



pairfam Data Manual

Release 5.0

Revised since Release 4.0, May 2013

Josef Brüderl, Simone Braun, Kristin Hajek, Michel Herzig,
Bernadette Huyer-May, Volker Ludwig, Bettina Müller,
Ulrike Müller, Jasmin Passet, Klaus Pforr, Mirte Scholten,
Philipp Schütze, Nina Schumann

University of Munich (LMU)

With contributions from:

Petra Buhr, Stefan Fiedrich, Christina Gschwendtner, Daniela
Klaus, Nadia Lois, Timo Peter, Veronika Salzburger, Markus Schaer,
Franziska Schmahl, Carolin Thönnissen, Barbara Wilhelm

May 15, 2014

Funded as long-term project by the German Research Foundation (DFG)

Contents

1	Introduction	3
1.1	Obtaining the data	3
1.2	Referencing the pairfam project	4
1.3	Quick Start Files	5
2	Overview of the data structure	6
2.1	Survey instruments and data sets	6
2.2	System of person identifiers	7
2.3	Variable names	8
3	DemoDiff	9
3.1	Introduction	9
3.2	Overview and data structure	9
3.3	Differences in detail	10
3.4	Codebook	10
4	Anchor data	11
4.1	Data editing	11
4.2	Generated variables and scales	23
4.3	Process-generated variables	46
4.4	Paradata	47
4.5	Weights	48
4.6	Generated data sets	51
5	Partner data	88
5.1	Data editing	88
5.2	Generated variables and scales	89
6	Parent data	92
6.1	Data editing	92
6.2	Generated variables and scales	94
7	Child data	98
7.1	Data editing	98
7.2	Generated variables and scales	100
8	Parenting data	103
8.1	Data editing	103
8.2	Generated variables and scales	104
9	Outlook	106

A	Appendix	109
A.1	DemoDiff-pairfam differences	109
A.2	Changes between releases	115
A.3	Information on longitudinal weights	130
A.4	Flag and tag variables	136
A.5	Changes in codebooks	150

1. Introduction

The aim of this manual is to facilitate work with pairfam data sets. The document provides information on the processing of the available wave 1 to wave 5 data. If you have questions concerning the data that are not covered by this manual, please contact the pairfam user service at support@pairfam.de.

The manual is structured as follows. Chapter 2 provides an overview of the pairfam survey instruments and the resulting data sets, of person numbers identifying the survey respondents, and of the system of variable names. In subsequent chapters, the processing of each of the provided data sets is described in detail. Chapter 3 gives an overview over the DemoDiff data which started as a complementary study and is integrated into the pairfam data from release 5.0 on. In chapter 4, we first describe the data editing process for the anchor data of waves 1 to 5. The subsequent sections 4.2 to 4.6 all deal with non-survey variables that have been added to the data set to improve the usability of the data. Constructs that are used frequently have been generated and added to the data set or are available as scales documented in syntax files. These variables are listed and described in section 4.2. Data generated from the original sample are described in section 4.3. Variables generated from the field process are listed in section 4.4. Section 4.5 deals with the computation of weights for the pairfam anchor data and gives advice on how to use these weights. Chapter 4.6 describes the computation of additional data sets, covering the anchors' partner and child relationships, their educational and occupational activities, information on their dwellings, composition of households and residences as well as on the anchor's biological, adoptive and stepparents. We processed the original data and reformatted them to ease event history and panel data analysis. Chapters 5 to 8 explain the data editing as well as the computation of generated variables and scales for the alteri data sets. The manual closes with a brief outlook on future developments (chapter 9).

1.1 Obtaining the data

Pairfam data release 5.0 contains four data sets from wave 1 (*anchor1*, *anchor1_DD*, *partner1* and *partner1_DD*), six data sets from wave 2 (*anchor2*, *anchor2_DD*, *partner2*, *parent2*, *child2*, and *parenting2*), five data sets from wave 3 (*anchor3*, *partner3*, *parent3*, *child3*, and *parenting3*), five data sets from wave 4 (*anchor4*, *partner4*, *parent4*, *child4*, and *parenting4*), five data sets from wave 5 (*anchor5*, *partner5*, *parent5*, *child5*, and *parenting5*) and ten generated data sets (*biopart*, *biochild*, *bioact*, *bioact_rtr*, *household*, *biomob_ehc*, *biomob_ehc_moves*, *biomob_rtr*, *biomob_rtr_parents*, and *bioparent*).¹ Some of the files encompassing data from wave 1 to 4 have been updated with data release 5.0.² Anchor and alteri data files come with English and German labels³, whereas the generated data sets are labeled in English only.

The data are made available to registered data users only. To become a registered user, interested researchers should follow the instructions on the homepage <http://www.pairfam.de>. The data are distributed by GESIS. Information on data errors detected after the release of the data will be posted on the website. Bug fixes will be available there as well. Data users should visit the homepage from

¹We provide the data in Stata and SPSS formats.

²For details see tables A.7 to A.19.

³Stata files contain both types of labels. To switch to English labels, enter `label language en` into Stata's command prompt. (To switch back to German, type `label language de`.) SPSS users will find two versions for each file on CD. Please choose the preferred data from the directory (\English or \German).

time to time to stay abreast of problems with the data.

From wave 4 on former CAPI children enter the main survey and become anchor respondents themselves. These data are not part of the current scientific use file but will be provided to the scientific audience at a later point in time. For further information please contact the pairfam user service.

In addition to the scientific use files, it is also possible to make use of process-generated data, most importantly, information on the anchor's place of residence (local municipality level, see chapter 4.3). These data are not included in the scientific use file, however, and can only be analyzed on-site. The requirements and contact information for on-site analyses are posted on the pairfam website.

1.2 Referencing the pairfam project

Receiving credit from data users is of vital interest to the pairfam project. Only with these references can we prove the scientific value of pairfam. Therefore, we kindly ask you to add the following points to all your publications that are based on pairfam data. Please visit the pairfam website where you find the ever-updated version of our citation rules in both English and German.

a) Refer to the study and the data version (currently release 5.0) by including the following (or a similar) phrase in your text:

Analyses are based on data from the first five waves of the German Family Panel (pairfam), release 5.0 (Nauck et al. 2014). A detailed description of the pairfam study can be found in Huinink et al. (2011).

b) The respective entries in the reference section are:

Nauck, Bernhard; Brüderl, Josef; Huinink, Johannes; Walper, Sabine (2014): The German Family Panel (pairfam). GESIS Data Archive, Cologne. ZA5678 Data file Version 5.0.0, doi: 10.4232/pairfam.5678.5.0.0

Huinink, Johannes; Brüderl, Josef; Nauck, Bernhard; Walper, Sabine; Castiglioni, Laura; Feldhaus, Michael (2011): Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual framework and design. Zeitschrift für Familienforschung - Journal of Family Research 23: 77-101.

The latest version of the reference paper can be downloaded from the website at <http://www.pairfam.de>.

c) When acknowledging the project, please include the following (or a similar) phrase:

This paper uses data from the German Family Panel (pairfam), coordinated by Josef Brüderl, Karsten Hank, Johannes Huinink, Bernhard Nauck, Franz Neyer, and Sabine Walper. pairfam is funded as a long-term project by the German Research Foundation (DFG).

There is no obligation to do so, but if you find this manual at all helpful, the authors would appreciate a reference:

Brüderl, Josef; Hajek, Kristin; Huyer-May, Bernadette; Ludwig, Volker; Müller, Bettina; Müller, Ulrike; Passet, Jasmin; Pforr, Klaus; Scholten, Mirte; Schütze, Philipp; Schumann, Nina (2014): pairfam Data Manual. Release 5.0, University of Munich, Technical report.

1.3 Quick Start Files

The pairfam team developed several commented Stata do-files (Quick Starts) that should allow the user a Quick Start with analyzing the pairfam data. The Quick Starts show how common analysis problems can be solved with the pairfam data and how different pairfam data sets can be merged. Using the Quick Starts, the user can start with analyses right away. Adapting the Quick Starts will be an easy way to start with your own analyses.

The following Quick Starts, using data of release 5.0, will be available soon on the pairfam homepage in the “Data / Documentation” section:

- “Sample Definition & Weighting”: information on defining the analysis sample and examples of how to use weights in pairfam and DemoDiff,
- “Matching”: introduction of the most common operations for merging separate pairfam data sets to combined multi-waves and/or multi-actor data sets,
- “Biopart”: information on how to use the generated data set “biopart”,
- “Biochild”: information on how to use the generated data set “biochild”,
- “Panel Analysis”: demonstration of an exemplary panel analysis using the pairfam data of wave 1 to 5.

A description on how to use pairfam data and the Quick Starts for analyzing wave 1 data is included in Brüderl et al. (2011a) which is also available as Brüderl et al. (2011b). The latter can be found on the pairfam homepage in the “Data / Documentation” section.

2. Overview of the data structure

The main goal of the pairfam study is to provide researchers with data for the longitudinal analysis of family processes. To serve this goal, pairfam is designed as a *panel* and a *multi-actor study*.¹ Thus there are two main principles behind the data structure: facilitating analysis of respondents' information over time and making it possible to match information of multiple respondents connected by family ties.

The following sections introduce the pairfam survey instruments and resulting data sets, the system of person identifiers, and the system of variable names.

2.1 Survey instruments and data sets

Table 2.1 gives an overview of the survey instruments and resulting data sets. In wave 1, we conducted interviews with our main respondents (called anchors) and (if available) their current partners. As of wave 2, we collect data from further alteri of the anchors (their parents and children). Furthermore, a new instrument has been devised to collect information on the anchor's and his or her partner's parenting. In the DemoDiff study only anchor respondents and their partners were surveyed (for details on the DemoDiff study see section 3).

As a general rule, we decided to store the information from each wave as separate cross-sections. Although we do not intend to produce long format panel data sets in the future, the data structure builds on the assumption that most users prefer to analyze data of this format. We acknowledge that in statistical software packages most built-in routines for panel data analysis operate on long format data. Therefore, we keep names of variables consistent across waves to allow for easy construction of long format data from the cross-sectional files (see section 2.3).²

The other organizing principle of the data structure is storage by survey instrument, i.e., we produce separate data sets for the anchor, partner, parents, child, and parenting surveys respectively. Data set names consist of an (invariable) radical identifying the survey instrument and a suffix indicating the panel wave (\$). The first-wave data release includes the data sets *anchor1*, *partner1*, *anchor1_DD*, and *partner1_DD*. The release of wave 2 data contains *anchor2*, *anchor2_DD*, *partner2*, *parent2*, *child2*, *parenting2*. The following data sets are published for wave 3 to 5: *anchor3*, *partner3*, *parent3*, *child3*, *parenting3* resp. *anchor4*, *partner4*, *parent4*, *child4*, *parenting4* resp. *anchor5*, *partner5*, *parent5*, *child5*, *parenting5*. With each release, previously published data sets are updated if necessary.³

¹For further information concerning the conceptual framework and design see the pairfam reference paper (Huinink et al. 2011) and the pairfam Technical Paper No. 01 (Arránz Becker et al. 2014).

²There were two main reasons why we did not opt for the long format solution. The first was that we did not want to force all users to start with a large data set including information for all waves and all variables. There will be questions which are asked only once or at intervals of several years, and researchers analyzing such variables should not have to deal with the full complexity of artificially inflated data sets. Secondly, some researchers prefer to work with wide format data sets, and for them, a long format data set presents the inconvenience of having to split the data into cross-sections by themselves.

³See table A.7, which documents the changes in the data sets between release 1.0 and release 2.0, tables A.8 and A.9 for changes in the various data sets between release 2.0 and release 3.0, tables A.10, A.11 and A.12 for changes between release 3.0 and release 3.1, and tables A.13, A.14 and A.15 for changes between release 3.1 and 4.0. and tables A.16, A.17, A.18 and A.19 for changes between release 4.0 and 5.0.

Table 2.1: Overview of survey instruments and resulting data sets

Survey	Start	Instrument	Respondents	Person ID	Data set
Anchor survey	2008/09	CAPI/CASI	Anchor	id	anchor\$
Partner survey	2008/09	PAPI	Anchor's partner	pid	partner\$
Parenting survey	2009/10	PAPI	Anchor, anchor's partner	id, pid	parenting\$
Parent survey	2009/10	PAPI	Anchor's biological (or adoptive) mother, biological (or adoptive) father, stepmother, stepfather	mid, fid, smid, sfid	parent\$
Child survey	2009/10	CAPI	Anchor's child(ren) between 8 and 15 years, living with anchor	cid	child\$

2.2 System of person identifiers

Each pairfam respondent was assigned a unique and invariant identifier according to the system described in table 2.2. The anchor identifier (variable *id*) is constructed on the basis of a 3-6 digit household number (*hhid*). The household number was assigned by *TNS Infratest* in order to (re-) identify the anchor in each wave. The anchor identifier simply adds three trailing zeros to the household number. The variable *id* is contained in each alteri data set. This allows matching of alteri information not only to anchor data, but (starting in wave 2) also among any two (or more) data sets of the alteri surveys.

The anchor id forms the basis for the construction of all other person identifiers (see table 2.2). For example, the identifier of anchor's wave 1 partner (*pid*) adds the trailing number 101 instead of trailing zeros to the household number. As of wave 2, the same *pid* will be kept as long as the anchor still is in a relationship with the same partner. New partners will be numbered consecutively⁴. Information contained in data sets *anchor\$* and *partner\$* is thus easily matched via the person identifier of the anchor. Since all data sets contain the anchor's *id*, this variable can always be used for merging purposes.

Following the concept of all other person identifiers, the anchor's siblings are also assigned a unique identifier, which is part of the parent data. The parent data include up to three parents per anchor respondent, providing information about their children - the anchor plus his or her siblings. In other words, information from different parents regarding the same child is available within each wave. The siblings identifier was generated through a process of record-linkage merging (Schnell et al. 2004) based on the child's name, gender, year of birth, and status as living or deceased. As a result, identical children in a family are identified in reference to their different parents through assignment of the so-called siblings identifier (*sibidx*) with the anchor identifier number (*id*) plus a trailing number 401 for the first sibling, 402 for the second, 403 for the third, and so on. The same procedure was undertaken for children where only one parent was interviewed in wave 2. In these cases, the *sibidx* will be continued with additional information in the longitudinal perspective.

In order to match information given by the same respondent over time, a wave identifier will be included in each cross-sectional data set. The variable *wave* is assigned value "1" in data sets of the first-wave data release, value "2" in second-wave data sets, and so forth.

⁴Please note that for some anchors the partner identifiers start with the trailing number 102. This is the case if the first current partner appeared in wave 2.

Table 2.2: Person identifiers of pairfam respondents

Respondent(s)	Person ID	(Range of) assigned IDs
Anchor	id	3-6 digit household identifier (hhid) · 1000
Anchor's partner	pid	wave 1: id + 101 wave 2: id + 101 if partner is still wave 1 partner, id + 102, . . . , 150 for new partner(s) as of wave 3: id + 101 if partner is first current partner so far
Anchor's child(ren)	cid(x)	id + 201, . . . , 299 (if defined as CAPI child)
Anchor's parents	mid	id + 301 for biological (or adoptive) mother,
	fid	id + 302 for biological (or adoptive) father,
	smid	id + 303, 305, . . . , 399 for stepmother(s),
	sfid	id + 304, 306, . . . , 398 for stepfather(s)
Anchor's siblings	sibidx	id + 401, . . . , 499

2.3 Variable names

The main principle in generating names of variables in the pairfam study is to maintain consistency across waves. For questions which are asked repeatedly and which comprise the same concept, names of wave 1 variables will therefore be maintained in later waves. This principle is applied not only to the anchor data, but also to each of the constituting surveys. If categories of answer lists or items change between waves, the numbers indicating the categories/items which no longer are valid are left out and new categories/items are added to the next free position of the answer list, i.e. to the end.⁵

If equivalent information is collected in two or more of the surveys, the names of the corresponding variables will be held constant across data sets as well. In general, a prefix identifying the survey instrument is added to the names of variables in the alteri data sets. In data sets *partner\$* all variable names (except person and wave identifiers) are prefixed by a p-, while the prefix par- is used in the parents survey, and c- indicates the variables collected in the child survey.

Names of variables within the anchor data consist of a *radical* and a *suffix*.

The radical is made up of:

- an acronym** indicating the substantial area of the study to which the variable belongs, e.g. sat- for "satisfaction", and if necessary
- a running number** e.g. *sat1-*, *sat2*, . . . , *sat6* to distinguish variables within areas.

The suffix consists of any of the following (or a combination thereof):

- qualifier -i** indicating one of several items measuring the same or closely related constructs, plus a running item number, e.g. *sat1i1*, *sat1i2*, etc.;
- qualifiers -b, -d, -e, -g, -h, -k, -m, -n, -o, -p, -r, -v, -y** indicating (in order of appearance in the anchor codebook of wave 1) day (-d), month (-m), year (-y), name (-n), gender (-g), episode (-e), beginning (-b), end (-e), partner (-p), kid (-k), open answer (-o), respondent (-r), variable (-v), place (-p), household (-h), person (-p), hours (-h), minutes (-m), e.g. *doby* contains the year of the anchor's date of birth; if necessary, running numbers are added to these qualifiers that indicate, for example, relationship episodes, previous partners, or children.

⁵See, for example, variables *pa3* or *per1_* in data sets *anchor\$*.

3. DemoDiff

3.1 Introduction

In 2009/2010 the Max Planck Institute for Demographic Research (MPIDR) initiated and funded DemoDiff (Demographic Differences in Life Course Dynamics in Eastern and Western Germany) as a panel study closely following the concept of the German Family Panel pairfam (Kreyenfeld et al. 2012). In the following years, DemoDiff was gradually integrated within pairfam, resulting in full integration beginning with wave 5 in 2012/13. From wave 5 onwards, DemoDiff has been conducted by pairfam, funded by the German Research Foundation (DFG).

The main differences between DemoDiff and pairfam can be summarized as follows:

Sample	DemoDiff only samples respondents who lived in Eastern Germany (excluding West Berlin) at the time of first interview.
Multi-Actor-design	It only samples the 1971-1973 and 1981-1983 cohorts. Between wave 1 and wave 4 it only surveyed anchor respondents and their partners.
Content	There are some minor differences for wave 1 and 3 ¹ .
Organization	Between wave 1 and 4 DemoDiff was conducted and funded by the Max Planck Institute for Demographic Research (MPIDR).

3.2 Overview and data structure

As depicted in table 3.1, the first wave of pairfam was conducted in 2008/09, and the first wave of DemoDiff was launched in 2009/10. In its first round, DemoDiff closely followed the questionnaire of the first wave of pairfam. In order to synchronize both studies, the second round of DemoDiff had the same content as wave three of pairfam. Consequently, the content of pairfam’s wave 2 was not part of DemoDiff. Only a few elements, like the childhood history, were adopted from pairfam’s wave 2 in round 3 of DemoDiff. Thus, pairfam and DemoDiff are synchronized from 2010/11 onwards, meaning also that the fieldwork for each wave has since then been conducted simultaneously. A full integration finally took place in wave 5. Thus, the pairfam team now coordinates both studies, and the multi-actor design was adapted for DemoDiff cases such that child, parenting, and parent data sets now exist.

Table 3.1: Synchronization of DemoDiff with pairfam

	2008/09	2009/10	2010/11	2011/12	2012/13
pairfam	W1	W2	W3	W4	W5
DemoDiff	—	W1	W3	W4	W5

Table 3.2 gives an overview of the valid number of respondents for each wave of DemoDiff. In wave 5, 987 anchor persons could be interviewed.

¹Before release 5.0, the second round of DemoDiff was named W2/3. However, because time-frame and content of the second round of DemoDiff is comparable to pairfam’s wave 3, it has been relabeled W3.

Table 3.2: Valid interviews for DemoDiff

	Wave 1 2009/10	Wave 3 2010/11	Wave 4 2011/12	Wave 5 2012/13
Anchor	1,489	1,173	1,074	987
Partner	684	578	550	490
Child	/	/	/	233
Parent	/	/	/	371
Parenting	/	/	/	349

In terms of data sets the DemoDiff set was implemented in the following way. Because the first waves of pairfam and DemoDiff (for the anchor and the partner) took place at different points in time, information was saved in separate data sets. For wave 1, the DemoDiff cases are stored in anchor1_DD and partner1_DD. Beginning with wave 3, all pairfam and DemoDiff cases are no longer stored separately. One single exception is the childhood history. This was collected for the DemoDiff cases in wave 3, but was part of the content of pairfam's wave 2, so it was stored in a separate data set named anchor2_DD.

3.3 Differences in detail

Some differences between the pairfam and the DemoDiff sets exist before wave 5. When a question from the original pairfam questionnaire was not part of DemoDiff, the corresponding variable was set to -10 "Not in DemoDiff." If a variable was not included in pairfam, but in DemoDiff, -11 "Not in pairfam" was indicated. Content-related differences only exist in waves 1 and 3. However, some non-content-related differences in wave 4 exist. These result from the fact that DemoDiff differs in the multi-actor approach between wave 1 and 4 (no child, parent, or parenting questionnaire).

Variables that were dropped from DemoDiff

In order to adhere to the maximum interview time agreed upon with INFRATEST, some questions had to be eliminated from DemoDiff. Tables A.1 to A.4 show the list of the variables that were deleted in waves 1 and 3. Note that these tables include both content-specific differences and missing variables caused by the differences in the multi-actor approach.

Additional variables in DemoDiff

There are a few additional variables included in DemoDiff which were not included in pairfam (see Tables A.5 and A.6). In wave 1, this applies to regional information regarding the anchor person and the partner's place of birth. In wave 3, this applies to information on partnership status at first childbirth, retrospective activity and residential history, and the childhood history that was saved in anchor2_DD, as mentioned above.

3.4 Codebook

For wave 1, separate codebooks (for anchor and partner, respectively) are available for pairfam and DemoDiff. For all other waves the differences are marked in the combined codebooks.

4. Anchor data

Processing of the anchor data will be described in the next section.

4.1 Data editing

This section describes the editing of the anchor data from waves 1 to 5.¹ The main steps in processing the data are:

- label variables and values
- define missing values
- produce preload information for the CAPI interview
- clean data of EHC variables (as of wave 2)
- code open answers
- make the data anonymous
- check value ranges
- check filters
- check consistency across answers and across waves; mark data inconsistencies
- compute user-friendly variables and episode/panel data
- produce English-language data sets

Data processing was done in Stata. The data management is designed such that a master do-file consecutively calls several Stata do-files. Each do-file opens the data set, manipulates the data, and saves a new data set (which is then opened by the next do-file). In this way, we produce and store interim versions of the data, a process corresponding roughly to the steps taken in editing the data.

Conceptually, editing the anchor data entails two main tasks. First, the raw data have to be cleaned and debugged. In principle, the data are then ready to use and the preloads for the following wave can be generated. Second, the cleaned data are enriched by both generated variables (“user-friendly” variables based on, e.g. integration of information from two or more variables, local context information from an external data base, etc.) and newly generated data sets (*biopart*, *biochild*, *bioact*, *bioact_rtr*, *household*, *biomob_ehc*, *biomob_ehc_moves*, *biomob_rtr*, *biomob_rtr_parents*, and *bioparent*).

The relevant steps taken to accomplish these tasks are described in the next subsections.

¹For information about the sampling procedure, fieldwork, and the response rate please refer to the methods reports of the survey institute (2009, 2010, 2011, 2012, 2013).

Variable and value labels

Every variable in the data set was assigned a label. Variable labels contain a short description of the variable and its position (i.e. question number) in the CAPI questionnaire. The values of all variables were labeled according to the CAPI questionnaire.

Missing values

For all variables of the anchor data, we defined a set of missing codes, which were applied throughout (see table 4.1). Missing values “-1 Don’t know”, “-2 No answer” were assigned if the respondent could not or did not want to answer a question. These codes are the only missing values also documented in the questionnaire.

Value “-3 Does not apply” was assigned if a respondent had not been asked the corresponding question, i.e., the person was filtered over the question. Only variables indicating person numbers (e.g. *pid*, *cid*) have system missings other than -3 if the respective person does not exist.

Errors in the CAPI program, which erroneously guided respondents to the wrong questions in the interview are indicated by missing code “-4 Filter error / Incorrect entry” as are incorrect data entries by the interviewers.

In order to detect inconsistencies between a respondent’s answers, we checked for logically impossible or empirically implausible combinations of values on two or more variables. Inconsistent values were then coded to “-5 Inconsistent value” if it was clear that the value was wrong (see below for inconsistencies that could not be resolved in this way).

For open answers that were not legible, we assigned value “-6 Unreadable answer”.

For generated data (variables and files), we used value “-7 Incomplete data” to indicate cases where we lacked the information necessary to compute a valid value.

For variables that are not available for the DemoDiff sample, we assigned value “-10 Not in DemoDiff”.

For variables that are not available for the pairfam sample, we assigned value “-11 Not in pairfam”.

