft, anderweitig nicht genannt
- 9811 Auszubildende mit (noch) nicht feststehendem Ausbildungsberuf
- 9821 Praktikanten/Praktikantinnen, Volontäre/ Volontärinnen mit (noch) nicht feststehendem Beruf
- 9911 Facharbeiter/innen, ohne nähere Tätigkeitsangabe
- 9921 Heimarbeiter/innen, ohne nähere Tätigkeitsangabe
- 9931 Vorarbeiter/innen, Gruppenleiter/innen, ohne nähere Tätigkeitsangabe
- 9971 Sonstige Arbeitskräfte, ohne nähere Tätigkeitsangabe


IS88$$

Variable label “4-digit ISCO-88 Occupation Code”
Value label ISCO$$ (see below)
Variable format 4-digit integer
$$ - Survey Years $$=10, 11, 12, 13

(1000) Legislators, senior officials, and managers
(1100) Legislators and senior officials
(1200) Corporate managers
(1300) Managers of small enterprises
(2000) Professionals
(2100) Physical, mathematical, and engineering science professionals
(2200) Life science and health professionals
(2300) Teaching professionals
(2400) Other professionals
(3000) Technicians and associate professionals
(3100) Physical and engineering science associate professionals
(3200) Life science and health associate professionals
(3300) Teaching associate professionals
(3400) Other associate professionals
(4000) Clerks
(4100) Office clerks
(4200) Customer services clerks
(5000) Service Workers and shop and market sales workers
(5100) Personal and protective services workers
(5200) Models, salespersons, and demonstrators
(6000) Skilled agricultural and fishery Workers
(6100) Skilled agricultural and fishery workers
(7000) Craft and related trades workers
(7100) Extraction and building trades workers
(7200) Metal, machinery, and related trades workers
(7300) Precision, handicraft, craft printing and related trades workers
(7400) Other craft and related trades workers
(8000) Plant and machine operators and assemblers
(8100) Stationary plant and related operators
(8200) Machine operators and assemblers
(8300) Drivers and mobile plant operators
(9000) Elementary occupations
(9100) Sales and services elementary occupations
(9200) Agricultural, fishery, and related laborers
(9300) Laborers in mining, construction, manufacturing, and transport

Comment This variable is designed to provide annual data on occupational activity for all employed persons according to the International Standard Classification of Occupations ISCO-88. Respondents answer the question on their current occupational title in their own words, and this response is entered into a blank in the questionnaire. Due to data protection regulations, this information cannot be provided to data users and was therefore completely recoded by Infratest Sozialforschung. This recoding has been documented in Hartmann/Schütz 2002.
**SpGen: Person related status variables and generated variables**

ISCO-88 is a strictly four-digit classification, and this variable is therefore coded in four-digit form. In contrast to the previous version of the classification system, ISCO-68, ISCO-88 does not use blanks if there is not adequate information for specific coding, but uses zeros instead. Thus 4000 stands for an unspecified office job; 2300 stands for teachers and 2000 stands for scientists, both without closer specification. There is no conversion key since the two classifications differ significantly. The SOEP data distribution 1984-2001 replaced ALL earlier data distributions with the ISCO-88-coding.

This question is asked – similar to the SOEP – on a biannual basis for those who did not change their job. Hence, this information is written forward in following waves.

**Detailed description:**

**NACE$$**

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Two-digit NACE Industry – Sector”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value label</td>
<td>NACE$$ (1-100) (see below)</td>
</tr>
<tr>
<td>Variable format</td>
<td>3-digit integer</td>
</tr>
<tr>
<td>$$ - Survey Years</td>
<td>$$=10, 11, 12, 13</td>
</tr>
</tbody>
</table>

( 1) Agriculture, Hunting, Related Service Activities  
( 2) Forestry, Logging, Related Service activities  
( 5) Fishing, Operation Of Fish Hatcheries And Fish Farms  
(10) Mining Of Coal And Lignite; Extraction Of Peat  
(11) Extraction Of Crude Petroleum And Natural Gas  
(12) Mining Of Uranium And Thorium Ores  
(13) Mining Of Metal Ores  
(14) Other Mining And Quarrying  
(15) Manuf. Food Products And Beverages  
(16) Manuf. Tobacco Products  
(17) Manuf. Textiles  
(18) Manuf. Wearing Apparel; Dressing And Dyeing Of Fur  
(19) Tanning, Dressing Of Leather; Manuf. luggage, Footwear  
(20) Manuf. Wood Products, Except Furniture  
(21) Manuf. Pulp, Paper And Paper Products  
(22) Publishing, Printing And Reproduction Of Recorded Media  
(23) Manuf. Coke, Refined Petroleum Prod, Nuclear Fuel  
(24) Manuf. Chemicals And Chemical Products  
(25) Manuf. Rubber And Plastic Products  
(26) Manuf. Other Non-metallic Mineral Products  
(27) Manuf. Basic Metals  
(28) Manuf. Fabricated Metal Prod., Ex. Machinery And Equip  
(29) Manuf. Machinery And Equipment NEC  
(30) Manuf. Office Machinery And Computers  
(31) Manuf. Electrical Machinery And Apparatus NEC  
(32) Manuf. Radio, Television And Communication Equipment  
(33) Manuf. Medical, Precision And Optical Instruments  
(34) Manuf. Motor Vehicles, Trailers And Semi-trailers  
(35) Manuf. Other Transport Equipment  
(36) Manuf. Furniture; Manufacturing NEC  
(37) Recycling  
(40) Electricity, Gas, Steam And Hot Water Supply  
(41) Collection, Purification And Distribution Of Water  
(45) Construction  
(50) Sale, Maint., Repair Motor Vehicles; Retail Car Gas  
(51) Wholesale Trade, Commission Trade, Ex. Motor Vehicles  
(52) Retail, Ex. Motor vehicles, Motorcycles; Repair  
(55) Hotels And Restaurants  
(60) Land Transport; Transport Via Pipelines  
(61) Water Transport  
(62) Air Transport  
(63) Supporting, Aux. Transport Activities; Travel agencies  
(64) Post And Telecommunications  
(65) Financial Intermediation, Ex. Insurance, Pension Funding  
(66) Insurance And Pension Funding, Ex. Compulsory SocSec
This variable is designed to provide annual data on the industry of economic activity for all employed persons according to the Statistical Classification of Economic Activities in the European Community (Nomenclature des statistiques des activités économiques de la Communauté européenne - NACE). Respondents answer the question in their own words regarding the industry in which they are currently working, and this response is entered into a blank in the questionnaire. In order to facilitate international comparability, the European industry standard classification system is used by Infratest Sozialforschung to recode this information. This recoding has been documented in Hartmann/Schütz 2002.

The codes in NACE Rev.1 also correspond to ISIC Rev.3 (International Standard Classification of All Economic Activities). Please note that special codes 96-98 as well as 100 were assigned by Infratest Sozialforschung whenever respondents did not provide a more detailed answer.

This question is asked – similar to the SOEP – on a biannual basis for those who did not change their job. Hence, this information is written forward in following waves.

### ISEI$$

**Variable label**
“ISEI-status 88 by Ganzeboom (based on ISCO88)"

**Variable format**
2-digit integer

**$$ - Survey Years**
$$=10, 11, 12, 13

**Comment**
This variable reflects the Standard International Socio-Economic Index of Occupational Status for all employed persons. The ISEI Index was developed in 1992 by Ganzeboom, De Graaf, Treiman, and De Leew based on information about income, education, and occupation. Technically, ISEI was created by scaling the ISCO88 classification. The values for the variable range between 16 and 90. In contrast to the prestige scores of Ganzeboom and Treiman (1996) and Wegener (1988), ISEI is a measure of socio-economic status.

This question is asked – similar to the SOEP – on a biannual basis for those who did not change their job. Hence, this information is written forward in following waves.

Please also see occupational prestige scores (SIOP$$, MP$$) and occupational class (EGP$$).

**Detailed description:**

### EGP$$

**Variable label**
“Erikson and Goldthorpe Class Category”

**Value label**
EGP$$

(1 ) high service
(2 ) low service
(3 ) routine non-manual
(4 ) routine service-sales
(5 ) self-employed with employees
(6 ) self-employed without employees
(8 ) skilled manual
(9 ) semi-unskilled manual
(10) farm labor
(11) self-employed farmer
(15) not working – unemployed
(18) not working – pensioner

**Variable format**
2-digit integer

**$$ - Survey Years**
$$=10, 11, 12, 13

**Comment**
This variable gives the occupational class for all employed persons. EGP$$ is derived from the Standard International Socio-Economic Index of Occupational Status (ISEI). Technically, the variable was created by scaling the ISCO-88 classification. In addition, it is based on
information about income, education and occupation. The EGP Index was documented by Ganzeboom/Treiman in 1996 and revised in 2003.

The values for the variable range between 1 and 11; additional categories are (15) not working – registered unemployed and (18) not working – pensioner.

Non-working persons are only assigned to the category “not working – pensioner” if they are recipients of retirement pension or if they are recipients of widow’s pension AND are older than 60 years. Moreover, if there is missing information on pension receipt, additional information from artkalen (retrospective information from the activity calendar for the previous year) will be used in future waves in the generation process to determine if a person was in retirement or early retirement (Vorruhestand) at the time of the interview. Hence, the category “not working – pensioner” in the most recent wave will be updated with retrospective information of the following wave. All other non-working persons are assigned to category (-2) “does not apply” as long as they are not registered as unemployed (category 15).

Annual information on the occupational position is used to generate the EGP-categories for the self-employed. In case no information on the number of employees is available, the EGPS$$-categories (5) and (6) contain information on the firm size for self-employed persons.

Based on the new classification developed by Ganzeboom/Treiman (2003), several ISCO values were recoded in EGPS$$ as follows:
- ISCO 2470 becomes EGP=1.
- ISCO 2500 becomes EGP=2.
- ISCO 4300, 4400, 4500 become EGP=4.
- ISCO 7900 becomes EGP=7.
- ISCO 9910-9990 become EGP=9.

This question is asked – similar to the SOEP – on a biannual basis for those who did not change their job. Hence, this information is written forward in following waves.

Please also see occupational status (ISEI$$) and occupational prestige scores (SIOP$$$$, MP$$$$).


**SIOPS$$**

Variable label: “Treimans Standard Int. Occupation Prestige Score”

Variable format: 2-digit integer

Survey Years: $$= 10, 11, 12, 13

Comment: This variable gives the occupational prestige score index for all employed persons. SIOPS$$ is based on ISCO-88 and was developed by Donald Treiman et al. The scale ranges from 6 to 78.

This question is asked – similar to the SOEP – on a biannual basis for those who did not change their job. Hence, this information is written forward in following waves.

Please also see occupational prestige scores (MPS$$), occupational status (ISEI$$), and occupational class (EGP$$).


---

**MPS$$**

Variable label: “Magnitude-Prestige Scala – Wegener”

Variable format: 5-digit real

Survey Years: $$= 10, 11, 12, 13

Comment: This variable gives the occupational prestige score developed by Wegener (1988) for all employed persons. Like the SIOPS$$ prestige score, Wegener’s prestige scala measures a person’s occupational prestige and was developed especially for use in the Federal Republic of Germany. MPS$$ is assigned based on the German Federal Statistical Office’s occupational classification of 1992 (KLASS$$). The procedure has been documented in Frietsch and Wirth (2001).

This question is asked – similar to the SOEP – on a biannual basis for those who did not change their job. Hence, this information is written forward in following waves.

Please also see occupational prestige scores (SIOPS$$), occupational status (ISEI$$), and occupational class (EGP$$).


AUTONO$$

Variable label  “Autonomy in occupational activity”

Value label
(0) apprentice, intern, unpaid trainee
(1) low autonomy
(2) low-medium autonomy
(3) medium autonomy
(4) medium-high autonomy
(5) high autonomy

Variable format  1 digit integer

$-$ Survey Years  $=$10, 11, 12, 13

Comment
This variable gives the occupational autonomy for all employed persons. It offers an alternative to the ISCO-based scales on occupational status (ISEI$$), class (EGP$$), or prestige (SIOP$$). AUTONO$$ is the simplest variable based on the scales of “occupational position” in terms of its construction, and strongly correlated with the Treiman Prestige Scale (SIOP$$).

The basis for the “autonomy in occupational activity” scale is the classification of occupational position. Self-employed persons are categorized according to the size of the company (with the exception of farmers, who are all classified within the same category of autonomy, independent of farm size in hectares). Civil servants are differentiated according to the civil service laws defining each kind of activity and the amount of autonomy connected to it. Workers are differentiated according to their vocational training, and thus categorized hierarchically according to the different tasks they can be expected to carry out and the different amounts of responsibility associated with each task. Similarly, salaried employees are classified according to how differentiated their tasks are and how much responsibility is associated with each.

The value “1” is assigned mainly to manual workers with a low level of status and a low level of autonomy. Group 2 encompasses work in production, services demanding a minimal level of specialization, and farm work. Activities that require completion of the middle track of secondary education and entail a limited amount of responsibility are classified in Group 3. Group 4 includes activities carried out either with or without supervision that require a degree from a college of applied sciences or university, but are not very high in prestige. Managers and freelance academics are both placed in Group 5 (highest autonomy). Depending on the number of employees, self-employed are categorized in Group 3, Group 4, or Group 5.

**AGRHRSS$$**

Variable label: “Agreed-upon weekly working hours”

Variable format: 3-digit real

Survey Years: $$=10, 11, 12, 13$$

Comment: This variable is designed to offer annual data on agreed weekly working hours. The variable takes into account only those persons who were in dependent employment (not self-employed) at the time of the survey. The value (“-2 Does not apply”) is assigned to employees without set hours and to self-employed people, including self-employed farmers, freelancers, persons helping out in family businesses, and other self-employed persons.

For implausible answers (agreed weekly working time of more than 80 hours per week) we assign the value (“-3 Answer implausible”). The value is rounded off and gives the number of working hours as a decimal number.

Please also see ACTHRSS$$ and OVRHRSS$$ for other variables on working hours.

**ACTHRSS$$**

Variable label: “Actual weekly working hours”

Variable format: 3-digit real

Survey Years: $$=10, 11, 12, 13$$

Comment: This variable is designed to offer annual data on actual weekly working hours (including overtime) for all persons employed at the time of the survey (including the self-employed). The data are obtained by asking respondents how many hours they work on average per week.

For implausible answers (actual weekly working hours of more than 80 hours per week), we assign the value (“-3 Answer implausible”). The variable is rounded off and gives the number of working hours as a decimal number.

Please also see AGRHRSS$$ and OVRHRSS$$.

**OVRHRSS$$**

Variable label: “Overtime per week”

Variable format: 3-digit real

Survey Years: $$=10, 11, 12, 13$$

Comment: This variable is designed to offer annual data on overtime per week for all persons in dependent employment at the time of the survey. The data are obtained by asking respondents how many overtime hours they worked in the month before the survey. The number of monthly overtime hours is then converted into weekly overtime by dividing the
number given by 4.3. Since OVRHRS$$ refers to weekly overtime during the last month, the number may deviate from the difference between average actual weekly working hours and the agreed weekly working hours.

The value (“-2 Does not apply”) is assigned for self-employment, including self-employed farmers, freelancers, persons helping out in family businesses, and other self-employed persons.

For implausible answers (agreed-upon weekly working time or actual weekly working time of more than 80 hours per week in addition to more than 10 overtime hours per week) we assign the value (“-3 Answer implausible”). The value is rounded off and gives the number of overtime hours as a decimal number.

Please also see AGRHRSS$ and ACTHRSS$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARTP$$</td>
<td>“Partner indicator (in HH)”</td>
</tr>
<tr>
<td>Value label</td>
<td>PARTP$$</td>
</tr>
<tr>
<td>(0) no partner (in HH), clearly</td>
<td></td>
</tr>
<tr>
<td>(1) spouse, clearly</td>
<td></td>
</tr>
<tr>
<td>(2) partner, clearly</td>
<td></td>
</tr>
<tr>
<td>(3) probably spouse</td>
<td></td>
</tr>
<tr>
<td>(4) probably partner</td>
<td></td>
</tr>
<tr>
<td>(5) spouse or partner, probably</td>
<td></td>
</tr>
<tr>
<td>(9) partner exists, identity unknown</td>
<td></td>
</tr>
<tr>
<td>Variable format</td>
<td>1-digit integer</td>
</tr>
<tr>
<td>$$ - Survey Years</td>
<td>$$=10, 11, 12, 13</td>
</tr>
</tbody>
</table>
| Comment | Partner indicators have the purpose of clearly defining spouse-(married) and partner- (unmarried) relationships in FiD households and thus enabling analyses on the couple level. The variable PARTP$$ generated in this context reveals whether a person in a FiD household has a partner in that household, and if so, the type of relationship existing between the partners. Relationships with persons outside the FiD household are not covered by this variable. Code 0 is automatically assigned to all persons born before 1993 or persons living in households in which there is clearly no partnership. These include (a) one-(adult)person households, (b) single-parent households, (c) household head living together with only one parent (or parent-in-law) and (d) youth (turning 17 in survey year). Codes 1 to 5 define the actual relationships. To assign Codes 1 and 2, the partnership has to be clearly defined from the perspective of both partners. This implies agreement between both partners in pointing to the respective partner within the household. No other contradiction may occur, e.g. a different indication by variable $STELL (= relationship to head of household in $Sprutto) or marital status in that wave. Those variables were also used to confirm or revise the partner indicator, so that the indicator turns to 3 or 4 where minor contradictions are found
but could be solved. If not solved, code -3 (“implausible answer”) is assigned. Code 3, 4 or 5 are assigned if partners are associated only via variable $STELL (= relationship to head of household in $pbrutto), e.g. the combination 0 (=head of household) and 1 (=spouse of household head) and amended by information on family status if available. Inconsistencies can occur between the answers provided by the two persons or between data on marital status and relationship to head of household. In those cases each person is examined individually within his or her household context and the marital history is taken into account. If the uncertainty remains, the codes 3 or 4 are assigned. Code 9 is assigned if at least two other household members may potentially be a particular person’s partner and thus no clear determination of partnership can be made.

Note that the variable PARTP$$ corresponds to PARTZ$$ in SOEP.

**PARTNO$$**
Variable label: “Person ID number of partner”
Variable format: 8-digit integer
$$ - Survey Years $$= 10, 11, 12, 13

Comment: Partner indicators have the purpose of clearly defining spouse- (married) and partner- (unmarried) relationships in FiD households and thus to make analyses on the couple level possible.
If PARTP$$ is coded 0 or 9, this person has no partner or the partner cannot be identified as such. The variable PARTNO$$ is assigned the missing code “-2” (does not apply) for these persons. An exception is made if a match of a couple can be clearly accomplished but both persons indicate to not be together any more with the matched partner, i.e. they live together but are separated.
If PARTP$$ is coded 1, 2, 3, 4 or 5, a partnership was defined and PARTNO$$ is assigned the value of the unchanging person ID number (=PERSNR) of the partner.
For analyses of partner relationships, this information can be used to clearly link all persons with their respective partners, and all information on both partners can also be stored in a common dataset. To give as much information as available, it is possible that there are partners or spouses identified, who are actually separated, but still live in the same household. In these cases, the PARTNO$$ is set, but the PARTP$$ is left at 0 (no partner).

**COUPST$$**
Variable label: “Partner status”
Value label: COUPST$$
(1) Married, spouse in household
(2) Married, spouse not in household
(3) Coupled, partner in household
(4) Coupled, partner not in household
(5) Single
(6) Registered same-sex partnership, living together
Registered same-sex partnership, living separately

Variable format: 1-digit integer
Survey Years: 10, 11, 12, 13

Comment: Partner status describes the existence of a partner within or outside the household. COUPST$ is based on information given by the respective person on his or her current relationship. For those whose partner was identified within the household, partner status is counter-checked with the information given by the partner, relation to the head of household as well as future reports on a given relationship. Thus note, that partner status in wave 1 can be different between data distributions due to consistency checks using up-to-date information from wave 2. When contradictions are not solvable, code “-3” (implausible answer) is assigned.

**COUPID$$**

Variable label: “Couple identifier within household”
Variable format: 4-digit integer
Survey Years: 10, 11, 12, 13

Comment: Partner indicators have the purpose of clearly defining spouse- (married) and partner- (unmarried) relationships in FiD households and thus to make possible analyses on the couple level. COUPID assigns a unique, unchangeable ID to an identified couple. For analyses of partner relationships, this information can be used to clearly link all persons with their respective partners, and all information on both partners can also be stored in a common dataset. COUPID$$ is completely consistent with the corresponding biomarsy and biocouply data. Note that the assigned COUPIDs do not follow a consistent rank order in time within a person’s life course.

**MARRST$$**

Variable label: “Marital status”
Value label:

1. Married
2. Married, separated
3. Single (never widowed/divorced)
4. Divorced (most recent event)
5. Widowed (most recent event)
6. Spouse abroad
7. Registered same-sex partnership
8. Registered same-sex partnership, separated

Variable format: 1-digit integer
Survey Years: 10, 11, 12, 13

Comment: Marital status is solely describing the institutional status of marriage at the time of the person interview. To check for the existence of a current relationship refer to COUPST$$. 
Marital status is based on information given by the respective person on his or her current relationship as well as on retrospective information about previous relationships asked in the biography questionnaire. Information on marital status when a child was born (provided in the biography information) is not used here, so contradicting information to biobirth might still be possible. For those whose partner was identified within the household, marital status is counter-checked with the information given by the partner. Where contradictions can be found, indication of the person information is compiled if reasonable. If no information is available, the indication by position related to head of household is deferred. Remaining contradictions are solved using information on marriage status when a child was born as well as future reports on a given relationship. Note that marital status in wave 1 can be different between data distributions due to consistency checks using up-to-date information from wave 2. Marital status is only available for people who are interviewed in the respective wave.

<table>
<thead>
<tr>
<th>NATION$S$</th>
<th>“Citizenship – nationality”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable label</td>
<td>NATION$S$ (see below)</td>
</tr>
<tr>
<td>Value label</td>
<td></td>
</tr>
<tr>
<td>Variable format</td>
<td>3-digit integer</td>
</tr>
<tr>
<td>$S$ - Survey Years</td>
<td>$S$=10, 11, 12, 13</td>
</tr>
</tbody>
</table>

( 1) Germany        (39) Israel        (75) Albania        
( 2) Turkey         (40) Japan         (76) Lebanon        
( 3) Ex-Yugoslavia  (41) Australia      (77) Kyrgyzstan     
( 4) Greece         (42) India          (78) Ukraine         
( 5) Italy          (43) Afghanistan     (79) Algeria         
( 6) Spain          (44) Thailand        (80) Mozambique      
( 7) Ex-GDR (Only Country Of Origin) (45) Jamaica      (81) Egypt      
(10) Austria        (46) Saudi Arabia    (82) Tajikistan      
(11) France         (47) Ethiopia       (83) Vietnam         
(12) Benelux        (48) Columbia       (84) Somalia         
(13) Denmark        (49) Ghana         (85) Pakistan        
(14) Great Britain  (50) Bangladesh     (86) South Africa    
(15) Sweden         (51) Venezuela      (87) UAE            
(16) Norway         (52) Tunisia        (88) El Salvador     
(17) Finland        (53) Mauritius      (89) Eritrea         
(18) USA            (54) Nigeria        (90) Jordan          
(19) Switzerland    (55) Canada         (91) Turkmenistan    
(20) Chile          (56) New Zealand     (92) Costa Rica      
(21) Romania        (57) Tanzania        (93) Singapore       
(22) Poland         (58) Cyprus         (94) Burkina Faso    
(23) Korea          (59) Cuba           (95) Zambia          
(24) Iran           (60) Iraq            (96) Ecuador         
(25) Indonesia      (61) Brazil         (97) Uzbekistan      
(26) Hungary        (62) Monaco         (98) No Nationality  
(27) Bolivia        (63) Hong Kong      (99) Puerto Rico     
(28) Portugal       (64) Peru            (100) Laos          
(29) Bulgaria       (65) Sri Lanka       (101) Estonia        
(30) Syria          (66) Nepal           (102) Angola         
(31) Czech Republic (67) Moroco         (103) Latvia         
(32) Russia         (68) China           (104) Malaysia       
(33) Empty (was Kurdistan) (69) Liechtenstein (105) Namibia     
(34) Mexico         (70) Iceland         (106) Montenegro     
(35) Argentina      (71) Ireland        (107) Belize         
(36) Cap Verde Is. (72) St. Lucia       (108) Dominican Republic 
(37) Benin          (73) Moldavia       (109) Nicaragua      
(38) Philippines    (74) Kazakhstan     (110) Costa Rica
| (110) Kenya | (132) Belarus | (154) Taiwan |
| (111) Libya | (133) Uruguay | (155) Turkmenistan |
| (112) Malta | (134) Bahamas | (156) Africa |
| (113) Botswana | (135) Uganda | (157) Guatemala |
| (114) Haiti | (136) Oman | (158) Sierra Leone |
| (115) Trinidad-Tobago | (137) Micronesia | (159) Panama |
| (116) Luxembourg | (138) Mali | (160) East Timor |
| (117) Belgium | (139) Cameroon | (161) Bahrain |
| (118) Holland | (140) Kosovo-Albania | (162) Senegal |
| (119) Croatia | (141) Georgia | (163) Maldives |
| (120) Bosnia-Herzegovina | (142) Sudan | (164) Kabarday |
| (121) Macedonia | (143) Congo | (165) Serbia |
| (122) Slovenia | (144) Togo | (166) Gambia |
| (123) Slovakia | (145) Mongolia | (167) Honduras |
| (124) Paraguay | (146) Lithuania | (168) Montenegro |
| (125) Guinea | (147) Chad | (169) Cambodia |
| (126) Kuwait | (148) Armenia | (170) Surinam |
| (127) Ivory Coast | (149) Kurdistan | (171) Guyana |
| (128) Malaysia | (150) Liberia | (172) Caucasus |
| (129) Samoa | (151) Yemen | (173) Zimbabwe |
| (130) Azerbaijan | (152) Palestine | (174) Madagascar |
| (131) Seychelles | (153) Freistaat Danzig | (175) Grenada |
| (222) Eastern Europe | | (333) Other, not specified foreign country |
Educational Variables in FiD

As in the SOEP, there are three categories of educational variables in the FiD distribution of $PGEN: those dealing with schooling degrees, those with vocational and university degrees as well as the recoded values for the time spent in education.

In general, the information on all education variables for FiD is taken from two sources 1) The biographical part of the questionnaire in question $L23 to $L40 and 2) the information asked about current education surveyed in every wave (as the question numbers are subject to change every year, we do not quote any specific questions here).

For schooling degrees (SCEDU$$, SCEDUE$$ and SCEDUA$$), we combine the questions on schooling attainment to identify the respondent’s secondary or tertiary school degree. While the respondents from the youth questionnaire are included (age 16 and older), we do not cover elementary schooling degrees. Whether a degree was obtained in Germany (pre or post unification), the former GDR (SCEDUE$$) or in another country (SCEDUA$$), is determined through the questionnaire as well. Note that the abroad category is only possible if the person has already returned to Germany to participate in FiD.

Higher educational degrees are covered in VCDEG$$, VCDEGE$$, VCDEGA$$, VCNONE$$ and COLLEG$$. Similar to the schooling attainment, they are based on the record in the $LELA files and updated by the more recent information if necessary. The information on attainment in the GDR (VCDEGE$$) and abroad (VCDEGA$$) comes from the biographical part, and – for the abroad category – also applies only, if the respondent has already returned to Germany.

Note that educational degrees coded from previous waves are written forward, unless the current wave contains new information. Previous waves are only updated if the new information also applies to that wave. E.g., if a respondent finishes a degree in wave 2, wave 1 data will remain on the old value. If it turns out, that the degree was already obtained before the wave 1 interview, wave 1 information is updated.
**SCEDU$$**

Variable label: “Secondary/tertiary school degree”

Value label: SCEDU$$

(1) Basic-track sec. school (9\textsuperscript{th} grade)
(2) Interm.-track sec. school (10\textsuperscript{th} grade)
(3) Technical secondary school (12\textsuperscript{th} grade)
(4) Academic-track sec. school (13\textsuperscript{th} grade)
(5) Other graduation diploma
(6) Left school without graduating
(7) Not yet graduated

Variable format: 1-digit integer

\$ - Survey Years $=10, 11, 12, 13$

Comment: All respondents are asked about diplomas/degrees attained for completion of secondary/tertiary education in the biographical section of the questionnaire. While the variables SCEDUE$$ and SCEDUA$$ provide school degrees for respondents educated in the former GDR or in a foreign country, respectively, this variable combines all school education information. This data will be regularly updated to take into account any changes in highest diploma/degree attained.

**SCEDUE$$**

Variable label: “Secondary school degree/diploma East Germany”

Value label: SCEDUE$$

(1) completion of 8\textsuperscript{th} grade
(2) completion of 10\textsuperscript{th} grade
(3) college entrance exam
(4) other degree/diploma
(5) dropout, no degree/diploma

Variable format: 1-digit integer

\$ - Survey Years $=10, 11, 12, 13$

Comment: As a supplement to the variable SCEDU$$ the highest secondary school degree/diploma in East Germany is provided as a separate variable. This information only originates from the biographical part of the questionnaire.

**SCEDUA$$**

Variable label: “Secondary school degrees/diplomas abroad”

Value label: SCEDUA$$

(1) secondary school, no degree/diploma attained
(2) secondary school, degree/diploma attained
(3) vocational school

Variable format: 1-digit integer

\$ - Survey Years $=10, 11, 12, 13$
Comment
As a supplement to the variable SCEDU$$, this variable provides annually updated data on the highest secondary school degree/diploma attained abroad.

VCDEG$$
Variable label “Vocational degree attained”
Value label VCDEG$$
(1) apprenticeship
(2) vocational school
(3) [health care school]
(4) technical school
(5) civil service training
(6) other training
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13
Comment
All respondents in all subsamples are asked about vocational degrees attained the first time they participate in FiD. This data is updated annually. While VCDEGE$$ and VCDEGA$$ provide information on vocational degrees obtained in the former GDR and abroad, respectively, VCDEG$$ provides this information on all respondents. VECDEG$$ captures the highest vocational degree attained, and in case of multiple degrees, VCDEG$$ and VCDEGA$$ may not correspond with VECDEG$$.
Note that code (3) “health care school” is a category present in the SOEP, which was kept in FiD to allow direct comparisons. In later waves of the SOEP, this category is integrated in category (4).

VCDEGE$$
Variable label “Vocational degree attained – East”
Value label VCDEGE$$
(1) vocational training
(2) master craftsman
(3) engineering, technical degree
(4) other training
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13
Comment
To supplement the variable VCDEG$$ the highest secondary school degree/diploma in East Germany is provided as a separate variable.

VCDEGA$$
Variable label “Vocational degree abroad”
Value label VCDEGA$$
(1) on-the-job training
(2) vocational training
(3) vocational school
(4) college
(5) other
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13
Comment As a supplement to the variable VCDEG$$, this variable gives (and updates) the highest-level vocational degree attained abroad.

VCNONE$$
Variable label “No vocational degree”
Value label VCNONE$$
(1) no vocational degree
(2) still doing an apprenticeship
(3) still in university
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13
Comment In connection with the question about vocational degrees (VCDEG$$ and VCDEGE$$), all first-time respondents to all subsamples are explicitly asked, whether they (still) do not possess a vocational degree. In subsequent years, this data will be carried forward or updated. The variable has the Missing Value Code “-2 Does not apply”, if one of the other two variables on vocational degree has a positive value.

COLLEG$$
Variable label “Completed college education”
Value label COLLEG$$
(1) technical college
(2) university, technical university
(3) college abroad
(4) engineering, technical school (East)
(5) university (East)
(6) doctorate degree
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13
Comment All respondents in all subsamples are asked about completed college education the first time they participate in FiD. To generate the variable, the different degrees/diplomas are integrated. Note that code “3 college abroad” is only assigned if the person reports to have obtained a degree when studying abroad. This is different to FiD version 1.2, where simply being abroad during college gave a code 3.

TIMEDU$$
Variable label “Amount of education or training (in years)”
Variable format 2-digit real
$$ - Survey Years $$=10, 11, 12, 13
Comment
The following statements describe the standard computation for schooling (including years of secondary vocational education). The re-coding is not very complicated. For example, special schools for health care professions or other kinds of specialized schools are included in the “technical school” label. However, in Germany, this code is the one commonly used when earnings functions based on human capital theory are estimated.

Computation
The $BILZEIT$ variables are computed using the education variables provided by the $PGEN$-files. The computation code is set to the following logic:

<table>
<thead>
<tr>
<th>Degree Type</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>no degree</td>
<td>7</td>
</tr>
<tr>
<td>lower school degree</td>
<td>9</td>
</tr>
<tr>
<td>intermediary school</td>
<td>10</td>
</tr>
<tr>
<td>degree for a professional college</td>
<td>12</td>
</tr>
<tr>
<td>high school degree</td>
<td>13</td>
</tr>
<tr>
<td>other</td>
<td>10</td>
</tr>
<tr>
<td>additional occupational training (includes universities)</td>
<td></td>
</tr>
<tr>
<td>apprenticeship</td>
<td>1.5</td>
</tr>
<tr>
<td>technical schools (incl. health)</td>
<td>2</td>
</tr>
<tr>
<td>civil servants apprenticeship</td>
<td>1.5</td>
</tr>
<tr>
<td>higher technical college</td>
<td>3</td>
</tr>
<tr>
<td>university degree</td>
<td>5</td>
</tr>
</tbody>
</table>


**ISCED**

<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable label</td>
<td>“ISCED-1997 classification”</td>
</tr>
<tr>
<td>Value label</td>
<td>ISCEDS$</td>
</tr>
<tr>
<td>(0)</td>
<td>in school</td>
</tr>
<tr>
<td>(1)</td>
<td>inadequately</td>
</tr>
<tr>
<td>(2)</td>
<td>general elementary</td>
</tr>
<tr>
<td>(3)</td>
<td>middle vocational</td>
</tr>
<tr>
<td>(4)</td>
<td>vocational + Abitur</td>
</tr>
<tr>
<td>(5)</td>
<td>higher vocational</td>
</tr>
<tr>
<td>(6)</td>
<td>higher education</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable format</th>
<th>1-digit integer</th>
</tr>
</thead>
<tbody>
<tr>
<td>$$$ - Survey Years</td>
<td>$$$=10, 11, 12, 13</td>
</tr>
</tbody>
</table>

Comment
To make the educational degrees and diplomas attained in different countries comparable, for all respondents an educational variable
(ISCED$$) is generated retroactively from 1984 on using the international classification scheme ISCED-1997 (International Standard Classification of Education). It creates the highest degree/diploma attained, taking into account degrees and diplomas attained in both general schooling and in vocational and university education. Here the higher-level vocational and university override lower-level school diplomas. Persons who, for example, have no values for the variables on secondary school degrees/diplomas but state that they have a university degree are placed in the highest ISCED category.

Please note that, due to a lack of more detailed information on tertiary degrees -- in particular on promotion -- we include all tertiary degrees in our ISCED category 6. Thus, the ISCED variable provided here is not comparable one-to-one with the ISCED levels as defined by the OECD, since we have included the original ISCED level 5A in our ISCED category 6. See below for more details.

Computation

The ISCED$$ variables are computed using the education variables provided by the SPGEN-files. For this we use the variables on secondary degrees/diplomas (SCEDU$$) and secondary degrees/diplomas abroad (SCEDAU$$), and the occupational education variables “vocational degree” (VCDEG$$), “university degree” (COLLEG$$) and “vocational degree abroad” (VCDEGA$$). We refrained from integrating the GDR-specific educational degrees/diplomas (SCEDUE$$ and VCDEGE$$).

<table>
<thead>
<tr>
<th>ISCED</th>
<th>SCEDU</th>
<th>SCEDUA</th>
<th>VCDEG</th>
<th>VCDEGA</th>
<th>COLLEG</th>
<th>ISCED97 – OECD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7. Noch kein Abschluss</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 noch kein Abschluss</td>
</tr>
<tr>
<td>1</td>
<td>5. Anderer Abschluss 6.Ohne Abschluss verlassen</td>
<td>1.Pflichtschule ohne Abschluss</td>
<td></td>
<td></td>
<td></td>
<td>1 ohne Abschluss verlassen</td>
</tr>
</tbody>
</table>

**CASMIN**

**Variable label**
“Highest degree/diploma according to CASMIN”

**Value label**
CASMIN

- (0) (0) in school'
- (1) (1a) inadequately completed'
- (1b) general elementary school'
- (2) (1c) basic vocational qualification'
- (3) (2a) intermediate general qualification'
- (4) (2b) intermediate vocational'
- (5) (2c_gen) general maturity certificate'
- (6) (2c_voc) vocational maturity certificate'
- (7) (3a) lower tertiary education'
- (8) (3b) higher tertiary education'

**Variable format**
1-digit integer

**$$ - Survey Years**
$$=10, 11, 12, 13

**Comment**
As an alternative to $ISCED, a second educational variable is generated ($CASMIN) that also enables comparison with international educational degrees/diplomas. Based on the modified CASMIN classification scheme (Comparative Analysis of Social Mobility in Industrial Nations), this variable has been computed retroactively from 1984 on for all respondents. Taken into account are both secondary-level and university/college-level degrees and diplomas. As with $ISCED, the higher-level occupational degrees override the lower-level secondary school degrees.

**Computation**
The ISCED$$ variables are computed using the education variables provided by the $PGEN-files. For this we use the variables on secondary degrees/diplomas (SCEDU$$) and secondary degrees/diplomas abroad (SCEDA$$), and the occupational education variables “vocational degree” (VCDEG$$), “university degree” (COLLEG$$) and “vocational degree abroad” (VCDEGA$$). We refrained from integrating the GDR-specific educational degrees/diplomas (SCEDUE$$ und VCDEGE$$).

<table>
<thead>
<tr>
<th>CASMIN</th>
<th>SCEDU Schulabschluss</th>
<th>SCEDUA Schule im Ausland</th>
<th>VCDEG Berufliche Ausbildung</th>
<th>VCDEGA Berufsbildung im Ausland</th>
<th>COLLEG Hochschulabschluss</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 0</td>
<td>7. Noch kein Abschluss</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 – 1a</td>
<td>5. Anderer Abschluss</td>
<td>1. Pflichtschule ohne Abschluss</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6. Ohne Abschluss verlassen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 – 1b</td>
<td>1. Hauptschulabschluss</td>
<td>2. Pflichtschule mit Abschluss</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6. Ohne Abschluss verlassen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### LABGROS$\$ Variable label

“Current gross labor income in euros (generated)”

### Variable format

5-digit integer

### $\$$ - Survey Years

$\$$=10, 11, 12, 13

### Comment

The variable LABGROS$\$ represents the imputed current gross labor income generated for all FiD respondents, who are employed in the respective wave. LABGROS$\$ is part of the individual imputation process conducted by multiple imputations in FiD (see documentation for details). Even though we recommend using all five implicates available in the dataset Snapinc, we include the first implicate for convenience reasons in Spgen.

The imputations for LABGROS$\$ are based on the variable LABNET$\$$. The difference between the two is imputed, and in case of a missing value in either of the two original variables, LABGROS$\$ is imputed by summing up the imputed values of LABNET$\$ and the imputed difference between LABNET$\$ and LABGROS$. This procedure leads to slightly more imputed values, but guarantees that the gross income does not fall below the net income.

Imputed values are flagged (IMPGROS$\$).
**IMPGROSS**

Variable label  
“Imputation flag for LABGROSS”

Value label  
IMPGROSS

(0) observed value
(1) imputed value

Variable format  
1-digit Integer

$\$$ - Survey Year  
$\$$=10, 11, 12, 13

Comment  
The variable IMPGROSS identifies imputations of item-nonresponse in the variable LABGROSS (current gross labor income). Note that there are slightly more imputations here than there are missing values due to the imputation process in LABGROSS.

**LABNET$$**

Variable label  
“Current net labor income (generated) in euros”

Variable format  
5-digit integer

$\$$ - Survey Years  
$\$$=10, 11, 12, 13

Comment  
The variable LABNET$$ represents the imputed current gross labor income generated for all FiD respondents, who are employed in the respective wave. LABNET$$ is part of the individual imputation process conducted by multiple imputations in FiD (see documentation for details). Even though we recommend using all five implicates available in the dataset $\text{Smipinc}$, we include the first implicate for convenience reasons in $\text{SPGEN}$.

Imputed values are flagged (IMPNET$$).

**IMPNET$$**

Variable label  
“Imputation flag for LABNET$$”

Value label  
IMPNET$$

(0) observed value
(1) imputed value

Variable format  
1-digit Integer

$\$$ - Survey Year  
$\$$=10, 11, 12, 13

Comment  
The variable IMPNET$$ identifies imputations of item-nonresponse in the variable LABNET$$ (current net labor income).

**FIELD$$**

Variable label  
“Field of tertiary education”

Value label  
FIELD$$

1-98 (see below)

Variable format  
2-digit Integer

$\$$ - Survey Year  
$\$$=10, 11, 12, 13
**Spgen: Person related status variables and generated variables**

<table>
<thead>
<tr>
<th>Field</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applied Linguistics and Cultural Studies</td>
<td>1</td>
</tr>
<tr>
<td>Protestant Theology</td>
<td>2</td>
</tr>
<tr>
<td>Catholic Theology</td>
<td>3</td>
</tr>
<tr>
<td>Philosophy</td>
<td>4</td>
</tr>
<tr>
<td>History</td>
<td>5</td>
</tr>
<tr>
<td>Library Science, Archival Studies, Journalism</td>
<td>6</td>
</tr>
<tr>
<td>Literary Studies, Linguistics</td>
<td>7</td>
</tr>
<tr>
<td>Classical Philology, Modern Greek</td>
<td>8</td>
</tr>
<tr>
<td>German Philology</td>
<td>9</td>
</tr>
<tr>
<td>English Studies</td>
<td>10</td>
</tr>
<tr>
<td>Romance Philology</td>
<td>11</td>
</tr>
<tr>
<td>Slavonic Studies</td>
<td>12</td>
</tr>
<tr>
<td>Non-European Languages and Cultural Studies</td>
<td>13</td>
</tr>
<tr>
<td>Cultural Studies</td>
<td>14</td>
</tr>
<tr>
<td>Psychology</td>
<td>15</td>
</tr>
<tr>
<td>Educational Science</td>
<td>16</td>
</tr>
<tr>
<td>Special Education</td>
<td>17</td>
</tr>
<tr>
<td>Sport Science</td>
<td>22</td>
</tr>
<tr>
<td>Law, Economics, Social Sciences</td>
<td>23</td>
</tr>
<tr>
<td>Regional Studies</td>
<td>24</td>
</tr>
<tr>
<td>Political Science</td>
<td>25</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>26</td>
</tr>
<tr>
<td>Social Work</td>
<td>27</td>
</tr>
<tr>
<td>Law</td>
<td>28</td>
</tr>
<tr>
<td>Public Management and Governance</td>
<td>29</td>
</tr>
<tr>
<td>Economics</td>
<td>30</td>
</tr>
<tr>
<td>Business Administration and Engineering</td>
<td>31</td>
</tr>
<tr>
<td>Mathematics, Natural Sciences</td>
<td>32</td>
</tr>
<tr>
<td>Mathematics</td>
<td>36</td>
</tr>
<tr>
<td>Computer Science</td>
<td>37</td>
</tr>
<tr>
<td>Physics, Astronomy</td>
<td>38</td>
</tr>
<tr>
<td>Chemistry</td>
<td>39</td>
</tr>
<tr>
<td>Pharmacology</td>
<td>40</td>
</tr>
<tr>
<td>Biology</td>
<td>41</td>
</tr>
<tr>
<td>Geosciences</td>
<td>42</td>
</tr>
<tr>
<td>Geography</td>
<td>43</td>
</tr>
<tr>
<td>Health Sciences</td>
<td>48</td>
</tr>
<tr>
<td>Medicine (without dentistry)</td>
<td>49</td>
</tr>
<tr>
<td>Dentistry</td>
<td>50</td>
</tr>
<tr>
<td>Veterinary Science</td>
<td>51</td>
</tr>
<tr>
<td>Landscape Conservation, Environmental Design</td>
<td>57</td>
</tr>
<tr>
<td>Agricultural Sciences, Food and Beverage Science</td>
<td>58</td>
</tr>
<tr>
<td>Forest Management, Wood Management</td>
<td>59</td>
</tr>
<tr>
<td>Nutrition and Home Economics</td>
<td>60</td>
</tr>
<tr>
<td>Engineering</td>
<td>61</td>
</tr>
<tr>
<td>Mining and Metallurgy</td>
<td>62</td>
</tr>
<tr>
<td>Mechanical Engineering</td>
<td>63</td>
</tr>
<tr>
<td>Electrical Engineering</td>
<td>64</td>
</tr>
<tr>
<td>Traffic Engineering, Nautical Science</td>
<td>65</td>
</tr>
<tr>
<td>Architecture, Interior Design</td>
<td>66</td>
</tr>
<tr>
<td>Regional Planning</td>
<td>67</td>
</tr>
<tr>
<td>Civil Engineering</td>
<td>68</td>
</tr>
<tr>
<td>Surveying and Mapping</td>
<td>69</td>
</tr>
<tr>
<td>Art, Aesthetics</td>
<td>74</td>
</tr>
<tr>
<td>Fine Arts</td>
<td>75</td>
</tr>
<tr>
<td>Design</td>
<td>76</td>
</tr>
<tr>
<td>Performance, Film and Television, Theater</td>
<td>77</td>
</tr>
<tr>
<td>Music, Musicology</td>
<td>78</td>
</tr>
<tr>
<td>Outside the structure of the university system</td>
<td>83</td>
</tr>
</tbody>
</table>

**Comment**

The variable is designed to provide information on the field of education of tertiary degrees which adds details to the information recorded in the variable `COLLEGE$$. While the latter variable records if a person holds a degree `FIELD$$ contains more detailed information on the type of the degree. The data of the generated variable `FIELD$$ stem from two sources: 1. Person questionnaire: Each year respondents are asked if they have left education since the beginning of the year prior to the survey and which degrees they have obtained. This part of the questionnaire contains an open question on the type and the field of newly obtained tertiary degrees. This information is coded and used for the generation of the variables `FIELD$$. 2. Biography questionnaire: Similar information is collected from respondents who fill in the biography questionnaire (usually during the first two years of participation in the panel). In contrast to the information from the person questionnaire the questions do not refer to currently obtained degrees but to degrees obtained during the time before being part of the FID sample.

In the variable `FIELD$$ we combine these two types of information. Information on the data source is stored in the variable `FDT_F$$.
Each year the variable contains the most recently collected information. Take for instance a person for whom we have observed a first degree in sociology in 2010 and a second degree in economics in 2012. For this person the variables FIELD$$ would be filled as follows:

2010-2011: 26 political/social science  
2012-today  30 economics

If you want to take into account that a person holds two degrees you have to combine the information from all available years. However, only a minority of the population holds more than one tertiary degree. In very few cases we encounter the problem that a respondent provides information on two different degrees in one survey year. This only happens in years when respondents fill in the person as well as the biography questionnaire. In these cases we prioritize the information from the person questionnaire as it refers to the current situation while the biography questionnaire contains retrospective information. Furthermore, there are cases who report an applied university degree and a university degree in the biography questionnaire. In these cases, the variable contains information on the university degree only.

The variable is coded according to the classification on fields of education ("Fächergruppen") provided by the Statistisches Bundesamt (2009).

**DEGREE$$**

Variable label  
“Type of tertiary degree”

Value label  
DEGREE$$

11-98 (see below)

Variable format  
2-digit Integer

$$_{-}$$ - Survey Year  
$$=10, 11, 12, 13$$

(11) Magister  
(12) Diplom (University)  
(13) Bachelor  
(14) Master  
(15) 1st State Examination  
(16) Other state examination  
(21) Diplom (at technical college, technical college for administration)  
(22) Bachelor (at technical college, technical college for administration)  
(31) Master (at technical college, technical college for administration)  
(32) Teacher training,BA,MA at elementary, lower secondary schools/primary level  
(33) Teacher training,BA,MA at 2ndary level 1/elementary schools/primary level  
(34) Teacher training,BA,MA at intermediate sendry schools/sendry level I  
(35) Teacher training, BA, MA at secondary level II and I  
(36) Teacher training,BA,MA at academic 2ndry schools,2ndry level 2,genrl school  
(37) Teacher training, BA, MA at special needs schools
The variable is designed to provide information on the type of tertiary degree (e.g., Diploma, Bachelor, Master) which adds details to the information recorded in the variable COLLEG$. While the latter variable records if a person holds a degree DEGREE$ contains more detailed information on the type of the degree. The data of the generated variable DEGREE$ stem from two sources: 1. Person questionnaire: Each year respondents are asked if they have left education since the beginning of the year prior to the survey and which degrees they have obtained. This part of the questionnaire contains an open question on the type and the field of newly obtained tertiary degrees. This information is coded and used for the generation of the variables DEGREE$. 2. Biography questionnaire: Similar information is collected from respondents who fill in the biography questionnaire (usually during the first two years of participation in the panel). In contrast to the information from the person questionnaire the questions do not refer to currently obtained degrees but to degrees obtained during the time before being part of the FID sample.

In the variable DEGREE$ we combine these two types of information. Information on the data source is stored in the variable FDT_F$.

Each year the variable contains the most recently collected information. Take for instance a person for whom we have observed first an applied university diploma in 210 and a university diploma in 2012. For this person the variables DEGREE$ would be filled as follows:

2010-2011: 21 diploma (applied university)
2012-today 12 diploma (university)

If you want to take into account that a person holds two degrees you have to combine the information from all available years. However, only a minority of the population holds more than one tertiary degree. In very few cases we encounter the problem that a respondent provides information on two different degrees in one survey year. This only happens in years when respondents fill in the person as well as the biography questionnaire. In these cases we prioritize the information from the person questionnaire as it refers to the current situation while the biography questionnaire contains retrospective information. Furthermore, there are cases who report an applied university degree and a university degree in the biography questionnaire. In these cases, the variables contain information on the university degree only.

The variable is coded according to a slightly collapsed version of the classification on types of tertiary degrees ("Prüfungsgruppen und Abschlussprüfungen") provided by the Statistisches Bundesamt (2009).

**TRAINASS**

**Variable label** “Apprenticeship – two-digit occupation KldB92”

**Value label** TRAINASS

1-99 (see below)

**Variable format** 2-digit Integer

**$\$ - Survey Year** $\$=10, 11, 12, 13

(1) Agricultural Occupations (Crops)
(2) Agricultural Occupations (Livestock)
(3) Administrative/Advisory/Technical Specialist In Agriculture
(5) Horticultural Occupations
(6) Forestry and Hunting Occupations
(7) Mineworkers
(8) Mineral Exploitation and Processing
(11) Manufacturers of Construction Materials
(12) Ceramists
(13) Glass Manufacturing Occupations
(14) Chemical Industry Occupations
(15) Plastics Manufacturing Occupations
(16) Paper Manufacturing and Processing
(17) Printing Occupations
(18) Wood and Woodworking, Wickerwork Occupations
(19) Occupations in Iron and Steelmaking, Semi-Finished Products
(20) Pouring and Casting Occupations
(21) Metal Processing Occupations (Chipless Forming)
(22) Metal Processing Occupations (Chip Forming)
(23) Occupations in the Metal Surface Treatment and Finishing Industry
(24) Metal Compounding Occupations
(25) Metal and Plant Construction Occupations
(26) Sheet Metal Manufacturing Occupations
(27) Mechanical Engineering and Maintenance Occupations
(28) Automotive and Aircraft Manufacturing and Maintenance Occupations
(29) Tool and Die Making Occupations
(30) Precision Engineering and Related Occupations
(31) Electrical Occupations
(32) Metalworkers, not otherwise mentioned
(33) Occupations in Spinning
(34) Occupations in Textile Production
(35) Occupations in Textile Processing
(36) Occupations in Textile Finishing
(37) Occupations in Leather Production and Processing
(39) Occupations in Baking, Confectionery, and Candy Production
(40) Butchers
(41) Chefs
(42) Occupations in Beverages, Alcohol, and Tobacco Manufacturing
(43) Other Occupations in Nutrition
(44) Occupations in Structural Engineering
(46) Occupations in Civil Engineering
(47) Unskilled Construction Workers
(48) Construction Finishing Occupations
(49) Interior Decorator, Upholsterers
(50) Occupations in Woodworking and Polymer Processing
(51) Painters, Varnishers, and Related Occupations
(52) Quality Control Inspectors, Mailing, and Dispatching Staff
(53) Unskilled Laborers, responsibilities not specified
(54) Machine Operators, not otherwise specified
(55) Machine Fitters, not otherwise mentioned
(60) Engineers, not otherwise mentioned
(61) Chemists, Physicists, Mathematicians
(62) Technicians, not otherwise mentioned
(63) Technical Specialists
(64) Technical Drafting, related occupations
(65) Industrial, Factory, Training Foremen/Forewomen
(66) Salespeople
(67) Wholesale and Retail Salespeople, Purchasing and Sales Staff
(68) Product Sales Staff, not otherwise specified, Sales Representatives
(69) Banking, Savings Association, Insurance Specialists
(70) Service Industry and Related Occupations
(71) Surface Transport Occupations
(72) Water and Air Traffic Occupations
(73) Communications Occupations
(74) Stock Clerks, Warehouse and Transport Workers
(75) Occupations in Management, Consulting, and Auditing
Comment:
The variable is designed to provide information on the occupation of vocational training which adds details to the information recorded in the variable VCDEG$$. In addition to the variable TRAINA$$ we provide the variables TRAINB$$, TRAINC$$ and TRAIND$$. All these variables record the occupation of vocational training. The difference is that TRAINA$$ contains information on vocational training within the German dual system which combines firm-based and school-based training (apprenticeship). TRAINB$$ is designed to provide information on the occupation of full-time school based vocational training. TRAINC$$ contains information on level vocational training (e.g., Meister, Techniker). TRAIND$$ is designed to provide information on the occupation of civil servant training (“Beamtenausbildung”). We describe in brief detail the construction of the variable TRAINA$$. TRAINB$$, TRAINC$$ and TRAIND$$ are constructed in an analogous manner.

The data of the generated variable TRAINA$$ stem from two sources: 1. Person questionnaire: Each year respondents are asked if they have left education since the beginning of the year prior to the survey and which degrees they have obtained. This part of the questionnaire contains an open question on the type and the field of newly obtained tertiary degrees. This information is coded and used for the generation of the variables TRAINA$$. 2. Biography questionnaire: Similar information is collected from respondents who fill in the biography questionnaire (usually during the first two years of participation in the panel). In contrast to the information from the person questionnaire the questions do not refer to currently obtained vocational qualifications but to qualifications obtained during the time before being part of the FID sample.

In the variable TRAINA$$ we combine these two types of information. Information on the data source is stored in the variable FDT_FS$$. Each year the variable contains the most recently collected information. Take for instance a person for whom we have observed a first vocational qualification as an electrician in 2010 and a second
qualification as a car mechanic in 2012. For this person the variables TRAINA$$ would be filled as follows:

2010-2011: 31 electrical occupation  
2012-today 28 automotive/flight industry occupation

If you want to take into account that a person holds two vocational qualifications you have to combine the information from all available years. In few cases we encounter the problem that a respondent provides information on two different apprenticeships in one survey year. This only happens once, namely in years when respondents fill in the person as well as the biography questionnaire. In these cases we prioritize the information from the person questionnaire as it refers to the current situation while the biography questionnaire contains retrospective information.

The variable is coded according to the classification of occupations at two-digit level (“Berufsgruppen”) provided by the Statistisches Bundesamt (1992).

**TRAINB$$**

**Variable label**  
“Vocational school – two-digit occupation KldB92”

**Value label**  
TRAINB$$

**Variable format**  
1-99 (see TRAINA$$)

**Variable format**  
2-digit integer

**$ - Survey Year**  
$=10, 11, 12, 13

**Comment**  
The variable is designed to provide information on the occupation of full-time school based vocational training (e.g., Berufsfachschule, Schule des Gesundheitswesens, Handelsschule). See the description of variable TRAINA$$ for more details on the construction and the values of the variable.

**TRAINC$$**

**Variable label**  
“Higher vocational school – two-digit occupation KldB92”

**Value label**  
TRAINC$$

**Variable format**  
1-99 (see TRAINA$$)

**Variable format**  
2-digit integer

**$ - Survey Year**  
$=10, 11, 12, 13

**Comment**  
The variable is designed to provide information on the occupation of higher level vocational training (e.g., Meister, Techniker). See the description of variable TRAINA$$ for more details on the construction and the values of the variable.
**TRAIND$$**

Variable label: “Civil servant training – two-digit occupation KldB92”

Value label: TRAIND$$
1-99 (see TRAINA$$)

Variable format: 2-digit Integer

$-$ Survey Year: $=$10, 11, 12, 13

Comment: The variable is designed to provide information on the occupation of civil servant training (“Beamtenausbildung”). See the description of variable TRAINA$$ for more details on the construction and the values of the variable.

**FDT_F$$**

Variable label: “Data source FIELD, DEGREE, TRAIN”

Value label: FDT_F$$
(1) person questionnaire
(2) person questionnaire (temporary drop-out)
(3) biography questionnaire
(4) various sources

Variable format: 1-digit integer

$-$ Survey Year: $=$10, 11, 12, 13

Comment: This is a flag variable which provides information on the data sources used for the construction of the variables FIELD$$, DEGREE$$, TRAINA$$, TRAINB$$, TRAINC$$ and TRAIND$$ (see the description of the respective variables for details).
Documentation $\text{Shgen}$

Household-related status variables and generated variables

Mathis Fräßdorf (geb. Schröder) and Malisa Zobel

This documentation is based on the comparable SOEP documentation on $\text{Shgen}$ and has benefited from previous work of Joachim Frick, Markus Grabka, Jan Goebel, Peter Krause, Olaf Groh-Samberg and Christian Schmitt. Please understand that for readability reasons, we do not specifically cite and specify text that has been used directly from the SOEP document.
General information

The $hgen$ datasets provide household level information for each wave of the survey. In some sense, the variables included in $hgen$ allow the researcher to have an easier access to household information. As the variable names and formats are consistent over all years, a combination of these files allows an easy comparison of different years. In addition, some variables are completely generated or imputed, and hence provide information which is not included in the regular questionnaire. The following documentation lists all variables in $hgen$ and provides information on their generating process. Information provided looks as follows for all variables:

<table>
<thead>
<tr>
<th>Variable Label</th>
<th>Provides the label of the variable as it is given in the dataset. Variables are given in CAPTIAL letters, even though they might appear in small letters in the dataset. This is simply for readability.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value Labels</td>
<td>$LBLNME$</td>
</tr>
<tr>
<td></td>
<td>In case $VARNME$ is categorical, $LBLNME$ specifies the labels for each category, and the value labels are listed here. Note that the standard missing value labels (-1: No answer; -2: Does not apply; -3: Not valid) are not listed, but apply to all variables in this dataset.</td>
</tr>
<tr>
<td>Variable format</td>
<td>Specifies the format for each variable, e.g. “1-digit integer” or “string”.</td>
</tr>
<tr>
<td>$$ - Survey Years</td>
<td>Specifies the years for which the variable is provided. This is provided for 2000+, such that “10” refers to 2010, etc.</td>
</tr>
<tr>
<td>Comment:</td>
<td>Provides more detailed information on the generating process, also on the population the variable is specified for, if necessary. Here, variables used, changes between waves, or any other anomalies are mentioned and their relevance explained.</td>
</tr>
</tbody>
</table>

The variables described in the following are in part status variables in this sense: information collected once will be carried forward to subsequent years if no address change has taken place since the previous year. This is the case for: CNSTYR$$, CONDIT$$, SIZE$$, ROOM$$, EQPKIT$$, EQPSHW$$, EQPIWC$$, EQPHEA$$, EQPTER$$, EQPBASS$$, EQPGAR$$, EQPWAT$$, EQPTEL$$, EQPALM$$, EQPSOL$$, EQPAIR$$, MOVEYR$$, RSUB$$ and REDUC$$. 

SOEP Survey Paper 332
**Comparison with SOEP**

The following variables, which are part of the SOEP distribution, are not part of FiD, as this information has not been collected. Whether the information will be included in a future wave of FiD, is not clear at the moment.