Table 4.1: Missing codes in data set *anchor*

Value	Label
-1	Don’t know
-2	No answer (also: I don’t want to answer that, answer refused)
-3	Does not apply
-4	Filter error / Incorrect entry
-5	Inconsistent value
-6	Unreadable answer
-7	Incomplete data
-10	Not in DemoDiff
-11	Not in pairfam

Preload variables for Dependent Interviewing - DI variables

In order to get more reliable responses in the CAPI of the second and the following waves, information from the previous wave was preloaded to the CAPI interview and presented to the respondents. As the previously reported information serves as a framework for the current wave, the quality of the responses is expected to improve. Further, the preload variables were used for routing respondents through the interview. The wave 2 preload information is based on the cleaned wave 1 data. The

preload information of wave 3 is based on wave 2 and wave 1 information depending on the participation in wave 2. The same procedure applies to wave 5. In total, 270 (wave 2), 237 (wave 3), 282 (wave 4) and 259 (wave 5) variables were preloaded to the interview.

The preloads or DI-variables are contained in the data sets *anchor2* to *anchor5*. They can be found at the beginning of the data sets. The variables are prefixed by a d- followed either by a 1- to 3-digit number or by person identifiers such as pid, smid or sfid. These variables are all made anonymous if necessary (strings and information on days). Please note that (unlike the other variables in the *anchor\$* data sets) these variables did not undergo the usual steps of data editing (e.g., the variables may contain system missings). Thus, the DI-variables should not be used for data analysis. A complete list of all DI-variables is contained in the anchor codebooks in the section "Preface".

Data cleaning of variables from the Event History Calendar (EHC)

The Event History Calendar used since wave 2 gathers information on four life domains: educational and occupational activities, children, partners, and residence. For a summary of all output variables created in the EHC, please refer to the anchor codebooks. Data preparation of the resulting variables (prefixed ehc-) contains detailed case-by-case analyses and corrections of inconsistent or implausible entries. In the following, we describe the data processing for each of the four domains.

EHC data cleaning: educational and occupational activities

The following steps were taken to prepare the information on educational and occupational activities (all *ehc19** variables). The general data cleaning contained checks of filters for open questions and, if possible, a recoding of open answers into existing categories (*ehc19i9o*, *ehc19i16o*). In spite of these checks, data were not manipulated in most cases since it is considered very difficult to find strong evidence for mistakes within the activity calendar. Activity biographies can contain many changes, gaps, and changing patterns. Therefore, entries by respondents and interviewers were generally considered credible.

Additionally, various checks were conducted to identify inconsistent episodes. A case was considered suspicious if interviewer notes suggested that problems appeared or that mistakes were made while filling in the EHC-activity calendar. Moreover, cases were considered suspicious if differences appeared between the month of the interview and the moment of the interview ($ehc19i^*mX \neq ehc19i^*$). If enough evidence was found, data were changed. In addition, the total number of activities and parallel activities per month in the calendar were checked.

EHC data cleaning: children

The data for information on children (all *vark** variables) were prepared as follows. First, we checked for repeatedly mentioned children of the same identity. Repetitions of exactly the same or very similar names were regarded as being the same child identities (e.g. by comparing dates of birth, sex and other information). All their variables (*vark**) were deleted (set on a missing value "-3") for children who represented repetitions within a wave.

Children with asynchronous positions between the waves were corrected by relocating the children (and all *vark**) falsely positioned according to their original position in preceding waves. Before any moving of a child to another position it was checked whether this new position was vacant. After moving, the variables on the old position were assigned a missing value ("-3"). In the event of unnecessary gaps between children (i.e., if a position between two children was empty), the children and all of their corresponding variables (*vark**) were moved to close the gaps. Please note that the gaps were closed only by moving children from a higher position to a lower gap position.

Additionally, the child's sex was checked using first names. If the first name indicated the sex unambiguously but did not match the particular child's indicated sex, the sex was changed to correspond with the first name (e.g. Herbert-female was changed to Herbert-male). Finally, for dead children

the variable for cohabitation was set to the missing value “-3”.

EHC data cleaning: partners

Concerning the information on partnerships gathered by means of the event-history calendar, we individually inspected all partnership biographies with suspicious entries such as deleted preloads, new partners with identical or similar names as the ones of the previous wave, implausible short cohabitation or marriage spells, marriages of anchor respondents belonging to the youngest cohort, and differences between the month of the interview (month 18 in wave 2, month 32 in wave 3, wave 4 and wave 5) and the date of the interview.

For clarification, we consulted answers to several additional questions from the anchor interview of the respective wave (if available). This refers to questions generating the partner’s name reported in the household grid, the fact whether the anchor has been employed in the partner’s business, answers to the single module, information on the new partner’s sociodemographics, the anchor’s satisfaction with his/her partnership, the separation module, questions on sexual behavior and parenting (did partner care for child?), the network module if available (partner’s name stated?), differences between the individual and the household income, and - finally - the fact whether the partner was present at the interview (as stated by the interviewer). If enough evidence was found, we changed the information stored in the data set *anchor\$*.

In the majority of cases, it could be inferred that partners from the previous wave were only mistakenly entered as new partners in the following wave. We recoded these partners as being the preloaded partner to indicate that the partnership with the partner from the previous wave still existed, at least at some time between the interviews². In addition, we recoded partners who had been entered as current partners if it was obvious that they were partners from the previous wave. If new partners were entered by mistake or if partners from the previous wave seemed to be implausible, they were deleted. Regarding all corrections, the auxiliary variables *hp** which are part of *anchor\$* as of wave 3 were adapted accordingly.

EHC data cleaning: residence

The EHC on the domain of residence is designed to collect information on the anchor’s current residence and the mobility between waves. It is also used to filter further questions about the current main (and second) household(s) at the time of the interview.

In wave 1, respondents were asked to report their current place(s) of residence.³ Respondents who mentioned more than one residence were asked which of them is their main residence. If more than two places of residence were mentioned, they were asked to indicate also the second residence (the one apart from the main residence where they spend most of their time).

From wave 2, respondents were presented with their information from the previous wave and then provided monthly information on all place(s) of residence since the previous interview. If the respondent indicated still living in the same city (or cities) as in the previous wave, he/she was asked whether he/she had moved to another dwelling in this city (cities). In case of an overlap of episodes (at least two consecutive months at the same place), the respondents were asked to indicate the main (and second) residence for each of the overlapping months. In wave 2 the CAPI program always shifted the information on the main residence at the time of the interview and the monthly information for this residence to the first position of the respective EHC variables (i.e. *ehc15p1*, whether the anchor currently lives there). If there was currently a second residence, this was always shifted to the second position *ehc15p2*. Please note that this is not valid since wave 3. Here, the first position of the

²The variable *tag_identp* as part of *anchor2* marks some additional cases which have been identified after cleaning the data and generating the preloads for the CAPI interview of wave 3. Consequently, no recoding was done and this tag variable was created instead (see table A.21).

³Specifically, the name of the city and the state of all their dwellings, i.e. “all rooms, apartments, or houses” where they “regularly stay overnight”. Respondents reporting to have more than one residence in the same city additionally were asked to provide information on the district or street to distinguish between dwellings.

respective EHC variables (i.e. *ehc28p1*) relates to the main dwelling of the previous wave. So the first position is not necessarily the current main residence. Since wave 4, the respondents should only name their main residence. After filling in the correct number of months, the respondents had been asked whether they have a second residence or not. Further questions about the second residence were asked later on after having finished the EHC module. Thus, second residences are no longer included in the event-history calendar data.

While processing the resulting EHC variables of wave 2, we encountered several problems with the data. In wave 1 some interviews apparently contained information on all residences instead of only the current residence. We preloaded this information regardless, as in most cases we were not able to ascertain whether the information was incorrect. In wave 2, some interviewers then seemed to have problems dealing with wrong preloads. Furthermore, the CAPI program always shifted the information on the main residence at the time of the interview and the monthly information for this residence to the first position of the respective EHC variables (i.e. *ehc15p1*, whether the anchor currently lives there). If there is currently a second residence, this was always shifted to the second position *ehc15p2*. Non-current residences (former first wave or between-wave households) were shifted to positions three and higher.

However, the program did not generate a pointer variable to indicate the main (and second) residence of wave 1. Moreover, when more than one current residence was mentioned, the information regarding which of them is the main residence was not stored by the program, and the same information for overlapping episodes was sometimes not stored correctly. Because of these problems, we analyzed cases in detail where there was a change of residence at the seam of wave 1 (after the wave 1 interview month), or at the seam of wave 2 (in the month of the wave 2 interview, or a difference between the interview month wave 2 and the current status). In addition, we analyzed all interviews with four or more residences mentioned.

Since wave 3 the data on residence was prepared as follows. First we checked if residences were indicated in which the anchor had never lived (all months and the current status were checked). If the residence of concern was not preloaded, all variables regarding the respective EHC residence was set to the missing value "-4". Furthermore, we merged two or more residences if the cities indicated had the same or a similar name (for example "München-Aubing" and "München-Schwabing") and neither of the residences was classified as second residence. We did not do this for "Berlin" because *ehc27p*i2* contains federal state information ("Berlin east" or "Berlin west") which would be lost. Apart from that we filled in gaps if the anchor had not indicated living at any residence in one or more months and if suitable information was provided by other variables. If a person lived nowhere in the last month before the interview and no current residence was named, we assumed that an entry mistake was made and filled in the month with the dwelling the anchor had lived in during the months before. In wave 3 we also used information from the retrospective questions (*rtr**), if available, to replace gaps.

For release 2.0 we recommend not to use the data contained in data set *anchor2* to conduct analyses of respondents' mobility. In wave 3 the quality of data was improved considerably, allowing for analysis of respondents' mobility. In this wave we also collected retrospective information on the respondents' mobility biography up to the interview of wave 3. Since release 4.0 four different panel data sets which contain information on the anchor's mobility are available: *biomob_ehc* (information on residences as collected through the EHC), *biomob_ehc_moves* (residences collected through the EHC, including moves within the same city), *biomob_rtr* (retrospective migration history since the age of 18), *biomob_rtr_parents* (moving out of the parental house) (for further information see chapter 4.6). For more detailed analyses of respondents' mobility over time we recommend using these generated data sets.

EHC data cleaning: household grid

As of wave 3, the EHC also collects information on individuals living at anchor's first or second household at the time of the interview. It is also used, in part, to filter questions on intergenerational

relationships later on.

During the editing process we checked if one of the household members is the anchor's partner, child or partner's child, because information on their cohabitation was already collected before and thus not relevant here. If this was the case, we assigned the missing value "-4". In wave 4 only individuals living at the main residence should be indicated. Individuals living with the anchor at the second residence are neglected. Since the partner and children information had been collected before, the respondents had the option to say that they live with them somewhere else, but not at the main residence.

In wave 4, household members' dates of birth could not be changed by the respondents. Therefore, the same person was entered again in some instances. This resulted in multiple identical persons and the problem was handled in this way: If similar persons exist, the case was checked in detail. If the person was the same, we deleted newly entered, not preloaded persons ("-4"). If information on date of birth etc. differed, we used the more recent information. If inconsistencies of date of birth or relationship status existed, the information was set to "-5". We filled in missing information if additional useful information had been given in previous waves. Relationships which occurred twice but could reasonably only occur once were set to "-5" or were deleted.

Coding open answers

In coding open answers, we adhered to the following procedure. First we checked the spelling of the entries and made corrections where necessary. Then we identified and recoded data errors. If a string variable contained information further qualifying the residual category of an answer list, we compared the open answer to the answer list. If appropriate, we recoded the open answer into an existing category and set the original value to missing (codes -4 or -6).⁴ Finally, all remaining open answers were coded to a single value, indicating merely that an open answer was provided. The actual string was deleted because of data protection.

Anonymity

Answers that might threaten our respondents' anonymity were deleted or recoded in the data set. Foremost, street addresses and respondents' names had already been dropped from the data set by *TNS Infratest*. We also deleted the information on exact dates, i.e. the day components, for privacy concerns.

By these means, all string variables in the data set were finally transformed to numeric variables. Thus the *anchor\$* data sets contain no string variables. Valid answers to open questions were recoded to value 1 throughout. The variables affected by the procedures to ensure anonymity are shown in table 4.2 along with the value labels.

Table 4.2: List of variables made anonymous

Variable	Variable label	Anonymous value label	Wave
cla4o	With whom did you live immediately after your birth? open entry	Other mentioned	2

continued on next page

⁴For information on the country of birth and nationality of the anchor, as well as of his or her partner and parents, we computed new variables where we grouped open answers into additional categories (see chapter 4.2).

continued from previous page

Variable	Variable label	Anonymous value label	Wave
cla6e1o, . . . , cla6e15 & cla6e7o, . . . , cla6e9o	Other: Lived with whom?	Other mentioned	2
crn12kxi14o	Child x: Other, namely:	Other mentioned	2, 5
crn13kxi13o	Child care morning child x: Other, own entry	Other mentioned	2 - 5
crn14kxi13o	Child care afternoon child x: Other, own entry	Other mentioned	2 - 5
crn18kxo	Child x: Custody - other, namely:	Other mentioned	2, 3, 5
crn1kxo	Other place of birth child x	Other mentioned	2, 3, 5
d1	Preload: Day of birth (Preload)	Day mentioned	2 - 5
d134, . . . , d143	City xst place of residence (Preload)	Residence mentioned	2 - 5
d14, . . . , d28	Name child x (Preload)	Name mentioned	2 - 5
d164, . . . , d173	Country xst place of residence (Preload)	Country mentioned	2 - 5
d176, . . . , d195	Name xst household member [main residence] (Preload)	Name mentioned	2 - 5
d218, . . . , d226	Name xst household member [second residence] (Preload)	Name mentioned	2
d267	Other education (Preload)	Yes	2 - 5
d274	Other type of employment (Preload)	Yes	2 - 5
d282	Current employment activity: open-ended answer (Preload)	Occupation mentioned	2 - 5
d398	Current partner's day of birth (Preload)	Day mentioned	2 - 5
d59, . . . , d73	Day of birth child x (Preload)	Day mentioned	2 - 5
d506	Name of marriage partner with whom no relationship at prev. wave (Preload)	First name mentioned	4, 5
d8	Name of current partner (Preload)	Name mentioned	2 - 5
dobd	Day of birth	Day mentioned	all
ehc12kxo	Other parent name child x (EHC)	Name mentioned	2 - 5
ehc14pxi1	City place of residence x (EHC)	Residence mentioned	2
ehc14pxi2o	Country place of residence x (EHC)	Country mentioned	2
ehc19i16mxo	Other type of employment, open entry in month x (EHC)	Other mentioned	2 - 5
ehc19i16o	Other type of employment, open entry currently (EHC)	Other mentioned	2 - 5
ehc19i22mxo	Other type of unemployment, open entry in month x (EHC)	Other mentioned	2 - 5

continued on next page

continued from previous page

Variable	Variable label	Anonymous value label	Wave
ehc19i22o	Other type of unemployment open entry currently (EHC)	Other mentioned	2 - 5
ehc19i9mxo	Other education open entry in month x (EHC)	Other mentioned	2 - 5
ehc19i9o	Other education open entry currently (EHC)	Other type of training mentioned	2 - 5
ehc1pxn	Name partner x (EHC)	Name mentioned	2 - 5
ehc20d	Day of birth of partner from previous wave (EHC)	Day mentioned	2 - 5
ehc22pxn	Name person x in household (EHC)	Name mentioned	3 - 5
ehc27pxi1	City place of residence x (EHC)	Residence mentioned	3 - 5
ehc27pxi2o	Country place of residence x (EHC)	Country mentioned	3 - 5
ehc7kxn	Name child x (EHC)	Name mentioned	2 - 5
ehc8kxd	Day of birth child x (EHC)	Day mentioned	2 - 5
frr13i14o	Reasons against child: Other Reason	Reason mentioned/ Other mentioned	1, 2, 4 & 5
hc1pxi1	Information x-th residence	Residence mentioned	1, 4
hc1pxi2o	Country x-th residence	Country mentioned	1, 4
hc8h1px	Main residence: Name person x	Name mentioned	1, 2
hc8h2px	Second Residence: Name person x	Name mentioned	1, 2
hcp1i1	Main residence partner	Residence mentioned	2 - 5
hcp1i2o	Country main residence partner	Country mentioned	2 - 5
hcp6i1	Second residence partner	Residence mentioned	3, 5
hcp6i2o	Country second residence partner	Country mentioned	3, 5
igr1d	Day of birth biological mother	Day mentioned	1
igr2d	Day of birth biological father	Day mentioned	1
igr3o	Other country of birth mother	Country mentioned	1
igr4o	Other country of birth father	Country mentioned	1
igr73i12o	Citizenship adoptive mother: Open entry other country	Other mentioned	3

continued on next page

continued from previous page

Variable	Variable label	Anonymous value label	Wave
igr74o	Open entry country of birth adoptive mother	Other mentioned	3
igr75o	Open entry (highest) level of school education of mother	Other mentioned	3
igr78o	Open entry country of birth adoptive father	Other mentioned	3
igr79o	Open entry (highest) level of school education of father	Other mentioned	3
igr82i12o	Citizenship mother's partner: Open entry other country	Other mentioned	3 - 5
igr83o	Open entry country of birth mother's partner	Other mentioned	3 - 5
igr85i12o	Citizenship father's partner: Open entry other country	Other mentioned	3 - 5
igr86o	Open entry country of birth father's partner	Other mentioned	3 - 5
int12o	Open entry suggestions or comments on the interview	Mentioned	3 - 5
int4i5o	Other persons, namely:	Other person mentioned	all
int9o	Reason partner won't participate	Reason mentioned	1
job1 / job20o	Current occupation	Occupation mentioned	all
mig1i12o	Country citizenship	Citizenship mentioned	1
mig4o	Other country of birth	Country mentioned	1
mig6i12o	Country other citizenship mother	Citizenship mentioned	1
mig7i12o	Country other citizenship father	Citizenship mentioned	1
netpxn	Name person x: Complete list name generator	Name mentioned	2, 4
rtr18kxd	Day of death child x	Day mentioned	1
rtr1pxn	Name partner x	Name mentioned	1
rtr23hx	Residence x	Residence mentioned	3
rtr24hxo	Open entry country of residence x	Other mentioned	3
rtr31i9o	Since 18th birthday: Open entry other education	Other mentioned	3
rtr35i7o	After school until W1: Open entry other type of employment	Other mentioned	3
sd14kxn	Name child x	Name mentioned	1
sd17kx	Name other parent child x	Name mentioned	1

continued on next page

continued from previous page

Variable	Variable label	Anonymous value label	Wave
sd19kxd	Day of birth child x	Day mentioned	1
sd23i16o	Other type of job, namely:	Occupation mentioned	1
sd23i9o	Other education, namely:	Type of training mentioned	1
sd32i5o	Since prev. wave: Other school leaving certificate, open entry	Other mentioned	2 - 5
sd4n	Name current partner	Name mentioned	1
sdp10i13o	Partner other type of employment	Occupation mentioned	all
sdp10i22o	Partner open entry other education	Other mentioned	3 - 5
sdp17o	Open entry other level of school education	Other mentioned	3 - 5
sdp1d	Day of birth current partner	Day mentioned	all
sdp2i12o	Country of other citizenship partner	Citizenship mentioned / Other mentioned	all
sdp5	In which country does current partner live	Country mentioned	all
sdp6o	Other country of birth current partner	Country mentioned	all
sep5kxo	Other custody of child x before separation	Other mentioned	2
sep7kxo	Other: Current custody of child x	Other mentioned	2, 3
sib2pxn	Name sibling x	Name mentioned	5

Value checks

As documented in the anchor questionnaire, a number of value checks were included in the CAPI program and therefore conducted already during the interview. In addition, we checked for each variable in the data whether the actual value range corresponded to the range of possible values listed in the questionnaire. Values out of range were regarded as incorrect entries and therefore recoded to -4.

Filter checks

We checked the filter of every variable as documented in the CAPI questionnaire and assigned the missing code -4 for data errors. On the one hand, a question may have been asked by mistake, or not asked by mistake if the filter was not implemented correctly in the CAPI program. These cases were both regarded as filter errors and the variable then set to value -4.⁵ On the other hand, values were classified as data errors if they were incorrectly entered by the interviewer. In these cases, we also assigned value -4. In checking the filters, we proceeded in the order in which questions were posed in the interview, to make sure that ensuing filter errors were detected.

⁵If a question is asked despite the filter, the resulting data will be of low quality in many instances, often because the question does not make sense to the respondent. Furthermore, it is often not only hard to determine whether the answers actually are useful, but it also requires considerable effort in terms of data inspection and documentation. Although we might delete useful information in some cases, we nonetheless decided to always recode these cases to -4.

Checks for consistency

As mentioned above, we conducted various checks to identify logically impossible or empirically implausible combinations of values on two or more variables.⁶ In some cases it was not possible to resolve the inconsistency by assigning the code -5, because we could not decide which of the variables under consideration was wrong. Consequently, the provided values were left unedited. Instead, a flag variable indicating the respective inconsistency was generated. Furthermore, some flags were produced to explain why the value -5 was assigned to the respective variable (see *flag12*). For each of these variables, code 0 indicates that the respective inconsistency is non-existent and code 1 that there is an inconsistency. The code -3 indicates that the flag variable does not apply to the respective wave. We performed a number of additional checks that did not provide an indication of any inconsistency. Table A.21 describes the flag variables in detail.

As of wave 2, we computed tag variables to indicate inconsistencies over time, i.e. over waves. These relate to the anchor's, the current partner's, and the children's sex and date of birth. All of these variables were preloaded and should be time-constant. The tag variables identify any inconsistent cases. Furthermore, the indicators whether the anchor or the anchor's partner is the biological parent of a child is tagged if the information from the previous and the current wave are contradictory. Table A.22 lists all of the generated tag variables that are part of the data sets *anchor2*, *anchor3*, *anchor4* and *anchor5*.

The flag and tag variables are provided as an additional service for users, to help them decide which information to use. We do not claim to have identified all major inconsistencies in the data. We strongly recommend using flagged values with caution. In many instances, it probably will be possible to assign plausible values rather than to exclude all inconsistent cases right away, but the users themselves are responsible for this. In the case of tagged values marking sex and date of birth, the respective generated identifiers explained in section 4.2 should be used.

Remark on variable flag_frt6 (wave 1 & wave 2)

In question 129 (wave 1), we asked respondents who already had children: "When you think realistically about having (additional) children, how many (more) children do you think you will have?" The intention was for respondents who already had or were expecting children to give the number of additional children. Unfortunately, some respondents seem to have overlooked the "more" (This is probably a framing effect, because in question 128 we had asked for the total number of children the respondent would ideally like to have). It seems that some respondents reported the total number of children they were thinking of having, including those already born or conceived. Thus, for those respondents the value of *frt6* is too large.

Accordingly, in wave 2 the wording of the question was changed to make our intention regarding that question more clear to respondents (question 154). An extra sentence was added for respondents who were pregnant/whose partner was pregnant/who already had children: "Here we mean children in addition to the ones you already have, or if you or your partner is pregnant, in addition to the child you are expecting." Unfortunately, after looking more closely at the data we must admit that the problem from wave 1 was still not solved. Some respondents still reported the total number of children they were thinking of having, still including those already born or conceived.

In wave 1 and in wave 2, there is no way of telling how respondents answered the question, whether in regard to additional children or to the total number of all children; no data editing procedure will solve the problem. Therefore, a flag variable was created to mark all respondents who potentially gave an incorrect answer for *frt6*. We flagged those respondents who already had children but reported a number of planned children greater or equal to the number of children they already had (wave 1: N=1,656; wave 2: N=1,043). This is a 'worst case scenario': Not all of these respondents

⁶This happened in addition to a number of checks that had already been implemented in the CAPI program (see anchor codebooks).

will have overreported the number of additional children they intended to have. But some certainly did.

When analyzing variable *ftr6* with this flagged subset of the sample, please proceed with caution. There are different options for analyzing these data:

1. Consider using the question on the intention to have a child within the next two years (*ftr7*) for your analyses instead of *ftr6*.
2. Use *ftr6* only for childless respondents who are not pregnant (unflagged values). For these cases there should be no problem with this question.

Note that *ftr6* is also used to filter some of the following questions. Respondents reporting false values on this variable may also have answered too many or too few of those other questions.

As of wave 3 we constructed the questionnaire in the way that *ftr6* was divided into three questions. In question *ftr26* respondents without children were asked how many children they will have realistically. Furthermore, to rule out wrong answers a filter variable was integrated (*ftr27*). Question *ftr27* asked pregnant respondents or respondents who already had children if they think that they will have additional children. Only if they indicated that they will have additional children were they asked how many additional children they plan to have (*ftr28*). Therefore *flag_ftr6* is not relevant for wave 3 and the following waves.

Remark on variable flag_igb (as of wave 2)

In the module on intergenerational relations (IGB), we asked respondents about parents who are identified as household members earlier in the questionnaire. Parents can be either biological parents, stepparents or adoptive parents, or combinations thereof (e.g. biological mother & stepfather). It was decided that adoptive parents only apply if respondents lived with adoptive parents before age 6. Therefore we did not consider combinations of one biological parent, and one adoptive parent or adoptive parent and stepparent with regard to the filtering and question wording in the IGB-module. It is unclear how respondents understood and answered certain questions in the module. The parent-combination help variables are used throughout the IGB-module for filtering; respondents may also have answered too many or too few other questions in this module. Thus we decided to provide a flag variable indicating these parent combinations. Users have to decide whether or not to use these cases in analyses.

Generated variables, scales and generated data sets

To generate variables and additional spell/panel data sets, we started from the cleaned data, i.e. the preliminary version 0.5 of wave 1 anchor data released in November 2009 and the preliminary versions of wave 2, wave 3, wave 4 and wave 5 anchor data used to compute the preloads for the next wave. By following the procedures we used to generate these data, users will be able to reconstruct the process.⁷ Please note that some generated variables and scales have been modified for some releases. The changes of the generated variables are documented in the appendix, see tables A.7 to A.15. The details on generated variables and scales are given in chapter 4.2, those on processing the generated files *biopart*, *biochild*, *bioact*, *bioact_rtr*, *household*, *biomob_ehc*, *biomob_ehc_moves*, *biomob_rtr*, *biomob_rtr_parents*, and *bioparent* in chapter 4.6.

English-language data

The final step in editing the data was to produce an English version of the data in order to enable non-German speakers to use pairfam data. To produce the English data set, all variable labels and value labels were translated according to the wording of the English version of the anchor codebook.

⁷Users should, however, start with the latest version of the wave 1 to wave 5 data (release 5.0) when running the do-files for the respective wave, since there have been minor changes in the data after the prior releases. These changes are documented in the appendix, see tables A.7 to A.15.

4.2 Generated variables and scales

In order to facilitate data analysis and to enhance comparability of results, the pairfam staff produced a number of variables that are of interest to many research projects. Table 4.3 shows a list of all generated variables. These variables are part of the delivered anchor data of all waves. This chapter describes the computation and content of the generated variables.

In general, we aim to provide the syntax written to produce these variables. Stata do-files are available as part of the scientific use file and on the project homepage <http://www.pairfam.de> for each wave separately. Users are invited to adapt the syntax to their special research needs.⁸ Before using any of the generated variables we strongly advice users to always check whether the respective generating procedures meet their specific needs. Please note that there is a common missing value “-7 Incomplete data” encompassing the original missing codes -1, -2, -4, -5, and -6. Changes in the syntax for generated variables between different releases are documented in tables A.7 to A.15.

Furthermore, table 4.17 displays all scales for the anchor data from the available waves. These variables are not part of the delivered data sets, but can be generated by users themselves. The corresponding syntax files (available for Stata and SPSS) are provided as part of the scientific use file and on the pairfam website. For additional information regarding these scales, please refer to the scales manual (Thönnissen et al. 2014).

Table 4.3: List of generated variables included in data set *anchor\$*

Construct	Variable name
Generated identifiers sex (anchor, partner, children)	sex_gen, psex_gen, k*sex_gen
Generated identifiers date of birth (anchor, partner, children, mother incl. adoptive mother, father incl. adoptive father, stepmother, stepfather)	dob*_gen, pdob*_gen, k*dob*_gen, mdob*_gen, fdob*_gen, smdob*_gen, sfdob*_gen
Age (anchor, partner, mother incl. adoptive mother, father incl. adoptive father, stepmother, stepfather)	age, page, mage, fage, smage, sfage
Age of children	k1age, ..., k10age
Age of anchor's youngest child living with anchor (in months)	ykage
Position of anchor's youngest child	ykid
Age of anchor's youngest CAPI child in months	ykagecapi
Position of anchor's youngest CAPI child	ykidcapi
Birth cohort	cohort
Country of birth (anchor, partner, mother, father, adoptive mother, adoptive father, stepmother, stepfather)	cob, pcob, mcob, fcob, amcob, afcob, smcob, sfcob
1st/2nd/3rd nationality (anchor, partner, mother, father, adoptive mother, adoptive father, stepmother, stepfather)	nat*, pnat*, mnat*, fnat*, amnat*, afnat*, smnat*, sf- nat*

continued on next page

⁸Running the do-files requires Stata version 11.0 or later.

continued from previous page

Construct	Variable name
Ethnicity	ethni
Migration status	migstatus
Relationship status	relstat
Marital status (anchor, partner)	marstat, pmarstat
Number of previous partners	np
Number of previous partners with whom anchor cohabitated	ncoh
Number of previous marriages	nmar
Months since anchor and current partner got to know each other	meetdur
Duration of current relationship, cohabitation and marriage	reldur, cohabdur, mardur
Anchor homosexual	homosex
Anchor and/or partner infertile	infertile
Anchor and/or partner pregnant	pregnant
Number of all kids born up to time of interview	nkids
Number of all biological kids born up to time of interview	nkidsbio
Number of all biological kids with partner born up to time of interview	nkidsp
Number of all kids alive	nkidsalv
Number of all biological kids alive	nkidsbioalv
Number of all biological kids with partner alive	nkidspalv
Number of all kids living with anchor	nkidsliv
Number of all biological kids living with anchor	nkidsbioliv
Number of all biological kids with partner living with anchor	nkidspliv
Number of all partner's biological kids alive	pnkidsbioalv
Type of children	k1type, . . . , k10type
Respondent has 2nd residence	res2nd
Household size (main residence)	hhsizemrd
Mother lives in household (main residence)	mmerd
Father lives in household (main residence)	fmerd
Partner lives in household (main residence)	pmerd
Number of children living in household (main residence)	childmrd
Number of others living in household (main residence)	othmrd
Household composition (main residence)	hhcomp
Enrollment in school or vocational qualification at time of interview (anchor, partner)	enrol, penrol
Highest school degree attained at time of interview (anchor, partner, mother, father)	school, pschool, mschool, fschool
Highest vocational degree attained at time of interview (anchor, partner, mother, father)	vocat, pvocat, mvocat, fvocat
ISCED-97, International Standard Classification of Education, no students (anchor, partner, mother, father)	isced, pisced, misced, fisced
ISCED-97, International Standard Classification of Education, including students (anchor, partner)	isced2, pisced2
CASMIN classification of educational attainment (1999) (anchor, partner, mother, father)	casmin, pcasmin, mcasmin, fcasmin
Years of schooling / vocational qualification (anchor, partner, mother, father)	yeduc, pyeduc, myeduc, fyeduc

continued on next page

continued from previous page

Construct	Variable name
KldB classification of occupation	kldb2010, kldb1992
ISCO classification of occupation	isco08, isco88
Erikson-Goldthorpe-Portocarero class schema (EGP)	egp
Int. Socio-Economic Index of Occupational Status (ISEI)	isei
Standard Int. Occupational Prestige Scale (SIOPS)	siops
MPS occupational prestige score	mps
Net equivalence income according to GCEE	hhincgcee
Net income (open and estimated information combined) (personal, household)	incnet, hhincnet
Consumer price index acc. to Nat. stat. off. (baseline 2010)	cpi
Current primary and secondary activity status (anchor, partner)	casprim, cassec, pcasprim, pcassec
Labor force status (anchor, partner)	lfs, plfs
Currently living in East Germany	east

Generated identifiers sex - sex_gen, psex_gen, k*sex_gen

To solve the problem of asynchronous information on the respondents' sex between the different informational sources⁹, we generated "best solution" sex variables for the anchor, the partner, and the children (*sex_gen*, *psex_gen*, *k*sex_gen*).

This was done according to the following rules: In general, self-reported sex information was preferred over proxy information. The *_gen variables were synchronized over the waves. Therefore a specific person has the same (best) sex information in all waves. The best sex information means that the value that was stated the most was used. If two values have been stated equally often the most frequently or (if not existent) the newest reported value was preferred. If the partner or a child was nonexistent, we assigned the code "-3 Does not apply". If the information was not available to derive the sex, we used the code "-7 Incomplete data". The Stata do-file identifiers.do, which can be found in the syntax folder of the latest wave, contains the syntax used to compute the variables. Note that in release 4.0 information on sex of 1 anchor person was set to -4 based on additional information from the anchor person about change of sex in wave 4.

Generated identifiers date of birth - dob*_gen, pdob*_gen, mdob*_gen, fdob*_gen, smdob*_gen, sfdob*_gen, k*dob*_gen

In order to cope with the problem of asynchronous information on the date of birth between different informational sources¹⁰, we generated best solution variables for the month and the year of birth of the anchor, the partner, the parents (incl. the adoptive parents), the stepparents, and the children (*dob*_gen*, *pdob*_gen*, *mdob*_gen*, *fdob*_gen*, *smdob*_gen*, *sfdob*_gen*, *k*dob*_gen*).

This was done according to the rules used for the generated identifiers for the gender (see above). The Stata do-file identifiers.do, which can be found in the syntax folder of the latest wave, contains the syntax used to compute the variables.

Age - age, page, mage, fage, smage, sfage, k*age

The variables *age*, *page*, *mage*, *fage*, *smage*, *sfage*, and *k*age* contain the anchor's, the partner's, the parents' (incl. the adoptive parents'), the stepparents', and the children's age, respectively. These variables were calculated on the basis of the generated date of birth variables (see above). The age values were calculated by subtracting the corresponding generated year of birth from the anchor's

⁹The various data sets of the anchor and the alteri, partially available for different waves.

¹⁰See footnote above.

year of interview (e.g. *age = inty-doby_gen*). Additionally, the generated month information was taken into account. Should a person not yet have had his/her birthday (*doby_gen < intm*), the built age variable was reduced by 1. For those persons with missing values in the generated month of birth variable, the calculation was done only on the basis of the year of birth information. In cases of ambiguous seasonal information on the month and valid answers for the year of a variable indicating a date, random values were imputed for the month. If a specific alter does not exist, the code “-3 Does not apply” was given. If the information necessary to derive the age was not available, the code “-7 Incomplete data” was given. The variables were generated by running the do-file *age.do*.

Age of youngest child - *ykage*, *ykid*, *ykagecapi*, *ykidcapi*

The variables *ykage* and *ykagecapi* show the age of the anchor person's youngest child (respectively the youngest CAPI child) living with the anchor. The variables *ykid* and *ykidcapi* contain the position of the youngest child/CAPI child corresponding to the variables *varX* in the anchor data sets (child number 1 to child number 10; as of wave 2: position in the EHC). Please note that there can be new CAPI children in each wave, so the youngest CAPI child might be a different one in the following wave. The variables were generated by running the do-file *age.do*. They are provided from release 4.0 on. Please note that *ykagecapi* and *ykidcapi* are not relevant for wave 1. Furthermore, *ykidcapi* is not relevant for wave 2. Since there was only one CAPI child in each household in wave 2 the variable *capikid* already gives the position of the youngest CAPI child.

Birth cohort - *cohort*

The variable *cohort* indicates to which of the three birth cohorts 1971-1973, 1981-1983 and 1991-1993 the anchor belongs. The information is derived from both the stated date of birth and the date of birth given by the register data. In cases of a contradiction, *Infratest* recontacted the respondents for clarification. The methods report of wave 1 (Suckow and Schneekloth 2009) states that the gross sample was drawn from the birth cohorts 1971-1973, 1981-1983 and 1991-1993. Due to reporting practices of some municipalities, however, a few respondents in directly adjacent birth cohorts entered the gross sample and also remain in the net sample. These respondents were assigned to the corresponding birth cohorts, e.g., a respondent with year of birth 1970 was assigned to the birth cohort 1971-1973. Since the birth cohort is a time-constant variable it is fixed throughout the waves. The variable was generated by running the do-file *cohort.do*.

Country of birth - *cob*, *pcob*, *mcob*, *fcob*, *amcob*, *afcob*, *smcob*, *sfcob*

These variables contain information on the country of birth of the anchor and his or her alteri (partner, mother and father, adoptive mother/father, stepmother/stepfather). The information on the alteri was given by the anchor. Open answers specifying a country not contained in the corresponding answer list were recoded and grouped into larger geographical or political regions (values “13” to “23”). Table 4.4 lists values and labels of all countries and regions.

Since country of birth is a time-constant concept, the anchor's as well as his/her parents' country of birth (*cob*, *mcob*, *fcob*) is based on information gathered in wave 1. The country of birth of the adoptive mother/father and the stepmother/stepfather is collected for the first time in wave 3. Accordingly their country of birth (*amcob*, *afcob*, *smcob*, *sfcob*) was built for the first time in wave 3 and will be copied for later waves. If there is a new stepmother/stepfather in wave 5 we use *igr86* and *igr83* from wave 5 to compute *smcob/sfcob*. If the partner mentioned in the previous wave is still the current partner, the generated variable *pcob* is copied from the previous wave. If the current partner is not the same one as in the previous wave, we use *sdp6* and *sdp6o* from wave 5 to compute *pcob*. If there is insufficient information in wave 5, information of the previous wave is used, if available. Due to privacy concerns, the syntax which generates these variables cannot be published. Please note that for wave 3 and 4, release 5.0, some corrections have been performed.

Nationality - nat*, pnat*, mnat*, fnat*, amnat*, afnat*, smnat*, sfnat*

These variables contain information on the nationality of the anchor and his or her alteri (partner, mother, father, adoptive mother/father, stepmother/stepfather). The variables contain information on the alteri given by the anchor. From wave 2 on, nationality is only asked for new partners. Additionally, in wave 3 the nationality of the adoptive mother/father is reported. The nationality of the stepmother/stepfather is stated from wave 3 on. Up to three nationalities are stored in variables **nat1*, **nat2* and **nat3*¹¹. Note that the order of storage does not convey any special meaning. **nat1* simply always contains the nationality with the lowest value (values are listed in table 4.4 below). Hence, if a person is a German citizen, this is always reflected in variable **nat1*. As with country of birth, open answers specifying a nationality not contained in the corresponding answer list were recoded and grouped into larger geographical or political regions (values “13” to “23”). As for country of birth, we do not publish the do-file to compute these variables.

Table 4.4: Values and labels of variables **cob* and **nat**

Value	Label
-7	Incomplete data
-3	Does not apply
1	Federal Republic of Germany
2	German Democratic Republic ^a
3	Turkey
4	Russian Federation
5	Poland
6	Italy
7	Serbia
8	Croatia
9	Greece
10	Romania
11	Kazakhstan
12	Bosnia-Herzegovina
13	Former Soviet Union
14	(South)Eastern Europe
15	former Yugoslavia
16	Southern Europe
17	North, West, Central Europe
18	Middle East and Hindu Kush
19	Asia
20	North America
21	Central and South America
22	Africa
23	Other country ^b
24	Stateless/nationality unsettled ^c

^a only for country of birth

^b “Other states” for nationality

^c only for nationality

Ethnicity and migration status of anchor - ethni, migstatus

Based mainly on the mother’s and father’s country of birth (generated variables *mcob* and *fcob*), the variable *ethni* contains the anchor’s ethnicity. The categories are: “German native, No migration background”, “Ethnic-German immigrant (Aussiedler)”, “Half-German”, “Turkish background”, “Other non-German background”. In addition, the variable *migstatus* reports the anchor’s migration status

¹¹So far, three nationalities exist for partners only.

and, if applicable, to which generation of migrants the anchor belongs. The categories are: “No migration status”, “1st generation”, “2nd generation”. Since these variables are time-constant, the computation is based on anchor data of wave 1 and the variables are copied for later waves. The syntax used to compute both variables is contained in the Stata do-file `migration.do`.

Relationship status - `relstat`

The variable `relstat` contains information about the anchor's detailed relationship status. The information was taken from the anchor interview. The categories are: “never married single”, “never married LAT”, “never married COHAB”, “married COHAB”, “married noncohabiting”, “divorced/separated single”, “divorced/separated LAT”, “divorced/separated COHAB”, “widowed single”, “widowed LAT”, and “widowed COHAB”. LAT stands for ‘living apart together’ and COHAB stands for ‘cohabiting’. ‘Married’, ‘divorced’ and ‘widowed’ also include ‘civil union’, respectively the ‘dissolution of a civil union’. Please note that ‘cohabiting’ and ‘married/civil union’ imply that the anchor also has a relationship. If a person is just ‘cohabiting/married’ but does not currently have a partnership, the person is regarded as being single. ‘Separated’ means being married but having no relationship with the current spouse. If the information necessary to derive the relationship status was not available, the code “-7 Incomplete data” was given. The syntax used to compute the variables is contained in Stata do-file `relstat.do` for the respective wave.

Marital status - `marstat`, `pmarstat`

The variables `marstat` and `pmarstat` contain the anchor's and his/her partner's marital status. The information on the anchor's marital status was taken from the anchor interview. The information on the partner's marital status was taken only from the partner questionnaire. The categories are: “never married”, “married/civil union”, “divorced/dissolved civil union” and “widowed/surviving partner in civil union”. For the partner's marital status, the code “-3 Does not apply” was given if no partner existed. If the information necessary to derive the marital status was not available, the code “-7 Incomplete data” was given. The syntax used to compute the variables is contained in the wave-specific Stata do-file `marstat.do`.

Number of previous partners, cohabitations and marriages - `np`, `ncoh`, `nmar`

The variables `np`, `ncoh` and `nmar` cover information on the number of previous, i.e. already terminated partnerships, cohabitations, and marriages. Note that the variable `ncoh` captures the number of previous partners with whom the anchor cohabited, not the total number of cohabitation episodes (i.e. one episode for each previous partner at most). Furthermore, the variable `np` counts the same previous partner only once although there has been a break or another partner in between. That is why the total number of previous partners of the current wave can be smaller than the total number of the wave before. The value “-7 Incomplete data” for `ncoh` and `nmar` indicates that respondents did not answer the relevant questions (“-1” or “-2”). The computation of these variables is based upon the latest version of the data set `biopart` (for further information see chapter 4.6). The do-file `npart.do` includes the syntax to compute these three variables.

Duration since having met current partner; duration of current relationship, cohabitation and marriage - `meetdur`, `reldur`, `cohabdur`, `mardur`

The variable `meetdur` is a measure of the number of months that have passed since the anchor and his or her current partner got to know each other. The code “-7 Incomplete data” was assigned if the anchor did not provide the date of the first meeting with his or her current partner (“-1” or “-2” for `pa2m` and `pa2y`)¹². The variables `reldur`, `cohabdur`, and `mardur` document the durations of the current relationship, cohabitation, and marriage in months. If the anchor reported more than

¹²If `pa2y` was known and only `pa2m` was “-1” or “-2” or contained an ambiguous seasonal information, we imputed a random variable between “1” and “12” for the month, taking into account sensible upper and lower boundaries (see do-file `durpart.do`).

one relationship or cohabitation episode with his or her current partner, the beginning of the latest episode was used to generate the variables *reldur* and *cohabdur*. Please pay attention to the fact that the computation of these variables (see do-file *durpart.do*) employs variables included in the data set *biopart* (see chapter 4.6).

Anchor homosexual - homosex

The variable *homosex* contains information on whether or not the anchor is homosexual. The information was taken from the anchor interview, combining information of all five waves. The gender information was taken from the generated gender variables (*sex_gen* and *psex_gen*). The anchor was defined as being homosexual if he or she reported to be homosexual or had a same-sex partner and heterosexual if he or she stated to be heterosexual or had a partner of the other sex in wave 5. If the anchor had no partner in wave 5 the gender of previous partners was used. If no information concerning a sexual preference was available in wave 5, information from the previous waves (anchor's reported homosexual/heterosexual preference or gender of partner) was added. Where only insufficient information (of neither wave) was available to derive the status, the code "-7 Incomplete data" was given. The syntax used to compute the variable is contained in the Stata do-file *homosex.do*.

Anchor and/or partner infertile - infertile

The variable *infertile* contains the information that either the anchor and/or the partner were infertile. The information was taken from the anchor and the partner interview. This variable is an indicator at the couple level if the anchor currently had a partner. Otherwise, it is an indicator at the individual level. Couples were seen to be infertile if at least one partner counted as infertile. A couple was fertile if both partners counted as fertile. A person counted as infertile if he or she stated it explicitly or if he or she used sterilization for contraception. If no self-reported information was available and the partner stated the infertility of his or her partner the couple was seen to be infertile. Generally self-reports dominated indirect reports but if no direct information was available the partner's information was used. A person was assigned a positive fertility status if he or she stated explicitly to be so or if information on a pregnancy could be found. If no information on fertility was available in wave 5 and if the information from previous waves indicated that the anchor or the couple unit was seen to be infertile (only for same couples as in the previous wave respectively), then the code for infertility status was given.

In three cases, the code "-7 Incomplete data" was assigned. First, it was assigned if the question on fertility was not posed (birth cohort 1991-1993) and simultaneously no information on a pregnancy was available. Second, the code "-7" was given if the information on a decision on the fertility status was insufficient. Third, value "-7" was given if there was conflicting information about a person, for example if the anchor indicated being fertile but also indicated use of sterilization for contraception. The syntax used to compute the variable is contained in the wave-specific Stata do-file *infertile.do*. While in wave 1, 3 and 5 the anchor as well as the partner were asked about their fertility status or if they use sterilization for contraception, in wave 2 and wave 4 only the anchor gave information about his and the partner's fertility status. Therefore the wave-specific do-files differ respectively.

Anchor and/or partner pregnant - pregnant

The variable *pregnant* indicates whether the anchor or his/her partner was currently pregnant. The variable is an indicator at the individual level for female singles and at the couple level for respondents with a partner. The information was taken from the anchor and the partner interviews (if available). A couple was defined as being pregnant if either the anchor or his/her partner explicitly indicated being pregnant, or if the anchor indicated that his/her partner is pregnant. Single female respondents were defined as pregnant if they reported to be so. Additionally, single female respondents who had never had sex were seen to be not pregnant. If a specific single or couple unit did not have the potential to have a pregnancy status (single men and male-male couples), the value "-3 Does not apply" was given. Where information was not sufficient to decide on the status of pregnancy, the

code “-7 Incomplete data” was assigned. The syntax used to compute the variable is contained in the Stata do-file `pregnant.do`.

Number of children - `nkids`, `nkidsbio`, `nkidsp`, `nkidsalv`, `nkidsbioalv`, `nkidspalv`, `nkidsliv`, `nkidsbioliv`, `nkidspliv`, `pnkidsbioalv`

These variables contain the number of children of each type listed in table 4.5. The information for all variables except for `pnkidsbioalv` is taken from the anchor interview. The information for the variable `pnkidsbioalv` is taken from the partner questionnaire. If a partner did not exist, the code “-3 Does not apply” was given for the variable `pnkidsbioalv`. In cases where the information was not available for deriving the number of the specific type of child, the code “-7 Incomplete data” was given. Please note that the concept for the variable `pnkidsbioalv` has changed since release 4. See Stata do-file `nkids.do` for the syntax used to compute the variables.

Table 4.5: Number of children - **nkids**

Variable	Label
<code>nkids</code>	Number of all kids born until time of interview
<code>nkidsbio</code>	Number of all biological kids born until time of interview
<code>nkidsp</code>	Number of all biological kids with partner born until time of interview
<code>nkidsalv</code>	Number of all kids alive
<code>nkidsbioalv</code>	Number of all biological kids alive
<code>nkidspalv</code>	Number of all biological kids with partner alive
<code>nkidsliv</code>	Number of all kids living with anchor
<code>nkidsbioliv</code>	Number of all biological kids living with anchor
<code>nkidspliv</code>	Number of all biological kids with partner living with anchor
<code>pnkidsbioalv</code>	Number of all partner’s biological kids alive

Type of children - *k*type*

For each child, we generated a variable *k*type* indicating to which category the child could be assigned. In the event that the type of a child could not be found out the code “-7 Incomplete data” was given. If a child did not exist, the code “-3 Does not apply” was assigned. The different types are shown in table 4.6. The do-file `ktype.do` generates the variables for each wave.

Table 4.6: Type of children - *k*type*

Code	Label
1	Adopted, step or foster child, deceased
2	Biological child not from current partner, deceased
3	Biological child from current partner, deceased
4	Adopted, step or foster child, living not with anchor
5	Biological child not from current partner, living not with anchor
6	Biological child from current partner, living not with anchor
7	Adopted, step or foster child, living with anchor
8	Biological child not from current partner, living with anchor
9	Biological child from current partner, living with anchor

Second residence - `res2nd`

Variable `res2nd` indicates whether the respondent reported having a second residence. In wave 4 & 5, the information on the domain of residence was not derived from the EHC (as it was the case in waves 1-3), but respondents were asked directly whether they had a second residence or not. The syntax used to compute the variable is contained in Stata dofile `hhsizedo`.

Household size of main residence - *hhsizemrd*

The variable *hhsizemrd* contains the household size of the anchor's main residence, i.e. the number of persons living in the household including the anchor. The information was derived from the household grid. The syntax used to compute the variable is contained in Stata do-file *hhsizemrd.do*.

Parents living at main residence - *mmerd*, *fmerd*

The variables *mmerd* and *fmerd* indicate whether the anchor's mother and father live at the anchor's main residence. The information was derived from the household grid. Biological, adoptive, step and foster parents were counted as parents. The syntax used to compute the variable is contained in Stata do-file *hhsizemrd.do*.