- **ACQUISS$** Means of acquiring dwelling
- **EQPTEL$$** Dwelling has telephone
- **EQPIWCS$$** Dwelling has indoor toilet
- **SUBSID$$** Government-subsidized housing payments

On the other hand, there are two variables for which information was collected for the first time (it was also collected in SOEP 2010):

- **EQPFHEA$$** Dwelling has floor heat
- **ELECTR$$** Amount of monthly electricity costs in Euro

If you have questions regarding *Shgen* data for the FiD-distribution, please contact Mathis Schröder.
List of Variables:

Contents

General Documentation ......................................................................................................... 1
  General information ........................................................................................................... 3
  The very basics of “Familien in Deutschland” ................................................................ 4
  Data structure: types of data files, names and data organization .................................... 5
  Combining data files ........................................................................................................ 7
  FiD data files .................................................................................................................. 9
    Basic Data Files .......................................................................................................... 9
    Original Data Files .................................................................................................... 11
    Generated Data Files ................................................................................................. 15
  Datasets not in FiD, which are known from the SOEP .................................................. 20

Documentation ppfad ....................................................................................................... 21
  General information ...................................................................................................... 22
  List of variables .......................................................................................................... 23
    HHNR ......................................................................................................................... 24
    PERSNR ..................................................................................................................... 24
    $SHHNR ................................................................................................................... 24
    PSAMPLE ............................................................................................................... 24
    SEX ......................................................................................................................... 25
    GEBJAHIR ............................................................................................................... 25
    GEBMONAT ............................................................................................................. 25
    GEBMOVAL ............................................................................................................. 25
    TODJAHIR .............................................................................................................. 26
    TODINFO ............................................................................................................... 26
    LOC1989 ................................................................................................................. 26
    $NETTO .................................................................................................................. 26
    $NETOLD ............................................................................................................... 27
    $CASEMAT ............................................................................................................... 28
    $POP ....................................................................................................................... 28
    $SAMPREG ............................................................................................................. 28
    GERMBOHN ............................................................................................................ 29
    IMMIYEAR .............................................................................................................. 29
    CORIGIN ............................................................................................................... 29
    MIGBACK .............................................................................................................. 30
    MIGININFO .......................................................................................................... 31

Documentation hpfad ...................................................................................................... 32
  General information ...................................................................................................... 33
  List of variables: .......................................................................................................... 34
    HHNR ......................................................................................................................... 35
    HHNRAKT .............................................................................................................. 35
    HSSAMPLE ............................................................................................................ 35
    $SHHNR ................................................................................................................... 35
    $SAMPREG ............................................................................................................. 36
$SHPOP................................................................................................................................. 36
$SHNETTO............................................................................................................................. 36
SINGPA ............................................................................................................................... 37
LRGFAM ............................................................................................................................. 37
LOWINC ............................................................................................................................... 37

Documentation $pgen .................................................................................................................. 39

General information ............................................................................................................ 39
List of variables ................................................................................................................... 41
EMPLST$$............................................................................................................................. 43
LFS$$ ........................................................................................................................................ 43
JOBCH$$ ............................................................................................................................. 44
EXPFT$$ ................................................................................................................................ 45
EXPPCT$$ ............................................................................................................................. 46
EXPUES$$ ............................................................................................................................. 46
TENURE$$............................................................................................................................... 47
JOBTRN$$................................................................................................................................ 48
REOQT$$ ................................................................................................................................... 48
CIVIL$$...................................................................................................................................... 49
CRS$$ ......................................................................................................................................... 49
CIVILSS.................................................................................................................................. 50
OCORPS.................................................................................................................................. 50
CLASS$$.................................................................................................................................. 51
IS88$$......................................................................................................................................... 53
NACE$$..................................................................................................................................... 54
ISEI$$......................................................................................................................................... 56
EGP$$.......................................................................................................................................... 56
SIOPS$$ ...................................................................................................................................... 58
MP$$.......................................................................................................................................... 58
AUTONOS.................................................................................................................................. 59
AGRHRSS.................................................................................................................................. 60
ACTRHRSS.................................................................................................................................. 60
OVHRRSS.................................................................................................................................. 60
PAR$$......................................................................................................................................... 61
PARTNOS$$............................................................................................................................... 61
COUPST$$.................................................................................................................................. 62
COUPID$$.................................................................................................................................. 62
MARRST$$.................................................................................................................................. 63
NATION$$.................................................................................................................................. 64

Educational Variables in FiD ..................................................................................................... 66
SCEDUS$$................................................................................................................................... 67
SCEDUESS.................................................................................................................................. 67
SCEDUA$$.................................................................................................................................. 67
VCDEGS$$................................................................................................................................... 68
VCDEGESS.................................................................................................................................. 68
VCDEGAS$$.................................................................................................................................. 68
VCNONSE$$.................................................................................................................................. 69
COLLEG$$................................................................................................................................... 69
TIME$$......................................................................................................................................... 69
ISCED$$....................................................................................................................................... 70
CASMIN$$................................................................................................................................... 72
List of Variables: .................................................................................................................. 85
Comparison with SOEP ....................................................................................................... 84

NORENT$$ .................................................................................................................... 103
FELECTR$$ ................................................................................................................... 103
FUTIL$$ ......................................................................................................................... 103
FRENT$$ ....................................................................................................................... 103
ELECTR$$ ..................................................................................................................... 102
HEAT$$ ......................................................................................................................... 102
RENT$$ ......................................................................................................................... 102
RSUB$$ .......................................................................................................................... 101
REVAL$$ ....................................................................................................................... 101
OSUB$$ ....................................................................................................................... 101
OWNER$$ ..................................................................................................................... 100
EQPLIF$$ ...................................................................................................................... 100
EQPNRJ$$ ....................................................................................................................... 99
EQPAIR$$ ....................................................................................................................... 99
EQPSOL$$ ....................................................................................................................... 99
EQPALM$$ ..................................................................................................................... 99
EQPWAT$$ ..................................................................................................................... 99
EQPGAR$$ ...................................................................................................................... 99
EQPBAS$$ ..................................................................................................................... 98
EQGARS$$ ................................................................................................................... 98
EQPWRAT$$ ................................................................................................................. 98
EQPAL$$ ....................................................................................................................... 98
EQPSOL$$ ...................................................................................................................... 98
EQPAIR$$ ....................................................................................................................... 97
EQPNRJ$$ ...................................................................................................................... 97
EQPLIF$$ ....................................................................................................................... 97
MOVYR$$ ...................................................................................................................... 97
OWNER$$ ....................................................................................................................... 97
OSUB$$ .......................................................................................................................... 96
REVAL$$ ....................................................................................................................... 96
RSUB$$ .......................................................................................................................... 96
REDUC$$ ....................................................................................................................... 96
RENT$$ .......................................................................................................................... 96
HEAT$$ .......................................................................................................................... 96
UTIL$$ .......................................................................................................................... 96
ELECTR$$ ...................................................................................................................... 96
FRENT$$ ....................................................................................................................... 96
FHEAT$$ ....................................................................................................................... 96
FUTIL$$ ....................................................................................................................... 96
NOREN$$ ...................................................................................................................... 96

Documentation $hgen: Household related status variables and generated variables$ FiD-Documentation

LABGROSS$ ..................................................................................................................... 73
IMPGROSS$ .................................................................................................................... 73
LABNET$ ....................................................................................................................... 74
IMPNET$ ....................................................................................................................... 74
FIELDS$ ....................................................................................................................... 74
DEGREE$ ....................................................................................................................... 76
TRAIN$ ......................................................................................................................... 78
TRAIN$ ......................................................................................................................... 80
TRAINC$ ....................................................................................................................... 80
TRAIND$ ....................................................................................................................... 81
FDT_F$ ......................................................................................................................... 81

Documentation $hgen: Household related status variables and generated variables$ FiD-Documentation

LABGROSS$ ..................................................................................................................... 73
IMPGROSS$ .................................................................................................................... 73
LABNET$ ....................................................................................................................... 74
IMPNET$ ....................................................................................................................... 74
FIELDS$ ....................................................................................................................... 74
DEGREE$ ....................................................................................................................... 76
TRAIN$ ......................................................................................................................... 78
TRAIN$ ......................................................................................................................... 80
TRAINC$ ....................................................................................................................... 80
TRAIND$ ....................................................................................................................... 81
FDT_F$ ......................................................................................................................... 81

List of Variables: .................................................................................................................. 85
CNSTYR$ ......................................................................................................................... 96
CONDIT$ ......................................................................................................................... 96
SIZES$ .......................................................................................................................... 96
FSIZE$ .......................................................................................................................... 96
ROOM$ .......................................................................................................................... 97
FROM$ .......................................................................................................................... 97
SEVAL$ .......................................................................................................................... 97
EQPKIT$ ......................................................................................................................... 97
EQPSSHW$ ..................................................................................................................... 98
EQPHEA$ ....................................................................................................................... 98
EQPFPHEA$ ................................................................................................................... 98
EQPTERS$ ...................................................................................................................... 98
EQPBA$ ........................................................................................................................ 98
EQPGRAS$ ................................................................................................................... 99
EQPWATS$ ................................................................................................................... 99
EQPALS$ ......................................................................................................................... 99
EQPSOLS$ ..................................................................................................................... 99
EQPARR$ ....................................................................................................................... 99
EQQNRJ$ ....................................................................................................................... 99
EQPLIFS$ .................................................................................................................... 100
MOVEYRI$ .................................................................................................................... 100
OWNER$ ......................................................................................................................... 100
OSUB$ .......................................................................................................................... 101
REVAL$ ......................................................................................................................... 101
RSUB$ .......................................................................................................................... 101
REDUC$ ....................................................................................................................... 101
RENT$ .......................................................................................................................... 102
HEAT$ .......................................................................................................................... 102
UTIL$ .......................................................................................................................... 102
ELECTR$ ...................................................................................................................... 102
FRENT$ ....................................................................................................................... 102
FHEAT$ ....................................................................................................................... 103
FUTIL$ ....................................................................................................................... 103
FELECTR$ ................................................................................................................... 103
NOREN$ ...................................................................................................................... 103
Variables in biobirth ........................................................................................................... 152
Information on data generation in biobirth ........................................................................ 149
General information ........................................................................................................... 149
Variables in biobirth........................................................................................................... 152
BIOYEAR ......................................................................................................................... 152
BIOAGE ......................................................................................................................... 152
BBSEX ........................................................................................................................... 152
SUMKIDS ...................................................................................................................... 152
BIOKIDS ........................................................................................................................ 153
NEWKIDS ..................................................................................................................... 153
KIDSOURCE[n] ............................................................................................................ 153
KIDPNR[n] .................................................................................................................... 153
KIDSEX[n] ................................................................................................................... 154
KIDYOB[n] .................................................................................................................. 154

Documentation bioage01-10 Files .............................................................. 108
Introduction ................................................................................................................... 110
Variables in bioage01-10 and bioage1l ................................................................. 111
Topics covered in the bioage files ........................................................................ 126
Data files and respondents ...................................................................................... 127
Number of observations ......................................................................................... 130
Generated Variables ................................................................................................. 130
AGE ............................................................................................................................. 130
BREASTFM ............................................................................................................... 130
PREBEG ...................................................................................................................... 131
PREEND ....................................................................................................................... 131
PREGY ......................................................................................................................... 131
PREGMO ...................................................................................................................... 132
SEX ............................................................................................................................... 132
SEXRESP .................................................................................................................... 132
CARE6 ......................................................................................................................... 132
CARE6H ....................................................................................................................... 134
CARE8 ........................................................................................................................ 134
CARE8H ....................................................................................................................... 134
CARE9H – CARE11H ................................................................................................. 134
CARE1H - CARE13H ................................................................................................. 134
BIOAGE ...................................................................................................................... 136
Different questionnaire versions .............................................................................. 137
bioage02 ..................................................................................................................... 137
bioage03 ..................................................................................................................... 137
bioage06 ..................................................................................................................... 139
bioage08 and bioage10 ............................................................................................ 139
bioage10 ..................................................................................................................... 142

Documentation biobirth ...................................................................................................... 147
General information ...................................................................................................... 149
Information on data generation in biobirth .................................................................... 149
Variables in biobirth ....................................................................................................... 152
BIOYEAR ...................................................................................................................... 152
BIOAGE ...................................................................................................................... 152
BBSEX ......................................................................................................................... 152
SUMKIDS ...................................................................................................................... 152
BIOKIDS ...................................................................................................................... 153
NEWKIDS ..................................................................................................................... 153
KIDSOURCE[n] ............................................................................................................ 153
KIDPNR[n] .................................................................................................................... 153
KIDSEX[n] ................................................................................................................... 154
KIDYOB[n] .................................................................................................................. 154
Documented variables:

- KIDMOB[n] .................................................. 154
- KIDMAR[n] .................................................. 154
- KIDLITO[n] .................................................. 154

Documentation *biomarsy* and *biocouply* .............................................. 156

General Information ................................................................. 158
Comparison with SOEP .............................................................. 158
*biocouply*: A yearly couple biography ........................................ 158
Variables in *biocouply* ............................................................... 160
- COUPID .................................................. 161
- SPELLNR .................................................. 161
- SPELLTYP .................................................. 161
- BEGINY .................................................. 161
- ENDY ..................................................... 161
- BEGIN ................................................... 161
- END ....................................................... 162
- PDEATH .................................................. 162
- DIVORCE ............................................... 162
- CENSOR .................................................. 162
- REMARK .................................................. 163
Sources of the couple history ...................................................... 163
Construction of couple history .................................................. 167
*biomarsy*: A yearly marital biography .......................................... 172
Variables in *biomarsy* ............................................................... 173
- SPELLNR .................................................. 173
- SPELLTYP .................................................. 173
- BEGINY .................................................. 173
- ENDY ..................................................... 173
- BEGIN ................................................... 173
- END ....................................................... 173
- CENSOR .................................................. 173
- REMARK .................................................. 174
Construction of marital histories .................................................. 174

Documentation *iskind* .............................................................. 175

General information ................................................................. 176
List of variables: ......................................................................... 176
- PSAMPLE .................................................. 177
- $WELLE .................................................. 177
- $KGJahr ................................................... 177
- $KGMon ................................................... 177
- $KSEX .................................................... 177
- $KStell .................................................... 178
- $KINHH .................................................. 179
- $HHGr .................................................... 179
- $KZahl .................................................... 179
- MOTHNO$ .............................................. 180
- FATHNOS ............................................. 181
- FATHPS .................................................. 181
- schltypES$ ............................................. 182
Appendix ............................................................................................................................... 182

Documentation *biojob* ........................................................................................................... 185

General information .................................................................................................................. 186

List of variables: ....................................................................................................................... 187

AGEFJOB .................................................................................................................................. 188
AGEINFO ................................................................................................................................. 188
NOJOB ....................................................................................................................................... 189
STILLFJ ...................................................................................................................................... 189
FULLTIME ................................................................................................................................. 190
OCCFJOB ................................................................................................................................... 190
FJBLUE ..................................................................................................................................... 190
FJSELF ...................................................................................................................................... 190
FJSEFSIZ .................................................................................................................................. 191
FJWHITE ................................................................................................................................... 191
FJCIVS ....................................................................................................................................... 191
STBA ......................................................................................................................................... 191
ISCO88 ..................................................................................................................................... 192
EGP .......................................................................................................................................... 192
ISEI ............................................................................................................................................ 192
MPS .......................................................................................................................................... 193
SIOPS ....................................................................................................................................... 193
CIVSFJ ...................................................................................................................................... 193
CURREMPL ............................................................................................................................. 193
YEARLAST ............................................................................................................................... 193
SCOPELJ ................................................................................................................................... 194
CIVILSLJ ................................................................................................................................... 194
NACEIJ ....................................................................................................................................... 194
OCCLJOB .................................................................................................................................. 195
LJBLUE ...................................................................................................................................... 195
LJSELF ...................................................................................................................................... 195
LJSEFSIZ ................................................................................................................................... 196
LJWHITE ................................................................................................................................... 196
LJCIVS ....................................................................................................................................... 196

Documentation *hhrf* and *phrf* ........................................................................................... 197

Hochrechnung in „Familien in Deutschland“ ........................................................................... 199
Gewichtungsansatz .................................................................................................................... 199
Stichproben in FiD .................................................................................................................... 201
Hochrechnung der Screening-Stichproben 2010 und 2011 .................................................... 202
Hochrechnung der Kohorten-Stichprobe ................................................................................. 209
Integration der FiD-Stichproben ............................................................................................. 212
Integration von SOEP und FiD ................................................................................................ 213
Querschnittsgewichte für Daten nach 2010 ............................................................................ 213
Längsschnittgewichte .............................................................................................................. 216
Nutzung der Hochrechnungsfaktoren ..................................................................................... 216
Anhang: Neue Gewichte 2010-2012 ab der Weitergabe FiDv3.1 ......................................... 218

Documentation *mipinc* and *mihinc* ................................................................................... 220

Introduction .............................................................................................................................. 222
## Methods

- Using multiple imputations .......................................................... 229

## Evaluation of imputations .......................................................... 228

## Using multiple imputations .......................................................... 229

### Variables in \$mipinc ................................................................. 231

<table>
<thead>
<tr>
<th>Variable</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>_MI</td>
<td>231</td>
</tr>
<tr>
<td>_MJ</td>
<td>231</td>
</tr>
<tr>
<td>PUNR$</td>
<td>231</td>
</tr>
<tr>
<td>$PNETINC</td>
<td>231</td>
</tr>
<tr>
<td>$PGROINC</td>
<td>231</td>
</tr>
<tr>
<td>$PMATBEN</td>
<td>232</td>
</tr>
<tr>
<td>$PALIMON</td>
<td>232</td>
</tr>
<tr>
<td>$PUEBEN</td>
<td>232</td>
</tr>
<tr>
<td>$PWIDOW</td>
<td>232</td>
</tr>
<tr>
<td>$PPENS</td>
<td>233</td>
</tr>
<tr>
<td>I_SPNETIC</td>
<td>233</td>
</tr>
<tr>
<td>I_F10PGROINC</td>
<td>233</td>
</tr>
<tr>
<td>I_$PMATBEN</td>
<td>233</td>
</tr>
<tr>
<td>I_$PALIMON</td>
<td>233</td>
</tr>
<tr>
<td>I_$PUEBEN</td>
<td>234</td>
</tr>
<tr>
<td>I_$PWIDOW</td>
<td>234</td>
</tr>
<tr>
<td>I_$PPENS</td>
<td>234</td>
</tr>
</tbody>
</table>

### Evaluation graphs for \$mipinc .............................................. 235

<table>
<thead>
<tr>
<th>Variable</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>F10PNETINC</td>
<td>235</td>
</tr>
<tr>
<td>F11PNETINC</td>
<td>236</td>
</tr>
<tr>
<td>F12PNETINC</td>
<td>237</td>
</tr>
<tr>
<td>F13PNETINC</td>
<td>238</td>
</tr>
<tr>
<td>F10PGROINC</td>
<td>239</td>
</tr>
<tr>
<td>F11PGROINC</td>
<td>240</td>
</tr>
<tr>
<td>F12PGROINC</td>
<td>241</td>
</tr>
<tr>
<td>F13PGROINC</td>
<td>242</td>
</tr>
</tbody>
</table>

### Variables in \$mihinc .......................................................... 243

<table>
<thead>
<tr>
<th>Variable</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>_MI</td>
<td>243</td>
</tr>
<tr>
<td>_MJ</td>
<td>243</td>
</tr>
<tr>
<td>$HHINC</td>
<td>243</td>
</tr>
<tr>
<td>$HCHDBEN</td>
<td>243</td>
</tr>
<tr>
<td>$HCHDADD</td>
<td>243</td>
</tr>
<tr>
<td>$HUEBEN2</td>
<td>244</td>
</tr>
<tr>
<td>$HCARBEN</td>
<td>244</td>
</tr>
<tr>
<td>$HHELEN</td>
<td>244</td>
</tr>
<tr>
<td>$HAGETRN</td>
<td>244</td>
</tr>
<tr>
<td>$HHOSBEN</td>
<td>244</td>
</tr>
<tr>
<td>$HRENT</td>
<td>245</td>
</tr>
<tr>
<td>$HUTIL</td>
<td>245</td>
</tr>
<tr>
<td>$HEHEAT</td>
<td>245</td>
</tr>
<tr>
<td>$SHELEC</td>
<td>245</td>
</tr>
<tr>
<td>$HCREDIT</td>
<td>245</td>
</tr>
<tr>
<td>I_$HHINC</td>
<td>246</td>
</tr>
<tr>
<td>I_$HCHDBEN</td>
<td>246</td>
</tr>
<tr>
<td>I_$HCHDADD</td>
<td>246</td>
</tr>
<tr>
<td>I_$HUEBEN2</td>
<td>246</td>
</tr>
</tbody>
</table>
SOEP Survey Paper 332

Shgen: Household related status variables and generated variables

List of variables:

I_SHCARBEN ................................................................. 247
I_SHHELBEN ................................................................. 247
I_SHAGETRN ................................................................. 247
I_SHHOSBEN ................................................................. 247
I_SHRENT ........................................................................ 247
I_SHUTIL ........................................................................ 248
I_SHHEAT ....................................................................... 248
I_SHHELEC ................................................................... 248
I_SHCREDIT ................................................................. 248

Evaluation graphs for $mihinc ........................................... 249
F10HHINC ...................................................................... 249
F11HHINC ..................................................................... 250
F12HHINC ..................................................................... 251
F13HHINC ..................................................................... 252
References ........................................................................ 253

Documentation pbiospe .................................................. 254
pbiospe ........................................................................... 255

Documentation pbrutto .................................................... 260
List of variables:

$GEBURT ................................................................. 261
$SEX ............................................................................ 262
$PNAT ................................................................. 262
$STISTAT ............................................................. 263
$BEFSTAT ............................................................ 263
$LINT ................................................................. 264
$UECKE ................................................................. 265
$ZUPAN ................................................................. 265
$PNRAKT ............................................................... 265
$STELL .................................................................... 265
$STELL (continued) .................................................. 266
$SPZUG ............................................................... 266
$PFORM ............................................................... 269
$PERG ................................................................. 269
$PERGZ ............................................................... 270
$PADER ............................................................... 270
$PADERQ ............................................................ 271
$SAUSZUGM .......................................................... 271
$SAUSZUGJ ............................................................ 271
$EINZUGM ............................................................. 272
$EINZUGJ ............................................................. 272
$ABWESM .............................................................. 272
$ABWESJ .............................................................. 272
$SPBIO ................................................................. 272
$DJ ................................................................. 272
$EINZUGM ............................................................. 273
$EINZUGJ ............................................................. 273
$SEX (SE1-SE6) ...................................................... 273

Documentation hbrutto .................................................. 275
List of Variables: ........................................................... 276
$HTYP ........................................................................ 277

FiD-Documentation
<table>
<thead>
<tr>
<th>Documentation</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>paradatal</td>
<td>288</td>
</tr>
<tr>
<td>General Information</td>
<td>290</td>
</tr>
<tr>
<td>Variables in paradatal</td>
<td>291</td>
</tr>
</tbody>
</table>

- HHNR .................................................. 291
- HHNRAKT ............................................. 291
- PERSNR .............................................. 291
- PERSNRK ............................................. 291
- SAMPLE1 ............................................ 291
- QSTNR ................................................ 292
- REQD .................................................. 292
- MISS ................................................... 292
- MODE ................................................... 293
- PROXY .................................................. 293
- INTID ................................................... 293
- DURA ................................................... 294
- MINT .................................................... 294
- DINT .................................................... 294
- DOW .................................................... 294

Table of frequencies ........................................................................................................... 296

Documentation bioage17 ............................................................................................... 297

General Information ........................................................................................................... 298
SOEP Survey Paper 332

Shgen: Household related status variables and generated variables

FiD-Documentation

Genesis and Target Population of the Youth Questionnaire .................................................. 298
Contents and Structure of the Data Set bioage17 ................................................................. 299
Special Features of Some Questions and Variables ............................................................... 300

Documentation $bioparen$ .................................................................................................. 308

Short summary .................................................................................................................... 309
How biography information has been collected in the FID .............................................. 309
How is bioparen generated? .............................................................................................. 309

List of Variables .................................................................................................................. 312
VNR / MNR ....................................................................................................................... 313
VGEBJ / MGEBJ ............................................................................................................... 313
VTODJ / MTODJ ............................................................................................................... 313
VAORT11 / MAORT11 ...................................................................................................... 314
VAORTAKT / MAORTAKT .............................................................................................. 314
VAORTUP / MAORTUP ................................................................................................. 315
VSBIL / MSBIL .................................................................................................................. 315
VBBIL / MBBIL .................................................................................................................. 316
VSINFO / MSINFO ......................................................................................................... 316
VBINFO / MBINFO ......................................................................................................... 316
VRELI / MRELI ................................................................................................................ 317
VNAT / MNAT ................................................................................................................... 317
VBSTELL / MBSTELL ...................................................................................................... 317
VBSINFO / MBSINFO ...................................................................................................... 318
VISCO88 / MISCO88 ....................................................................................................... 318
VISEI / MISEI ................................................................................................................... 318
VMPS / MMPMS .............................................................................................................. 318
VSIOPS / MSIOPS .......................................................................................................... 318
VEGP / MEGP ................................................................................................................... 319
VBLASS / MBKLAS ........................................................................................................ 319
ORTKINDH / ORTKIND1 .............................................................................................. 319
LIVING1 - LIVING8 ......................................................................................................... 319
VSTREIT / MSTREIT ...................................................................................................... 320
BIOYEAR ......................................................................................................................... 320
BIO .................................................................................................................................. 320
ALTER / VALTER / MALTER .......................................................................................... 321
VORIGIN / MORIGIN ...................................................................................................... 321
GESCHW ........................................................................................................................ 322
GESCHWUP ..................................................................................................................... 322
NUMS .............................................................................................................................. 322
NUMB .............................................................................................................................. 322
TWIN ............................................................................................................................... 323

Änderungen in den Versionen 1.1-4.0 der FiD Datenweitergaben ........................................ 324

1.1 .................................................................................................................................. 324
1.2 .................................................................................................................................. 326
2.0 .................................................................................................................................. 332
2.1 .................................................................................................................................. 335
3.0 .................................................................................................................................. 339

SOEP Survey Paper 332 94
3.1 ........................................................................................................................................... 342
4.0 ........................................................................................................................................... 348
**CNSTYR$$**

Variable label: “Year house was constructed”

Value labels:

- (1) before 1919
- (2) 1919 to 1948
- (3) 1949 to 1971
- (4) 1972 to 1980
- (5) 1981 to 1990
- (6) 1991 to 2000
- (7) 2001 or later
- (8) 2011 or later

Variable format: 1 digit integer

$\$$ - Survey Years: $\$$=10, 11, 12, 13

Comment: Classified statement of year the building was constructed in which a household lives at the time the survey was built.

**CONDIT$$**

Variable label: “Condition of building”

Value labels:

- (1) in good condition
- (2) partial renovation needed
- (3) major renovation needed
- (4) ready for demolition

Variable format: 1 digit integer

$\$$ - Survey Years: $\$$=10, 11, 12, 13

Comment: Respondent’s assessment of the condition of the building.

**SIZE$$**

Variable label: “Size of housing unit in square meters”

Variable format: 3-digit Integer

$\$$ - Survey Years: $\$$=10, 11, 12, 13

Comment: Size of housing unit could be missing due to non-response. However, the value is imputed for all missing values using the household multiple imputation procedure (see imputation documentation for details). The first implicate of these imputations is used to impute SIZE$$$. See FSIZE$$ for the imputation indicator. If, however, the household reported the size in any wave and never moved in between, the reported value is used instead of the imputation. This is not indicated by the imputation flag FSIZE$$. See documentation on imputations for more detail on the multiple imputations.

**FSIZE$$**

Variable label: “Imputation flag for size of housing unit”
**Value labels**

FSIZE$$
- (0) observed value
- (1) imputed value

**Variable format**

1-digit Integer

**$\$$ - Survey Years**

$\$$=10, 11, 12, 13

---

**ROOM$$**

**Variable label**

"Number of rooms larger than 6 square meters"

**Variable format**

2-digit Integer

**$\$$ - Survey Years**

$\$$=10, 11, 12, 13

**Comment:**

Number of rooms could be missing due to non-response. However, the value is imputed for all missing values using the household multiple imputation procedure (see imputation documentation for details). The first imputation of these imputations is used to impute ROOM$$$. See FROOM$$ for the imputation indicator. If, however, the household reported the number of rooms in any wave and never moved in between, the reported value is used instead of the imputation. This is not indicated by the imputation flag FROOM$$.

See documentation on imputations for more detail on the multiple imputations.

---

**FROOM$$**

**Variable label**

"Imputation flag for size of housing unit"

**Value labels**

FROOM$$
- (0) observed value
- (1) imputed value

**Variable format**

1-digit Integer

**$\$$ - Survey Years**

$\$$=10, 11, 12, 13

---

**SEVAL$$**

**Variable label**

"Adequacy of living space in the housing unit"

**Value label**

SEVAL$$
- (1) much too small
- (2) a bit too small
- (3) just right
- (4) a bit too large
- (5) much too large

**Variable format**

1-digit integer

**$\$$ - Survey Years**

$\$$=10, 11, 12, 13

**Comment:**

Assessment of adequacy by respondent (household head).

---

**EQPKIT$$**

**Variable label**

"Dwelling has kitchen"

**Value label**

EQPKIT$$
- (1) yes
(2) no
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13

**EQPSHW$$**

Variable label “Dwelling has indoor bath/shower”
Value label EQPSHW$$
(1) yes
(2) no
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13

**EQPHEA$$**

Variable label “Dwelling has central heating”
Value label EQPHEA$$
(1) yes
(2) no
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13

**EQPFHEA$$**

Variable label “Dwelling has floor heating”
Value label EQPHEA$$
(1) yes
(2) no
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13

Comment: This information has not been asked in the SOEP before 2010.

**EQPTER$$**

Variable label “Dwelling has balcony/terrace”
Value label EQPTER$$
(1) yes
(2) no
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13

**EQPBAS$$**

Variable label “Dwelling has basement”
Value label EQPBAS$$
(1) yes
(2) no
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Value labels</th>
<th>Format</th>
<th>Survey Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQPGAR$$$</td>
<td>&quot;Dwelling has garden&quot;</td>
<td>(1) yes, (2) no</td>
<td>1-digit integer</td>
<td>$$=10, 11, 12, 13</td>
</tr>
<tr>
<td>EQPWAT$$$</td>
<td>&quot;Dwelling has hot water/boiler&quot;</td>
<td>(1) yes, (2) no</td>
<td>1-digit integer</td>
<td>$$=10, 11, 12, 13</td>
</tr>
<tr>
<td>EQPALM$$$</td>
<td>&quot;Dwelling has alarm device&quot;</td>
<td>(1) yes, (2) no</td>
<td>1-digit integer</td>
<td>$$=10, 11, 12, 13</td>
</tr>
<tr>
<td>EQPSOL$$$</td>
<td>&quot;Dwelling has solar collector&quot;</td>
<td>(1) yes, (2) no</td>
<td>1-digit integer</td>
<td>$$=10, 11, 12, 13</td>
</tr>
<tr>
<td>EQPAIR$$$</td>
<td>&quot;Dwelling has air conditioning&quot;</td>
<td>(1) yes, (2) no</td>
<td>1-digit integer</td>
<td>$$=10, 11, 12, 13</td>
</tr>
<tr>
<td>EQPNRJ$$$</td>
<td>&quot;Dwelling has other alternative energy source&quot;</td>
<td>(1) yes, (2) no</td>
<td>1-digit integer</td>
<td>$$=10, 11, 12, 13</td>
</tr>
</tbody>
</table>
Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13

**EQPLIF$$

Variable label “Dwelling has lift”
Value label EQPLIF$$
(1) yes
(2) no

Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13

**MOVEYR$$

Variable label “Year moved into dwelling”
Variable format 4-digit integer
$$ - Survey Years $$=10, 11, 12, 13

Comment: The information given in MOVEYR$$ was collected on an individual level in the biographical section of the person questionnaire. It was then collapsed on the household level, such that the individual living in the household the longest provided the year of moving in. Beginning with survey year 2011 for households at their old address, data are carried forward. For new households and for old households that have moved, the variable is based on newly collected data. In case the information is missing and an old household has moved that year or the previous year, MOVEYR$$ will take the value of the year of the respective wave. The carrying forward of data entails the possibility that the year of moving into the new dwelling may lie before the year of birth of the oldest household member.

**OWNER$$

Variable label “Tenant or owner of dwelling”
Value label OWNER$$
(1) Owner
(2) Main tenant
(3) Subtenant
(4) Tenant
(5) Institutional resident (Nursing home etc.)

Variable format 1-digit integer
$$ - Survey Years $$=10, 11, 12, 13

Comment: Code 4 is used if the original variable is coded as -1 (“No answer”) but if at least one answer that is specific to tenants, was given. If the original variable is “-1” but at least one answer specific to owners was given, then the “-1” is recoded to “1” (“Owner”). Code 4 is also used if a change in ownership (from owner to tenant) has taken place, but no original information for OWNER$$ was given.
OSUBSS$  
Variable label  “Received government housing subsidies last year”  
Value label  OSUBSS$  
(1) yes  
(2) no  
Variable format  1-digit integer  
$\$ - Survey Years  $\$=10, 11, 12, 13  
Comment:  This information is given for owners-occupiers only. OSUBSS$ contains information on cash housing subsidies received from the government during the year prior to the interview. This information is based solely on the respondent’s report, no other information is used. If missing, the information is not carried forward.

REVALSS$  
Variable label  “Evaluation of rent paid”  
Value label  REVALSS$  
(1) very inexpensive  
(2) inexpensive  
(3) reasonable  
(4) slightly too expensive  
(5) too expensive  
Variable format  1-digit integer  
$\$ - Survey Years  $\$=10, 11, 12, 13  
Comment:  This information is given for tenant-occupiers only. It contains the subjective assessment by the respondent (household head). The corresponding information from the previous year is not carried forward longitudinally due to the possibility of changes in rent and income, residential moves, and change in the person responding.

RSUBSS$  
Variable label  “Government subsidized housing”  
Value label  RSUBSS$  
(1) yes, with subsidy  
(2) yes, with expired subsidy  
(3) no  
Variable format  1-digit integer  
$\$ - Survey Years  $\$=10, 11, 12, 13  
Comment:  This information is given for tenant-occupiers only. Beginning with survey year 2011 information will be carried forward from previous years for immobile households. The code “2” was introduced to indicate expired subsidization for subsequent waves.

REDUCSS$  
Variable label  “Rent-reduced dwelling”
### Value label
- REDUC$$
  1. yes
  2. no

### Variable format
- 1-digit integer

### $-$ Survey Years
- $=$10, 11, 12, 13

### Comment:
This information is given for tenant-occupiers only. Beginning with survey year 2011, information will be carried forward from the previous years for old households residing at their old address; for new households and for old households that have moved, newly collected data will be used.

#### RENT$$
- Variable label: “Amount of rent, no heating, with utilities (EURO)”
- Variable format: 4-digit integer
- $-$ Survey Years: $=$10, 11, 12, 13

#### HEAT$$
- Variable label: “Amount of monthly heating costs (EURO)”
- Variable format: 4-digit integer
- $-$ Survey Years: $=$10, 11, 12, 13

#### UTIL$$
- Variable label: “Amount of monthly utility costs (EURO)”
- Variable format: 4-digit integer
- $-$ Survey Years: $=$10, 11, 12, 13

#### ELECTR$$
- Variable label: “Amount of monthly electricity costs (EURO)”
- Variable format: 4-digit integer
- $-$ Survey Years: $=$10, 11, 12, 13

### Comment for RENT$$, HEAT$$, UTIL$$, ELECTR$$
All four variables are only provided for tenants-occupants who pay rent (NORENT$$ not equal “1”). In the questionnaire, the information of all four variables is asked separately for all tenants. Respondents report their monthly payments for rent, heating, electricity and utilities. For heating the respondents further specify whether it is included in rent or not. RENT$$ is then generated depending on the answers to the inclusion question. If included, then the amount of heating costs is subtracted from the answer given in to the rent question.

If any of the above values is missing, they are part of the household multiple imputation procedure (see imputation documentation for details). Although we recommend using the multiple imputations to fully use their distributional properties, as elsewhere in Shgen, we use the first of five implicants for those users who do not wish to deal with multiple imputations. All calculations regarding RENT$$ are done with the imputed first implicate as well. The variables FRENT$$, FHEAT$$, FUTIL$$, FELECTR$$ contain the respective imputation flags.
**FRENT$$**

**Variable label**  
“Imputation flag gross rent”

**Value labels**  
FSIZE$$
- (0) observed value
- (1) imputed value

**Variable format**  
1-digit Integer

**$$ - Survey Years**  
$$=10, 11, 12, 13

**Comment:**
This flag captures whether any component of the rent calculation was imputed. As RENT$$ is a combination of the rental value, heating costs and possibly the utility costs, the number of values including any imputations is larger than the number of imputed rents. The variable i._$hrent in the file *Smithinc* depicts how many of the actual rental values have been imputed.

**FHEAT$$**

**Variable label**  
“Imputation flag for heating costs”

**Value labels**  
FHEAT$$
- (0) observed value
- (1) imputed value

**Variable format**  
1-digit Integer

**$$ - Survey Years**  
$$=10, 11, 12, 13

**FUTIL$$**

**Variable label**  
“Imputation flag for utility costs”

**Value labels**  
FUTIL$$
- (0) observed value
- (1) imputed value

**Variable format**  
1-digit Integer

**$$ - Survey Years**  
$$=10, 11, 12, 13

**FELECTR$$**

**Variable label**  
“Imputation flag for electricity costs”

**Value labels**  
FELECTR$$
- (0) observed value
- (1) imputed value

**Variable format**  
1-digit Integer

**$$ - Survey Years**  
$$=10, 11, 12, 13

**NORENT$$**

**Variable label**  
“Pays no rent”

**Value label**  
NORENT$$
- (1) Does not pay rent

**Variable format**  
1-digit Integer

**$$ - Survey Years**  
$$=10, 11, 12, 13
Comment: Coded as “1” if the household does not pay rent; e.g., if living space is provided by relatives at no cost. Note that in these cases, the information on gross cold rent (RENT$$), heating costs (HEAT$$), utilities (UTIL$$) and electricity (ELECTR$$) is coded “-2” (“not applicable”). Information will not be carried forward.

**TYP1HH$$**

Variable label “Household typology (1-digit)”

Value label TYP1HH$$

(1) 1-person HH
(2) Childless couple
(3) Single parent
(4) Couple with children <= 16 yrs.
(5) Couple with children > 16 yrs.
(6) Couple with children <= 16 and > 16 yrs
(7) Multiple generations HH
(8) Other combination

Variable format 1-digit integer

$$ - Survey Years $$= 10, 11, 12, 13

**TYP2HH$$**

Variable label “Household typology (2-digit)”

Value label TYP2HH$$

(11) 1-P-HH Man < 35
(12) 1-P-HH Man 35-<60
(13) 1-P-HH Man >=60
(14) 1-P-HH Woman < 35
(15) 1-P-HH Woman 35-<60
(16) 1-P-HH Woman >=60
(21) Childless couple
(31) Single parent + 1 child
(32) Single parent + 2 or more children
(33) Single parent + 1 EK.
(34) Single parent + 2 or more EK.
(35) Single parent + 2 (E)K.
(36) Single parent + 3 or more(E)K.
(41) Couple + 1 child
(42) Couple + 2 children
(43) Couple + 3 or more children
(51) Couple + 1 EK.
(52) Couple + 2 EK.
(53) Couple + 3 or more EK.
(61) Couple + 2 (E)K.
(62) Couple + 3 or more (E)K.
(71) 3-generation HH
(72) 4-generation HH
(73) Grandparents-Children HH
(81) Other combination without children
(82) Other combination + 1 or more children

Variable format  2-digit integer
$$ - Survey Years $$=10, 11, 12, 13

Comment: Generated variable created by combining the relationships of all persons living in the household to the head of household (Variable $STELL$ in the file $PBRUTTO$) at the time of the survey. $TYP1HH$ is an aggregation of $TYP2HH$ (first column of the two-digit code). Single households are differentiated in $TYP2HH$ according to both gender and age.

Legend:
- $K =$ children up to the age of 16;
- $EK =$ adult children age 17 and older;
- $(E)K =$ children both below and above age 16;
- $1-P-HH =$ one-person households.

$HINC$$

Variable label  “Monthly net household income (EURO)”
Variable format  5-digit integer
$$ - Survey Years $$=10, 11, 12, 13

Comment: This variable contains the current monthly net household income asked for in the household questionnaire, always provided in Euros. Income is reported by the respondent (head of household).

$AHINC$$

Variable label  “Adjusted monthly net household income (EURO)”
Variable format  5-digit integer
$$ - Survey Years $$=10, 11, 12, 13

Comment on $AHINC$$
This variable is based on the current monthly net household income asked for directly in the household questionnaire (“screener”). Since everyone in FiD over the age of 16 is also interviewed personally, this income can be calculated based on the current individual monthly incomes of all household members. Possible underestimation in “screener” can thus be assessed and corrected. However, in the case of item-nonresponse in the original screener, the procedure is only used for households surveyed completely, without item-non-response on the variables in question.

For personal income, we use monthly net income (from dependent employment and self-employment), extra earnings, pensions, widow’s pensions, unemployment benefits or relief, maintenance payments, early retirement payments, maternity benefits, BaFoeG (state higher education grants), military or civil service pay, compulsory child support, as well as other forms of support from the $P$ files.

Civil servants’ pension income is taxed at the flat rate of 20% and multiple entries on the use of employment office services are corrected for by calculating a median value.
We add all the individual incomes of all interviewed household members, also adding to this sum all income from the household context (housing subsidies, child benefits, welfare and home nursing subsidies, social assistance, unemployment Benefit II or Social Benefit ("ALG II")).

When no answer was provided on net household income, we use the net household income calculated as described above, under the condition that all household members gave valid answers.

If the net household income generated in this way is higher than the household income stated in the questionnaire, we correct the value upwards. If the generated household income is lower, we stay with the value originally stated.

When no answer was given for the different components of income, we set the value of the particular component to zero. An overview of the different components (excluding net monthly income, which is available each year) is provided in the table below, in which “x” indicates that the particular component was taken into account in that particular year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Additional earnings</th>
<th>Old-age pensions</th>
<th>Widow's pensions</th>
<th>Unemployment benefits</th>
<th>Unemployment relief</th>
<th>Unemployment payments while in higher education</th>
<th>Maintenance payments</th>
<th>BaFög (state higher education grants)</th>
<th>Military or civil service pay</th>
<th>Compulsory child support</th>
<th>Support payments</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2011</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2012</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2013</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

**I_HINC$$**

Variable label: “Imputed monthly net household income (EURO)”

Variable format: 5-digit integer

$$ - Survey Years $$=10, 11, 12, 13

Comment on I_HINC$$

FiD uses a household multiple imputation procedure which jointly imputes all necessary values for the household. See the documentation on multiple imputations for detailed information. The variables I1HINC$$ to I5HINC$$ are the result of these multiple imputations and are provided here in the wide format (i.e. the number of observations per household is not changed). Complete imputation results are provided in the dataset $mihinc$. The imputation flag FHINC$$ identifies households with imputed incomes.

Analyzing multiply imputed data

For analyzing multiple imputed data, you do not necessarily need special methods, however such tools exist and simplify the use of multiply imputed data. Below a short overview of some useful tools for various statistical packages is given. These tools estimate the parameters of a regression model by combining the estimates across the several replicates of imputation. Point estimates from multiple imputations are then the arithmetic mean of the several point estimates obtained from analysis on each imputed data. Standard errors are obtained by
combining the average of the squared standard errors of the several (m) estimates with the within- and between-imputation variance.

- Since Stata™ version 11.0, there is a complete mi module available with various useful tools for analyzing multiple imputed data, but also for imputation itself.
- Within SAS, PROC MIANALYZE combines the results of analyses on the data sets.
- IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module regress to perform multiple imputation analysis.

**FHINC$$**

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Imputation flag for household income”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value labels</td>
<td>FHINC$$</td>
</tr>
<tr>
<td>(0)</td>
<td>observed value</td>
</tr>
<tr>
<td>(1)</td>
<td>imputed value</td>
</tr>
<tr>
<td>Variable format</td>
<td>1-digit Integer</td>
</tr>
</tbody>
</table>

**NUTS$$**

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“NUTS-systematic 1 (Federal State)”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value label</td>
<td>NUTS$$</td>
</tr>
<tr>
<td>(1)</td>
<td>Baden-Wuerttemberg</td>
</tr>
<tr>
<td>(2)</td>
<td>Bavaria</td>
</tr>
<tr>
<td>(3)</td>
<td>Berlin</td>
</tr>
<tr>
<td>(4)</td>
<td>Brandenburg</td>
</tr>
<tr>
<td>(5)</td>
<td>Bremen</td>
</tr>
<tr>
<td>(6)</td>
<td>Hamburg</td>
</tr>
<tr>
<td>(7)</td>
<td>Hesse</td>
</tr>
<tr>
<td>(8)</td>
<td>Mecklenburg-Western Pomera</td>
</tr>
<tr>
<td>(9)</td>
<td>Lower Saxony</td>
</tr>
<tr>
<td>(10)</td>
<td>North Rhine-Westphalia</td>
</tr>
<tr>
<td>(11)</td>
<td>Rhineland-Palatinate</td>
</tr>
<tr>
<td>(12)</td>
<td>Saarland</td>
</tr>
<tr>
<td>(13)</td>
<td>Saxony</td>
</tr>
<tr>
<td>(14)</td>
<td>Saxony-Anhalt</td>
</tr>
<tr>
<td>(15)</td>
<td>Schleswig-Holstein</td>
</tr>
<tr>
<td>(16)</td>
<td>Thuringia</td>
</tr>
</tbody>
</table>

**NUTS$$**

<table>
<thead>
<tr>
<th>Variable format</th>
<th>2-digit integer</th>
</tr>
</thead>
<tbody>
<tr>
<td>$$ - Survey Years</td>
<td>$$=10, 11, 12, 13</td>
</tr>
</tbody>
</table>

Comment: FHINC$$ is a dummy variable indicating whether an observation was missing on HINC$$ and was therefore imputed or not.

Comment: The NUTS$$ variable provides the basic classification of the German “Bundesländer”.
Documentation bioage01-10 Files

Generated variables from the Parent-Questionnaires

Alexander Raith, Nina Scherner, Malisa Zobel, Moritz Mannschreck, Linda Wittbrodt and Mathis Fräßdorf (geb. Schröder)
Table of Contents

Generated variables from the Parent-Questionnaires .......................................................... 108
Introduction .......................................................................................................................... 110
Variables in bioage01-10 and bioagel ................................................................................. 111
Topics covered in the bioage files ...................................................................................... 126
Data files and respondents .................................................................................................. 127
Number of observations ...................................................................................................... 130
Generated Variables ............................................................................................................ 130
Different questionnaire versions .......................................................................................... 137
  bioage02 ............................................................................................................................. 137
  bioage03 ............................................................................................................................. 137
  bioage06 ............................................................................................................................. 139
  bioage08 and bioage10 ....................................................................................................... 139
  bioage10 ............................................................................................................................. 142

List of Tables and Figures

Table 1: Overview variables in bioage files in FiD and SOEP ...................................................... 112
Table 2: Overview of topics .................................................................................................. 126
Table 3: Overview questionnaire names and corresponding age group ................................. 128
Table 4: Overview number of observations in bioage08 p1 and p2 ........................................ 129
Table 5: Overview number of observations in bioage10 p1 and p2 ........................................ 129
Table 6: Overview number of observations per wave .......................................................... 130
Table 7: Changes in question ordering, retrospective and non-retrospective birth questions ... 137
Table 8: Additions to questions and changes in order of questions in parent questionnaire 2 . 138
Table 9: bioage08/bioage10 differences in SCOLON across samples .................................... 140
Figure 1: Screenshot Parent Questionnaire 5 Screening Sample 2010 ...................................... 141
Figure 2: Screenshot Parent Questionnaire 5 Cohort Sample 2010 and 2011 sample ............... 141
Table 10: Recoding scheme for SDQ questions in bioage10 (BEHAV) .................................... 142
Table 11: Frequencies of original and recoded variables BEHAV1-BEHAV25 ......................... 143
**Introduction**

The *bioage* data files are generated using information collected in the “Parent-Questionnaires”. There are six different “Parent-Questionnaires” (for an overview see table 3) and accordingly six different *bioage* data files with the data file specific suffix indicating the age group of the children (e.g. children in *bioage01* are at most 23 months old, those in *bioage02* are between 24 and 35 months old, etc.).

The aim of the “Parent-Questionnaires” is to follow and observe future generations of the population in *Familien in Deutschland* (FiD) and collect all information in age specific files, even though the data come from different survey years. As we try to make this process as comprehensive and gap-free as possible, we start following and documenting the development of children in FiD households from birth onwards. Consequently, *bioage* data files contain information regarding child birth, pregnancy, child’s health, child care and further aspects related to raising children (for an overview of subjects covered in *bioage01-10*, see table 2).

Since many questions overlap, all *bioage* data files are documented within this chapter, rather than dedicating each *bioage* data file a single documentation file. By doing so, we are able to provide an overview of all variables covered in the *bioage* data files.

The *bioage* files in FiD have their conceptual match in the *bioage* files known from the SOEP – however, the FiD parent questionnaires cover slightly more subjects than those in the SOEP (where they are named “MuKi”-files, for “Mother-Child” questionnaires). A direct comparison of FiD and SOEP content is provided in table 1. Also, FiD covers one age group not captured in the SOEP so far – the 1-2 year-olds. Starting in the data collection of 2012 (wave BC), the SOEP also collects information about 9-10 year olds. However, data about the children are only collected from one parent, such that less information is available per child.

The rules used to generate the variables from the questionnaires are consistent over the various *bioage* data files. For the most part, the variables in the *bioage* files are simply renamed from their respective parent-questionnaires. This ensures identity over the years, even if different questionnaires (in terms of question ordering or content) are used. The renaming also allows the important comparison across the different age groups, as variable names are identical (up to their prefix), if their contents match, see table 1. A few variables are generated by combining the information from two or more variables. The ‘common origin’ (i.e. the common question) of a range of variables is indicated by using the same stem

---

6 You will find two *bioage08* and two *bioage10* data files, with suffix p1 and p2. We refer to those as one dataset here, respectively, as the variable content is identical. See below for the specifics of *bioage08* and *bioage10*. 
word (e.g. “MVMN”), followed by a number (e.g. “MVMN2”). In some cases those slight variations match the difference between questionnaires for younger and older children, in which case the letter “y” for younger (e.g. CHAR6Y).\footnote{Note that since FiD v2.0 from March 2012 we do not name variables, which are similar over the datasets, with additional letters anymore. Variable names such as “MOV5” and “MOV5A” did not prove to be helpful for users, but rather confusing.}

Additionally, this documentation provides detailed information on variables that were generated using information from other data files (such as the p-files). You can find an overview of the specific variables and the rules by which they were generated in section 6 of this chapter.

Note that while parents (see section 4 on specifics) are the respondents, data are organized according to the never-changing personal identification number of the child (PERSNR). In case of \textit{bioage08} and \textit{bioage10} this leads to two possible observations for each child (one from the ‘father respondent’ and one from the ‘mother respondent’), stored in two separate data files. As such, you will only find one observation per PERSNR in any \textit{bioage} data file. The only exception to this rule is the dataset \textit{bioagel}, introduced with FiD v2.1, which contains the combination of all \textit{bioage} files.

\section*{Variables in \textit{bioage01-10} and \textit{bioagel}}

The following table provides an overview of all variables contained in the FiD \textit{bioage} files in alphabetical order, as well as a comparison with the structure of variables in the SOEP. The variables are shown without their prefix, to increase readability and allow the comparison over the different files.\footnote{Note that in FiDv2.0 from March 2012 as well as in FiDv3.1 from September 2013, several variables in the bioage files have been renamed. In addition, several variable labels have been shortened or changed. In case of questions, please contact the FiD-Team.}

\begin{table}
\centering
\begin{tabular}{|l|l|}
\hline
\textbf{Variable} & \textbf{Description} \\
\hline
\textit{bioagel} & Contains the combination of all \textit{bioage} files. \footnote{Cells with an “(x)” indicate that the variable is included in the SOEP, but either the variable name is different or the information stored does not match. Such differences will be harmonized within the SOEP later. Users should refer to the variable labels when combining these variables over FiD and SOEP.} \\
\hline
\end{tabular}
\end{table}
Table 1: Overview variables in *bioage* files in FiD and SOEP

<table>
<thead>
<tr>
<th>Name</th>
<th>Label</th>
<th>Bioage01</th>
<th>Bioage02</th>
<th>Bioage03</th>
<th>Bioage06</th>
<th>Bioage08</th>
<th>Bioage10</th>
<th>SOEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>hhnrakt</td>
<td>Current Wave HH Number</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>sample1</td>
<td>Subsample</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>persnr</td>
<td>Never Changing Person ID of child</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>persnrresp</td>
<td>Never Changing Person ID of Mother/Respondent</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>syear</td>
<td>Survey year</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>intid</td>
<td>Interviewer ID</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

**VARIABLE PREFIX IS FILE SPECIFIC**

<table>
<thead>
<tr>
<th>Name</th>
<th>Label</th>
<th>b01</th>
<th>b02</th>
<th>b03</th>
<th>b06</th>
<th>b08</th>
<th>b10</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>actcare1</td>
<td>Seeking ch. care: Waitinglist</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>actcare2</td>
<td>Seeking ch. care: In contact w. nanny/day care</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>actcare3</td>
<td>Seeking ch. care: In contact w. welfare office</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>actcare4</td>
<td>Seeking ch. care: Other</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ1</td>
<td>Last 14 ds: Singing children songs</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ2</td>
<td>Last 14 ds: Taking walks outdoors</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ3</td>
<td>Last 14 ds: Drawing or doing arts/crafts</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ4</td>
<td>Last 14 ds: Reading/telling stories (German)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ5</td>
<td>Last 14 ds: Looking at picture books</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ6</td>
<td>Last 14 ds: Going to playground</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ7</td>
<td>Last 14 ds: Visiting oth. families w. children</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ8</td>
<td>Last 14 ds: Going shopping w. child</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ9</td>
<td>Last 14 ds: Watching TV/videos/DVD w. child</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ10</td>
<td>Last 14 ds: Reading/telling stories (not German)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ11</td>
<td>Last 14 ds: Actions outside (walks o. similar)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ12</td>
<td>Last 14 ds: Card games or similar</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ13</td>
<td>Last 14 ds: Playing computer game w. child</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>activ14</td>
<td>Last 14 ds: Playing computer game w. child</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>adpccare1</td>
<td>Probl. adapting to day care center: Contact with other children</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>adpccare2</td>
<td>Probl. adapting to day care center: Relationship to child care workers</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>adpccare3</td>
<td>Probl. adapting to day care center: Separation from home</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>adpccare4</td>
<td>Probl. adapting to day care center: Adjustment to fixed daily rhythm</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>adpscl1</td>
<td>Probl. adapting to primary school: Contact with other children</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Name</td>
<td>Label</td>
<td>Bioage01</td>
<td>Bioage02</td>
<td>Bioage03</td>
<td>Bioage06</td>
<td>Bioage08</td>
<td>Bioage10</td>
<td>SOEP</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>adpscl2</td>
<td>Probl. adapting to primary school: Contact with teachers</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adpscl3</td>
<td>Probl. adapting to primary school: Want of familiar environment</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adpscl4</td>
<td>Probl. adapting to primary school: Abiding by school's rules</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adpscl5</td>
<td>Probl. adapting to primary school: School performance</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>Age of child in months</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>allowpm</td>
<td>Child allowance per month</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>allowpw</td>
<td>Child allowance per week</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>behav1</td>
<td>Ch. is considerate</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>behav2</td>
<td>Ch. is anxious, overactive, cannot sit quietly</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>behav3</td>
<td>Ch. often complains a. headache/stomach sickness</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>behav4</td>
<td>Ch. likes to share with other children</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>(x)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>behav5</td>
<td>Ch. has rage attacks, is choleric</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>(x)</td>
<td></td>
</tr>
<tr>
<td>behav6</td>
<td>Ch. is a loner, plays mostly alone</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>behav7</td>
<td>Ch. is compliant, does what adults ask for</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>behav8</td>
<td>Ch. has a lot of worries, is depressed</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>behav9</td>
<td>Ch. is helpful if anothers are injured/ill/unhappy</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>behav10</td>
<td>Ch. is fidgety</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>behav11</td>
<td>Ch. has at least one good friend</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>behav12</td>
<td>Ch. quarrels with other children, mobs them</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>behav13</td>
<td>Ch. is often unhappy or depressed, weeps often</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>behav14</td>
<td>Ch. is popular with other children</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>behav15</td>
<td>Ch. is easily distracted, unconcentrated</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>behav16</td>
<td>Ch. is nervous, clinging in new situations</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>behav17</td>
<td>Ch. is nice to younger children</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>behav18</td>
<td>Ch. often lies or cheats</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>behav19</td>
<td>Ch. is mobbed by others</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>behav20</td>
<td>Ch. offers its help voluntary</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>behav21</td>
<td>Ch. takes things w/o asking for permission</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>behav22</td>
<td>Ch. gets along better w. adults than w. children</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>behav23</td>
<td>Ch. has many fears, is easily afraid</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>behav24</td>
<td>Ch. finishes tasks, can concentrate for long time</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bepar1</td>
<td>Be parents: sacrifice own wishes</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Label</td>
<td>Bioage01</td>
<td>Bioage02</td>
<td>Bioage03</td>
<td>Bioage06</td>
<td>Bioage08</td>
<td>Bioage10</td>
<td>SOEP</td>
</tr>
<tr>
<td>--------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bepar2</td>
<td>Be parents: disobedience is aimed to bother me</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bepar3</td>
<td>Be parents: child upbringing is all that is left</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>bepar4</td>
<td>Be parents: wish was less captured in responsibility</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bepar5</td>
<td>Be parents: problems in upbringing is fault of child</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bepar6</td>
<td>Be parents: whn with child, nothing is better</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bepar7</td>
<td>Be parents: would bear anything for child</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bepar8</td>
<td>Be parents: child misbehavior is intentional</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bepar9</td>
<td>Be parents: often put everything aside to support child</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bepar10</td>
<td>Be parents: look foward to spending time w. child</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>biochild</td>
<td>Biological child</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>biopar</td>
<td>Respondent is mother/father</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>birthm</td>
<td>Month of birth, child</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>birthy</td>
<td>Year of birth, child</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>birthpw</td>
<td>Pregnancy week at birth</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>breastf</td>
<td>Breast-feeding</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>breastfc</td>
<td>Currently breast-feeding</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>breastfm</td>
<td>Breast-feeding time in months</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>care1</td>
<td>Cared for by partner</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>care1h</td>
<td>Cared for by partner (hrs/wk)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>care2</td>
<td>Cared for by moth/fath.(if not in household)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>care2h</td>
<td>Cared for by moth/fath.(if not in household)(hrs/wk)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>care3</td>
<td>Cared for by grandparents</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>care3h</td>
<td>Cared for by grandparents (hrs/wk)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>care4</td>
<td>Cared for by older siblings</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>care4h</td>
<td>Cared for by older siblings (hrs/wk)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>care5</td>
<td>Cared for by other relatives</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>care5h</td>
<td>Cared for by other relatives (hrs/wk)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>care6</td>
<td>Cared for by nanny/day care (not in-house)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>care6h</td>
<td>Cared for by babysitter/nanny (hrs/wk)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>care7</td>
<td>Cared for by babysitter/nanny (in-house)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>care7h</td>
<td>Cared for by babysitter/nanny (in-house)(hrs/wk)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>care8</td>
<td>Cared for in crib/day care center</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Label</td>
<td>Bioage01</td>
<td>Bioage02</td>
<td>Bioage03</td>
<td>Bioage06</td>
<td>Bioage08</td>
<td>Bioage10</td>
<td>SOEP</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>care8h</td>
<td>Cared for in crib/day care center (hrs/wk)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>care9</td>
<td>Cared for at school</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>care9h</td>
<td>Cared for at school (hrs/wk)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>care10</td>
<td>Cared for in after-school hoard</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>care10h</td>
<td>Cared for in after-school hoard (hrs/wk)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>care11</td>
<td>Cared for in social institution, center</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>care11h</td>
<td>Cared for in social institution, center (hrs/wk)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>care12</td>
<td>Cared for by others</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>care12h</td>
<td>Cared for by others (hrs/wk)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>care13</td>
<td>No person or institution</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>care13h</td>
<td>Cared for in social institution, center (hrs/wk)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>careaft</td>
<td>Child in school: Problems with afternoon child care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>careaftw</td>
<td>Child n.i. school: Worries about afternoon child care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>caremn</td>
<td>Start current child care, month</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>caretyp</td>
<td>Type of child care, if any</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>careyr</td>
<td>Start current child care, year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ccare</td>
<td>Child care: day care center/nanny (y/n)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ccarec</td>
<td>Never changed child care institution</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ccarecn</td>
<td>Number of changes of child care institution</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ccarefr</td>
<td>Child care: freq of day care center/nanny</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>change1</td>
<td>Life circumstances have greatly changed</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>change2</td>
<td>Child provides happiness and joy</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>change3</td>
<td>Often close to running out of strength</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>change4</td>
<td>Very satisfied with the role of (being) a mother</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>change5</td>
<td>Often unable to cope with (new) tasks/responsibil.</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>change6</td>
<td>Have made new contacts through the child</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>change7</td>
<td>Suffer f. being limited to role of mother/father</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>change8</td>
<td>Important to provide the child w. much affection</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>char1</td>
<td>Ch. is communicative vs. quiet</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>char2</td>
<td>Ch. is untidy vs. tidy</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>char3</td>
<td>Ch. is sweet-tempered vs. short-tempered</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>char4</td>
<td>Ch. is not interested vs. eager to learn</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>char5</td>
<td>Ch. has self-confidence vs. is insecure</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
</tbody>
</table>
### bioage01-10: Generated variables from the parent questionnaires

<table>
<thead>
<tr>
<th>Name</th>
<th>Label</th>
<th>Bioage01</th>
<th>Bioage02</th>
<th>Bioage03</th>
<th>Bioage06</th>
<th>Bioage08</th>
<th>Bioage10</th>
<th>SOEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>char6</td>
<td>Ch. is withdrawn vs. sociable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>(x)</td>
</tr>
<tr>
<td>char6y</td>
<td>Child tends to be timid vs. sociable</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>char7</td>
<td>Ch. is concentrated vs. easily distracted</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>char7y</td>
<td>Child tends to be concentrated vs. distracted</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>char8</td>
<td>Ch. is defiant vs. obedient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>char8y</td>
<td>Child tends to be defiant vs. obedient</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>char9</td>
<td>Ch. comprehends fast vs. needs more time</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>char9y</td>
<td>Child tends to comprehend fast vs. needs more time</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>char10</td>
<td>Ch. is afraid vs. unafraid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>chhealth</td>
<td>Child health status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>circum</td>
<td>Head circumference of the child at birth in cm</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clofrie</td>
<td>Number of close friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conscho1</td>
<td>Contact w. school: Reg. partic. at parent-teacher evening</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conscho2</td>
<td>Contact w. school: Reg. parent-teacher meeting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conscho3</td>
<td>Contact w. school: See teacher individually</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conscho4</td>
<td>Contact w. school: Parent representative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conscho7</td>
<td>Contact w. school: None of these</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>curscol1</td>
<td>Cur. school: Primary School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>curscol2</td>
<td>Cur. school: Special educational concepts</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>curscol3</td>
<td>Cur. school: Special needs school</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>curscol4</td>
<td>Cur. school: 'Hauptschule'</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>curscol5</td>
<td>Cur. school: 'Realschule'</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>curscol6</td>
<td>Cur. school: 'Gymnasium'</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>curscol7</td>
<td>Cur. school: 'Gesamtschule'</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>curscol8</td>
<td>Cur. school: Other school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>delivcs</td>
<td>Delivery by caesarean section</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>delivpl</td>
<td>Place of delivery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dint</td>
<td>Day of interview</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dis0</td>
<td>Child is restricted in abilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>disord</td>
<td>Child has confirmed disorders</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disord1</td>
<td>Confirm. disorders: Sensory (vision&amp;hearing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>disord2</td>
<td>Confirm. disorders: Motor functions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Label</td>
<td>Bioage01</td>
<td>Bioage02</td>
<td>Bioage03</td>
<td>Bioage06</td>
<td>Bioage08</td>
<td>Bioage10</td>
<td>SOEP</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>disord3</td>
<td>Confirm. disorders: Neurological</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disord4</td>
<td>Confirm. disorders: Speech</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disord5</td>
<td>Confirm. disorders: Regulatory</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disord6</td>
<td>Confirm. disorders: Chronic disease</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disord7</td>
<td>Confirm. disorders: Physical disability</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disord8</td>
<td>Confirm. disorders: Mental disability</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disord9</td>
<td>Confirm. disorders: Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>disordu</td>
<td>First U exam confirming disorder</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>dvlpm1</td>
<td>Problems in development/behavioral problems</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dvlpm2</td>
<td>In therapy/consulting because of behav./devel. problems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh1</td>
<td>Edu. behav.: show love</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh2</td>
<td>Edu. behav.: criticize</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh3</td>
<td>Edu. behav.: ask what he/she has experienced</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh4</td>
<td>Edu. behav.: punish when disobedient</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh5</td>
<td>Edu. behav.: threaten to punish, but do not punish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh6</td>
<td>Edu. behav.: know where child is</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh7</td>
<td>Edu. behav.: rather strict with child</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh8</td>
<td>Edu. behav.: comfort child when child is sad</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh9</td>
<td>Edu. behav.: shout when child makes mistakes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>edbeh10</td>
<td>Edu. behav.: child is not grateful, b/c is disobedient</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh11</td>
<td>Edu. behav.: do not talk to child when disobedient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh12</td>
<td>Edu. behav.: tell child to not disobey</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh13</td>
<td>Edu. behav.: praise my child</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh14</td>
<td>Edu. behav.: insult child b/c I am upset w. him/her</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh15</td>
<td>Edu. behav.: try to influence childs friendships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh16</td>
<td>Edu. behav.: reduce punishmnt or cancel it</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh17</td>
<td>Edu. behav.: disappointed about bad behavior</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edbeh18</td>
<td>Edu. behav.: hard to be consistent in upbringing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>edgoal1</td>
<td>Child should: be a good student</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edgoal2</td>
<td>Child should: get along with other kids</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edgoal3</td>
<td>Child should: be interested why things happen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edgoal4</td>
<td>Child should: act like normal girl/boy</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Label</td>
<td>Bioage01</td>
<td>Bioage02</td>
<td>Bioage03</td>
<td>Bioage06</td>
<td>Bioage08</td>
<td>Bioage10</td>
<td>SOEP</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>edgoal5</td>
<td>Child should: be honest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal6</td>
<td>Child should: have good manners</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal7</td>
<td>Child should: have good self-control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal8</td>
<td>Child should: be responsible</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal9</td>
<td>Child should: treat others with respect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal10</td>
<td>Child should: obey the parents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal11</td>
<td>Child should: have good ability to judge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal12</td>
<td>Child should: be clean and neat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal13</td>
<td>Child should: strive to achieve aims</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal14</td>
<td>Child should: fit in well in groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal15</td>
<td>Child should: learn to fight obstacles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal16</td>
<td>Child should: be satisfied with self</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal17</td>
<td>Child should: learn to avoid risks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>edgoal18</td>
<td>Child should: be lovable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>famofreq</td>
<td>Freq. child sees mother/father</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>famoinhh</td>
<td>Bio. fath./moth. (of child) lives in househ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>fathfreq</td>
<td>Freq. child sees father</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>fathinhhh</td>
<td>Bio. father of child lives in household</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>feeling1</td>
<td>Phys. condition in final trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>feeling2</td>
<td>Phys. condition in 1st 3 months after birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>feeling3</td>
<td>Mental state in final trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>feeling4</td>
<td>Mental state in 1st 3 months after birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact1</td>
<td>Freq. of activities: Watching tv, video, dvd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact2</td>
<td>Freq. of activities: Computer games</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact3</td>
<td>Freq. of activities: Surfing the web/chatting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact4</td>
<td>Freq. of activities: Listening to music</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact5</td>
<td>Freq. of activities: Playing music</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact6</td>
<td>Freq. of activities: Doing sports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact7</td>
<td>Freq. of activities: Doing sth. w. family</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact8</td>
<td>Freq. of activities: Dancing, theater</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact9</td>
<td>Freq. of activities: Metal-/woodwork</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact10</td>
<td>Freq. of activities: Painting/handicraft</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Name</td>
<td>Label</td>
<td>Bioage01</td>
<td>Bioage02</td>
<td>Bioage03</td>
<td>Bioage06</td>
<td>Bioage08</td>
<td>Bioage10</td>
<td>SOEP</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>freqact11</td>
<td>Freq. of activities: Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact12</td>
<td>Freq. of activities: Hanging out, dreaming</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact13</td>
<td>Freq. of activities: Being together with friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>freqact14</td>
<td>Freq. of activities: Going to church</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>health</td>
<td>Concerns about the child health</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>height</td>
<td>Height of child in cm</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>heightb</td>
<td>Height of child at birth in cm</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>helphmwk1</td>
<td>Frequency help with homework by mother</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>helphmwk2</td>
<td>Frequency help with homework by father</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>helphmwk3</td>
<td>Frequency help with homework by siblings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>helphmwk4</td>
<td>Frequency help with homework by friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>helphmwk5</td>
<td>Frequency help with homework by tutor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>helphmwk6</td>
<td>Frequency help with homework by homework supervision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>sathmwk</td>
<td>Satisfaction help with homework</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>hospital3mb</td>
<td>Days in hospital 1st 3 months after birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>hospital12m</td>
<td>Days in hospital last 12 months</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>idegrad1</td>
<td>Ideal grad: low ('Hauptschule')</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>idegrad2</td>
<td>Ideal grad: medium ('Realschule')</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>idegrad3</td>
<td>Ideal grad: high ('Abitur')</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ill2</td>
<td>Disease: Otitis media</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ill3</td>
<td>Disease: Hay fever</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>ill4</td>
<td>Disease: Neurodermatitis</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ill5</td>
<td>Disease: Ametropia</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ill6</td>
<td>Disease: Hardness of hearing</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ill7</td>
<td>Disease: Nutritional disturbances</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ill8</td>
<td>Disease: Disturbance of motoric functions</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ill9</td>
<td>Disease: Other</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ill10</td>
<td>Disease: respiratory disease</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ill11</td>
<td>Disease: Asthma</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ill12</td>
<td>Disease: Chronic bronchitis</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ill13</td>
<td>Disease: Akute bronchitis</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>ill14</td>
<td>Disease: Croup syndrome</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>
### bioage01-10: Generated variables from the parent questionnaires