Partner living at main residence - *pmrd*

The variable *pmrd* indicates whether the anchor's partner currently lives at the anchor's main residence. The information was derived from the household grid. The syntax used to compute the variable is contained in Stata do-file *hhsizemrd.do*.

Other household members at main residence - *othmrd*

The variable *othmrd* indicates how many household members live at the anchor's main residence. Other household members are all persons other than partner, children, and parents. The information was derived from the household grid. The syntax used to compute the variable is contained in Stata do-file *hhsizemrd.do*.

Children living at main residence - *childmrd*

The variable *childmrd* indicates how many children lived at the anchor's main residence. Only cohabiting children of the anchor are included, i.e. biological, adopted, step and foster children. The information was derived from the household grid. The syntax used to compute the variable is contained in Stata do-file *hhsizemrd.do*.

Household composition (partner, children, other HH-members) - *hhcomp*

The variable *hhcomp* combines the information on cohabitation with a partner, with at least one parent, with at least one child, and with at least one other household member into one variable. The 16 values of the variable are derived from a cross-classification of these four facts. The syntax used to compute the variable is contained in Stata do-file *hhcomp.do*.

Enrollment in school or vocational qualification - *enrol*, *penrol*

Variables *enrol* and *penrol* capture information on the anchor's and his or her partner's current enrollment in the educational system. Both general schooling and vocational qualifications were considered. Values and labels of the variables are listed below (see table 4.8).

The computation of the variable *enrol* relies on information about the anchor's current activities, and, for pupils and students, on their school type. In case of multiple activities, a dominance structure was applied in which more concrete information was preferred to less concrete information, enrollment in vocational institutions was preferred to general schooling, and more valuable vocational degrees were preferred to less valuable ones. Please note that we put people who do "Retraining / Further education" into the category "16 Other education". If there is no information at all, previous wave data is used.

The default operation in constructing partner's enrollment uses proxy information given by the anchor. However, if relevant information was given in wave 1, 3 and wave 5 by the partner in the partner

Table 4.7: Household composition (partner, kid(s), parent(s), others) - *hhcomp*

Code	Label
1	w - w - w - w
2	w - w - w -w/o
3	w - w - w/o -w
4	w - w - w/o - w/o
5	w - w/o - w - w
6	w - w/o - w - w/o
7	w - w/o - w/o - w
8	w - w/o - w/o - w/o
9	w/o - w - w - w
10	w/o - w - w - w/o
11	w/o - w - w/o - w
12	w/o - w - w/o - w/o
13	w/o - w/o - w - w
14	w/o - w/o - w - w/o
15	w/o - w/o - w/o - w
16	w/o - w/o - w/o - w/o

survey, this information was used instead. In wave 2 and wave 4, only proxy information given by the anchor was available and was therefore used to construct the variable *penrol*.

Note that information on partner's enrollment deviates from anchor's information for several reasons. In wave 1 neither the anchor nor the partner was asked about the kind of school the partner was attending. Thus, partners enrolled in the German school system of general education were assigned value "9 General school without further information". Moreover, anchors could report that their partner was attending vocational school (berufsorientierte Schule). In this case, we assigned value "19 Vocational school without further information" as a default. If the partner gave other, more detailed information, this information was used instead.

Table 4.8: Values and labels of variables *enrol*, *penrol*

Value	Label
-7	Incomplete data
-3	Does not apply
0	Not enrolled
1	Hauptschule
2	Realschule
3	Gymnasium
4	Gesamtschule
5	Sonderschule / Förderschule
6	Other school
7	Kolleg, 2. Bildungsweg
9	General school w/o further information
11	Vocational training (berufl. Ausbildung)
12	Fachschule
13	Berufsakademie
14	University, FH
15	Berufsvorbereitende Maßnahme
16	Other education
19	Vocational school w/o further information

In wave 2 to wave 5 we have more differentiated answers concerning the type of school the partner of the respondent is visiting at time of interview while information on vocational schooling is missing in wave 2. That is why in wave 2 the same coding as in wave 1 was applied for vocational training. Another difference between information on anchors and partners in wave 2 derives from the fact that concerning enrollment in vocational training, we can differentiate only between vocational schooling (non-specific), vocational training, and enrolled in university (including Ph.D.) for partners. Wave 3 to wave 5 data is more detailed in respect of enrollment in vocational training. The category “-3 Does not apply” only applies for anchors without partners. The syntax used to compute the variables *enrol* and *penrol* is contained in the Stata do-file *education.do*, which is provided for each wave.

Highest school degree - *school*, *pschool*, *mschool*, *fschool*

The variables *school* and *pschool* contain information about the highest school degree achieved by the anchor and his or her partner. The information on highest school degree is updated every year if the anchor and his or her partner have achieved another degree. A general dominance rule was used to decide which information given should be used: 1. information on school degree, 2. currently enrolled in general schooling (without information on school degree), 3. left school without degree, 4. incomplete data. By applying this dominance rule, more concrete information is always preferred to less concrete information. Note that a new degree overwrites an old degree only if it is higher. If multiple school degrees are mentioned, the highest degree is used.

In wave 2 and wave 4 only proxy information given by the anchor was available to construct the variable *pschool*. In wave 1, wave 3 and wave 5, the partner is questioned about his/her highest school degree. This direct partner information has priority over the indirect information the anchor gave for generating the variable *pschool*.

Please note that there probably is a certain portion of (not identifiable) respondents who were enrolled in general schooling at the time of interview and who had completed a degree earlier, but nevertheless were classified as students. This is due to the filtering in wave 1, where these respondents were not further asked about their school attainment, and in wave 2 to wave 5, where only respondents who finished at least one episode of education in the EHC activity calendar were asked if and what kind of degree they had achieved. Information on these respondents will become available only in later waves, as they leave the educational system. The category “-3” applies only for anchors without partners. The Stata do-file *education.do* (one for each wave) contains the syntax used to compute both variables.

Note that the codes “4: Intermediate GDR, POS 8./9.” and “5: Intermediate GDR, POS 10.” are only asked in wave 1 for the anchor. The highest school degree of the partner always contains the mentioned codes.

Table 4.9: Values and labels of variables *school*, *pschool*, *mschool*, *fschool*

Value	Label
-7	Incomplete data
-3	Does not apply
0	Currently enrolled
1	Left school w/o degree
2	Lower, Volks-/Hauptschulabschluss
3	Intermediate, Realschulabschluss / mittlere Reife
4	Intermediate GDR, POS 8./9.
5	Intermediate GDR, POS 10.
6	Upper, Fachhochschulreife / FOS
7	Upper, allg. Hochschulreife / EOS
8	Other school degree

Information on highest school degree is also provided for respondents' biological and adoptive parents (variables *mschool/fschool*) in each wave. If parents' information is available in multiple waves the newest information is used. In wave 4, only parents who did not participate in wave 2 or wave 3 or who did not indicate their school degree were asked again. Proxy information on parents' highest school degree given by the anchor in wave 3 is used to complete the variables *mschool* and *fschool* for respondents without direct information from their parents in wave 2 to wave 5. The variables were computed using the do-file *education_parents.do*.

Highest vocational degree - *vocat*, *pvocat*, *mvocat*, *fvocat*

The highest vocational degree completed by the anchor and his or her partner is stored in variables *vocat* and *pvocat*, respectively. Only proxy information given by the anchor was available for the partner and was therefore used to construct the variable *pvocat*.

The information on highest vocational degree is updated every year if the anchor and his or her partner have achieved another degree. A general dominance rule similar to that for the variables *school/pschool* was imposed: 1. information on vocational degree, 2. currently enrolled in general or vocational schooling, 3. no degree, 4. incomplete data. By applying this dominance rule, more concrete information is always preferred to less concrete information. Note that a new degree overwrites an old degree only if it is higher. If multiple vocational degrees are mentioned, the highest degree is used. The category "-3" applies for anchors without partners. The variables were generated by running the wave-specific do-file *education.do*.

Information on highest vocational degree is also provided for respondents' biological and adoptive parents (variables *mvocat/fvocat*). If parents' information is available in multiple waves the newest information is used. In wave 4, only parents who did not participate in wave 2 or wave 3 or who did not indicate their vocational degree were asked again. Proxy information on parents' highest vocational degree given by the anchor in wave 3 is used to complete the variables *mvocat* and *fvocat* for respondents with no direct information from their parents in wave 2 to wave 5. The category "-3" applies for anchors with stepparents. The variables were computed using the do-file *education_parents.do*.

Table 4.10: Values and labels of variables *vocat*, *pvocat*, *mvocat*, *fvocat*

Value	Label
-7	Incomplete data
-3	Does not apply
0	Currently enrolled
1	No degree
2	Vocational training (Lehre)
3	Vocational school (Berufsfachschule, Handelsschule, Schule d. Ges.wesens)
4	Technical school (Fachschule)
5	Civil service training (Beamtenausbildung)
6	Technical college (Fachhochschule, Berufsakademie)
7	University
8	Doctoral degree

ISCED classification of educational attainment - *isced*, *pisced*, *miscd*, *fiscd*, *iscd2*, *pisced2*

The generated variables *isced*, *pisced*, *miscd* and *fiscd* carry information on anchors', partners', and parents' educational attainment classified according to the International Standard Classification of Education (ISCED-97) (UNESCO 2006). We applied a revised version of the scheme adapted to

the German institutional context as proposed by Schneider (2008). While the original scheme was designed to classify persons enrolled in the educational system, the revision applied here aims to classify persons according to their educational attainment. Hence, persons currently enrolled were assigned value “0”, and the ISCED-categories were applied only to persons who had earned an educational certificate. As suggested by Schneider (2008), we introduced a further category “1b” (value “1”) for persons who did not achieve any formal degree. The variables were computed using the do-files `education.do` (available for each wave) and `education_parents.do`.

The variables `isced` and `pisced` provided the basis for two additional variables (`isced2`, `pisced2`), where afterwards respondents and partners currently enrolled in the German educational system were included. Thus it was assumed that those currently enrolled would complete their education and attain the corresponding degree. This can be helpful, since the percentage of those currently enrolled was at approximately 30 percent. Values and labels of the resulting variables are shown in table 4.11. The syntax used to compute these variables is contained in the Stata do-files `education.do` (available for each wave) and `education_parents.do`.

Table 4.11: Values and labels of variables *isced*, *pisced*, *misced*, *fiscd*

Value	Label
-7	Incomplete data
-3	Does not apply
0	Currently enrolled
1	No degree (1b)
2	Lower secondary education (2b)
3	Lower secondary education (2a)
4	Upper secondary education vocational (3b)
5	Upper secondary education general (3a)
6	Post-secondary non tertiary education general (4a)
7	First stage of tertiary education (5)
8	Second stage of tertiary education (6)

CASMIN classification of educational attainment - `casmin`, `pcasmin`, `mcasmin`, `fcasmin`

Variables `casmin`, `pcasmin`, `mcasmin`, and `fcasmin` contain respondents' educational attainments according to the classification schema originally set up by the project on Comparative Analysis of Social Mobility in Industrial Nations (CASMIN) (König et al. 1988). We applied the updated version of the original classification (Brauns and Steinmann 1999). Both school and vocational degrees were considered, i.e., the variables used to assign values of the CASMIN categories were the generated variables `school` and `vocat` (`pschool` and `pvocat` for the partner, `mschool/fschool` and `mvocat/fvocat` for the parents). The complete classification schema - including additional values “-7 Incomplete data” and “0 Currently enrolled” - is shown in table 4.12. The category “-3” applies for anchors without partners and anchors with stepparents. The syntax used to compute these variables is contained in the Stata do-files `education.do` (one for each wave) and `education_parents.do`.

Years in education, schooling and vocational qualification - `yeduc`, `pyeduc`, `myeduc`, `fyeduc`

Variables `yeduc`, `pyeduc`, `myeduc`, and `fyeduc` are linear measures of the anchor's, his or her partner's, and parents' educational attainment. Values assigned were derived from the typical duration of educational careers, i.e., values are the years it usually takes to earn a specific degree. With regard to school attainment, we assigned

- 8 years of education for persons who left school without a degree,

Table 4.12: Values and labels of variables *casmin*, *pcasmin*, *mcasmin*, *fcasmin*

Value	Label
-7	Incomplete data
-3	Does not apply
0	Currently enrolled
1	Inadequately completed (1a)
2	General elementary education (1b)
3	Basic vocational qualification (1c)
4	Intermediate vocational qualification (2a)
5	Intermediate general qualification (2b)
6	General maturity certificate (2c_gen)
7	Vocational maturity certificate (2c_voc)
8	Lower tertiary education (3a)
9	Higher tertiary education (3b)

- 9 years for lower secondary degree (Volks-/Hauptschulabschluss, Abschluss Polytechnische Oberschule 8./9.),
- 10 years for intermediate secondary degree (Realschulabschluss, mittlere Reife, Abschluss Polytechnische Oberschule 10., other school degree)
- 12 years for upper secondary degree, vocational track (Fachhochschulreife, Abschluss Fachoberschule),
- 13 years for upper secondary degree, general track (allgemeine Hochschulreife, erweiterte Oberschule).

To years of schooling, we added for vocational qualifications

- 1.5 years for vocational training (Lehre) or civil service training (Beamtenausbildung),
- 2 years for vocational school (Berufsfachschule, Handelsschule, Schule des Gesundheitswesens),
- 3 years for technical school (Fachschule, Meister-/Technikerabschluss),
- 4 years for technical college (Fachhochschulabschluss/Berufsakademie),
- 5 years for university degree,
- 7 years for doctoral degree.

Hence, valid values of *yeduc*, *pyeduc*, *myeduc*, and *fyeduc* range from “8 No school degree and no vocational degree” to “20 Upper secondary and doctoral degree”. Persons without information on either schooling or vocational qualification were assigned value “-7 Incomplete data”. Respondents who are enrolled in education were assigned value “0 Currently enrolled”. The variables were computed using the do-files *education.do* (available for each wave) and *education_parents*.

KldB classification of occupation - *kldb2010*

Variable *kldb2010* contains information according to the classification of occupations proposed by the German Statistical Office (Klassifikation der Berufe, KldB). The KldB schema was designed to fit the German occupational system better than the international ISCO scale does (see below). The classification is based on variable *job1* in wave 1 and *job200* in waves 2 to 5, which originally carried open answers about the anchor’s current occupation (The original answers have been made anonymous as described in section 4.1). Note that in waves 1 to 5, release 5.0, variable *kldb2010* is based on the new 5-digit classification schema KldB 2010. Variable *kldb1992* based on the former 4-digit classification is additionally provided in the anchor data sets of waves 1 to 3, release 5.0.

Variable *flag_isco08_kldb2010*¹³ indicates that a respondent's answer was ambiguous and could not be recoded to exactly one KldB (or ISCO) code, but could equally well have been assigned two or more codes. The variable distinguishes between two coding problems. Value "1" indicates that the occupation requiring the lowest level of qualification has been assigned. Value "2" flags respondents where the most frequent of all possible occupations has been assigned. Recoding was done by *TNS Infratest*. Variable *flag_isco08_kldb2010* based on the new classification schema KldB2010 (ISCO-08) is provided in waves 1 to 5, release 5.0. (see Hartmann et al. (2012) for more details). Variable *flag_isco88_kldb1992* based on the former classification is provided in wave 1, release 5.0 (see Hartmann et al. (2010), Hartmann et al. (2011a), and Hartmann et al. (2011b) for more details).

ISCO classification of occupation - isco08

Variable *isco08* contains the anchor's occupation classified according to the 4-digit International Standard Classification of Occupations (ISCO-08) schema. The variable was derived from variable *kldb2010* (see above). Recoding was done by *TNS Infratest* (see Hartmann et al. (2012) for details). Note that in waves 1 to 5, release 5.0, the variable *isco08* is based on the new classification ISCO-08. Variable *isco88* based on the former classification ISCO-88 is additionally provided in the anchor data sets of waves 1 to 3, release 5.0 (see Hartmann et al. (2010) for details).

EGP class schema - egp

The variable *egp* contains information on the anchor's social class according to the Erikson-Goldthorpe-Portocarero (EGP) class schema (Erikson et al. 1979). In constructing the variable, we applied the revised schema developed by Ganzeboom and Treiman (2003). EGP categories were assigned based on the ISCO-88 codes first. In a second step, two variables indicating self-employment and supervisory status were used to differentiate further within occupations. The full procedure is described in Ganzeboom and Treiman (2003). Please note that information on the assignment of Erikson-Goldthorpe-Portocarero (EGP) class category codes to ISCO-08 codes is not yet available. The variable *egp* is therefore only provided for waves 1 to 3 and is based on ISCO-88 codes.

When using the variable *egp*, please take into account that we used proxy information on supervisory function derived from the occupational status (variable *job2*). At this point, the procedure relies on quite strong assumptions.¹⁴ The Stata code we used to compute the variable *egp* is available as part of the scientific use file and online (see do-file *egp.do*).

ISEI occupational prestige score - isei

The variable *isei* is a prestige measure carrying information on the respondents' prestige according to the International Socio-Economic Index of Occupational Status (ISEI). Variable *isei* was computed by assigning prestige values as described by Ganzeboom and Treiman (2010). The assignment of values is documented in do-file *isei.do*. Note that in waves 1 to 5, release 5.0, coding of the variable *isei* is based on the new classification ISCO-08.

SIOPS occupational prestige score - siops

Variable *siops* is a prestige measure containing information on the respondents' prestige as determined by Treiman's Standard International Occupational Prestige Scale (SIOPS). The variable was computed by assigning prestige values as detailed in Ganzeboom and Treiman (2010). The assignment of scores is documented in do-file *siops.do*. Note that in waves 1 to 5, release 5.0, coding of the variable *siops* is based on the new classification ISCO-08.

¹³This variable is a renamed copy of variable *beruprob*, which was originally provided by *TNS Infratest*.

¹⁴More specifically, for employed persons, we assumed that the following positions indicated supervision of 1 to 9 employees: civil servant following the upper career track (*gehobener Dienst*), industry and works foreman (*Industriemeister*), employees with highly qualified duties or managerial functions, foreman, and master craftsman. For civil servants on the higher career track (*höherer Dienst*) and for employees with extensive managerial duties, we assumed supervision of 10 or more employees.

MPS occupational prestige score - mps

Wegener's Magnitude Prestige Scale (MPS) is an alternative to the ISCO-based internationally comparable ISEI and SIOPS prestige measures. It was particularly designed to better fit the German occupational structure (Wegener 1984). Variable *mps* contains prestige values based on the variable *kldb1992*. The procedure is described in Frietsch and Wirth (2001). The assignment of values is documented in do-file *mps.do*. Please note that information on the assignment of Wegener's Magnitude Prestige Scale (MPS) prestige values to KldB 2010 codes is not yet available. The variable *mps* is therefore only provided for waves 1 to 3 and is based on the former KldB 1992 classification.

Net equivalence income according to the German Council of Economic Experts (GCEE) - hhincgee

For this variable, the net equivalence income according to the German Council of Economic Experts (GCEE) was generated by dividing the household net income (*hhincnet*) by the square root of the size of the household at the main residence. If the household size or the household income was missing, the net equivalence income was coded to "-7 Incomplete data". The syntax used to compute the variable is contained in Stata do-file *incnet.do*.

Equivalence income according to the (modified) OECD equivalence scale assigns a value of 1 to the first household member, of 0.7 (0.5) to each additional adult and of 0.5 (0.3) to each child below the age of 14. Hence, calculation requires information on the age of all household members. For partners, children, and parents living in the household the variables *page*, *k*age*, *mage*, *fage*, *smage*, *sfage* can be used. Information for calculating the age of other household members can be found in the do-files *household2.do* and *household3.do* and *age.do*.

Personal and household net income - incnet, hhincnet

The variables *incnet* and *hhincnet* contain the information of both open-ended and categorized answers to questions on personal and household net income. The categorized answers were coded as midpoints of the categories. At the higher end, personal net income was coded 2,500 Euro for wave 1, and 5,000 Euro as of wave 2. For the household net income the value 5,000 Euro was assigned at the higher end for both waves. The codings are shown in table 4.13 and table 4.14. If an individual was not gainfully employed the value "-3 Does not apply" was assigned to the variable *incnet*.¹⁵ Where the information was not available to derive the income, the code "-7 Incomplete data" was given. The syntax used to compute the variables is contained in the Stata do-file *incnet.do*.

Table 4.13: Coding scheme for the categorized personal and household net income (wave 1)

personal net income		household net income	
Categories	Codes	Categories	Codes
[0, 400)	200	[0, 800)	400
[400, 600]	500	[800, 1150]	975
(600, 750)	675	(1150, 1450)	1300
[750, 900)	825	[1450, 1700)	1575
[900, 1100]	1000	[1700, 2000]	1850
(1100, 1350)	1225	(2000, 2300]	2150
(1350, 1600)	1475	(2300, 2800)	2550
[1600, 2000]	1800	[2800, 3500]	3150
(2000, ∞)	2500	(3500, ∞)	5000

¹⁵This was not necessary for variable *hhincnet* as the question concerning the household income was not filtered, i.e., was posed to all anchors.

Table 4.14: Coding scheme for the categorized personal and household net income (since wave 2)

personal net income		household net income	
Categories	Codes	Categories	Codes
[0, 250)	125	[0, 250)	125
[250, 500)	375	[250, 500)	375
[500, 750)	675	[500, 750)	675
[750, 1000)	825	[750, 1000)	825
[1000, 1250)	1125	[1000, 1250)	1125
[1250, 1500)	1375	[1250, 1500)	1375
[1500, 1750)	1625	[1500, 1750)	1625
[1750, 2000)	1875	[1750, 2000)	1875
[2000, 2500)	2250	[2000, 2500)	2250
[2500, 3000)	2750	[2500, 3000)	2750
[3000, 3500)	3250	[3000, 3500)	3250
[3500, 4000)	3750	[3500, 4000)	3750
[4000, 4500)	4250	[4000, 4500)	4250
[4500, ∞)	5000	(4500, ∞)	5000

Consumer price index - *cpi*

The variable *cpi* contains the consumer price index on a month-to-month basis according to the National Statistical Office. The information was taken from a query of www.destatis.de. Note that this variable was constant for all observations. The syntax used to compute the variable is contained in the Stata do-file *cpi.do*. In waves 1 to 5, release 5.0, coding of the variable *cpi* is based on the new baseline 2010.

Current primary and secondary activity status (anchor, partner) - *casprim*, *cassec*, *pcasprim*, *pcassec*

The variables *casprim*, *cassec*, *pcasprim*, and *pcassec* contain the anchor's and his or her partner's current primary and secondary activity status. For the anchor, the information was taken from the anchor interview. Information on the partner was taken from the partner questionnaire if available, otherwise from the anchor interview. This, however, was only possible for wave 1, wave 3 and wave 5. In wave 2 and wave 4, we had only proxy information given by the anchor person. Concerning educational attainment, this information is less rich than for anchors.¹⁶

The possibly multiple activities were reduced to the two most important ones. For the anchor this reduction was accomplished by a series of reductions that follow specific dominance rules. The same rules were used for partners.

The following dominance rules were applied:

- If only one activity was stated, it was defined as the primary activity status.
- If more than one activity and
 - only employment activities were stated:
 - primary activity was defined according to the dominance rule:
 vocat. training > full-time empl. > self-employ. > part-time empl. > marginal empl. > internship > occasionally empl. > other type of job
 - only educational activities were stated¹⁷:
 - primary activity is defined according to the dominance rule:

¹⁶For partners, we only know if they are involved in general schooling, vocational training, or university.

¹⁷Due to filtering and question wording partners cannot have multiple education activities.

- general secondary school > UCE¹⁸ > college/university > evening school > technical/professional school > pre-vocat. training > vocat. retraining > other education
- only unemployment activities were stated:
primary activity was defined according to the dominance rule:
maternity/paternity leave > military service > unemployed > retired > housewife > other, non employed
 - both employment and unemployment activities were stated:
primary activity was defined according to the dominance rule:
maternity/paternity leave > military service > unemployed > retired > vocat. training > full-time employ. > self-employ. > part-time empl. > marginal empl. > internship > housewife > occasionally empl.
 - both unemployment and educational activities were stated:
primary activity was defined according to the dominance rule:
maternity/paternity leave > military service > unemployed > retired > general secondary school > UCE > college/university > evening school > technical/professional school > pre-vocat. training > vocat. retraining > housewife
 - both employment and educational activities were stated:
primary activity was defined according to the dominance rule:
general secondary school > UCE > college/university > evening school > technical/professional school > pre-vocat. training > vocat. retraining > vocat. training > full-time empl. > self-empl. > part-time empl. > marginal empl. > internship > occasionally empl.
 - employment, unemployment, and educational activities were stated:
primary activity was defined according to the dominance rule:
maternity/paternity leave > military service > unemployed > retired > general secondary school > UCE > college/university > evening school > technical/professional school > pre-vocat. training > vocat. retraining > vocat. training > full-time empl. > self-empl. > part-time empl. > marginal empl. > internship > housewife > occasional empl.
- If more than one activity was stated, the secondary activity was defined according to the same dominance rules and one additional dominance rule:
if only activities out of “Other education”, “Other type of job” and “Other, not employed”:
“Other, not employed” > “Other education” > “Other type of job”
 - If at least one employment activity was stated, but neither primary nor secondary activity status, case-by-case decision under consideration of occupation, occupational status, work hours, educational attainment, personal income, and age¹⁹.

After this reduction process, the defined activity statuses were checked for inconsistencies. In table 4.15 inconsistent combinations of the primary and secondary activity status are marked with an X. These cases are marked in the data with the flag variable *flag_cas*.

If only one activity status was given, the secondary activity status was coded as “-3”. If there was no partner, the partner’s activity statuses were coded as “-3”. If no information was available to code an activity status, it was coded as “-7”. The syntax used to compute the variables is also contained in Stata do-file *cas.do*.

¹⁸University of Cooperative Education (“Berufsakademie”)

¹⁹For partners only occupational status, educational attainment, work hours, and age are considered.

Table 4.15: Inconsistent activity status combinations

Codes	Codes																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	X		X	X	X	X	X	X		X								X	X		X	
2		X																				
3	X		X		X					X									X		X	
4	X			X	X		X	X										X				
5	X		X	X	X	X	X	X		X								X			X	
6	X				X	X	X	X										X				
7	X			X	X	X	X			X								X			X	
8	X			X	X	X		X		X								X			X	
9									X													
10	X		X		X		X	X		X								X	X		X	
11											X											
12												X						X	X			
13													X									
14														X								
15															X							
16																X						
17																	X	X			X	
18	X			X	X	X	X	X		X		X					X	X	X			
19	X		X							X		X						X	X			
20																				X		
21	X		X		X		X	X		X							X				X	
22																						X

Labor force status (anchor, partner) - lfs, plfs

The variables *lfs* and *plfs* contain the anchor's and his or her partner's labor force status. The variables were directly derived from the activity statuses (see above) in the following way:

Employment activities are defined as one of the following activities:

- vocational training,
- full-time employment,
- self-employment,
- part-time employment,
- internship, trainee, work experience etc.,
- marginal part-time employment, mini-job, "Ein-Euro-Job" ("one-euro job", when receiving unemployment benefits),
- occasionally or irregularly employed,
- other type of job.

If only the primary or only the secondary activity status was an employment activity, this was taken for the labor force status. If both activity statuses were employment activities, the primary activity status was used. If both activity statuses were unemployment activities, the primary activity status was used. This derived single activity status was recoded to the labor status. The coding scheme is shown in table 4.16.

Table 4.16: Coding scheme for labor force status

Label	Code	Recoded activity status codes
nw, education	1	1,2,4,...,9
nw, parental leave	2	17
nw, homemaker	3	20
nw, unemployed	4	19
nw, military service	5	18
nw, retired	6	21
nw, other	7	22
w, vocational training	8	3
w, full-time employment	9	10
w, part-time employment	10	12
w, marginal employment (geringfügige Beschäftigung)	11	14
w, self-employed	12	11
w, other	13	13,15,16

For the anchor, an inconsistency check was conducted. If the labor force status was “w, vocational training” and working hours, occupation, and the occupational status were missing, the labor force status was coded “nw, education”. If the labor force status was another working category and working hours, occupation, and the occupational status were missing, and more than one activity was stated, a case-by-case decision about a possible recoding was employed, considering activities, occupation, occupational status, work hours, and income.

If there was no partner, the partner’s labor force status was coded as “-3”. If no information was available to code a status, it was coded as “-7”. The syntax used to compute the variables is also contained in the Stata do-file *lfs.do* (available for each wave). Furthermore, researchers also need the Stata do-file *cas.do*, as the labor force status is based on the primary and secondary activity statuses.

Currently living in Eastern Germany - east

To differentiate whether the anchor was currently living in Eastern or Western Germany, we generated the dummy variable *east*. Value “1” indicates that the anchor was living in Eastern Germany, while “0” means that the anchor resided in Western Germany. Note that the computation of this variable is based on the process-generated variable *bula* (see section 4.3). The relevant do-file *east.do* is available for each wave.

Scales

All of the previously described generated variables are included in the anchor data sets. In contrast, scale variables (see table 4.17) are not part of the delivered data set. They can be generated by running the syntax files *scales_anchor\$* (available in Stata and SPSS format), which are available for each wave as part of the scientific use file and on the pairfam website. For further information, refer to the scales handbook (Thönnissen et al. 2014), which contains a detailed description of all scales.

Table 4.17: List of scales included in data set *anchor\$*

Construct	Variable name	Wave
[t] Traditional concept of marriage	tradmarr	1,3,5
Value of Partnership - Negative expectations	vopneg/vopneg2	1/3,5
Value of Children: Benefit of stimulation	vocbstim	1,2,4
Value of Children: Costs of comfort	vocccomf	1,2,4
Single: Interest in partnership	partint	1 - 3
Single: Desire for partnership	partdes	all
Single: Broad exploration	siexplbr/siexplbr2	1,2/3 - 5
Single: In-depth exploration	siexplde	all
Single: Mating confidence	chanpm	all
Getting to know each other: In-depth exploration	npexplde	all
Ambivalence: moving in together	ambcoh	1,3,5
Ambivalence: marriage	ambmarr	1,3,5
Partnership: Conflict	confl_apd	all
Partnership: Intimacy	intim_aps	all
Partnership: Admiration	admir_apo	all
Partnership: Dominance	domin_apo	all
Partnership: Emotional ambivalence	ambiv_apd	1 - 3,5
Partnership: Anxiety about loss of love	lovewitanx_apd	1 - 3,5
Partnership: Anxiety about being absorbed	enganx_apd	1 - 3,5
Partnership: Autonomy	indep_apd	1 - 3,5
Partnership: Feelings of competence in the partnership	comppart/comppart2	1/2,3,5
Partnership: Own partnership satisfaction (global scale)	satpart	1,3,5
Partnership: Instability of partnership	instab_apd	all
Partnership: Readiness to sacrifice Self	sacrif_aps	1,3,5
Partnership: Hostile attribution Self	hostattr_aps	1,3,5
Partnership: Future orientation	comfut_apd	1 - 3,5
Partnership: Tolerance of conflicts	comctol_apd	1 - 3,5
Partnership: Orientation of reciprocity Self	reciproc_aps	1,3,5
Partnership: Frequency of manifest conflicts	confldom_apd	1,5
	confldom2_apd	3
Partnership: Verbal aggression Partner	verbaggr_apo	all
Partnership: Verbal aggression Self	verbaggr_aps	all
Partnership: Constructive behavior Partner	constrbh_apo	all
Partnership: Constructive behavior Self	constrbh_aps	all
Partnership: Withdrawal Partner	withdraw_apo	all
Partnership: Withdrawal Self	withdraw_aps	all
Partnership: Manipulation Partner	manipul_apo	1,3,5
Partnership: Manipulation Self	manipul_aps	1,3,5

continued on next page

continued from previous page

Construct	Variable name	Wave
Partnership: Dyadic coping Partner	dycop_apo	1,3,5
Partnership: Dyadic coping Self	dycop_aps	1,3,5
Intergenerational Relationships: Conflict with mother	confl_amd	all
Intergenerational Relationships: Conflict with partner of mother	confl_asfd	2 - 5
Intergenerational Relationships: Conflict with father	confl_afd	all
Intergenerational Relationships: Conflict with partner of father	confl_asmd	2 - 5
Intergenerational Relationships: Intimacy mother	intim_ams	all
Intergenerational Relationships: Intimacy partner of mother	intim_asfs	2 - 5
Intergenerational Relationships: Intimacy father	intim_afs	all
Intergenerational Relationships: Intimacy partner of father	intim_asms	2 - 5
Intergenerational Relationships: Admiration mother	admir_amo	2,4
Intergenerational Relationships: Admiration partner of mother	admir_asfo	2,4
Intergenerational Relationships: Admiration father	admir_afo	2,4
Intergenerational Relationships: Admiration partner of father	admir_asmo	2,4
Intergenerational Relationships: Dominance mother	domin_amo	2,4
Intergenerational Relationships: Dominance partner of mother	domin_asfo	2,4
Intergenerational Relationships: Dominance father	domin_afo	2,4
Intergenerational Relationships: Dominance partner of father	domin_asmo	2,4
Intergenerational Relationships cohort1: Negative Communication Mother/Partner of father	negcomm_amo	2
Intergenerational Relationships cohort1: Negative Communication Father/Partner of mother	negcomm_afo	2
Intergenerational Relationships cohort1: Successful individuation Mother/Partner of father	sucindiv_amd	2
Intergenerational Relationships cohort1: Successful individuation Father/Partner of mother	sucindiv_afd	2
Intergenerational Relationships cohort1: Fear of love withdrawal Mother/Partner of father	lovewitanx_amd	2
Intergenerational Relationships cohort1: Fear of love withdrawal Father/Partner of mother	lovewitanx_afd	2
Intergenerational Relationships cohort1: Ambivalence Mother/Partner of father	ambiv_amd	2
Intergenerational Relationships cohort1: Ambivalence Father/Partner of mother	ambiv_afd	2
Sexual competence	compsex	2,3
Sexual communication	sexcom	2,3
Explosiveness and tendency to anger	explosive	1,5
Shyness	shyness	1,5

continued on next page

continued from previous page

Construct	Variable name	Wave
Emotional autonomy	emotautn	1,5
Self-worth	selfworth	all
Depressiveness	depressive	2 - 5
Anger	anger	3 - 5
Activity	activ/activ2	3/4,5
Stress	stress	4,5
Anxiety	anxiety	4,5
Prevention	prevent	4
Promotion	promot	4
BIG 5: Neuroticism	neurot	2
BIG 5: Extraversion	extrav	2
BIG 5: Agreeableness	agreeable	2
BIG 5: Conscientiousness	conscient	2
BIG 5: Openness	openness	2
Economic deprivation parents	ecodep_par/ecodep2_par	2/3,4
Economic deprivation anchor	ecodep_a/ecodep2_a	2/3,4
Economic deprivation household	ecodep_hh/ecodep2_hh	2/3 - 5
Newborn temperament child1	temperc1	2 - 5
Newborn temperament child2	temperc2	2 - 5
Newborn temperament child3	temperc3	2 - 5
Newborn temperament child4	temperc4	4,5
Newborn temperament child5	temperc5	4,5
Newborn temperament child6	temperc6	4,5
Newborn temperament child7	temperc7	4,5
Unspecific strain	unspstrain	3
Autonomy in the parenting role (3 Item Scale)	autonom	2
Autonomy in the parenting role (4-Item Scale)	autonom	4
Autonomy in the parenting role (2-Item Scale)	autonom2	4
Pleasure in the parenting role (newborn module)	pleasure_newborn	2
Pleasure in the parenting role (children till 15 years)	pleasure	4
Parental Self Efficacy/Competence	comperz	2,4
Coparenting	coparent	2,4
Coparenting with ex-partner	coparent_ex	3
Coparenting with ex-partner1	coparent_exp1	5
Coparenting with ex-partner2	coparent_exp2	5
Coparenting with ex-partner3	coparent_exp3	5
Coparenting with ex-partner - capikid1	coparent_opk1	3,5
Coparenting with ex-partner - capikid2	coparent_opk2	3,5
Coparenting with ex-partner - capikid3	coparent_opk3	5
Coparenting with ex-partner - capikid4	coparent_opk4	5
Coparenting with ex-partner - capikid6	coparent_opk6	5

continued on next page

continued from previous page

Construct	Variable name	Wave
Parenting goals: Status	pgoalstatus	2 - 5
Parenting goals: Autonomy	pgoalautn	2 - 5
Parenting goals: Competence	pgoalcomp	2 - 5
Parenting: Partner support	partnersup	3,5
Parenting: Overprotection	overprotect	3,5
Parenting: Social support	socialsup	3,5
Parenting: Readiness to make sacrifices	sacrif_pacs	3,5
Demands on parenting: Perfectionism	pstandperfect	5
Demands on parenting: Pragmatism	pstandpragtam	5
SDQ Behavior: Conduct problems childx	conduct_aco1...	5
	conduct_aco7	
SDQ Behavior: Hyperactivity childx	hyper_aco1,....,	5
	hyper_aco7	
SDQ Behavior: Emotional symptoms childx	emotion_aco1,....,	5
	emotion_aco7	
Emotional warmth childx	warmth_acs1,....,	5
	warmth_acs1	
Negative communication childx	negcomm_acs1,....,	5
	negcomm_acs7	
Inconsistent parenting childx	inconsist_acs1,....,	5
	inconsist_acs7	

4.3 Process-generated variables

This chapter describes all process-generated variables listed in table 4.18.

Variables *bula*, *gkpol*, *gkz*, *bik* and the microm data contain information about the respondents' place of residence, i.e. the address where *TNS Infratest* contacted participants (*Meldeadresse*).

Table 4.18: List of process-generated variables included in data set *anchor\$*

Construct	Variable name
State / Bundesland	<i>bula</i>
Size of community in 7 categories	<i>gkpol</i>
Official key of the municipality	<i>gkz</i>
Settlement structure	<i>bik</i>

Variable *bula* contains information on the state (*Bundesland*).

Variable *gkpol* carries information on the size (population) of the community, divided into 7 categories (see table 4.19).

Table 4.19: Values and labels of variable *gkpol*

1	1,000 - 2,000 inhabitants
2	2,000 - 5,000 inhabitants
3	5,000 - 20,000 inhabitants
4	20,000 - 50,000 inhabitants
5	50,000 - 100,000 inhabitants
6	100,000 - 500,000 inhabitants
7	500,000 + inhabitants

Variable *gkz* contains the official key of the municipality (*Gemeindekennziffer*) where the anchor respondent resides.

Variable *bik* contains information on the settlement structure (*Siedlungsstruktur*, BIK-Typ), divided into 10 categories (see table 4.20).

Table 4.20: Values and labels of variable *bik*

0	City Center - population 500,000+
1	Periphery - population 500,000+
2	City Center - population 100,000-500,000
3	Periphery - population 100,000-500,000
4	City Center - population 50,000-100,000
5	Periphery - population 50,000-100,000
6	Region - population 20,000-50,000
7	Region - population 5,000-20,000
8	Region - population 2,000-5,000
9	Region - population < 2,000

Pairfam data can be enriched with data delivered by microm consumer marketing. This marketing research enterprise delivers information on the housing environment of street addresses gathered by a variety of sources in order to target and profile customers. Information is based on a group of addresses clustered around the respondent's actual street address. The data include the Mosaic Types, Mosaic Milieus and other lifestyle typologies, sociodemographic variables, types of residence, age, unemployment, and ethnic composition (see table 4.21).

Table 4.21: Microm information available

Mosaic Typology	Typology of neighborhoods
Mosaic Sozio	sociodemographic variables
Mosaic Bebauung	sizes of and number of commercially used properties
Mosaic Mobilität	likelihood and types of mobility
Mosaic Milieus	lifestyle typology of residents
Mosaic Lebensphasen	life phases of respondents
Wohnen	information on home ownership quota
Arbeitslosenquote	unemployment quota
Einwohner nach Altersklassen	age group of residents
Ethno	ethnic composition of neighborhoods
Marktzellentypologie	cluster of communities

Microm data and the variable *gkz* are not part of the scientific use file for licensing due to privacy concerns. Researchers interested in analyzing microm data and the official municipal key number (*Gemeindekennziffer*) can do so on-site at the available pairfam locations. For more information on on-site use please refer to the pairfam website.

4.4 Paradata

We added information on the survey process to the data set for all valid and completed interviews. Paradata are, however, somewhat limited in pairfam. The available variables are listed in table 4.22. For all addresses contacted, the total number of contacts was reported by the interviewers (variable *intcont*). Interviewer characteristics, gender, and age are on file (variables *intsex* and *intage*). Interviewer's school degree is available for wave 5 (variable *intedu*). The variables *intm*, *intd*, and *inty* contain information on the date of the interview (month, day, year). In addition, a unique number

identifies each interviewer (variable *intid*).

We also provided the length of the interview in minutes (variable *intdur*). In wave 1 the relevant variables for computing the total length were provided by the institute *TNS Infratest*. From this information, we constructed the interview duration. We trimmed the duration by recoding implausible values above 240 minutes or below 10 minutes to “-5 Inconsistent value”. For some respondents, the computer did not record the start or end of the interview completely.²⁰ In this case, we assigned value “-7 Incomplete data”. For wave 2 to 5 the duration in minutes was calculated and provided directly by the institute. If the duration seemed implausible according to specific criteria, it was not provided.²¹ Therefore, we could not differentiate between inconsistent values and incomplete data and assigned the value “-7 Incomplete data” in both cases.

Table 4.22: Available paradata in data set *anchor5*

Construct	Variable name
Total number of interviewer contacts with respondent	intcont
Interviewer’s sex	intsex
Interviewer’s age	intage
Interviewer’s school degree (wave 5)	intedu
Interviewer-ID	intid
Duration of CAPI interview in minutes	intdur
Date of interview (month, day, year)	intm, intd, inty

4.5 Weights

Weights are provided account for the disproportionate stratified sample and the systematic nonresponse in wave 1. By design, the three birth cohorts of the pairfam panel are of about equal size (c.f. Huinink et al. (2011)). The proportions in the respective population are naturally different across the cohorts. Therefore, the sample of the first wave is a disproportionate stratified sample. Within the birth cohorts, however, the sample is random. Furthermore, the overall nonresponse rate in the first wave is about 35% and is distributed unequally among the sample. To enable the data user to correct for the resulting discrepancies from the population, we provide three weighting factors:

dweight Design weight to correct for disproportionate sampling across cohorts

psweight Post-stratification weight to correct for systematic nonresponse

dxpsweight Combined design and post-stratification weight.

For the DemoDiff sample with release 5.0 the design weight *ddweight*, the post-stratification weight *caweight* and the combined design and post-stratification weight *ddcaweight* were computed according to the weights in the pairfam sample. For analyses based on the combined pairfam and DemoDiff sample, we provide the design weight *d1weight*, the post-stratification weight *ca1weight* and the combined design and post-stratification weight *d1ca1weight*. For details see Brix and Pfister (2014).

In addition to these cross-sectional weights, we provide a longitudinal weight *lweight* to account for attrition.

²⁰Some digits were missing, and it was not possible to recover the information whether it was the hours or the seconds missing.

²¹Duration was classified implausible if the duration of a single module exceeded 60 minutes, if the interview was not completed on the same day, or if interviewers indicated by comments or notes that the length of interview measured by the computer was wrong for some reason (long telephone call by respondent, dinner for children prepared, etc.).

Design weight

The design weight is the factor by which the birth cohorts are under- or overrepresented in the gross sample as compared to the population. According to German National Statistics Office data²², the population sizes of the three birth cohorts 1991-1993, 1981-1983 and 1971-1973 are 2,544,451, 2,994,725 and 3,126,100, respectively. According to the methods report (Suckow and Schneekloth 2009), the gross respective sample sizes are 9,648, 16,810 and, 15,616. Considering the reduced net sample size, the design weight factors are therefore 1.225, .828, and .930.²³

Post-stratification weights

The post-stratification weight is a raking weight. The weighting factor is calculated with an iterative procedure so that the marginal distributions of the weighting variables are adjusted to the corresponding population distributions.²⁴ The population distributions are taken from a special analysis of the German census Mikrozensus 2008. The weighting variables are federal states, age, gender, urban agglomeration type BIK (ten groups), marital status (with the categories unmarried, married, divorced and widowed), and whether the respondent has children. From these variables seven marginal distributions are calculated and used to construct the weighting factors:

- Federal state x gender x age (96 cells)
- Federal state x BIK (121 cells)
- BIK (10 cells)
- Federal state (16 cells)
- Marital status x gender (8 cells + 1 cell for “no answer” to marital status)
- Marital status x BIK (30 cells + 1 cell for “no answer” to marital status)
- Federal state x gender x having children (64 cells)
- Gender x age (6 cells).

The two distributions that include marital status were not used for the youngest cohort, as this cohort was almost completely unmarried. In the birth cohort 1981-1983 the categories “divorced” and “widowed” were combined to avoid cells which were too small. Due to data privacy regulations, cell sizes under 5,000 are not presented in the Mikrozensus data. Therefore, some combinations of federal state and BIK are not provided in the Mikrozensus data and cannot be used to calculate these factors. Here the affected combinations were combined. Further, the variable “has children” was used only for women of the birth cohorts 1971-1973 and 1981-1983, as respondents of the youngest cohort almost never had children and for men this information was not available.²⁵ The degree to which the factors fit to the distribution of these variables is shown in table 4.23.

²²Query at www.destatis.de at 01/20/2010.

²³The weighting factor w_i for the cohort i is calculated the following way:

$$w_i = \frac{N_{i,P} N_{T,NS}}{N_{i,GS} \sum_{j=1}^3 \left(\frac{N_{j,P} N_{j,NS}}{N_{j,GS}} \right)}$$

with $N_{i,P}$ being the size of cohort i in the population, $N_{T,P}$ the total size of all cohorts in the population, $N_{i,GS}$ the size of cohort i in the gross sample, $N_{T,GS}$ the total size of all cohorts in the gross sample, $N_{i,NS}$ the size of cohort i in the net sample, and $N_{T,NS}$ the total size of all cohorts in the net sample.

²⁴According to the methods report by Suckow and Schneekloth (2009), an iterative proportional fitting method is used. As a constraint additional to the marginal distributions of the weighting variables, the weighting factors have to lie within the interval [0.2, 4.9]. Further, the weighting procedures are calculated separately for each birth cohort. Note, that the differences between the figures in this text and those in the methods report are the result of an updated calculation of the weighting factors after the methods report was issued.

²⁵Note that this information was not derived from the problematic question about the number of children in the household but from the question to female respondents about all children ever born, which was added in the Mikrozensus 2008 (c.f. Statistisches Bundesamt 2009).

Table 4.23: Degree of fitness to weighting variable distributions

	Birth cohorts		
	1991-1993	1981-1983	1971-1973
Federal state x gender x age	99.56%	99.26%	99.50%
Federal state x BIK ²⁶	94.58%	96.05%	96.28%
BIK	98.96%	99.73%	99.82%
3Federal state	99.99%	99.50%	99.83%
Marital status x gender		99.86%	99.83%
Marital status x BIK		99.92%	99.95%
Federal state x gender x having children		99.87%	99.99%
Gender x age	100.00%	100.00%	100.00%
Number of iterations	9	16	8
Effectiveness ²⁷	85%	77%	75%

Combined weight

The combined weight combines the information of the design weight and the post-stratification weighting factor. It was computed as the product of the two factors, considering that the combined weight sums to the net sample size. For further information on the computation of the combined weights, see the Stata do-file `weight.do`.

Longitudinal weight

The longitudinal weight aims to correct for attrition. Attrition is defined as any pattern of loss of individual records over time, i.e. over waves. The weighting factor is based on interview characteristics of both the previous and the current wave. In wave 5 the weighting factors of temporary dropouts (people who skip a maximum of one wave) refer to wave 4 characteristics. The calculation is done in three stages. Consequently, there are three predicted probabilities:

- probability of acceptance of survey participation given in the previous wave for the current wave (*ppanel*)
- probability of contact, given an accepted survey participation (*pcontact*)
- probability of response, given a successful contact (*panswer*).

These probabilities are results of stepwise logistic regressions, which were run separately for each birth cohort. Note that the independent variables of the regression models vary in their categories between each cohort (see table A.20, Appendix). Furthermore, to control for temporary dropouts the variable *temp_dropout* is added to the logistic regression models. The inverse of the three multiplied probabilities defines the longitudinal weight. Values greater than 5 are set to 5.

With release 5.0 the longitudinal weight is also available for DemoDiff waves 3, 4, and 5. Sample information is added to the logistic regression models of the respective waves (see table A.20, Appendix).

Best implementation of weights

We strongly recommend using the design weight *if and only if* you combine more than one birth cohort in your analyses. Thus, if you use cases from more than one cohort and you do not allow for a

²⁶The low fitness to the marginal distribution of federal state x BIK in all three cohorts follows from the necessary cell combinations due to data privacy regulations.

²⁷Weight effectiveness is a measure of the deviation from the unweighted distributions. The higher the effectiveness is, the better is the quality of the net sample. Effectiveness values above 70% are considered as very good. The values were calculated as follows: $E = \left(\frac{\sum_{n=1}^N g_n}{N} \right)^2 / \left(\frac{\sum_{n=1}^N g_n^2}{N} \right)$ with E = effectiveness ($0 \leq E \leq 1$), g_n = weighting factor of the n -th case, N = unweighted sample size.

complete interaction of the birth cohorts on your other independent variables or run separate analyses for all cohorts used, you need to use the design weight. If you use only one cohort or more than one and allow for complete interaction of the birth cohorts on all effects, you do not need to use the design weight. As the use of post-stratification weights is highly controversial in the research community, we leave the decision to use a post-stratification weight to the data user. If you want to use the longitudinal weight, you will need to decide between the analysis of a single birth cohort and that of multiple cohorts. If the latter is the case, a multiplication of the combined weight (*dxpsweight*) and the longitudinal weight is needed. To analyze one specific birth cohort, a combination of the *psweight* and the longitudinal weight is sufficient. Tables 4.24 and 4.25 show when to use which weight. Note that, at least to our knowledge, no statistical package handles raking weights properly as here the weighting factors vary within strata. Therefore, we recommend treating the post-stratification weight, the combined weight, and the longitudinal weight as a design weight. Examples of how to use the weights are provided in the Quick Start do-file “Sample Definition & Weighting”.