<table>
<thead>
<tr>
<th>Name</th>
<th>Label</th>
<th>Bioage01</th>
<th>Bioage02</th>
<th>Bioage03</th>
<th>Bioage06</th>
<th>Bioage08</th>
<th>Bioage10</th>
<th>SOEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>illno</td>
<td>No confirmed disease</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>imper1a</td>
<td>1st person: child/adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper1b</td>
<td>1st person: relative/not related</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper1c</td>
<td>1st person: gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper1d</td>
<td>1st person: age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper1e</td>
<td>1st person: German origin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper1f</td>
<td>1st person: education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper2a</td>
<td>2nd person: child/adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper2b</td>
<td>2nd person: relative/not related</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper2c</td>
<td>2nd person: gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper2d</td>
<td>2nd person: age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper2e</td>
<td>2nd person: German origin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper2f</td>
<td>2nd person: education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper3a</td>
<td>3rd person: child/adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper3b</td>
<td>3rd person: relative/not related</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper3c</td>
<td>3rd person: gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper3d</td>
<td>3rd person: age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper3e</td>
<td>3rd person: German origin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imper3f</td>
<td>3rd person: education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inshelp1</td>
<td>Used help: family education facility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inshelp2</td>
<td>Used help: family counseling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inshelp3</td>
<td>Used help: support by youth-welfare-center</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inshelp4</td>
<td>Used help: support services by social workers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inshelp5</td>
<td>Used help: day-cares services by social workers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inshelp6</td>
<td>Used help: consultation of nurse/educator</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inshelp7</td>
<td>Used help: school-counselor/therapist</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inshelp8</td>
<td>Used help: other aid of youth-welfare-center</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inshelp9</td>
<td>Used help: other aid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lamark</td>
<td>Mark/grade: German</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>language</td>
<td>Language spoken with child at home</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>language1</td>
<td>Language 1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>language2</td>
<td>Language 2</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Name</td>
<td>Label</td>
<td>Bioage01</td>
<td>Bioage02</td>
<td>Bioage03</td>
<td>Bioage06</td>
<td>Bioage08</td>
<td>Bioage10</td>
<td>SOEP</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>language3</td>
<td>Language 3</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>lstmedex</td>
<td>Last med. examination (U-exam)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>maincare</td>
<td>Respondent(mother) is main child care provider</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>matmark</td>
<td>Mark/grade: Math</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>medaid3m</td>
<td>Medical help: # of times in last 3 months</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>medaid3mb</td>
<td>Medical help: # of times 1st 3 months after birth</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>mint</td>
<td>Month of interview</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>mvmm1</td>
<td>Child walks down stairs forward</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>mvmm2</td>
<td>Child uses door handle to open doors</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>mvmm3</td>
<td>Child uses climbing frames + high playground equip.</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>mvmm4</td>
<td>Child uses scissors to cut paper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>mvmm5</td>
<td>Child paints/draws recognizable forms on paper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>mvmm6</td>
<td>Child jumps with one foot</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>mvmm7</td>
<td>Child holds pens correctly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>mvmm8</td>
<td>Child climbs stairs w. alternating feet</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>mvmm9</td>
<td>Child can run w/o falling down</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>mvmm10</td>
<td>Child is able to jump forwards at least 3 times</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>mvmm11</td>
<td>Child can color simple forms and figures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>mvmm12</td>
<td>Child can solve puzzles w. a min. of 2 parts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>nactcar1</td>
<td>Not seek. child care: but concrete future plans</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>nactcar2</td>
<td>Not seek. child care: take it as it comes</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>nactcar3</td>
<td>Not seek. child care: general disapproval of ch. care</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>nchild</td>
<td>Child birth order</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>noccare1</td>
<td>Reason no care: Child too young</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>noccare2</td>
<td>Reason no care: Want to raise child by myself</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>noccare3</td>
<td>Reason no care: At home anyway and can take care</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>noccare4</td>
<td>Reason no care: Costs are too high</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>noccare5</td>
<td>Reason no care: No spots available</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>noccare6</td>
<td>Reason no care: Distance too far</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>noccare7</td>
<td>Reason no care: Opening hours not suitable</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>noccare8</td>
<td>Reason no care: Transfers too time-consuming</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>noccare9</td>
<td>Reason no care: Child has a chronic disease/disorder</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Name</td>
<td>Label</td>
<td>Bioage01</td>
<td>Bioage02</td>
<td>Bioage03</td>
<td>Bioage06</td>
<td>Bioage08</td>
<td>Bioage10</td>
<td>SOEP</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>noccare10</td>
<td>Reason no care: Child should spend time w. siblings</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>noccare11</td>
<td>Reason no care: Good experiences caring at home</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prebeg</td>
<td>Spell Begin Pregnancy (Month, 01.83=1)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>preend</td>
<td>Spell End Pregnancy, Birth (Month,01.83=1)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pregmo</td>
<td>Mother: pregnancy month at interview</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>pregplan</td>
<td>Pregnancy unintended/intended</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pregy</td>
<td>Mother: pregnant at interview, year</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>probgra1</td>
<td>Prob. of grad: low ('Hauptschule')</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>probgra2</td>
<td>Prob. of grad: medium ('Realschule')</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>probgra3</td>
<td>Prob. of grad: high ('Abitur')</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reccare1</td>
<td>Reason child care: Employment</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reccare2</td>
<td>Reason child care: Increased work schedule</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reccare3</td>
<td>Reason child care: start/contin. acadm. studies</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reccare4</td>
<td>Reason child care: Positive impact on child's develop.</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reccare5</td>
<td>Reason child care: Want more time for myself</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reccare6</td>
<td>Reason child care: Other</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>relapare</td>
<td>Relationship to biological mother/father</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>saccare1</td>
<td>Satisf. child care: with group size</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>saccare2</td>
<td>Satisf. child care: with size of staff</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>saccare3</td>
<td>Satisf. child care: with opening hours</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>saccare4</td>
<td>Satisf. child care: with costs</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>saccare5</td>
<td>Satisf. child care: with flexibility</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>saccare6</td>
<td>Satisf. child care: with conduct of staff</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>saccare7</td>
<td>Satisf. child care: with activities/education</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>saccare8</td>
<td>Satisf. child care: with parental participation</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>saccare9</td>
<td>Satisf. child care: with contact to nurse/teacher</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>saccare10</td>
<td>Satisf. child care: with preparation for school</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>saticare</td>
<td>General satisfaction with child care</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>scactiv1</td>
<td>Other school activity: athletics</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scactiv2</td>
<td>Other school activity: acting/dancing</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scactiv3</td>
<td>Other school activity: choir/orchestra/music group</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scactiv4</td>
<td>Other school activity: other</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Label</td>
<td>Bioage01</td>
<td>Bioage02</td>
<td>Bioage03</td>
<td>Bioage06</td>
<td>Bioage08</td>
<td>Bioage10</td>
<td>SOEP</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>sclenrolm</td>
<td>Start of school (ever), month</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sclenroln</td>
<td>Not yet in school</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sclenroly</td>
<td>Start of school (ever), year</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sclr1</td>
<td>Child calls familiar people by name</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>sclr2</td>
<td>Child plays games with other Children</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sclr3</td>
<td>Child participates in role-playing (‘as if’)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sclr4</td>
<td>Child shows particular liking for certain friends</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sclr5</td>
<td>Child specif. own feelings, e.g. sad, happy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sclr6</td>
<td>Child is happy or worried for others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>sclr7</td>
<td>Child takes turns when playing w/o being asked</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sclr8</td>
<td>Child plays games with other Children</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>sclr9</td>
<td>Child shows desire to make others feel happy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sclr10</td>
<td>Child shares toys if asked for</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>sclr11</td>
<td>Child continues playing w/o crying whn parents leave</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sclr12</td>
<td>Child creatively plays with household stuff</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>sclr13</td>
<td>Child seeks to make friends with same-age kids</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sclr14</td>
<td>Child fulfills simple requests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scolcon1</td>
<td>Schl. cond.: likes to go to school</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scolcon2</td>
<td>Schl. cond.: does not get along with classmates</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scolcon3</td>
<td>Schl. cond.: thinks school is waste of time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scolcon4</td>
<td>Schl. cond.: does not take school work seriously</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scolcon5</td>
<td>Schl. cond.: follows lessons well</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scolcon6</td>
<td>Schl. cond.: does not get along w. teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scolcon7</td>
<td>Schl. cond.: likes to study/has a zeal for learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scolcon8</td>
<td>Schl. cond.: gets along w. classmates (SC 2010 only)</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scolcon9</td>
<td>Schl. cond.: gets along w. teacher (SC 2010 only)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scoldura</td>
<td>Time in school in months</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sex</td>
<td>Gender of child</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>skill1</td>
<td>Child eats w. spoon w/o spilling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>skill2</td>
<td>Child blows nose w/o assistance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>skill3</td>
<td>Child uses toilet to do number two</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>skill4</td>
<td>Child puts on pants and underpants frontwards</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Name</td>
<td>Label</td>
<td>Bioage01</td>
<td>Bioage02</td>
<td>Bioage03</td>
<td>Bioage06</td>
<td>Bioage08</td>
<td>Bioage10</td>
<td>SOEP</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td>skll5</td>
<td>Child brushes teeth w/o assistance</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>skll6</td>
<td>Child closes press buttons w/o assistance</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>skll7</td>
<td>Child covers mouth and nose when sneezing</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>skll8</td>
<td>Child takes off sweater/jacket w/o assistance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>skll9</td>
<td>Child drinks f. a cup/glass w/o spilling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>skll10</td>
<td>Child washes and dries its face w/o assistance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>spch1</td>
<td>Child understands brief instructions</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>spch2</td>
<td>Child forms sentences with at least two words</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>spch3</td>
<td>Child speaks in full sentences (at least four words)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spch4</td>
<td>Child attends orders given by someone 5 minutes ago</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>spch5</td>
<td>When asked, child gives first and last name</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>spch6</td>
<td>Child is able to inform others w. simple messages</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>spch7</td>
<td>Child listens attentively to story f. at least 15 min.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>spch8</td>
<td>Child follows brief instructions w. simple tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spch9</td>
<td>Child listens attentively to story f. at least 5 min.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>spch10</td>
<td>Child can identify at least 5 body parts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spch11</td>
<td>Child can identify daily-life-objects in a book</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spch12</td>
<td>Child answers questions in words, or tries to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spch13</td>
<td>Child can name at least 10 objects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spch14</td>
<td>Child can describe things in simple words</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>specned1</td>
<td>Special needs: learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>specned2</td>
<td>Special needs: linguistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>specned3</td>
<td>Special needs: behavior/experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>specned4</td>
<td>Special needs: mental</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>specned5</td>
<td>Special needs: amlyopia (eyes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>specned6</td>
<td>Special needs: hearing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>specned7</td>
<td>Special needs: physical/motoric</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>specned8</td>
<td>Special needs: long-lasting disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>specned9</td>
<td>Special needs: none</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stcaremn</td>
<td>Start of child care (ever), month</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>stcareyr</td>
<td>Start of child care (ever), year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>suppartn</td>
<td>Feel supported by partner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Name</td>
<td>Label</td>
<td>Bioage01</td>
<td>Bioage02</td>
<td>Bioage03</td>
<td>Bioage06</td>
<td>Bioage08</td>
<td>Bioage10</td>
<td>SOEP</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>temp1</td>
<td>Child generally happy and satisfied</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>temp2</td>
<td>Child is easily irritated, cries often</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>temp3</td>
<td>Child hard to console</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>temp4</td>
<td>Child is curious and active</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>temp5</td>
<td>Child tends to be shy</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>temp6</td>
<td>Child likes to talk</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>temp7</td>
<td>Child shows empathy when others are sad</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>timfamod</td>
<td>Time father/mother sees child (days)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>timfamoh</td>
<td>Time father/mother sees child (hours)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>timfathd</td>
<td>Time father sees child (days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>timfathh</td>
<td>Time father sees child (hours)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tvhrs</td>
<td>Watches video/tv alone (h/week)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>tvyn</td>
<td>Child may watch tv unattended</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>vistfrnd</td>
<td>Time child spends at friends (alone)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight</td>
<td>Weight of child in kilograms</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>weightb</td>
<td>Weight of child at birth in grams</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Topics covered in the bioage files

The bioage data files cover 18 different subjects regarding various aspects of children’s upbringing. The following table provides an overview of which subject is covered in which file. For more specific information please refer to the parent-questionnaires.

Table 2: Overview of topics

<table>
<thead>
<tr>
<th>Subjects covered</th>
<th>Bioage File:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>01</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
</tr>
<tr>
<td>Age group (years)</td>
<td>0-1</td>
</tr>
<tr>
<td>Information regarding child birth</td>
<td>x</td>
</tr>
<tr>
<td>Physical and mental well-being during pregnancy</td>
<td></td>
</tr>
<tr>
<td>and at the time of giving birth</td>
<td>x</td>
</tr>
<tr>
<td>Changed circumstances and first experiences of</td>
<td>x</td>
</tr>
<tr>
<td>living with a child</td>
<td></td>
</tr>
<tr>
<td>Child’s health, medical examinations and</td>
<td>x</td>
</tr>
<tr>
<td>medical treatment</td>
<td></td>
</tr>
<tr>
<td>Relationship to biological mother/father</td>
<td>x</td>
</tr>
<tr>
<td>Activities done with the child</td>
<td></td>
</tr>
<tr>
<td>Leisure-time</td>
<td></td>
</tr>
<tr>
<td>Child care</td>
<td></td>
</tr>
<tr>
<td>School</td>
<td></td>
</tr>
<tr>
<td>Peer groups</td>
<td></td>
</tr>
<tr>
<td>Child’s network</td>
<td></td>
</tr>
<tr>
<td>Child’s temper and personality</td>
<td>x</td>
</tr>
<tr>
<td>Child’s adaptive behavior</td>
<td></td>
</tr>
<tr>
<td>Child’s socio-emotional behavior</td>
<td></td>
</tr>
<tr>
<td>Language spoken in household</td>
<td></td>
</tr>
<tr>
<td>Allowance</td>
<td></td>
</tr>
<tr>
<td>Parental ambitions regarding the education of the</td>
<td></td>
</tr>
<tr>
<td>child</td>
<td></td>
</tr>
<tr>
<td>Being parents</td>
<td></td>
</tr>
<tr>
<td>Upbringing: preferences/principles</td>
<td></td>
</tr>
<tr>
<td>Educational aims</td>
<td></td>
</tr>
</tbody>
</table>
Data files and respondents

Data files are named according to the age of a child found within a specific bioage data file, corresponding with the file specific suffix (e.g. bioage06 includes children turning six in the survey year, giving a range between 5 years and 1 month and 6 years and 11 months, depending on month of birth and interview month). For the correspondence between age groups and questionnaires please consult table 3. Whether the mother or the father or both answer the questionnaire and thus are the (proxy) respondent for the bioage data depends on the respective questionnaire.

bioage01:  ‘Parent-Questionnaire 1: 0-1 year old’

The questionnaire is given to all women who gave birth to their own child in the current survey year or the year before. In addition, all women whose non-biological child was born in this time period receive the questionnaire. In contrast to parent-questionnaires for older children, because of some of the contents, the “Parent-Questionnaire 1” is meant to be answered by the mother (in a very few cases fathers have answered the questionnaire).

bioage02:  ‘Parent-Questionnaire 2: 1-2 years old’

Principally, this questionnaire is given to all mothers whose child turns two in the current survey year (this applies to biological, as well as non-biological children). Note, however that in case the father is a single parent he may answer the questionnaire as well.

bioage03:  ‘Parent-Questionnaire 3: 2-3 years old’

The questionnaire is given to all parents whose child turns three in the current survey year. It can be answered either by the mother or father.

bioage06:  ‘Parent-Questionnaire 4: 5-6 years old’

The questionnaire is given to all parents with children turning six years old during the survey year. It should be answered either by the mother or by the father.

bioage08:  ‘Parent-Questionnaire 5: 7-8 years old’

The questionnaire is given to both mothers and fathers of children turning eight years old in the current survey year. Two questionnaires will be filled out only if either (biological or non-biological) parents live in the household.
**bioage10:** ‘Parent-Questionnaire 6: 9-10 years old’

The questionnaire is given to *both* mothers and fathers of children turning eight years old in the current survey year. Two questionnaires will be filled out only if either (biological or non-biological) parents live in the household.

**Table 3:** Overview questionnaire names and corresponding age group

<table>
<thead>
<tr>
<th>Bioage File</th>
<th>Classification by Age</th>
<th>Quest. Field Name Wave 2010</th>
<th>Quest. Field Name Wave 2011 and later</th>
<th>Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>bioage01</td>
<td>Parent-Quest. 0-1 years old</td>
<td>Parent-Quest.1A</td>
<td>Parent-Quest. 1</td>
<td>Mother only</td>
</tr>
<tr>
<td>bioage02</td>
<td>Parent-Quest. 1-2 years old</td>
<td>Parent-Quest.1B</td>
<td>Parent-Quest. 2</td>
<td>Mother or father</td>
</tr>
<tr>
<td>bioage03</td>
<td>Parent-Quest. 2-3 years old</td>
<td>Parent-Quest. 2</td>
<td>Parent-Quest. 3</td>
<td>Mother or father</td>
</tr>
<tr>
<td>bioage06</td>
<td>Parent-Quest. 5-6 years old</td>
<td>Parent-Quest. 3</td>
<td>Parent-Quest. 4</td>
<td>Mother or father</td>
</tr>
<tr>
<td>bioage08</td>
<td>Parent-Quest. 7-8 years old</td>
<td>Parent-Quest. 4</td>
<td>Parent-Quest. 5</td>
<td>Mother and father</td>
</tr>
<tr>
<td>bioage10</td>
<td>Parent-Quest. 9-10 years old</td>
<td>Parent-Quest. 5</td>
<td>Parent-Quest. 6</td>
<td>Mother and father</td>
</tr>
</tbody>
</table>

**Specifics of bioage08 & bioage10**

Due to the fact that parent-questionnaires 5 and 6 are given to both parents of the child, the dataset of *bioage08* and *bioage10* was split into *bioage08p1/bioage10p1* for the first respondent and *bioage08p2/bioage10p2* for the second respondent. Typically, the first respondent is the biological mother or a single parent. Thus, for each child there exists a second observation provided the second parent filled out the respective questionnaire. The actual number of children is congruent with the number of observations in *bioage08p1/bioage10p1*, while the number of observations in *bioage08p2/bioage10p2* are those observations, which are additionally available for the children of that age group (note that this applies only to children for whom two questionnaires have been filled out).
Table 4: Overview number of observations in bioage08 p1 and p2

<table>
<thead>
<tr>
<th></th>
<th>Bioage08p1</th>
<th>Bioage08p2</th>
<th>Bioage08 (total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother</td>
<td>1809</td>
<td>0</td>
<td>1809</td>
</tr>
<tr>
<td>Father</td>
<td>55</td>
<td>1276</td>
<td>1331</td>
</tr>
<tr>
<td>Total</td>
<td>1864</td>
<td>1276</td>
<td>3140</td>
</tr>
</tbody>
</table>

Table 5: Overview number of observations in bioage10 p1 and p2

<table>
<thead>
<tr>
<th></th>
<th>Bioage10p1</th>
<th>Bioage10p2</th>
<th>Bioage10 (total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother</td>
<td>1782</td>
<td>0</td>
<td>1782</td>
</tr>
<tr>
<td>Father</td>
<td>67</td>
<td>1144</td>
<td>1211</td>
</tr>
<tr>
<td>Total</td>
<td>1849</td>
<td>1144</td>
<td>2993</td>
</tr>
</tbody>
</table>

**bioagel:** ‘Combined dataset of all bioage files’

Starting with the data distribution FiD v2.1, a combined dataset of all bioage files is distributed, analogously to the SOEP dataset first distributed with SOEP v28. This file is called **bioagel**, where the “l” characterized the dataset as being in the “long” format, i.e. multiple observations per child are present in the dataset, where variables do not have a prefix anymore, but are identically named in case the question across the parent questionnaires are identical\(^\text{10}\). The **bioagel** dataset allows following children and their developments within one dataset: e.g., one can now see how children have grown from one interview to the next. The data are organized on a child-dataset level, i.e. the variables PERSNR and BIOAGE (see below) identify unique observations. Note that any variable not asked because the child was not in the respective age group will receive the missing code “-5 Question not included”. The **bioagel** dataset will become more and more useful with every year of additional observations.

\(^{10}\) Note that from FiD v3.0 the variables B10BEHAV in bioage10 are no longer an exception to this rule and all these variables are presented in a three-point scale (see below).
**Number of observations**

Table 6: Overview number of observations per wave

<table>
<thead>
<tr>
<th>Filename</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>bioage01</td>
<td>1,321</td>
<td>207</td>
<td>212</td>
<td>167</td>
<td>1,907</td>
</tr>
<tr>
<td>bioage02</td>
<td>787</td>
<td>647</td>
<td>568</td>
<td>187</td>
<td>2,189</td>
</tr>
<tr>
<td>bioage03</td>
<td>871</td>
<td>741</td>
<td>555</td>
<td>523</td>
<td>2,690</td>
</tr>
<tr>
<td>bioage06</td>
<td>473</td>
<td>486</td>
<td>425</td>
<td>656</td>
<td>2,040</td>
</tr>
<tr>
<td>bioage08p1</td>
<td>425</td>
<td>529</td>
<td>501</td>
<td>409</td>
<td>1,864</td>
</tr>
<tr>
<td>bioage08p2</td>
<td>257</td>
<td>373</td>
<td>348</td>
<td>298</td>
<td>1,276</td>
</tr>
<tr>
<td>bioage10p1</td>
<td>403</td>
<td>510</td>
<td>477</td>
<td>459</td>
<td>1,849</td>
</tr>
<tr>
<td>bioage10p2</td>
<td>242</td>
<td>310</td>
<td>291</td>
<td>301</td>
<td>1,144</td>
</tr>
<tr>
<td>Total</td>
<td>4,779</td>
<td>3,803</td>
<td>3,377</td>
<td>3,000</td>
<td>14,959</td>
</tr>
</tbody>
</table>

**Generated Variables**

This section provides additional information on variables, which have been generated from combinations of variables from other datasets than the parent questionnaires.

**AGE**

- **Variable label**: “Child's age in months”
- **Variable format**: 2-digit integer
- **Bioage File**: 01-10

**Comment**: The variable AGE provides the child’s age in months, as a combination of month of birth information and the interview month. As the exact day of birth remains unknown, information is only an approximation and may vary by one month.

**BREASTFM**

- **Variable label**: “Breast-feeding time in months”
- **Variable format**: 2-digit integer
- **Bioage File**: 01; 02; 03
Comment: The breast-feeding time in months is taken directly from the questionnaire, if the child is no longer breast-fed and the mother stated the number of months. If the baby is currently breast-fed, BREASTFM is set to “-2 Does not apply”, because the information is not identical. However, assuming that there were no gaps in breast-feeding, users can easily generate BREASTFM for children who are currently breast-fed by taking the age of the child in months (AGE).

**PREBEG**

**Variable label**: “Spell Begin Pregnancy (Month, 01.83=1)”

**Variable format**: 3-digit integer

**Bioage File**: 01

Comment: This variable is based on the exact month of birth (BIRTHM) and the duration of childbearing in weeks (BIRTHPW). Accordingly, information is available only for women who completed the “Parent-Questionnaire 1” and for whom the duration of the pregnancy is known. Note that the month of conception may vary by one month, as the exact day of birth remains unknown.

The variable PREBEG contains information on the beginning of childbearing (e.g. the month of conception). Information in FiD is given in the regular SOEP spell format: values start with 1 for January 1983 (e.g. earliest spell in bioage01, survey year 2010, is 304, which equals April 2008). This spell design allows for direct comparison with spell information in the SOEP data files (e.g. biomarsm, artkalen).

**PREEND**

**Variable label**: “Spell End Pregnancy, Birth (Month, 01.83=1)”

**Variable format**: 3-digit integer

**Bioage File**: 01

Comment: This variable is based on the exact month of birth (BIRTHM) and the duration of childbearing in weeks (BIRTHPW). Accordingly, information is available only for women who completed the “Parent-Questionnaire 1” and for whom the duration of the pregnancy is known. Variable PREEND contains information on the end of childbearing (e.g. the month of birth). Information in FiD is given in the regular SOEP spell format: values start with 1 for January 1983 (e.g. earliest spell in bioage01, survey year 2010, is 304, which equals April 2008). This spell design allows for direct comparison with spell information in the SOEP data files (e.g. biomarsm).

**PREGY**

**Variable label**: “Mother: pregnant at interview, year”

**Variable format**: 4-digit integer

**Bioage File**: 01
Comment

This variable is based on information the mother gave on her then current pregnancy status in the previous year in the person-questionnaire. In case a pregnancy is stated (or was unknown) and a child was born, the year in which the interview was done is recorded in PREGY. Hence this information is available only for those women in the sample for at least two years with completed personal interview in the first and a completed parent-questionnaire 1 in the second year. Please note that future mother need not be aware of their pregnancy in the early stages of childbearing.

**PREGMO**

Variable label “Mother: pregnancy month at interview”
Variable format 2-digit integer
Bioage File 01

Comment

This variable is based on the exact month of birth (BIRTHM), the duration of childbearing in weeks (BIRTHPW) and the interview month of the previous year’s personal interview. Hence this information is available only for those women in the sample for at least two years. As the exact day of birth is unknown, this variable remains a close estimation.

**SEX**

Variable label “Gender of child”
Variable format 1-digit integer
Bioage File 01-10

Comment

The sex of the child has not been asked for in the Parent-Questionnaires. Information for this variable stems from the *ppfad*.

**SEXRESP**

Variable label “Gender of respondent (parent)”
Variable format 1-digit integer
Bioage File 01-10

Comment

This variable can be used to identify whether the mother or father of the child has answered the respective questionnaire. The information for this variable comes from the *ppfad* file.

**CARE6**

Variable label “Cared for by nanny/day care (not in-house)”
Variable format 1-digit integer
Bioage File 01; 02; 03

Comment

Within the parent questionnaires 1-3 there are some discrepancies when comparing the first question on child care (related to day care and nannies only, variables $E121b$, $E224$, $E324$) with the second one.
(dealing with child care in general, variables $E128, $E232, $E332). Specifically, instances of nannies (not in-house) and day care centers were reported less in the latter questions, which also does not correspond with information given in the household questionnaire (dataset $kind$, variables $K065A, $K065B, $K066A, $K066B, $K070D). We cannot be sure of the cause of these differences, we assume, however that it was not clear to some respondents why they were meant to give partially identical information in the two questions in the parent questionnaire. This assumption is based on the CAPI interviews, where the differences are more pronounced. In CAPI mode, first the list of care givers is asked in questions $E128, $E232, and $E332, making the items almost identical to the first question. Only then the respondents are asked to provide the number of hours each care giver spends with the child. In PAPI mode, the respondents immediately see that they are asked to provide the hours as well, and hence understand the purpose of the question. When the person filling out the household questionnaire is identical to the one doing the parent questionnaire, the problem is also slightly more present. We interpret this again in the direction that individuals do not understand why they should give the same information twice (or even three times).

These problems are tackled by combining all information available for nannies (CARE6) and day care centers (CARE8) as well as hours spent at day care centers (CARE8H). We use the household information for each child (from $kind$) together with the information from the respective parent questionnaire to provide more reliable measures. The most reliable information is assumed to be the household questionnaire information – hence hours and incidence are taken from here if there is information available. Only when the parent questionnaire provides additional information, it is used and coded in the respective variable. Of course the solution we provide here is only one of many. As all information from parent and household questionnaires are available to the user, other ways to deal with these issues are possible. However, we advise to use caution when using the variables $E128E1, $E128E2, $E128F1, and $E128F2 in parent questionnaire 1, $E232F1, $E232F2, $E232H1, and $E232H2 in parent questionnaire 2, and $E332F1, $E332F2, $E332H1, and $E332H2 in parent questionnaire 3.

To cope with the problems mentioned above, the questionnaire was slightly changed in 2013, where the question on nannies and dare care centers was asked only once (e.g. questions 21 and 29 in parent questionnaire 1). In addition to whether the child was cared for by a nanny or day care center, information on the duration of care per day was collected. In the later question (where information on other sources of care is collected) day care and nannies are not included any more. This led to two changes in the coding of the variable: first, CARE6 and CARE8 had to be generated via CCAREFR (frequency of child care). Second, as CARE6H and CARE8H are now captured in hours per day, the values were multiplied by five to adapt this information to the structure of the previous waves (hours per week).
**CARE6H**
Variable label: “Cared for by nanny/day care (not in-house) (hrs/wk)”
Variable format: 3-digit integer
Bioage File: 01; 02; 03
Comment: See comment for CARE6

**CARE8**
Variable label: “Cared for in crib/day care center”
Variable format: 1-digit integer
Bioage File: 01; 02; 03
Comment: See comment for CARE6

**CARE8H**
Variable label: “Cared for in crib/day care center (hrs/wk)”
Variable format: 3-digit integer
Bioage File: 01; 02; 03
Comment: See comment for CARE6

**CARE9H – CARE11H**
Variable label: “Cared for at school/ in after-school hoard/ in social institution, center (hrs/wk)”
Variable format: 3-digit integer
Bioage File: 08; 10
Comment: In 2013 the information on CARE9H, CARE10H and CARE11H was captured in terms of hours per day. To adapt this information to the structure of the previous waves (hours per week) we multiplied it by five. However, this created some inconsistencies which probably derive from the nature of the question. Since the previous questions were asked in terms of hours per week, it seemed that some respondents still answered the CARE9H – CARE11H for the whole week. Therefore the answers were only multiplied by five when they did not exceed 12 hours, assuming that no child spends more than 12 hours a day in care facilities and that other answers were meant to reflect the weekly hours. Thus, answers over 12 remained unchanged.

**CARE1H - CARE13H**
Variable Label: “Cared for by <care giver> (hrs/wk)”
Variable format: 3-digit integer
Bioage File: 01; 02; 03; 06; 08; 10
Comment: For all positive hours given for the different care givers, we sum them up and check whether they are larger than 168 (the maximum hours per
week). If so, we reduce each positive number of hours according to the percentage they make up from the total and round them to the nearest full hour. This is done after any corrections are applied to CARE8H.
**BIOAGE**

<table>
<thead>
<tr>
<th>Variable label</th>
<th>Value labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIOAGE</td>
<td>BIOAGE</td>
</tr>
<tr>
<td></td>
<td>(1) Bioage01</td>
</tr>
<tr>
<td></td>
<td>(2) Bioage02</td>
</tr>
<tr>
<td></td>
<td>(3) Bioage03</td>
</tr>
<tr>
<td></td>
<td>(6) Bioage06</td>
</tr>
<tr>
<td></td>
<td>(81) Bioage08p1</td>
</tr>
<tr>
<td></td>
<td>(82) Bioage08p2</td>
</tr>
<tr>
<td></td>
<td>(101) Bioage10p1</td>
</tr>
<tr>
<td></td>
<td>(102) Bioage10p2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable format</th>
<th>Bioage File</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-digit integer</td>
<td>bioagel</td>
</tr>
</tbody>
</table>

**Comment**
The variable BIOAGE is only given in the long version of the `bioage` data, in `bioagel`. It allows identifying the source of the information, i.e. from which `bioage` dataset a variable was drawn. For a distribution of BIOAGE, refer to table 6.
**Different questionnaire versions**

While the main part of the questionnaires has remained constant across the different samples and across the two waves of data collection, there have been a few changes, which are documented here.

**bioage02**

Information on birth circumstances for children of 2010, who were already covered by the parent-questionnaires 1, is no longer collected in the following waves. To be able to separate this (purely retrospective) part visually for the respondents, to some small changes in question ordering were needed (see table 7).

| Table 7: Changes in question ordering, retrospective and non-retrospective birth questions |
|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| **Parent questionnaire 2** | **2010 Cohort Sample** | **2011 and 2012 (all samples)** |
| Planned Pregnancy | Question 6 | Question 6, retrospective |
| Place of delivery | Question 7 | Question 7, retrospective part |
| Pregnancy week of delivery | Question 8 | Question 8, retrospective part |
| Height, weight and head circumference at birth | Question 9 | Question 9, retrospective part |
| Breast-feeding | Question 10 | Question 12, not in retrospective part |
| Disorders at medical exams | Question 11 | Question 13, not in retrospective part |
| Last U-exam | Question 12 | Question 14, not in retrospective part |
| Medical help needed during 1st 3 months after birth | Question 13 | Question 10, retrospective part |
| Hospital visits during 1st 3 months after birth | Question 14 | Question 11, retrospective part |

Note also that the parent-questionnaire 2 was fielded by mail for the screening sample in 2010 after the data collection had already finished (it is the only questionnaire done by mail in FiD).

**bioage03**

The parent-questionnaire 3 in the screening sample of 2010 lacked information on birth circumstances of the child. This was later added to the questionnaire in the cohort sample in 2010, and is now a fixed part of this questionnaire. In addition, similar to the parent-
questionnaire 2, the retrospective information on birth circumstances is not asked a second time. These changes bring about new information as well as changes in question ordering documented in table 8. Note that information missing due to questions not being asked is coded “[-2] Does not apply”.

**Table 8: Additions to questions and changes in order of questions in parent questionnaire 3**

<table>
<thead>
<tr>
<th>Parent questionnaire 3</th>
<th>2010 Screening Sample</th>
<th>2010 Cohort Sample</th>
<th>2011/2012 (all samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child’s birth order</td>
<td>Not included</td>
<td>Question 3</td>
<td>Question 3, not in retrospective part</td>
</tr>
<tr>
<td>Biological child</td>
<td>Not included</td>
<td>Question 4</td>
<td>Question 4, not in retrospective part</td>
</tr>
<tr>
<td>Changes after birth</td>
<td>Not included</td>
<td>Question 5</td>
<td>Question 5, not in retrospective part</td>
</tr>
<tr>
<td>Planned Pregnancy</td>
<td>Not included</td>
<td>Question 6</td>
<td>Question 6, retrospective</td>
</tr>
<tr>
<td>Place of delivery</td>
<td>Not included</td>
<td>Question 7</td>
<td>Question 7, retrospective</td>
</tr>
<tr>
<td>Pregnancy week of delivery</td>
<td>Not included</td>
<td>Question 8</td>
<td>Question 8, retrospective</td>
</tr>
<tr>
<td>Height, weight and head circumference at birth</td>
<td>Not included</td>
<td>Question 9</td>
<td>Question 9, retrospective</td>
</tr>
<tr>
<td>Breast-feeding</td>
<td>Question 3</td>
<td>Question 10</td>
<td>Question 12, not in retrospective part</td>
</tr>
<tr>
<td>Disorders at medical exams</td>
<td>Not included</td>
<td>Question 11</td>
<td>Question 13, not in retrospective part</td>
</tr>
<tr>
<td>Last U-exam</td>
<td>Not included</td>
<td>Question 12</td>
<td>Question 14, not in retrospective part</td>
</tr>
<tr>
<td>Medical help needed during 1st 3 months after birth</td>
<td>Not included</td>
<td>Question 13</td>
<td>Question 10, retrospective</td>
</tr>
<tr>
<td>Hospital visits during 1st 3 months after birth</td>
<td>Not included</td>
<td>Question 14</td>
<td>Question 11, retrospective</td>
</tr>
<tr>
<td>Current height and weight</td>
<td>Question 7</td>
<td>Question 15</td>
<td>Question 15</td>
</tr>
<tr>
<td>Hospital visits in last 12 months</td>
<td>Question 4</td>
<td>Question 16</td>
<td>Question 16</td>
</tr>
<tr>
<td>Medical help needed in last 3 months</td>
<td>Question 5</td>
<td>Question 17</td>
<td>Question 17</td>
</tr>
<tr>
<td>Illnesses or disorders confirmed by doctor</td>
<td>Question 6</td>
<td>Question 18</td>
<td>Question 18</td>
</tr>
<tr>
<td>Rest of the questionnaire corresponds one-to-one</td>
<td>Question 8 through question 28</td>
<td>Question 19 through question 39</td>
<td>Question 19 through question 39</td>
</tr>
</tbody>
</table>
The parent-questionnaire 4 for children aged five to six the information on birth order and whether the child was biological or not was missing. This information was added starting with the data collection in 2011, changing the order of questions. Question 4 now covers the birth order, question 5 whether the child is the respondent’s biological child or not. All other questions starting with former question 4, child’s height and weight, are moved back by two positions.

bioage08 and bioage10
The bioage08 and bioage10 files are based on the parent-questionnaires 5 and 6, respectively. In the screening sample of 2010, question 13 was used in a slightly different version. This question covers the school conduct in both questionnaires and provides the information for the variable SCOLON1 through SCOLON9. Due to this change, valid observations for variables SCOLON8 and SCOLON9 exist for the screening sample in 2010 only, and hence these two variables are marked with a “SC 2010” in the variable label. An overview of items and respective differences is provided in table 9. As before, if a question is specific to one sample, observations stemming from the other sample will show the value “[-2] Does not apply” for this variable. Figures 1 and 2 show the questions as they appear in the respective questionnaires.
Table 9: *bioage08/bioage10* differences in SCOLON across samples

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Label Text</th>
<th>Item in Questionnaires 5/6</th>
<th>All but Screening sample asked in 2010</th>
<th>Screening sample, asked in 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCOLCON1</td>
<td>likes to go to school</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>SCOLCON2</td>
<td>doesn't get along with classmates</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCOLCON3</td>
<td>thinks school is a waste of time</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCOLCON4</td>
<td>never takes school work seriously</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCOLCON5</td>
<td>is able to follow the lessons</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>SCOLCON6</td>
<td>doesn't get along with current teacher</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCOLCON7</td>
<td>likes to study/has a zeal for learning</td>
<td>7</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>SCOLCON8</td>
<td>gets along with classmates (SC 2010)</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCOLCON9</td>
<td>gets along with current teacher (SC 2010)</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Figure 1:** Screenshot Parent Questionnaire 5 Screening Sample 2010
(Parent Questionnaire 6 similar)

13. Inwieweit treffen die folgenden Aussagen auf das Kind zu?

Das Kind...

- versteht sich gut mit seinen Klassenkameraden
- geht gerne in die Schule
- kommt im Unterricht gut mit
- kommt mit dem jetzigen Lehrer oder der Lehrerin gut aus
- lernt gerne

Options:
- trifft voll zu
- trifft eher zu
- trifft eher nicht zu
- trifft gar nicht zu
- weiß nicht

**Figure 2:** Screenshot Parent Questionnaire 5 Cohort Sample 2010 and 2011 sample
(Parent Questionnaire 6 similar)

13. Inwieweit treffen die folgenden Aussagen auf das Kind zu?

Das Kind...

- geht gerne in die Schule
- versteht sich nicht gut mit seinen Klassenkameraden
- empfindet Schule als reine Zeitverschwendung
- nimmt Arbeit in der Schule nie ernst
- kommt im Unterricht gut mit
- kommt mit dem jetzigen Lehrer oder der Lehrerin nicht gut aus
- lernt gerne

Options:
- trifft voll zu
- trifft eher zu
- trifft eher nicht zu
- trifft gar nicht zu
- weiß nicht
bioage10

Until the wave 2012, the questions on the child’s socio-economic behavior (Strengths-and-Difficulties, SDQ\textsuperscript{11}, captured in the variables B10BEHAV1-B10BEHAV25) have been asked on a 7-point scale ranging from 1 (does not apply at all) to 7 (fully applies). However, the international standard requires a 3-point scale, which will be implemented starting in 2013. In 2012, both scales have been used to find a way to easily “translate” the 7-point-scale into the 3-point-scale, mainly for the previous waves. To accomplish this, two questionnaire versions containing either scale were randomly distributed across households. The information from 2012 wave was then used to recode the information from 2010 and 2011 into a 3-point-scale in the bioage10 datasets, while the original information remains available in the respective Parent-Questionnaires (Seltern6). Based on the distribution of each single item on the 3-point-scale, the 7-point-scale items can be categorized into the following 5 groups:

### Table 10: Recoding scheme for SDQ questions in bioage10 (BEHAV)

<table>
<thead>
<tr>
<th>Group</th>
<th>Recoding scheme of item values</th>
<th># of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>“1” =&gt; “1”; “2”-“5” =&gt; “2”; “6”-“7” =&gt; “3”</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>“1”-“3” =&gt; “1”; “4”-“5” =&gt; “2”; “6”-“7” =&gt; “3”</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>“1”-“2” =&gt; “1”; “3”-“5” =&gt; “2”; “6”-“7” =&gt; “3”</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>“1”-“2” =&gt; “1”; “3”-“6” =&gt; “2”; “7” =&gt; “3”</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>“1”-“3” =&gt; “1”; “4”-“6” =&gt; “2”; “7” =&gt; “3”</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>“1”-“1” =&gt; “1”; “3”-“4” =&gt; “2”; “5”-“7” =&gt; “3”</td>
<td>1</td>
</tr>
</tbody>
</table>

Following these schemes, all 25 variables were recoded. Before 2012, the questions on the child’s socio-economic behavior only comprise 18 variables but were expanded to 25 variables in 2012. The “new” questions B10BEHAV3, B10BEHAV7-8, B10BEHAV11, B10BEHAV17-18 and B10BEHAV21 were not included in the bioage10 datasets in 2010 and 2011 and have a missing code of -5 (Questions not included for sample) for the years 2010 and 2011 in the bioage10 datasets from 2012 on. The following tables show the distributions in 2012 for a) the 7-point scales, b) the recoded 7-point-scales in 3-point-scale format, and c) the distribution for the 3-point-scale as found in the field.

\textsuperscript{11} More information on the SDQ including translations from German to English can be found at http://www.sdqinfo.org/.
Table 11: Frequencies of original and recoded variables BEHAV1-BEHAV25

<table>
<thead>
<tr>
<th>B10BEHAV1</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.02</td>
<td>1.02</td>
<td>1.88</td>
</tr>
<tr>
<td>2</td>
<td>2.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4.59</td>
<td>41.84</td>
<td>44.77</td>
</tr>
<tr>
<td>4</td>
<td>13.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>21.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>35.71</td>
<td>57.14</td>
<td>53.35</td>
</tr>
<tr>
<td>7</td>
<td>21.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV2</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>23.92</td>
<td>59.80</td>
<td>56.45</td>
</tr>
<tr>
<td>3</td>
<td>12.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>14.76</td>
<td>33.84</td>
<td>33.87</td>
</tr>
<tr>
<td>5</td>
<td>11.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6.36</td>
<td>6.36</td>
<td>9.68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV3</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41.58</td>
<td>68.88</td>
<td>67.02</td>
</tr>
<tr>
<td>2</td>
<td>27.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10.97</td>
<td>26.53</td>
<td>26.01</td>
</tr>
<tr>
<td>4</td>
<td>8.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.81</td>
<td>4.59</td>
<td>6.97</td>
</tr>
<tr>
<td>7</td>
<td>1.79</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV5</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25.26</td>
<td>46.17</td>
<td>51.61</td>
</tr>
<tr>
<td>2</td>
<td>20.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>17.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>14.80</td>
<td>43.62</td>
<td>39.52</td>
</tr>
<tr>
<td>5</td>
<td>11.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8.16</td>
<td>10.20</td>
<td>8.87</td>
</tr>
<tr>
<td>7</td>
<td>2.04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV6</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>42.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20.90</td>
<td>74.23</td>
<td>72.39</td>
</tr>
<tr>
<td>3</td>
<td>9.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9.20</td>
<td>18.11</td>
<td>22.25</td>
</tr>
<tr>
<td>5</td>
<td>9.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5.77</td>
<td>7.65</td>
<td>5.36</td>
</tr>
<tr>
<td>7</td>
<td>2.96</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV7</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.79</td>
<td>6.38</td>
<td>5.11</td>
</tr>
<tr>
<td>2</td>
<td>4.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>16.33</td>
<td>47.70</td>
<td>58.06</td>
</tr>
<tr>
<td>5</td>
<td>22.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>33.93</td>
<td>45.92</td>
<td>36.83</td>
</tr>
<tr>
<td>7</td>
<td>11.99</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV8</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>28.46</td>
<td>71.79</td>
<td>67.92</td>
</tr>
<tr>
<td>3</td>
<td>11.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>14.62</td>
<td>23.59</td>
<td>26.68</td>
</tr>
<tr>
<td>5</td>
<td>8.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.33</td>
<td>4.62</td>
<td>5.39</td>
</tr>
<tr>
<td>7</td>
<td>1.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B10BEHAV9</td>
<td>7-pt-scale</td>
<td>3-pt recoded</td>
<td>3-pt-scale orig</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>--------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>1</td>
<td>2.29</td>
<td>2.29</td>
<td>2.96</td>
</tr>
<tr>
<td>2</td>
<td>3.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.31</td>
<td>24.68</td>
<td>22.37</td>
</tr>
<tr>
<td>4</td>
<td>3.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>14.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>35.37</td>
<td>73.03</td>
<td>74.66</td>
</tr>
<tr>
<td>7</td>
<td>37.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV10</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>22.76</td>
<td>67.52</td>
<td>63.98</td>
</tr>
<tr>
<td>3</td>
<td>11.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>12.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7.93</td>
<td>28.39</td>
<td>28.76</td>
</tr>
<tr>
<td>6</td>
<td>8.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4.09</td>
<td>4.09</td>
<td>7.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV11</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.02</td>
<td>4.33</td>
<td>6.70</td>
</tr>
<tr>
<td>4</td>
<td>1.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.31</td>
<td>92.62</td>
<td>91.42</td>
</tr>
<tr>
<td>6</td>
<td>13.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>75.32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV12</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>45.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>30.61</td>
<td>75.77</td>
<td>74.53</td>
</tr>
<tr>
<td>3</td>
<td>8.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8.93</td>
<td>23.98</td>
<td>24.66</td>
</tr>
<tr>
<td>5</td>
<td>4.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.26</td>
<td>0.26</td>
<td>0.80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV13</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>44.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>29.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV14</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.29</td>
<td>2.29</td>
<td>1.61</td>
</tr>
<tr>
<td>2</td>
<td>1.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>15.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>38.42</td>
<td>67.43</td>
<td>71.77</td>
</tr>
<tr>
<td>7</td>
<td>29.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV15</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>16.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>15.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>12.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8.65</td>
<td>8.65</td>
<td>13.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV16</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>16.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>15.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B10BEHAV17</td>
<td>7-pt-scale</td>
<td>3-pt recoded</td>
<td>3-pt-scale orig</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>--------------</td>
<td>----------------</td>
</tr>
<tr>
<td>1</td>
<td>2.30</td>
<td>4.86</td>
<td>3.23</td>
</tr>
<tr>
<td>2</td>
<td>2.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.28</td>
<td>16.11</td>
<td>16.94</td>
</tr>
<tr>
<td>4</td>
<td>5.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>27.88</td>
<td>79.03</td>
<td>79.84</td>
</tr>
<tr>
<td>7</td>
<td>51.15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV21</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>75.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12.24</td>
<td>90.82</td>
<td>90.32</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>3.06</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.32</td>
<td>7.91</td>
<td>9.14</td>
</tr>
<tr>
<td>6</td>
<td>1.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.28</td>
<td>1.28</td>
<td>0.54</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV18</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.87</td>
<td>64.80</td>
<td>68.92</td>
</tr>
<tr>
<td>2</td>
<td>33.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12.50</td>
<td>33.93</td>
<td>28.65</td>
</tr>
<tr>
<td>4</td>
<td>11.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.28</td>
<td>1.28</td>
<td>2.43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV22</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38.62</td>
<td>60.10</td>
<td>61.56</td>
</tr>
<tr>
<td>2</td>
<td>21.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>9.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>19.95</td>
<td>36.32</td>
<td>34.14</td>
</tr>
<tr>
<td>5</td>
<td>6.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.81</td>
<td>3.58</td>
<td>4.30</td>
</tr>
<tr>
<td>7</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV19</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36.69</td>
<td>65.63</td>
<td>70.70</td>
</tr>
<tr>
<td>2</td>
<td>28.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11.89</td>
<td>31.52</td>
<td>25.27</td>
</tr>
<tr>
<td>4</td>
<td>10.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.84</td>
<td>2.84</td>
<td>4.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV23</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32.48</td>
<td>60.87</td>
<td>64.34</td>
</tr>
<tr>
<td>2</td>
<td>28.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>13.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>12.79</td>
<td>32.48</td>
<td>29.76</td>
</tr>
<tr>
<td>5</td>
<td>6.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5.37</td>
<td>6.65</td>
<td>5.90</td>
</tr>
<tr>
<td>7</td>
<td>1.28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV20</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.79</td>
<td>5.88</td>
<td>3.49</td>
</tr>
<tr>
<td>2</td>
<td>4.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5.88</td>
<td>37.34</td>
<td>48.26</td>
</tr>
<tr>
<td>4</td>
<td>13.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>17.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>30.95</td>
<td>56.78</td>
<td>51.74</td>
</tr>
<tr>
<td>7</td>
<td>25.83</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B10BEHAV24</th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.58</td>
<td>14.25</td>
<td>12.60</td>
</tr>
<tr>
<td>2</td>
<td>4.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>25.70</td>
<td>49.36</td>
<td>52.01</td>
</tr>
<tr>
<td>5</td>
<td>23.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>21.63</td>
<td>36.39</td>
<td>35.39</td>
</tr>
<tr>
<td>7</td>
<td>9.67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### B10BEHAV25

<table>
<thead>
<tr>
<th></th>
<th>7-pt-scale</th>
<th>3-pt recoded</th>
<th>3-pt-scale orig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.58</td>
<td>4.58</td>
<td>6.43</td>
</tr>
<tr>
<td>2</td>
<td>4.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10.18</td>
<td>64.12</td>
<td>63.27</td>
</tr>
<tr>
<td>4</td>
<td>25.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>23.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>21.63</td>
<td>31.30</td>
<td>30.29</td>
</tr>
<tr>
<td>7</td>
<td>9.67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Documentation biobirth

Generated birth history for men and women

Stefan Damerow
(Based on the SOEP Documentation by Joachim Frick and Christian Schmitt)
## Contents

**General information** ........................................................................................................ 149

**Information on data generation in biobirth** ................................................................. 149

**Variables in biobirth** ..................................................................................................... 152

<table>
<thead>
<tr>
<th>Variable</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIOYEAR</td>
<td>152</td>
</tr>
<tr>
<td>BIOAGE</td>
<td>152</td>
</tr>
<tr>
<td>BBSEX</td>
<td>152</td>
</tr>
<tr>
<td>SUMKIDS</td>
<td>152</td>
</tr>
<tr>
<td>BIOKIDS</td>
<td>153</td>
</tr>
<tr>
<td>NEWKIDS</td>
<td>153</td>
</tr>
<tr>
<td>KIDSOURCE[n]</td>
<td>153</td>
</tr>
<tr>
<td>KIDPNR[n]</td>
<td>153</td>
</tr>
<tr>
<td>KIDSEX[n]</td>
<td>154</td>
</tr>
<tr>
<td>KIDYOB[n]</td>
<td>154</td>
</tr>
<tr>
<td>KIDMOB[n]</td>
<td>154</td>
</tr>
<tr>
<td>KIDMAR[n]</td>
<td>154</td>
</tr>
<tr>
<td>KIDLITO[n]</td>
<td>154</td>
</tr>
</tbody>
</table>
**General information**

The data file *biobirth* provides a detailed overview of each individual’s birth history, i.e. an account of biological children a woman has born or a man has fathered. While this information is gathered through the biographical questionnaire on the one hand, children are on the other hand also born during the panel life and then captured through the regular person interviews of mother and/or father. The parent questionnaires and partnership information from *biocouply* provide an additional set of information. *biobirth* combines these different sources into one consistent dataset. Thus, *biobirth* can be described as an accumulative data set, which contains the entire birth biography of all FiD respondents.

The *biobirth* file is based on every person, who had at least one successful FiD interview. It includes all individuals turning 17 during the survey year. *biobirth* covers the following information:

1. birth year and sex for each biological child up to the last date of interview
2. sum of children captured over the panel, split by their source – biography or other
3. person identifier (KIDPNR) for each child, provided the child was ever living within a FiD household
4. marital status and living circumstances (related to the biological mother/father) at the time of birth

**Information on data generation in biobirth**

As mentioned above, there are various sources for the birth history. The variable KIDSOURCE[n] provides information on which source was used to generate the information for every of the respondent’s children.

The main basis of the individual birth biography in *biobirth* is the information collected by the biography questionnaire. It provides information on the number of children, the birth date and the sex of each respondent’s biological children. In addition, the relationship status at the time of each child’s birth is recorded for every respondent. The biographical information is

---

12 Up to FiD v2.0, the information where a child was living was included as well (variable KIDHOME). However, starting with distribution v2.1, this information was dropped, because it cannot be kept up to date. The information remains available in the *Slela* files. Whether a child still lives in the same household can be gathered using the yearly *Sprutto* datasets.

13 While the wave specific files *Skind* present the social and thus time-dependent parent-child relationships for children aged 16 or younger in the household, *biobirth* documents only biological parent-child relationships. This is the major difference to the SOEP version of this dataset, where the distinction between biological and adopted children is not possible.

14 This information is not part of the SOEP distribution of *biobirth*, as it is not asked in the questionnaire.
covered in the data set \textit{Slela}, filled out by each respondent once during his/her panel life. The birth history information is collected in the first part of the FiD \textit{Slela} questionnaires ($\text{SLELTYP}=1$ or $3$).

The \textit{Slela} data sets do not contain the person number for children from the biographical information who are part of FiD (i.e. living in a household belonging to the FiD-sample). The dataset \textit{Skind} provides the person number of the mother and father (MOTHNOS/FATHNOS) for every child in FiD. In combination with the indicator MOTHPS/FATHPS, biological child-parent relations can be identified to assign the child’s person number (KIDPNR[n]) in \textit{biobirth}. Since \textit{Skind} includes only children under 17 living in the household, the process generating these variables is extended to the whole data (see documentation of \textit{Skind} for detailed information). Note that due to additional information with every new wave, the indicator MOTHPS/FATHPS from previous waves may change in rare cases. In this case the birth biography is updated retrospectively.

The generating process of MOTHNOS/FATHNOS and MOTHPS/FATHPS is based not only on information from the birth biography. Parent – child relations are additionally found by the various parent questionnaires and in the variable $\text{STELL}$ (relationship to household head) from \textit{Sprbrutto} alone and in combination with partnership information from \textit{biocouply}. As mentioned, the differences are stored in KIDSOURCE[n] for every child.

For children found by other sources except the birth biography, the information in KIDLITO[n] and KIDMAR[n] need to be obtained from other data. The dataset \textit{biocouply} describes the partnership history of the respondent, so that it is possible to determine periods of marriage and coupled households. In case the child was born during a marriage and/or a partnership (same household), KIDLITO[n] and KIDMAR[n] are specified respectively.

The variable BIOKIDS sums up all respondent’s children found by biography information while NEWKIDS covers the other sources.

The youth questionnaire is relevant for all young adults turning 17 during the survey year. This questionnaire is the first individual questionnaire a FiD-member answers for herself. In terms of \textit{biobirth}, it is important to note that the youth questionnaire – which itself does not contain any relevant information for \textit{biobirth} – is answered instead of the biographical questionnaire. Assuming that only very few individuals give birth or father children before the age of 17, and that they can be identified within the household context (as long as they remain within FiD), this does not pose any problems for compiling the birth-biography of the respondents.
Once the children are found through the various ways, they are ordered chronologically, starting with the oldest child. In case of twins, the child with the lower person identifier is put first. If the twins are not part of the household, the order the respondent chose when reporting the children is maintained.

Finally, information on the birth history can be found through updates from the yearly questionnaire. Each individual is asked every year, whether a child “has come/born” into the household. With the help of the parent questionnaires and the previous observations, it is possible to identify new-born children in a household and relate them to their parents, and add information on the variables KIDLITO[n] and KIDMAR[n].
Variables in biobirth

BIOYEAR
Variable label “Year of biography survey”
Value labels BIOYEAR
Variable format 4-digit integer
Comment In BIOYEAR, “-2 Does not apply” is given to individuals without the biographical information and for those who have answered the youth questionnaire instead of the biographical questionnaire. Accordingly the variables BIOAGE and BIOKIDS are set to “-2 Does not apply”. Except for respondents with children from other sources, the same missing code is assigned to SUMKIDS and the children related variables (KIDPNR[n], KIDSEX[n], KIDYOB[n], KIDMOB[n], KIDMAR[n], KIDLITO[n], KIDSOURCE[n]).

BIOAGE
Variable label “Age when surveyed biography”
Value labels BIOAGE
Variable format 2-digit integer
Comment For women/men who have not filled out the birth biographical questionnaire (yet), the code “-2 Does not apply” is used. Similarly, individuals entering the dataset via the youth questionnaire are assigned this missing code.

BBSEX
Variable label “Gender of respondent”
Value labels SEX
(1) male
(2) female
Variable format 1-digit integer

SUMKIDS
Variable label “Total number of births”
Variable format 1-digit integer
Comment The variable is the total number of children identifiable within FiD by combining all available data up to the time of the last observation (SUMKIDS=BIOKIDS+NEWKIDS). If children are not mentioned by the respondent, or the source of information is not available (missing biographical information) the information in this variable may be underestimated. This is especially true for those respondents who did not fill in the biographical information.
**BIOKIDS**

Variable label: “Number of births from biography”
Variable format: 2-digit integer

Comment: For women/men who have not filled out the birth biographical questionnaire (yet), the code “-2 Does not apply” is used. Similarly individuals entering the dataset via the youth questionnaire are assigned this missing code.

**NEWKIDS**

Variable label: “Number of births not from biography”
Variable format: 2-digit integer

Comment: NEWKIDS provides the number of children identified through sources other than the birth biography. Hence all children born after the birth biography are stored here.

**KIDSOURCE[n]**

Variable label: “Source of birth biography”

Value labels: KIDSOURCE[n]
(1) birth biography
(2) parent questionnaire (E1 – E3 direct)
(3) parent questionnaire (E5/E6 direct)
(4) parent questionnaire (E1 – E3 indirect)
(5) biocouply
(6) $stell-variable ($pbrutto)

Variable format: 1-digit integer

Comment: The parent questionnaires contain direct and indirect information on the relationship to the child. On the one hand the respondent gives info about the own relationship (direct) to the child and furthermore his/her partner (indirect) when in household. The different sources in KIDSOURCE[n] can be ranked in relation to the quality of information. The best source is described by the direct info from parent questionnaire (E1 – E3, E5/E6) followed by the birth biography, then the indirect information, then biocouply and the information from the $STELL variable on the relationship to the household head from $pbrutto.

**KIDPNR[n]**

Variable label: “Persnr [n]th child”

Variable format: 8-digit integer

Comment: Provides the identifier for the first child [01] up to the fifteenth child [15], if the child is included in the FiD-sample, i.e. living in one of the FiD-households. If not, i.e. the child has already left the household, is
living with a different parent, or has already died, and has never been part of FiD, this variable is set to “-2 Does not apply”.

**KIDSEX[n]**
Variable label  “Sex [n]th child”
Value labels  KIDSEX[n]
   (1) male
   (2) female
Variable format  1-digit integer
Comment  n runs from 01 to 15.

**KIDYOB[n]**
Variable label  “Year of birth child [n]”
Variable format  4-digit integer
Comment  n runs from 01 to 15. If inconsistencies occur between $lela$ and $spfad$, $spfad$ is used as reference.

**KIDMOB[n]**
Variable label  “Month of birth of child [n]”
Variable format  2-digit integer
Comment  n runs from 01 to 15. If inconsistencies occur between $lela$ and $spfad$, $spfad$ is used as reference.

**KIDMAR[n]**
Variable label  “Married at birth [n]th child”
Value labels  KIDMAR[n]
   (1) Yes
   (2) No
Variable format  1-digit integer
Comment  n runs from 01 to 15.

**KIDLITO[n]**
Variable label  “Lived with partner at birth [n]th child”
Value labels  KIDLITO[n]
   (1) Yes
   (2) No
Variable format  1-digit integer
Comment  n runs from 01 to 15.
Note that for the variables KIDPNR[n], KIDSEX[n], KIDYOB[n], KIDMOB[n], KIDMAR[n] and KIDLITO[n] identical missing codes apply. The code “-2 Does not apply” is assigned, if there is no [n]th child found for this person. The code “-1 No answer” is used if information about the [n]th child is found but the specific information is missing.

For every woman/man a maximum of 15 entries for children is provided, although the biography questionnaire enables only ten possible entries regarding birth information. If there have been additional births up to the time the biography questionnaire is collected, they are recorded separately by the interviewer and are included in biobirth. The sequence of children within biobirth is recorded with regards to the age of the children. The oldest child is recorded under KIDPNR01 the second oldest under KIDPNR02 and so forth.
Documentation *biomarsy* and *biocouply*

**The Marital and Couple History Variables**

Juliana Werneburg / Stefan Damerow

This documentation is based on the comparable SOEP documentation *biomarsy* and *biomarsm* and has benefited from the work by Olaf Groh-Samberg and Florian R. Hertel. For readability reasons, we do not specifically cite and specify text that has been used directly from the SOEP document.

---

Contents

General Information ............................................................................................................... 158
Comparison with SOEP ......................................................................................................... 158
biocouply: A yearly couple biography ................................................................................ 158
Variables in biocouply ......................................................................................................... 160
Sources of the couple history ............................................................................................. 163
Construction of couple history ........................................................................................... 167
biomarsy: A yearly marital biography ................................................................................. 172
Variables in biomarsy ......................................................................................................... 173
Construction of marital histories ........................................................................................ 174
**General Information**

“Familien in Deutschland” (FiD) provides individual marital and partnership histories in two data files, *biomarsy* and *biocouply*. *Biomarsy* is a spell dataset containing annual spells of the individuals’ marital status, while *biocouply* provides annual spells on the partnership status. Both files comprise data on marital and couple biographies of respondents with a personal interview and additionally of adults living in an interviewed household without their own interview.