Table 4.24: When to use which weight

		Want to use post-stratification weight	
		No	Yes
Combine more than one cohort in analysis	No	—	<i>psweight</i>
	Yes	<i>dweight</i>	<i>dxpsweight</i>

Table 4.25: When to use which weight

		Want to do longitudinal analysis	
		No	Yes
Combine more than one cohort in analysis	No	see table 4.24	<i>psweight</i> * <i>lweight</i> (W2)* <i>lweight</i> (W3) ...
	Yes	see table 4.24	<i>dxpsweight</i> * <i>lweight</i> (W2)* <i>lweight</i> (W3) ...

To use a specific weight factor *weightvar* in *Stata*, you use the following code lines:

```
* Put this before relevant command lines
svyset [pweight=weightvar]

* Commands
svy: command ...
```

To use a specific weight factor *weightvar* in *SPSS*, you use the following code lines:

```
* Put this before relevant command lines.
weight by weightvar.

* Commands.
command ... .
```

4.6 Generated data sets

In wave 1, pairfam anchor respondents answered questions on their relationship history since the age of 14 and on their entire fertility history. As of wave 2, an electronic event-history calendar has been used to collect retrospective and/or prospective information on the life domains: partnership, children,

place(s) of residence, and education/employment activity. The information on these aspects of the individual life history as well as information on the parents was compiled into separate, user-friendly data sets to allow convenient analysis of biographical information. As of release 5.0, the data sets *biopart*, *biochild*, *biomob_ehc*, *biomob_ehc_moves*, and *bioact* include respondents of the DemoDiff sample as well.

The episode data set *biopart* provides information on the individual relationship, cohabitation, and marriage history, including both retrospective and prospective information on a monthly basis. In addition, the episode data set *bioact* contains monthly information on the anchor's educational and occupational activities as of the month of the wave 1 interview. The data set *bioact_rtr* covers retrospective information collected in wave 3 about education and work starting at the age of eighteen.

The data set *biochild* is a panel data set which also includes retrospective episode data collected in wave 1. It covers respondents' fertility biography, information on their (biological, step, adoptive, and foster) children, and episodes of cohabitation with children. In contrast, the data set *household* is a pure panel data set and contains information on the dwelling and the household members of the anchor at the time of the interview.

Four panel data sets contain information on the anchor's mobility: *biomob_ehc* (information on residences as collected through the EHC), *biomob_ehc_moves* (residences collected through the EHC, including moves within the same city), *biomob_rtr* (retrospective migration history since the age of 18), and *biomob_rtr_parents* (moving out of the parental house).

In addition, the data set *bioparent* provides retrospective and prospective information on the anchor's biological, adoptive, and stepparents covering waves 1-4.

Please note that the data set *household* does not contain information collected in wave 4 or wave 5. Since no conceptual changes in the generating process of the respective information were implemented in these waves, the do files of the previous waves provide instructions how to enrich that data set with information from wave 4 and 5.

The data set *biopart* is organized in "long" form, i.e., it includes one row for each partner. In comparison, *biochild* and *household* comprise one row for each child or household for each wave. Thus, their format is "long-long". The data sets *bioact* and *bioact_rtr* are also "long-long". However the rows are not based on waves, but on activity episodes. If one activity is performed with one or more interruptions, there is one row for each of these separate episodes. Each of the four *biomob* data sets is organized in "long" format (one row for every episode). The data set *bioparent* is provided in "long-long" format (one row per wave for each parent).

In order to ease the calculation of durations, dates within the generated data sets (with the exception of *household*) are stored in a numerical variable combining both month and year. The value of this variable represents the number of months that have passed since January 1900. We chose this date as a reference point to avoid negative values due to dates previous to January 1960, which is the baseline date in Stata. The following formula is used to calculate this value: $((Year\ of\ respective\ date - 1900) * 12 + Month\ of\ respective\ date - 1)$. To reconvert this information into the original months and years, a new variable has to be generated by subtracting 720 (60*12 months; difference between 1900 and the baseline 1960 in Stata). Subsequently, this variable has to be transformed into format "%td" by using the function "dofm": "gen new variable=dofm(old variable)". Finally, two variables containing the respective year and month can be generated using the functions "year()" and "month()"²⁸.

²⁸Consequently, the complete command for year is: `gen year=year(dofm(old variable-720))` and for month: `gen month=month(dofm(old variable-720))`.

In the data sets *biopart*, *biochild*, and *bioparent*, dates with values “-1 Don’t know”, “-2 No answer”, “-4 Filter error / Incorrect entry”, “-5 Inconsistent value”, “-6 Unreadable answer” are recoded as missings with the value “-7 Incomplete data”. The end of episodes which are ongoing as of the interview date (e.g. the anchor’s current relationship) was assigned the value “-99 Right-censored, ongoing”. In the data set *bioact*, information on censoring is given by a separate variable.

If the respondent provided information on the year and only the information on the month was missing, the value of the month was randomly imputed for all data sets (except for *household*), taking into account potential lower or upper boundaries. Please note that in cases of ambiguous seasonal information on the month and valid answers for the year of a variable indicating a date, random values within the following bounds were imputed for the month:

21 Beginning of the year / winter	→ random value between 1 and 2
24 Spring / Easter	→ random value between 3 and 5
27 Middle of the year / summer	→ random value between 6 and 8
30 Fall	→ random value between 9 and 11
32 End of the year	→ 12

If time-constant variables (such as date of birth) differed between waves due to this procedure of randomly imputing a missing month, the value of the last available wave was retained. The usual missing value definition (see table 4.1) is retained for variables that do not provide episodes or dates. The value “-3 Does not apply” was used for all variables in all data sets.

All generated data sets contain the following basic information on the anchor respondent:

- Person number (ID)
- Date of interview of all (relevant) waves.

Since the variables included in these files have been corrected, the data are more accurate than and differ from the raw data. We therefore strongly recommend using these edited data sets instead of the original information contained in data sets *anchor\$*.

The Stata do-files to produce the respective data sets start from data sets *anchor\$*. They are provided as part of the scientific use file and available on the homepage. For some consistency checks, original non-anonymized data (mainly first names given during the interview) were used. The respective commands were commented out to ensure users are able to run these do-files.

Anchor-partner episode data - *biopart*

The data set *biopart* contains retrospective as well as prospective information on the anchor’s partnership, cohabitation, and marriage episodes since the age of 14. There is one row for each partner (“long” format).²⁹ With each wave, *biopart* is updated, i.e., the information on already existing partners is potentially renewed and new partners are added. Information from each existing wave was used.

For each partner, the beginning of the very first, as well as the end of the last relationship or cohabitation episode were coded as overall beginning and end of this relationship or cohabitation. If the anchor reported more than two episodes with the respective partner, the beginning and end of each of the breaks are stored in the data set. Breaks have been sorted in ascending order according to their beginning date³⁰. In addition, we included the beginning and end dates of each marriage. Since only one marriage episode with each partner was recorded, there were no breaks for marriage episodes.

²⁹This is the major difference compared to *biopart* of release 1.0, which was organized in “wide” format.

³⁰Please note that breaks start at the last month of the preceding relationship or cohabitation episode and end at the first month of the next relationship or cohabitation episode.

Table 4.26 displays the full set of variables which are part of this data set.

In addition to the date of birth (variable *dob*, based on the generated identifier *dob_gen*), the variables *sex* (based on the generated identifier *sex_gen*) and *homosex_p*³¹ are part of this data set. We generated the variables *respwx* (with x=number of wave) as dummy variables to indicate whether the anchor took part in the interview of the respective wave (no attrition).

The variable *index* was generated to indicate the correct ascending order of the reported relationships (including the current relationship) with respect to their first beginning. If the beginning of the relationship was missing (“-7”) for at least one partner, the original position of these episodes as provided by the anchor was retained.

The variables *pnow\$* contain the original serial number for each partner in the respective wave (value “x” of variables “varpx” in the anchor data sets). This facilitates the use of additional information from the anchor data set. In wave 1, partners were numbered consecutively; the current partner was assigned the value “0”. In wave 2, the current partner who was already the current partner in the previous wave received the number “1”, the current new partner received the number “2”, and partners who were partners between two interviews received the numbers “3-5”. As of wave 3, the auxiliary variable *hpnr* as part of the data set *anchor\$* displays the running number of the current partner.

To indicate whether a specific partner was the current partner in wave x, a dummy variable *currwx* (with x=number of wave) was included.³²

For all partners, the variable *sexp* contains the “best” information on the partner’s sex (also see chapter 2.2). For (formerly) current partners, the partner’s date of birth was known and included (variable *dobp_gen*).³³ The identification number *pid* only exists for (formerly) current partners who (potentially) took part in the partner survey of the respective wave.

If a partner had died, the variable *dodp* provides the date of the partner’s death. The value “-66” for variables indicating the end of an episode shows that this episode was terminated by the partner’s death.

Please note that it was not possible to re-identify new partners in wave 2 as previous partners mentioned in wave 1. As of wave 3, the variable *pa30* specifies whether the anchor had already had a relationship with the supposedly new partner. We employed this information to detect identical partners for the data set *biopart*. As of wave 4, the variables *pa31*, *pa32*, *pa33m*, and *pa33y* in the anchor data set indicate whether and, if so, how and when a marriage reported in a previous wave ended meanwhile. This information regarding previous partners was integrated in *biopart* as well.

Table 4.26: List of variables included in data set *biopart*

Variable	Variable label	Values	Value labels
<i>General information</i>			
<i>id</i>	Person number anchor	(see table 2.2)	-
<i>demodiff</i>	DemoDiff sample	0	pairfam
		1	demodiff
<i>intdatwx</i>	Date of interview wave x	-3	Does not apply
		<i>date</i>	-

continued on next page

³¹Generated variable “Homosexual union” based on variables *sex* and *sexp*.

³²Note that in case there is more than one current partner, this variable marks just the current partner for whom additional information was collected during the anchor interview. If there is an additional current partner, the variable is 0 for this additional partner.

³³In case of inconsistencies between different versions of *dobp_gen* in different waves for the same partner, the latest version is employed.

continued from previous page

Variable	Variable label	Values	Value labels
sex	Sex anchor	1 2	Male Female
dob	Date of birth anchor	<i>date</i>	-
homosex_p	Homosexual union	-7 0 1	Incomplete data Heterosexual union Homosexual union
respwx	Respondent in wave x	0 1	No Yes
<i>Ascending order of relationships</i>			
index	Correct order of relationships: Number partner	-3 1 2 ... 13	Does not apply 1st partner 2nd partner ... 13th partner
<i>(Formerly) current partners</i>			
pid	Person number current partner	(see table 2.2)	-
dobp	Date of birth partner	-7 -3 <i>date</i>	Incomplete data Does not apply -
<i>Relationship history</i>			
pnwx	Number partner in wave x	-3 <i>number</i>	Does not apply -
currwx	Current partner in wave x	0 1	No Yes
sexp	Sex partner	1 2	Male Female
dodp	Date of death partner	-7 -3 <i>date</i>	Incomplete data Does not apply -
beg	Beginning relationship	-7 -3 <i>date</i>	Incomplete data Does not apply -
end	End relationship	-99 -66 -7 -3 <i>date</i>	Ongoing Death partner Incomplete data Does not apply -
bkbeg	Beginning break k relationship	-7 -3 <i>date</i>	Incomplete data Does not apply -
bkend	End break k relationship	-7 -3 <i>date</i>	Incomplete data Does not apply -

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
<i>Cohabitation history</i>			
cohbeg	Beginning cohabitation	-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
cohend	End cohabitation	-99	Ongoing
		-66	Death partner
		-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
bkcohbeg	Beginning break k cohabitation	-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
bkcohend	End break k cohabitation	-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
<i>Marriage history</i>			
marbeg	Beginning marriage	-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
marend	End marriage	-99	Ongoing
		-66	Death partner
		-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
marcer	Marriage ceremony	-7	Incomplete data
		-3	Does not apply
		-2	No answer
		-1	Don't know
		1	Only a civil ceremony
		2	A civil and a religious ceremony
		3	Only a religious ceremony
<i>imputations</i>			
imp_unionbeg	Imputed date of beginning union / end of break of union	0	No imputation
		1	Only year information
		2	Only season information
imp_unionend	Imputed date of end union / end of break of union	0	No imputation
		1	Only year information
		2	Only season information
imp_cohabbeg	Imputed date of beginning cohabitation / end of break of cohabitation	0	No imputation
		1	Only year information
		2	Only season information

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
imp_cohabend	Imputed date of end	0	No imputation
	cohabitation / end of break of cohabitation	1	Only year information
		2	Only season information
imp_marbeg	Imputed date of beginning marriage	0	No imputation
		1	Only year information
		2	Only season information
imp_marend	Imputed date of end marriage	0	No imputation
		1	Only year information
		2	Only season information
<i>biopart flag variables</i>			
biopartflag1	Inconsistency biopart: Marriage earlier than beginning of relationship	0	No inconsistency
		1	Inconsistency
biopartflag2	Inconsistency biopart: Overlapping cohabitation episodes with different partners	0	No inconsistency
		1	Inconsistency
biopartflag3	Inconsistency biopart: Beginning current and end previous marriage	0	No inconsistency
		1	Inconsistency
biopartflag4	Inconsistency biopart: Year of birth partner	0	No inconsistency
		1	Younger than 10 years old

Please note that the variables *end*, *cohend*, and *marend* may be -99 even if respondents did not participate in the latest wave and we cannot be sure whether this information is still correct. In case they did participate in the latest interview, the value -7 is assigned if there is some ambiguity with respect to the end of the episode.

Various checks have been conducted to identify *inconsistent episodes and information*:

- Negative durations of relationships, cohabitations, marriages (end before beginning)
- Inconsistencies across breaks in episodes (relationships and cohabitation):
 - subsequent episode (beginning and end) prior to beginning and end of preceding episode
 - subsequent episode (beginning and end) between beginning and end of preceding episode
 - beginning of subsequent episode prior to beginning of preceding episode
 - beginning of subsequent episode prior to end of preceding episode
 - end of subsequent episode prior to beginning of preceding episode
 - end of subsequent episode prior to end of preceding episode
 - identical beginning and end of two episodes
- Beginning of marriage before beginning of relationship (also see *flag5*, table A.21)
- Overlapping episodes of cohabitation with same or different partners (also see *flag6* and *flag7*, table A.21)
- End of previous marriage after beginning of current marriage (also see *flag8*, table A.21)
- Divorce from partner to whom never married (also see *flag10*, table A.21)
- Separation through death/divorce current spouse (also see *flag11*, table A.21)

- Separation before beginning relationship (current partner; also see *flag14*, table A.21)

With respect to information collected in wave 1 (retrospective partnership history and prospective information), we also checked whether the beginning of the relationship/cohabitation/marriage was prior to the first meeting of the anchor and the respective partner. Since the date of the first meeting was not included in *biopart*, please also see *flag20*, *flag21*, and *flag22* as parts of the data set *anchor\$* as of wave 2. They mark inconsistencies regarding the first meeting with the current partner.

Solution to these inconsistencies:

We tried to eliminate inconsistencies as far as possible. In addition to sorting the relationship episodes, beginning with the first provided date (see above), some months were changed slightly if no (“-1/-2”) or no precise (“21-32”) information was available or if a change seemed plausible and necessary. Therefore, new variables are included in the *biopart* data set to mark these imputations in the dates of beginning or end of relationship/cohabitation/marriage episodes. If episodes had exactly identical dates, one of them was dropped. Moreover, episodes which were completely contained in another one with the same partner were also deleted. Any breaks coded with “-1” or “-2” for both the beginning and ending dates were dropped as well.

If it was not possible to eliminate inconsistencies, new variables *biopartflag1*, *biopartflag2*, *biopartflag3*, and *biopartflag4* have been created to mark these inconsistencies.

Partners were dropped (including the respective information on cohabitation episodes and marriages) in the following cases:

- Current partner mistakenly provided as previous partner as well; see interviewer’s note
- Name previous partner = name current partner and beginning of the relationships identical or almost identical (often if end of previous relationship = date of interview or “-1/-2”).

Sources for additional information regarding the respective partners:

As of wave 2, three “types” of partners have to be differentiated:

1. Retrospective partners: Partners whom the anchor respondent mentioned as part of the retrospective partnership history but who were no longer current partners at the time of the interview of wave 1.
2. In-between-waves partners: Partners listed as previous partners in the course of the event-history calendar (EHC), which captures the time between the previous and the current wave, but who were no longer current partners at the time of the interview in wave 2, wave 3, or wave 4.
3. (Formerly) current partners: Partners who were current partners at the time of at least one interview.

For retrospective and in-between-waves partners, the anchor data set additionally contains information on whether the respective partner is the parent of an anchor’s biological child (wave 1: variable *sd16kx*; as of wave 2: variable *ehc12kx*) and (if so) how often the child sees the other parent (as of wave 2: *crn17kx*).

For current partners, the variable *currwx* states the wave in which questions regarding the current partner or regarding the current relationship included in the anchor data set refer to this partner or to the relationship with this partner. The information from the partner survey (PAPI questionnaire) of the respective wave is saved as data set *partner\$* and can be merged using the identification variable *id*. If applicable, the data set *parenting\$* includes information from the partner’s parenting survey and can also be matched using the variable *id*.

The questions in the anchor data set that are part of the module “Meeting the current partner” (variables *pa1_ -pa3*, *sdp1-sdp21*) refer to current partners whose value for *currw\$* is “0” in the previous wave and “1” in the current wave (*currw\$*: “0” at time t-1 and “1” at time t).

The module “Separation from the ex-partner” (variables starting with the acronym “sep” as of wave 2) relates to partners who were current partners in the previous wave, but no longer are current partners in the respective wave and who had not died since the previous wave (*currw\$*: “1” at time t-1 and “0” at time t).

For further information on the computation of the data set, please see the Stata do-files *biopart1-2.do* (data from waves 1 and 2), *biopart3.do* (data from wave 3), *biopart4.do* (data from wave 4) and *biopart5.do* (data from wave 5). The pairfam Technical Paper No. 3 (Schnor and Bastin 2014) offers helpful information on how to combine the two data sets *biopart* and *biochild*.

Anchor-child panel/episode data - biochild

The data set *biochild* contains retrospective and prospective information on the anchor’s children³⁴ and on episodes of living together with the children in the same household. The data set is in long-long format. This means that there is one row per child and for each child there exists one row per wave (that child was mentioned in). Every wave, *biochild* is updated, i.e., the information on existing children is renewed and new children are added. The current data set contains information of wave 1 to wave 5 of pairfam and DemoDiff.

Table 4.27 depicts the variables which are included in this file.

To identify the children, the variable *number* contains the value X on the X-th child corresponding to the variables *varX* in the anchor data sets.³⁵ The variable *surveykid* is a dummy variable that indicates if a child was interviewed as part of the CAPI child survey. Thus, this variable provides information on the success of merging the anchor files with the data sets *child2* to *child5*, respectively.

For all children, the variable *sexk* contains the “best” information on the children’s sex (also see section 4.2).³⁶ The variable *dobk* depicts the “best” information on a child’s date of birth. This variable is provided as number of months passed since January 1900. In case of seasonal or missing monthly information with a non-missing year of birth, the monthly information was randomly imputed. For further details on this principle and the generation of *dobk**, see chapter 4.6 above.

The variable *index* documents the sequence of children within this data set according to the children’s age (referred to the variable *dobk*). The oldest child was assigned code “1 1st child”, the second oldest code “2 2nd child” and so forth. In the case of a missing date of birth of a child, the value “-7 Incomplete data” was assigned, and the order was built corresponding to the remaining information of the other children (only if existent).

The variables *currliv* and *currliv_detail* contain information on the status of cohabitation of the child and the anchor. As of wave 2, details on the cohabitation status, recorded in variable *currliv*, are available. Further details on cohabitation for those children who lived with the “anchor and elsewhere” can be found in variable *currliv_detail*.

³⁴According to the anchor questionnaires, the term “children” refers to biological children, adopted children, children of a partner, or foster children.

³⁵For example, if one child is the third one mentioned, the variable *number* has the value “3”. And if you want to know the status of the child according to the anchor data set in wave 3, you will find the information in the variable *ehc9k3*.

³⁶This “best” information was taken from the generated identifiers *k*sex_gen*.

Similarly to the variables *pnowx* in the data set *biopart*, the variable *pno* carries the original serial number for each partner in the respective wave (value X of variables *varpX* in the anchor data sets).

The variable *parentid* contains the identification number of the child's second biological parent. Note that this variable can differ from the variable *pid*, as the anchor's relationship to the second biological parent can result in separation or a new partnership can emerge.

Retrospective information collected in wave 1:

Note that the variables *livk*beg*, *livk*end*, *b1livk*beg*, *b1livk*end*, and *livk*birth* contain retrospective information and these variable are only filled in wave 1. "Living together" (abbreviation "liv") refers to episodes in which the respective child had lived with the anchor in the same household (see anchor codebook wave 1, questions 49 et seq.).

Table 4.27: List of variables included in data set *biochild*

Variable	Variable label	Values	Value labels
id	Person number anchor	(see table 2.2)	-
demodiff	DemoDiff sample	0 1	pairfam demodiff
intdat	Date of interview (months since january 1900)	<i>date</i>	-
wave	Survey year	1 2 3 4 5	Wave 1: 2008/09 ^a Wave 2: 2009/10 Wave 3: 2010/11 Wave 4: 2011/12 Wave 5: 2012/13
number	Pointer on Xth child (corresponding varkX in anchor data)	1...10	-
cid	Person number CAPI-child	(see table 2.2)	-
surveykid	Indicator : child was asked in CAPI child survey	-7 -3 0 1	Incomplete data Does not apply No Yes
index	Correct order of children (corresponding to date of birth)	-7 -3 1 ... 10	Incomplete data Does not apply 1st child (oldest) ... 10th child
dobk	Date of birth of child (months since january 1900)	-7 -3 <i>date</i>	Incomplete data Does not apply -
sexk	Gender of child	1 2	Male Female

continued on next page

^aFor technical reasons wave 1 of DemoDiff in *biochild* was labeled 2008/09, although data was collected in 2009/10. For detailed information on DemoDiff, see chapter 3.

continued from previous page

Variable	Variable label	Values	Value labels
statusk	Status of child	-7	Incomplete data
		-3	Does not apply
		1	Biological child
		2	Adopted child
		3	Partner's child/stepchild
currliv	Cohabitation with child	4	Foster child
		-7	Incomplete data
		-3	Does not apply
		1	Only with anchor
		2	With anchor and elsewhere
		9	Only alone/flat share
		10	Only with other parent unit
		11	Only with other relative
		12	Children's home
		13	Only elsewhere
currliv_detail	Details on [currliv] category <2. With anchor and elsewhere>	-3	Does not apply
		2	With anchor but also alone/flat share
		3	With anchor but also with other parent unit; mainly with anchor
		4	With anchor but also with other parent unit; namely in equal shares with anchor and with other parent unit
		5	With anchor but also with other parent unit; mainly with other parent unit
		6	With anchor but also with other relative
		7	With anchor but also at children's home
		8	With anchor but also elsewhere
pno	Partnerindex: second biological parent	-7	Incomplete data
		-3	Does not apply
		0 ... 9	-
		97	Another person
pid	Person number partner	(see table 2.2)	-
parentid	Person number second biological parent	see variable pid	-
mid	Person number mother	(see table 2.2)	-
fid	Person number father	(see table 2.2)	-
smid	Person number stepmother	(see table 2.2)	-

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
sfid	Person number stepfather	(see table 2.2)	-
livkbeg	Beginning living together with child	-7	Incomplete data
		-3	Does not apply
livkend	End living together with child	<i>date</i>	-
		-99	Ongoing
		-7	Incomplete data
		-3	Does not apply
b1livkbeg	Beginning break 1 living together with child	<i>date</i>	-
		-7	Incomplete data
		-3	Does not apply
b1livkxend	End break 1 living together with child	<i>date</i>	-
		-7	Incomplete data
		-3	Does not apply
livkbirth	Living together since birth with child	<i>date</i>	-
		-7	Incomplete data
		-3	Does not apply
		0	Not living together since birth
dodk	Date of death child (months since january 1900)	1	Living together since birth
		-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-

Various additional checks have been implemented to identify inconsistent cases.

Checks for episodes of living together with children:

- Date of child's birth after beginning of living together
- Beginning of living together after breaks in living together
- Beginning after end of living together
- Inconsistent order of breaks in living together
- Breaks in living together after end of living together
- End of living together after date of interview
- End of living together after child's death
- Identical beginning and end of living together
- End of living together = date of interview.