Both files contain whole relationship biographies starting at the year of birth. Thus, they mainly include retrospective information. In addition, they are extended by information given in every subsequent personal interview. The data file *biomarsm*, which is known to SOEP users, is not provided in FiD, because there is only very little monthly information on relationship histories available yet. Additionally, the marital and couple status in the *Spgen* data files, stored as MARRST$$ and COUPST$$, are derived from *biocouply* for the time of the interview.

This documentation proceeds with a brief description of the two data files *biocouply* and *biomarsy*. Users interested in more details may read further on how information on couple histories was collected in FiD and on the editing process of constructing logically consistent marital and couple histories.

**Comparison with SOEP**

FiD gathers more detailed information on a person’s former relationships than the SOEP did up to 2011. For that reason, instead of only providing marital histories, FiD extended the data generation to whole couple histories provided in the dataset *biocouply*. For the first time, the starting year of a relationship – independent of a later marriage – is available. In addition, a specific identification number for couples, COUPID, was created which will ease data handling for longitudinal analyses on relationships. Couple histories differentiate between being coupled vs. being married as well as between living together vs. living apart. Furthermore, separation periods for former couples who stay married afterwards can be analyzed. For convenience reasons the known dataset *biomarsy* is made available for FiD as well.

**biocouply: A yearly couple biography**

The spells in the data file *biocouply* contain retrospectively collected information on couple history since a respondent’s year of birth on an annual basis. FiD’s retrospective part of the person questionnaire covers up to four relationships in addition to the current status.
Information provided in the dataset is not only made available for respondents themselves. For most adults living in the same household, even if they were not interviewed, the current couple status can be reconstructed. In the case a respondent reported a relationship with a non-interviewed person living in the same household, information was copied to the non-responding partner. Persons aged 17 and interviewed only via Youth Questionnaire are not included in any of the datasets.

The data file contains thirteen variables: the household, couple and individual identifiers HHNR, COUPID and PERSNR as well as ten spell specific variables. The variable SPELLTYP documents the couple status with the possible categories ‘married, spouse in household’, ‘married, spouse not in household’, ‘coupled, partner in household’, ‘coupled, partner not in household’, ‘single’, ‘separated’, ‘registered same-sex partnership, living together’, ‘registered same-sex partnership, living separately’. For same-sex partnerships it is only asked since wave 2 (2011), thus, this information cannot be reconstructed retrospectively. ‘Separated’ spells apply only to former couples who are still married, i.e. who are separated but not yet divorced. Note that ‘separated’ is a redundant spell in terms of completeness of couple histories: it always overlaps with other spells that contain the actual couple status(es) over the entire separation episode. Hence, by deletion of all separation spells the seamless couple history is preserved. The additionally assigned codes ‘unknown’ and ‘unit nonresponse’ indicate a lack of information for the respective period.

The variable SPELLNR is a chronological index number for each individual’s spell during the observation period. The variables BEGINY and ENDY provide the years in which a spell begins and ends, whereas the variables BEGIN and END indicate respondent’s age for users’ convenience. In biocouply, spell systems for each individual always start with the respondent’s birth. Due to the fact that the spells’ duration is measured in years, it is important to note that an individual may encounter several events in the same year. In this case the variable SPELLNR allows the user to order spells with respect to the respondent’s life course. The SPELLTYP of the first spell per definition is ‘single’.

In addition, the indicator variables PDEATH and DIVORCE are provided. PDEATH indicates whether a respective spell ends with the death of a person’s partner. Single and gap spells are assigned a “(-2)” (does not apply). Please note that indicator PDEATH is not restricted to married persons, thus does not only refer to widowhood. By summing up PDEATH over the person’s life course, the number of subsequent states of widowhood can be retrieved easily (be careful to change missings into system missing values if you plan to sum up these indicators column-wise). DIVORCE works in a similar fashion. It indicates whether
the last marriage spell, that is the separated spell, ended in divorce. Hence, if it did not end in divorce it is coded as a still ongoing marriage. In this case ENDY is updated by year of last interview.

**Variables in biocouply**

The following provides an overview of the variables in *biocouply*. Missing codes are not listed separately, as the usual conventions apply: “-1” refers to missing answers (don’t know or refusal), “-2” indicates, that the question did not apply, whereas “-3” specifies implausible answers. There are some missing values (-1), (-2) or (-3) in BEGINY as well as in ENDY (BEGIN and END) indicating that we do not know the exact year of change in the couple status. Missing dates may simply indicate that the year was either not reported (-1), not asked for (-2) or implausible (-3), i.e. contradictory to other information. A gap might also have occurred: a) due to unit nonresponse or b) lack of substantial detail on an announced relationship or c) because it was not asked for. In order to differentiate the reasons for missing information the user can utilize variables REMARK and CENSOR.

The variable REMARK provides information on whether we had to edit or supplement original information provided by respondents in order to construct consistent couple biographies. Spells in *biocouply* are marked as ‘edited’ (in contrast to ‘original’) if the editing process involved substitution of or additions to original information as reported in the questionnaire. This happened, for example, if a respondent failed to report a relationship but valid information from a partner was available. Similarly, we inserted a divorce between two marriages, even if it was not specified, just because two marriages to separate persons at the same time are not legal (see later section for more details on editing). Furthermore, ‘first spells’ and ‘gap spells’ are marked separately. Note that inserted ‘single’ episodes between two consecutive reported relationships are edited as ‘original spell’.

The variable CENSOR indicates whether a spell is left-censored, right-censored or censored on both sides. Furthermore, there is information included for the reasons of censoring. In principle, spells might be censored if they precede or follow a gap spell or if BEGIN or END is missing. The last spell for each person is marked as right-censored if a person is still in FiD and the current marital status is open (‘last spell’).

**Table 1: Coding of variable CENSOR in biocouply**

<table>
<thead>
<tr>
<th>Left: not censored</th>
<th>Right:</th>
<th>not censored</th>
<th>censored missing</th>
<th>censored before gap</th>
<th>censored last spell</th>
<th>censored death</th>
</tr>
</thead>
<tbody>
<tr>
<td>not censored</td>
<td></td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
### COUPID

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Couple identifier”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable format</td>
<td>4-digit integer</td>
</tr>
</tbody>
</table>

### SPELLNR

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Consecutive Spell Number”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable format</td>
<td>2-digit integer</td>
</tr>
<tr>
<td>Comment</td>
<td>Range from (1) to (20)</td>
</tr>
</tbody>
</table>

### SPELLTYP

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Type of spell/event”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value labels</td>
<td>SPELLTYP</td>
</tr>
<tr>
<td>(1) married, spouse in household</td>
<td></td>
</tr>
<tr>
<td>(2) married, spouse not in household</td>
<td></td>
</tr>
<tr>
<td>(3) coupled, partner in household</td>
<td></td>
</tr>
<tr>
<td>(4) coupled, partner not in household</td>
<td></td>
</tr>
<tr>
<td>(5) single</td>
<td></td>
</tr>
<tr>
<td>(6) separated</td>
<td></td>
</tr>
<tr>
<td>(7) registered same-sex partnership, living together</td>
<td></td>
</tr>
<tr>
<td>(8) registered same-sex partnership, living separately</td>
<td></td>
</tr>
<tr>
<td>(98) unknown</td>
<td></td>
</tr>
<tr>
<td>(99) unit nonresponse</td>
<td></td>
</tr>
<tr>
<td>Variable format</td>
<td>2-digit integer</td>
</tr>
</tbody>
</table>

### BEGINY

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Year spell begins”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable format</td>
<td>4-digit integer</td>
</tr>
<tr>
<td>Comment</td>
<td>Range from 1924 to 2013</td>
</tr>
</tbody>
</table>

### ENDY

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Year spell ends”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable format</td>
<td>4-digit integer</td>
</tr>
<tr>
<td>Comment</td>
<td>Range from 1939 to 2013</td>
</tr>
</tbody>
</table>

### BEGIN

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Age spell begins”</th>
</tr>
</thead>
</table>

---

**SOEP Survey Paper 332**

161
Variable format 2-digit integer
Comment Range from 0 to 86

END

Variable label “Age spell ends”
Variable format 2-digit integer
Comment Range from 6 to 86

PDEATH
Variable label “Death indicator”
Value labels PDEATH
(0) spell does not end with death of partner
(1) spell ends with death of partner
Variable format 1-digit integer

DIVORCE
Variable label “Divorce indicator”
Value labels DIVORCE
(0) spell does not end with divorce
(1) spell ends with divorce
Variable format 1-digit integer

CENSOR
Variable label “Censoring information”
Value labels CENSOR
(0) not censored
(1) LC no beginy
(2) LC after gap
(3) RC no endy
(4) RC before gap
(5) RC last spell
(6) RC death
(7) LC+RC no beginy a. no endy
(8) LC+RC no beginy a. before gap
(9) LC+RC no beginy a. last spell
(10) LC+RC no beginy a. death
(11) LC+RC after gap a. no endy
(12) LC+RC after gap a. before gap
(13) LC+RC after gap a. last spell
(14) LC+RC after gap a. death
(See Table above for detailed description on the censoring codes.)
Variable format 2-digit integer
REMARK
Variable label “Error Code”
Value labels REMARK
(1) original spell
(2) edited spell
(3) gap spell
(4) first spell
Variable format 1-digit integer

Sources of the couple history
For the construction of individual couple histories, information on the biography from the Biography Questionnaire $lela$ was mostly used. Figures 1 to 5 show those parts of the Biography and Person Questionnaire for the survey year 2011 which aimed at collecting respondents’ couple history. In addition, the couple identifier COUPID and the generated partner pointer PARTP$$ of the generated dataset $pgen$ were used to link current partners living in the same household, to compare their answers, and supplement them if necessary. Only long-term relationships, defined as lasting for at least six months, should be mentioned in the questionnaire (see Figure 1). For those interviewed again any changes that occurred during the last year were asked for (see Figure 2). From the Biography Questionnaire we obtain information on the current (see Figure 3) and up to three other marriages or relationships (see Figure 4) or three relationships plus one marriage (see Figure 5) that took place prior to the interview. Hence, all in all, up to five relationships are possible to record.

Figure 1: Introduction to marriage part of the Biography Questionnaire (translated)

The following part deals with relationships and marriage

All questions are related to relationships no matter whether you are married or not.

Whether you married in this relationship will be asked later.

These questions ask about both your current and your previous relationships.

We start with your current relationship. Afterwards, we ask retrospectively for the long-term relationship you had previously to the last-mentioned.

Concerning previous relationships, we call them long-term, if they lasted at least six months or longer.

When we ask about the previous long-term relationship, please always consider the relationship you had before the last mentioned, which lasted at least six months.
The amount of information collected about the different relationships – the current, the second to fourth previous relationship, and a former marriage if applicable – is not identical. Only in 2010, the year of moving together was not retrieved for the current relationship. For this reason, the year of moving together was estimated by the year of moving into the current household for current relationships. In those cases where both partners responded and gave different answers on the year of moving in, the most recent year was chosen as the date of moving in.

**Figure 2:** Questions on changes during last year (translated)

<table>
<thead>
<tr>
<th>Event</th>
<th>Yes</th>
<th>2011</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>in month</td>
<td>in month</td>
</tr>
<tr>
<td>Started a new relationship</td>
<td>☐</td>
<td>…☐☐</td>
<td>…☐☐</td>
</tr>
<tr>
<td>Got married</td>
<td>☐</td>
<td>…☐☐</td>
<td>…☐☐</td>
</tr>
<tr>
<td>Moved in with my partner</td>
<td>☐</td>
<td>…☐☐</td>
<td>…☐☐</td>
</tr>
<tr>
<td>I separated from my spouse / partner</td>
<td>☐</td>
<td>…☐☐</td>
<td>…☐☐</td>
</tr>
<tr>
<td>I got divorced</td>
<td>☐</td>
<td>…☐☐</td>
<td>…☐☐</td>
</tr>
<tr>
<td>My spouse / partner died</td>
<td>☐</td>
<td>…☐☐</td>
<td>…☐☐</td>
</tr>
</tbody>
</table>

**Figure 3:** Questions on the current relationship (translated, Person Questionnaire 2011)
<table>
<thead>
<tr>
<th>Question</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>L43. What about the present: Do you currently have a long-term relationship?</td>
<td>Yes...☐ (→ L44)  No...☐ (→ Go to question L53)</td>
</tr>
<tr>
<td>L44. When did the relationship with this partner start?</td>
<td>Year ☐☐☐☐</td>
</tr>
<tr>
<td>L47. Does your partner live in this household?</td>
<td>Yes...☐ (→ L48)  No...☐ (→ Go to question L49)</td>
</tr>
<tr>
<td>L47a. When did you move in with your partner? [Asked in 2011 only]</td>
<td>Year ☐☐☐☐</td>
</tr>
<tr>
<td>L48. Please tell the first name of your partner.</td>
<td>____________________</td>
</tr>
<tr>
<td>L49. Did you live together with that partner in the past?</td>
<td>Yes...☐ (→ L50)  No...☐ (→ Go to question L45)</td>
</tr>
<tr>
<td>L50. When did you give up your common accommodation or when did you or your partner move out?</td>
<td>Year ☐☐☐☐</td>
</tr>
<tr>
<td>L45. Are you married to this partner?</td>
<td>Yes...☐ (→ L46)  No...☐ (→ Go to question L51)</td>
</tr>
<tr>
<td>L46. When did you marry?</td>
<td>Year ☐☐☐☐</td>
</tr>
</tbody>
</table>

**Figure 4:** Questions on previous relationships (translated)
Now we cover your previous relationship – that is the one prior to the current / the last mentioned relationship.  
*(Analogously: 2nd to 4th relationship)*

<table>
<thead>
<tr>
<th>Question</th>
<th>Description</th>
</tr>
</thead>
</table>
| L53. | Did you have a previous long-term relationship?  
Yes...☐ (→ L54)  
No...☐ (→ Go to question L63) |
| L54. | When did that relationship start?  
Year ☐☐☐☐ |
| L55. | When and in which way did that relationship end?  
Year ☐☐☐☐  
Through separation...................☐ (→ L56)  
Through death.......................☐ |
| L56. | Did you live with this partner?  
Yes...☐ (→ L57)  
No...☐ (→ Go to question L59) |
| L57. | When did you move in with this partner?  
Year ☐☐☐☐ |
| L58. | And when did you give up your common accommodation or when did you or your partner move out?  
Year ☐☐☐☐  
The common domicile was not vacated..................☐ |
| L59. | Did you marry this partner?  
Yes...☐ (→ L60)  
No...☐ (→ Go to question L63) |
| L60. | When did you marry?  
Year ☐☐☐☐ |
| L61. | Did you get a divorce?  
Yes...☐ (→ L62)  
No...☐ (→ Go to question L63) |
| L62. | When were you divorced?  
Year ☐☐☐☐ |
moving in together. Information on a previous marriage is just elementary (see also Figure 5). Moreover, it was asked for only in the case that three relationships had already been mentioned and none of these were marriages.

**Construction of couple history**

Information on couple history mainly stems from respondents’ retrospective reports on their own history. Thus, no other benchmark on the substance of these reports exists. This led to inconsistencies – for example, overlaps of two relationships occurred, or the ordering of becoming a couple, moving together and marrying was reported differently than expected. Even though many different patterns of the reported histories occurred, sometimes with unusual appearances, they are still possible for the most part. Hence, no verified decision between measurement error and uncommon reality can be made. For that reason, whenever possible, couple histories are left as they were reported. Only in rare cases, restrictions, corrections on orderings of events or changes of reported years are conducted (read the following carefully for details). Little original information was changed (note: to retrieve original data, see the data file originating from the Biography Questionnaire, $lela$). Further corrections to smooth out irregularities are thus left to the user. Note that consistency checks between waves are done as well. That way, changes between data distributions are possible for former waves.
A clear ordering of spells needs to be sustained if information on timing is missing. Besides, information was asked on an annual basis. Thus, ordering of those spells cannot be done empirically, but has to be decided in advance. For that reason the following default rules were obeyed to obtain logically consistent histories if no other information forced to do otherwise:

1. Every individual couple history starts with the state ‘single’. In general it is assumed to last at least until the age of 15. However, we did not restrict age within a relationship, that is, unmarried relationships are allowed to start anytime between the age of 0 and 15. We restricted the age of marriage to be at least 15, though.

2. Every spell set for a certain couple starts with the state ‘coupled, partner not in household’. One exception exists: if respondents report a year of moving together that lies before the start of their relationship, this specific couple history starts with ‘coupled, partner in household’. Note that in this case the information when this couple moved together is not available in biocouply anymore. You would have to look it up in the original source stored in the data file Slela.

3. If there is no evidence to the contrary, it is assumed that married couples live together and moved together before marriage. That is: for married couples their specific couple history starts with ‘coupled, partner not in household’ and is followed by a spell ‘coupled, partner in household’ before their marriage spell ‘married, spouse in household’ starts. Thus, if a couple moved together in the same year they married, a spell ‘coupled, partner in household’ is included anyway. Additionally, this assumption applies to a marriage reported as the fourth/fifth reported relationship (see Figure 4). Note that dates of becoming a couple and moving together or whether they moved together at all are not known in that case.

4. To ensure that any information possible is included into the spell dataset, a spell ‘married, spouse not in household’ is created even if the date of moving out and the end of a relationship fall into the same year or if either of the two dates is not known. Likewise, if a partner moves out of the joint household in the year of marriage, the spell ‘married, spouse in household’ is included anyway, even though it may be redundant. Note that the date of moving out is lost if it was reported to come after the end of a relationship. (It can still be retrieved from the original source stored in the data file Slela.)

5. Any (formerly) married couple that is not an active relationship anymore, i.e. married who are separated but not yet divorced, ends with a spell ‘separated’. As long as they are not divorced yet, the end date of those separated spells is the same as their last
interview year. Again, one exception exists: the separation spell is not added if marriage clearly ended with the death of the respondent’s partner. If not known, a ‘separated’ spell is added. Note that ‘separated’ is a redundant spell in terms of couple history: it always and fully overlaps with other spells that contain the actual couple status(es) over the entire separation episode. Hence, by deletion of all separation spells, the seamless couple history is preserved. It should be noted that information on how a marriage ended – via death or divorce – is saved in the separation spells in PDEATH and DIVORCE, hence care should be applied when deleting those spells.

6. Because biocouply documents couple statuses and not marital statuses, it is possible to become single after marriage (in biomarsy the only possible change from ‘divorced’ or ‘widowed’ is to ‘married’).

7. As it was possible to mention another relationship in the questionnaire (up to four were allowed, see above), it is assumed that periods between those relationships mentioned were ‘single’ states and thus filled accordingly. Note that this relates to long-term relationships only.

8. Information on the death of a partner or a divorce is stored in variables PDEATH and DIVORCE. If not applicable, that is if a person is currently single, PDEATH and DIVORCE are set to “-2”, does not apply. If a respective couple is not married, DIVORCE is set to “-2”, does not apply, as well. (Be careful to change missings into system missing values if you plan to sum up these indicators column-wise.)

For many non-interviewed persons information on the current relationship can be reconstructed, because the respective partner is available from his or her personal interview. Thus, those non-interviewed adults remain in the biocouply dataset, even though information is not directly retrieved from them. If information on their current couple status was given by their partner, it is fully copied to these non-respondents as well. Concerning their prior history, the previously mentioned rules are followed: from age 0 to 15 they are stated ‘single’. From age 15 until the current relationship a gap – a spell with SPELLTYP ‘unit nonresponse’ (99) – is included. Non-interviewed persons, for whom no information is available get a gap – a spell with SPELLTYP ‘unit nonresponse’ (99) – from age 15 onwards.

Consistency checks are possible in several ways: a) contradictions between information given by a respondent and her relationship to the head of household (see $PSTELL in the dataset Spbrutto). This information was given by the head of household in the Household interviewed. Again, one exception exists: the separation spell is not added if marriage clearly ended with the death of the respondent’s partner. If not known, a ‘separated’ spell is added. Note that ‘separated’ is a redundant spell in terms of couple history: it always and fully overlaps with other spells that contain the actual couple status(es) over the entire separation episode. Hence, by deletion of all separation spells, the seamless couple history is preserved. It should be noted that information on how a marriage ended – via death or divorce – is saved in the separation spells in PDEATH and DIVORCE, hence care should be applied when deleting those spells.

6. Because biocouply documents couple statuses and not marital statuses, it is possible to become single after marriage (in biomarsy the only possible change from ‘divorced’ or ‘widowed’ is to ‘married’).

7. As it was possible to mention another relationship in the questionnaire (up to four were allowed, see above), it is assumed that periods between those relationships mentioned were ‘single’ states and thus filled accordingly. Note that this relates to long-term relationships only.

8. Information on the death of a partner or a divorce is stored in variables PDEATH and DIVORCE. If not applicable, that is if a person is currently single, PDEATH and DIVORCE are set to “-2”, does not apply. If a respective couple is not married, DIVORCE is set to “-2”, does not apply, as well. (Be careful to change missings into system missing values if you plan to sum up these indicators column-wise.)

For many non-interviewed persons information on the current relationship can be reconstructed, because the respective partner is available from his or her personal interview. Thus, those non-interviewed adults remain in the biocouply dataset, even though information is not directly retrieved from them. If information on their current couple status was given by their partner, it is fully copied to these non-respondents as well. Concerning their prior history, the previously mentioned rules are followed: from age 0 to 15 they are stated ‘single’. From age 15 until the current relationship a gap – a spell with SPELLTYP ‘unit nonresponse’ (99) – is included. Non-interviewed persons, for whom no information is available get a gap – a spell with SPELLTYP ‘unit nonresponse’ (99) – from age 15 onwards.

Consistency checks are possible in several ways: a) contradictions between information given by a respondent and her relationship to the head of household (see $PSTELL in the dataset Spbrutto). This information was given by the head of household in the Household interview year. Again, one exception exists: the separation spell is not added if marriage clearly ended with the death of the respondent’s partner. If not known, a ‘separated’ spell is added. Note that ‘separated’ is a redundant spell in terms of couple history: it always and fully overlaps with other spells that contain the actual couple status(es) over the entire separation episode. Hence, by deletion of all separation spells, the seamless couple history is preserved. It should be noted that information on how a marriage ended – via death or divorce – is saved in the separation spells in PDEATH and DIVORCE, hence care should be applied when deleting those spells.

6. Because biocouply documents couple statuses and not marital statuses, it is possible to become single after marriage (in biomarsy the only possible change from ‘divorced’ or ‘widowed’ is to ‘married’).

7. As it was possible to mention another relationship in the questionnaire (up to four were allowed, see above), it is assumed that periods between those relationships mentioned were ‘single’ states and thus filled accordingly. Note that this relates to long-term relationships only.

8. Information on the death of a partner or a divorce is stored in variables PDEATH and DIVORCE. If not applicable, that is if a person is currently single, PDEATH and DIVORCE are set to “-2”, does not apply. If a respective couple is not married, DIVORCE is set to “-2”, does not apply, as well. (Be careful to change missings into system missing values if you plan to sum up these indicators column-wise.)

For many non-interviewed persons information on the current relationship can be reconstructed, because the respective partner is available from his or her personal interview. Thus, those non-interviewed adults remain in the biocouply dataset, even though information is not directly retrieved from them. If information on their current couple status was given by their partner, it is fully copied to these non-respondents as well. Concerning their prior history, the previously mentioned rules are followed: from age 0 to 15 they are stated ‘single’. From age 15 until the current relationship a gap – a spell with SPELLTYP ‘unit nonresponse’ (99) – is included. Non-interviewed persons, for whom no information is available get a gap – a spell with SPELLTYP ‘unit nonresponse’ (99) – from age 15 onwards.

Consistency checks are possible in several ways: a) contradictions between information given by a respondent and her relationship to the head of household (see $PSTELL in the dataset Spbrutto). This information was given by the head of household in the Household
Questionnaire and processed in the partner pointer PARTP$. b) for the current relationship answers of both partners can be compared if a possible partner was identified via COUPID and both were interviewed; c) dates within a personal couple history might appear plausible or not. a) and b) are accomplished and single-case corrections done if reasonable. Concerning c) it is tried to leave as much of the original information as it is given by the respondent to allow for analyses of uncommon patterns as well. Some overlaps and inconsistencies might still appear in the dataset, which you may want to remove. Generally, inconsistencies were dealt with the following way:

1. If a respondent reported an early marriage in the Biography Questionnaire, we restricted its beginning to age 15. But we did not restrict age within a relationship, that is, unmarried relationships are allowed to start anytime between the age of 0 and 15.

2. Within a persons’ reported history, dates are not edited or overwritten (such as dates of marriage or begin of relationships). Thus, it is possible that a person is married more than once at the same time or has multiple, overlapping relationships in a certain period. That is, relationships do not need to be consecutive. The variable SPELLNR provides the sorting order, whereby the most recent spell comes last, that is, the relationship that ended most recently (no matter which are the starting dates). If no clear dates were available, the sorting in SPELLNR reflects the order of reporting by the respondent.

3. Within a specific couples’ reported history, dates are not edited or overwritten. E.g. it was not corrected if the start of a marriage was reported to be before the start of a relationship. This way 7 spells occur with a starting date coming after the end date of the spell. As mentioned above, it was accepted if the date of moving together was before the year of becoming a couple. In this case, this specific couple history starts with ‘coupled, partner in household’, while the information when this couple moved together is not available in biocouply anymore (it is available via the data file Slela).

4. Contradictions in both partners’ answers are left as they were, even if both were interviewed in the current wave. E.g. information on starting date or the date of marriage of their joint relationship often differs. As before, neither date is edited or overwritten between both partners of a current couple. To check for contradictions you can link them via couple identifier COUPID.

5. For very few cases, there were contradictions about the actual partnership between to individuals in the sample. E.g., if one part of a couple or the position to the head of household suggested a couple in a household, but one or both answered to be single or
coupled with someone else (outside the household), it was tried to link them by COUPID in their respective former relationship. Note that those single case corrections are not flagged if they just concern the couple identifier COUPID.

Some respondents refused to answer some or all questions in the relationship part. To some extent, missing values and contradiction are similar in the generating sense. Hence, some of the following rules applied to missing values may seem familiar:

1. In some cases it is possible that a respondent might have had earlier relationships, but they could not be named anymore due to the above mentioned questionnaire restriction. Thus, it is not known whether there were other partners before the last mentioned. In these cases, gaps were introduced, that is a spell ‘unknown’ (98) was inserted.

2. If a respondent stated to be married in a relationship, but it is not known whether the couple lived together as well, the above mentioned default course is assumed. In some rare cases details on the relationship are not known, as just a relationship was reported. There thus exists no information whether the couple moved together or whether they were married. Here, gaps, (98) ‘unknown’, are introduced as well.

3. For current relationships with both partners living together, it is not known when they moved together. In these cases the year of moving together was substituted by the year of moving into the current household. If both partners were respondents and gave a differing answer on the year of moving in, the most recent year was chosen as the date of moving together. If the date of moving into the household comes before the year the relationship started, it is assumed that both partners moved together the very same year they became a couple. The user may decide herself whether a very long time span between becoming a couple and the estimated year of moving together might indicate that the couple already changed households together via moving.

4. Some interviewed persons did not report a current but – given other information – likely relationship. This contradiction is apparent either by the information of the related partner, by the relation to the head of household or seen in case-to-case checks, e.g. via common children identified through the parent questionnaires. For these cases, the relationship is included into the person’s relationship record with a default record fitting to the identified marital status, but all dates are set to (-1), ‘no answer’. The decision to copy the dates from the corresponding reporting partner is left to the user.
**biomarsy: A yearly marital biography**

The data file **biomarsy** provides retrospectively collected information on marital history starting with the respondent’s year of birth. **biomarsy** in FiD is fully comparable to the version given in the SOEP: data are presented in spell format on an annual basis. **biomarsy** is derived by collapsing the dataset **biocouply** while applying only minor changes to the format. Thus, both datasets are consistent with each other, where **biomarsy** is a subset of **biocouply**, focusing only on marriage, widowhood and divorce. Like **biocouply**, **biomarsy** includes both respondents and non-responding adults living in an interviewed household.

The **biomarsy** file comprises eleven variables, including the individual and household identifiers HHNR and PERSNR as well as SPELLTYP. The variable SPELLTYP documents the marital status as known from the SOEP. It has the possible categories ‘not married’, ‘married’, ‘divorced’, widowed or divorced’ and ‘gap’. Once married, a later spell ‘not married’ is not assigned anymore. Note that we renamed the known SOEP code “1” ‘single’ to ‘not married’. This is to indicate that it is possible the respondent might have a partner anyway. If you are interested in this information, we recommend using **biocouply** instead of **biomarsy**.

SPELLTYP has one category ‘divorced or widowed’ in **biomarsy**, which indicates that a marriage definitely ended, although we do not know whether this happened due to divorce or due to the death of the spouse, as this information is missing in the Biography Questionnaire. Because spell durations are measured in years, it is important to note that an individual may encounter several events in the same year. In this case the variable SPELLNR allows the user to order the spells with respect to the respondent’s life course. The variables BEGINY and ENDY provide the years in which a spell begins and ends, while the variables BEGIN and END indicate the respective age of the respondent for users’ convenience. The spell system for each individual in **biomarsy** always starts with the birth of the respondent. We thus created a first spell for each individual ever interviewed in the SOEP starting in the year of birth and continuing at least until the year in which a person reaches the age of 15. The SPELLTYP of the first spell per definition is ‘not married’. Even if a respondent reported an earlier marriage in the Biography Questionnaire, we restricted its beginning to age 15.

Variables REMARK and CENSOR are constructed as in **biocouply**. Please see the explanations above.

The sources of **biomarsy** are identical to those in **biocouply**. Please consult the documentation of **biocouply** for further details.
Variables in biomarsy

**SPELLNR**
- Variable label: “Consecutive Spell Number”
- Variable format: 1-digit integer
- Comment: Range from (1) to (8)

**SPELLTYP**
- Variable label: “Marital status”
- Value labels:
  1. not married
  2. married
  3. divorced
  4. widowed
  5. divorced or widowed
  6. registered same-sex partnership
  9. gap
- Variable format: 1-digit integer

**BEGINY**
- Variable label: “Year spell begins”
- Variable format: 4-digit integer
- Comment: Range from 1924 to 2013

**ENDY**
- Variable label: “Year spell ends”
- Variable format: 4-digit integer

**BEGIN**
- Variable label: “Age spell begins“
- Variable format: 2-digit integer
- Comment: Range from 0 to 86

**END**
- Variable label: “Age spell ends”
- Variable format: 2-digit integer
- Comment: Range from 6 to 86

**CENSOR**
- Variable label: “Censoring information”
- Value labels: CENSOR
(0) not censored
(15) LC no beginy
(16) LC after gap
(17) RC no endy
(18) RC before gap
(19) RC last spell
(20) RC death
(21) LC+RC no beginy a. no endy
(22) LC+RC no beginy a. before gap
(23) LC+RC no beginy a. last spell
(24) LC+RC no beginy a. death
(25) LC+RC after gap a. no endy
(26) LC+RC after gap a. before gap
(27) LC+RC after gap a. last spell
(28) LC+RC after gap a. death
(See Table above for detailed description on the censoring codes.)

Variable format 2-digit integer

REMARK
Variable label “Error Code”
Value labels REMARK
(1) original spell
(2) edited spell
(3) gap spell
(4) first spell

Variable format 1-digit integer

Construction of marital histories
Marital histories are derived directly from *biocouply*. Hence, the information used is identical to *biocouply*, although shorter and sometimes in a different format. The following rules hold:

1. Every individual marital history starts with the state ‘not married’. We did not allow a person to be married before age 15.
2. From ‘not married’, one can only change to ‘married’.
3. There is no possible return to ‘not married’ once a person was ever ‘married’. The only possible change from ‘married’ is to ‘divorced’ or ‘widowed’.
4. The only possible change from ‘divorced’ or ‘widowed’ is to ‘married’.
Documentation $kind$

Person-related variables on children (up to the age of 16) within the household

By Stefan Damerow and Moritz Mannschreck
(Based on the SOEP Documentation by Joachim Frick)
**General information**

The variables from the annual *Skind* files described in the following are not based on answers provided by the children themselves but by answers to the household questionnaire provided by the respondent within the household (head of household). This data are disaggregated on the person level and saved as child-specific entries in the file *Skind*.

The annual *Skind* datasets also contain additional information on institutional care, school attendance and extracurricular activities for children up to age 16. Questions on the costs of childcare as well as non-institutional care arrangements are also asked. These variables are not described here. The variable names correspond to the numeration in the household questionnaire.

The persons in *Skind* are identifiable with the Codes “120 to 126” in the wave-specific variable $NETTO (in the file ppfad).

**List of variables:**

- PSAMPLE .............................................................................................................................. 177
- SWELLE ................................................................................................................................ 177
- KGJAHR .............................................................................................................................. 177
- KGMON ............................................................................................................................... 177
- KSEX ................................................................................................................................... 177
- KSTELL ............................................................................................................................... 177
- KINHH ................................................................................................................................. 178
- HHGR ................................................................................................................................... 179
- KZAHL ................................................................................................................................ 179
- MOTHNO$ ............................................................................................................................ 180
- MOTHP$ ................................................................................................................................ 180
- FATHNOS ........................................................................................................................... 180
- FATHPS ................................................................................................................................. 181
- SCHLTYPES ......................................................................................................................... 182

**Appendix**

- Conversion of $STELL from 2012 to 2010/2011 Version .................................................... 182
- $STELL-Combinations to identify mother-child relationships ............................................. 184
PSAMPLE
Variable label
“Subsample”
Value label
PSAMPLE
(61) FiD 2007 Birth Cohort
(62) FiD 2008 Birth Cohort
(63) FiD 2009 Birth Cohort
(64) FiD 2010 Birth Cohort
(65) FiD Screening (sampled 2010)
(66) FiD Screening (sampled 2011)
Variable format
2-digit integer
Comment
Note that this variable is included in all datasets, and provides information whether the household or person originates from the cohort or the screening sample in FiD. In *ppfad* it is also named PSAMPLE, in *hpfad* it is named HSAMPLE; in all other datasets it is called SAMPLE1, which is analogous to the SOEP notation.

$WELLE$
Variable label
“Wave”
Variable format
4-digit integer
$- Wave$
F10, F11, F12, F13
Comment
This variable reports the survey year.

$KGJAHR$
Variable label
“Year of birth”
Variable format
4-digit integer
$- Wave$
F10, F11, F12, F13
Comment
Year of birth of the child (4-digit). The value of $KGJAHR$ can in some cases differ from the longitudinally tested data in the central file *ppfad* (variable GEBJAHR).

$KGMON$
Variable label
“Month of birth”
Variable format
4-digit integer
$- Wave$
F10, F11, F12, F13
Comment
The child’s month of birth. The value of $KGMON$ can in some cases differ from the longitudinally tested data in the central file *ppfad* (variable GEBMONAT).

$KSEX$
Variable label
“Sex”
Value label
$KSEX
(1) male
(2) female

Variable format  1-digit integer
$ - Wave  F10, F11, F12, F13

Comment  Child’s sex. The value of $KSEX can in some cases differ from the longitudinally tested data in the central file *ppfad* (variable SEX).

**$KSTELL**

Variable label  “Relationship to head of household”
Value label  $KSTELL

Waves F10 and F11
(0) Head Of Household (HH)
(1) Spouse of HH head
(2) Partner of HH head
(3) Son, daughter of HH head
(4) Foster child of HH head
(5) Son, daughter-in-law of HH head
(6) Father, mother of HH head
(7) Parent-in-law of HH head
(8) Sister, brother of HH head
(9) Grandchild of HH head
(10) Other relative of HH head
(11) Not related to HH head
(12) Stepchild of HH head
(13) Same-sex spouse
(99) Unknown

Wave F12 and F13
(0) Household (HH) head
(11) Spouse of HH head
(12) Same-sex spouse
(13) Life partner of HH head
(21) Son, daughter of HH head
(22) Stepchild of HH head
(23) Adopted child of HH head
(24) Foster child of HH head
(25) Grandchild of HH head
(26) Great-grandchild of HH head
(27) Son, daughter-in-law of HH head
(31) Father, other of HH head
(32) Stepmother, stepfather of HH head
(33) Adoptive mother, father of HH head
(34) Foster mother, father of HH head
(35) Parent-in-law of HH head
(36) Grandmother, gradfather of HH head
(41) Sister, brother of HH head
(42) Half-sister, -brother of HH head
(43) Stepsister, -brother
(44) Adopted sister, brother of HH head
(45) Foster sister, brother of HH head
(51) Sister-in-law/brother-in-law 1: spouse or life partner of HH head's sibling
(52) Sister-in-law/brother-in-law 2: sibling of HH head's spouse or life partner
(61) Aunt/uncle of HH head
(62) Niece/nephew of HH head
(63) Cousin of HH head
(64) Other relative of HH head
(71) Not related to HH head
(99) Unknown

Variable format 2-digit integer
$ - Wave F10, F11, F12, F13

Comment Copy of the data from $STELL from the wave-specific file Spbrutto. Starting in F12, $STELL was changed and includes more values than in previous waves, which allows more precise identifications of relationships within the households. The values (0)-(2) as well as (6)-(7) in 2010 and 2011 and (0)-(13) and (31)-(36) in 2012 and 2013 are by definition not a part of the file Skind.

$KINHH
Variable label “Household Membership”
Value label $KINHH
Variable format 2-digit integer
$ - Wave F10, F11, F12, F13

Comment Copy of the data from $PZUG from the wave-specific file Spbrutto. In Skind some of the value labels do not apply due to the age of child.

$HHGR
Variable label “Number of persons in HH”
Variable format 2-digit integer
$ - Wave F10, F11, F12, F13

Comment Copy of the variables $HHGR from the wave-specific files Shbrutto.

$KZAHL
Variable label “Number of children in the HH”
Variable format 2-digit integer
$ - Wave F10, F11, F12, F13

Comment For each child in Skind, the variable $KZAHL provides the total number of children up to age 16 in the current household
**MOTHONO$**

Variable label: “Never Changing Person ID (mother)”
Variable format: 8-digit integer
$\text{\$\$ - Survey Years: }$ $\text{\$\$=10, 11, 12, 13}$

Comment: Person ID number of the child’s mother. MOTHNO$ is comparable to the SOEP Variable $\text{\$KMUTTI}$, except that in addition to biological and adoptive mothers, it includes social and foster mothers (for definitions see $\text{\$MOTHP}$).

The identification of mother-child relation starts with the wave-specific files $\text{\$lela}$, in which the birth biography includes information about birth date, sex and the first name (not in data distribution) of the respondent’s children. Since this information is unique for all members of a household, it is possible to identify children in the wave-specific files $\text{\$pbrutto}$ and assign the child’s PERSNR. Within that procedure, problems occur if specific information is missing or obviously wrong in the biographical data. With the help of $\text{\$STELL}$ (= relationship to HH head in $\text{\$pbrutto}$, see appendix) potential children can be found. If a child could not be matched directly but is the “right” child, based on the name, birth date and sex, the person ID is matched. Furthermore, $\text{\$STELL}$ is used as the second source for the mother-child relation as $\text{\$lela}$ may not be available for every potential mother (see appendix).

To obtain the person ID of social mothers, the biological father’s partner is derived from the variable $\text{\$PARTNR\$}$ (=person ID number of spouse or partner), which is taken from the wave-specific files $\text{\$pgen}$ and the $\text{\$STELL}$ variable. Furthermore, the variable $\text{\$STELL}$ identifies foster mothers.

**MOTHP$**

Variable label: “Indicator for mother’s relationship to the child”
Value label: MOTHP$
- (1) biological
- (2) social
- (3) adoptive
- (4) foster
- (5) social, same-sex partners
- (6) unknown

Variable format: 1-digit integer
$\text{\$\$ - Survey Years: }$ $\text{\$\$=10, 11, 12, 13}$

Comment: Indicator for the mother-child relation in the household. As described in MOTHNO$ the mother-child relation are linked with the person ID of the mother and the child. Within the procedures of identification the indicators (for mothers and fathers) are defined according to the source as follows:

If data from the wave-specific file $\text{\$lela}$ is available, first, it is possible to differentiate between “biological” and “adoptive” (Variable $\text{\$l064g*}$ in $\text{\$lela}$). Secondly, the mother can be defined as the social mother of her partner’s biological children when she has not mentioned them in $\text{\$lela}$, but lives in the same household.
The identification by $STELL in the wave-specific file $sprutto can only be used to identify foster mothers, because it is not possible to make a difference between biological, adoptive and step-children in $STELL (see appendix). Thus, potential biological relations for whom information is only found in $STELL are defined as “unknown”. For further identification of biological mothers the individual parent questionnaires are used as they include information about motherhood (as generated variable BIOCHILD in files $bioage01, $bioage02, $bioage03, and $bioage06 and variable BIOPAR in file $bioage08 and $bioage10).

Finally the dataset biocouply gives information about the partnerships, specifically their beginnings. Unknown mother-child relations found by $STELL are replaced by the indicator “biological” when the child’s date of birth falls in the history of the current partnership.

Due to additional information with every new wave, the indicator from previous waves may change in rare cases. In this case the MOTHPS$ is updated retrospectively.

Note that there are a few cases of same-sex partnerships in FiD. In the event that these are women, MOTHNOS is set to the PERSNR of the woman with the highest value in MOTHPS$, while FATHNOS is taken from the other partner (i.e. the biological mother would be found in MOTHNOS, the social mother in FATHNOS). If it is a male couple, FATHNOS is the PERSNR of the man with the highest value in FATHPS$, while MOTHNOS is taken from the other partner. As these cases are rare, we refrained from adding a sex indicator for MOTHPS$ and FATHPS$. This information can always be obtained through ppfad.

### FATHNOS$

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Never Changing Person ID (father)”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable format</td>
<td>8-digit integer</td>
</tr>
<tr>
<td>$ - Survey Years</td>
<td>$=10, 11, 12, 13</td>
</tr>
</tbody>
</table>

Comment: Variable generated identically to MOTHNOS$, see description above.

### FATHPS$

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Indicator for father’s relationship to the child”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value labels</td>
<td>FATHPS$</td>
</tr>
<tr>
<td>(1) biological</td>
<td></td>
</tr>
<tr>
<td>(2) social</td>
<td></td>
</tr>
<tr>
<td>(3) adoptive</td>
<td></td>
</tr>
<tr>
<td>(4) foster</td>
<td></td>
</tr>
<tr>
<td>(5) social, same-sex partners</td>
<td></td>
</tr>
<tr>
<td>(6) unknown</td>
<td></td>
</tr>
<tr>
<td>Variable format</td>
<td>1-digit integer</td>
</tr>
<tr>
<td>$ - Survey Years</td>
<td>$=10, 11, 12, 13</td>
</tr>
</tbody>
</table>

Comment: Indicator for the relationship father-child in the household.
Variable generated according to MOTHPS$, see description above.
**schltyp$**

**Variable label**  
“Schooltype (generated)”

**Value labels**  
SCHLTYPE$  
(1) primary school  
(2) secondary general school  
(3) intermediate school  
(4) upper secondary  
(5) comprehensive school  
(6) vocational school  
(7) school with special pedagogical concepts  
(8) special needs school  
(9) other

**Variable format**  
1-digit integer

**$S$ - Survey Years**  
$S$=10, 11, 12, 13

**Comment**  
Question 57a of the household questionnaire (survey years 2010/2011) and question 62a (survey years 2012/2013) specify the school the respective child goes to, with the option of an open answer in “other”. As many respondents use this field to put in specific schools like Waldorf, Montessori or schools for children with special needs, this additional information is recoded in this variable. Note that the recoded cases all have “-2 Does not apply” in the original variable.

---

**Appendix**

**Conversion of $STELL from 2012/2013 to 2010/2011 Version**

<table>
<thead>
<tr>
<th>$STELL in waves F10 and F11</th>
<th>$STELL in wave F12</th>
</tr>
</thead>
<tbody>
<tr>
<td>0  Head Of Household (HH)</td>
<td>0  Household (HH) head</td>
</tr>
<tr>
<td>1  Spouse of HH head</td>
<td>11  Spouse of HH head</td>
</tr>
<tr>
<td>2  Partner of HH head</td>
<td>13  Life partner of HH head</td>
</tr>
<tr>
<td>3  Son, daughter of HH head</td>
<td>21  Son, daughter of HH head</td>
</tr>
<tr>
<td>4  Foster child of HH head</td>
<td>23  Adopted child of HH head</td>
</tr>
<tr>
<td>5  Son, daughter-in-law of HH head</td>
<td>27  Son, daughter-in-law of HH head</td>
</tr>
<tr>
<td>6  Father, mother of HH head</td>
<td>31  Father, other of HH head</td>
</tr>
<tr>
<td></td>
<td>33  Adoptive mother, father of HH head</td>
</tr>
<tr>
<td></td>
<td>34  Foster mother, father of HH head</td>
</tr>
<tr>
<td>7  Parent-in-law of HH head</td>
<td>35  Parent-in-law of HH head</td>
</tr>
<tr>
<td>8  Sister, brother of HH head</td>
<td>41  Sister, brother of HH head</td>
</tr>
<tr>
<td></td>
<td>42  Half-sister, -brother of HH head</td>
</tr>
<tr>
<td></td>
<td>43  Stepsister, -brother</td>
</tr>
<tr>
<td></td>
<td>44  Adopted sister, brother of HH head</td>
</tr>
<tr>
<td></td>
<td>45  Foster sister, brother of HH head</td>
</tr>
<tr>
<td></td>
<td>51  Sister-in-law/brother-in-law 1: spouse or life partner of HH head's sibling</td>
</tr>
<tr>
<td>Code</td>
<td>Description</td>
</tr>
<tr>
<td>------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>52</td>
<td>Sister-in-law/brother-in-law 2: sibling of HH head's spouse or life partner</td>
</tr>
<tr>
<td>9</td>
<td>Grandchild of HH head</td>
</tr>
<tr>
<td>25</td>
<td>Grandchild of HH head</td>
</tr>
<tr>
<td>10</td>
<td>Other relative of HH head</td>
</tr>
<tr>
<td>26</td>
<td>Great-grandchild of HH head</td>
</tr>
<tr>
<td>32</td>
<td>Stepmother, stepfather of HH head</td>
</tr>
<tr>
<td>36</td>
<td>Grandmother, gradfather of HH head</td>
</tr>
<tr>
<td>61</td>
<td>Aunt/uncle of HH head</td>
</tr>
<tr>
<td>62</td>
<td>Niece/nephew of HH head</td>
</tr>
<tr>
<td>63</td>
<td>Cousin of HH head</td>
</tr>
<tr>
<td>64</td>
<td>Other relative of HH head</td>
</tr>
<tr>
<td>11</td>
<td>Not related to HH head</td>
</tr>
<tr>
<td>71</td>
<td>Not related to HH head</td>
</tr>
<tr>
<td>12</td>
<td>Stepchild of HH head</td>
</tr>
<tr>
<td>22</td>
<td>Stepchild of HH head</td>
</tr>
<tr>
<td>13</td>
<td>Same-sex spouse</td>
</tr>
<tr>
<td>12</td>
<td>Same-sex spouse</td>
</tr>
<tr>
<td>99</td>
<td>Unknown</td>
</tr>
<tr>
<td>99</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
### SSTELL-Combinations to identify mother-child relationships

Potential mother-child relationships as a combination of the variable SSTELL in waves F10 and F11 (accordingly for father-child relationships, conversion table above can be used to translate mother-child relationships to waves F12 and F13)

<table>
<thead>
<tr>
<th>SSTELL of the woman another person</th>
<th>Potential mother-child relationship (reference person = head of the HH) In this case the person is the…</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1</td>
<td>Child of reference person</td>
</tr>
<tr>
<td></td>
<td>Child of the wife of reference person</td>
</tr>
<tr>
<td>1 1</td>
<td>Child of the wife of reference person, but not child of reference person</td>
</tr>
<tr>
<td>2 2</td>
<td>Child of “life companion” of reference person and of reference person</td>
</tr>
<tr>
<td></td>
<td>Child of “life companion” of reference person but not of reference person</td>
</tr>
<tr>
<td>3 4</td>
<td>Child of daughter of reference person</td>
</tr>
<tr>
<td>5 5</td>
<td>Child of foster child of reference person</td>
</tr>
<tr>
<td></td>
<td>Child of daughter in law of reference person (3 generation household)</td>
</tr>
<tr>
<td>6 6</td>
<td>Child is reference person, lives with his mother in the same household</td>
</tr>
<tr>
<td></td>
<td>Child is the sister / brother of reference person, the siblings live with their mother in the same household</td>
</tr>
<tr>
<td>7 7</td>
<td>Child is spouse of reference person and lives together with spouse and mother in the same household</td>
</tr>
<tr>
<td></td>
<td>Child is daughter / son of the mother in law of reference person, but not the spouse of the reference person rather the sister in law / brother in law of reference person</td>
</tr>
<tr>
<td>8 9</td>
<td>Child is niece / nephew of reference person, mother is sister / sister in law of reference person</td>
</tr>
<tr>
<td>9 10</td>
<td>Child is another relation to reference person, great grandchild of reference person</td>
</tr>
<tr>
<td>10 11</td>
<td>Mother and child have another relation to reference person</td>
</tr>
<tr>
<td>11 11</td>
<td>Child and mother are in no way related to reference person</td>
</tr>
</tbody>
</table>
Documentation *biojob*

Detailed Information on First and Last Job

*Moritz Mannschreck*

This documentation is based on the comparable SOEP documentation on *biojob* and has benefited from previous work of Tanja Schmidt, Hansjoerg Haas, Anita Kottwitz, Daniel Wachtlin, Mathis Schroeder and Thorsten Schneider. For readability reasons we do not specifically cite and specify text that has been used directly from the SOEP document.
General information

**biojob** provides detailed information on the respondent’s first and last job. The relevant information is taken from the $slela-files which contain biographical information for all FiD respondents who are 18 or older in their first wave. **biojob** consists of generated variables as well as plain questionnaire information. Concerning different sources of information, the following priority scheme is applied: First the plain information stemming directly from questions on the relevant topic in the latest valid $slela-file is used. In case of inconsistencies, which will be explained later on, the latest valid information stemming from the pbiospe-file is also used. The pbiospe-file consists of spell data concerning the retrospective question ‘what did you do since the age of 15’ in the Biography Questionnaire as well as the question on activities in the last year in the Individual Questionnaire (for detailed information see pbiospe documentation). In contrast to the SOEP, **biojob** does not contain information from the Youth Questionnaire which is answered by respondents aged 16 to 17. Due to missing information, **biojob** contains fewer variables than the SOEP-version of **biojob**.

For all variables, the information provided looks as follows:

- **Variable label**: Provides the label of the variable as it is given in the dataset. Variables are given in CAPTIAL letters, even though they might appear in small letters in the dataset. This is simply for readability.

- **Value labels**: LBLNME
  
  In case VARNME is categorical, LBLNME specifies the labels for each category, and the value labels are listed here. Note that the standard missing value labels (-1: No answer; -2: Does not apply; -3: Not valid) are not listed, but apply to all variables in this dataset.

- **Variable format**: Specifies the format for each variable, e.g. “1-digit integer” or “string”.

- **Comment**: Provides more detailed information on the generating process, also on the population the variable is specified for, if necessary. Here, variables used, changes between waves, or any other anomalies are mentioned and their relevance explained.

If you have questions regarding **biojob** data for the FiD-distribution, unless noted otherwise, please contact Mathis Schröder at +49 (0)30 / 89789 - 222.
List of variables:

- AGEFJOB
- AGEINFO
- NOJOB
- STILLFJ
- FULLTIME
- OCCFJOB
- FJBLUE
- FJSELF
- FJSEFSIZ
- FJWHITE
- FJCIVS
- STBA
- ISCO88
- EGP
- ISEI
- MPS
- SIOPS
- CIVSFJ
- CURREMPL
- YEARLAST
- SCOPELJ
- CIVILSLJ
- NACELJ
- OCCLJOB
- LJBLUE
- LISELF
- LJSEFSIZ
- LJWHITE
- LICIVS
**AGEFJOB**

Variable label  
“Age at first job”

Variable format  
2-digit integer

Comment  
The variable AGEFJOB provides the age at entry into the working force.

In the Biography Questionnaire people either have to give information on their age at entry into the working force or have to state that they have never worked before the time of the interview. The latter information is used in the variable NOJOB. Some respondents have low values on AGEFJOB. Since most of the respective respondents worked either in low skilled blue collar or white collar jobs or within the family business, we consider the information as valid.

**AGEINFO**

Variable label  
“Info Source Age First Job”

Value labels  
AGEINFO

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LELA-Files (case (a) below)</td>
</tr>
<tr>
<td>2</td>
<td>PBIOSPE Age First Job LT 15 BEGIN GT 15 (c)</td>
</tr>
<tr>
<td>3</td>
<td>PB Not Worked And Later Begin (b)</td>
</tr>
<tr>
<td>4</td>
<td>PB N.W. And Earlier Begin (d)</td>
</tr>
<tr>
<td>5</td>
<td>PB Age First Job GT 31 And Earlier Begin (e)</td>
</tr>
<tr>
<td>6</td>
<td>Inconsistent Info (h)</td>
</tr>
<tr>
<td>7</td>
<td>PB N.W.,Age First Job Not Answ (f)</td>
</tr>
<tr>
<td>8</td>
<td>PB N.W.,Age First Job,Spell N.A. (g)</td>
</tr>
<tr>
<td>9</td>
<td>Completely Missing</td>
</tr>
<tr>
<td>10</td>
<td>SP No Info In LELA And PBIOSPE (i)</td>
</tr>
<tr>
<td>11</td>
<td>YOUTH-files (y)</td>
</tr>
<tr>
<td>12</td>
<td>PB Empl. And Earlier Begin (y)</td>
</tr>
<tr>
<td>13</td>
<td>N.W. And Later Begin (y)</td>
</tr>
<tr>
<td>14</td>
<td>Inval. Info And Later Empl. (y)</td>
</tr>
<tr>
<td>15</td>
<td>Empl. N.A. And Spell (y)</td>
</tr>
<tr>
<td>16</td>
<td>Completely Missing</td>
</tr>
<tr>
<td>17</td>
<td>Inconsistent Info (y)</td>
</tr>
<tr>
<td>18</td>
<td>SP No Info YOUTH And PBIOSPE (y)</td>
</tr>
</tbody>
</table>

Variable format  
2-digit integer

Comment  
AGEINFO is a pointer variable indicating the source of the age information for AGEFJOB. For *lela*-respondents, the following coding procedure is applied:

a) For people who are or have ever been employed at the time of answering the biographical questions their age at the time of entry into the working force is taken from the *lela*-files.

b) When we observe, that the person has not been in the working force at the time of responding, but starts to work later on, data of the *pbiospe*-file is used. Using the spell information in *pbiospe*, we are able to collect the age at the first job.
c) A replacement of the lela-data takes place, when respondents state that they have worked before the age of fifteen, but have a spell entry later than the age of fifteen. This rule is not applied when the spell starts at the age of fifteen, since this is the minimum value for spell data in the questionnaires.

d) The same procedure is applied, when people answer, that they have never worked at the time of the interview, but have a spell which starts before the first interview.

e) In some cases the AGEFJOB value is higher than the start of the corresponding working spell in pbiospe. In general, the AGEFJOB value is maintained. Only when the value is greater than 27, is it replaced by the pbiospe data. (95% of these cases have an AGEFJOB below 27.)

f) If we observe item non response concerning AGEFJOB and NOJOB, but spell information is available, the missing value is replaced by the corresponding pbiospe spell data.

g) If even the ‘What did you do since you were 15’ question had not been answered, there still was a chance to extract similar information out of the pbiospe-file by considering the question ‘What did you do every month last year’.

h) If we still had no valid information, the value of AGEFJOB was left out of the dataset.

i) Due to the fact that pbiospe information are collected only until the end of the year preceding the actual wave (in this version of biojob: December 2011), for respondents without first job information from both the biography questionnaire and pbiospe we further look for a first job using information from the current wave individual questionnaire.

As described above, biojob does not contain information from the Youth Questionnaire therefore there are no observations for values (11) – (18).

**NOJOB**

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Never been employed”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value labels</td>
<td>NOJOB</td>
</tr>
<tr>
<td>(1) Yes</td>
<td></td>
</tr>
<tr>
<td>Variable format</td>
<td>1-digit integer</td>
</tr>
<tr>
<td>Comment</td>
<td>The underlying question for the variable NOJOB is ‘I have never been employed up to this date’. If NOJOB has a missing value, in general there should exist AGEFJOB information, for special cases, see above.</td>
</tr>
</tbody>
</table>

**STILLFJ**

<table>
<thead>
<tr>
<th>Variable label</th>
<th>“Still Employed In First Job”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value labels</td>
<td>STILLFJ</td>
</tr>
<tr>
<td>(1) Yes</td>
<td></td>
</tr>
<tr>
<td>(2) No</td>
<td></td>
</tr>
<tr>
<td>Variable format</td>
<td>1-digit integer</td>
</tr>
<tr>
<td>Comment</td>
<td>This variable is based on the question ‘Are you still employed in the same job and at the same place?’. It applies only to lela-respondents who do not state ‘I have never been gainfully employed.</td>
</tr>
</tbody>
</table>
FULLTIME

Variable label: “First Job Full Time”
Value labels:
   (1) Full time job
   (2) Part time job
Variable format: 1-digit integer

Comment: The FULLTIME-variable is used to indicate, whether the first job of a person was a full-time or a part-time job. This variable is generated out of the pbiospe-file for all respondents. For persons with first job information stemming from the Biography Questionnaires, FULLTIME possibly does not refer to the declared first job if pbiospe does not contain the respective job spell (i.e. due to item non response or incomplete answering of the activity biography within the Biography Questionnaire).

OCCFJOB

Variable label: “Occ. Position First Job”
Value labels:
   (1) Blue Collar W.
   (2) Self-Employed
   (3) White Collar W.
   (4) Civil Servant
Variable format: 1-digit integer

Comment: The variable OCCFJOB provides information on the occupational position at the first job. The group ‘Farmers’ is included in the blue collar worker group. Due to the fact that the pbiospe-file is used for the coding of AGEFJOB in certain cases (see above) there is less information on OCCFJOB than on AGEFJOB.

FJBLUE

Variable label: “First Job Blue Collar”
Value labels:
   (1) Unskilled Worker
   (2) Semiskilled Worker
   (3) Skilled Worker
Variable format: 1-digit integer

Comment: The FJBLUE variable provides detailed information on the first occupational status if the person was a blue collar worker.

FJSELFE

Variable label: “First Job Self Employed”
Value labels:
   (1) Independent Farmer
Variable format 1-digit integer

Comment The FJSELF variable provides detailed information on the first occupational status if the person was self-employed.

**FJSEFSIZ**
Variable label “No. of Employees First Job Self Employed”
Value labels FJSEFSIZ
(1) No employees
(2) <=9 employees
(3) >=10 employees
Variable format 1-digit integer

Comment FJSEFSIZ gives the number of employees in the respondent’s firm.

**FJWHITE**
Variable label “First Job White Collar”
Value labels FJWHITE
(1) Unskilled Labour, Without Degree
(2) Unskilled Labour, With Degree
(3) Skilled Labour
(4) Professional Labour
Variable format 1-digit integer

Comment FJWHITE gives detailed information on persons, who were first employed as white collar workers.

**FJCIVS**
Variable label “First Job Civil Servant”
Value labels FJCIVS
(1) Low Level Civil Servant
(2) Middle Level Civil Servant
(3) High Level Civil Servant
(4) Executive Civil Servant
Variable format 1-digit integer

Comment FJCIVS provides detailed information on first employment as a public servant

**STBA**
Variable label “StaBua Vocational Classification First Job”
Value labels STBA
**biojob: Detailed Information on First and Last Job**

**FiD-Documentation**

Variable format 4-digit integer

Comment STBA provides information on the StaBua Vocational Classification of the first job. STBA builds on information from the Biography Questionnaire. Respondents answer the question on their first occupational title in their own words, and this response is entered into a blank in the questionnaire. Due to data protection regulations, this information cannot be provided to data users and was therefore completely recoded by Infratest Sozialforschung. For detailed information on the generation and coding of STBA see the documentation on *pgen*.

**ISCO88**

Variable label “4 Digit ISCO-88 Occupation Code First Job”
Value labels ISCO88
Variable format 4-digit integer

Comment ISCO88 provides information on the ISCO-88 Occupation Code of the first job. ISCO88 builds on information from the Biography Questionnaire. Respondents answer the question on their first occupational title in their own words, and this response is entered into a blank in the questionnaire. Due to data protection regulations, this information cannot be provided to data users and was therefore completely recoded by Infratest Sozialforschung. For detailed information on the generation and coding of ISCO88 see the documentation on *pgen*.

**EGP**

Variable label “EGP Class Category ISCO-88 First Job”
Value labels EGP
Variable format 2-digit integer

Comment EGP provides information on the EGP Class Category of the first job. For detailed information on the generation and coding of EGP see the documentation on *pgen*.

**ISEI**

Variable label “Ganzeboom ISEI-Status88 First Job”
Value labels ISEI
Variable format 2-digit integer

Comment ISEI provides information on the ISEI Class Category of the first job. For detailed information on the generation and coding of ISEI see the documentation on *pgen*. 
**biojob: Detailed Information on First and Last Job**

**FiD-Documentation**

**MPS**
- **Variable label**: “Magnitude Prestige Scale First Job”
- **Value labels**: MPS
- **Variable format**: 5-digit real

**Comment**: MPS provides information on the Magnitude Prestige Scale of the first job. For detailed information on the generation and coding of MPS see the documentation on pgen.

**SIOPS**
- **Variable label**: “Treiman Standard Int. Occ. Prestige First Job”
- **Value labels**: SIOPS
- **Variable format**: 2-digit integer

**Comment**: SIOPS provides information on the Treiman Standard International Occupational Prestige Scale of the first job. For detailed information on the generation and coding of SIOPS see the documentation on pgen.

**CIVSFJ**
- **Variable label**: “First Job In Civil Service”
- **Value labels**: CIVSFJ
  - (1) Yes
- **Variable format**: 1-digit integer

**Comment**: CIVSFJ indicates if the first job was assigned to the civil service or not. Note that this variable only applies to public servants (German: Beamte) and not to other employees in the public sector.

**CURREMPL**
- **Variable label**: “Employed At Time Of Bio Interview”
- **Value labels**: CURREMPL
  - (1) Yes
  - (2) No
- **Variable format**: 1-digit integer

**Comment**: This variable is based on the question ‘Are you gainfully employed at the current time?’. The question applies only to Stele respondents who do not state ‘I have never been gainfully employed’ or ‘Still employed in the first job’.

**YEARLAST**
- **Variable label**: “Year Of Last Employment”
- **Variable format**: 4-digit integer
Comment This variable is based on the question ‘When was the last time you were gainfully employed?’. The question applies only to Stela respondents who do not make at least one of the following statements in their biography interview:
‘I have never been gainfully employed.’
‘Still employed in the first job’
‘Gainfully employed at the current time’.

**SCOPELJ**
Variable label “Last Job Full-/Part-Time”
Value labels

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FT Employed</td>
</tr>
<tr>
<td>2</td>
<td>PT Employed</td>
</tr>
<tr>
<td>3</td>
<td>Marg./Irreg.Empl.</td>
</tr>
</tbody>
</table>

Variable format 1-digit integer

Comment SCOPELJ indicates if the last job was a full time or part time job. Information is only provided for respondents who answer the respective question within the Biography Questionnaires. The respective question applies only to respondents who do not make at least one of the following statements:
‘I have never been gainfully employed.’
‘Still employed in the first job’
‘Gainfully employed at the current time’.

**CIVILSLJ**
Variable label “Last Job In Civil Service”
Value labels

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
</tr>
</tbody>
</table>

Variable format 1-digit integer

Comment CIVILSLJ indicates if the last job was assigned to the civil service or not. Information is only provided for respondents who answer the respective question within the Biography Questionnaires. The respective question applies only to respondents who do not make at least one of the following statements:
‘I have never been gainfully employed.’
‘Still employed in the first job’
‘Gainfully employed at the current time’.