Since no such inconsistencies exist, there was no need for any changes.

If the same child was reported twice (duplicated name and date of birth), we dropped the second entry. Moreover, we checked whether any second biological parent was identified as inconsistent and dropped it while generating the data set *biopart* (for more information see above). If this discrepancy occurred, the correct value (in all cases "0 Current partner") was assigned.

Checks for consistency across wave 1 to wave 5:

- Child's date of death (*dodk*):
If death occurred in wave t but the information from wave t+1 does not reflect this fact, then wave t+1 information is preferred to that of wave t. If two different dates of death occur, the newer information of wave t+1 is preferred.

- Child's status:
If there are unrealistic differences between the information of wave 1 to wave 5, the newest information is preferred. Such unrealistic differences were changes in status, for example from or to the status "biological child".

Matching with further information on children:

One general possibility to merge information from the other data sets is to use the anchor's identification variable *id*.

By using the anchor's parents' identification variables (*mid*, *fid*, *smid*, and *sfid*), the data set *parent\$* can be matched, and information reported by the grandparents of the child can be added. Note that this provides only information on the CAPI children. To get information on the parents, their identification variables (*id*, *parentid*, and *pid*) make it possible to match the data sets *anchor\$*, *parent\$*, and *parenting\$*.

Please note: Before using the variable *parentid* for matching purposes, it has to be temporarily renamed to *pid*. Otherwise, the matching process will fail, as the other data sets do not contain this special variable. This variable exists only for a fast differentiation between the general partner identification number (*pid*) and identification number of the second biological parent (*parentid*).

The do-file *biochild1-2.do* documents how the first parts (wave 1 and wave 2) of the data were modified and how the checks to identify inconsistencies in the retrospective parts were generated. The do-files *biochild3.do* to *biochild5.do* show the preparation of the wave 3 to wave 5 data, respectively, as well as the junction of the data from all waves which finally results in the data set *biochild*.

Anchor-activities episode data - *bioact* and *bioact_rtr*

The data set *bioact* contains basic information on all of the anchor's activities in the areas of education and work, starting at the time of the interview of wave 1, i.e. with the release of wave 2 data. It contains information from *pairfam* as well as *DemoDiff*. It covers the period between the survey data of the first and the data of the current wave. The data set is provided in "long-long" format, which means that it consists of one row for each education and/or work episode. There can be more than one activity per respondent and several instances of one activity, depending on the number of activities mentioned. Wave 3 data also contains retrospective information about education and work starting at the age of eighteen. This information can be found in *bioact_rtr*.

The variable *activity* displays the kind of activity the anchor respondent reported. Table 4.28 displays all possible activities covered by the questionnaire.

The anchor data sets beginning with wave 2 contain variables that store information about education and work of the respondent separately for each month. For the activity calendar, the respondent was asked to indicate what he/she had been doing in each month after the preceding interview wave up to the current interview wave. The idea is to acquire a full overview of what has happened concerning education and employment during the period of - on average - the last year.

The *bioact* data set provides the following information about the duration of each activity: The variables *beg* and *end* indicate the first and last month in which an activity was reported by the respondent. The variables *currw1*, *currw2*, *currw3*, *currw4*, and *currw5* mark the current activity at time of the interview of the respective wave. Additionally, *spell* presents the number of separated episodes per activity. The variable *censor* indicates if and in which way these spells were censored, i.e. if the episode began before or lasted longer than the covered period or if information about the month before or after the episode was missing. For censor details see table 4.28.

Furthermore, *bioact* and *bioact_rtr* data include the anchor's date of birth (*dob*) and the interview dates of the current and preceding waves for identification reasons (at present *intdatw1*, *intdatw2*, *intdatw3*, *intdatw4* and *intdatw5*). The dummy variables *respw2*, *respw3*, *respw4* and *respw5* indicate whether the anchor took part in the interview of wave 2 and/or in wave 3 (identification variable for temporary dropouts).

The structures of the two data sets *bioact* and *bioact_rtr* are mostly identical. The retrospective information in wave 3 was not collected with the event history calendar, thus there is ambiguous seasonal information for the beginning and ending of episodes. In such cases random values were imputed. The original ambiguous seasonal information is provided through the variables *flag_a* (beginning) and *flag_b* (ending). When the respondents answered "none of the above" to the question on education or employment, that information can be found in the variables *rtr31i10* (education) and *rtr35i14* (occupation), because no further data on the beginning or ending was collected in these cases.

Please note: Originally, the number of activities of categories 12-16 was known (n=1..5). For simplicity, this was not taken into account for the *bioact* data set. The information on the number of activities is still available in the anchor data set.

Table 4.28 displays the full set of variables which are part of these data sets, in contrast to table 4.29, which shows all relevant variables included in anchor data. The variables can be easily merged with variables from anchor data by using the key variable *id*.

Concerning the variables *ehc19i23* and *ehc19i23m** of the anchor data set, which were used to generate the *bioact* data, additional alterations had to be carried out. If a respondent had gaps in his/her calendar and clicked the "finish"-button, then a pop-up appeared and he/she was reminded that the calendar should be filled out completely. Additionally, an extra line/activity appeared in the calendar labeled "don't know/cannot remember". Information entered in this line/activity was stored in variables *ehc19i23* and *ehc19i23m** in anchor data. These variables were constructed to act as "gap fillers" for months when respondents could not recall what they had done.

In the progress of data editing *ehc19i23* and *ehc19i23m** were recoded to "1" for all cases with gaps in the activity calendar, if no activity applied in a given month. In contrast, the variables were recoded "0" if at least one activity was mentioned in a month. Thus, in the final data it works as a "gap filler" as originally intended. However, we recommend considering and checking *ehc19i23* and *ehc19i23m** when using monthly activity information contained in anchor data other than the *bioact* data.

In wave 2/3 of DemoDiff for all variables for which the wave is given in the name "3" is used instead of "2," so for all DemoDiff cases the variables *intdatw2* and *currw2* are coded "-3 does not apply." The do-files *bioact2.do*, *bioact3.do* and *bioact4-5.do* document in detail how the data set *bioact* was computed.

Table 4.28: List of variables included in data sets *bioact* and *bioact_rtr*

Variable	Variable label	Values	Value labels
<i>General information</i>			
<i>id</i>	Person number anchor	(see table 2.2)	-

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
demodiff	DemoDiff sample	0 1	pairfam demodiff
intdatw1	Date of interview wave 1	<i>date</i>	-
intdatw2	Date of interview wave 2	<i>date</i>	-
intdatw3	Date of interview wave 3	<i>date</i>	-
intdatw4	Date of interview wave 4	<i>date</i>	-
intdatw5	Date of interview wave 5	<i>date</i>	-
dob	Date of birth	<i>date</i>	-
<i>Activity history</i> activity	Type of activity	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18	Evening school, working on a school leaving certificate for adults Vocational training (apprenticeship, business school etc.) Vocational retraining / further edu- cation University of cooperative education (also "Berufsakademie") University of applied sciences, col- lege, university General secondary school (first edu- cation) Pre-vocational training Technical/professional school Other education Full-time employment Self-employment Part-time employment (also multi- ple part-time jobs) Internships, traineeship, including unpaid work Marginal part-time employment, mini-job, "Ein-Euro-Job" Occasional or irregular employment Other type of employment, namely Maternity or paternity leave or other leave of absence for childcare Military service, alternative civilian service, voluntary social service year

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
		19	Unemployed, seeking employment
		20	Housewife / Househusband
		21	Early retirement, retirement, occupational disability
		22	Other type of non-employment
		23	Don't know, can't remember
rtr31i10 (bioact_rtr only)	Education: None of the above	0 1	Not mentioned Mentioned
rtr35i14 (bioact_rtr only)	Occupation: None of the above	0 1	Not mentioned Mentioned
flag_a (bioact_rtr only)	Ambiguous information on beginning	-3 21 24 27 30 32	Does not apply Original information on the month Original information on the month Original information on the month Original information on the month Original information on the month
flag_b (bioact_rtr only)	Ambiguous information on ending	-3 21 24 27 30 32	Does not apply Original information on the month Original information on the month Original information on the month Original information on the month Original information on the month
spell	Counter - Number of spells per activity	<i>number</i>	-
beg	Beginning of activity in month	<i>date</i>	-
end	End of activity in month	<i>date</i>	-
censor	Indicator for censored spells	-3 0 1 2 3 4 5	No censoring (only for ehc19i23*) Uncensored Left-censored Right-censored, end of episode is missing Right-censored, ongoing episode Combination of 1 & 2 Combination of 1 & 3

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
currw1	Current activity in wave 1	0 1	No Yes
currw2	Current activity in wave 2	0 1	No Yes
currw3	Current activity in wave 3	0 1	No Yes
currw4	Current activity in wave 4	0 1	No Yes
currw5	Current activity in wave 5	0 1	No Yes
respw2	Respondent in wave 2	0 1	No Yes
respw3	Respondent in wave 3	0 1	No Yes
respw4	Respondent in wave 4	0 1	No Yes
respw5	Respondent in wave 5	0 1	No Yes

Table 4.29: List of available variables covering education and work in anchor data set

Variable	Description	Connection
sd32i*	Attained educational certificate/voc. qualification last year	Between waves
rtr31i*	Education experience after 18th birthday until first interview date	Before first interview
sd33	Attending same school as in previous wave	Currently
sd25	Type of school currently attending	Currently
sd26	Grade in school	Currently
rtr35i*	Work experience after 18th birthday until first interview date	Before first interview
job19	Same occupation as in previous wave	Currently
job20o	Current occupation	Currently
job21	Same occupational status as in previous wave	Currently
job2	Current occupational status	Currently
job3-job14	Additional information on current job(s)	Currently
inc2 / inc21	Net income (earnings) last month	Currently

Anchor-household panel data - household

The data set *household* contains information on the anchor's main residence, on the household members, and on household income. The data set uses information from the anchor data sets throughout the waves and combines them into a panel data set where one row is reserved per household per wave. Since we used only information on the main residence so far, there is only one household per wave, and thus there is one row in the data for each wave the anchor participated. The processing of the data is documented in the Stata do-files *household2.do* and *household3.do*.

Table 4.30 shows the variables contained in the data set. For each household member mentioned, we generated variables indicating the relationship of the anchor to the household member, the sex of the household member, and their date of birth (year and month). In contrast to wave 2 we generated variables for the different relationships. In wave 3 we distinguish between characteristics of partner (*p**), children (*c**), parents (*pa**), and other household members (*other**). Variables *p*_rel* contain

the information of cohabitation with partners. We assigned value 1 to variable *p1_rel* if the anchor reported that he/she lives still (or again) with the wave 1 or wave 2 partner. We assigned value 1 to variable *p2_rel* if the anchor cohabited with a new partner.

The anchor's children were assigned value 10 (biological children) or 11 (step/adopted/foster children) on *c*_rel* if the anchor reported to live with them (*ehc9k*=1,...,4 & ehc10k*h1=1*).

Next we assigned the values for parents (*pa*_rel=2,...,9*) and other household members (*other*_rel=12,...,21*) from the household grid variables. In wave 3, this information is stored in variables *ehc21p**, *ehc22p**, and *ehc25p*h1*.

For all persons for which we had the information (partners, children, parents), we used the generated variables on the birth month and year (**doby_gen*, **dobm_gen*, **sex_gen*) to fill in the respective variables. If this information was not available for some household members, we used the information gathered in the household grid instead.³⁷

Furthermore, from the *anchor3* data set, we matched the person number of the current partner to variables *p*_id* and the person number of biological, adopted, or step parents (*mid*, *fid*, *smid*, *sfid*) to the variables *pa*_id*. Since wave 3 we also matched the person number of the Anchor's child(ren) (*cid*) to variables *c*_id*. Note that this id is only available for so-called "CAPI children". Additionally, we generated variables *c*_point*, which include information about the position of the children in the event history calendar. Moreover, we assigned valid values on sex, date of birth, and person number forwards and backwards for household member present in both waves.

Table 4.30: List of variables included in data set *household*

Variable	Variable label	Values	Value labels
id	Person number anchor	(see table 2.2)	-
wave	Survey year	1 2 3	Wave 1: 2008/09 Wave 2: 2009/10 Wave 3: 2010/11
pid	Person number partner	(see table 2.2)	-
mid	Person number mother	(see table 2.2)	-
fid	Person number father	(see table 2.2)	-
smid	Person number stepmother	(see table 2.2)	-
sfid	Person number stepfather	(see table 2.2)	-
cid	Person number capic child	(see table 2.2)	-
intm	Month of interview	(see section 4.4)	-

continued on next page

³⁷Information on sex was collected only in wave 1, information on dates of birth only in wave 2. For mothers and fathers, however, we could simply infer sex from the relationship.

continued from previous page

Variable	Variable label	Values	Value labels
inty	Year of interview	(see section 4.4)	-
sex_gen	Generated sex anchor	1 2	Male Female
doby_gen	Generated year of birth anchor	<i>see dob*_gen</i>	-
dobm_gen	Generated month of birth anchor	<i>see dob*_gen</i>	-
hhsizemrd	Number of hh members main residence	<i>see hhsizemrd</i>	-
pmrd	Partner lives in household main residence	-7 0 1	Incomplete data Partner does not live at main residence Partner lives at main residence
mmrd	Mother lives in household main residence	-7 0 1	Incomplete data Mother does not live at main residence Mother lives at main residence
fmrdr	Father lives in household main residence	-7 0 1	Incomplete data Father does not live at main residence Father lives at main residence
childmrd	Number of children in household main residence	-7 0...10	Incomplete data -
othmrd	Number of other hh members main residence	-7 0...11	Incomplete data -
dwtype	Type of household (own, parental, shared, dorm., other)	-7 1 2 3 4 5	Incomplete data Own household Parental household (father/mother/step/foster parents) Shared dwelling with roommates/housemates Dormitory, student dormitory, boarding school, or similar Other type of household

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
owner	Home-ownership	-7	Incomplete data
		-3	Does not apply
		1	sublease
		2	rented apartment/house
		3	own exclusive property
		4	property of parent(s)
		5	property of another person
		6	property of partner
		7	joint property with partner
nights	Nights spent at residence	8	Other
		-7	Incomplete data
		-3	Does not apply
rent	Monthly rent for dwelling (euros)	1...6	-
		-7	Incomplete data
		-3	Does not apply
expown	Monthly expenditures for self-owned apartment / house (euros)	-7	Incomplete data
		-3	Does not apply
dwsiz	Size of dwelling (square meters)	-7	Incomplete data
		-3	Does not apply
rooms	Number of rooms of dwelling	-7	Incomplete data
		-3	Does not apply
res2nd	Respondent has 2nd residence	-7	Incomplete data
		0	No 2nd residence
		1	2nd residence
hhincnet	Household net income (open and estimated information combined)	<i>see hhincgcee</i>	-
hhcomp	Household Composition (partner, kid(s), parent(s), others)	-7	Incomplete data
		1	<i>Partner-Kids-parent-other</i>
		2	w - w - w - w
		3	w - w - w -w/o
		4	w - w - w/o -w
		5	w - w - w/o - w/o
		6	w - w/o - w - w
		7	w - w/o - w - w/o
		8	w - w/o - w/o - w
		9	w - w/o - w/o - w/o
		10	w/o - w - w - w
		w/o - w - w - w/o	

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
		11	w/o - w - w/o - w
		12	w/o - w - w/o - w/o
		13	w/o - w/o - w - w
		14	w/o - w/o - w - w/o
		15	w/o - w/o - w/o - w
		16	w/o - w/o - w/o - w/o
hhincgcee	Net equivalence income (GCEE)	see <i>hhincgcee</i>	-
pX_rel	Anchor's relation to partner X	-3 1	Does not apply partner
pX_sex	Sex of partner X	-7 -3 1 2	Incomplete data Does not apply Male Female
pX_byear	Year of birth of partner X	-7 -3	Incomplete data Does not apply
pX_bmonth	Month of birth of partner X	-7 -3	Incomplete data Does not apply
pX_id	Person number of partner X	-7 -3	Incomplete data Does not apply
cX_rel	Anchor's relation to child X	-3 10 11	Does not apply biological child step/adopt./foster child
cX_sex	Sex of child X	-7 -3 1 2	Incomplete data Does not apply Male Female
cX_byear	Year of birth of child X	-7 -3	Incomplete data Does not apply
cX_bmonth	Month of birth of child X	-7 -3	Incomplete data Does not apply
cX_id	Person number of child X	-7 -3	Incomplete data Does not apply
paX_rel	Anchor's relation to parent X	-3 2 3 4 5 6 7	Does not apply biological mother biological father adoptive mother adoptive father stepmother / father's partner stepfather / mother's partner

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
		8	foster mother
		9	foster father
paX_sex	Sex of parent X	-7	Incomplete data
		-3	Does not apply
		1	Male
		2	Female
paX_byear	Year of birth of parent X	-7	Incomplete data
		-3	Does not apply
paX_bmonth	Month of birth of parent X	-7	Incomplete data
		-3	Does not apply
paX_id	Person number of parent X	-7	Incomplete data
		-3	Does not apply
otherX_rel	Anchor's relation to partner X	-7	Incomplete data
		-3	Does not apply
		12	sister or brother
		13	half-sister or half-brother
		14	stepsister or stepbrother
		15	adopt./foster sister or adopt./foster brother
		16	grandmother or grandfather
		17	Another relative or someone similar
		18	mother- or father-in-law (partner's parents)
		19	A friend
		20	Room- or housemate
		21	Other
otherX_sex	Sex of partner X	-7	Incomplete data
		-3	Does not apply
		1	Male
		2	Female
otherX_byear	Year of birth of partner X	-7	Incomplete data
		-3	Does not apply
otherX_bmonth	Month of birth of partner X	-7	Incomplete data
		-3	Does not apply
otherX_id	Person number of partner X	-7	Incomplete data
		-3	Does not apply

Regarding the identification of same household members throughout the waves we used the person number for identification of same partners and parents and the position in the event-history calendar for identification of same children. These positions remain constant in each wave (unlike the position in the household grid). To identify the other persons in the household we generated the variables *other*_point*, which include information about the position in the household grid in the first two waves. Using the information of the household members' first name we detected same household members throughout the waves and assigned the respective value to the variables *other*_point*. If the first name was not available we used their day of birth to identify same household members. If both variables were not available we decided individually. Please notice that in wave 3 household members' sex is unknown. So in some cases only the relationship to the anchor is known. We accepted the possibility of potential mistakes (wrong sex) because we wanted to avoid classifying them as "new" household members. Later, we used the variables *other*_point* to match the wave 3 information to the former waves. Consistency checks showed that matching did not work if names were modified.

Again, we decided individually how to classify these cases.

After completing the editing of the information on household members, we generated aggregate household characteristics. The data can be merged with information from other data sets by using the person number of the anchor, or the person number of household members (partner, child, parents) who are potentially respondents of the respective multi-actor surveys. Please note that the data set *household* does not encompass information from wave 4.

Anchor-mobility panel data - *biomob_ehc*, *biomob_ehc_moves*, *biomob_rtr* and *biomob_rtr_parents*

Besides information on educational and occupational activities, partners, and the household grid, information on the current main, second and further residences, as well as on former residences is gathered through the Event History Calendar (EHC). The respondents are asked to provide information on all residences in which they have lived or are still living since the last interview. During the interviews of the second and third wave the respondents were asked to give additional information on the second and all further residences. In the fourth wave they were only asked for information on the main residences. Thus, information on the second residence can only be found in the data sets *anchor2* and *anchor3*. From this data gathered through the EHC the data sets *biomob_ehc* and *biomob_ehc_moves* are generated.

In wave 3 the respondents answered questions on their migration history since the age of 18. Besides the retrospective migration history anchor persons were asked if and when they left their parental home for the first time and if and when they moved back if additional moves out of the parental house occurred. This data is compiled in the data sets *biomob_rtr* (retrospective migration history) and *biomob_rtr_parents* (moves out of the parental house).

Although all these data sets contain information on migration and residential changes, the data has to be partitioned into more than one data set since the retrospective information does not include second residences and a residence at the parental house does not mean automatically that no spatial mobility happened.

According to the generated data sets *biopart*, *bioact*, and *biochild* the time variables in the data sets *biomob_ehc*, *biomob_rtr*, and *biomob_rtr_parents* are calculated according to the following formula: $((Year\ of\ respective\ date - 1900) * 12 + Month\ of\ respective\ date - 1)$. Ambiguous seasonal or missing information on the month is imputed in the same bounds as in *biopart* and *biochild*. The end of ongoing episodes is coded to 97.

In the two data sets *biomob_ehc* and *biomob_rtr* new data is added: the BIK-Classification *bik* and the migration distance *Mig_Dis*. In order to calculate the migration distance between two consecutive places of residence and the BIK-Classification, information on the places of residence is synchronized with the "Gemeindeverzeichnis" published by the Federal Statistical Office of Germany. By doing so, the BIK-classification and geo-coordinates for the calculation can be attached. The variable *bik* equates to the variable *bik* described in chapter 3.3. For places outside the borders of the Federal Republic of Germany the geo-coordinates have been investigated individually.

Migration distance *Mig_Dis* is calculated according to the following formula for orthodromes:

$$acos(\sin(\Phi_A) \sin(\Phi_B) + \cos(\Phi_A) \cos(\Phi_B) \cos(\lambda_A - \lambda_B)) * 6370.$$

Φ_A and Φ_B represent the latitudes of the two consecutive places of residence; λ_A and λ_B are representing the corresponding longitudes of the two places. The number 6370 equates to the radius of earth in kilometers. The calculated migration distances are rounded accurately to kilometers. If a place or residence is not detectable *Mig_Dis* is coded to -99. If the calculation of distance is not possible the code -98 is assigned. If there is a break between two episodes at a second residence

Mig_Dis is coded to -97. If there is no migration *Mig_Dis* is coded to “-3 does not apply”. If the respondent has not answered the questions or didn’t know the place of residence *Mig_Dis* is coded to “-2 No answer” or “-1 Don’t know”.

The person number of the anchor, dates of interview, date of birth, and sex are included in all data sets. Further information from anchor data sets can easily be attached by using the person number.

Biomob_ehc

The data set *biomob_ehc* contains information on residences as collected through the EHC in pairfam and DemoDiff. The EHC was first used in the second wave of pairfam. The data covers the time between the first and the last interview. The data set is provided in the long format, so there is one row for every episode at a residence. For waves 2 and 3 information on the second residence is included. In wave 4, no data on second and further residences has been collected. For each episode the beginning and the end (variables *begin* and *end*) are stored in the data set. The variable *end* is coded to “97 ongoing” for the respondents’ last episode.

Table 4.31 displays the full set of variables which are part of this data set.

The variables *index_lm* and *index_zw* are generated to indicate the ascending order of episodes at the main and the second residences separately. In some cases it is not possible to determine the main residence based on the information delivered by the anchor. If this is the case the variable *biomobflag1* is coded to 1. The same problem appears with some second residences so the variable *biomobflag2* is coded to 1. If it was possible to indicate a main or second residence for a certain period of time, both variables have been coded to 0. In few cases respondents had only incorrect entries. In this case *biomobflag3* is coded to 1.

As described in chapter 3.1 some problems occurred during wave 2. To improve the quality of the data, information gathered through the EHC in wave 2 was combined with retrospective data on the residence of *anchor3*. Retrospective questioning regarding residences was conducted until the respondent mentioned a date after the date of the first interview. Because there has to be an overlap between the first episode mentioned in the EHC of *anchor2* and the episode last mentioned in the retrospective questionnaire in wave 3, data from the two sources was compared. Whenever there was a difference between information on the places of residences the respective cases were checked separately. Five possible types of differing information occurred:

- one of the two places of residence is missing (*biomobflag4* is coded “1 no answer”)
- neither the information on the place of residence of *anchor2* nor the retrospective information of *anchor3* can be allocated to an existing place (*biomobflag4* is coded “2 no allocation possible”).
- the place mentioned in the EHC of *anchor2* is a district of the municipality mentioned in the third wave and vice versa (*anchor3 biomobflag4* is coded “3 district”).
- the mismatch results from the different spelling of names (*biomobflag4* coded “4 spelling”).
- information of *anchor2* and *anchor3* indicates two different places (*biomobflag4* coded “5 difference”)

After this, the beginning of the first episode given in the EHC and the beginning of the last episode of the retrospective module of *anchor3* was compared. The difference is stored in *biomobflag5*. If one of the two answers was “97 ongoing” *biomobflag5* is also coded to 97. It is assumed that the retrospective data is not as error-prone as the EHC data on the residences of *anchor2*. So *biomob_ehc* is modified as follows: If the difference between the ends of the two episodes is smaller than 4 or greater than -4 information on the place of residence is changed according to the retrospective data from *anchor3*. If there is a difference between the places of residence and the difference between the ends of the

two episodes is too large, no geo-coordinates are attached and therefore no migration distances are calculated. In this case the migration distance is coded “-99 indeterminate place”. The variables *biomobflag4* and *biomobflag5* are kept in the data-set to make the conducted changes transparent since the syntax in the do-files contains the complete information on the places of residence. Thus the syntax is not contained in the public do-file *biomob_ehc*. In wave 2/3 of DemoDiff for all variables for which the wave is given in the name “3” is used instead of “2,” so for all DemoDiff cases the variable *inttime_w2* is coded “-3 does not apply.” For further information on the computation of the data set, please see the Stata do-file *biomob_ehc.do*.