**NACELJ**
Variable label “2 Digit NACE Industry,Sector (Last Job)”
Value labels NACELJ
Variable format 3-digit integer

Comment NACELJ provides information on the NACE Industry Code on the industry sector the respondent was employed in during the last job.
Respondents answer the question in their own words regarding the industry in which they are currently working, and this response is entered into a blank in the questionnaire. For detailed information on the generation and coding of NACELJ see the documentation on pgen.

**OCCLJOB**
Variable label **“Occ. Position Last Job”**
Value labels OCCLJOB
(1) Blue Collar W.
(2) Self-Employed
(3) White Collar W.
(4) Civil Servant
Variable format 1-digit integer

Comment The variable OCCLJOB provides information on the occupational position at the last job. Information is only provided for respondents who answer the respective question within the Biography Questionnaires. The respective question applies only to respondents who do not make at least one of the following statements:

‘I have never been gainfully employed.’
‘Still employed in the first job’
‘Gainfully employed at the current time’.

**LJBLUE**
Variable label **“Last Job Blue Collar”**
Value labels LJBLUE
(1) Unskilled Worker
(2) Semiskilled Worker
(3) Skilled Worker
Variable format 1-digit integer

Comment The LJBLUE variable provides detailed information on the last occupational status if the person was a blue collar worker.

**LJSELFIE**
Variable label **“Last Job Self Employed”**
Value labels LJSELFIE
(1) Independent Farmer
(2) Freelance
(3) Other Self Employed
(4) Within Family Business
Variable format 1-digit integer

Comment The LJSELFIE variable provides detailed information on the last occupational status if the person was self-employed.
**LJSEFSIZ**

Variable label: **“No. of Employees Last Job Self Employed”**

Value labels:
- LJSEFSIZ
  - (1) No employees
  - (2) <=9 employees
  - (3) >=10 employees

Variable format: 1-digit integer

Comment: LJSEFSIZ gives the number of employees in the respondent’s firm.

**LJWHITE**

Variable label: **“Last Job White Collar”**

Value labels:
- LJWHITE
  - (1) Unskilled Labour, Without Degree
  - (2) Unskilled Labour, With Degree
  - (3) Skilled Labour
  - (4) Professional Labour

Variable format: 1-digit integer

Comment: LJWHITE gives detailed information on persons, who were last employed as white collar workers.

**LJCIVS**

Variable label: **“Last Job Civil Servant”**

Value labels:
- LJCIVS
  - (1) Low Level Civil Servant
  - (2) Middle Level Civil Servant
  - (3) High Level Civil Servant
  - (4) Executive Civil Servant

Variable format: 1-digit integer

Comment: LJCIVS provides detailed information on last employment as a public servant.
Documentation \textit{hhrf} and \textit{phrf}

Calculating person and household level weights

\textit{Rainer Siegers}

\textit{Note that this documentation is in German. Translations will be provided if necessary.}
Inhalt

Inhalt .............................................................................................................................. 198
Hochrechnung in „Familien in Deutschland“ ............................................................... 199
Gewichtungsansatz ....................................................................................................... 199
Stichproben in FiD ...................................................................................................... 201
Hochrechnung der Screening-Stichproben 2010 und 2011 ........................................... 202
Hochrechnung der Kohorten-Stichprobe ................................................................. 209
Integration der FiD-Stichproben .............................................................................. 212
Integration von SOEP und FiD ................................................................................. 213
Querschnittsgewichte für Daten nach 2010 .............................................................. 213
Längsschnittgewichte .............................................................................................. 216
Nutzung der Hochrechnungsfaktoren ..................................................................... 216
Anhang: Neue Gewichte 2010-2012 ab der Weitergabe FiDv3.1 .............................. 218
**Hochrechnung in „Familien in Deutschland“**


**Gewichtungsansatz**

Die Nutzung von Hochrechnungsfaktoren basiert auf der Idee, dass jeder beobachtete Fall der Stichprobe einen Teil der Grundgesamtheit repräsentiert. Grundgesamtheit von FiD sind alle Haushalte in Deutschland, die zum Zeitpunkt der Stichprobenziehung in einer der oben genannten Familienkonstellationen leben sowie sämtliche Personen in diesen Haushalten. Im Idealfall einer einfachen Zufallsziehung von Haushalten und der realisierten Teilnahme jedes einzelnen Haushaltes würde jeder Fall die gleiche Anzahl an Haushalten repräsentieren und alle den gleichen Hochrechnungsfaktor bekommen, der dem Kehrwert seiner Ziehungswahrscheinlichkeit entspricht. Tatsächlich führen sowohl das Ziehungsdesign als auch Unit-Nonresponse, also die Verweigerung eines Interviews durch Befragte, zu
Unterschieden in der Wahrscheinlichkeit, Teil der Stichprobe zu sein. Ziel der Hochrechnung ist die Schätzung des Kehrwerts dieser Wahrscheinlichkeit für jede Erhebungseinheit. Das Hochrechnungsverfahren besteht dabei grundsätzlich aus drei Schritten:


**Stichproben in FiD**

Die Haushalte des FiD-Datensatzes wurden in drei voneinander unabhängigen Stichproben gezogen. Die praktische Umsetzung des Ziehungsdesigns und die Feldarbeit wurden von TNS Infratest durchgeführt.


**Hochrechnung der Screening-Stichproben 2010 und 2011**


---

Bereitschaft für weitere Befragungen erklärt haben. Es besteht also ein Selektionseffekt darin, nur wiederbefragungswillige Personen in der Ausgangsstichprobe zu haben. Über das Ausmaß dieser Einschränkungen stehen keine genauen Informationen zur Verfügung. Daher erlangt bei den Screening-Stichproben der abschließende Kalibrierungsschritt eine stärkere Bedeutung als üblich. Migranten-Haushalte können aufgrund der Konzentration der Busse auf Deutsche bzw. deutschsprachige Personen in der Screening-Stichprobe nicht repräsentativ erfasst werden und werden bei der Gewichtung nicht gesondert berücksichtigt.

Abbildung 1: Stichprobendesign für die Screening-Stichprobe 2010, Quelle: Methodenbericht TNS Infratest

17 Unter der „Integration“ ist zu verstehen, dass die unterschiedlichen Stichproben (bzw. Datensätze) mit Hochrechnungsfaktoren versehen werden, die eine gemeinsame Nutzung ermöglichen.

Abbildung 2a: Boxplots der Hochrechnungsfaktoren der Screening-Stichprobe 2010 nach Haushaltstypen (NE: Niedrigereinkommenshaushalte, AE: Alleinerziehendenhaushalte, MK: Mehrkindhaushalte)

Abbildung 2b: Boxplots der Hochrechnungsfaktoren der Screening Stichprobe 2011 nach Haushaltstypen
(AE: Alleinerziehendenhaushalte, MK: Mehrkindhaushalte)

Modell 3: Unter den kontaktierten Haushalten des Telefon-Screenings gibt es solche, die die Teilnahme am Screening verweigerten. Von den 10.396 kontaktierten Personen in 2010

---

18 für weitere Informationen über die zur Verfügung stehenden Variablen siehe


Modell 4: In 2010 erklärten sich von den 3.701 Haushalten, die das Screening durchlaufen haben und zur Zielpopulation gehören, 3.261 Haushalte grundsätzlich bereit, bei der FiD-Studie teilzunehmen. 51 Haushalte wurden als Streuadressen nicht kontaktiert. Sie wurden bei der Gewichtung als qualitätsneutrale Ausfälle gewertet und nicht weiter berücksichtigt. 61 Haushalte gaben falsche bzw. fehlerhafte Adressen an und wurden bei der Gewichtung als heimliche Verweigerer aufgefasst. Für 2010 fließen in Modell 4 also 3650 Fälle ein.

Für die Haupterhebung 2010 gibt es ein Ausgangs-Brutto von 3.149 Haushalten (2011: 1.154), die sich zur Teilnahme bereit erklärt haben. Auf dieser Modellstufe wurde die Wahrscheinlichkeit zur Teilnahmebereitschaft an der Haupterhebung geschätzt, diesmal allerdings auf Basis der aktualisierten individuellen Informationen, erneut ergänzt durch Gemeindeinformationen von INKAR.


Modell 7: Ausgehend von den kalibrierten Haushaltshochrechnungsfaktoren werden im Anschluss Hochrechnungsfaktoren auf Personenebene erstellt, die sich durch eine weitere iterative Randanpassung ergeben. Genutzt wurden die Randverteilungen für Alter, Geschlecht, Erwerbstatus, Bildung, Zugehörigkeit zu den Haushaltstypen, Überschneidungen mit den Haushalten der Kohorten-Stichprobe und einer Korrektur für Partial Unit-Non-Response (PUNR) nach Art der Haushaltszusammensetzung. PUNR-Fälle selbst, also Personen aus teilnehmenden Haushalten, die die Teilnahme an der Befragung verweigern, bekomen ein Gewicht von Null. Tabelle 1 zeigt eine zusammenfassende Übersicht über die einzelnen Schritte bei der Gewichtung der Screening-Stichprobe.

\begin{align*}
W_{\text{korr}} = \begin{cases} 
2.5 \cdot \bar{W}_{\text{org}} + \ln(W_{\text{org}} + 1 - (2.5 \cdot \bar{W}_{\text{org}})) & \text{, wenn } W_{\text{org}} > 2.5 \cdot \bar{W}_{\text{org}} \\
W_{\text{org}} & \text{, sonst}
\end{cases}
\end{align*}

\textsuperscript{20} nach folgender Formel

\textsuperscript{21} Besonderer Dank gilt Robert Herter-Esche, der uns diese recht aufwändige Sonderauswertung des Statistischen Bundesamtes zusammengestellt hat.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Modell 1</td>
<td>Auswahlwkt. für das Screening aus Infratest-Bussen</td>
<td>Eingeschränkte Informationen aus den Bus-Befragungen</td>
<td>23193</td>
<td>13653</td>
<td>19041</td>
<td>8400</td>
</tr>
<tr>
<td>Modell 2</td>
<td>Kontaktwahrscheinlichkeit bei Screening</td>
<td>Angaben aus den Bus-Befragungen sowie INKAR-Daten</td>
<td>13653</td>
<td>10396</td>
<td>8400</td>
<td>6461</td>
</tr>
<tr>
<td>Modell 3</td>
<td>Teilnahmewahrscheinlichkeit am Screening</td>
<td>Angaben aus den Bus-Befragungen sowie INKAR-Daten</td>
<td>10396</td>
<td>6656</td>
<td>6461</td>
<td>3515</td>
</tr>
<tr>
<td>Modell 4</td>
<td>Wkt. für Bereitschaft zur Teilnahme an Haupterhebung</td>
<td>Angaben aus den Bus-Befragungen, dem Screening sowie INKAR-Daten</td>
<td>3650</td>
<td>3149</td>
<td>1272</td>
<td>1154</td>
</tr>
<tr>
<td>Modell 5</td>
<td>Wkt. für tatsächliche Teilnahme an Haupterhebung</td>
<td>Angaben aus den Bus-Befragungen, dem Screening, durch Interviewer sowie INKAR- und MICROM-Daten</td>
<td>3149</td>
<td>2500</td>
<td>1154</td>
<td>924</td>
</tr>
<tr>
<td>Modell 6</td>
<td>Kalibrierung auf Haushaltsebene</td>
<td>Angaben aus dem Mikrozensus</td>
<td>2263</td>
<td>2263</td>
<td>915</td>
<td>915</td>
</tr>
<tr>
<td>Modell 7</td>
<td>Kalibrierung auf Personenebene</td>
<td>Angaben aus dem Mikrozensus</td>
<td>7994</td>
<td>7994</td>
<td>3548</td>
<td>3548</td>
</tr>
</tbody>
</table>

Tabelle 1: Übersicht über die einzelnen Schritte der Gewichtung der Screening-Stichproben

**Hochrechnung der Kohorten-Stichprobe**

Die folgende Grafik veranschaulicht den Auswahlprozess der Kohorten-Stichprobe bis hin zu den tatsächlich im Datensatz enthaltenen Netto-Haushalten. Im Gewichtungsprozess spiegeln sich die einzelnen Schritte im Wesentlichen wider. Im Folgenden werden die einzelnen Stufen im Hinblick auf die Gewichtung der Kohorten-Stichprobe betrachtet (Tabelle 2 gibt den Überblick über die einzelnen Modelle).

Modell 1: Die Kohorten-Stichprobe wurde in einem geschichteten, mehrstufigen Verfahren gezogen. In einem ersten Schritt wurden zufällig Einwohnermeldeämter gezogen. Innerhalb der Daten der Einwohnermeldeämter wurden in einem weiteren Schritt zufällig Kinder der interessierenden Geburtskohorten ausgewählt. Um dem bei Surveys bekannten Phänomen zu begegnen, dass Personen mit Migrationshintergrund deutlich geringere Responseraten als die

Abbildung 3: Stichprobendesign für die Kohorten-Stichprobe, Quelle: Methodenbericht TNS Infratest

Modell 2: Von den 5.366 zur Befragung vorgesehenen Haushalten nahmen schließlich 2.074 Haushalte an der Befragung teil. Die anderen Haushalte der Brutto-Stichprobe teilen sich in
604 qualitätsneutrale Ausfälle, 1.018 Ausfälle, die nicht kontaktiert werden konnten sowie 1.670 Ausfälle durch Verweigerung der Teilnahme. Im zweiten Modell wird die Kontaktwahrscheinlichkeit geschätzt. Dabei kann, wie schon bei der Hochrechnung der Screening-Stichproben, neben den von den Interviewern erhobenen Informationen auf externe Daten von INKAR und MICROM zurückgegriffen werden.


Modell 5: Die Erstellung der Hochrechnungsfaktoren auf Personenebene geschieht analog zu dem Verfahren der Screening-Stichprobe (dort Modell 7).

Tabelle 2 zeigt eine Übersicht über die einzelnen Schritte bei der Gewichtung der Kohorten-Stichprobe.

<table>
<thead>
<tr>
<th>Gegenstand der Schätzung</th>
<th>Genutzte Information</th>
<th>Fallzahl vor Modellschritt</th>
<th>Fallzahl nach Modellschritt</th>
</tr>
</thead>
</table>

Modell 1: Oversampling von Migranten
Angaben der Einwohnermeldeämter
5366  5366

Modell 2: Kontaktwahrscheinlichkeit bei Haupterhebung
Angaben durch Interviewer sowie INKAR- und MICROM-Daten
4762  3744

Modell 3: Teilnahmewahrscheinlichkeit bei Haupterhebung
Angaben durch Interviewer sowie INKAR- und MICROM-Daten
3744  2074

Modell 4: Kalibrierung auf Haushaltsebene
Angaben aus dem Mikrozensus
2074  2074

Modell 5: Kalibrierung auf Personenebene
Angaben aus dem Mikrozensus
7670  7670

Tabelle 2: Übersicht über die einzelnen Schritte der Gewichtung der Kohorten-Stichprobe

Integration der FiD-Stichproben
Mit den bisher beschriebenen Hochrechnungsfaktoren lassen sich Analysen für die jeweilige Population durchführen, also beispielsweise für die Haushalte, die 2010 als Niedrigereinkommenshaushalte definiert wurden und sich in der Screening-Stichprobe befinden. Allerdings nutzt man dann nicht das volle Potential, weil Fälle nicht berücksichtigt werden, die aus den anderen Stichproben ebenfalls zur interessierenden Population gehören. Ein Niedrigereinkommenshaushalt mit Kleinkind könnte beispielsweise sowohl als Teil der Screening- als auch der Kohorten-Stichprobe gezogen worden sein. Die Integration der Screening- und Kohorten-Stichproben mit gemeinsamen Hochrechnungsfaktoren löst dieses Problem, so dass bei Analysen zusätzliche Fälle durch die Nutzung der anderen Stichproben gewonnen werden können.

Aufgrund der unterschiedlichen Ziehungsjahre erfolgt die Integration der einzelnen Stichproben nicht einheitlich. Für das Ziehungsjahr 2010 und die Integration der Screening-Stichprobe 2010 mit der Kohorten-Stichprobe werden als Ausgangswerte für die Integration die Gewichte dieser beiden Stichproben vor dem Kalibrierungsschritt herangezogen und gemeinsam auf der Haushaltsebene kalibriert. Da bei der Kalibrierung die verwendeten Eingangsgewichte Einfluss auf die Verteilung der resultierenden Hochrechnungsfaktoren haben, werden sie normiert, so dass die überschneidenden Populationen beider Stichproben mit dem gleichen durchschnittlichen Eingangsgewicht in den ersten Kalibrierungsschritt laufen. Die verwendeten Variablen für die Randanpassung entsprechen inhaltlich denen, die
für die stichprobenspezifischen Kalibrierungen herangezogen werden. In einem zweiten Schritt werden auf Basis der integrierten Hochrechnungsfaktoren auf Haushaltsebene die Personen-Hochrechnungsfaktoren erstellt.


**Integration von SOEP und FiD**


**Querschnittsgewichte für Daten nach 2010**


1. in 2010 im Niedrigeinkommensbereich waren (Screening 2010), oder
2. in 2010 oder 2011 Mehrkindhaushalte waren (Screening 2010 oder 2011), oder
3. in 2010 oder 2011 alleinerziehend waren (Screening 2010 oder 2011), oder
4. in 2010 Kinder der Geburtsjahrgänge 2007-2010 hatten.

Eine so definierte Population von FiD- und SOEP-Fällen hat ein bestimmtes Bevölkerungsäquivalent, das sich mit Hilfe der gemeinsamen Hochrechnungsfaktoren errechnen lässt. Wenn nun die SOEP-Fälle aus der Population entfernt werden, müssen die Hochrechnungsfaktoren der verbleibenden FiD-Fälle angepasst werden, damit sich bei einer erneuten Hochrechnung weiterhin das Bevölkerungsäquivalent ergibt.

Dieses Verfahren lässt sich für die Merkmale (1), (2) und (3), die im Prinzip die Screening Stichproben beschreiben, nicht anwenden, weil sie im Zeitverlauf stärkeren Veränderungen unterworfen sind. Entsprechend wird hier mithilfe der Haushalte, die in beiden Jahren vorliegen, ein Logit-Modell²⁴ geschätzt, das anhand von 2011 erhobenen Informationen die Wahrscheinlichkeit vorhersagt, ob der Haushalt in 2010 zur Screening Stichprobe gehörte oder nicht. Die Ergebnisse dieses Models werden dann genutzt, um die Zugehörigkeit der ausschließlich in 2011 befragten Haushalte zu schätzen.

Mit dem skizzierten Verfahren lassen sich nun alle in 2011 befragten FiD- und SOEP-Haushalte in die definierten Populationsmerkmale (1)-(4) einordnen, woraus sich auch das angestrebte Bevölkerungsäquivalent ergibt. Ein Entfernen der SOEP-Fälle aus dieser Gruppe bedeutet, dass die verbleibenden FiD-Fälle im Durchschnitt höhere Hochrechnungsfaktoren erhalten müssen, damit das Bevölkerungsäquivalent sich nicht wesentlich verändert. Am einfachsten wäre hier eine Multiplikation mit einem einzigen Faktor, was aber aufgrund der unterschiedlichen Verteilung der Merkmale zwischen FiD und SOEP nicht sinnvoll ist (vgl. auch Abbildung 2a und 2b). Stattdessen wird nun mit einem Logit-Modell die Wahrscheinlichkeit berechnet²⁵, zur FiD-Population zu gehören, ähnlich wie in einer Ausfallanalyse die Wahrscheinlichkeit ermittelt wird, in der folgenden Welle die Studie zu verlassen. Die Kehrwert dieser Wahrscheinlichkeit ist dann der haushalts- bzw. personenspezifische Faktor, mit dem die Hochrechnungsfaktoren multipliziert und erhöht werden (entspricht der inversen Bleibewahrscheinlichkeit bei der Ausfallanalyse).


Auch für die dritte FiD-Welle (2012) werden FiD-spezifische Querschnittsgewichte bereitgestellt, die nach dem beschriebenen Prinzip aus den gemeinsamen SOEP-FiD-

---

Hochrechnungsfaktoren gewonnen werden. Die Basis bilden hier Hochrechnungsfaktoren, die ohne das im SOEP 2012 neu erhobene Sample K gebildet wurden. Auf diese Weise wird verhindert, für die Haushalte des Sample K die FiD-Populationszugehörigkeit für zwei zurückliegende Jahre schätzen zu müssen. Die Genauigkeit einer solchen Schätzung verringert sich mit zunehmendem Zeitabstand, so dass der Verzicht auf das neue Sample vorgezogen wird.


**Längsschnittgewichte**

Die Erhebung der zweiten Welle markiert den Übergang in den Längsschnitt für die in 2010 gezogenen Stichproben. Um Fälle im Längsschnitt richtig hochzurechnen, müssen die Fälle, die in beiden Wellen teilnehmen in ihren Gewichten um die Fälle korrigiert werden, die nach der ersten Welle ausfallen. Diese Korrektur erfolgt über sogenannte Bleibewahrscheinlichkeiten, die sich aus den Schätzungen der Wahrscheinlichkeiten des Wiederauffindens und der erneuten Teilnahme der Haushalte ergeben. Diese Wahrscheinlichkeiten werden mittels logistischer Regression geschätzt, wobei ein Vorteil gegenüber der Berechnung der ursprünglichen Gewichte ist, dass sämtliche Haushalte der Analyse bereits im Vorjahr teilgenommen haben, und entsprechend viele eigene Daten vorliegen, die die Teilnahme in der zweiten Welle erklären können. Dennoch werden auch bei der Schätzung der Bleibewahrscheinlichkeiten Informationen von MICROM genutzt. Der Kehrwert der geschätzten Teilnahmewahrscheinlichkeit bildet dann das Bleibegewicht, das als Korrektur auf das ursprüngliche Gewicht multipliziert wird, um Analysen im Längsschnitt hochrechnen zu können.

**Nutzung der Hochrechnungsfaktoren**

Das Hochrechnungskonzept von FiD erlaubt es dem Nutzer, die FiD-Stichprobe separat zu analysieren oder sie in Kombination mit den Daten des SOEP zu verwenden und gemeinsam

---

25 Die Variablen dieses Modells bilden die Zusammensetzung der Haushalte bezüglich der FiD-Merkmale in den Jahren 2010 und 2011 ab und werden ergänzt durch die Wohnregion (Ost / West) sowie im Fall der Personengewichte um Altersgruppen.
hhrf & phrf: Berechnung und Nutzung der Gewichte

FiD-Dokumentation

hhrf & phrf: Berechnung und Nutzung der Gewichte

FiD-Dokumentation

Die Datensätze *hhrf* und *phrf* enthalten jeweils die Identifikatoren (HHNR und HHNRAKT bzw. PERSNR), die es erlauben eine Beobachtung zu identifizieren und die Gewichte anderen Daten zuzuordnen. Die folgenden Gewichte sind hier enthalten:

**hhrf**

<table>
<thead>
<tr>
<th>Code</th>
<th>Beschreibung</th>
</tr>
</thead>
<tbody>
<tr>
<td>hhrf*</td>
<td>HRF für die gesamte FiD-Stichprobe 20$$ auf Haushaltsebene</td>
</tr>
<tr>
<td>f11hbleib</td>
<td>Inverse Bleibwahrscheinlichkeit für 2011 auf Haushaltsebene</td>
</tr>
<tr>
<td>f12hbleib</td>
<td>Inverse Bleibwahrscheinlichkeit für 2012 auf Haushaltsebene</td>
</tr>
<tr>
<td>f13hbleib</td>
<td>Inverse Bleibwahrscheinlichkeit für 2013 auf Haushaltsebene</td>
</tr>
<tr>
<td>f10hhrf_sc10</td>
<td>HRF für die Screening-Stichprobe 2010 auf Haushaltsebene (nur 2010)</td>
</tr>
<tr>
<td>f10hhrf_co10</td>
<td>HRF für die Kohorten-Stichprobe 2010 auf Haushaltsebene (nur 2010)</td>
</tr>
<tr>
<td>f11hhrf_sc11</td>
<td>HRF für die Screening-Stichprobe 2011 auf Haushaltsebene (nur 2011)</td>
</tr>
</tbody>
</table>

**phrf**

<table>
<thead>
<tr>
<th>Code</th>
<th>Beschreibung</th>
</tr>
</thead>
<tbody>
<tr>
<td>phrf*</td>
<td>HRF für die gesamte FiD-Stichprobe 20$$ auf Personenebene</td>
</tr>
<tr>
<td>f11pbleib</td>
<td>Inverse Bleibwahrscheinlichkeit für 2011 auf Personenebene</td>
</tr>
<tr>
<td>f12pbleib</td>
<td>Inverse Bleibwahrscheinlichkeit für 2012 auf Personenebene</td>
</tr>
<tr>
<td>f13pbleib</td>
<td>Inverse Bleibwahrscheinlichkeit für 2013 auf Personenebene</td>
</tr>
<tr>
<td>f10phrf_sc10</td>
<td>HRF für die Screening-Stichprobe 2010 auf Personenebene (nur 2010)</td>
</tr>
<tr>
<td>f10phrf_co10</td>
<td>HRF für die Kohorten-Stichprobe 2010 auf Personenebene (nur 2010)</td>
</tr>
<tr>
<td>f11phrf_sc11</td>
<td>HRF für die Screening-Stichprobe 2011 auf Personenebene (nur 2011)</td>
</tr>
</tbody>
</table>

**hhrf_fidsoep**

<table>
<thead>
<tr>
<th>Code</th>
<th>Beschreibung</th>
</tr>
</thead>
<tbody>
<tr>
<td>hhrf_soep</td>
<td>HRF SOEP integriert mit FiD 20$$ auf Haushaltsebene</td>
</tr>
<tr>
<td>f11hbleib</td>
<td>Inverse Bleibwahrscheinlichkeit für 2011</td>
</tr>
<tr>
<td>f12hbleib</td>
<td>Inverse Bleibwahrscheinlichkeit für 2012</td>
</tr>
</tbody>
</table>

**phrf_fidsoep**

<table>
<thead>
<tr>
<th>Code</th>
<th>Beschreibung</th>
</tr>
</thead>
<tbody>
<tr>
<td>phrf_soep</td>
<td>HRF SOEP integriert mit FiD 20$$ auf Personenebene</td>
</tr>
<tr>
<td>f11pbleib</td>
<td>Inverse Bleibwahrscheinlichkeit für 2011</td>
</tr>
<tr>
<td>f12pbleib</td>
<td>Inverse Bleibwahrscheinlichkeit für 2012</td>
</tr>
</tbody>
</table>

* In Version 4.0 vorläufige Hochrechnungsfaktoren für 2013
**Anhang: Neue Gewichte 2010-2012 ab der Weitergabe FiDv3.1**

Folgende Punkte führen zu Veränderungen in den Gewichten ab der Weitergabeversion 3.1

1) Änderung im Umgang mit temporären Ausfällen im SOEP und in FiD
2) Neue Gewichte für die Kohorten-Stichprobe der im 1. Quartal 2010 geborenen Kinder

---

1. *Temporäre Ausfälle*


2. *Neue Gewichte für die Kohorten-Stichprobe der im 1. Quartal 2010 geborenen Kinder*

Die Kohorten-Stichprobe der 2010 geborenen Kinder besteht aus Haushalten, die 2010 aus Einwohnermeldedaten gezogen wurden und mindestens ein Kind beinhalten, das zwischen Januar und März 2010 geboren wurde. Die Hochrechnung (siehe auch Dokumentation „hhrf_phrf“ der Version FiDv3.0, S.15) erfolgte bisher unter der Maßgabe, die vorliegenden Haushalte auf die gesamte Geburtskohorte hochzurechnen und damit die Kinder der ersten drei Monate als repräsentativ für das gesamte Jahr zu betrachten. Berechnungen der FiD-
Arbeitsgruppe am SOEP haben jedoch eine Revision nahegelegt: dabei werden die Haushalte der Kohorte 2010 entsprechend ihrem realen Anteil gewichtet und deren Hochrechnungsfaktoren im Schnitt auf ca. ein Viertel des vorherigen Wertes sinken. Dadurch kann es bei der Betrachtung bestimmter Variablen zu Unterschieden gegenüber der bisherigen Gewichtung kommen, die sich auf die hochgerechnete Anzahl an Kindern bzw. an Haushalten mit Kindern im ersten Geburtsquartal 2010 zurückführen lassen.


Veränderungen lassen sich auch für andere Variablen erwarten, in denen bestimmte Charakteristika der Kohorten-Stichprobe der im ersten Quartal 2010 geborenen Kinder, insbesondere zusammenhängend mit dem Alter dieser Kinder, eine Rolle spielen. Die bisherigen Tests und Vergleiche legen allerdings nahe, dass sich alle Änderungen innerhalb des durch die Konfidenzbänder vorgegebenen Rahmens bewegen und somit keine substantiellen Änderungen an bisherigen Ergebnissen zu erwarten sind.
Documentation $mipinc$ and $mihinc$

Multiple imputation of income variables

Mathis Fräßdorf (geb. Schröder)
## Contents

- Introduction ................................................................................................................... 222
- Methods......................................................................................................................... 225
- Evaluation of imputations ............................................................................................. 228
- Using multiple imputations ........................................................................................... 229
- Variables in `mipinc` ...................................................................................................... 231
- Evaluation graphs for `mipinc` ..................................................................................... 235
- Evaluation graphs for `mihinc` .................................................................................... 249
- References ..................................................................................................................... 253
**Introduction**

FiD employs multiple multivariate imputations for income and other monetary variables in the data. This is on two levels – the household and the individual level. For multiple imputations, it is important to jointly impute all variables, to allow for cross-variable dependencies. Hence, within each level, all variables of interest are imputed at the same time. Table 1 shows the variables which are imputed and included in the datasets *Smipinc* and *Smihinc* over data collections in 2010-2013, along with the respective number of values that had to be imputed in comparison to the non-missing values.²⁶

**Table 1a: Imputed variables on the individual level (dataset *Smipinc*)**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Ratio of missings</th>
<th>Imputed values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$PNETINC</td>
<td>Net labour income</td>
<td>0.105  0.054  0.046 0.037</td>
<td>520  288  239  187</td>
</tr>
<tr>
<td>$PGROINC</td>
<td>Gross labour income</td>
<td>0.192  0.069  0.06 0.056</td>
<td>947  370  309  286</td>
</tr>
<tr>
<td>$PMATBEN</td>
<td>Maternity benefits</td>
<td>0.028  0.014  0.015 0.000</td>
<td>27  4  3  0</td>
</tr>
<tr>
<td>$PALIMON</td>
<td>Alimony payments</td>
<td>0.036  0.013  0.012 0.008</td>
<td>31  12  10  6</td>
</tr>
<tr>
<td>$PPENS</td>
<td>Pensions/Retirement benefits</td>
<td>0.063  0.037  0.042 0.064</td>
<td>7  5  6  10</td>
</tr>
<tr>
<td>$PUEBEN</td>
<td>Unemployment benefits (ALG1)</td>
<td>0.043  0.036  0.027 0.026</td>
<td>7  4  3  3</td>
</tr>
<tr>
<td>$PWIDOW</td>
<td>Widow(er) pensions</td>
<td>0.039  0.060  0.068 0.036</td>
<td>3  5  6  3</td>
</tr>
</tbody>
</table>

Note: Non-responding household members (PUNR) are not considered in the table, but values are imputed. Given that large parts of the data are missing, these imputations should be handled with care. Also note that “imputed values” may vary by implicate, as the receipt of an income group is also subject to missing values and hence imputation.

Source: FiD 2014. No further distribution without the explicit consent of the author.

Table 1 shows quite substantial differences in the size of the missing value problem depending on the variable of interest as well on the year of the survey. In general, the missings rates decrease with increasing years and thus fewer values need to be imputed, which is not unusual for the later waves of a panel survey. All in all, the values are largely comparable to the missing ratios in the SOEP. In the individual data, the variable $PGROINC has the largest amount of missing values in each year. Part of this is due to the calculation procedure for this variable. Instead of only imputing the gross income, we also restrict it to be equal or larger than the net income. This leads to an imputed value, whenever $PGROINC or

²⁶ Note that we use the complete FiD distribution for the imputations and the descriptions as shown here, i.e. including those cases with a weight of zero. Users should be careful to be aware of these cases when using the data. Note also that the variable f10plsyrinc, “last year’s gross income”, is no longer included in the imputations, as users should prefer to use the stated by the respondent or its imputed value.
$\text{SPNETINC}$ are missing. Such a procedure is unique for $\text{SPGROINC}$, and leads to an increase in the missing percentage. On the other hand, the missing ratios for maternity benefits and child support are very low, as are the total numbers which had to be imputed for these variables. This is important for the general evaluation purpose of FiD, as reported data are usually preferred to imputed data.

### Table 1b: Imputed variables on the household level (dataset $\text{Smihinc}$)

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Ratio of missings</th>
<th>imputed values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{SHHINC}$</td>
<td>Monthly net household income</td>
<td>0.053</td>
<td>0.025</td>
</tr>
<tr>
<td>$\text{SHCHDBEN}$</td>
<td>Monthly child benefits</td>
<td>0.010</td>
<td>0.002</td>
</tr>
<tr>
<td>$\text{SHCHDADD}$</td>
<td>Monthly added child benefits</td>
<td>0.232</td>
<td>0.071</td>
</tr>
<tr>
<td>$\text{SHUEBEN2}$</td>
<td>Monthly long-term UE benefits</td>
<td>0.034</td>
<td>0.011</td>
</tr>
<tr>
<td>$\text{SHHOSBEN}$</td>
<td>Monthly housing benefits</td>
<td>0.057</td>
<td>0.015</td>
</tr>
<tr>
<td>$\text{SHCARBEN}$</td>
<td>Monthly care benefits</td>
<td>0.085</td>
<td>0.047</td>
</tr>
<tr>
<td>$\text{SHHELBEN}$</td>
<td>Monthly other benefits</td>
<td>0.182</td>
<td>0.029</td>
</tr>
<tr>
<td>$\text{SHAGETRN}$</td>
<td>Monthly age benefits</td>
<td>0.043</td>
<td>0.087</td>
</tr>
<tr>
<td>$\text{SHRENT}$</td>
<td>Monthly rent payments</td>
<td>0.006</td>
<td>0.002</td>
</tr>
<tr>
<td>$\text{SHUTIL}$</td>
<td>Monthly utility payments</td>
<td>0.597</td>
<td>0.580</td>
</tr>
<tr>
<td>$\text{SHHEAT}$</td>
<td>Monthly heating payments</td>
<td>0.221</td>
<td>0.176</td>
</tr>
<tr>
<td>$\text{SHELECTR}$</td>
<td>Monthly electricity payments</td>
<td>0.090</td>
<td>0.071</td>
</tr>
<tr>
<td>$\text{SHCREDIT}$</td>
<td>Monthly credit payments</td>
<td>0.041</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Note: “imputed values” may vary by implicate, as the receipt of an income group is also subject to missing values and hence imputation.

Source: FiD 2014. No further distribution without the explicit consent of the author.

On the household level, first note that $\text{SHRENT}$, $\text{SHUTIL}$, $\text{SHHEAT}$ and $\text{SHELECTR}$ are imputed only for those renting the accommodation they live in. The last three variables for renters have relatively large missing percentages in all years, with the highest on the utility payments, i.e. payments for trash removal, water, etc. One reason for this high percentage is
that - due to questionnaire routing - utilities have to be imputed whenever the household head states that rent does not include utilities. However, other variables have low missing percentages – notably household income, with 5.3% missing in 2010 and 2.5% or less missing since 2011.

Why then use multiple imputations? Missing information itself is a problem in any analysis if it is not missing at random, i.e. if the sample that is left by the missing data generation process is selective or different from the sample that contains missing data. Imputations per se are meant to deal with this problem, where the usual assumption is, that data from the non-missing population can be used to infer the values for the missing observations (see e.g. Starick and Watson, 2007, or Little, 1992). However, single imputations are problematic in certain aspects – while the predictions are usually good, the standard errors in analyses are reduced (see e.g., Rubin, 1996, or Schafer and Graham, 2002). This is a feature of “pretended” certainty as only one value per missing observation is imputed. Multiple imputations on the other hand try to mitigate this issue by not only imputing one, but multiple values (so-called implicates) for each missing observation. This leads to better variance estimates, as the uncertainty of the data is kept. Even though the computational effort is somewhat increased – both for the data provider and for the researcher using the imputations later – we believe that the benefits outweigh these costs.

The rest of the documentation describes the procedures applied for the multiple imputations on both levels. We then provide some measures which we used to evaluate the performance of our imputations. Finally, we briefly describe how to use the multiple imputations in practice.
Methods

The main procedure for the imputations is identical for the household and the individual level:

1. The first step is to generate a dataset, which contains all necessary information for the imputations.
2. Then, as we are using a method of chained equations (see below), for each variable that is to be imputed, a parsimonious model needs to be specified.
3. The third step then is the main part – imputing all variables according to the models in step 2.
4. The imputations are evaluated in the fourth step.
5. Finally, the data are prepared for their distribution to the user.

In step 1 the basic datasets are used to create variables which could be used in the imputation procedure. Data are taken from $h$, $p$, $pbrutto$, $hbrutto$, and $lela$. All variables were taken from the original questionnaires, although some of them are then restructured similarly to the generating process. For all income variables natural logarithms are used instead of the actual values because of the skewed distributions. Some household level characteristics are added to the imputations on the individual level. These include size of the household, number of children, whether the household belongs to the Cohort-Sample as well as the cross-sectional weights. On the household level, individual variables are used as well: these included whether any household member receives certain incomes, is working, as well as the partner structure in the household.27 Table 2 provides some more detail on the variables included on each level.

Step 2 then involves setting up equations for each variable with missing values. Here the focus is on the predictive power of each equation, not on substantial modelling. Hence, the imputations do not consider endogeneity or simultaneity issues, but are solely interested in better predictions, as long as the convergence criteria are met. For example, we tested whether the incomes could be used for prediction, i.e. use net income to predict gross income and vice versa. However, convergence was a problem in these models, likely because of the large overlap of missing values in both incomes on the one hand and the high predictive power of the variables for each other on the other hand. The predictive power did not suffer too much by not including them in the respective regressions. Starting with 2011, it is possible to use information from other years in the imputation. In principle, there are no prior reasons not to

27 Until FiD v3.1, the household imputations were independent of the individual imputations. This meant that individual variables, which were aggregated on the household level, had to be imputed in case one person’s information on that variable was missing. To reduce the computational effort, we use the results from the individual imputations (i.e. the first implicate) for household aggregated values starting with FiD v4.0.
take all available years into account in any year’s imputation. However, we only consider the previous year in the imputations (i.e. 2011 uses 2010, 2012 uses 2011, 2013 uses 2012), because the computational effort proved to be too large when considering multiple years.

The selection of equations is conducted with the stepwise command in Stata™, again keeping in mind that good prediction is the key. In specifying the stepwise command, variables with a significance level above 0.10 are removed from the model. This leads to a considerable reduction in complexity for most variables that needed imputation. With these specified models, the imputation for all variables is executed with the mi impute chained command.

We use Stata™ (release 13) for the imputations in step 3, and here the mi package first introduced with Stata 12. The imputations are done in a Multivariate Imputation by Chained Equations (or mice) procedure (see also van Buuren et al., 2006 for a detailed description). The basic idea of the mice procedure is to use a regression equation for variable $x$ to predict missing values in $x$, use these predicted values for the next equation when modelling variable $y$. The predicted values for the missing observations are then used in the next equation, until all equations have been estimated and one “cycle” is completed. The next cycle starts with the same sequence, but uses the predicted values from the first cycle. These cycles or iterations are repeated, until some measure of convergence is achieved. Also see the Stata (2013) for details.

Table 2a: Variables used in the individual-level imputations

<table>
<thead>
<tr>
<th>Context</th>
<th>Variables included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>Natural logarithms of: net labour income; gross labour income; current maternity benefits; current child support</td>
</tr>
<tr>
<td>Job characteristics</td>
<td>Hours worked; overtime; trained for job; job satisfaction; working full or part-time; being in vocational training; tenure; firm size; occupation</td>
</tr>
<tr>
<td>Demographics</td>
<td>Age; gender; foreigner; east/west; education; health; marital status</td>
</tr>
<tr>
<td>Household characteristics</td>
<td>Age, gender, and foreign status of household head; number of children; household size; number of non-respondents in household; cohort dummy; household weight, federal state; size of community (BIK)</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Current satisfaction with different aspects in the respondent’s live</td>
</tr>
</tbody>
</table>

The imputations themselves are restricted to those cases where imputation is useful. For example, labour income is not imputed for individuals who do not work, and similarly, rent is not imputed for owners. Note that if there are missing values in any of the independent
variables, imputations are imperative for these variables as well. For example, education has to be imputed for some cases. We do not distribute these variables, as they usually concern only a few cases and including them would increase datasets by an unnecessary extent.

The imputations for the individual income variables are done with overall 750 cycles and five implicates, using nearest neighbour matching with a matchpool of 3 observations to impute the data. While this takes considerable time, the evaluation shown below provides evidence that this large amount of cycles is quite useful to reach convergence. On the household level, we use 750 cycles, while also providing five implicates for each imputed variable.

<table>
<thead>
<tr>
<th>Context</th>
<th>Variables included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>Natural logarithms of: household income; monthly child benefits; added child benefits; long term unemployment benefits; housing benefits; rent; utility payments; heating payments; electricity payments; credit payments. Additionally the categories of household income from the unfolding bracket sequence.</td>
</tr>
<tr>
<td>Expenditures</td>
<td>Natural logarithms of: rent; utility payments; heating payments; electricity payments; credit payments; payments on durable and non-durable goods</td>
</tr>
<tr>
<td>Individual incomes</td>
<td>Natural logarithms of all income sources.</td>
</tr>
<tr>
<td>Individual aggregates</td>
<td>Any employed; working status; highest education; any married, divorced, or widowed;</td>
</tr>
<tr>
<td>Household characteristics</td>
<td>Age, gender, and foreign status of household head; number of children; household size; number of non-respondents in household; cohort dummy; household weight; number of rooms; square meters of housing unit; federal state; size of community (BIK)</td>
</tr>
</tbody>
</table>

The *mi chained* procedure itself does not check convergence or goodness of fit of the imputed variables, so this is up to the researcher. We use two main criteria in step 4 to evaluate the performance of our imputed values. First, we implemented the *Gelman-Rubin-Brooks* criterion (see Brooks and Gelman, 1998) in Stata™ to test convergence, which compares the within-variance within a chain of imputations with the between variance across the chains (see the next section for details). We then compared the distribution of imputed variables to the non-missing observations. Some caution needs to be taken here. By comparing the distribution of non-imputed values with those which are imputed, one should not expect them to be equal, as individuals with missing values may be different than those without missing

Table 2b: Variables used in the household-level imputations
values. Hence a comparison of distributions may only serve as a rough eyeball test for whether there are significant problems or not. The following section provides some details on the evaluation as well as the results.

Finally, in step 5, the data are prepared for the distribution. This involves generating the imputation flags which indicate which values are imputed. One variable, $PGROINC has to be specifically constructed from the imputations of other variables, as mentioned above. Some variables are integrated into $pgen and $shgen: these are $PNETINC and $PGROINC which are renamed to LABNET$$ and LABGROSS in $pgen. $HRENT, $HUTIL, $HHEAT and $HHELEC become part of $shgen. In addition, the variables ROOM$$ and SIZE$$ are taken from the imputations and added to $shgen as well. For each of the previous variables, the first and only the first implicate is used, even though this counteracts the initial purpose of the multiple imputations. We do not recommend using only one implicate, but recognize that some users may find it easier to deal with only one implicate instead of five. Also, we follow the SOEP logic with this step. As in the SOEP, for $HHINC five implicates are included for each household as I1HINC to I5HINC in $shgen. Variables included in the two multiple imputation datasets ($f10mipinc and $f10mihinc) are listed and described below.

**Evaluation of imputations**

The imputations for each variable are evaluated by criteria concerning the convergence of the different iterations on the one hand and the final distribution of the imputed values on the other.

For the evaluation of convergence we use the criteria as specified in Brooks and Gelman, 1998. The authors identify the following measures, for which we specify $m as the number of implicates, in our case $m=5. $n is the number of cycles or iterations the imputations have run until convergence is reached; in the individual case, $n=7500, while in the household case, $n=750. Then the between-implicate variance $B/n is defined as follows (see Brooks and Gelman, page 436):

\[
B/n = \frac{1}{m-1} \sum_{j=1}^{m} \left( \bar{\phi}_j - \bar{\phi} \right)^2
\]

and the within-implicate variance, $W, is

\[
W = \frac{1}{m(n-1)} \sum_{j=1}^{m} \sum_{t=1}^{n} \left( \bar{\phi}_{jt} - \bar{\phi}_j \right)^2
\]

Brooks and Gelman then continue to define the posterior variance estimate:

---

28 Note that in version 4.0 of the FiD data, no imputations were necessary for the number of rooms in either year.
Finally, the potential scale reduction factor, or PSRF, is given as the Ratio of $V$ and $W$:

\[
\hat{R} = \frac{\hat{V}}{W} = \frac{n-1}{n} W + \frac{m+1}{mn} B
\]

(Note that Brooks and Gelman apply a correction factor $d$ to their statistic, which for simplicity is ignored here.)

Brooks and Gelman define convergence to be achieved if:

a) $\hat{V}$ stabilizes as a function of $n$,

b) $W$ stabilizes as a function of $n$, and

c) $\hat{R}$ approaches 1, implying that $\hat{V}$ and $W$ converge to one another.

The convergence is monitored graphically, where the number of iterations is plotted against the three statistics above. (Note that we use the same approach as Brooks and Gelman in their paper, where we only consider certain steps in the convergence process, and hence restrict the number of points to 40 at all times.) $\hat{R}$, $\hat{V}$, and $W$ can be computed for different statistics of the respective variable. We consider the two most common here, the mean and the standard deviation.

To determine the performance of our final imputations, we compare the distribution of the five implicates with the original, non-imputed distribution. Note, however, that an argument can be made for the distributions not to match, as the population with missing observations may be different in the variable(s) of interest to the population with non-missing values. E.g., if people with high incomes are more likely not to report their income, the distribution for the imputed values should be shifted right compared to the non-imputed observations.

The graphs for the Gelman-Rubin-Brooks criterion for mean and standard deviation as well as the comparison of the densities can be found with the variable descriptions below for the most important variables: net and gross individual income as well as household net income. Graphs for other variables are available on request.

**Using multiple imputations**

The datasets of multiply imputed variables are distributed separately from all others. They only contain the imputed variables and their imputation flags, indicating whether a value had been imputed or not. Individual data is identified via the PERSNR identifier, while households (in both datasets) are identified through $HHNR$. To use the imputation datasets for analysis, they need to be merged with other data using the respective identifiers.
As multiple imputations become more common in social science datasets, statistical packages now allow for their use. When using multiple imputations, means, coefficients and variances have to be adjusted to take into account the different implicates. Basically, any analysis has to be conducted as many times as there are implicates, and then coefficients can be combined and joint variances calculated. Following Rubin (1987), the coefficient of $m$ regressions using each of the implicates $j=1,\ldots,m$ lead to a combined estimate of

$$
\bar{\beta}_m = \frac{1}{m} \sum_{j=1}^{m} \hat{\beta}_j
$$

For the variance, the within-imputation variance needs to be combined with the between-imputation variance. The former is derived as the mean of the variances from each imputation regression

$$
\bar{W}_m = \frac{1}{m} \sum_{j=1}^{m} \text{Var} \left( \hat{\beta}_j \right),
$$

while the latter is computed as the variance of the coefficients over the $m$ implicates:

$$
B_m = \frac{1}{m-1} \sum_{j=1}^{m} (\hat{\beta}_j - \bar{\beta}_m)^2
$$

The joint variance is then given by

$$
T_m = \bar{W}_m + \frac{m+1}{m} B_m
$$

While the user may calculate these statistics him or herself, statistical software provides these and more applications for multiple imputations. Since the distribution of Stata 11 a complete package on multiple imputations is included – allowing to impute as well as using the imputations in analyses. See the $mi$ command in that package. In addition, there are several added programs for analyzing multiple imputed data, for example $mim$. Within SAS, PROC MIANALYZE combines the results of analyses on the different data sets. IVEware is a set of routines that can be launched from SAS or run independently using data from many sources. You can use the IVEware module $regress$ to perform multiple imputation analysis.
Variables in $mipinc$

_MI
Variable label “observation number”
Variable format 5-digit integer
Comment _mi identifies observations belonging to the same household. For each person, there are 6 observations – one original and 5 imputed values.

_MJ
Variable label “imputation number”
Variable format 4-digit integer
Comment _MJ identifies the implicates for each person and variable. The range is from 0 to 5, where 0 denotes the original observation and 1-5 identify the respective imputations.

PUNR$
Variable label “Non-responding household member”
Value label PUNR$
   (0) Regular respondent
   (1) Non-respondent
Variable format 1-digit integer
Comment This variable identifies, whether an observation stems from a non-responding household member (partial unit non-response, PUNR) or not. Note that values are imputed for all household members, even if they decide not to participate in the survey.

$PNETINC
Variable label “monthly net labor earnings (imputed)”
Variable format 5-digit integer
Comment The original question in the person questionnaire is about the net income from labor earnings for each working individual. Hence “-2” (does not apply) is set for all non-working individuals. The natural logarithm of this variable was imputed, and then converted back to the monetary values.

$PGROINC
Variable label “monthly gross labor earnings (imputed)”
Variable format 5-digit integer
Comment The original question in the person questionnaire is about the gross income from labor earnings for each working individual. Hence “-2” (does not apply) is set for all non-working individuals. The imputation
was conducted using the difference between net and gross income, imputing it, and then adding it to the imputed net income. As mentioned above, this leads to more imputed values than were originally missing. However, the consistency here seemed more important.

$PMATBEN
Variable label  “maternity benefits (imputed)”
Variable format  4-digit integer
Comment  The origin of this variable is the question for the maternity benefit while being on some sort of maternity leave (“Elterngeld”, “Erziehungsgeld”). “-2” (does not apply) is set for all individuals who state not to receive any maternity benefits. The natural logarithm of this variable was imputed, and then converted back to the monetary values.

$PALIMON
Variable label  “child alimony (imputed)”
Variable format  4-digit integer
Comment  This variable originates in the question for child support received individually (“Kindesunterhalt”). Note that “-2” (does not apply) is set for all individuals who do not receive any child support. The natural logarithm of this variable was imputed, and then converted back to the monetary values.

$PUEBEN
Variable label  “unemployment benefits(imputed)”
Variable format  4-digit integer
Comment  This variable originates in the question for unemployment benefits (“Arbeitslosengeld”). Note that this is not “ALG2”, which is received on the household level. Note that “-2” (does not apply) is set for all individuals who do not receive any benefits. The natural logarithm of this variable was imputed, and then converted back to the monetary values.

$PWIDOW
Variable label  “widow(er) pensions (imputed)”
Variable format  4-digit integer
Comment  This variable originates in the question for pensions received as a widow or orphan (“Witwen/Waisenrente”). Note that “-2” (does not apply) is set for all individuals who do not receive any pensions. The natural logarithm of this variable was imputed, and then converted back to the monetary values.
**$PPENS**
Variable label: “pensions/retirement (imputed)”
Variable format: 4-digit integer

Comment: This variable originates in the question for pensions and retirement benefits received individually (“Rente/Pension”). Note that “-2” (does not apply) is set for all individuals who do not receive any pensions. The natural logarithm of this variable was imputed, and then converted back to the monetary values.

**I_$PNETIC**
Variable label: “pnetic is imputed”
Value label: I_$PNETIC
(0) Observed value
(1) Imputed value
Variable format: 1-digit integer

Comment: Note that “-2” (does not apply) is set for all non-working individuals.

**I_F10PGROINC**
Variable label: “pgroinc is imputed“
Value label: I_$PGROINC
(0) Observed value
(1) Imputed value
Variable format: 1-digit integer

Comment: Note that “-2” (does not apply) is set for all non-working individuals.

**I_$PMATBEN**
Variable label: “pmatben is imputed”
Value label: I_$PMATBEN
(0) Observed value
(1) Imputed value
Variable format: 1-digit integer

Comment: Note that “-2” (does not apply) is set for all individuals who do not receive any maternity benefits.

**I_$PALIMON**
Variable label: “palimon is imputed”
Value label: I_$PALIMON
(0) Observed value
(1) Imputed value
Variable format: 1-digit integer
Comment Note that “-2” (does not apply) is set for all individuals who do not receive any child support.

\textbf{I\_$PUEBEN}

Variable label “$pueben is imputed”
Value label I\_$PUEBEN
(0) Observed value
(1) Imputed value
Variable format 1-digit integer
Comment Note that “-2” (does not apply) is set for all individuals who do not receive any unemployment benefits.

\textbf{I\_$PWIDOW}

Variable label “$pwidow is imputed”
Value label I\_$PWIDOW
(0) Observed value
(1) Imputed value
Variable format 1-digit integer
Comment Note that “-2” (does not apply) is set for all individuals who do not receive any widow(er) benefits.

\textbf{I\_$PPENS}

Variable label “$ppens is imputed”
Value label I\_$PPENS
(0) Observed value
(1) Imputed value
Variable format 1-digit integer
Comment Note that “-2” (does not apply) is set for all individuals who do not receive any pension benefits.
**Evaluation graphs for $mipinc**

**F10PNETINC**

Gelman-Rubin Brooks Criteria: Mean

Gelman-Rubin Brooks Criteria: Standard Deviation

Comparison of Densities between original distribution (blue) and the five implicate
F11PNETINC

Gelman-Rubin Brooks Criteria: Mean

Gelman-Rubin Brooks Criteria: Standard Deviation

Comparison of Densities between original distribution (blue) and the five implicates
F12PNETINC

Gelman-Rubin Brooks Criteria: Mean

Gelman-Rubin Brooks Criteria: Standard Deviation

Comparison of Densities between original distribution (blue) and the five implicates
F13PNETINC
Gelman-Rubin Brooks Criteria: Mean

Gelman-Rubin Brooks Criteria: Standard Deviation

Comparison of Densities between original distribution (blue) and the five implicates
F10PGROINC

Gelman-Rubin Brooks Criteria: Mean (log-difference of net and gross income)

Gelman-Rubin Brooks Criteria: Standard Deviation (log-difference of net and gross income)

Comparison of Densities between original distribution (blue) and the five implicates
F11PGROINC

Gelman-Rubin Brooks Criteria: Mean

Gelman-Rubin Brooks Criteria: Standard Deviation

Comparison of Densities between original distribution (blue) and the five implicates
F12PGROINC

Gelman-Rubin Brooks Criteria: Mean (log-difference of net and gross income)

Gelman-Rubin Brooks Criteria: Standard Deviation (log-difference of net and gross income)

Comparison of Densities between original distribution (blue) and the five implicates
**F13PGROINC**

Gelman-Rubin Brooks Criteria: Mean (log-difference of net and gross income)

![Gelman-Rubin Brooks Diagnostics for lnroin13_mean](image1)

Gelman-Rubin Brooks Criteria: Standard Deviation (log-difference of net and gross income)

![Gelman-Rubin Brooks Diagnostics for lnroin13_sd](image2)

Comparison of Densities between original distribution (blue) and the five implicates

![Densities of lnroinc13](image3)
Variables in $mihinc$

_MI
Variable label          “observation number”
Variable format         4-digit integer

Comment   _mi identifies observations belonging to the same household. For each household, there are 6 observations – one original and 5 imputed values.

_MJ
Variable label         “imputation number”
Variable format         1-digit integer

Comment   _mj identifies the implicates for each household and variable. The range is from 0 to 5, where 0 denotes the original observation and 1-5 identify the respective imputations.

$HHINC
Variable label         “monthly net household income”
Variable format         5-digit integer

Comment   This variable is the so-called income “screener” and records the overall household net income at the current interview month. Note that all households are “eligible” to have a household income, hence the code “-2” (does not apply) is only set for this variable in the original variable (where _MJ=0). The natural logarithm of this variable was imputed, and then converted back to the monetary values.

$HCHDBEN
Variable label         “monthly child benefits (‘Kindergeld’)”
Variable format         4-digit integer

Comment   This variable originates from the current report of child benefits. Note that “-2” (does not apply) is set for all households which do not receive child benefits (“Kindergeld”).

$HCHDADD
Variable label         “monthly added child benefits (‘Kinderzuschlag’)”
Variable format         3-digit integer

Comment   The “Kinderzuschlag” question concerns the specific benefits given to households with low income, which are not receiving any ALG2 benefits. It is not quite clear, whether individuals have correctly understood this question, as the name of the benefit is somewhat ambiguous. Note that “-2” (does not apply) is set for all households
which do not receive the added child benefit for low income households.

$\text{HUEBEN2}$
Variable label: “monthly long-term UE benefits (‘ALG2’)”
Variable format: 4-digit integer
Comment: This variable is based on the household receipt of long term unemployment benefits (‘ALG2’ or “Hartz IV”). Note that “-2” (does not apply) is set for all households which do not receive any long-term unemployment benefits.

$\text{HCARBEN}$
Variable label: “monthly care benefits ‘PflegeVers’ (imputed)”
Variable format: 3-digit integer
Comment: This variable covers the imputations for the household care benefits (‘Pflegeversicherung’). Note that “-2” (does not apply) is set for all households which do not report any care benefits.

$\text{HHELBEN}$
Variable label: “monthly other help benefits ‘Hilfe Lebenslagen’ (imputed)”
Variable format: 3-digit integer
Comment: This variable covers the imputations for the household’s other help benefits (‘Hilfe in besonderen Lebenslagen’). Note that “-2” (does not apply) is set for all households which do not report to receive any of these benefits.

$\text{HAGETRN}$
Variable label: “monthly age transfer benefit ‘Grundsicherung Alter’ (imputed)”
Variable format: 3-digit integer
Comment: This variable covers the imputations for the household receipt of age transfer benefits (‘Grundsicherung im Alter und bei Erwerbsminderung’). Note that “-2” (does not apply) is set for all households which do not report to receive any of these benefits.

$\text{HHOSBEN}$
Variable label: “monthly housing benefits (‘Wohngeld’)”
Variable format: 3-digit integer
Comment: The origin is in the question about housing benefits received by the household (‘Wohngeld’). Note that “-2” (does not apply) is set for all households which do not receive any housing benefits.
$HRENT
Variable label
“monthly rent payments”
Variable format
4-digit integer
Comment
The origin of this variable is the question about rent payments. As concerns the imputation, there is no distinction between gross and net rent, although the information, whether utilities are included in the rent payments is part of the imputation model and thus is accounted for. Note that “-2” (does not apply) is set for all owner-occupier households and renter-occupier households which do not pay rent. Note also that this variable does not correspond to the variable rent$$ in Shgen.

$HUTIL
Variable label
“monthly utility payments”
Variable format
3-digit integer
Comment
The variable originates in the report of the monthly utility costs. There is no distinction between utility costs which are partially or fully included in the rent, although a corresponding indicator variable is included in the imputation process. Note that “-2” (does not apply) is set for all owner-occupier households and households reporting not to pay rent or any utilities.

$HHEAT
Variable label
“monthly heating payments”
Variable format
3-digit integer
Comment
The variable originates in the report (or the respondent’s estimate) of the monthly heating costs. Note that “-2” (does not apply) is set for all owner-occupier households and households reporting not to pay rent.

$HELEC
Variable label
“monthly electricity payments”
Variable format
3-digit integer
Comment
Electricity payments are derived from the report of each renter on these costs. Note that “-2” (does not apply) is set for all owner-occupier households and households reporting not to pay rent.

$HCREDIT
Variable label
“monthly credit payments”
Variable format
3-digit integer
Comment: This variable covers the imputations for the household credit payments which are not mortgages. Note that “-2” (does not apply) is set for all households which do not report any credit burden.

$I_{\text{HHINC}}$
Variable label: “$\text{hhinc is imputed}$”
Value labels:
- (0) $I_{\text{HHINC}}$: Observed value
- (1) $I_{\text{HHINC}}$: Imputed value
Variable format: 1-digit integer

Comment: Note that all households are “eligible” to have a household income, hence the code “-2” (does not apply) is only set for this variable in the original variable (where _MJ=0).

$I_{\text{HCHDBEN}}$
Variable label: “$\text{hchdsup is imputed}$”
Value labels:
- (0) $I_{\text{HCHDBEN}}$: Observed value
- (1) $I_{\text{HCHDBEN}}$: Imputed value
Variable format: 1-digit integer

Comment: Note that “-2” (does not apply) is set for all households which do not receive child benefits (“Kindergeld”).

$I_{\text{HCHDADD}}$
Variable label: “$\text{hchdadd is imputed}$”
Value labels:
- (0) $I_{\text{HCHDAAD}}$: Observed value
- (1) $I_{\text{HCHDAAD}}$: Imputed value
Variable format: 1-digit integer

Comment: Note that “-2” (does not apply) is set for all households which do not receive the added child benefit for low income households (“Kinderzuschlag”).

$I_{\text{HUEBEN2}}$
Variable label: “$\text{hueben is imputed}$”
Value labels:
- (0) $I_{\text{HUEBEN2}}$: Observed value
- (1) $I_{\text{HUEBEN2}}$: Imputed value
Variable format: 1-digit integer

Comment: Note that “-2” (does not apply) is set for all households which do not receive any long-term unemployment benefits (“ALG2”).
**I\_HCARBEN**

**Variable label**

"$hcarben is imputed"

**Value labels**

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Observed value</td>
</tr>
<tr>
<td>1</td>
<td>Imputed value</td>
</tr>
</tbody>
</table>

**Variable format**

1-digit integer

**Comment**

Note that "-2" (does not apply) is set for all households which do not receive any care benefits (“Pflegeversicherung”).

---

**I\_HHELBEN**

**Variable label**

"$hhelben is imputed"

**Value labels**

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Observed value</td>
</tr>
<tr>
<td>1</td>
<td>Imputed value</td>
</tr>
</tbody>
</table>

**Variable format**

1-digit integer

**Comment**

Note that "-2" (does not apply) is set for all households which do not receive any help benefits (“Hilfe in besonderen Lebenslagen”).

---

**I\_HAGETRN**

**Variable label**

"$hagetrn is imputed"

**Value labels**

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Observed value</td>
</tr>
<tr>
<td>1</td>
<td>Imputed value</td>
</tr>
</tbody>
</table>

**Variable format**

1-digit integer

**Comment**

Note that "-2" (does not apply) is set for all households which do not receive age transfer benefits ('Grundsicherung im Alter und bei Erwerbsminderung')

---

**I\_HHOSBEN**

**Variable label**

"$hhosben is imputed"

**Value labels**

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Observed value</td>
</tr>
<tr>
<td>1</td>
<td>Imputed value</td>
</tr>
</tbody>
</table>

**Variable format**

1-digit integer

**Comment**

Note that "-2" (does not apply) is set for all households which do not receive any housing benefits (“Wohngeld”).

---

**I\_HRENT**

**Variable label**

"$hrent is imputed"

**Value labels**

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Observed value</td>
</tr>
<tr>
<td>1</td>
<td>Imputed value</td>
</tr>
</tbody>
</table>
Variable format 1-digit integer
Comment Note that “-2” (does not apply) is set for all owner-occupier households and renter-occupier households which do not pay rent.

I_$HUTIL
Variable label “Shutil is imputed”
Value labels
I_$HUTIL
(0) Observed value
(1) Imputed value
Variable format 1-digit integer
Comment Note that “-2” (does not apply) is set for all owner-occupier households and households reporting not to pay rent or any utilities.

I_$HHEAT
Variable label “$hheat is imputed”
Value labels
I_$HHEAT
(0) Observed value
(1) Imputed value
Variable format 1-digit integer
Comment Note that “-2” (does not apply) is set for all owner-occupier households and households reporting not to pay rent.

I_$HELEC
Variable label “$helec is imputed”
Value labels
I_$HELEC
(0) Observed value
(1) Imputed value
Variable format 1-digit integer
Comment Note that “-2” (does not apply) is set for all owner-occupier households and households reporting not to pay rent.

I_$HCREDIT
Variable label “$hcredit is imputed”
Value labels
I_$HCREDIT
(0) Observed value
(1) Imputed value
Variable format 1-digit integer
Comment Note that “-2” (does not apply) is set for all households which do not report any credit burden.
Evaluation graphs for $mihinc$

F10HHINC

Gelman-Rubin Brooks Criteria: Mean

Gelman-Rubin Brooks Criteria: Standard Deviation

Comparison of Densities between original distribution (blue) and the five implicates
F11HHINC

Gelman-Rubin Brooks Criteria: Mean

Gelman-Rubin Brooks Criteria: Standard Deviation

Comparison of Densities between original distribution (blue) and the five implicates
**F12HHINC**

Gelman-Rubin Brooks Criteria: Mean

![Gelman-Rubin Brooks Criteria: Mean](image1)

Gelman-Rubin Brooks Criteria: Standard Deviation

![Gelman-Rubin Brooks Criteria: Standard Deviation](image2)

Comparison of Densities between original distribution (blue) and the five implicates

![Comparison of Densities](image3)
F13HHINC
Gelman-Rubin Brooks Criteria: Mean

Gelman-Rubin Brooks Criteria: Standard Deviation

Comparison of Densities between original distribution (blue) and the five implicates
References


Documentation pbiospe

Activity Biography

Stefan Damerow

This documentation is based on the comparable SOEP documentation on pbiospe and has benefited from previous work by Henning Lohmann. Please understand that for readability reasons, we do not specifically cite and specify text that has been used directly from the SOEP document.
The spell file pbiospe is based on the information on activity status over the life course, which is collected as a matrix from every respondent using the Biography Questionnaire (in FiD, this belongs to the second part of the biography, i.e. Slela with $LELTYP=2 \text{ or } 3$). The observations start at the age of 15 and end at the current age (up to age 65). The information on the activity status covers only the period up to the time the biography is collected. To update the ongoing occupational career in pbiospe, information from the yearly Person Questionnaire is also used. In this questionnaire, respondents are always asked their occupational status for every month of the previous year. Therefore, the information collected on a monthly basis and stored in the file artkalen is aggregated into yearly values and combined with the information gathered from the Biography Questionnaire.

In the following, the method of combining the data is described. First of all we provide a brief overview of the contents of pbiospe. Table 1 contains a list of all the variables in the dataset. The variables BEGIN and END indicate the beginning and the end of a spell. These variables are age entries. There are also variables that refer to calendar years: BEGINY and ENDY (Y stands for Year). The variable SPELLTYP contains information on the activity status during the spell, e.g., employed full-time or unemployed. The SPELLNR is a serial identifier of spells of a given person. Missing information on the beginning or end of a spell cause what are known as censoring problems. There are two types of missing data. First, data can be missing on periods outside the observation window (before the age of 15 and after the age of 65). Second, data can be missing on years within the observation window due to item non-response in particular years or due to temporary drop-outs (the latter applies to calendar information only). In this case, we speak of “gaps.” There are nine different patterns (variable ZENSOR):

1. uncensored: beginning observed, end observed
2. right-censored: beginning observed, end not observed
3. right-censored (gap): beginning observed, end not observed because of gap
4. left-censored: beginning not observed, end observed
5. left- and right-censored: beginning not observed, end not observed
6. left-censored and right-censored (gap): beginning not observed, end not observed because of gap
7. left-censored (gap): beginning not observed because of gap, end observed
8. left-censored (gap) and right-censored: beginning not observed because of gap, end not observed

9. left-censored (gap) and right-censored (gap): beginning not observed because of gap, end not observed because of gap

Table 1: Contents of pbiospe (variables)\(^\text{29}\)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>hhnr</td>
<td>Original Household Number</td>
</tr>
<tr>
<td>persnr</td>
<td>Never Changing Person ID</td>
</tr>
<tr>
<td>spellnr</td>
<td>Serial Number of the Event per Person</td>
</tr>
<tr>
<td>spelltyp</td>
<td>Type of Event</td>
</tr>
<tr>
<td>begin</td>
<td>Age spell begins</td>
</tr>
<tr>
<td>end</td>
<td>Age spell ends</td>
</tr>
<tr>
<td>beginy</td>
<td>Year spell begins</td>
</tr>
<tr>
<td>endy</td>
<td>Year spell ends</td>
</tr>
<tr>
<td>zensor</td>
<td>Zensor Variable</td>
</tr>
<tr>
<td>spellinf</td>
<td>Spell construction information</td>
</tr>
<tr>
<td>erhebj</td>
<td>Survey year biography data</td>
</tr>
<tr>
<td>kalyear</td>
<td>First observation year calendar</td>
</tr>
<tr>
<td>beginb1</td>
<td>Age spell begins, 1. initial biography spell</td>
</tr>
<tr>
<td>endb1</td>
<td>Age spell ends, 1. initial biography spell</td>
</tr>
<tr>
<td>begink1</td>
<td>Age spell begins, 1. initial calendar spell</td>
</tr>
<tr>
<td>endk1</td>
<td>Age spell ends, 1. initial calendar spell</td>
</tr>
<tr>
<td>beginyb1</td>
<td>Year spell begins, 1. initial biography spell</td>
</tr>
<tr>
<td>endyb1</td>
<td>Year spell ends, 1. initial biography spell</td>
</tr>
<tr>
<td>beginyk1</td>
<td>Year spell begins, 1. initial calendar spell</td>
</tr>
<tr>
<td>endyk1</td>
<td>Year spell ends, 1. initial calendar spell</td>
</tr>
<tr>
<td>beginb2</td>
<td>Age spell begins, 2. initial biography spell</td>
</tr>
<tr>
<td>endb2</td>
<td>Age spell ends, 2. initial biography spell</td>
</tr>
<tr>
<td>beginyb2</td>
<td>Year spell begins, 2. initial biography spell</td>
</tr>
<tr>
<td>endyb2</td>
<td>Year spell ends, 2. initial biography spell</td>
</tr>
<tr>
<td>begink2</td>
<td>Age spell begins, 2. initial calendar spell</td>
</tr>
<tr>
<td>endk2</td>
<td>Age spell ends, 2. initial calendar spell</td>
</tr>
<tr>
<td>beginyk2</td>
<td>Year spell begins, 2. initial calendar spell</td>
</tr>
<tr>
<td>endyk2</td>
<td>Year spell ends, 2. initial calendar spell</td>
</tr>
</tbody>
</table>

As mentioned above, pbiospe combines information collected in the biography questionnaire and the calendar matrix of the individual questionnaire. The two types of information are merged into pbiospe following a number of rules. First of all, it is important to acknowledge that the Biography Questionnaire Matrix as well as the Individual Questionnaire Matrix allow

\(^{29}\) In SOEP the data file pbiospe also contains the variables beginb3 – endyk4. These additional variables become relevant for longer panel durations only (for further explanation see below).
for multiple activity statuses for a given year or month. No concept of main activity is used. A common combination is, for instance, “housewife/-husband” and “working part-time”. There are a number of other plausible combinations, but also combinations that are less plausible. However, a list of valid combinations of activity statuses defined according to legal or similar constructs would need to be based on very strong assumptions. In addition—in particular in case of the yearly matrix in the Biography Questionnaire—activities are reported that took place in a calendar year in consecutive months, which makes it impossible to exclude combinations of activities. Therefore, no data cleaning is performed at this stage. As a consequence, the data may contain information on more than one activity for a given point in time.

This also defines the rules for aggregating the monthly *artkalen* data into yearly values. Take, for example, a person who was in full-time employment from January to November 2007, and unemployed in December 2007. The exact months are recorded in the dataset *artkalen*. In the aggregated data, which is merged with the yearly data from the Biography Questionnaire, you find the information that the person worked full-time and was also unemployed in the year 2007. There is a second level of aggregation of *artkalen* information as the data on type of activity, which is recorded in the variable SPELLTYP is more detailed than in *pbiospe*. The respective information is aggregated as described in Table 2.

**Table 2: Aggregation of *artkalen* spell information into *pbiospe***

<table>
<thead>
<tr>
<th><em>pbiospe</em></th>
<th><em>artkalen</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 School/University</td>
<td>School, College (1)</td>
</tr>
<tr>
<td>2 Apprenticeship/Training</td>
<td>Vocational Training (4), First Job Training, Apprenticeship (13), Continuing Education, Retraining (14)</td>
</tr>
<tr>
<td>3 Military/Civilian service</td>
<td>Military, Community Service (9)</td>
</tr>
<tr>
<td>4 Full-time employed</td>
<td>Full-Time Employment (1), Short Work Hrs (2)</td>
</tr>
<tr>
<td>5 Part-time employed</td>
<td>Part-Time Employment (3), Second Job (11), Mini-job (up to 400 euros) (15)</td>
</tr>
<tr>
<td>6 Unemployed</td>
<td>Unemployed (5)</td>
</tr>
<tr>
<td>7 House-Husband/Wife</td>
<td>Housewife, Husband (10)</td>
</tr>
<tr>
<td>8 Retired</td>
<td>Retired (6)</td>
</tr>
<tr>
<td>9 Other</td>
<td>Maternity Leave (7), Other (12)</td>
</tr>
<tr>
<td>99 Gap</td>
<td>Information on gaps in <em>artkalen</em> is not used. Gaps are calculated on the basis of the merged dataset.</td>
</tr>
</tbody>
</table>

After merging the information from the Biography Questionnaire and *artkalen*, the data are transformed into spells, whereby each spell is defined by the duration of a given status. A
question that arises when merging the data is how to handle overlapping pieces of information. The basic principle is to assign a value of a given status in a given year if the status is recorded in the calendar or in the biography information or both. An example might help to illustrate this: the calendar records full-time employment for the years 2010 and 2011 while the biography records full-time employment for the period from 2000 up to 2011. The merged data from pbiospe contains a spell that begins in 2000 and ends in 2011. However, the initial information is restored by including additional variables, which allows for alternative ways of merging the data (see below). The variables SPELLINF, ERHEBJ, and KALYEAR contain general information on the sources of the information captured in a given spell. Table 3 shows that the majority of spells are based on biography information only (64.3 percent). Almost one sixth of all spells (15.5 percent) are not observed in the Biography Questionnaire but only in the calendar data. The remainder of spells contains information from biography as well as calendar data. Usually these spells combine one period observed in the Biography Questionnaire with a period observed in the calendar.

### Table 3: Sources of pbiospe spells

<table>
<thead>
<tr>
<th>Spell construction information</th>
<th>n</th>
<th>%</th>
<th>cum. %</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-2] Does not apply</td>
<td>118</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>[1] Biography only</td>
<td>32,606</td>
<td>61.75</td>
<td>61.97</td>
</tr>
<tr>
<td>[2] Calendar only</td>
<td>9,809</td>
<td>18.58</td>
<td>80.54</td>
</tr>
<tr>
<td>[3] 1 Biography + 1 Calendar spell</td>
<td>9,950</td>
<td>18.84</td>
<td>99.39</td>
</tr>
<tr>
<td>[4] 2+ Biography and 1 Calendar spell(s)</td>
<td>164</td>
<td>0.31</td>
<td>99.7</td>
</tr>
<tr>
<td>[5] 1 Biography and 2+ Calendar spell(s)</td>
<td>150</td>
<td>0.28</td>
<td>99.98</td>
</tr>
<tr>
<td>[6] 2+ Biography and 2+ Calendar spell(s)</td>
<td>10</td>
<td>0.02</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>52,807</td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

Source: FiDv4.0

The variables BEGINB1-ENDYB2 document the initial information from the two different sources and are probably not of interest to the majority of users. However, on the basis of these variables, users are able to fully separate the Biography data from the aggregated artkalen data. This is advisable if you want to use the more detailed artkalen information and combine it with the yearly information from pbiospe for earlier years only. The variable names indicate the “source” of the original information utilized (B: Biography -Questionnaire

---

30 Again, the variety of combinations of biography and calendar spells will increase in upcoming waves.
or K: calendar information from the yearly survey). As an example from SOEP v27, we discuss one of the spells that combines information on more than one period from any of the two sources. The spell number 4 of person 9205 starts in 1983 and ends in 1994 (SPELLTYP=4: full-time employment). As the variable SPELLINF (=5) shows, this a spell that combines one period from the biography data with two periods from the calendar data. According to the biography data, the person worked full-time from 1983 (BEGINYB1) until 1992 (ENDYB1). There is overlapping information from the calendar data available from 1986 onwards (KALYEAR). According to these data, the person worked full-time from 1986 (BEGINYK1) to 1990 (ENDYK1) and from 1993 (BEGINYK2) to 1994 (ENDYK2). During the years 1991 and 1992, no full-time employment is recorded in the calendar data, which contradicts the information from the biography data.

Table 5: Example of combined spell

<table>
<thead>
<tr>
<th>persnr</th>
<th>spellnr</th>
<th>spelltyp</th>
<th>beginy</th>
<th>endy</th>
<th>spellinf</th>
<th>erhebj</th>
<th>kalyear</th>
<th>beginyb1</th>
<th>endyb1</th>
<th>beginyk1</th>
<th>endyk1</th>
<th>beginyk2</th>
<th>endyk2</th>
</tr>
</thead>
</table>

Source: SOEP v27 (*pbiospe*).

In *pbiospe*, no attempt is made to “resolve” such contradictions, as this would require rather strong assumptions. More important, such assumptions would differ according to the research question, which makes it even more difficult to provide a standard solution. Therefore, in such cases, we generate spells in the same manner as in less difficult cases, namely by combining the information from the calendar and the biography data. However, users who are interested in combining biography and calendar data in a different manner can use the variables BEGINB1-ENDYB2 to fully separate the two types of data and to recombine the data on the basis of different rules of aggregation.
Documentation *pbrutto*

Gross information on all persons in the household

*Note that part of this information is in German, based on the codebook provided by TNS-Infratest for FiD and SOEP.*
List of variables:

$GEBURT .............................................................................................................................. 262
$SEX ...................................................................................................................................... 262
$PNAT ............................................................. 262
$STISTAT .............................................................................................................................. 263
$BEFSTAT ............................................................................................................................. 263
$LINT ............................................................. 264
$LUECKE .............................................................................................................................. 265
$ZUPAN ............................................................. 265
$PNRAKT .............................................................................................................................. 265
$STELL ................................................................................................................................. 265
$SPZUG ............................................................. 268
$SPFORM .............................................................................................................................. 269
$SPERG ................................................................................................................................. 269
$SPERGZ ................................................................................................................................. 270
$SPADER ................................................................................................................................. 270
$SPADERQ ................................................................................................................................. 271
$SAUSZUGM ........................................................................................................................... 271
$SAUSZUGJ ........................................................................................................................... 271
$SEINZUGM ........................................................................................................................... 272
$SEINZUGJ ........................................................................................................................... 272
$SABWESM ........................................................................................................................... 272
$SABWESJ ........................................................................................................................... 272
$PBIO ................................................................................................................................. 272
$SDJ ................................................................................................................................. 273
$SEWSTATU ........................................................................................................................... 273
$EX (SE1-SE6) ........................................................................................................................... 273
$GEBURT
Variable Label „Geburtsjahr“
Value Labels $GEBURT
Variable format 4-digit integer
$ - Wave F10, F11, F12, F13
Comment Contains the year of birth the sample member. Is used to determine how the sample member should be treated (e.g. new or old respondent). See also $BEFSTAT.

$SEX
Variable Label „Geschlecht“
Value Labels $SEX
   (1) männlich
   (2) weiblich
Variable format 1-digit integer
$ - Wave F10, F11, F12, F13
Comment Respondent gender. The code “9 unbekannt” is not set in FiD.

$PNAT
Variable Label „Staatsangehörigkeit“
Value Labels $PNAT

(001) Deutschland (035) Argentinien (069) Liechtenstein
(002) Türkei (036) Cap Verde (070) Island
(003) Jugoslawien/ (037) Benin (071) Irland
   Montenegro/
   Serbien
(004) Griechenland (038) Philippinen (072) St.Lucia
(005) Italien (039) Israel (073) Moldawien
(006) Spanien (040) Japan (074) Kasachstan
(007) Ex-DDR (041) Australien (075) Albanien
(010) Österreich (042) Indien (076) Libanon
(011) Frankreich (043) Afghanistan (077) Kirgistan
(012) Benelux (044) Thailand (078) Ukraine
(013) Dänemark (045) Jamaika (079) Algerien
(014) Großbritannien (046) Saudi Arabien (080) Mosambik
(015) Schweden (047) Äthiopien (081) Ägypten
(016) Norwegen (049) Ghana (082) Tadschikistan
(017) Finnland (050) Bangladesch (083) Vietnam
(018) USA (051) Venezuela (084) Somalia
(019) Schweiz (052) Tunesien (085) Pakistan
(020) Chile (053) Mauritius (086) Südafrika
(021) Rumänien (054) Nigeria (087) Ver.Arabische Emirate
(022) Polen (055) Kanada (088) ElSalvador
(023) Korea (056) Neuseeland (089) Eritrea
(024) Iran (057) Tansania (090) Jordanien
(025) Indonesien (058) Zypern (092) CostaRica
(026) Ungarn (059) Kuba (093) Singapur
(027) Bolivien (060) Irak (094) Burkina Faso
(028) Portugual (061) Brasilien (095) Sambia
(029) Bulgarien (062) Monaco (096) Ecuador
(030) Syrien (063) Hongkong (097) Usbekistan
(031) Tschechien (064) Peru (098) Staatenlos
(032) Russland (065) SriLanka (099) Puerto Rico
(034) Mexiko (066) Nepal (100) Laos
(036) Kap Verde (067) Marokko (101) Estland
(037) Benin (068) China (102) Angola

SOEP Survey Paper 332 262
(103) Lettland  (129) Samoa  (153) Freistaat Danzig
(105) Namibia  (130) Aserbaidschan  (154) Taiwan
(107) Belize  (131) Seychellen  (155) Turkmenistan
(108) Dominikanische Republik  (132) Weißrußland  (156) Afrika
(109) Nicaragua  (133) Uruguay  (157) Guatemala
(110) Kenia  (134) Bahamas  (158) Sierra Leone
(111) Libyen  (135) Uganda  (Westafrika)
(112) Malta  (136) Oman  (159) Panama
(113) Botswana  (137) Mikronesien  (160) Osttimor
(114) Haiti  (138) Mali  (161) Bahrain
(115) Trinidad, Tobago  (139) Kamerun  (162) Senegal
(116) Luxemburg  (140) Kosovo-Albaner  (163) Malediven
(117) Belgien  (141) Georgien  (164) Hawaii
(118) Holland  (142) Sudan  (165) Serbien
(119) Kroatien  (143) Kongo  (166) Gambia
(120) Bosnien-Herzegowina  (144) Togo  (167) Honduras
(121) Makedonien  (145) Mongolei  (168) Montenegro
(122) Slowenien  (146) Litauen  (169) Kambodscha
(123) Slowakei  (147) Tschat  (170) Surinam
(124) Paraguay  (148) Armenien  (171) Guyana
(125) Guinea  (149) Kustaritan  (172) Kaukasus
(126) Kuwait  (150) Liberia  (173) Simbabwe
(127) Elfenbeinküste  (151) Jemen  (174) Madagaskar
(128) Malaysia  (152) Palästina  (175) Grenada

Variable format  3-digit integer
$-Wave F10, F11, F12, F13

Comment  Respondent’s nationality.

$STISTAT
Variable Label „Stichprobenstatus“
Value Labels
1. Stichprobenmitglied Stammperson alt
2. Nicht-Stichprobenmitglied (zugeordnete Person) alt
3. Stichprobenmitglied neu
4. Nicht-Stichprobenmitglied neu

Variable format  1-digit integer
$-Wave F10, F11, F12, F13

Comment  Sample Status
A person’s sample status “Stichprobenstatus“ is a fixed person characteristic given at the first contact with the study. All individuals living in the household in the first wave are sample persons. Individuals entering the study in any following wave are usually non-sample members. Children born into an old panel household usually receive code 3 and are thus considered panel members. If a family member returns from abroad, she is also considered a panel member and receives cod 3. All other persons receive code 4 (usually individuals moving into a household from within Germany and/or children from non-sample members. In a person’s initial wave, code 3 or 4 is given, which in the following wave is changed to code 1 or 2.

$BEFSTAT
Variable Label „Befragungsstatus“
Value Labels

$BEFSTAT

1. Erneut zu befragen
2. Erstmals zu befragen, da bisher temporärer Ausfall
3. Erstmals zu befragen, da Befragungsalter erreicht
4. Erstmals zu befragende neue Person
5. Noch nicht zu befragende alte Person
6. Noch nicht zu befragende neue Person
7. Nicht zu befragen, da harter Verweigerer in Vorwelle/n
8. Umwandlung von Code 7, da doch wieder teilgenommen

Variable format 1-digit integer

$ - Wave F10, F11, F12, F13

Comment

Respondent’s question status

„Befragungsstatus“ allows the survey agency (and thus the interviewer) to know in advance, what type of interview (if any) is to be expected in this wave. Codes 1, 2, 3, 5, 7 are set before the field period starts based on information from the previous years. These codes are not changed during the field work, even if a household member moves into a new household. Codes 4 and 6 are set manually for all new individuals (also for those in new samples) considering the following rules: Code 4 is set, when the person is eligible in terms of age, i.e. 17 years old or older. Code 6 is set whenever individuals are not age eligible. Code 8 is given only for the case when a household member refused explicitly to respond in a previous wave. These individuals receive a code 7 initially. However, in some cases respondents change their mind in a following wave and decide to participate. In that case, they receive the code 8, removing the prior code 7.

$LINT

Variable Label „Letztes Interview in Welle“

Value Labels $LINT

(0) Befragungsperson, jedoch noch nie ein Personeninterview gegeben

Variable format 4-digit integer

$ - Wave F10, F11, F12, F13

Comment

This code is set after the end of the field work on the basis of the result of the interview. If an interview was conducted, the current year will appear in this wave.

Refusing individuals of the previous year ($BEFSTAT=7) are not followed in case they move (see also SPADER=7). However, in case they move with a person still in the panel, a person dataset is created for them (SPADER=8).
$LUECKE
Variable Label "Lückenbearbeitung"
Value Labels $LUECKE
(6) Lücke ist entstanden, da Person in aktueller Welle wieder teilgenommen hat.
(7) Keine Lücke entstanden, da Person auch in aktueller Welle nicht teilgenommen hat.
(8) Lücke über mehrere Vorwellen
Variable format 1-digit integer
$ - Wave F10, F11, F12, F13
Comment This variable documents whether a gap exists in the person’s panel life. Note that for most persons, this variable takes the code “-2 does not apply”, as they participated in a previous wave.

$ZUPAN
Variable Label "Zugangswelle zum Panel"
Variable format 4-digit integer
$ - Wave F10, F11, F12, F13
Comment $ZUPAN shows in which wave a person entered the panel, i.e. is first listed in the gross sample. This variable is thus constant over time.

$PNRAKT
Variable Label "Aktuelle Personennummer"
Variable format 2-digit integer
$ - Wave F10, F11, F12, F13
Comment This variable depicts the order in which the household members are listed in the address protocol, where the order is set usually when the interviewer enters the household. This number is a useful identifier in cases the PERSNR of a person answering the questionnaire is unknown in case of a proxy interview (i.e. in the household files, Sh, $PNRAKT allows to identify the person answering the questionnaire via the corresponding variable $AUSKU).

$STELL
Variable Label "Stellung zum Haushaltivorstand"
Version f10, f11
Value Labels $STELL
(00) Haushaltivorstand
(01) Ehepartner des HV
(02) Lebenspartner des HV
(03) Kind (auch Adoptivkind) HV
(04) Pflegekind des HV
(05) Schwiegersohn/-tochter HV
(06) Eltern HV
(07) Schwiegereltern HV
(08) Geschwister und Schwager/Schwägerin HV
(09) Enkel des HV
(10) Onkel, Tante, Neffe, Nichte u.ä. HV
(11) Nicht verwandte/verschwägerte Personen HV
(12) Kind vom Lebenspartner des HV
(13) gleichgeschlechtliche Ehepartner des HV nach dem
Lebenspartnergesetz
(99) Stellung zum HV unbekannt

$STELL (continued)

Version f12 and beyond

Value Labels $STELL
(00) Haushaltsvorstand
(11) Ehegatte des HV
(12) Gleichgeschl. Ehepartner
(13) Lebenspartner des HV
(21) Sohn,Tochter des HV
(22) Stiefkind des HV
(23) Adoptivkind des HV
(24) Pflegekind des HV
(25) Enkel des HV
(26) Urenkel des HV
(27) Schwsohn,-tochter HV
(31) Vater, Mutter des HV
(32) Stiefmutter -vater/Partner d. leibl. Elternteils HV
(33) Adoptivmutter, -vater HV
(34) Pflegemutter, -vater HV
(35) Schwvater,-mutter HV
(36) Grossmutter, -vater HV
(41) Schwester, Bruder HV
(42) Halbschwester, -bruder HV
(43) Stiefschwester, -bruder HV
(44) Adoptivschwester, -bruder HV
(45) Pflegeschwester, -bruder HV
(51) Schwaegerin/Schwager 1: Ehegatten oder
Lebenspartner von Geschwistern HV
(52) Schwaegerin/Schwager 2: Geschwister von Ehegatten
oder Lebenspartner HV
(61) Tante/Onkel des HV
(62) Nichte/Neffe des HV
(63) Cousine/Cousin des HV
(64) Sonst.mit HV verw.
(71) Mit HV nicht verw.
(99) Stellung zu HV unbekannt

Variable format 2- digit integer
$ - Wave F10, F11, F12, F13

Comment All households have household head (Haushaltsvorstand), usually the
person most equipped to answer questions about the financial situation
of the household. Whenever possible, the household head remains constant over the years, unless a move or death requires to define a new household head. In that case, relationships via $STELL need to be newly identified. Since 2012, the $STELL variable is asked in more detail to allow for more precise definitions. Given the table below, the new version can be “translated” into the old one.

### $STELL (continued)

<table>
<thead>
<tr>
<th>$STELL in waves F10 and F11</th>
<th>$STELL in wave F12 and F13</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Haushaltsvorstand (HV)</td>
</tr>
<tr>
<td>1</td>
<td>Ehepartner</td>
</tr>
<tr>
<td>2</td>
<td>Lebenspartner</td>
</tr>
<tr>
<td>3</td>
<td>Kind (auch Adoptivkind) HV</td>
</tr>
<tr>
<td>4</td>
<td>Pflegekind des HV</td>
</tr>
<tr>
<td>5</td>
<td>Schwiegersohn/-tochter HV</td>
</tr>
<tr>
<td>6</td>
<td>Eltern HV</td>
</tr>
<tr>
<td>7</td>
<td>Schwiegereltern HV</td>
</tr>
<tr>
<td>8</td>
<td>Geschwister und Schwager/Schwägerin HV</td>
</tr>
<tr>
<td>9</td>
<td>Enkel des HV</td>
</tr>
<tr>
<td>10</td>
<td>Onkel, Tante, Neffe, Nichte u.ä. HV</td>
</tr>
<tr>
<td>11</td>
<td>Nicht verwandte/verschwägerte Personen HV</td>
</tr>
<tr>
<td>12</td>
<td>Stiefkind des HV</td>
</tr>
<tr>
<td>13</td>
<td>Gleichgeschl. Ehepartner</td>
</tr>
<tr>
<td>14</td>
<td>Unbekannt</td>
</tr>
</tbody>
</table>

Since 2012, the $STELL variable is asked in more detail to allow for more precise definitions. Given the table below, the new version can be “translated” into the old one.
$PZUG

Variable Label  „Zugehörigkeit der Personen zum Haushalt“
Value Labels

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>Lebt noch im Haushalt</td>
</tr>
<tr>
<td>01</td>
<td>Vorübergehend abwesend: Bundeswehr</td>
</tr>
<tr>
<td>02</td>
<td>Vorübergehend abwesend: Ausbildung/Studium</td>
</tr>
<tr>
<td>03</td>
<td>Vorübergehend abwesend: Beruf/Montage</td>
</tr>
<tr>
<td>04</td>
<td>Vorübergehend abwesend: Krankenhaus/Kur</td>
</tr>
<tr>
<td>05</td>
<td>Vorübergehend abwesend: Längere Zeit verreist</td>
</tr>
<tr>
<td>06</td>
<td>Vorübergehend abwesend: Sonstiges</td>
</tr>
<tr>
<td>07</td>
<td>Verstorben</td>
</tr>
<tr>
<td>08</td>
<td>Verzogen, jedoch keine Weiterverfolgung</td>
</tr>
<tr>
<td>09</td>
<td>Personen, die in Vorwellen irrtümlich als Haushaltsmitgliedergeführt worden sind</td>
</tr>
<tr>
<td>10</td>
<td>Personen, die in Vorwellen irrtümlich als Haushaltsmitgliedergeführt worden sind</td>
</tr>
<tr>
<td>11</td>
<td>Geboren</td>
</tr>
<tr>
<td>12</td>
<td>Zuzug aus Westdeutschland</td>
</tr>
<tr>
<td>13</td>
<td>Zuzug aus dem Ausland</td>
</tr>
<tr>
<td>14</td>
<td>In den Haushalt zurückgekehrt</td>
</tr>
<tr>
<td>15</td>
<td>Zuzug aus dem Inland vor letzter Befragung und erstmals genannt</td>
</tr>
<tr>
<td>16</td>
<td>Zuzug aus dem Ausland vor letzter Befragung und erstmals genannt</td>
</tr>
<tr>
<td>17</td>
<td>Geboren vor letzter Befragung/schon immer im Haushalt geführt und erstmals genannt</td>
</tr>
<tr>
<td>18</td>
<td>Zuzug aus Ostdeutschland</td>
</tr>
<tr>
<td>19</td>
<td>Keine Information</td>
</tr>
<tr>
<td>20</td>
<td>Lebt im neuen Haushalt</td>
</tr>
<tr>
<td>21</td>
<td>Vorübergehend abwesend: Bundeswehr</td>
</tr>
<tr>
<td>22</td>
<td>Vorübergehend abwesend: Ausbildung/Studium</td>
</tr>
<tr>
<td>23</td>
<td>Vorübergehend abwesend: Beruf/Montage</td>
</tr>
<tr>
<td>24</td>
<td>Vorübergehend abwesend: Krankenhaus/Kur</td>
</tr>
<tr>
<td>25</td>
<td>Vorübergehend abwesend: Längere Zeit verreist</td>
</tr>
<tr>
<td>26</td>
<td>Vorübergehend abwesend: Sonstiges</td>
</tr>
<tr>
<td>27</td>
<td>Verstorben</td>
</tr>
<tr>
<td>28</td>
<td>Verzogen, jedoch keine Weiterverfolgung</td>
</tr>
<tr>
<td>30</td>
<td>Lebt in diesem Haushalt schon länger</td>
</tr>
<tr>
<td>31</td>
<td>Geboren</td>
</tr>
<tr>
<td>32</td>
<td>Zuzug aus Westdeutschland</td>
</tr>
<tr>
<td>33</td>
<td>Zuzug aus dem Ausland</td>
</tr>
<tr>
<td>34</td>
<td>Zuzug aus Ostdeutschland</td>
</tr>
<tr>
<td>35</td>
<td>Keine Information</td>
</tr>
<tr>
<td>40</td>
<td>Aktuelle Haushaltszugehörigkeit konnte in dieser Welle nicht geklärt werden.</td>
</tr>
</tbody>
</table>

Variable format  2-digit integer
$ - Wave  F10, F11, F12, F13
Generally, any person appearing in the address protocol receives a code from the list above, i.e. there are no missing values for this variable.

Some notes on the different codes:
Code 0-19 apply to all households in which interviews have already taken place. Code 7 is set even when the whole household is deceased. Code 8 is set only, when a single person has moved, not when the entire household has moved.
Codes 20-39 apply to all persons in new households, i.e. those who lived in a household as part of a new sample, those previously interviewed founding a new household and those who moved into a newly founded household.
Codes 40-48 apply to all persons moving from an old household into a household founded earlier by another person (e.g. a child moving in with the father in wave t, where the father had moved in t-1).

**$PFORM**
Variable Label  „Bearbeitungsform Person“
Value Labels  $PFORM
(1) Nur Interviewer
(2) Interviewer nach positivem Telefonkontakt
(3) Schriftlich/postalisch nach positivem Telefonkontakt
(4) Nur Telefonkontakt (negativ)
(5) Telefoninterview
(6) Schriftlich/postalisch ohne Telefonkontakt
(8) Keine Bearbeitung
(9) CAPI-Interview
(10) CAWI (Computer Assisted Web Interview)

Variable format  2-digit integer
$ - Wave  F10, F11, F12, F13

Comment  Note that this variable records the form of contact regardless of the result. In general, the result on the household level is identical to the one on the individual level (see $HFORMS). Hence only the last form of contact is coded.

**$PERG**
Variable Label  „Bearbeitungsergebnis 1-Steller“
Value Labels  $PERG
(1) Interview Realisiert
(2) Derzeit nicht durchführbar
(3) Derzeit nicht bereit
(4) Endgültige Verweigerung
(5) Ins Ausland verzogen
(6) Verstorben
(8) Während der Feldzeit nicht auffindbar
(9) Endgültig nicht auffindbar

Variable format  1-digit integer
$ - Wave  
F10, F11, F12, F13

Comment  
$PERG$ shows whether an interview was conducted or not and provides basic information about possible non-response. Codes 5 and 6 mean that the person has left the sampling frame and is no longer eligible for the study. The other codes apply to the eligible (or unknown) population.

$PERGZ$
Variable Label  
„Bearbeitungsergebnis 2-Steller“
Value Labels  
$PERGZ$
(10) Interview Realisiert
(11) Realisiert auch in Vorwelle
(16) Person hat teilgenommen und Lückendaten nacherhoben
(17) Person hat teilgenommen und Lückendaten nicht nacherhoben
(18) Person hat teilgenommen, kein Haushaltsinterview vor
(20) Interview derzeit nicht durchführbar:
(21) Alt und krank
(22) In der Schlussphase der Feldzeit nicht mehr erreicht
(23) Ausländer: Längere Zeit im Heimatland
(24) Krankheit/Krankenhaus über Feldende hinaus
(25) Während der gesamten Feldphase nicht erreichbar
(26) Geistig nicht in der Lage/Framebogen nicht auswertbar
(29) Sonstige unklare Fälle
(30) Zur Teilnahme derzeit nicht bereit:
(31) Zusage für Telefonbearbeitung, kein Fragebogen ausgefüllt
(32) Keine Zeit/Lust, jedoch im nächsten Jahr wieder ansprechbar
(39) Sonstige unklare Fälle
(46) Wie Code 26, jedoch endgültig nicht mehr in der Lage
(47) Sprachprobleme
(50) Ins Ausland verzogen
(60) Verstorben
(80) Nicht auffindbar während der Feldzeit
(90) Endgültig nicht auffindbar

Variable format  
2-digit integer
$ - Wave  
F10, F11, F12, F13

Comment  
$PERGZ$ provides more detail on the interview result and allows coding more elaborate reasons for non-response.

$PADER$
Variable Label  
„Ergebnis der Adressenermittlung Personen“
Value Labels  
$PADER$
(1) umgezogen, neue Adresse ermittelt
(2) unbekannt verzogen
(3) trotz Bestätigung der alten Adresse durch Post/ Einwohnermeldeamt, dort nicht auffindbar
(4) an alter Adresse nicht auffindbar und bei Einwohnermeldeamt nicht registriert
*(5) ins Ausland verzogen
(7) verzogen, wird aber nicht weiterverfolgt (Kind/ harte Verweigerer)
(8) umgezogen mit einer Person, die weiterverfolgt wird
(9) in bestehenden Panelhaushalt zurückgekehrt (laufende Personennummer ist reserviert)*

Variable format
1-digit integer
S - Wave F10, F11, F12, F13

Comment $PADER documents the whereabouts of individuals from previously interviewed households. As most people stay in their old household, most individuals receive a “-2 Does not apply”. Code 1 results in an interview request, all other depict why finding the respondent’s new address was not be determined.

$PADERQ
Variable Label „Informationsquelle der Adressenermittlung Personen“
Value Labels $PADERQ
(1) Interviewer
(2) Post
(3) Einwohnermeldeamt
(4) Befragungsperson

Variable format
1-digit integer
S - Wave F10, F11, F12, F13

Comment $PADERQ documents where the information from $PADER comes from.

$AUSZUGM
Variable Label „Monat des Auszugs einer Person“
Value Labels $AUSZUGM (months)
Variable format 2-digit integer
S - Wave F10, F11, F12, F13

Comment Shows the month in which a person left the household (the year is coded in $AUSZUGJ).

$AUSZUGJ
Variable Label „Jahr des Auszugs einer Person“
Value Labels $AUSZUGJ (years)
Variable format 4-digit integer
S - Wave F10, F11, F12, F13

Comment Shows the year in which a person left the household (the month is coded in $AUSZUGM).
**$EINZUGM**

Variable Label: „Monat des Einzuges einer Person“
Value Labels: $EINZUGM (months)
Variable format: 2-digit integer
$ - Wave: F10, F11, F12, F13

Comment: Shows the month in which a person moved into the household (the year is coded in $EINZUGJ).

**$EINZUGJ**

Variable Label: „Jahr des Einzuges einer Person“
Value Labels: $EINZUGJ (years)
Variable format: 4-digit integer
$ - Wave: F10, F11, F12, F13

Comment: Shows the year in which a person moved into the household (the month is coded in $EINZUGM).

**$ABWESM**

Variable Label: „Monat der Abwesenheit einer Person“
Value Labels: $ABWESM (months)
Variable format: 2-digit integer
$ - Wave: F10, F11, F12, F13

Comment: Shows the month in which a person’s temporary absence begins (see also $PZUG).

**$ABWESJ**

Variable Label: „Jahr der Abwesenheit einer Person“
Value Labels: $ABWESM (years)
Variable format: 4-digit integer
$ - Wave: F10, F11, F12, F13

Comment: Shows the year in which a person’s temporary absence begins (see also $PZUG).

**$PBIO**

Variable Label: „Biographie“
Value Labels:
(0) Kein Zusatzfragebogen nötig, da seit Welle 1984 teilgenommen
(1) Zusatzfragebogen ausgefüllt
(2) Keinen Zusatzfragebogen ausgefüllt und Grund nicht bekannt
(3) Kein Zusatzfragebogen nötig, da in akt. Welle erst ins Befragungsalter gekommen
Befragungsalter bekomen
(4) Zusatzfragebogen ausdrücklich verweigert
(5) aus 1 umgesetzt: Zusatzfragebogen vorhanden
Variable format 1-digit integer
$ - Wave F10, F11, F12, F13

Comment In the SOEP, this variable denotes whether a person filled out the extra biography questionnaire or the youth questionnaire. Given that in FiD, the biography part is integrated into the p-questionnaire, this variable refers to the youth questionnaire only.

$DJ
Variable Label „Kognitiver Test „Lust auf DJ““
Value Labels $DJ
Variable format 1-digit integer
$ - Wave F10, F11, F12, F13

Comment This variable is set “-5 not asked in this sample” for all cases, as no tests were administered in FiD.

$EWSTATU
Variable Label „Erwerbstatus von Nicht-Teilnehmern“
Value Labels $EWSTATU
(1) In Vollzeit erwerbstätig
(2) In Teilzeit erwerbstätig
(3) Arbeitslos gemeldet
(4) In Schule/Studium/Ausbildung
(5) In Rente
(6) Sonstiges
Variable format 1-digit integer
$ - Wave F10, F11, F12, F13

Comment This Information is collected for non-responding individuals in participating households (i.e. $BEFSTAT = 3).

$EX ($E1-$E6)
Variable Label „Elternfragebogen X“ (X=1,2,3,4,5,6)
Value Labels $EX
(1) Zusatzfragebogen ausgefüllt
(2) Keinen Zusatzfragebogen ausgefüllt und Grund nicht bekannt
(4) Zusatzfragebogen ausdrücklich verweigert
(5) aus 1 umgesetzt: Zusatzfragebogen vorhanden
(6) aus 2 umgesetzt: Kein Zusatzfragebogen vorhanden
(8) aus 4 umgesetzt: Zusatzfragebogen ausdrücklich verweigert
Variable format 1-digit integer
$ - Wave F10, F11, F12, F13
The parent questionnaires 1-6 are cohort specific questionnaires, i.e. in every wave, one questionnaire is meant to be answered by mothers (and/or fathers) of children in a specific birth year. The following table shows the distribution over the years.

<table>
<thead>
<tr>
<th>Parent Questionnaire</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Questionnaire 2</td>
<td>2008</td>
<td>2009</td>
<td>2010</td>
<td>2011</td>
</tr>
<tr>
<td>Parent Questionnaire 3</td>
<td>2007</td>
<td>2008</td>
<td>2009</td>
<td>2010</td>
</tr>
<tr>
<td>Parent Questionnaire 4</td>
<td>2004</td>
<td>2005</td>
<td>2006</td>
<td>2007</td>
</tr>
<tr>
<td>Parent Questionnaire 5</td>
<td>2002</td>
<td>2003</td>
<td>2004</td>
<td>2005</td>
</tr>
<tr>
<td>Parent Questionnaire 6</td>
<td>2000</td>
<td>2001</td>
<td>2002</td>
<td>2003</td>
</tr>
</tbody>
</table>
Documentation *hbrutto*

Gross information on all households

*Note that some of this information is in German, based on the codebook provided by TNS-Infratest for FiD and SOEP.*
List of Variables:

$HTYP ................................................................................................................................... 277
$HPMAX ............................................................................................................................... 277
$DATUMTG .......................................................................................................................... 277
$DATUMMO ......................................................................................................................... 277
$DATUMY ............................................................................................................................ 278
$BULA ................................................................................................................................... 278
$SHADER ............................................................................................................................. 278
$SHADQ ................................................................................................................................ 279
$INTZA .................................................................................................................................. 279
INTID ..................................................................................................................................... 279
$INTK .................................................................................................................................... 280
$TELK1 .................................................................................................................................. 280
$TELK2 .................................................................................................................................. 280
$SCHK ................................................................................................................................... 280
$HFORM1 .............................................................................................................................. 281
$HERG1 ................................................................................................................................... 281
$HFORMS ............................................................................................................................. 282
$HERGS ................................................................................................................................ 282
$HSTU ................................................................................................................................... 283
$SPLIT ................................................................................................................................... 284
$HHGR .................................................................................................................................. 285
$SWUM1 ................................................................................................................................ 285
$SWUM2 ................................................................................................................................ 285
$SWUM3 ................................................................................................................................ 286
$WEIN ................................................................................................................................... 286
$HTEL .................................................................................................................................... 286
$INTEINS .............................................................................................................................. 286
$EMAIL ................................................................................................................................. 286
$MKZ1 ................................................................................................................................... 287
$MKZ2 ................................................................................................................................... 287
$HTYP

Variable Label

„Haushaltstyp“

Value Labels

$HTYP
(1) Alter Haushalt mit unveränderter Adresse
(2) Umgezogener alter Haushalt
(3) Alter Haushalt, im Vorjahr umgezogen und temporärer Ausfall
(4) Neuer Haushalt der Vorwelle, im Vorjahr temporärer Ausfall
(5) Neuhaushalt der aktuellen Welle

Variable format
1-digit integer

$ - Wave
F10, F11, F12, F13

Comment
For all splits, the basic rule holds that the part remaining at the old address becomes the “old household” (code 1), whereas anyone moving will move into a “new household” (all other codes). If there are changes to codes 3, 4, or 5 within a wave, these codes remain unchanged. If the respective household ceases to exist, there is only a code 1. Code 2 is set if all members of an old household move to a new address. If a household moves AND splits at the same time, the person who was household head (or the person with the lower identification number) is defined to live in a code 2 household (and all other members receive a code 5).

$HPMAX

Variable Label

„Höchste vergebene Personennummer im Haushalt“

Variable format
2-digit integer

$ - Wave
F10, F11, F12, F13

Comment
This variable shows, how many individual datasets have been set up in this household, i.e. how many different individuals have ever been associated with the respective household number. In case a new individual moves into the household, $HPMAX increases by one, if a household number moves out, there is no change to $HPMAX. Note that $HPMAX is very different from $HHRGR, the current household size.

$DATUMTG

Variable Label

„Tag des letzten Haushaltskontakts“

Value Labels

$DATUMTG (day of month)

Variable format
2-digit integer

$ - Wave
F10, F11, F12, F13

Comment
Refers to the day of the last contact of the household, regardless of their final state. In case the household was never in the field, the case receives code 0.

$DATUMMO

Variable Label

„Monat des letzten Haushaltskontakts“

Value Labels $DATUMMO (month)
Variable format 2-digit integer
$ - Wave F10, F11, F12, F13
Comment Refers to the month in which the household was last contacted, regardless of their final state. In case the household was never in the field, the case receives code 0.

$DATUMY
Variable Label „Interviewjahr“
Value Labels $DATUMY (year)
Variable format 4-digit integer
$ - Wave F10, F11, F12, F13
Comment Refers to the year in which the household was to be contacted, regardless of their final state.

$BULA
Variable Label „Bundesland“
Value Labels $BULA
(01) Schleswig-Holstein
(02) Hamburg
(03) Niedersachsen
(04) Bremen
(05) Nordrhein-Westfalen
(06) Hessen
(07) Rheinland-Pfalz
(08) Baden-Württemberg
(09) Bayern
(10) Saarland
(11) Berlin
(12) Brandenburg
(13) Mecklenburg-Vorpommern
(14) Sachsen
(15) Sachsen-Anhalt
(16) Thüringen
Variable format 2-digit integer
$ - Wave F10, F11, F12, F13
Comment $BULA refers to the federal state the household is located in during the field work. In case the household moved, the regional code is changed if necessary. In case the new address remains unknown ($HADER=2 – 4), the regional code stays as it was. Similar for the deceased ($HADER=5 – 7).

$HADER
Variable Label „Adressermittlung Haushalt“
Value Labels $HADER
(1) umgezogen, neue Adresse ermittelt  
(2) unbekannt verzogen  
(3) an alter Adresse nicht auffindbar; Post /Einwohnermeldeamt bestätigen

*alte Adresse*

(4) an alter Adresse nicht auffindbar; Einwohnermeldeamt hat die Person nicht registriert  
(5) ins Ausland verzogen  
(6) verstorben  
(7) Haushaltsauflösung, Mitglieder sind in anderen Panelhaushalt verzogen  
(8) Auskunftssperre gemäß Meldegesetz

**Variable format** 1-digit integer  
$ - $Wave F10, F11, F12, F13

**Comment**  
This variable is set for all households that are not found at their previous address. All others receive a code “-2 Does not apply”. Reasons are generally a moves of address within Germany, death or moves abroad.

In case of one or more persons moving out of a remaining household, there is no code in $HADER$, but in $SPADER$.

**$HADQ**

**Variable Label** „Informationsquelle Adressermittlung“  
**Value Labels** $HADQ$  
(1) Interviewer  
(2) Post  
(3) Einwohnermeldeamt  
(4) Zielperson  

**Variable format** 1-digit integer  
$ - $Wave F10, F11, F12, F13

**Comment**  
$HADQ$ provides the source of the information given in $HADER$. It is based on the last result of the address search.

**$INTZA**

**Variable Label** „Zahl der eingesetzten Interviewer“  

**Variable format** 1-digit integer  
$ - $Wave F10, F11, F12, F13

**Comment**  
Shows the numbers of interviewers that worked this case.

**INTID**

**Variable Label** „Interviewer-Nummer“  

**Variable format** 6-digit integer  
$ - $Wave F10, F11, F12, F13
INTID shows the unique identifier for the interviewer, which is consistently defined over all waves. In case of more than one interviewer working the household, the one conducting the household interview is the one listed in INTID. In case other interviewers did some other interviews in the same household (personal, parent), their interviewer identifier will be listed in the respective dataset.

### $\text{INTK}$
- **Variable Label**: „Interviewer-Kontakte“
- **Value Labels**: $\text{INTNK}$ (number of contacts)
- **Variable format**: 1-digit integer
- **$S$ - Wave**: F10, F11, F12, F13
- **Comment**: Provides the number of contacts until the successful interview or the refusal.

### $\text{TELK1}$
- **Variable Label**: „Telefonkontakte 1“
- **Value Labels**: $\text{TELK1}$
  - (0) Telefonischer Bearbeitungsversuch, jedoch keine Telefonnummer zu ermitteln
  - (1) Telefonisch niemanden erreicht bzw. Situation nicht geklärt.
  - (2) Telefonisch geklärt, ob und (wenn ja) auf welche Weise weitere Bearbeitung möglich ist.
  - (3) Bei schriftlich-postalischer Weiterbearbeitung: Erneute telefonische Bearbeitung, da Fragebogen nach vereinbarter Frist nicht zurückgesandt wurde.
- **Variable format**: 1-digit integer
- **$S$ - Wave**: F10, F11, F12, F13
- **Comment**: Note that in FiD, no phone contact with the household is allowed to conduct the interview. Hence all cases are set to “-2 Does not apply”.

### $\text{TELK2}$
- **Variable Label**: „Telefonkontakte 2“
- **Value Labels**: $\text{TELK2}$
  - (0) Keine Telefonnummer zu ermitteln
  - (1) Nacherhebung ohne Erfolg
  - (2) Nacherhebung mit Erfolg
- **Variable format**: 1-digit integer
- **$S$ - Wave**: F10, F11, F12, F13
- **Comment**: Note that in FiD, no phone contact with the household is allowed to conduct the interview. Hence all cases are set to “-2 Does not apply”.

### $\text{SCHK}$
- **Variable Label**: „Schriftliche Kontakte“
**$SCHK**

Value Labels

1. Angeschrieben ohne Reaktion
2. Angeschrieben und Antwort erhalten

Variable format 1-digit integer

$ - Wave F10, F11, F12, F13

Comment Rarely, the household is contacted via mail, if neither the interviewer nor a phone call is successful in making contact.

**$HFORM1**

Variable Label „Bearbeitungsform Haushalt“

Value Labels

1. Nur Interviewer
2. Interviewer nach positivem Telefonkontakt
3. Schriftlich/postalisch nach positivem Telefonkontakt
4. Telefonkontakt mit negativem Ergebnis bzw. nicht erreicht
5. Telefoninterview
6. Schriftlich-postalisch ohne Telefonkontakt
7. Mischform
8. Keine Bearbeitung möglich
9. CAPI
10. CAWI (Computer Assisted Web Interview)

Variable format 2-digit integer

$ - Wave F10, F11, F12, F13

Comment This variable codes the mode in which the household is first contacted to be interviewed. In case no interview is possible, but the interviewer was working the case, code 1 is set.

**$HERG1**

Variable Label „Bearbeitungsergebnis Haushalt“

Value Labels

0. Teilweise realisiert
1. Vollständig realisiert
2. Interview derzeit nicht durchführbar
3. Haushalt derzeit zur Teilnahme nicht bereit
4. Endgültige Verweigerung
5. Ins Ausland verzogen
6. Verstorben
7. Auflösung des Haushalts wegen Verzug in anderen Panelhaushalt
8. Während der Feldzeit nicht angetroffen / aufgefunden
9. Haushalt endgültig nicht auffindbar

Variable format 1-digit integer

$ - Wave F10, F11, F12, F13

Comment Documents the grade of completion in the household. Code 0 is set, if any eligible person has refused an interview request. Code 1 is only set
if all eligible persons ($BEFSTAT=1,2,3,4, or 8) were interviewed or no longer in the household ($BEFSTAT=5,6,7).

$\text{HFORMS}$

Variable Label: „Schlusscode Bearbeitungsform“

Value Labels:

1. Nur Interviewer
2. Interviewer nach positivem Telefonkontakt
3. Schriftlich / postalisch nach positivem Telefonkontakt
4. Telefonkontakt mit negativem Ergebnis
5. Telefoninterview
6. Schriftlich-postalisch ohne Telefonkontakt
7. Mischform wegen Nachbearbeitung in anderer Form
8. Keine Bearbeitung möglich
9. CAPI
10. CAWI (Computer Assisted Web Interview)

Variable format: 2-digit integer

$\text{- Wave}$: F10, F11, F12, F13

Comment: This variable codes the mode in which the household is actually interviewed (as opposed to $\text{HFORM1}$). In case there is only one approach to the household (as in most cases) $\text{HFORMS}$ and $\text{HFORM1}$ are identical.

$\text{HERGS}$

Variable Label: „Schlusscode Bearbeitungsergebnis“

Realisierte Haushalte

(0) Teilweise realisiert (d.h. nicht alle Befragungspersonen haben teilgenommen)
1. Ohne weitere Hinweise auf Folgewelle
2. Mit Hinweis: In nächster Welle nicht mehr zur Teilnahme bereit
3. Mit Hinweis: In nächster Welle unter keinen Umständen mehr bereit

(1) Vollständig realisiert

11. Ohne weitere Hinweise auf Folgewelle
12. Mit Hinweis: In nächster Welle nicht mehr zur Teilnahme bereit
13. Mit Hinweis: In nächster Welle unter keinen Umständen mehr bereit
14. Vollständig realisiert ohne Personen, die BEFSTAT 7 haben
18. Haushaltsinterview, aber kein Personeninterview vorhanden
19. Realisierte Haushalte, die nicht gültig sind (z.B. wegen Wertgrenzen, Fälschungen)

Nicht realisierte Haushalte

(2) Interview derzeit nicht durchführbar

21. Alt und krank
22. In der Schlussphase der Feldarbeit nicht mehr erreichbar
23. Ausländer: Längere Zeit im Heimatland

SOEP Survey Paper 332
(24) Krankheit / Krankenhaus über Feldende hinaus
(25) Während der gesamten Feldphase nicht erreichbar
(26) Geistig nicht in der Lage / Fragebogen nicht auswertbar
(27) Zweimal temporärer Ausfall, soll trotzdem ins Brutto der Folgewelle aufgenommen werden
(29) Sonstige / unklare Fälle

(3.) Haushalt derzeit zur Teilnahme nicht bereit
(31) Zusage zum Ausfüllen des Fragebogens in der telefonischen Bearbeitung, jedoch nicht ausgefüllt
(32) Diverse Begründungen: z.B. keine Zeit, Trauerfall; wird in Folgewelle wieder angesprochen
(39) Sonstige / unklare Fälle

(4.) bis (7.) Endgültige Ausfälle
(40) Endgültige Verweigerung
(41) Zweimal hintereinander temporärer Ausfall
(46) Endgültig nicht mehr in der Lage teilzunehmen (z.B. geistig nicht in der Lage/Pflegefall)
(47) Sprachprobleme
(48) Ganzer Point ohne Bearbeitung
(49) Einzelhaushalt ohne Bearbeitung
(50) Ins Ausland verzogen
(60) Verstorben
(70) Haushalt aufgelöst (in anderen Panel-Haushalt verzogen)

(8.) Während der Feldzeit nicht angetroffen (Adressenprobleme)
(81) Verfügt laut Auskunft Dritter über zweiten Wohnsitz, Adresse konnte jedoch nicht ermittelt werden
(82) Ständig unterwegs, auf Reisen, beruflich im Ausland
(83) Lebt laut Auskunft von Dritten überwiegend bei Freunden oder anderen Familienmitgliedern
(84) Haushalt in Adressermittlung über Feldende hinaus
(85) Neue Adresse ermittelt, jedoch nach Feldende
(89) Sonstige / unklare Fälle
(90) Haushalt endgültig nicht auffindbar
(98) QNA (Qualitätsneutraler Ausfall)
(99) Haushalt ohne Bearbeitung wegen unterschrittener Einkommensgrenze (nur Stichprobe G / Welle 2003)

Variable format 2-digit integer
$ - Wave F10, F11, F12, F13

Comment $HERGS shows the final result for each household. The first digit in this case is identical to the one for $HERG1, the second digit provides some more information about the code. Code 99 is never set for FiD cases.

$$HSTU$$

Variable Label „Bearbeitungsstufen Haushalt“
Value Labels $HSTU$
   (0) Keine Bearbeitung
   (1) Eine Bearbeitungsstufe
   (2) Zwei Bearbeitungsstufen
(3) Drei Bearbeitungsstufen

Variable format 1-digit integer
$ - Wave F10, F11, F12, F13

Comment This variable documents different stages of the field work. It is related to the variable $HFORMS, which refers to the last mode of interviewing used for this household before the field period closed. Note that in FiD, most cases receive code 1, as there are no mode switches allowed.

$SPLIT

Variable Label „Startbearbeitungsform aktuelle Welle“
Value Labels $SPLIT

(10) Feldbearbeitung: Erfolgreiche Bearbeitung durch Interviewer in Vorwelle sowie Einzelfälle, die nicht vom Interviewer in der Vorwelle realisiert wurden (Termingründe)
(61) Erstbearbeitung über Telefonkontakt: Temporärer Ausfall, in der Vorwelle in der telefonischen oder schriftl. Bearbeitung
(66) Erstbearbeitung über Telefonkontakt: Temporärer Ausfall, in der Vorwelle in der Interviewerbearbeitung
(68) Diverse Bearbeitungsformen: Nicht mehr im Ausgangsbrutto der aktuellen Welle enthaltene Haushalte, die aus unterschiedlichsten Gründen wieder auftauchen
(69) Erstbearbeitung über Interviewer: Ausfälle aus dem Vorjahr
(70) Erstbearbeitung über Telefonkontakt: Teilnehmer in Vorwelle, jedoch mit Hinweis auf künftige Verweigerung
(75) Keine Bearbeitung in aktueller Welle: Vor Beginn der aktuellen Welle ins Ausland verzogen
(76) Keine Bearbeitung in aktueller Welle: Vor Beginn der aktuellen Welle verstorbem
(77) Keine Bearbeitung in aktueller Welle: Vor Beginn der aktuellen Welle in anderen Panelhaushalt zurückgezogen
(78) Keine Bearbeitung in aktueller Welle: In letzter Welle Teilnehmer, jedoch für Folgewellen ausdrücklich verweigert
(81) Erstbearbeitung über Telefonkontakt: Letzte Welle telefonisch/schriftlich befragt
(88) Erstbearbeitung über Telefonkontakt: Letzte Welle Telefoninterview (nur in Ausnahmefällen)
(89) Erstbearbeitung schriftlich: Letzte Welle schriftlich befragt
(90) Diverse Bearbeitungsformen: Neuhaushalte aktuelle Welle
(99) Feldbearbeitung im Erstversand: Neuhaushalte aktuelle Welle

Variable format 2-digit integer
$ - Wave F10, F11, F12, F13

Comment This variable shows the choice of the most promising form to contact the household in this wave, based on previous wave experience.
$HHGR
Variable Label  „Zahl der im Haushalt lebenden Personen“
Value Labels  $HHGR (number)
Variable format  2-digit integer
$ - Wave  F10, F11, F12, F13
Comment  $HHGR shows the number of people in the household at the time of the interview, not counting moved-out or deceased persons. Note that a household that has been dissolved is coded with a “0”.

$WUM1
Variable Label  „Wohnumfeld1 - Haushaltstyp“
Value Labels  $WUM1
(1) Landwirtschaftliches Wohngebäude
(2) Freistehendes Ein-/Zweifamilienhaus
(3) Ein-/Zweifamilienhaus als Reihenhaus oder Doppelhaus
(4) Wohnhaus mit 3 - 4 Wohnungen
(5) Wohnhaus mit 5 - 8 Wohnungen
(6) Wohnhaus mit 9 und mehr Wohnungen (aber höchstens 8 Stockwerke; also kein Hochhaus)
(7) Hochhaus, 9 und mehr Stockwerke, Wohnungen unbegrenzt
(8) Sonstiges
Variable format  1-digit integer
$ - Wave  F10, F11, F12, F13
Comment  Note that the information for this variable is retrieved from the household questionnaire (as it is for the next two, $WUM2 and $WUM3). Only new or moved household are to answer these questions. If households do not participate, the interviewer is asked to provide the information on these variables. For old households which have stayed in the old address, there are no changes.

$WUM2
Variable Label  „Wohnumfeld2 – Privat/Anstaltshaushalt“
Value Labels  $WUM2
(1) Privathaushalt/Kein Wohnheim
(2) Schüler-/Jugendlichenwohnheim
(3) Studentenwohnheim
(4) Berufstätigen-/Ledigenwohnheim
(5) Altenheim
(6) Altenwohnheim
(7) Sonstiges Heim/Unterkunft
(8) Hotel / Pension
Variable format  1-digit integer
$ - Wave  F10, F11, F12, F13
Comment  Similar to $WUM1, these data are drawn either from the household questionnaire or (in case of refusals) provided by the interviewer. This information is used to identify non-private households.
$WUM3
Variable Label „Wohnumfeld3 – Quartier“
Value Labels $WUM3
(1) Reines Wohngebiet mit überwiegend Altbauten
(2) Reines Wohngebiet mit überwiegend Neubauten
(3) Mischgebiet mit Wohnungen und Geschäften/Gewerbe
(4) Geschäftszentrum (Läden, Banken, Verwaltungen) mit wenig Wohnungen
(5) Gewerbe- und Industriegebiet mit wenig Wohnungen
(6) Sonstiges
Variable format 1-digit integer
$ - Wave F10, F11, F12, F13
Comment Similar to $WUM1, these data are drawn either from the household questionnaire or (in case of refusals) provided by the interviewer.

$WEIN
Variable Label „Welle des Einzugs in die Aktuelle Adresse“
Value Labels $WEIN (year)
Variable format 4-digit integer
$ - Wave F10, F11, F12, F13
Comment $WEIN provides the information since when the household resides at its current address.

$HTEL
Variable Label „Telefon“
Value Labels $HTEL
(1) Telefonnummer bekannt
(-2) Telefonnummer unbekannt
Variable format 1-digit integer
$ - Wave F10, F11, F12, F13
Comment Shows whether the household has a known phone number.

$INTEINS
Variable Label „Nummer des ersteingesetzten Interviewers“
Variable format 1-digit integer
$ - Wave F10, F11, F12, F13
Comment In combination with $INTID, this variables shows whether there has been a change in interviewers during the field period.

$EMAIL
Variable Label „Haushalt mit e-mail Anschluss“
Value Labels $EMAIL
**$hbrutto**

<table>
<thead>
<tr>
<th>Variable format</th>
<th>1-digit integer</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ - Wave</td>
<td>F10, F11, F12, F13</td>
</tr>
</tbody>
</table>

Comment: Shows whether the household has a known phone number.

**$MKZ1**

Variable Label: „Migrantenkennzeichen1“

Value Labels: $MKZ1

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adresse ohne eindeutigen Migrationshintergrund</td>
</tr>
<tr>
<td>2</td>
<td>Adresse mit Migrationshintergrund laut EMA</td>
</tr>
<tr>
<td>3</td>
<td>Adresse mit Migrationshintergrund laut Onomastik</td>
</tr>
</tbody>
</table>

Variable format: 1-digit integer

$ - Wave: F10

Comment: $MKZ1 defines in the cohort sample, drawn 2010, whether the household was classified as a migration household before the field period or not. This was done via nationality (code 2) and onomastic method (code 3). For the screening sample, the code is “-2 Does not apply”. This variable is only set in 2010 so far.

**$MKZ2**

Variable Label: „Migrantenkennzeichen2“

Value Labels: $MKZ2

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Deutsch</td>
</tr>
<tr>
<td>2</td>
<td>Migrant/ Ausländer Basis</td>
</tr>
<tr>
<td>3</td>
<td>Migrant/ Ausländer Top Up</td>
</tr>
</tbody>
</table>

Variable format: 1-digit integer

$ - Wave: F10

Comment: $MKZ2 defines in the cohort sample, drawn 2010, whether the household belonged to the Top-up sample or to the “original” migration sample. The top-up was drawn from the originally sampled households in FiD to boost the number of migrant households (see also documentation on FiD). For the screening sample, the code is “-2 Does not apply”. This variable is only set in 2010 so far.
Additional information regarding interview circumstances

by Mathis Fräßdorf (geb. Schröder) and Malisa Zobel

Note that this file was called “Sintview” up to FiD v1.2. As this name exists in the SOEP data collection but covers different information (namely information on the interviewer), this dataset was renamed to avoid confusion. The Sintview dataset with the according information on the interviewers will likely be part of a future data distribution.

Up to FiDv3.0, there were wave specific file Sparadata. Starting with distribution 3.1, only a longitudinal dataset is included.
Contents

General Information ...................................................................................................... 289
Variables in *paradatal* .................................................................................................. 291
   HHNR .......................................................................................................................... 291
   HHNRAKT .................................................................................................................... 291
   PERSNR ....................................................................................................................... 291
   PERSNRK ..................................................................................................................... 291
   SAMPLE1 ...................................................................................................................... 291
   QSTNR .......................................................................................................................... 291
   REQD ............................................................................................................................ 292
   MISS ............................................................................................................................. 293
   MODE ............................................................................................................................ 293
   PROXY .......................................................................................................................... 293
   INTID ............................................................................................................................ 293
   DURA ............................................................................................................................. 294
   MINT .............................................................................................................................. 294
   DINT .............................................................................................................................. 294
   DOW .............................................................................................................................. 294
Table of frequencies ...................................................................................................... 296
General Information

The data file provides information regarding:

- Questionnaire
- Interview Mode
- Duration of Interview
- Day Interview was conducted
- Month Interview was conducted
- Interviewer

The aim of the *paradatal* data file is to provide additional information regarding the circumstances under which the interview took place. Information is available for households, as well as individuals, based on their never-changing personal ID. With the distribution of FiDv3.1, only a longitudinal version of these datasets is available. The wave specific data can still easily be extracted by restricting to certain years using SYEAR.
## Variables in paradatal

### HHNR
- **Variable Label**: “Original household number”
- **Variable Format**: 7-digit integer
- **Comment**: HHNR is the original household number. It can be linked to the never-changing person ID (PERSNR), thereby showing the household in which the individual first entered the panel.

### HHNRAKT
- **Variable Label**: “Current wave household number”
- **Variable Format**: 7-digit integer

### PERSNR
- **Variable Label**: “Never-changing person ID”
- **Variable Format**: 8-digit integer
- **Comment**: PERSNR in this dataset refers to the person answering the questionnaire. Hence, in the household questionnaire, this is the household head; in the parent questionnaires, PERSNR refers to the person giving information about the child.

### PERSNRK
- **Variable Label**: “Child’s never-changing person ID”
- **Variable Format**: 8-digit integer
- **Comment**: PERSNRK is the identifier complementing PERSNR in case of a parent questionnaire. It refers to the child for whom the parent questionnaire has been answered. It is set to “(-2) Does not apply” in case of the other questionnaires.

### SAMPLE1
- **Variable label**: “Subsample”
- **Value label**:
  - (61) FiD 2007 Birth Cohort
  - (62) FiD 2008 Birth Cohort
  - (63) FiD 2009 Birth Cohort
  - (64) FiD 2010 Birth Cohort
  - (65) FiD Screening (sampled 2010)
  - (66) FiD Screening (sampled 2011)
- **Variable format**: 2-digit integer
- **Comment**: Note that this variable is included in all datasets, and provides information whether the household or person originates from the cohort.
or the screening sample in FiD. In \textit{ppfad} and \textit{Skind} it is named PSAMPLE, in \textit{hpfad} it is named HSAMPLE.

**QSTNR**

<table>
<thead>
<tr>
<th>Variable Label</th>
<th>“Questionnaire”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value Label</td>
<td>QSTNR</td>
</tr>
<tr>
<td>(10)</td>
<td>Parent-Qunaire 1 (0-1 yrs)</td>
</tr>
<tr>
<td>(20)</td>
<td>Parent-Qunaire 2 (1-2 yrs)</td>
</tr>
<tr>
<td>(30)</td>
<td>Parent-Qunaire 3 (2-3 yrs)</td>
</tr>
<tr>
<td>(40)</td>
<td>Parent-Qunaire 4 (5-6 yrs)</td>
</tr>
<tr>
<td>(50)</td>
<td>Parent-Qunaire 5 (7-8 yrs)</td>
</tr>
<tr>
<td>(60)</td>
<td>Parent-Qunaire 6 (9-10 yrs)</td>
</tr>
<tr>
<td>(70)</td>
<td>Person-Qunaire (only)</td>
</tr>
<tr>
<td>(71)</td>
<td>Person-Qunaire (with Bio 1)</td>
</tr>
<tr>
<td>(72)</td>
<td>Person-Qunaire (with Bio 2)</td>
</tr>
<tr>
<td>(73)</td>
<td>Person-Qunaire (with Bio 1+2)</td>
</tr>
<tr>
<td>(80)</td>
<td>Household-Qunaire</td>
</tr>
<tr>
<td>(90)</td>
<td>Youth-Qunaire</td>
</tr>
<tr>
<td>(100)</td>
<td>Luecke-Qunaire (Gap in prev. year)</td>
</tr>
</tbody>
</table>

Variable Format: 3-digit integer

Comment: QSTNR indicates which questionnaire was used during the interview. Starting with the data distribution 3.1, the codes include specific information about which person questionnaire is filled out, i.e. whether any of the biography questionnaires are included or not. (For this reason, the codes were also extended to 3-digits.) Also, a code for the gap questionnaires is included since version 3.0. Note that even though the \textit{Sluecke} files collect data about the previous year (and are stored in the previous year’s folder), \textit{paradatal} is motivated by the surroundings of the interview (such as date or interviewer). For this reason, interview information on \textit{Sluecke} is kept in the year the data are collected.

**REQD**

<table>
<thead>
<tr>
<th>Variable Label</th>
<th>“Questions required (not code -2,-4,-5,-6)”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value Label</td>
<td>REQD</td>
</tr>
<tr>
<td>Variable Format</td>
<td>3-digit integer</td>
</tr>
</tbody>
</table>

Comment: REQD counts the number of questions a respondent would have had to answer. This information varies by respondent, as some individuals may go through different parts of the questionnaire than others. E.g. employed respondents have a whole set of job related questions which individuals not in the labour market do not have to answer. This count is based on questions which are not “missing by design”, i.e. “-2 does not apply”, “-4 invalid multiple answers”, “-5 Question not asked in sample”, and “-6 Sample specific filters” are excluded.
**MISS**

Variable Label: **MISS**  
Value Label: **Missing answers (code -1 or -3)**  
Variable Format: 2-digit integer  

Comment: MISS provides additional information on the quality of the interview by providing the number of missing answers, i.e. counting the total of “-1 No answer” and “-3 Answer improbable” in a questionnaire.

**MODE**

Variable Label: **MODE**  
Value Label: **Interview mode**  
(1) CAPI  
(2) PAPI  
(3) Mail (special)  
Variable Format: 1-digit integer  

Comment: MODE provides information on the mode of the interview. Note that there have been mail-outs after the field phase for parent-questionnaire 2 in the screening sample of 2010. All other interviews from 2010 to 2013 have been conducted face-to-face.

**PROXY**

Variable Label: **PROXY**  
Value Label: **Proxy Interview**  
(1) Condition met  
(2) Condition not met  
Variable Format: 1-digit integer  

Comment: PROXY denotes that the indicated person was not able to answer the questionnaire and responses are given by a proxy. These are very few cases in FiD for the personal interviews. Note that while all parent-questionnaires are proxy interviews, we do not specifically set this variable for these questionnaires.

**INTID**

Variable Label: **INTID**  
Value Label: **Interviewer Identification Number**  
Variable Format: 6-digit integer  

Comment: INTID provides the interviewer identification number, as given by the survey institute (TNS-Infratest). Note that there are no interviewer IDs for parent-questionnaire 2 in the screening sample of 2010, as these have been completed by mail, not with an interviewer present. These cases are coded with “(-2) Does not apply”.
**DURA**

Variable Label: “Duration of interview”

Variable Format: 3-digit integer

Comment: DURA provides information on the duration of the interview in minutes. Note that this is information provided by the interviewer, not information from the CAPI program. For all questionnaires available in PAPI (pen-and-paper) mode, the interview duration is not collected and hence set to “-2 Does not apply”.

**MINT**

Variable Label: “Month of interview”

Value Label: MINT

(1) January
(2) February
(3) March
(4) April
(5) May
(6) June
(7) July
(8) August
(9) September
(10) October
(11) November
(12) December

Variable Format: 2-digit integer

Comment: MINT provides information on the month of the interview. Note that this is information provided by the interviewer, not information from the CAPI program. In case of a PAPI questionnaire (only possible in the parent questionnaires, QSTNR=1-6), the date recorded may be different from the date the respondent answered the questions.

**DINT**

Variable Label: “Day of interview”

Variable Format: 2-digit integer

Comment: DINT provides information on the day of the month of the interview. Note that this is information provided by the interviewer, not information from the CAPI program. In case of a PAPI questionnaire (only possible in the parent questionnaires, QSTNR=1-6), the date recorded may be different from the date the respondent answered the questions.

**DOW**

Variable Label: “Day of week of interview”

Value Label: DOW
(1) Monday  
(2) Tuesday  
(3) Wednesday  
(4) Thursday  
(5) Friday  
(6) Saturday  
(7) Sunday  

Variable Format  1-digit integer  
Comment  DOW provides information on the day of the week the interview took place. Note that this is information provided by the interviewer, not information from the CAPI program. In case of a PAPI questionnaire (only possible in the parent questionnaires, QSTNR=1-6), the date recorded may be different from the date the respondent answered the questions.
### Table of frequencies

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10) Parent-Qunaire 1 (0-1 yrs)</td>
<td>1,321</td>
<td>207</td>
<td>212</td>
<td>167</td>
</tr>
<tr>
<td>(20) Parent-Qunaire 2 (1-2 yrs)</td>
<td>787</td>
<td>647</td>
<td>568</td>
<td>187</td>
</tr>
<tr>
<td>(30) Parent-Qunaire 3 (2-3 yrs)</td>
<td>871</td>
<td>741</td>
<td>555</td>
<td>523</td>
</tr>
<tr>
<td>(40) Parent-Qunaire 4 (5-6 yrs)</td>
<td>473</td>
<td>486</td>
<td>425</td>
<td>656</td>
</tr>
<tr>
<td>(50) Parent-Qunaire 5 (7-8 yrs)</td>
<td>682</td>
<td>902</td>
<td>849</td>
<td>707</td>
</tr>
<tr>
<td>(60) Parent-Qunaire 6 (9-10 yrs)</td>
<td>647</td>
<td>820</td>
<td>768</td>
<td>760</td>
</tr>
<tr>
<td>(70) Person-Qunaire (only)</td>
<td></td>
<td>5,361</td>
<td></td>
<td>6,747</td>
</tr>
<tr>
<td>(71) Person-Qunaire (with Bio 1)</td>
<td>7,807</td>
<td>1,414</td>
<td>143</td>
<td>0</td>
</tr>
<tr>
<td>(72) Person-Qunaire (with Bio 2)</td>
<td></td>
<td>6,250</td>
<td>1,519</td>
<td>0</td>
</tr>
<tr>
<td>(73) Person-Qunaire (with Bio 1+2)</td>
<td></td>
<td></td>
<td>154</td>
<td>96</td>
</tr>
<tr>
<td>(80) Household-Qunaire</td>
<td>4,574</td>
<td>4,529</td>
<td>4,186</td>
<td>3,923</td>
</tr>
<tr>
<td>(90) Youth-Qunaire</td>
<td>190</td>
<td>264</td>
<td>293</td>
<td>310</td>
</tr>
<tr>
<td>(100) Luecke-Qunaire (Gap in prev. year)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>17,352</td>
<td>16,260</td>
<td>15,262</td>
<td>14,347</td>
</tr>
</tbody>
</table>
Documentation \textit{bioage17}

Detailed Information on Youths

This documentation is based on the comparable SOEP documentation on \textit{bioage17} and has benefited from previous work of Marco Giesselman, Mila Staneva, Henning Lohmann and Sven Witzke. For readability reasons we do not specifically cite and specify text that has been used directly from the SOEP document.
**General Information**

A special group of first time respondents are young persons living in a panel household, who reach the surveying age of 17 years. From this specific group of panel entrants, we are able to obtain some more detailed information on youth and socialisation than from other new sample members. At the same time, certain life-course dimensions (as the partnership- or employment biography) have not yet developed in 17 year-olds. With regard to these specifics, the standard biography questionnaire is not appropriate to this group. Thus, we use an independent questionnaire for this special group of first time respondents: the Youth Questionnaire. This instrument was used in the SOEP since the year 2000 and was introduced in FiD from the first wave on in 2010. It can be understood as an alternative version of the Biography Questionnaire, collecting more comprehensive information on relationships with parents, leisure-time activities, and past achievements in school, as well as on personality characteristics. In addition, there are numerous prospective questions about educational plans and plans for further training, as well as questions about expectations for future career and family.

A number of statements regarding specific circumstances—including the expectations for the future mentioned above—are directly related to the time at which the questionnaire was completed. However, they provide a multifaceted background for long-term analyses since these young people will continue to be interviewed in subsequent years like other SOEP respondents. The Youth Questionnaire also contains retrospective questions, for example, at what age the teenager started his or her first job or first music lessons, what recommendations he or she received regarding choice of secondary school level, and which grades he or she repeated.

**Genesis and Target Population of the Youth Questionnaire**

The Youth Questionnaire is aimed at youths who have reached the surveying age of 17 years and are therefore being interviewed for the first time. This questionnaire takes the place of the supplementary Biography Questionnaire, since the latter does not apply to the young people’s family or career situations. As a rule, information on social origin can be obtained from the parents’ Individual Questionnaire, in case the youth lives together with the respective parent. If the teenager does not live with either parent, the Youth Questionnaire collects information on the missing parent(s). Young people who immigrated to Germany are also given the standard questions on immigration from the supplementary Biography Questionnaire. This guarantees that all important information collected in the Biography Questionnaire is also available on these young people.

---

33 More precisely, this refers to youths who live in an already existing panel household and are or will turn 17 years old in the year of the survey. They are therefore 16 or 17 years old at the time of the interview.
The Youth Questionnaire is used in all FiD samples, however, as Table 1 shows, the focus on young children in the Cohort Sample leads to slightly fewer youths being present in those samples. For the years currently available, 747 data on youths are provided.

Table 1: Target Population for the Youth Questionnaire by year, sample and age

<table>
<thead>
<tr>
<th>Survey year</th>
<th>Cohort</th>
<th>Screening 2010</th>
<th>Screening 2011</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>14</td>
<td>176</td>
<td>-</td>
<td>190</td>
</tr>
<tr>
<td>2011</td>
<td>20</td>
<td>171</td>
<td>73</td>
<td>264</td>
</tr>
<tr>
<td>2012</td>
<td>16</td>
<td>193</td>
<td>84</td>
<td>293</td>
</tr>
<tr>
<td>2013</td>
<td>27</td>
<td>205</td>
<td>78</td>
<td>310</td>
</tr>
<tr>
<td>Total</td>
<td>77</td>
<td>745</td>
<td>235</td>
<td>1,057</td>
</tr>
</tbody>
</table>

**Contents and Structure of the Data Set bioage17**

From a technical perspective, four different types of questions are asked in the Youth Questionnaire:

**A)** Questions used to complete certain biographical files (bioparen, bioimmig in the SOEP). These questions are identical to questions in the standard Biography Interview. This applies to the topic blocks ‘Origin’ (questions 60 to 71) and ‘Childhood and Parents’ House’ (questions 72-85). The corresponding variables are not included in bioage17, but combined with biographical information from non-youth new entrants in the file bioparen. The information on a person’s migration background are used for the variable MIGBACK in ppfad, and are also available from the original dataset Jugend.

**B)** Questions that are similar to items in the standard Biography Interview, but go further into detail. This applies to the topic blocks ‘Relationships’ (questions 12-14), ‘Leisure and Sport’ (questions 15-25) and ‘Education and Career plans’ (questions 26-55). These variables are stored in bioage17. Corresponding Variables obtained from other new sample members (with a standard Biography Interview) are included in the original dataset Slela (in the SOEP, these data appear equivalently in biosoc. Depending on the complexity and scope of the analysis, the user might want to combine corresponding data from bioage17 and Slela in order to access all panel members.

**C)** Questions that specifically relate to young persons and therefore have no equivalent in the standard Biography Interview. This applies to the topic blocks ‘Residence’ (questions 1-3), ‘Jobs and Money’ (questions 4-11), ‘Future’ (question 59) and ‘Attitudes and Opinions’
(questions 86-87). These Variables are stored in bioage17 and have no equivalent for other panel entrants.

D) Selected time-variant questions from the regular individual questionnaire are added to the Youth Questionnaire. This refers to the questions 56 to 58, 89, 90 and the topic block ‘personality’ (questions 91 to 99). This data is not included in bioage17, but can be accessed in the original dataset $jugend$.

The design of the dataset bioage17 is patterned after the SOEP Youth Questionnaire, which is the standard version for subsequent years. As in the biographical data survey, every youth answers the Youth Questionnaire only once. The data is therefore presented in column form, just as it would be in a cross-sectional record. The variable ERHEBJ makes it possible to quickly identify the year of the survey.

Table 2 (at the end of this chapter) lists all variables for the dataset bioage17. The first column contains the name of each variable, the second a brief specification of its content, and the third the number of the question as it appears in the Youth Questionnaire. The variables containing the identification of the person surveyed and the interview situation have no corresponding number because they do not originate from the regular section of the Youth Questionnaire.

**Special Features of Some Questions and Variables**

The question regarding the support received by these young people from their parents (question 14) is based on the Supportive Parenting Scale of Simons et al. (1992)\(^{35}\), which was transformed for Germany by Schwarz and Walper (1997)\(^{36}\). The instrument used to compile career orientation (question 54) was taken from Kracke (1996)\(^{37}\).

If the question on school attendance in the Youth Questionnaire is answered with ‘yes’ when at the same time information about vocational degrees is provided, a recoding is undertaken. In this case the variable BYSCHBES is changed to the value -3 (-3: Entry deleted after intensive examination).

In question 51, young people are asked whether they know what career they would like to start. If they give a positive answer (‘yes, with some certainty’, ‘yes, with a lot of certainty’), then they are asked to specify the occupation in plain text. This plain-text entry is coded according to the classification of occupations of the Federal Statistical Office, Germany, (Statistisches Bundesamt), version 1992, and according to the ISCO 1988. In addition, the

\(^{34}\) The first ten items in question 90 are still stored in BIOAGE17, for details see below.


values for Ganzeboom’s International Socio-Economic Index of Occupational Status (ISEI), for Treiman’s Standard International Occupational Prestige Scale (SIOPS) for Erikson’s and Goldthorpe’s Class Category (EGP) as well as Wegener’s Magnitude Prestige Scale (MPS) are also given.


Table 2: Description of the data set BIOAGE17

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Content of the Variable</th>
<th>Number of Question in Youth Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHNR</td>
<td>Original household identifier (invariant)</td>
<td></td>
</tr>
<tr>
<td>HHNRAKT</td>
<td>Actual household identifier</td>
<td></td>
</tr>
<tr>
<td>PERSNR</td>
<td>Personal identifier</td>
<td></td>
</tr>
<tr>
<td>BEFRPER</td>
<td>Respondent identifier</td>
<td></td>
</tr>
<tr>
<td>ERHEBJ</td>
<td>Survey year</td>
<td></td>
</tr>
<tr>
<td>BYGEBJAH</td>
<td>Year of birth</td>
<td></td>
</tr>
<tr>
<td>BYMNR</td>
<td>identifier of mother (taken from BIOPAREN; social, not necessarily biological relationship)</td>
<td></td>
</tr>
<tr>
<td>BYVNR</td>
<td>identifier of father (taken from BIOPAREN; social, not necessarily biological relationship)</td>
<td></td>
</tr>
<tr>
<td><strong>Residence</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BYWOELT</td>
<td>Residing in parents’ household (HH)</td>
<td>01</td>
</tr>
<tr>
<td>BYWOZIM</td>
<td>Own room</td>
<td>02</td>
</tr>
<tr>
<td>BYWOWEI</td>
<td>Additional apartment outside of parents’ HH</td>
<td>03</td>
</tr>
<tr>
<td><strong>Jobs and Money</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BYVDEIG</td>
<td>Own income</td>
<td>04</td>
</tr>
<tr>
<td>BYVDART</td>
<td>Type of income</td>
<td>05</td>
</tr>
<tr>
<td>BYJBFRUE</td>
<td>Worked before (on holiday or while in school)</td>
<td>06</td>
</tr>
<tr>
<td>BYJBALT</td>
<td>Age by first job (on holiday or while in school)</td>
<td>07</td>
</tr>
<tr>
<td>BYJBGRUN</td>
<td>Reason for working</td>
<td>08</td>
</tr>
<tr>
<td>BYTGELD</td>
<td>Allowance</td>
<td>09</td>
</tr>
<tr>
<td>BYTGELDW</td>
<td>Amount of allowance per week</td>
<td>10</td>
</tr>
<tr>
<td>BYTGELDM</td>
<td>Amount of allowance per month</td>
<td>10</td>
</tr>
<tr>
<td>BYSPAR</td>
<td>Saving money</td>
<td>11</td>
</tr>
<tr>
<td>BYSPARM</td>
<td>Amount saved every month</td>
<td>11</td>
</tr>
<tr>
<td>BYSPARUN</td>
<td>Sporadic saving</td>
<td>11</td>
</tr>
<tr>
<td><strong>Relationships</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BYWIVA</td>
<td>Father</td>
<td>12</td>
</tr>
<tr>
<td>BYWIMU</td>
<td>Mother</td>
<td>12</td>
</tr>
<tr>
<td>BYWIBS</td>
<td>Brother, Sister</td>
<td>12</td>
</tr>
<tr>
<td>BYWIWV</td>
<td>Other related persons</td>
<td>12</td>
</tr>
<tr>
<td>BYWIFFR</td>
<td>Serious boy/girlfriend</td>
<td>12</td>
</tr>
<tr>
<td>BYWIBFR</td>
<td>Best friend</td>
<td>12</td>
</tr>
<tr>
<td>BYWILEHR</td>
<td>Teacher</td>
<td>12</td>
</tr>
<tr>
<td>BYWICLQ</td>
<td>Clique</td>
<td>12</td>
</tr>
<tr>
<td>BYWISON</td>
<td>Other person</td>
<td>12</td>
</tr>
<tr>
<td><strong>Frequency of fights with:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BYSTRVA</td>
<td>Father</td>
<td>13</td>
</tr>
<tr>
<td>BYSTRMU</td>
<td>Mother</td>
<td>13</td>
</tr>
<tr>
<td>BYSTRBS</td>
<td>Brother, Sister</td>
<td>13</td>
</tr>
<tr>
<td>BYSTRFFR</td>
<td>Serious boy/girlfriend</td>
<td>13</td>
</tr>
<tr>
<td>BYSTRBFR</td>
<td>Best friend</td>
<td>13</td>
</tr>
<tr>
<td>BYBZ01MU</td>
<td>Talk with mother about personal experiences</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ01VA</td>
<td>Talk with father about personal experiences</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ02MU</td>
<td>Mother addresses problems</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ02VA</td>
<td>Father addresses problems</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ03MU</td>
<td>Mother asks opinion before a decision is made</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ03VA</td>
<td>Father asks opinion before a decision is made</td>
<td>14</td>
</tr>
<tr>
<td>Variable Name</td>
<td>Content of the Variable</td>
<td>Number of Question in Youth Questionnaire</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>BYBZ04MU</td>
<td>Mother shows approval</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ04VA</td>
<td>Father shows approval</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ05MU</td>
<td>Solve problems together with mother</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ05VA</td>
<td>Solve problems together with father</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ06MU</td>
<td>Mother shows trust</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ06VA</td>
<td>Father shows trust</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ07MU</td>
<td>Mother asks opinion on family issues</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ07VA</td>
<td>Father asks opinion on family issues</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ08MU</td>
<td>Mother justifies decision</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ08VA</td>
<td>Father justifies decision</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ09MU</td>
<td>Mother shows love</td>
<td>14</td>
</tr>
<tr>
<td>BYBZ09VA</td>
<td>Father shows love</td>
<td>14</td>
</tr>
</tbody>
</table>

**Leisure and Sports**

- Frequency of free time activities:
  - BYFZFERN  TV, Video  15
  - BYFZPC    Computer games  15
  - BYFZMUSH  Listen to music  15
  - BYFZMUSS  Play music  15
  - BYFZSPRT  Do sports  15
  - BYFZTANZ  Dance, Theatre  15
  - BYFZTECH  Technical work, Programming  15
  - BYFZLESE  Read  15
  - BYFZEHRE  Volunteer activities  15
  - BYFZABH  Do nothing, hang around, day dream  15
  - BYFZMFFR  Spend time with boy/girlfriend  15
  - BYFZMBFR  Spend time with best friend  15
  - BYFZMCLQ  Spend time with clique  15
  - BYFZINT  Internet/chatting  15
  - BYFZJUGZ  visiting youth center  15
  - BYFZRELI  go to church/religious activities  15
  - BYMUSSP  Actively make music  16
  - BYMUSART  Style of music made  17
  - BYMUSMW  Play music with whom  18
  - BYMUSALT  Age starting playing music  19
  - BYMUSUNT  Paid music lessons  20
  - BYSPRTR  Participate in sports  21
  - BYSPRTAR  Favourite sport  22
  - BYSPRTLAL  Age started favourite sport  23
  - BYSPRTMW  Where and with whom favourite sport  24
  - BYSPRTWE  Participation in competitions  25
**Variable Name** | **Content of the Variable** | **Number of Question in Youth Questionnaire**
---|---|---
BYSCHBES | School attendance | 26
BYSCHEND | Last year of school | 27
BYSCHABS | Type of school certificate | 28
BYSCHZUK | Strive for further school certificate | 29
BYSCHZAR | Type of further school certificate | 30
BYFMD1 | 1. foreign language | 31
BYFMD2 | 2. foreign language | 31
BYSCHAUSS | School attendance in foreign country | 32
BYSCHPRI | Attendance in a private school | 33

**Activities in school:**

- BYENKSPR | Class representative | 34
- BYENSSPR | School representative | 34
- BYENSZTG | School newspaper | 34
- BYENTHEA | Theatre, Dance group | 34
- BYENCHOR | Choir, Music | 34
- BYENSPRT | Sport group | 34
- BYENSONS | Other groups | 34
- BYENNEIN | No activities | 34

- BYZFINS | Satisfaction with effort at school (overall) | 35
- BYZFDEUT | Satisfaction with effort in German | 35
- BYZFMA | Satisfaction with effort in math | 35
- BYZFFMD1 | Satisfaction with effort in 1. foreign language | 35
- BYEMPFEH | Recommendation after elementary school | 36
- BYNTDEUT | Last grade in German | 37
- BYNTMATH | Last grade in math | 37
- BYNTFMD1 | Last grade in 1. foreign language | 37
- BYPTDEUT | Total points in German | 37
- BYPTMATH | Total points in math | 37
- BYPTFMD1 | Total points in 1. foreign language | 37
- BYGSDEUT | Level of German at comprehensive school | 37
- BYGMATH | Level of math at comprehensive school | 37
- BYLKDEUT | Complementary / main subject in German | 37
- BYLMATH | Complementary / main subject in math | 37
- BYLFMD1 | Complementary / main subject in 1. foreign language | 37
- BYKLWDJA | Class repeated | 38
- BYKLWD1 | Class level 1. repeated | 39
- BYKLWD2 | Class level 2. repeated | 39
- BYNACHHI | Paid tutor lessons | 40

---

40 Students normally receive grades ranging from 1 to 6, whereby 1 is the best and 6 the worst. This system of assigning grades is used up to the 11th or 12th grade (level II of upper secondary or comprehensive school) depending on the federal state. After that, a new grading system is used. To make the data set more user-friendly, the information given for school grades and the information on points transformed into grades is stored in this variable. Note: No corrections have been made when a person has reported both grades and point scores and when the two types of information do not correctly correspond.

41 From the 11th or 12th grade on, pupils are awarded points in upper secondary or comprehensive school ranging from 0 to 15, whereby 15 points are the best, 0 points the worst. The link between points and grades is as follows: 0 points: 6; grade of 1 to 3 points: grade of 5; 4 to 6 points: grade of 4; 7 to 9 points: grade of 3; 10 to 12 points: grade of 2; 13 to 15 points: grade of 1.

42 The subjects German, math and the first foreign language are split up into different levels during the secondary school level I in comprehensive schools. Level A is the highest. The number of levels differs between the federal states.

43 From the 11th or 12th grade on, pupils can choose their main subjects. At this stage, German, math and foreign languages can be downgraded from major to minor subjects.
<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Content of the Variable</th>
<th>Number of Question in Youth Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>BYELKUEM</td>
<td>Parents care about efforts at school</td>
<td>41</td>
</tr>
<tr>
<td>BYELHAUS</td>
<td>Parents help with homework</td>
<td>42</td>
</tr>
<tr>
<td>BYELDIFF</td>
<td>Problems with parents because of effort at school</td>
<td>43</td>
</tr>
<tr>
<td>BYELBEN</td>
<td>Parents attend parents’ evening</td>
<td>44</td>
</tr>
<tr>
<td>BYELSPRE</td>
<td>Parents go to parents’ day</td>
<td>44</td>
</tr>
<tr>
<td>BYELLEHR</td>
<td>Parents go to see a teacher</td>
<td>44</td>
</tr>
<tr>
<td>BYELVERT</td>
<td>Active as parent representative</td>
<td>44</td>
</tr>
<tr>
<td>BYELNIDA</td>
<td>Parents do not participate in any of these activities</td>
<td>44</td>
</tr>
<tr>
<td>BYKLAUSL</td>
<td>Number of foreign classmates</td>
<td>45</td>
</tr>
<tr>
<td>BYBAABGE</td>
<td>Vocational education, Internship, training</td>
<td>46</td>
</tr>
<tr>
<td>BYBABGJ</td>
<td>Vocational introductory year (“Berufgrundschul- / Berufsvorbereitungsjahr”)</td>
<td>47</td>
</tr>
<tr>
<td>BYBABEGL</td>
<td>Vocational integration training (“Beruf. Eingliederungslehrgaenge”)</td>
<td>47</td>
</tr>
<tr>
<td>BYBALEH</td>
<td>Vocational education, apprenticeship (“Berufsausbildung, Lehre”)</td>
<td>47</td>
</tr>
<tr>
<td>BYBABFS</td>
<td>Full-time vocational school/ School for public health (“Berufsfachschule / Schule des Gesundheitswesen”)</td>
<td>47</td>
</tr>
<tr>
<td>BYBAPRAK</td>
<td>Internship (“Praktikum, Voluntariat”)</td>
<td>47</td>
</tr>
<tr>
<td>BYZAJA</td>
<td>Vocational / university degree is aspired</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Type of aspired vocational / university degree:</td>
<td></td>
</tr>
<tr>
<td>BYZALEH</td>
<td>Apprenticeship (“Lehre”)</td>
<td>49</td>
</tr>
<tr>
<td>BYZABFS</td>
<td>Full-time vocational school/ School for public health (“Berufsfachschule / Schule des Gesundheitswesen”)</td>
<td>49</td>
</tr>
<tr>
<td>BYZAFSC</td>
<td>Technical school, school for master of a trade (“Fachschule, Meister-, Technikerschule”)</td>
<td>49</td>
</tr>
<tr>
<td>BYZABEA</td>
<td>Training for civil servants (officer) (“Beamtenausbildung”)</td>
<td>49</td>
</tr>
<tr>
<td>BYZABAK</td>
<td>Approved vocational academy (“anerkannte Berufskademie”)</td>
<td>49</td>
</tr>
<tr>
<td>BYZAFH</td>
<td>Advanced technical college (“Fachhochschule”)</td>
<td>49</td>
</tr>
<tr>
<td>BYZAUNI</td>
<td>University</td>
<td>49</td>
</tr>
<tr>
<td>BYSLBALT</td>
<td>Desired age for financial independence</td>
<td>50</td>
</tr>
<tr>
<td>BYBWUNJA</td>
<td>Occupation is aspired</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Occupation categories, encoded:</td>
<td></td>
</tr>
<tr>
<td>BYKLAS</td>
<td>Classification of career according to the Federal Statistical Office, Germany, (Statistisches Bundesamt), version 1992</td>
<td>52</td>
</tr>
<tr>
<td>BYISCO88</td>
<td>International Standard Classification of Occupation 1988 (ISCO88)</td>
<td>52</td>
</tr>
<tr>
<td>BYEGBP</td>
<td>Erikson and Goldthorpe’s Class Category (EGP)</td>
<td>52</td>
</tr>
<tr>
<td>BYISEI</td>
<td>International Socio-Economic Index of Occupational Status after Ganzeboom (ISEI)</td>
<td>52</td>
</tr>
<tr>
<td>BYSIOPS</td>
<td>Treiman’s Standard International Occupational Prestige Scale (SIOPS)</td>
<td>52</td>
</tr>
<tr>
<td>BYMPS</td>
<td>Magnitude Prestige Scale after Wegener (MPS)</td>
<td>52</td>
</tr>
<tr>
<td>BYZBINF</td>
<td>Information level of planned career</td>
<td>53</td>
</tr>
<tr>
<td>BYZBELT</td>
<td>Influence of the parents on career choice</td>
<td>54</td>
</tr>
<tr>
<td>BYZBLAS</td>
<td>No specific career in mind</td>
<td>54</td>
</tr>
<tr>
<td>BYZBBES</td>
<td>Intensive thoughts about various careers</td>
<td>54</td>
</tr>
<tr>
<td>BYZBRAU</td>
<td>Still looking for a career</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Important aspects for the career choice:</td>
<td></td>
</tr>
<tr>
<td>BYWBSICH</td>
<td>Secure job</td>
<td>55</td>
</tr>
<tr>
<td>BYWBEINK</td>
<td>High income</td>
<td>55</td>
</tr>
</tbody>
</table>
### Variable Name | Content of the Variable | Number of Question in Youth Questionnaire
---|---|---
BYWBAUF | Promotion opportunities | 55
BYWBANE | Established profession | 55
BYWBFR1 | Enough free time | 55
BYWBINT | Interesting activities | 55
BYWBSEL | Working independently | 55
BYWBKONT | Contact with persons | 55
BYWBGSL | Relevant to society | 55
BYWBGSND | Healthy conditions at work | 55
BYWBFA | Flexibility for family | 55
BYWBHEL | Help others | 55

### Future
Probability of future career related and private events:

BYWAUSP | To be accepted for a desired apprenticeship / place at university | 59
BYWAERFA | To complete training/ university successfully | 59
BYWAARB | Job in desired career | 59
BYWAERF | Job-related success | 59
BYWAUBL | Longer unemployment | 59
BYWAREUR | From family related reasons held back in career | 59
BYWASELB | Self-employed | 59
BYWAUSL | Work in foreign country | 59
BYWAHEIR | To marry | 59
BYWAPART | Live together with partner (not married) | 59
BYWAKE | Have one child | 59
BYWAKE | Have two or more children | 59

### Attitudes and Opinions

BYGLPART | Happiness: live with/without partner | 86\(^{44}\)
BYGLKIND | Happiness: with/without children | 87\(^{45}\)
BYEFFLEI | Studiousness | 88\(^{46}\)
BYEFAUSN | Exploitation of others | 88
BYEFIN | Intelligence | 88
BYEFFAM | Family’s origin | 88
BYEFFACH | Technical know-how | 88
BYEFHEL | Money | 88
BYEFSABS | School education | 88
BYEFHART | Being inconsiderate and hard | 88
BYEFSEO | Networking | 88
BYEFPOI | Political activities | 88
BYEFMANN | Sex/ ‘being a man’ | 88
BYEFILI | Being dynamic and taking initiative | 88
BYESEVER | What happens in life, depends on me | 90\(^{47}\)
BYESESRRE | Did not reach, what I deserve | 90
BYESEGLUE | What you achieve, is a matter of luck | 90
BYESEND | Others decide about my life | 90
BYESEHART | You have to work hard for success | 90
BYESEZWEI | By difficulties, doubt about own abilities | 90
BYESESO | Chances are determined by social circumstances | 90

\(^{44}\) Question 88 in 2012/2013  
\(^{45}\) Question 89 in 2012/2013  
\(^{46}\) Question 90 in 2012/2013  
\(^{47}\) Question 92 in 2012/2013
<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Content of the Variable</th>
<th>Number of Question in Youth Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>BYESFAEH</td>
<td>Abilities are more important than efforts</td>
<td>90</td>
</tr>
<tr>
<td>BYESKNTR</td>
<td>Little control over events in my life</td>
<td>90</td>
</tr>
<tr>
<td>BYESENGA</td>
<td>Change of social circumstances through social/political activities</td>
<td>90</td>
</tr>
</tbody>
</table>

**Specification of Interview Situation**

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Content of the Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>BYINTA</td>
<td>Type of interview</td>
</tr>
<tr>
<td>BYDAUER1</td>
<td>Duration of personal interview</td>
</tr>
<tr>
<td>BYDAUER2</td>
<td>Duration of interview filled out independently</td>
</tr>
<tr>
<td>BYANW</td>
<td>Presence of other persons</td>
</tr>
<tr>
<td>BYTAGIN</td>
<td>Day of the interview</td>
</tr>
<tr>
<td>BYMONIN</td>
<td>Month of the interview</td>
</tr>
<tr>
<td>INTID</td>
<td>Identifier of the interviewer</td>
</tr>
</tbody>
</table>
Documentation $bioparen$

Biography Information for the Parents of FID-Respondents

by Linda Wittbrodt

(This documentation is based on the SOEP versions of the bioparen documentation and has benefited from work by Anne Fromm, Sebastian Frischholz, Daniel D. Schnitzlein, Charlotte Büchner, Stefanie Lenuweit, Katharina Mahne, Matthias Pollmann-Schult, Jürgen Schupp and Verena Tobsch. Please understand that for readability reasons we do not mark text that has been directly taken from earlier work.)
**Short summary**

The aim of the data file bioparen is to make the biography entries on the parents and on the social origin of the respondent available.

**How biography information has been collected in the FID**

Since the first wave of FID in 2010 the respondents received a separate Biography Questionnaire (‘Lebenslauf-Fragebogen’) in addition to the Individual Questionnaire. In 2010 it included only questions about the history of the surveyed persons themselves. In 2011 (the second wave of FID) another part of the Biography Questionnaire was handed out, where intergenerational aspects of the persons surveyed were included by means of a special group of questions. This deals with statements made about the education or professional training of the parents, the parents’ residency, and their year of birth and death. In 2012 and 2013 the complete collection of biography questions was included in only one Biography Questionnaire for individuals surveyed for the first time. However, those who had only answered either of the two separate parts handed out in wave one and two were given the other, still unanswered part.

Therefore there are persons who filled in the two Biography Questionnaire parts separately as well as persons who were given the complete questionnaire. In addition there are respondents who only answered one of the two parts because they dropped out of the study at some point. This is the reason why, in contrast to the SOEP, there are persons who answer two Biography Questionnaires (but different parts of it) in two different years.

In addition to the Biography Questionnaire, there is an independent questionnaire (Youth Questionnaire) in FID for the group of survey participants who are 17 years old and are being interviewed for the first time; this questionnaire is mostly identical to the Biography Questionnaire.

**How is bioparen generated?**

The information available in bioparen is obtained in two different ways. On the one hand, bioparen includes the children’s proxy entries on the parents from the Biography Questionnaire and the Youth Questionnaire. On the other hand, it contains the direct entries from the parents in the case the respondent lives in the same household as his parents.

Every respondent is asked for information on the regional mobility of the children, adolescents also on the religious affiliation of the parents. However, information on the year
of birth, as well as the education and occupational training of the parent, additional to the professional position and occupation of father or mother are, due to the filter command in the questionnaire, not collected when the parent lives in the same household as the child at the time of the survey. In this case, the direct entries of the parents are used.

The identification of the parents occurs first of all through the variable $STELL$ (relationship to head of household). The possible values of the variable $STELL$ are listed in Table 2. The combinations of these characteristics of the $STELL$-variable and their assigned interpretation for the generation of parent identifiers are describe in Table 3.

The second source of information is the population of the file $kind$, which includes all children under the age of 16. The file contains the personal number of the mother, as well as the personal number of the father. Through both variables the latest mother, as well as the father are identified, ideally, at the time when the child is 16 years old and thus one year before the first survey of the child. In the case the parents could not be identified by the $STELL$ variable, this information is used.

In a further step the biological mother or father are identified through the parent-child relationship in the file $biobirth$. In the event that still no personal number for the mother or the father exists, the number from $biobirth$ is used.

Table 1: Number of observations in BIOPAREN

<table>
<thead>
<tr>
<th>Year of data collection</th>
<th>Sample</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>61</td>
<td>62</td>
</tr>
<tr>
<td>2010</td>
<td>165</td>
<td>188</td>
</tr>
<tr>
<td>2011</td>
<td>757</td>
<td>778</td>
</tr>
<tr>
<td>2012</td>
<td>56</td>
<td>55</td>
</tr>
<tr>
<td>2013</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>991</td>
<td>1,031</td>
</tr>
</tbody>
</table>
Table 2: Characteristics of the variable SSTE LL “relationship of the person to the head of the household”

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>HH</td>
</tr>
<tr>
<td>1</td>
<td>Marital partner of the HH</td>
</tr>
<tr>
<td>2</td>
<td>partner of the HH</td>
</tr>
<tr>
<td>3</td>
<td>Daughter/son (also adopted/stepchild) of the HH</td>
</tr>
<tr>
<td>4</td>
<td>Foster child of the HH</td>
</tr>
<tr>
<td>5</td>
<td>Daughter/son–in-law of the HH</td>
</tr>
<tr>
<td>6</td>
<td>Father/mother of the HH</td>
</tr>
<tr>
<td>7</td>
<td>Father/mother-in-law of the HH</td>
</tr>
<tr>
<td>8</td>
<td>Brother/sister, brother/sister–in-law of the HH</td>
</tr>
<tr>
<td>9</td>
<td>Grandchild of the HH</td>
</tr>
<tr>
<td>10</td>
<td>Other relationship to the HH</td>
</tr>
<tr>
<td>11</td>
<td>Not related to the HH</td>
</tr>
<tr>
<td>12</td>
<td>Daughter/son of the partner of the HH</td>
</tr>
<tr>
<td>13</td>
<td>Marital partner of the HH (same sex)</td>
</tr>
</tbody>
</table>

Starting in F12, SSTE LL was changed and includes more values than in previous waves, which allows more precise identifications of relationships within the households. For details on the conversion of SSTE LL from 2012 to the Version of 2010/2011 see the documentation on Skind.

Table 3: Possible Parent-Child Relationships based on SSTE LL

<table>
<thead>
<tr>
<th>Relationship of the child to the HH</th>
<th>Relationship of the parent to the HH</th>
<th>Person is ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>Child of HH</td>
</tr>
<tr>
<td>3</td>
<td>1 or 2</td>
<td>Child of marital/ partner of HH</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>Foster child of HH</td>
</tr>
<tr>
<td>4</td>
<td>1 or 2</td>
<td>Foster child of marital/ partner of HH</td>
</tr>
<tr>
<td>12</td>
<td>2 or 3</td>
<td>Child of partner of HH</td>
</tr>
<tr>
<td>9</td>
<td>3 or 4</td>
<td>Child of child/foster child of HH</td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>Child is HH, lives with parents in same household</td>
</tr>
<tr>
<td>1 or 2</td>
<td>7</td>
<td>Marital partner/partner of HH (child of in laws of HH)</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>Grandchild of HH (child of son/daughter-in-law of HH)</td>
</tr>
</tbody>
</table>
### List of Variables:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>VNR / MNR</td>
<td>313</td>
</tr>
<tr>
<td>VGEBJ / MGEBJ</td>
<td>313</td>
</tr>
<tr>
<td>VTODJ / MTODJ</td>
<td>313</td>
</tr>
<tr>
<td>VAORT11 / MAORT11</td>
<td>314</td>
</tr>
<tr>
<td>VAORTAKT / MAORTAKT</td>
<td>314</td>
</tr>
<tr>
<td>VAORTUP / MAORTUP</td>
<td>315</td>
</tr>
<tr>
<td>VSBIL / MSBIL</td>
<td>315</td>
</tr>
<tr>
<td>VBBIL / MBBIL</td>
<td>316</td>
</tr>
<tr>
<td>VSINFO / MSINFO</td>
<td>316</td>
</tr>
<tr>
<td>VBINFO / MBINFO</td>
<td>316</td>
</tr>
<tr>
<td>VRELI / MRELI</td>
<td>317</td>
</tr>
<tr>
<td>VNAT / MNAT</td>
<td>317</td>
</tr>
<tr>
<td>VBSTELL / MBSTELL</td>
<td>317</td>
</tr>
<tr>
<td>VBSINFO / MBSINFO</td>
<td>318</td>
</tr>
<tr>
<td>VISCO88 / MISCO88</td>
<td>318</td>
</tr>
<tr>
<td>VISEI / MISEI</td>
<td>318</td>
</tr>
<tr>
<td>VMPS / MMPS</td>
<td>318</td>
</tr>
<tr>
<td>VSIOPS / MSIOPS</td>
<td>318</td>
</tr>
<tr>
<td>VEGP / MEGP</td>
<td>319</td>
</tr>
<tr>
<td>VBLKAS / MBKLAS</td>
<td>319</td>
</tr>
<tr>
<td>ORTKINDH / ORTKIND1</td>
<td>319</td>
</tr>
<tr>
<td>LIVING1 - LIVING8</td>
<td>319</td>
</tr>
<tr>
<td>VSTREIT / MSTREIT</td>
<td>320</td>
</tr>
<tr>
<td>BIOYEAR</td>
<td>320</td>
</tr>
<tr>
<td>BIO</td>
<td>320</td>
</tr>
<tr>
<td>ALTER / VALTER / MALTER</td>
<td>321</td>
</tr>
<tr>
<td>VORIGIN / MORIGIN</td>
<td>321</td>
</tr>
<tr>
<td>GESCHW</td>
<td>322</td>
</tr>
<tr>
<td>GESCHWUP</td>
<td>322</td>
</tr>
<tr>
<td>NUMS</td>
<td>322</td>
</tr>
<tr>
<td>NUMB</td>
<td>322</td>
</tr>
<tr>
<td>TWIN</td>
<td>323</td>
</tr>
</tbody>
</table>
### VNR / MNR

**Variable Label**
Personal number of the father of the respondent / Personal number of the mother of the respondent

**Variable Format**
8-digit integer

**Values**
- (-1) PERSNR father / mother unknown
- (-2) Does not apply
- (-3) Answer improbable

**Description**
The personal ID of the parents (VNR and MNR) is generated in three steps.
1. The parents of the respondent are identified by the relationship to the head of the household ($STELL$ in Sbrutto). Ideally, the children’s parents are identified at the time of the first survey of the child, i.e., when the child is 17 years old. Furthermore, the social parents and not necessarily the biological parents are identified.
2. The parents of the respondent are identified via the mother’s and father’s ID in $kind$. By using these variables the “oldest” parents are identified. Ideally, these are the parents at the time the child is 16 years old (one year before the first survey).
3. The mother-ID as well as the father-ID of the respondent can be identified in $biobirth$.

As $bioparen$ aims at identifying the parents that live in the household when the child is 17 years old, the steps above are carried out in the hierarchy 1-3 with step 1 having the highest priority. If one is interested in only biological parents, please have a look at the information in $biobirth$.

### VGEBJ / MGEBJ

**Variable Label**
Year of birth of the father / Year of birth of the mother

**Variable Format**
4-digit integer

**Description**
In a first step the information of the year of birth comes from the Biography Questionnaire. Due to a filter command, the children’s proxy entries are only available for these variables when the parents or one parent and the child do not live in the same household at the time of the survey.
After the parents’ personal numbers have been identified the information can be compared with the entries in PPFAD. If there are differences of +/- two years the VNR / MNR will be set to missing. The same applies when parents were aged less than 10 at the time of birth.
For the missing entries the information of the parents’ year of birth is taken from PPFAD.

### VTODJ / MTODJ

**Variable Label**
Year of death of the father / Year of death of the mother

**Variable Format**
4-digit integer

**Description**
The variables are generated as usual using the information from the Youth Questionnaire or the Biography Questionnaire and the parents’ direct-entries from PPFAD. As a next step the annual proxy information on a parent’s death
from the $P$-files are used. Furthermore we use information of the month of death of a parent from the year before. With this data a wrong marking as “no death in 2010” / “death in 2011” can be corrected if there is data from 2011 indicating that one parent died e.g. in October 2010. The variables VTODJ and MTODJ will be updated with new survey information. They are updated as long as the father or the mother is part of the FID sample. We additionally use the annual proxy information of respondents about reported life events of the last year.

**VAORT11 / MAORT11**

Variable Label | Residency of the father/ Residency of the mother in 2011  
Variable Format | 2-digit integer  
Values  
(0) Has Died  
(1) Lives In Same HH  
(2) Lives In Same Housing  
(3) Lives Neighborhood  
(4) Lives Same Town  
(5) Lives Other Town  
(6) Lives Elsewhere In Germany  
(7) Lives Elsewhere  
(8) Lives Else E Germany  
(9) Lives Else W Germany  
(10) Lives Foreign Country

Description The information on the residency of the parents stems from the Youth and Biography Questionnaires as well as from the Person Questionnaire. The information from the $P$-files have a higher priority. For more details see also the information on VAORTAKT / MAORTAKT.

**VAORTAKT / MAORTAKT**

Variable Label | Father's place of residence /Mother's place of residence  
Variable Format | 2-digit integer  
Values  
(0) Has Died  
(1) Lives In Same HH  
(2) Lives In Same Housing  
(3) Lives Neighborhood  
(4) Lives Same Town  
(5) Lives Other Town  
(6) Lives Elsewhere In Germany  
(7) Lives Elsewhere  
(8) Lives Else E Germany  
(9) Lives Else W Germany  
(10) Lives Foreign Country

Description The variables VAORTAKT and MAORTAKT contain the latest available information about the parents’ residence and on whether or not they are deceased, respectively. For persons without identified parents who answered the biography questionnaire in 2011, the information from the Person Questionnaire in 2011 was assumed. For those persons whose parents are identified in the FID, the information on the year of death in PPFAD was used for updating. If the year of death lies
chronologically after the latest available information, VAORTAKT and
MAORTAKT were put on “deceased”.

**VAORTUP / MAORTUP**

Variable Label Year of update of VAORTAKT/MAORTAKT
Variable Format 4-digit integer

Description The variable contains the year, in which the information stored in
VAORTAKT and MAORTAKT has been updated.

**VSBIL / MSBIL**

Variable Label Education of the father / Education of the mother
Variable Format 1-digit integer

Values

(0) Do Not Know
(1) Secondary School Degree
(2) Intermediate School Degree
(3) Technical School Degree
(4) Upper Secondary School Degree
(5) Other Degree
(6) No School Degree
(7) School Not Attended

Description The parents’ education is generated with information from the Youth
Questionnaire, the Biography Questionnaire and direct entries from the $pgen$-
files. Due to the filter command, the children’s proxy entries are only available
for VSBIL / MSBIL when the parents or one parent and the child do not live in
the same household at the time of the survey.

---

49 In gathering the information from different data sets, inconsistencies occurred. On the one hand, some parents
had been reported as deceased in the early waves, while information about their residence at a later date was
available. In this case, the information about the parents’ residence was not accepted.
**VBBIL / MBBIL**

Variable Label: Vocational training of the father / Vocational training of the mother

Variable Format: 2-digit integer

Values:

- (0) Do Not Know
- (10) No Vocational Degree
- (20) Vocational Degree
- (21) Trained in Foreign Company
- (22) Trained long Time in Foreign Company
- (23) Foreign Vocational School
- (24) Trade, Farming Apprentice
- (25) Business Apprentice
- (26) Health Care School
- (27) Special Technical School
- (28) Civil Service Training
- (29) Tech Engineer School
- (30) Foreign Collage
- (31) Foreign Collage
- (32) College, University
- (33) College, University
- (34) College, University
- (35) College, University
- (36) College, University
- (37) College, University
- (38) College, University
- (39) College, University
- (40) Other Training
- (50) Currently in Vocational Training
- (51) Currently in Schooling

Description: The parents’ vocational training is generated the same way as the education variables (see VSBIL / MSBIL).

**VSINFO / MSINFO**

Variable Label: Origin of the information on father’s education / Origin of the information on mother’s education

Variable Format: 1-digit integer

Values:

- (0) Do Not Know
- (1) Biography-Proxy
- (2) $P$-Individual Info

Description: The variable contains the origin of the information on parental education.

**VBINFO / MBINFO**

Variable Label: Origin of the information on father’s vocational training / Origin of the information on mother’s vocational training

Variable Format: 1-digit integer

Values:

- (0) Do Not Know
- (1) Biography-Proxy
- (2) $P$-Individual Info

Description: The variable contains the origin of the information on parental vocational training.
**VRELI / MRELI**

**Variable Label**  Religious affiliation of the father / Religious affiliation of the mother

**Variable Format**  1-digit integer

**Values**

(0) Do Not Know – Proxy
(1) Catholic
(2) Protestant
(3) Other Christian Denomination
(4) Islamic Denomination
(5) Other Denomination
(6) No Denomination

**Description**  The questions about the religious affiliation of the parents are only asked to youngsters who are not living in the household of their parents.

**VNAT / MNAT**

**Variable Label**  Nationality of the father / Nationality of the mother

**Variable Format**  1-digit integer

**Values**

(1) German
(2) Other

**Description**  The information on the parents’ nationality is generated similar to VRELI / MRELI. The question is only asked to youngsters who are not living in the same household as their parents. In addition, the parents’ personal numbers are used to match information on parents’ nationality with data from the $PGEN$-files in case there are missing entries.

**VBSTE LLP / MBSTE LLP**

**Variable Label**  Professional position of the father
(when the respondent was 15 years old) / Professional position of the mother
(when the respondent was 15 years old)

**Variable Format**  3-digit integer

**Description**  The children’s proxy entries on professional position and occupation of the father or mother (VBSTE LLP / MBSTE LLP) as well as VISCO88 /MISCO88 and all prestige scores are available when the parent and the child do not live in the same household at the time of the survey and if the parent lived in Germany when the child was 16 years old. Besides the proxy entries parents’ direct information from the $P$-files are used.
VBSINFO / MBSINFO

Variable Label  Origin of the information on the professional position of the father / 
Origin of the information on the professional position of the mother
Variable Format  1-digit integer

Values
(0) do Not Know-Proxy
(1) Biography-Proxy
(2) $P-$Individual Info

Description  The variables VBSINFO / MBSINFO are indicator variables. They tell whether the information is from the Biography or Youth or Person Questionnaires. This information is generated at the same steps as it is done with the VBSTELL / MBSTELL variables.

VISCO88 / MISCO88

Variable Label  Professional occupation of the father 
(when the respondent was 15 years old) / 
Professional occupation of the mother 
(when the respondent was 15 years old)
Variable Format  4-digit integer

Description  The variables contain the ISCO88 code for the father and mother. See also documentation on $pgen (variable IS88$$).

VISEI / MISEI

Variable Label  Prestige score of father – concept of Ganzeboom / 
Prestige score of mother – concept of Ganzeboom
Variable Format  2-digit integer

Description  The variables contain the ISEI code for the father and mother. See also documentation on $pgen (variable ISEI$$).

VMPS / MMPS

Variable Label  Prestige score of father – Magnitude scale – Wegener / 
Prestige score of mother – Magnitude scale – Wegener
Variable Format  5-digit real

Description  The variables contain the prestige scores (magnitude scale - Wegener) for the father and mother. See also documentation on $pgen (variable MPS$$).

VSIOPS / MSIOPS

Variable Label  Prestige score of father – Treiman standard score / 
Prestige score of mother – Treiman standard score
Variable Format  2-digit integer
Description The variables contain the prestige scores (Treiman standard score) for the father and mother. See also documentation on $pgen$ (variable SIOPSS$).

**VEGP / MEGP**

| Variable Label | Prestige score of father – Erikson – Goldthorpe class category/
| Prestige score of mother – Erikson – Goldthorpe class category |
| Variable Format | 2-digit integer |
| Values | (-1) No answer  
(-2) Does not apply  
(-3) Answer improbable |

Description The variables contain the prestige scores (EGP) for the father and mother. See also documentation on $pgen$ (variable EGP$).

**VBKLAS / MBKLAS**

| Variable Label | Occupational coding scheme father according German statistical office/
| Occupational coding scheme mother according German statistical office |
| Variable Format | 4-digit integer |

Description: The variables contain the occupational code for the father and mother according to the coding scheme of the German statistical office. See also documentation on $pgen$ (variable CLASS$).

**ORTKINDH / ORTKIND1**

| Variable Label | ORTKINDH: Place of childhood /
| ORTKIND1: Still lives in place of childhood? |
| Variable Format | 1-digit integer |
| Values | ORTKINDH:  
(1) Large City  
(2) Medium City  
(3) Small City  
(4) Countryside  
ORTKIND1:  
(1) Yes, Still  
(2) Yes, Again  
(3) No  

Description The variables provide information on the place of childhood.

**LIVING1 - LIVING8**

| Variable Label | LIVING1: No. of years living with both parents  
LIVING2: No. of years living alone with mother  
LIVING3: No. of years living with mother and new partner of mother  
LIVING4: No. of years living alone with father  
LIVING5: No. of years living alone with father and new partner of father  
LIVING6: No. of years living with other relatives  
LIVING7: No. of years living with foster parents  
LIVING8: No. of years living in youth center |
Variable Format 2-digit integer
Description The variables show the total number of years for different categories of where the child lived during his childhood.

VSTREIT / MSTREIT
Variable Label Conflict with father /
Conflict with mother
Variable Format 1-digit integer
Values
(-1) No answer
(-2) Does not apply
(-3) Answer improbable
(1) Very Often
(2) Often
(3) Sometimes
(4) Seldom
(5) Never
(6) Person Not Present

Description The variables provide information on the frequency of conflicts with the parents. It is only asked in the Youth Questionnaire.

BIOYEAR
Variable Label Year of the Biography Survey
Variable Format 4-digit integer
Values
(-1) No answer
(-2) Does not apply
(-3) Answer improbable

Description The variable BIOYEAR provides the year in which the information was surveyed. Since in FiD, the Biography Questionnaire is split up into two parts for individuals first interviewed in 2010 and 2011, respondents generally answered the biography in two different years. BIOYEAR then provides the year when the second part (in which the main part about a respondent’s parents is collected) is answered. If the second part was never answered, the year of filling out the first biography part is stored in BIOYEAR. For first time respondents of 2012 or later, the biography is not split up. Youths do not fill out a biography questionnaire, so the year of filling out the youth questionnaire appears in BIOYEAR. For all cases, BIOYEAR is changed to the most recent year in which any sort of information was taken into the data.

BIO
Variable Label Form of Biography Questionnaire
Variable Format 1-digit integer
Values
(-1) No answer
(-2) Does not apply
(-3) Answer improbable  
(1) Youth  
(2) Biolela blue  
(3) Only first part of Biography Questionnaire  
(4) Only second part of Biography Questionnaire  
(5) Both parts of Biography Questionnaire

Description The variable BIO is generated to indicate the origin of the information in BIOPAREN (Youth Questionnaire or one or both parts of the Biography Questionnaire).

**ALTER / VALTER / MALTER**

Variable Label Age of the respondents / Age of the respondent’s father / Age of the respondent’s mother

Variable Format 2-digit integer

Values (-1) No answer  
(-2) Does not apply  
(-3) Answer improbable

Description The variable ALTER gives the age of the respondent at the moment of the interview. VALTER gives the age of the respondents’ father when the respondent answered the Biography Questionnaire or the Youth Questionnaire. The same was applied for the mothers with the variable MALTER. In order to generate the variables the information for the parents who are identified in the FID was gained with data from PPFAD. The proxy entries from BIOPAREN were used when there weren’t any information of the respondents parents available.

**VORIGIN / MORIGIN**

Variable Label Country of origin of the respondent’s father / Country of origin of the respondent’s mother

Variable Format 3-digit integer

Values (-1) No answer  
(-2) Does not apply  
(-3) Answer improbable

Description These variables give information about the country of origin of the respondents mother (MORIGIN) and father (VORIGIN). This information is collected in the Youth and the Biography Questionnaire. Another source of information can be found in PPFAD by the direct-entries of the parents in the variable CORIGIN. These two kinds of information, proxy- and direct-entries, are used to generate MORIGIN and VORIGIN. In a first step we use the proxy-information for all the parents whose children made an entry in the Youth Questionnaire. For all the parents where there are no proxy-information available, we then use the direct-entries of the parents from the PPFAD-variable CORIGIN.
GESCHW
Variable Label: Siblings yes/no
Variable Format: 1-digit integer

Values
(-1) No answer
(-2) Does not apply
(-3) Answer improbable
(1) Yes
(2) No

Description: GESCHW contains the information whether a respondent has siblings or not. The question was asked in the Youth Questionnaire in 2010 and since 2011 it is asked in both Youth and Biography Questionnaire.

GESCHWUP
Variable Label: Time of update – siblings
Variable Format: 4-digit integer

Values
(-1) No answer
(-2) Does not apply
(-3) Answer improbable

Description: GESCHWUP contains the year, in which the latest sibling information was surveyed.

NUMS
Variable Label: Number of sisters
Variable Format: 2-digit integer

Values
(-1) No answer
(-2) Does not apply
(-3) Answer improbable

Description: NUMS contains the number of sisters. The question was asked in the Youth Questionnaire in 2010 and since 2011 it is asked in both Youth and Biography Questionnaire.

NUMB
Variable Label: Number of brothers
Variable Format: 2-digit integer

Values
(-1) No answer
(-2) Does not apply
(-3) Answer improbable

Description: NUMB contains the number of brothers. The question was asked in the Youth Questionnaire in 2010 and since 2011 it is asked in both Youth and Biography Questionnaire.
**TWIN**

Variable Label: Twin sister/brother  
Variable Format: 1-digit integer  

Values:  
- (-1) No answer  
- (-2) Does not apply  
- (-3) Answer improbable  
- (1) Yes, monozygotic  
- (2) Yes, dizygotic  
- (3) No  

Description: TWIN contains information whether the respondent has a twin sibling. The question was asked in the Youth Questionnaire in 2010 and 2011
# Änderungen in den Versionen 1.1-4.0 der FiD Datenweitergabensetzungen

<table>
<thead>
<tr>
<th>Nummer</th>
<th>Datensatz</th>
<th>Beschreibung</th>
<th>in Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1</td>
<td>bioage03</td>
<td>Die Variablen &quot;b03stcaremy&quot; und &quot;b03stcareyry&quot; bezeichneten nicht den Monat und das Jahr, in dem die Betreuung angefangen wurde.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.2</td>
<td>bioage03</td>
<td>Die Variablen &quot;b03spe6&quot; und &quot;b03spe5&quot; sind bei der Screening-Stichprobe miteinander vertauscht worden.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.3</td>
<td>bioage03</td>
<td>Die Labels der Variablen &quot;b03spe7&quot;, &quot;b03spe8b&quot;, &quot;b03spe6&quot;, und &quot;b03spe5&quot; waren vertauscht.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.4</td>
<td>bioage03</td>
<td>Die Variablen &quot;b03health8&quot;, &quot;b03health1&quot; und &quot;b03health7&quot; sind in der Screening-Stichprobe miteinander vertauscht worden.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.5</td>
<td>bioage10p1/p2</td>
<td>Die Variablen &quot;b10imper1e&quot;-&quot;b10imper3e&quot; wurden neu gelabeled.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.6</td>
<td>bioage10p1/p2</td>
<td>Die Variablen &quot;b10biopar&quot; und &quot;b10sexresp&quot; waren nicht aufeinander abgestimmt.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.7</td>
<td>f10eltern1</td>
<td>Geburtsmonate wurden verändert.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.8</td>
<td>f10eltern2</td>
<td>Geburtsmonate wurden verändert.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.9</td>
<td>f10eltern3</td>
<td>Geburtsmonate wurden verändert.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.10</td>
<td>f10eltern3</td>
<td>Die Variablen &quot;f10e337e&quot;, &quot;f10e337f&quot; und &quot;f10e337g&quot; sind in der Screening-Stichprobe miteinander vertauscht worden.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.11</td>
<td>f10eltern3</td>
<td>Die Variable &quot;f10e335a6&quot; und &quot;f10e335a7&quot; sind bei der Screening-Stichprobe miteinander vertauscht worden.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.12</td>
<td>f10eltern4</td>
<td>Durch falsche Bespaltung sind die Variablen &quot;f10e404b&quot; (Größe) und &quot;f10e404a&quot; (Gewicht) nicht korrekt eingelesen worden.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.13</td>
<td>f10eltern5</td>
<td>Geburtsmonate wurden verändert.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.14</td>
<td>f10eltern5</td>
<td>Geburtsmonate wurden verändert.</td>
<td>1.1</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>1.1.15</td>
<td>f10eltern6</td>
<td>Die Variablen &quot;f10e621e1&quot;-&quot;f10e621e3&quot; wurden neu gelabeled.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.16</td>
<td>f10eltern6</td>
<td>Geburtsmonate wurden verändert.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.17</td>
<td>f10jugend</td>
<td>Bei den Fragen nach den Freizeitbeschäftigungen (f10j015c,e,f,g,i,o) wurden bei der Dateneingabe einzelne Items vertauscht, die aber einfach umgesetzt werden können. c=&gt;n; d=&gt;c; e=&gt;d; f=&gt;e; g=&gt;f; h=&gt;g; i=&gt;h; n=&gt;j; j=&gt;i</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.18</td>
<td>f10lela</td>
<td>Durch einen Softwarefehler wurde die Frage nach der Anerkennung eines Abschlusses im Ausland in der Screening Stichprobe nicht gestellt. Entsprechend ist die Variable &quot;f10l038c&quot; für diese Stichprobe auf &quot;-2&quot; zu setzen. Erst nach dem 18.6.2010 wurde diese Frage in der Kohorte gestellt - entsprechend werden auch hier alle anderen Werte auf &quot;-2&quot; gesetzt.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.19</td>
<td>f10lela</td>
<td>Die Variable &quot;sample&quot; (jetzt &quot;sample1&quot;) war mit falschen Werten belegt.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.20</td>
<td>f10mihinc</td>
<td>Umkodierung der nicht-imputierten Werte auf -1 anstelle von -2 in den Originaldaten.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.21</td>
<td>f10mipinc</td>
<td>Umkodierung der nicht-imputierten Werte auf -1 anstelle von -2 in den Originaldaten.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.22</td>
<td>f10p</td>
<td>Die Kalenderdaten aus f10p070 waren in der ersten Datenlieferung nicht enthalten.</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1.23</td>
<td>f10pgen</td>
<td>Die Variable &quot;nation10&quot; hat aufgrund falscher Bespaltung nicht die korrekten Ausprägungen.</td>
<td>1.1</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>1.2.1</td>
<td>alle</td>
<td>In allen Datensätzen bezeichnet &quot;sample1&quot; die Stichprobenzugehörigkeit. Ausnahmen sind <em>fi0pbrutto</em> und <em>fi0kind</em> (&quot;psample&quot;) und <em>fi0hbrutto</em> (&quot;hsample&quot;)</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.2</td>
<td>alle</td>
<td>Sämtliche value Labels sind gleichlautend wie der Variablenname (einschließlich Groß- und Kleinschreibung).</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.3</td>
<td>alle</td>
<td>Die Ausprägungen in Variablen mit nur einem Wert wurden gelabeled.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.4</td>
<td><strong>Neue Gewichte</strong></td>
<td>Die Gewichte, die eine gemeinsame Hochrechnung von SOEP und FiD erlauben, befinden sich im Datensatz hhrf_fidsoep (Haushaltsebene) und phrf_fidsoep (Personenebene). Sie enthalten die Variablen zum Anspielen an andere Datensätze („hhnrakt“ sowie „persnr“) und jeweils drei Gewichtungsfaktoren pro Welle (in der Datenweitergabeversion 1.2 also nur drei). Diese drei Gewichtungsfaktoren ermöglichen eine Gewichtung einer kombinierten SOEP-FiD Population mit den Gewichtungsfaktoren „f10hhrf_soep“ (Haushalte) und „f10phrf_soep“ (Personen). Für den Fall, dass nur die FiD-Screening-Stichprobe bzw. nur die FiD-Kohorten-Stichprobe mit dem SOEP zusammen analysiert werden, sollten die jeweiligen Gewichte mit den Kürzeln „_scr“ (Screening) bzw. „_coh“ (Kohorte) verwendet werden.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.5</td>
<td>bioage01</td>
<td>In den Variablen zu Störungen (&quot;b01disord1&quot;-&quot;b01disord8&quot;) wurde noch der Code für &quot;-1 no answer&quot; hinzugefügt (vorher &quot;-2&quot;).</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.6</td>
<td>bioage01</td>
<td>Die Variable &quot;b01care&quot; wurde in zwei Variablen &quot;b01ccare&quot; und &quot;b01ccarei&quot; umgesetzt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.7</td>
<td>bioage01</td>
<td>Die Variable &quot;b01noccar10&quot; wurde in &quot;b01noccare10&quot; umbenannt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.8</td>
<td>bioage01</td>
<td>Die Variable &quot;b01breastf&quot; wurde auf eine &quot;1&quot;/&quot;2&quot;-Kodierung umgestellt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.9</td>
<td>bioage01</td>
<td>Die Variable &quot;b01breastfm&quot; wurde für die Mütter, die momentan noch stillen auf &quot;-2 Does not apply&quot; gesetzt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.10</td>
<td>bioage01</td>
<td>Die Variable &quot;b01medaid3m&quot; wurde in &quot;b01medaid3m&quot; umbenannt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.11</td>
<td>bioage01</td>
<td>Die Variable &quot;b01age&quot; wurde neu kodiert: Nachdem Kinder vorher mindestens einen Monat alt waren, ist nun ein Alter von 0 Monaten möglich.</td>
<td>1.2</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>1.2.12</td>
<td>bioage01</td>
<td>Die Variable &quot;b01suppartn&quot; wurde ans SOEP angepasst und das Labeling geändert: &quot;0&quot; is nun &quot;No partner in hh&quot;, &quot;1&quot; - &quot;No support at all&quot;, &quot;2&quot; - &quot;Some support&quot;, &quot;3&quot; - &quot;Strong support&quot;, &quot;4&quot; - &quot;Very strong support&quot;.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.13</td>
<td>bioage01</td>
<td>Durch Fehler in der Kodierung wurden einige Werte in &quot;b01preend&quot; und &quot;b01prebeg&quot; geändert.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.14</td>
<td>bioage02</td>
<td>Die Variable &quot;b02breastf&quot; wurde auf eine &quot;1&quot;/&quot;2&quot;-Kodierung umgestellt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.15</td>
<td>bioage02</td>
<td>Die Variable &quot;b02breastfm&quot; wurde für die Mütter, die momentan noch stillen auf &quot;-2 Does not apply&quot; gesetzt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.16</td>
<td>bioage02</td>
<td>Die Variable &quot;b02age&quot; wurde analog zu &quot;b01age&quot; neu kodiert.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.17</td>
<td>bioage02</td>
<td>Die Variable &quot;b02care&quot; wurde in zwei Variablen &quot;b02care&quot; und &quot;b02ccarei&quot; umgesetzt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.18</td>
<td>bioage02</td>
<td>Für die Variable &quot;b02breastfp&quot; wurden missing codes (&quot;-1&quot;, &quot;-2&quot;) eingefügt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.19</td>
<td>bioage02</td>
<td>Für die Variable &quot;b02lstmedex&quot; wurde die Ausprägung &quot;0&quot; mit missing codes (&quot;-1&quot;, &quot;-2&quot;) kodiert.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.20</td>
<td>bioage02</td>
<td>Die Variablen &quot;b02nactcar1&quot;, &quot;b02nactcar2&quot;, &quot;b02nactcar3&quot; wurden in &quot;b02nactcare1&quot;, &quot;b02nactcare2&quot;, &quot;b02nactcare3&quot; umbenannt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.21</td>
<td>bioage03</td>
<td>Die Variablen &quot;b03change1&quot;-&quot;b03change8&quot; wurden für die Screening-Stichprobe auf &quot;-2 Does not apply&quot; gesetzt (vorher &quot;0&quot;).</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.22</td>
<td>bioage03</td>
<td>Die Variable &quot;b03care&quot; wurde in zwei Variablen &quot;b03care&quot; und &quot;b03ccarei&quot; umgesetzt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.23</td>
<td>bioage03</td>
<td>Die Variable &quot;b03breastf&quot; wurde auf eine &quot;1&quot;/&quot;2&quot;-Kodierung umgestellt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.24</td>
<td>bioage03</td>
<td>Für die Variable &quot;b03nmedaid&quot; wurde jetzt eine &quot;0&quot; kodiert, wenn keine Arztbesuche vorlagen.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.25</td>
<td>bioage03</td>
<td>Für die Variablen &quot;b03hospital&quot; und &quot;b03hospital3m&quot; wurde jetzt eine &quot;0&quot; kodiert, wenn kein Krankenhausaufenthalt vorlag. (In der Screening-Stichprobe bleibt die Variable &quot;b03hospital3m&quot; auf &quot;-2 Does not apply&quot;.)</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.26</td>
<td>bioage03</td>
<td>Für die Variable &quot;b03lstmedex&quot; wurde die Ausprägung &quot;0&quot; mit missing codes (&quot;-1&quot;, &quot;-2&quot;) kodiert.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.27</td>
<td>bioage03</td>
<td>Die Variable &quot;b03age&quot; wurde analog zu &quot;b01age&quot; neu kodiert.</td>
<td>1.2</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>1.2.28</td>
<td>bioage06</td>
<td>Die Variablen &quot;b06weight&quot; und &quot;b06height&quot; waren durch einen Fehler beim Einlesen des <em>f10eltern4</em> Datensatzes falsch kodiert (siehe auch dort).</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.29</td>
<td>bioage06</td>
<td>Die Variablen &quot;b06char6o&quot;, &quot;b06char7o&quot;, &quot;b06char8o&quot;, &quot;b06char9o&quot; wurden in &quot;b06char6&quot;, &quot;b06char7&quot;, &quot;b06char8&quot; und &quot;b06char9&quot; umbenannt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.30</td>
<td>bioage08p1/p2</td>
<td>Die Variable &quot;b08waychi&quot; wurde in &quot;b08chhealth&quot; umbenannt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.31</td>
<td>bioage08p1/p2</td>
<td>Die Variable &quot;b08allownce&quot; wurde gerundet auf ganze Zahlen.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.32</td>
<td>bioage08p1/p2</td>
<td>In den Variablen &quot;b08lamark&quot; und &quot;b08matmark&quot; wurden die System-Missings auf &quot;-1&quot; gesetzt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.33</td>
<td>bioage08p1/p2</td>
<td>Durch eine neue Geschlechtszuordnung verschiebt sich ein Fall aus p1 in p2 und umgekehrt, wodurch sich in einem Fall jeweils mehrere Variablenausprägungen ändern.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.34</td>
<td>bioage10p1/p2</td>
<td>Die Variablen &quot;b10freqfre1&quot;-&quot;b10freqfre14&quot; wurden in &quot;b10freqact1&quot;-&quot;b10freqact14&quot; umbenannt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.35</td>
<td>bioage10p1/p2</td>
<td>Die Variable &quot;b10allownce&quot; wurde gerundet auf ganze Zahlen.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.36</td>
<td>bioage10p1/p2</td>
<td>In den Variablen &quot;b10lamark&quot; und &quot;b10matmark&quot; wurden die System-Missings auf &quot;-1&quot; gesetzt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.37</td>
<td>bioage10p1/p2</td>
<td>Durch eine neue Geschlechtszuordnung verschiebt sich ein Fall aus p1 in p2 und umgekehrt, wodurch sich in einem Fall jeweils mehrere Variablenausprägungen ändern.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.38</td>
<td>biobirth</td>
<td>Der Datensatz wurde auf Personen mit einem Personeninterview sowie die Jugendlichen beschränkt, nachdem vorher versucht wurde, aus anderen Informationen im Haushalt auf eigene Kinder rückzuschließen.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.39</td>
<td>biocouply</td>
<td>Durch Verbesserungen in der Kodierung der Partnerschaftsbeziehungen (Berücksichtigung der Nicht-Antwortenden) erhöhen sich die Fallzahlen.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.40</td>
<td>biomarsy</td>
<td>Durch Verbesserungen in der Kodierung der Partnerschaftsbeziehungen (Berücksichtigung der Nicht-Antwortenden) erhöhen sich die Fallzahlen.</td>
<td>1.2</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>1.2.41</td>
<td>f10hbrutto</td>
<td>Einige Fälle (19) waren in &quot;f10bula&quot; als &quot;Rheinland-Pfalz&quot; codiert, kommen aber eigentlich aus dem &quot;Saarland&quot;.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.42</td>
<td>f10hbrutto</td>
<td>Die Variable &quot;f10intza&quot; (Zahl der eingesetzten Interviewer) wurde eingeführt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.43</td>
<td>f10hgen</td>
<td>Die Variable &quot;cnstry10&quot; wurde umbenannt in &quot;cnstyr10&quot;</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.44</td>
<td>f10hgen</td>
<td>Umbenennung der Variablen &quot;elec10&quot; in &quot;electr10&quot;, sowie &quot;felec&quot; in &quot;felectr10&quot; (SOEP Standard seit 2010).</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.45</td>
<td>f10hgen</td>
<td>Die imputierten Variablen (&quot;size10&quot;, &quot;room10&quot;, &quot;rent10&quot;, &quot;heat10&quot;, &quot;util10&quot;, &quot;electr10&quot;, &quot;i1hinc10&quot;, &quot;i2hinc10&quot;, &quot;i3hinc10&quot;, &quot;i4hinc10&quot;, &quot;i5hinc10&quot;) haben durch die neuen Imputationen neue Werte. Außerdem wurde die Imputationsroutine für &quot;util10&quot; geändert, so dass &quot;futil10&quot; zusätzliche Impuationen aufweist.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.46</td>
<td>f10hgen</td>
<td>Einige Fälle (89) waren in &quot;nuts10&quot; als &quot;Rheinland-Pfalz&quot; codiert, kommen aber eigentlich aus dem &quot;Saarland&quot; (siehe f10hbrutto).</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.47</td>
<td>f10hgen</td>
<td>Zur Variable &quot;typ2hh10&quot; wurde die Ausprägung &quot;73 Grandparent(s)-grandchild(ren) hh&quot; hinzugefügt, so dass die Abgrenzungen genauer werden. Dadurch verändert sich auch typ1hh10 in einigen Fällen.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.48</td>
<td>f10hgen</td>
<td>Bei der Berechnung von &quot;ahinc&quot; wurde jetzt auch das Einkommen aus Vermietung und Verpachtung berücksichtigt (betrifft 130 Fälle).</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.49</td>
<td>f10hgen</td>
<td>Die Variablen &quot;singpa&quot;, &quot;lrgfam&quot; und &quot;lowinc&quot; wurden als stichprobenpezifische und wellenunveränderliche Variablen in hpfad überführt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.50</td>
<td>f10intview</td>
<td>In der Variable &quot;f10dura&quot; wurden alle Elternfragebögen von system-missing (&quot;.&quot;.) auf &quot;-2&quot; gesetzt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.51</td>
<td>f10jugend</td>
<td>Die Ausprägungen der Variablen &quot;f10j062&quot;, &quot;f10j066b&quot;, &quot;f10j068&quot;, &quot;f10j079d&quot; wurden gelabeled.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.52</td>
<td>f10kind</td>
<td>Es wurden einige Geburtsmonatsveränderungen mit aufgenommen (Vergleich mit f10pbrutto).</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.53</td>
<td>f10kind</td>
<td>Die Indikatoren der Mutter- und Vaterzeiger wurden verbessert, und benutzen nun auch die Informationen aus den Beziehungenhistorien (biocouply). Dadurch konnten zahlreiche unbekannte Eltern-Kind-Beziehungen zugeordnet werden.</td>
<td>1.2</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>1.2.54</td>
<td>f10lela</td>
<td>Die Ausprägungen Variablen &quot;f10l003&quot; (Geburtsland), &quot;f10l008b&quot; (Land zweiter Staatsangehörigkeit) und &quot;f10l010&quot; (Alte Staatsangehörigkeit) waren nicht gelabeled.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.55</td>
<td>f10lela</td>
<td>Umkodierung von Beziehungsvariablen im Haushalt mit der &quot;hhnrakt&quot; 20000876 (siehe f10pbrutto).</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.56</td>
<td>f10lela</td>
<td>Umkodierung der Variable &quot;f10l064e2&quot; auf 2003 im Haushalt mit der &quot;hhnrakt&quot; 20000876.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.57</td>
<td>f10mihinc</td>
<td>Verbesserte Imputationen.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.58</td>
<td>f10mipinc</td>
<td>Verbesserte Imputationen, insbesondere Imputationen der nicht-antwortenden Haushaltsmitglieder (PUNRs). Diese sind durch die zusätzliche Variable &quot;punr&quot; gekennzeichnet.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.59</td>
<td>f10mipinc</td>
<td>Die Variablen &quot;f10pchdsup&quot; bzw. &quot;i_f10pchdsup&quot; wurden in &quot;f10palimon&quot; bzw. &quot;i f10palimon&quot; umbenannt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.60</td>
<td>f10p</td>
<td>Der Variable &quot;f10p102n&quot; (Sonstige Sorgen) wurden labels hinzugefügt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.61</td>
<td>f10pbrutto</td>
<td>Die Variable &quot;f10stell&quot; wurde im Haushalt mit der &quot;hhnrakt&quot; 20000876 geändert.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.62</td>
<td>f10pbrutto</td>
<td>Die Variable &quot;f10hhnrold&quot; wurde für alle Fälle von &quot;0&quot; auf &quot;-2&quot; gesetzt.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.63</td>
<td>f10pgen</td>
<td>Die Variablen &quot;agrhrs10&quot;, &quot;acthrs10&quot; und &quot;ovrhrs10&quot; wurden mit einem neuen Umrechnungsfaktor für Wochenstunden berechnet.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.64</td>
<td>f10pgen</td>
<td>Durch Verbesserungen in biomars und biocouply wurden die Variablen &quot;partp10&quot;, &quot;partno10&quot;, &quot;coupst10&quot;, &quot;coupid10&quot; und &quot;marrst10&quot; in einigen Fällen verändert.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.65</td>
<td>f10pgen</td>
<td>Durch genauere Kodierung der Abschlüsse in der ehemaligen DDR haben sich in den Variablen &quot;scedu10&quot;, &quot;scedue10&quot;, &quot;vcedege10&quot;, &quot;colleg10&quot;, &quot;timedu10&quot;, &quot;casmin10&quot; und &quot;isced10&quot; einige Fälle verändert.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.66</td>
<td>f10pgen</td>
<td>Durch falsche Kodierung in der Variable &quot;occpos10&quot; waren 486 Individuen falsch als Rentner (Wert &quot;13&quot;) kodierte, die entweder nicht erwerbstätig (&quot;10&quot;) sind, keine Angabe gemacht haben (&quot;-1&quot;) oder auf die die Frage nicht zutrifft (&quot;-2&quot;). Entsprechend ändert sich auch &quot;egp10&quot;.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.67</td>
<td>f10pgen</td>
<td>Die arbeitenden Jugendlichen bekommen jetzt einen Erwerbstatus zugewiesen, so dass sich die Variable &quot;lfs10&quot; ändert.</td>
<td>1.2</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>----------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>1.2.68</td>
<td>f10pgen</td>
<td>Durch die neuen Imputationen verändern sich die Variablen &quot;labgro10&quot; und &quot;labnet10&quot;.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.69</td>
<td>hbrutt10_fid</td>
<td>Die Variablen &quot;f10apro&quot;, &quot;f10wuma1&quot;, &quot;f10wuma2&quot;, &quot;f10wuma3&quot;, &quot;f10wuma401&quot;, &quot;f10wuma402&quot;, &quot;f10wuma403&quot;, &quot;f10wuma404&quot;, &quot;f10wuma4ka&quot;, &quot;f10wuma5&quot;, &quot;f10wuma6&quot; und &quot;f10wuma7&quot; wurden für die Screening Stichprobe auf &quot;,-2 Trifft nicht zu&quot; gesetzt (vorher &quot;0&quot;).</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.70</td>
<td>hpfad</td>
<td>Die Variablen &quot;singpa&quot; (Alleinerziehend bei erstem Kontakt), &quot;Irgfam&quot; (Mehrkindfamilie bei erstem Kontakt) und &quot;lowinc&quot; (Niedrigkommenshaushalt bei erstem Kontakt) wurden aus f10hgen in hpfad übernommen, weil es sich hierbei um Informationen handelt, die nicht über die Wellen ändern und außerdem Informationen über die Stichproben geben.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.71</td>
<td>ppfad</td>
<td>Durch Verbesserungen in den Mutter- und Vaterzeigern konnten zahlreiche unbekannte Beziehungen in den Variablen &quot;germborn&quot;, &quot;immiyear&quot;, &quot;corigin&quot; und &quot;migback&quot; neu zugordnet werden.</td>
<td>1.2</td>
</tr>
<tr>
<td>1.2.72</td>
<td>ppfad</td>
<td>In Vorbereitung auf die neue Welle bekommt die Variable &quot;f10netto&quot; neue labels: &quot;110 FiD: Interviewee P-Interview&quot;, &quot;111 FiD: Interviewee P-Interview &amp; Biography I&quot;, &quot;112 FiD: Interviewee P-Interview &amp; Biography II&quot;, &quot;113 FiD: Interviewee P-Interview &amp; Biography I+II&quot;</td>
<td>1.2</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>2.0.1</td>
<td>bioage01-10</td>
<td>Die Sprachen, in denen mit den Kindern im Haushalt gesprochen wird, sind nun codiert nach ISO-639-1 Standard in den Datensätzen enthalten als &quot;$language01&quot;-&quot;$language03&quot;.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.2</td>
<td>bioage01-10</td>
<td>Die Variablen &quot;care6&quot;, &quot;care8&quot;, und &quot;care8h&quot; wurden wegen Problemen in der Befragung neu berechnet. Eine genauere Beschreibung findet sich in der bioage01-10 Dokumentation.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.3</td>
<td>bioage01-10</td>
<td>Die Variablen &quot;care1h&quot;-&quot;care12h&quot; wurden so berechnet, dass die Gesamtsumme der Betreuungszeit maximal 168 Stunden beträgt.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.4</td>
<td>bioage01-10</td>
<td>Es wurden einige Variablennamen in den bioage Files geändert; sämtliche neuen Namen finden sich in der entsprechenden Tabelle in der bioage01-10 Dokumentation. Ebenso wurden einige Variablen und Value Labels geändert, um eine bessere Benutzung zu ermöglichen.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.5</td>
<td>bioage06 -bioage10</td>
<td>In diesen drei bioage Datensätzen wurde die Variable &quot;bXXage&quot; (Alter des Kindes in Monaten) hinzugefügt, so dass jetzt sämtliche bioage files diese Variable enthalten.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.6</td>
<td>bioage08, bioage10</td>
<td>In den Variablen &quot;scoldura&quot;, &quot;hospital&quot; und &quot;nmedaid&quot; wurden kleinere Codierungsfehler beseitigt.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.7</td>
<td>biomearsy</td>
<td>Durch neu eingeführte Fragen im P-Fragebogen nach gleichgeschlechtlichen Partnerschaften haben sich auch die &quot;speltyp&quot; Variablen in biomearsy und biocouply entsprechend geändert.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.8</td>
<td>f10eltern1</td>
<td>Die Benennung der Variablen aus der Frage 28 (Betreuungseinrichtungen/-personen) wurde an das FiD-Prinzip angeglichen. Vorherige Variablen &quot;f10e128a&quot;-&quot;f10e128a28o&quot; heißen jetzt &quot;f10e128a1&quot;, &quot;f10e128a2&quot; bis &quot;f10e128g1&quot;, &quot;f10e128g2&quot;, &quot;f10e128h&quot;.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.9</td>
<td>f10eltern1- f10eltern3</td>
<td>Für die U-Untersuchungen in &quot;f10e109a/b&quot;, &quot;f10e212a/b&quot; und &quot;f10e312a/b&quot; wurden Value Labels eingesetzt.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.10</td>
<td>f10h</td>
<td>Die Variablen &quot;f10h042a&quot;-&quot;f10h042u&quot; (Transferleistungen im letzten Jahr) wurden in ihrer Benennung an die FiD-Logik angepasst, und in &quot;f10h042a1&quot;, &quot;f10h042a2&quot; bis &quot;f10h042g3&quot; umbenannt. Gleiches gilt für &quot;f10h044a&quot;-&quot;f10h044o&quot; (aktuelle Transferleistungen), die zu &quot;f10h042a1&quot;-&quot;f10h042g2&quot; umbenannt wurden.</td>
<td>2.0</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>2.0.11</td>
<td>f10hbrutto</td>
<td>Die Variable f10mkz2 (Migrantenkennzeichen) wurde neu kodiert, &quot;0&quot; wird zu &quot;1&quot;, &quot;1&quot; zu &quot;2&quot;, &quot;2&quot; zu &quot;3&quot;.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.12</td>
<td>f10hgen</td>
<td>Die Variable &quot;rent10&quot; war durch einen Kodierungsfehler falsch berechnet. Sie ist jetzt korrigiert und sollte ausschließlich benutzt werden.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.13</td>
<td>f10hgen</td>
<td>Durch neue Imputationen haben sich alle imputierten Werte (&quot;rent10&quot;, &quot;heat10&quot;, &quot;util10&quot;, electr10, i_hinc10, size10, room10) verändert.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.14</td>
<td>f10lela</td>
<td>Das Label des Wertes &quot;1 Trennung&quot; in der Variable &quot;f10l062f&quot; (Ende der Beziehung) von auf &quot;2 Scheidung&quot; geändert.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.15</td>
<td>f10lela</td>
<td>In der Variable &quot;f10l062f&quot; (Ende der Beziehung) wurden drei Fälle nachträglich auf -.2 Trifft nicht zu&quot; umgesetzt, deren Ehe noch Bestand hatte (&quot;f10l062d&quot;=1), die aber aus technischen Gründen Scheidung bzw. Tod des Ehepartners angegeben hatten.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.16</td>
<td>f10mihinc</td>
<td>Mit den neu berechneten Imputationen ergeben sich auch neue Werte für &quot;i_hinc&quot; in hgen, so dass ausschließlich f10hgen der neuen Distribution genutzt werden sollte.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.17</td>
<td>f10mipinc</td>
<td>Die Variable &quot;f10psyrinc&quot; (imputiertes Einkommen des letzten Jahres) wurde bei den neuen Imputationen nicht mehr berücksichtigt.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.18</td>
<td>f10mipinc</td>
<td>Mit den neu berechneten Imputationen ergeben sich auch neue Werte für &quot;labgro10&quot; und &quot;labnet10&quot; in f10pgen, so dass ausschließlich f10pgen der neuen Distribution genutzt werden sollte.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.19</td>
<td>f10p</td>
<td>In der Variable &quot;f10p030&quot; (Art des Arbeitsvertrags) wurden die Value Labels korrigiert, auf jetzt &quot;1 unbefristet&quot; und &quot;2 befristet&quot;.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.20</td>
<td>f10pgen</td>
<td>Wegen identisch benannter Variablen wurde &quot;fsize10&quot; in &quot;cosize10&quot; umbenannt, sowie &quot;fcsize10&quot; in &quot;crsize10&quot;.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.21</td>
<td>f10pgen</td>
<td>Die Variablen &quot;expft10&quot;, &quot;exppt10&quot; und &quot;expue10&quot; sowie &quot;tenure10&quot; sind rückwirkend aus den Daten aus 2011 berechnet und eingeführt worden.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.22</td>
<td>f10pgen</td>
<td>Durch neue Imputationen haben sich die imputierten Einkommen (&quot;labgro10&quot;, &quot;labnet10&quot;) verändert.</td>
<td>2.0</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>2.0.23</td>
<td>f10pgen</td>
<td>Durch neu eingeführte Fragen im P-Fragebogen nach gleichgeschlechtlichen Partnerehen haben sich auch die Variablen &quot;marrst10&quot; und &quot;coupst10&quot; entsprechend geändert.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.24</td>
<td>f10pkal</td>
<td>Die Variable &quot;f10p1o&quot; (Short-Time working Jan-Dec 2009) und die dazugehörigen Variablen &quot;f10p1o001&quot;-&quot;f10p1o012&quot; wurden umbenannt in &quot;f10p1k&quot; bzw. &quot;f10p1k001&quot;-&quot;f10p1k012&quot;. Damit heißen diese Variablen wie im SOEP 2010.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.25</td>
<td>f10pkal</td>
<td>Die Indikator-Variablen &quot;f10p1a01&quot;-&quot;f10p1n01&quot; sowie die dazugehörigen Monatsangaben &quot;f10p1a02&quot;-&quot;f10p1n02&quot; wurden in die Weitergabe mit aufgenommen.</td>
<td>2.0</td>
</tr>
<tr>
<td>2.0.26</td>
<td>hhrf/phfr</td>
<td>Durch die Hinzunahme der Screening Stichprobe 2011 wurde eine Unterscheidung der Hochrechnungsfaktoren für die Screening Stichproben nötig. Daher gibt es jetzt die Variablen &quot;f10phrf_sc10&quot; (Screening 2010), &quot;f10phrf_ch10&quot; (Kohorte) und &quot;f11phrf_sc11&quot; (Screening 2011). Entsprechend auch für die Variablen &quot;hhrf_sc10&quot;, &quot;hhrf_ch10&quot; und &quot;hhrf_sc11&quot;.</td>
<td>2.0</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Neue Gewichte</td>
<td>In dieser Version der Datenweitergabe gibt es erneut Gewichte, die eine gemeinsame Nutzung von FiD und SOEP ermöglichen. Sie finden sich in den Datensätzen hhrf_fidsoep und phrf_fidssoep. Es werden darüber hinaus Querschnittsgewichte für 2011 für die FiD-Stichproben bereitgestellt, die eine Analyse der FiD-Daten alleine ermöglichen. Diese Hochrechnung erfolgte nach einem komplett neuen Verfahren. Die Dokumentation der Gewichte enthält nähere Details.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.2</td>
<td>bioage01-03</td>
<td>Die Variablen &quot;health1&quot;-&quot;health8&quot; wurden umbenannt: &quot;health1&quot; heißt jetzt &quot;health&quot;, und ist bezieht sich auf die Sorgen der Mutter um die Gesundheit des Kindes. &quot;health2&quot;-&quot;health6&quot; wurden zu &quot;temp1&quot;-&quot;temp5&quot;, &quot;health7&quot; wurde zu &quot;temp7&quot; und &quot;health8&quot; wurde zu &quot;temp6&quot;. Damit folgt FiD der Logik im SOEP Datensatz <em>bioagel</em> (siehe 2.1.5), der Längsschnittversion der Bioagedaten im SOEP.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.3</td>
<td>bioage10</td>
<td>Die Variablen &quot;imper1a&quot;, &quot;imper2a&quot; und &quot;imper3a&quot; hatten bisher Werte von 3, die eigentlich als &quot;-2 Does not apply&quot; hätten kodiert werden müssen. Dies wurde jetzt korrigiert.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.4</td>
<td>bioage10</td>
<td>In sämtlichen Variablen &quot;imperXa&quot;-&quot;imperXe&quot; waren bisher Werte auf &quot;-2 Does not apply&quot;, auch wenn es sich um eine sogenannte &quot;gesammelte fehlende Angabe&quot; handelte, die Person also die Antwort verweigert hatte. Dies war in &quot;imperXf&quot; als &quot;-1 No answer&quot; festgehalten. Jetzt wurden bei diesen Fällen sämtliche Antworten auf &quot;-1 No answer&quot; gesetzt.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.5</td>
<td>bioagel</td>
<td>Dieser Datensatz wurde neu in die Weitergabe aufgenommen. Er enthält die kombinierte Fassung der <em>bioage01-bioage10</em> Datensätze. Variablen werden hier ohne ihren prefix (bXX) angegeben, so dass Variablen, die über die Fragebögen hinweg identisch erhoben wurden, direkt für ein Kind verglichen werden können.</td>
<td>2.1</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>2.1.6</td>
<td>biobirth</td>
<td>Der Datensatz biobirth wurde komplett neu aufgesetzt. Dadurch kommt es zu Veränderungen, hauptsächlich gibt es zahlreiche neue Fälle, weil erstmals auch nicht-teilnehmende Personen berücksichtigt wurden. Es wurde außerdem die Variable &quot;kidhome&quot; entfernt, weil sie zeitveränderlich ist; ob ein Kind im Haushalt ist, lässt sich über seine Personennummer und den jahresspezifischen pbrutto Datensatz anspielen. Außerdem wird jetzt zur höheren Genauigkeit die Variable &quot;kidsource&quot; einzeln für jedes Kind aufgeführt.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.7</td>
<td>biocouply/biomarsy</td>
<td>Durch leichte Veränderungen in der Kodierung (die SOEP-Änderungen folgen) ist es zu einer Erhöhung der Fallzahlen um rund 1000 Spells in beiden Datensätzen gekommen. Dadurch haben sich auch manche Variablen geändert.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.8</td>
<td>f10pgen/ f11pgen</td>
<td>Die Kodierung der variablen &quot;mps$$&quot; war fehlerhaft und wurde jetzt korrigiert.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.9</td>
<td>f10pgen/ f11pgen</td>
<td>In der Weitergabeversion 2.0 waren die Variablenlabels der Variablen &quot;partp$$&quot; und &quot;partno$$&quot; falsch und wurden geändert.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.10</td>
<td>f10pgen/ f11pgen</td>
<td>In der Variable &quot;coupst10&quot; (Partner Status) waren 4 Fälle mit gleichgeschlechtlicher Partnerschaft falsch zugeordnet, in &quot;coupst11&quot; betraf dies 5 Fälle.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.11</td>
<td>f10pgen/ f11pgen</td>
<td>Durch die Veränderungen in biocouply (siehe 2.1.7) hat sich für gut 80 Fälle die Variable &quot;marrst&quot; von &quot;3 single&quot; auf &quot;4 geschieden&quot; verändert.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.12</td>
<td>f10pgen/ f11pgen</td>
<td>Die Variable &quot;tenure10&quot;/&quot;tenure11&quot; (Betriebszugehörigkeit) wurde in wenigen Fällen geändert. Fälle mit unbekanntem Monat des Arbeitsbeginns hatten den Wert &quot;.1&quot; und bekommen nun einen Wert, sofern das Jahr des Arbeitsbeginns bekannt ist. Der entsprechende Monat wurde dabei zufällig besetzt.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.13</td>
<td>f10pkal/ f11pkal</td>
<td>Bei den Variablen &quot;f10p2n01&quot; und &quot;f11p2n01&quot; (keine dieser Einkunftsarten im Vorjahr) wurden die Werte nach SOEP-Logik umgesetzt: &quot;-2 Trifft nicht zu&quot; wird gesetzt wenn eine Einkommensart vorliegt.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.14</td>
<td>f10pkal/ f11pkal</td>
<td>In den Variablen &quot;p1a01&quot;-&quot;p1n01&quot; waren die Werte &quot;1 Yes&quot; und &quot;2 No&quot; vertauscht, was in der neuen Version behoben wurde.</td>
<td>2.1</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>2.1.15</td>
<td>f11h</td>
<td>Die Variable &quot;f10h034b&quot; (Absetzen von Verlusten im Vorjahr) war mit &quot;f10&quot; falsch benannt und wurde jetzt korrigiert auf &quot;f11h034b&quot;.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.16</td>
<td>f11hbrutto</td>
<td>Die Variable &quot;f11hhnrold&quot; (HH-Nummer im Vorjahr) wurde dem Datensatz jetzt hinzugefügt. Bisher war diese Information ausschließlich im Datensatz hpfad enthalten.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.17</td>
<td>f11pbrutto</td>
<td>Die Variable &quot;f11pnrold&quot; (laufende Personennummer im Vorjahr) wurde jetzt auf maximal 2 Stellen reduziert und ist so mit der aktuellen laufenden Nummer vergleichbar. (Dies betrifft in keiner Weise die unveränderliche Personennummer, &quot;persnr&quot;.)</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.18</td>
<td>hhrf / phfr / hhrf_fidsoep / phfr_fidsoep</td>
<td>Zur Verbesserung der Übersichtlichkeit wird ab Version 2.1 darauf verzichtet, spezifische Hochrechnungsfaktoren für die Kohorten- bzw. Screening-Stichprobe bereitzustellen. Bei Bedarf können sie auf individuelle Nachfrage zur Verfügung gestellt werden.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.19</td>
<td>ppfad</td>
<td>Durch die Veränderungen in biobirth (siehe 2.1.6), denen verbesserte Zuordnungen von Kindern zu ihren Eltern zugrunde liegen, kommt es bei den Variablen zum Migrationshintergrund (&quot;germborn&quot;, &quot;immiyear&quot;, &quot;corigin&quot;, &quot;migback&quot; und &quot;miginfo&quot;) zu Unterschieden zur Vorversion. Diese beziehen sich ausschließlich auf Kinder und nicht-teilnehmende Personen.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.20</td>
<td>f10paradata / f11paradata</td>
<td>In den Datensätzen zur Interviewsituation wurde eine neue Variable hinzugefügt, die den Wochentag der Befragung angibt (f$dow).</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.21</td>
<td>f10p/f11p</td>
<td>Die kodierten Berufe und Branchen, die bisher in f10pgen und f11pgen vorlagen, wurden nun auch in f10p und f11p hinzugefügt (Variablen &quot;f10p021_is88&quot;, &quot;f10p021_klas&quot;, &quot;f10p060_is88n&quot;, &quot;f10p060_klasn&quot;, &quot;f10p026_nace&quot;, &quot;f11p022_is88&quot;, &quot;f11p022_klas&quot;, &quot;f11p060_is88n&quot;, &quot;f11p060_klasn&quot;, &quot;f11p027_nace&quot;).</td>
<td>2.1</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>2.1.22</td>
<td>f10p/f11p</td>
<td>Die Variablen zum Einkommen im Vorjahr (&quot;f10p071&quot; und &quot;f11p075&quot;) wurden an die SOEP Benennung angepasst. Dabei wurden die Buchstaben, die das Einkommen kennzeichnen, geändert. Unter anderem erhalten die FiD-spezifischen Variablen zu Kindesunterhalt (jetzt &quot;u&quot; statt &quot;i&quot;), Betreuungsunterhalt (&quot;v&quot; statt &quot;j&quot;) und nachehelicher Unterhalt (&quot;w&quot; statt &quot;k&quot;) neue Zuordnungen.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.23</td>
<td>f10p/f11p</td>
<td>Die Variablen zu Sondervergütungen (&quot;f10p072&quot; und &quot;f11p076&quot; wurden jetzt nach der FiD-Logik benannt, d.h. mit Buchstaben für das Item sowie Zahl für Antworten innerhalb des Items.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.24</td>
<td>mehrere Datensätze</td>
<td>Die Variable, die angibt, durch welchen Interviewer das Interview durchgeführt wurde, wurde in allen Datensätzen, sofern die Information vorkommt, einheitlich in &quot;intid&quot; umbenannt. Diese Änderung entspricht der neuen Benennung im SOEP, Version 28.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.25</td>
<td>f10kind/f11kind</td>
<td>Die Veränderungen in biobirth (siehe 2.1.6) führen zu einem Update der Elternzeiger, also der Variablen &quot;mothn$$&quot;, &quot;mothp$$&quot;, &quot;fathn$$&quot; sowie &quot;fathp$$&quot;.</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1.27</td>
<td>f10jugend / f11jugend</td>
<td>Die Variablen &quot;f$j037g&quot;, &quot;f$j037h&quot;, &quot;f$j037i&quot; wurden hinzugefügt. Sie geben das Leistungsniveau von Gesamtschülern in den Hauptfächern an.</td>
<td>2.1</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>3.0.1</td>
<td>Alle Elterndatensätze aller Wellen</td>
<td>Die Variable zum Interview-Mode wurde in 2011 von &quot;f11pinta&quot; in &quot;f11e#inta&quot; umbenannt. 2010 wurde sie hinzugefügt (&quot;f10e#inta&quot;).</td>
<td>3.0</td>
</tr>
<tr>
<td>3.0.2</td>
<td>f12eltern6</td>
<td>Die Fragen zum Kind (Stärken und Schwächen, SDQ) wurden im Fragenbogen Eltern 6 randomisiert mit verschiedenen Skalen (7-Punkt und 3-Punkt) abgefragt (siehe auch 3.0.4). Darüber hinaus haben &quot;f12e6a23a&quot;-&quot;f12e6a23y&quot; (3-Punkt-Skala) sowie &quot;f12e6b23a&quot;-&quot;f12e6b23y&quot; (7-Punkt-Skala) ab 2012 den vollen Umfang von 25 Variablen. In 2010 und 2011 wurden im Elternfragebogen 6 nur 18 Variablen (&quot;f11e622a&quot;-&quot;f11e622r&quot;) abgefragt. Die neuen Variablen in <strong>f12e6</strong> sind benannt als &quot;f12e6#23s&quot;-&quot;f12e6#23y&quot;, was nicht der Reihenfolge im Fragebogen entspricht.</td>
<td>3.0</td>
</tr>
<tr>
<td>3.0.3</td>
<td>f12eltern6</td>
<td><strong>f12e6</strong> ist die Frage zu den schulischen Aktivitäten des Kindes außerhalb des Unterrichts (&quot;f12e614a&quot;-&quot;f12e614d&quot;) neu. Daher wurden die folgenden Fragen nach hinten verschoben. Zudem wurden die Fragen zu den Freizeitbeschäftigungen des Kindes (&quot;f12e615a&quot;-&quot;f12e615n&quot;) und zum besonderen pädagogischen Förderbedarf (&quot;f12e616a&quot;-&quot;f12e616i&quot;) vertauscht. Die Fragen &quot;f12e615a&quot;-&quot;f12e615n&quot; entsprechen demnach &quot;f11e618a&quot;-&quot;f11e618n&quot; und die Fragen &quot;f12e616a&quot;-&quot;f12e616i&quot; entsprechen &quot;f11e614a&quot;-&quot;f11e614i&quot;.</td>
<td>3.0</td>
</tr>
<tr>
<td>3.0.4</td>
<td>bioage10</td>
<td>Die Fragen zum Kind (Stärken und Schwächen, SDQ) wurden im Fragenbogen Eltern 6 mit verschiedenen Skalen (7-Punkt und 3-Punkt) abgefragt, um die international übliche 3-Punkt Skala auch für die vorherigen Jahre anwenden zu können. Entsprechend haben sich für alle Jahre die BEHAV Variablen auf die 3-Punkt-Skala reduziert. Nähere Information zum Verfahren finden sich in der Dokumentation zu <strong>bioage10</strong>. Diese Änderung findet sich auch im Längsschnittdatensatz <strong>bioagel</strong>.</td>
<td>3.0</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>--------------</td>
<td>------------</td>
</tr>
</tbody>
</table>
| 3.0.5  | f12p      | Es gibt im Datensatz *f12p* (Personenfragebogen 2012) einige Personen, die vorher noch den gesamten Biographiefragebogen (oder einen Teile) beantworten. Folgende Variablen liegen für diese Personen im *f12lela* vor, während sie im *f12p* den Wert "-6 Sample-spez. Filter" annehmen:  
<p>|          | <em>f12lela</em> | <em>f12p</em> |          |
|        | &quot;f12l002&quot; | &quot;f12p122&quot; |          |
|        | &quot;f12l006&quot; | &quot;f12p125&quot; |          |
|        | &quot;f12l007&quot; | &quot;f12p129&quot; |          |
|        | &quot;f12l008&quot; | &quot;f12p126&quot; |          |
|        | &quot;f12l009&quot; | &quot;f12p127&quot; |          |
|        | &quot;f12l045&quot; | &quot;f12p137&quot; |          |
|        | &quot;f12l046&quot; | &quot;f12p138&quot; |          |
|        | &quot;f12l061&quot; | &quot;f12p116&quot; |          |
|        | &quot;f12l063&quot; | &quot;f12p117&quot; |          |
|        | &quot;f12l064&quot; | &quot;f12p118&quot; |          |
|        | &quot;f12l065&quot; | &quot;f12p119&quot; |          |
|        | &quot;f12l066&quot; | &quot;f12p120&quot; |          |
| 3.0.6  | f11luecke | In dieser Welle wird erstmals die Datensatz <em>f11luecke</em> weitergegeben, der für diejenigen Personen, die in einem Jahr nicht teilgenommen haben, eine Kurzinformation über dieses Jahr erhoben wird. Hier werden Informationen über das Jahr 2011 erhoben, und die Daten daher im Ordner „2011“ abgelegt. | 3.0 |
| 3.0.7  | f12pbrutto / f12kind | Analog zum SOEP wurde in FiD die variable &quot;$stell&quot; (bzw. &quot;$kstell&quot; in <em>$kind</em>) in 2012 umgestellt, um die Beziehungen der Personen im Haushalt zum Haushaltsvorstand besser erfassen zu können als bisher. Sämtliche Codes der Vorwellen sind direkt abbildbar, eine Tabelle zur Überführung findet sich in der Dokumentation zu <em>$kind</em>. | 3.0 |
| 3.0.8  | $pgen    | Die Werte zur Berufserfahrung (&quot;expft$$&quot;, &quot;exppt$$&quot;, &quot;expue$$&quot;) wurden neu codiert, was zu Änderungen in den 2010 gezogenen Stichproben führte. Außerdem liegen nun erstmals Informationen für die 2011 gezogenen Screening Stichprobe vor. | 3.0 |
| 3.0.9  | $pgen    | Die Berechnung der Variable &quot;tenure$$&quot; (Betriebszugehörigkeit) wurde neu codiert, was zu leichten Veränderungen führt. | 3.0 |</p>
<table>
<thead>
<tr>
<th>Nummer</th>
<th>Datensatz</th>
<th>Beschreibung</th>
<th>in Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0.10</td>
<td>$pgen</td>
<td>Die Variablen rund um die Berufsklassifizierung (&quot;class$$&quot;, &quot;is88$$&quot;, &quot;nace$$&quot;, &quot;isei$$&quot;, &quot;egp$$&quot;, &quot;siops$$&quot;, &quot;mps$$&quot;) werden ab dieser Welle mit den Werten der Vorwelle ersetzt, wenn diese vorliegen. Aufgrund der Filterführung werden diese Variablen nur alle zwei Jahre erfragt, sofern kein Berufswechsel angegeben wurde.</td>
<td>3.0</td>
</tr>
<tr>
<td>3.0.12</td>
<td>f12pkal</td>
<td>Die Variable &quot;f12p2m03&quot; ist ausschließlich mit missings besetzt und wird auf &quot;.-2&quot; gesetzt. Die entsprechende Variable im P-Fragebogen &quot;f12p084o3&quot; wurde in 2012 nicht erhoben. Möglichkeit zur Nachbesserung besteht über Frage &quot;f12p088&quot;.</td>
<td>3.0</td>
</tr>
<tr>
<td>3.0.13</td>
<td>f10mihinc / f11mihinc</td>
<td>Neu berechnete Imputationen, die nun auch die in 2012 erhobenen Daten berücksichtigen. Mit den neu berechneten Imputationen ergeben sich auch neue Werte für &quot;i_hinc&quot; in $hgen, so dass ausschließlich die neue Distribution genutzt werden sollte.</td>
<td>3.0</td>
</tr>
<tr>
<td>3.0.14</td>
<td>f10mipinc / f11mipinc</td>
<td>Neu berechnete Imputationen, die nun auch die in 2012 erhobenen Daten berücksichtigen. Mit den neu berechneten Imputationen ergeben sich auch neue Werte für &quot;labgro10&quot; und &quot;labnet10&quot; in $pgen, so dass ausschließlich die neue Distribution genutzt werden sollte.</td>
<td>3.0</td>
</tr>
<tr>
<td>3.0.15</td>
<td>f12paradata</td>
<td>Mit der Aufnahme der $luecke Datensätze (siehe auch 3.0.6) gibt es einen zusätzlichen Code in der Variable &quot;qstnr&quot;, und zwar &quot;10 Luecke-Qunaire (Gap in prev. year)&quot;. Obwohl der Datensatz $luecke inhaltlich zur Vorwelle gehört, wird er in $paradata in der Erhebungswelle (hier f12) abgelegt.</td>
<td>3.0</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>3.1.2</td>
<td>bioage files</td>
<td>Im Zuge der Angleichung an das SOEP wurden mehrere Variablen in allen <em>bioage</em> Files umbenannt: PERSNRM -&gt; PERSNRRESP (persnr der ausfüllenden Person) SVYYEAR &gt; SYEAR (Jahr der Befragung) NMEDAID -&gt; MEDAID3M (Ärztliche Hilfe in den letzten 3 Monaten) HOSPITAL -&gt; HOSPITAL12M (Krankenhausaufenthalte in den letzten 12 Monaten) Andere Änderungen beziehen sich auf einzelne Datensätze (siehe 3.1.3-3.1.7).</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.3</td>
<td>bioage01</td>
<td>In den Variablen &quot;B01PREGMO&quot; (Schwangerschaftsmonat der Mutter bei Interview im Vorjahr) sowie PREGMY kam es durch einen Fehler in der Kodierung zu falschen Werten in 83 Fällen aus 2012. Dieser Fehler wurde nun behoben.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.4</td>
<td>bioage01, bioage03, bioage03</td>
<td>Im Zuge der Angleichung an das SOEP wurde die Variable MEDAID3M (ärztliche Hilfe in den ersten 3 Monaten nach der Geburt) in MEDAID3MB umbenannt. Nutzer sollten darauf achten, diese Variable nicht mit MEDAID3M (ehemals NMEDAID) zu verwechseln.</td>
<td>3.1</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>3.1.5</td>
<td>bioage08, bioage10</td>
<td>Bisher wurde in diesen beiden bioage files die Variable ALLOWNCE (Taschengeld im Monat) nur auf Monatsebene weitergegeben. In Übereinstimmung mit dem SOEP wird nun auch das wöchentliche Taschengeld berechnet. Nun gibt es zwei Variablen: ALLOWPM (Taschengeld pro Monat) und ALLOWPW (Taschengeld pro Woche).</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.6</td>
<td>bioage08, bioage10</td>
<td>In Übereinstimmung mit dem SOEP wurden die Variablen CURSCOL5 (noch nicht in der Schule) durch SCLENROLN, STSCOLYR durch SCLENROLY, und STSCOLMN durch SCLENROLM ersetzt. Außerdem wurden die Variable CURSCOL4 (sonstige Schule) in CURSCOL8 umbenannt, und die Variablen CURSCOL4 (Hauptschule), CURSCOL5 (Realschule), CURSCOL6 (Gymnasium) und CURSCOL7 (Gesamtschule) hinzugefügt, wobei letztere (anders als beim SOEP) aus dem Haushaltsfragebogen gewonnen werden.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.7</td>
<td>bioage10</td>
<td>In Übereinstimmung mit dem SOEP wurde die Variable CONSCHO5 (Kein Kontakt mit der Schule) in CONSCHO7 umbenannt. Außerdem wurden die Variablen ALLOW (Taschengeld) und NOMARK (keine Noten) hinzugefügt.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.8</td>
<td>biojob</td>
<td>Aufgrund von Unklarheiten bei der Generierung des Files biojob wurde er aus der Weitergabe 3.1 herausgenommen. Auf Anfrage kann er Nutzern zur Verfügung gestellt werden.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.9</td>
<td>f10jugend</td>
<td>Die Werte der Variablen F10J015I, F10J015J und F10J015O waren untereinander vertauscht und wurden nun korrigiert.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.11</td>
<td>f11hgen</td>
<td>Bei der Berechnung der Variablen AHINC11 wurde bei 36 Fällen statt des Witwenpensions eine falsche Komponente addiert. Die Unterschiede sind für diese Fälle gering.</td>
<td>3.1</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>3.1.12</td>
<td>f12hgen</td>
<td>Bei der Berechnung der Variablen AHINC12 kann im Gegensatz zu den Vorjahren nicht mehr auf die Pensionen zurückgegriffen werden, weil diese Frage nicht mehr gestellt wurde. Entsprechend erfolgt die Berechnung von AHINC12 nun ohne diese Komponente. Irrtümerlicherweise war in der Version 3.0 der Fehler aus 2011 (siehe 3.1.11) wiederholt worden, so dass nun 34 Fälle korrigiert wurden.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.13</td>
<td>f12eltern2</td>
<td>Für die Person mit der PERSNR 21504004 war in der Weitergabe 3.0 in der Variable F12E205G irrtümlich der Wert &quot;6&quot; angegeben. Dieser wurde nun auf &quot;-1 Keine Angabe&quot; umgesetzt.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.14</td>
<td>$kind</td>
<td>Durch Verbesserungen bei der Partnerdefinition kann es vereinzelt zu Umcodierungen bei den Variablen MOTHPP$ und FATHPP$ (Indikator für das Verhältnis zwischen Kind und Eltern) kommen, die die Information &quot;6 - unbekannt&quot; revidieren.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.15</td>
<td>f12kind</td>
<td>Folgende Variablen wurden in der Weitergabe 3.0 aufgrund von Verschiebungen im Fragebogen falsch benannt und nun korrigiert: F12K056 -&gt; F12K061, F12K057a, b -&gt; F12K062a, b, F12K058 -&gt; F12K063, F12K059 -&gt; F12K064, F12K060 -&gt; F12K065, F12K061a, b -&gt; F12K066a, b, F12K062 -&gt; F12K067, F12K063 -&gt; F12K068, F12K064a, b -&gt; F12K069a, b, F12K065a, b -&gt; F12K070a, b, F12K066a, b -&gt; F12K071a, b, F12K067 -&gt; F12K072, F12K068a, b -&gt; F12K073a, b, F12K069a, b -&gt; F12K074a, b, F12K070a, b, c, d, e -&gt; F12K075a, b, c, d, e, F12K071a, b -&gt; F12K076a, b.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.16</td>
<td>f12kind</td>
<td>In Version 3.0 der Daten waren die neuen Items in Frage 77 der Kindermatrix nicht in ihrer Reihenfolge im Haushaltsfragebogen enthalten. Dies wurde nun korrigiert, so dass sich neue Zuordnungen der Items zu Variablen ergeben. Außerdem wurden hier die Labels der Items so geändert, dass sie die vier Gruppen (bis 6, Kita ja/nein und ab 6, Schule ja/nein) widerspiegeln.</td>
<td>3.1</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>3.1.17</td>
<td>f12lela</td>
<td>Die Geburtsländer der Eltern (F12L078A2, F12L078B2) waren in der Version 3.0 nicht mit den jeweiligen Codes besetzt. Dieser Fehler wurde behoben, so dass an den Labels das Geburtsland der Eltern abzulesen ist.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.18</td>
<td>f12lela</td>
<td>Folgende Umbenennungen wurden vorgenommen, um Konsistenz mit den vorherigen Datensätzen (f10lela, f11lela) herzustellen: F12L001A -&gt; F12L001C (Geschlecht) F12L001B -&gt; F12L001A (Geburtsjahr) F12L001C -&gt; F12L001B (Geburtsmonat)</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.19</td>
<td>$p, $pgen</td>
<td>Durch Veränderungen in den Berufsverkodungen durch TNS Infratest verändern sich die entsprechenden Variablen (ISCO, KLAS, NACE, etc.) in $p und $pgen. Dies betrifft insgesamt nur wenige Fälle.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.20</td>
<td>f12p</td>
<td>Bereits in der gleichen Welle im Lebenslauf erhobene Variablen werden im Personenfragebogen überfiltert. Bisher wurde bei betroffenen Personen der Missing-Code &quot;-6&quot; vergeben und die entsprechende Information lag nur in f12lela vor. Seit Version 3.1 werden diese Informationen zusätzlich direkt in den entsprechenden Variablen des f12p-Files abgelegt. Änderungen treten dabei in 297 Fällen auf.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.21</td>
<td>f12p</td>
<td>Die Berufsverkodungen (F12P023_IS88, F12P023_KLAS und F12P028_NACE), die vorher nur im f12pgen verfügbar waren, wurden in der Version 3.1 hinzugefügt.</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.22</td>
<td>ppfad</td>
<td>Bei 168 Fällen wurde in der Weitergabew 3.0 bei der Variable GEBMONAT nicht auf die neueste Information (aus 2012) zurückgegriffen. Dies wurde in dieser Weitergabe korrigiert und betrifft auch 147 Fälle in der Variable GEBMOVAL.</td>
<td>3.1</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>3.1.24</td>
<td>$pgen</td>
<td>Aufgrund von Änderungen in der Berechnung von COUPST und MARRST11 kommt es in allen Jahren zu leichten Veränderungen. (Die Variable COUPI D wird neu vergeben und ist entsprechend nicht notwendigerweise identisch über die Weitergaben.)</td>
<td>3.1</td>
</tr>
<tr>
<td>3.1.25</td>
<td>$pgen</td>
<td>Im SOEP werden seit 2011 neue Bildungsvariablen weitergegeben, hierbei handelt es sich um: FIELD$$ (Fachrichtung) DEGREE$$ (Hochschulabschluss) TRAIA$$ (Ausbildungsberuf, Lehre) TRAINB$$ (Ausbildungsberuf, Berufsfachschule) TRAINC$$ (Ausbildungsberuf, Fachschule) TRAIND$$ (Ausbildungsberuf, Beamtenausbildung) FDT_F$$ (Quelle FIELD, DEGREE, TRAIN) Mit der Weitergabe 3.1 von FiD sind diese Variablen (auch rückwirkend) in FiD enthalten. (Die Variablenlabels sind dabei in deutscher Sprache.)</td>
<td>3.1</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>3.1.26</td>
<td>$paradata</td>
<td>Die Datensätze $paradata werden ab dieser Weitergabe nur noch im long-Format als <em>paradatal</em> weitergegeben. Dadurch fallen die Wellenkürzel (f10, f11, f12) bei den Variablen weg und die Variable SYEAR (Befragungsjahr) wird aufgenommen. Außerdem wird der Datensatz um zwei Variablen erweitert: REQD, die Anzahl der zu beantwortenden Fragen und MISS, die Anzahl der verweigerten oder nicht-validen Antworten, jeweils auf Interviewebene.</td>
<td>3.1</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>4.0.1</td>
<td>f13eltern1-3</td>
<td>Die Frage, ob das Kind regelmäßig oder zeitweise durch eine Tagesmutter oder KiTa betreut wird, wurde um die Anzahl der Stunden pro Tag erweitert (F13E1012, F13E218, F13E318). Dadurch wurde in der späteren Abfrage nach der Betreuungssituation auf Tagesmutter und KiTa verzichtet (F13E129, F13E225, F13E325). Siehe auch Punkt 4.0.6.</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.2</td>
<td>f13eltern1-4</td>
<td>Es wurden hier neue Fragen eingeführt, in denen die Eltern die Eingewöhnungsphase ihrer Kinder in die KiTa bewerten sollen. Dies führt auch zu neuen Variablen in <strong>bioage01, bioage02, bioage03, und bioage06</strong> (siehe Punkt 4.0.7).</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.3</td>
<td>f13eltern2-3</td>
<td>Die Fragen zur Geburt des Kindes, die in den vorherigen Elternfragebögen 2 und 3 noch gestellt wurden, werden für 2013 nicht mehr erhoben - für fast alle Kinder liegen diese Information aus den vorherigen Jahren vor. Siehe auch Punkt 4.0.8.</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.4</td>
<td>f13eltern2-4</td>
<td>Die Aktivitäten, die mit dem Kind gemacht werden (F13E227, F13E327, F13E412), wurden um den Skalenpunkt &quot;4 seltener&quot; erweitert, um die bisherige Lücke zwischen &quot;einmal die Woche&quot; und &quot;gar nicht&quot; zu überbrücken. Dadurch verschiebt sich in dieser Frage abweichend zu vorher &quot;gar nicht&quot; auf den Wert &quot;5&quot;. Siehe auch Punkt 4.0.9.</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.5</td>
<td>f13eltern5-6</td>
<td>Es wurden hier neue Fragen eingeführt, in denen die Eltern die Eingewöhnungsphase ihrer Kinder in die Schule bewerten sollen. Dies führt auch zu neuen Variablen in <strong>bioage08 und bioage10</strong> (siehe Punkt 4.0.10).</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.6</td>
<td>bioage01, bioage02, bioage03</td>
<td>Durch die Umstellung der Frage nach der Betreuung durch KiTa und Tagesmutter in 2013 (siehe Punkt 4.0.1) ändert sich auch die Kodierung für CARE6, CARE6H, CARE8, CARE8H. Hierbei werden die Angaben zur Betreuung pro Tag auf die Woche umgerechnet. Näheres dazu in der Dokumentation zu den <strong>bioage files</strong> (6_bioage01-10.pdf).</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.7</td>
<td>bioage01, bioage02, bioage03, bioage06</td>
<td>Erstmals wurde über die Probleme bei der Eingewöhnung in die KiTa gefragt, diese Variablen heißen ADPCCAR1-ADPCCAR4. Sie sind ausschließlich im Jahr 2013 besetzt und haben sonst den Wert &quot;.5&quot;.</td>
<td>4.0</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>4.0.8</td>
<td>bioage02, bioage03</td>
<td>Da die Information zur Geburt des Kindes ab 2013 nicht mehr in den jeweiligen Elternfragebögen erhoben wird, werden die entsprechenden Variablen für dieses Jahr auf &quot;-5 Question not included&quot; gesetzt. (Die Information liegt in den meisten Fällen aus den vorherigen Jahren vor.)</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.9</td>
<td>bioage02, bioage03, bioage04</td>
<td>In den Variablen ACTIV1-ACTIV14 wurden durch die Änderungen in f13eltern2-4 (siehe Punkt 4.0.4) ebenfalls die Skalen geändert. Der Wert 4 (bis Version 3.1 &quot;Never&quot;) wurde nun zu 5 umkodiert. Neu eingeführt wurde der Wert &quot;4 Less often&quot;, der nur in 2013 besetzt ist.</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.10</td>
<td>bioage08, bioage10</td>
<td>Erstmals wurde über die Probleme bei der Eingewöhnung in die Schule gefragt, diese Variablen heißen ADPSCL1-ADPSCL5. Sie sind ausschließlich im Jahr 2013 besetzt und haben sonst den Wert &quot;-5&quot;.</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.11</td>
<td>$jugend</td>
<td>Die in den Jugendfragebogen abgefragten Klartexte zum eigenen Berufswunsch sowie zu den letzten Berufen der Eltern werden nun als ISCO-Codes und Klassifizierung des Statistischen Bundesamts bereitgestellt.</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.12</td>
<td>f13jugend</td>
<td>In Frage 15 wurde ein neues Item eingeführt (Nutzen sozialer Online-Netzwerke). Um die Längsschnittkonsistenz zu wahren, wird dieses Item entgegen den Regeln für die Variablenbenennung als &quot;F13j015Q&quot; eingefügt.</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.13</td>
<td>$lela</td>
<td>Die im 2. Teil des Biographiefragebogens als Klartexte erfassten Klassifikationen des ersten Jobs bzw. der letzten Branche werden nun kodiert weitergegeben.</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.15</td>
<td>$mipinc / $mihinc</td>
<td>Neu berechnete Imputationen, die nun auch die in 2013 erhobenen Daten berücksichtigen. Mit den neu berechneten Imputationen ergeben sich auch neue Werte für &quot;LABGRO$&quot; und &quot;LABNET$&quot; in $pgen, sowie neue Werte für imitierte Variablen in $hgen, so dass ausschließlich die neue Distribution genutzt werden sollte. Es gibt außerdem leichte Veränderungen in der Art der Imputation; genaueres findet sich in der entsprechenden Dokumentation.</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.16</td>
<td>f11hgen, f12hgen</td>
<td>Die Variable MOVEYR wurde neu kodiert, nachdem vorher Umzüge von einigen Personen nicht korrekt berücksichtigt worden waren.</td>
<td>4.0</td>
</tr>
<tr>
<td>Nummer</td>
<td>Datensatz</td>
<td>Beschreibung</td>
<td>in Version</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>4.0.17</td>
<td>ppfad</td>
<td>Die Nettocodes in $NETTO wurden überarbeitet und wie im SOEP ergänzt: 180 (Person ohne aktuelle Angabe ohne Austritt), 181 (Vormals Befragte ohne aktuelle Angabe), 188 (Rückkehrer, zuvor Ausland), 189 (Rückkehrer, zuvor Ausfall) sind nun besetzt.</td>
<td>4.0</td>
</tr>
<tr>
<td>4.0.18</td>
<td>ppfad</td>
<td>Die Variable LOC1989 (Wohnort Ost oder West 1989) ist nun auf &quot;-1 Keine Angabe&quot; gesetzt, wenn der entsprechende Teil der Biographie nie erfragt wurde und das Geburtsjahr der Person vor 1989 liegt. Vorher war hier eine &quot;-2 Trifft nicht zu&quot; kodiert worden.</td>
<td>4.0</td>
</tr>
</tbody>
</table>