Table 4.31: List of variables included in data set *biomob_ehc*

Variable	Variable label	Values	Value labels
id	Person number anchor	(see table 2.2)	-
demodiff	DemoDiff sample	0 1	pairfam demodiff
inttime_w1	Date of interview wave 1	<i>date</i>	
inttime_w2	Date of interview wave 2	-3 <i>date</i>	Does not apply
inttime_w3	Date of interview wave 3	-3 <i>date</i>	Does not apply
inttime_w4	Date of interview wave 4	-3 <i>date</i>	Does not apply
inttime_w5	Date of interview wave 5	-3 <i>date</i>	Does not apply
sex	Sex	-4 1 2	Filter error / Incorrect entry Male Female
dob	Date of birth	<i>date</i>	-
BL	Federal State (Bundesland)	-99 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17	Indeterminate place Inconsistent value Filter error / Incorrect entry Does not apply No answer Don't know Berlin (West) Schleswig-Holstein Hamburg Niedersachsen (Lower Saxony) Bremen Nordrhein-Westfalen (North Rhine-Westfalia) Hessen (Hesse) Rheinland-Pfalz (Rhineland-Palatinate) Baden-Württemberg Bayern (Bavaria) Berlin (East) Brandenburg Mecklenburg-Vorpommern (Mecklenburg-Western Pomerania) Sachsen (Saxony) Sachsen-Anhalt (Saxony-Anhalt) Thüringen (Thuringia) Saarland Country, please specify
begin	Beginning	<i>date</i>	-

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
end	End	97	ongoing
		<i>date</i>	
Wohnsitz	Main or second residence	1	Main residence
		2	Second residence
Index_lm	Number of main residence	-3	Does not apply
		1	1st main residence
		2	2nd main residence
	
		26	26th main residence
Index_zw	Number of second residence	-3	Does not apply
		1	1st second residence
		2	2nd second residence
	
		16	16th second residence
Mig_Dis	Migration: Distance	-99	Indeterminate place
		-98	Calculation not possible
		-97	Second residence with break
		-3	Does not apply
		-2	No answer
		-1	Don't know
		<i>number</i>	
bik	BIK size of region (10)	0	City Center - population 500,000+
		1	Periphery - population 500,000+
		2	City Center - population 100,000-500,000
		3	Periphery - population 100,000-500,000
		4	City Center - population 50,000-100,000
		5	Periphery - population 50,000-100,000
		6	Region - population 20,000-50,000
		7	Region - population 5,000-20,000
		8	Region - population 2,000-5,000
		9	Region - population < 2,000
biomobflag1	No indicated main residence	0	allocated
		1	not allocated
biomobflag2	No indicated second residence	0	allocated
		1	not allocated
biomobflag3	Only Filter error/Incorrect entry	0	allocated
		1	not allocated
biomobflag4	Inconsistency	0	No inconsistency
	Biomob: Difference	1	no answer
	ehc rtr; names	2	no allocation possible
		3	district
		4	spelling
		5	difference

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
biomobflag5	Inconsistency Biomob: Difference ehc rtr; duration	0 <i>number</i>	No inconsistency

Biomob_ehc_moves

The ETC was intended to record migration only between separate cities. The interviewers were advised not to enter a new place of residence into the EHC if the respondents changed their place of residence within the same city. Questions on such moves were posed only when a person indicated that he or she had been living continuously in the same place since the last interview. Only in this case were respondents asked whether they had moved within the city. But in some cases the respondents mentioned a district of a town already mentioned, causing the interviewers to register this district of the same town as a new place of residence in the EHC. These cases occurred after the synchronization with the "Gemeindeverzeichnis," therefore all the names of districts were allocated to towns. These changes were indicated by a migration distance of zero kilometers, and because the EHC was intended to only cover "real" migration these changes of places were removed from the *biomob_ehc*. In order to make use of all collected data through DemoDiff and pairfam, the data set *biomob_ehc_moves* was generated to include all these moves.

Table 4.32 displays the full set of variables which are part of this data set.

The variable *mig0_w1_w2_lm* contains information about how many of the moves described above occurred for the main residence between the interview date of wave 1 and the interview date of wave 2. If the respondent did not participate in wave 2 but did participate in wave 3, the variable *mig0_w1_w3_lm* indicates how many moves occurred between wave 1 and wave 3. The remaining variables - including the variables for the second residences - are computed according to the same logic. If the respondents were asked if they moved their main or second residence within the same town (ehc18 in anchor2 and ehc29 since anchor3) the variables *within_w1_w2_lm*, *within_w1_w3_lm*, *within_w2_w3_lm* etc. are generated in a similar way. In these cases the variables do not indicate the number of moves, but rather if at least one move occurred. The respondents were asked only if they moved within a city, not how often. In wave 2/3 of DemoDiff for all variables for which the wave is given in the name "3" is used instead of "2," so for all DemoDiff cases the variable *inttime_w2* is coded "-3 does not apply."

For further information on the computation please see Stata do-file *biomob_ehc_moves*.

Table 4.32: List of variables included in data set *biomob_ehc_moves*

Variable	Variable label	Values	Value labels
id	Person number anchor	(see table 2.2)	-
demodiff	DemoDiff sample	0 1	pairfam demodiff

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
inttime_w1	Date of interview wave 1	<i>date</i>	
inttime_w2	Date of interview wave 2	-3 <i>date</i>	Does not apply
inttime_w3	Date of interview wave 3	-3 <i>date</i>	Does not apply
inttime_w4	Date of interview wave 4	-3 <i>date</i>	Does not apply
inttime_w5	Date of interview wave 5	-3 <i>date</i>	Does not apply
sex	Sex	-4 1 2	Filter error/ Incorrect entry Male Female
dob	Date of birth	<i>date</i>	-
mig0_w1_w2_lm	No. of moves (Mig_Dis=0) w1 to w2: Main residence	-3 <i>number</i>	Does not apply
mig0_w1_w3_lm	No. of moves (Mig_Dis=0) w1 to w3: Main residence	-3 <i>number</i>	Does not apply
mig0_w2_w3_lm	No. of moves (Mig_Dis=0) w2 to w3: Main residence	-3 <i>number</i>	Does not apply
mig0_w2_w4_lm	No. of moves (Mig_Dis=0) w2 to w4: Main residence	-3 <i>number</i>	Does not apply
mig0_w3_w4_lm	No. of moves (Mig_Dis=0) w3 to w4: Main residence	-3 <i>number</i>	Does not apply
mig0_w3_w5_lm	No. of moves (Mig_Dis=0) w3 to w5: Main residence	-3 <i>number</i>	Does not apply
mig0_w4_w5_lm	No. of moves (Mig_Dis=0) w4 to w5: Main residence	-3 <i>number</i>	Does not apply
mig0_w1_w2_zw	No. of moves (Mig_Dis=0) w1 to w2: Second residence	-3 <i>number</i>	Does not apply
mig0_w1_w3_zw	No. of moves (Mig_Dis=0) w1 to w3: Second residence	-3 <i>number</i>	Does not apply
mig0_w2_w3_zw	No. of moves (Mig_Dis=0) w2 to w3: Second residence	-3 <i>number</i>	Does not apply
within_w1_w2_lm	Within city (ehc18) w1 to w2: Main residence	-3 0 1	Does not apply No Yes
within_w1_w3_lm	Within city (ehc29) w1 to w3: Main residence	-3 0 1	Does not apply No Yes
within_w2_w3_lm	Within city (ehc29) w2 to w3: Main residence	-3 0 1	Does not apply No Yes

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
within_w2_w4_lm	Within city (ehc29) w2 to w4:	-3	Does not apply
	Main residence	0	No
		1	Yes
within_w3_w4_lm	Within city (ehc29) w3 to w4:	-3	Does not apply
	Main residence	0	No
		1	Yes
within_w3_w5_lm	Within city (ehc29) w3 to w5:	-3	Does not apply
	Main residence	0	No
		1	Yes
within_w4_w5_lm	Within city (ehc29) w4 to w5:	-3	Does not apply
	Main residence	0	No
		1	Yes
within_w1_w2_zw	Within city (ehc18) w1 to w2:	-3	Does not apply
	Second residence	0	No
		1	Yes
within_w1_w3_zw	Within city (ehc29) w1 to w3:	-3	Does not apply
	Second residence	0	No
		1	Yes
within_w2_w3_zw	Within city (ehc29) w2 to w3:	-3	Does not apply
	Second residence	0	No
		1	Yes

Biomob_rtr

The data set *biomob_rtr* contains information on all main residences after the 18th birthday. All main residences inhabited by the respondents up to the first date of interview are covered. The respondents were asked about former residences (before the date of the first interview) in chronological order, about the date on which they started to live there, and about the date they left. This data was gathered during the interviews of wave 3. The *biomob_rtr* data set is provided in the long format; every episode of a person is represented by a row.

Table 4.33 displays all variables which are part of this data set.

BLand contains information on the federal state in which the respondent lived during the respective time. In the original data set, the values for Rhineland-Palatinate and Saarland are merged. Whenever the original place of residence could be allocated to Saarland, the value was changed to Saarland. Now it is possible to differentiate between Saarland and Rhineland-Palatinate.

The variables *begin* and *end* represent the beginning and end of an episode at a certain place of residence. By definition, the month of the 18th birthday is the beginning of the first episode (*begin* = *dob* + 216). If only ambiguous seasonal information exists, the beginning was coded to a random value within the bounds mentioned above. If there was a complete missing (“-1 don’t know” or “-2 no answer”) for the month but the year of the beginning or end of an episode was reported, the values were changed to a random value between 1 to 12. If such random information for the month was generated the original ambiguous seasonal information is saved in the variable *biomobflag_rtr_flag1*. If the information on the month was coded to “-1 don’t know” or “-2 no answer” the original information was stored in *biomob_rtr_flag2*.

Geo-coordinates were collected for places of residence outside the borders of Germany. If there was information on the country but not on the town in the respective country, or the city or town could not be located, the geographical center of the country was used to calculate the migration distances.

In this case the variable *biomob_rtr_flag3* is coded to “1 Geographical center point”.

The do-file *biomb_rtr.do* documents in detail the computation of *biomob_rtr*.

Table 4.33: List of variables included in data set *biomb_rtr*

Variable	Variable label	Values	Value labels
id	Person number anchor	(see table 2.2)	-
inttime_w1	Date of interview wave 1	<i>date</i>	
inttime_w2	Date of interview wave 2	-3 <i>date</i>	does not apply
inttime_w3	Date of interview wave 3	-3 <i>date</i>	does not apply
sex	Sex	-5 -4 -3 -1 1 2	Inconsistent value Filter error / Incorrect entry Does not apply Don't know/No answer Male Female
dob	Date of birth	<i>date</i>	-
BLand	Federal State (Bundesland)	-5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17	Inconsistent value Filter error / Incorrect entry Does not apply No answer Don't know Berlin (West) Schleswig-Holstein Hamburg Niedersachsen (Lower Saxony) Bremen Nordrhein-Westfalen (North Rhine-Westfalia) Hessen (Hesse) Rheinland-Pfalz (Rhineland-Palatinate) Baden-Württemberg Bayern (Bavaria) Berlin (East) Brandenburg Mecklenburg-Vorpommern (Mecklenburg-Western Pomerania) Sachsen (Saxony) Sachsen-Anhalt (Saxony-Anhalt) Thüringen (Thuringia) Saarland Country, please specify
begin	Beginning	-2 -1 <i>date</i>	No answer Don't know -

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
end	End	-2	No answer
		-1	Don't know
		97	Ongoing
Index	Number of main residence	<i>date</i>	-
		-3	Does not apply
		1	1st main residence
		2	2nd main residence
	
		26	26th main residence
Mig_Dis	Migration: Distance	-99	Indeterminate place
		-98	Calculation not possible
		-3	Does not apply
		-2	No answer
		-1	Don't know
		<i>number</i>	
BIK_10	BIK size of region (10)	0	City Center - population 500,000+
		1	Periphery - population 500,000+
		2	City Center - population 100,000-500,000
		3	Periphery - population 100,000-500,000
		4	City Center - population 50,000-100,000
		5	Periphery - population 50,000-100,000
		6	Region - population 20,000-50,000
		7	Region - population 5,000-20,000
		8	Region - population 2,000-5,000
		9	Region - population < 2,000
biomob_rtr_flag1	Unknown information on month	-3	Does not apply
		21	Original information on the month
		24	Original information on the month
		27	Original information on the month
		30	Original information on the month
biomob_rtr_flag2	Month: No answer/Don't know	32	Original information on the month
		-3	No Change
		-2	Original information on the month
		-1	Original information on the month
biomob_rtr_flag3	geographical centre point of the country entry	-3	Does not apply
		1	Geographical centre point

Biomob_rtr_parents

The data set *biomob_rtr_parents* contains all information on the episodes inside the parental house. The respondents were asked to provide the date of their first move out of the parental house and all following moves out of the parental house and as well as all following moves back into the parental house. This data set also is in the long format. Every episode is represented by a row.

All variables contained in the data set are shown in Table 4.34.

The beginning of the first episode has been defined as the respondent's date of birth. It has been assumed that the respondents had lived in their parents' house since birth. If there was a change because of ambiguous seasonal information, the original information is stored in *flag_b_begin*. In the case of missing information on the month, the original information is stored in the *flag_b_begin*. Analogous to this procedure, the original ambiguous seasonal information for the end of an episode is stored in the variable *flag_a_end* and the original missing variable is represented in *flag_b_end*.

The do-file *biomb_rtr_parents.do* documents the computation of this data set.

Table 4.34: List of variables included in data set *biomob_rtr_parents*

Variable	Variable label	Values	Value labels
id	Person number anchor	(see table 2.2)	-
inttime_w1	Time of interview: wave 1	<i>date</i>	
inttime_w2	Time of interview: wave 2	-3 <i>date</i>	does not apply
inttime_w3	Time of interview: wave 3	-3 <i>date</i>	does not apply
sex	Sex	-5 -4 -3 -1 1 2	Inconsistent value Filter error / Incorrect entry Does not apply Don't know/No answer Male Female
dob	Date of birth	<i>date</i>	-
begin	Beginning of episode	-4 -2 -1 <i>date</i>	Filter error / Incorrect entry No answer Don't know
end	End of episode	-4 -2 -1 97 <i>date</i>	- Filter error / Incorrect entry No answer Don't know Ongoing
index	Index	1 2 3 4 5	- First episode in parents' house Second episode in parents' house Third episode in parents' house Fourth episode in parents' house Fifth episode in parents' house
flag_a_begin	Flag_a: Beginning	21 24 27	21 original information: month 24 original information: month 27 original information: month

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
flag_a_end	Flag_a: End	30	30 original information: month
		32	32 original information: month
		21	21 original information: month
		24	24 original information: month
		27	27 original information: month
flag_b_begin	Flag_b: Beginning	30	30 original information: month
		32	32 original information: month
		-3	-3 does not apply
flag_b_end	Flag_b: End	-2	-2 original information: month
		-1	-1 original information: month
		-3	-3 does not apply
		-2	-2 original information: month
		-1	-1 original information: month

Anchor-parent panel data - bioparent

The file *bioparent* contains retrospective and prospective information on the anchor's biological, adoptive, or stepparents. The data set is provided in long-long format. This means that for each parent there is one row per wave as long as the respective parent is relevant, e.g. if a biological parent breaks up with a stepparent, the respective stepfather or stepmother is no longer considered. *Bioparent* is updated every wave by renewing the information on existing parents and/or by adding information on new (step)parents. The available data are based on the survey waves 1 to 4.

Table 4.35: List of variables included in data set *bioparent*

Variable	Variable label	Values	Value labels
id	Person number anchor	(see table 2.2)	-
parid	Person number parent (parent survey)		-
parentid	Person number parent		-
intdat	Date of interview		-
wave	Survey year	1 2 3 4	2008/09 2009/10 2010/11 2011/12

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
typepar	Type of parent	1	Biological mother
		2	Biological father
		3	Adoptive mother
		4	Adoptive father
		5	Stepmother #1
		6	Stepfather #1
		7	Stepmother #2
		8	Stepfather #2
	
surveypar	Participation in parents survey	-7	Incomplete data
		-3	Does not apply
		0	No
		1	Yes
dobpar	Date of birth	-7	Incomplete data
dodpar	Date of death	-7	Incomplete data
		-3	Does not apply
marriedpar	Parents are married	-7	Incomplete data
		-3	Does not apply
		0	No
		1	Yes
cohabpar	Parents living together in one household	-7	Incomplete data
		-3	Does not apply
		0	No
		1	Yes
marstatpar	Marital status	-7	Incomplete data
		-3	Does not apply
		1	Single
		2	Married
		3	Civil union
		4	Divorced or dissolved civil union
partnerpar	Parent has a new partner	5	Widowed or surviving partner in a civil union
		-7	Incomplete data
		-3	Does not apply
		0	No
samepartner	Same partner as last wave	1	Yes
		-7	Incomplete data
		-3	Does not apply
		0	No
cohabpartner	Living together with this partner	1	Yes
		-7	Incomplete data
		-3	Does not apply
		0	No
		1	Yes

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
begpartner	Beginning of this partnership	-7	Incomplete data
		-3	Does not apply
livanchor	Living together with anchor	-7	Incomplete data
		-3	Does not apply
		0	No
		1	Yes
contact	Contact with anchor	-7	Incomplete data
		-3	Does not apply
		0	No
		1	Yes
		9	Never had contact
begnocontact	Since when no contact	-7	Incomplete data
		-3	Does not apply
flag1	Inconsistency: identification of new stepparent	0	No inconsistency
		1	Inconsistency

In addition to the anchor's ID (*id*), the data set contains two identification numbers for the parents. First, *parid* identifies those parents who took part in the parents' survey (in any wave). This variable can be used for matching *bioparent* with specific waves of the parents' survey. Additionally, the variable *surveypar* was included in order to identify parents who participated in the parents' survey in the respective wave. Second, *parentid* serves as ID for all parents, no matter whether they took part in the parents' survey or not. The *parentid* variable allows for a differentiation between biological, adoptive, and social parents, whereas *parid* only distinguishes between biological/adoptive versus social parents. Thus, *parid* and *parentid* can differ for the same parent.

The variable *typepar* indicates the sex and the status of the parent, i.e., whether the parent is an adoptive, biological, or social mother or father. If a stepparent drops out of the survey due to separation or death and the biological father or mother begins a new partnership, then the new stepparent will be consecutively numbered (e.g. stepfather #1, stepfather #2). If a distinction between old (#1) and new (#2) stepparent is not clearly possible, the variable *flag1* is assigned the value 1.

The variables *dobpar* and *dodpar* indicate the date of birth and, if relevant, the date of death of the respective parent. These variables are provided as the number of months passed since January 1900. In cases when seasonal information of birth (or death) is available but month of birth (or death) is not, the monthly information was randomly generated by imputation.

The variables *marriedpar*, *cohabpar*, and *marstatpar* are generated for adoptive and biological parents only. Accordingly, the value "-3 does not apply" is assigned for all social parents. While *marriedpar* covers the question of whether the parent is married, *cohabpar* provides information on whether the parent is living together with the other biological or adoptive parent of the anchor in one household. Finally, *marstatpar* further characterizes the legal marital status of the parent.

The variables *partnerpar*, *samepartner*, and *cohabpartner* indicate whether the anchor's adoptive or biological parent has a new partner (if this parent is separated from the other anchor's biological or adoptive parent), whether this new partner is the same partner as in the last wave, and whether the parent is living together with this new partner in one household. Information on the beginning of the new partnership is given in the variable *begpartner*. Like all other episode data, *begpartner* is coded as number of months passed since January 1900.

Aspects of the anchor-parent relationship are covered by the variables *livanchor*, *contact*, and *begnocontact*. The variable *livanchor* indicates whether the anchor and the parent are living together in one household at a certain wave. The variable *contact* tells us whether there is or ever was contact between the anchor and the parent. If there is no contact in a certain wave, *begnocontact* indicates the beginning of this episode.

5. Partner data

TNS Infratest compiled a raw data set from the returned PAPI questionnaires. Data preparation followed closely the procedures for cleaning the anchor data.

5.1 Data editing

This section describes the editing of wave 1-5 partner data. Data editing includes general information on names of variables, value labels, and missing values. Furthermore, this section illustrates how open answers and the issue of depersonalized data were handled, as well as how various checks were performed.

Variable and value labels

As mentioned in chapter 2, all variables in the partner data start with the prefix “p-”. To facilitate the analysis, variables derived from questions that were also included in the anchor survey were labeled with the same variable name (plus the p). Variables unique to the partner survey received unique names according to our system of variable names. Values were labeled according to the partner survey codebook.

Missing values

Missings were defined and labeled as in the anchor data set, with three exceptions. First, code “-9 Invalid multiple answer” was assigned if the respondent had checked more than the allowed number of boxes. Second, codes “-6” and “-4” were collapsed into the single code “-4 Filter error / Incorrect entry/Unreadable open answer”. And finally, we did not check the partner data for consistency across variables. Therefore, we did not assign missing code “-5 Inconsistent value”. Table 5.1 shows the missing codes and value labels assigned to the partner data.

Table 5.1: Missing codes in data sets *partner\$*

Value	Label
-1	Don't know
-2	No answer
-3	Does not apply
-4	Filter error / Incorrect entry / Unreadable answer
-9	Invalid multiple answer

Open answers

Only two string variables had to be recoded in the partner data in wave 1; one had to be recoded in waves 2, 3 & 5, none in wave 4. For variable *psd27o* (included in wave 1, 3 & 5 data) we assigned the correct category contained in the answer list, if appropriate, and set the original answer to missing (“-4”) afterwards. For waves 1 & 2, we assigned the reason (or reasons) against having children listed in items *pfri13i1*, . . . , *pfri13i13* if possible, and recoded the open answer. The original answer was set to missing (-4) afterwards. The remaining open answers were then depersonalized.

Anonymity

Open answers stored in the string variables *psd27o* and *pfrt13i14o* that could not be assigned to a category of the corresponding answer list were recoded to value “1” (“Other certificate mentioned” and “Other reason mentioned”, respectively). In the partner data, depersonalization affected only one additional variable: the partner’s day of birth (*pdobd*), which was recoded to value “1” (“Day mentioned”).

Value and filter checks

To check value ranges and filters, we followed the same procedures as for the anchor data.

Checks for consistency

One difference of the partner data compared to the anchor data is that we did not check data consistency across answers.

English data

As a final step, we produced an English-language data set in which variable and value labels have been defined corresponding to the English partner codebook.

5.2 Generated variables and scales

Another major difference to the editing of the anchor data is that, with few exceptions, we did not produce user-friendly partner data.

The exceptions for wave 2 are:

In wave 1 we asked if the respondent had own or adopted children. In wave 2 we asked about the number of own, adopted, step, and foster children. We used different variable names for these concepts in wave 1 (*psd9*) and wave 2 (*psd190*). In wave 2 we additionally created a new variable *pkid* which differentiates only between having children and having no children. However, *psd9* (wave 1) and *pkid* (wave 2) are not fully congruent because *psd9* refers only to own and adopted children while *pkid* refers to own, adopted, step, and foster children.

In wave 2, we also created the additional variable *pigr27*, which indicates whether father and mother are married to each other, and *pigr28*, which indicates whether mother and father live together in one household.

The Stata do-file `genvars_partner.do` contains the syntax used to compute the variables for wave 2.

Note that to produce some of the generated variables included in the anchor data set, we used information given by the partner (see chapter 4.2).

We refrained from constructing weights for this sample.

The generated variables described above are included in second-wave partner data. In contrast, the scale variables (see table 5.2) are not part of the delivered data set. They can be generated by running the syntax file `scales_partner$` which is available for each wave on the pairfam website and as part of the scientific use file. For further information, refer to the scales handbook (Thönnissen et al. 2014) which contains a detailed description of all scales.

Table 5.2: List of scales included in data sets *partner\$*

Construct	Variable name	Wave
Traditional concept of marriage	ptradmarr	1,3,5
Value of Partnership: Negative expectations	pvpneg/pvpneg2	1/3,5
Value of Children: Benefit of stimulation	pvocbstim	1,2,4
Value of Children: Costs of comfort	pvocccomf	1,2,4
Partnership: Conflict	pconfl_apd	all
Partnership: Intimacy	pintim_aps	all
Partnership: Esteem	padmir_apo	all
Partnership: Dominance	pdomin_apo	all
Partnership: Emotional ambivalence	pambiv_apd	1,3,5
Partnership: Anxiety about loss of love	plovewitanx_apd	1,3,5
Partnership: Anxiety about being absorbed	penganx_apd	1,3,5
Partnership: Independence	pindep_apd	1 - 3,5
Partnership: Own partnership satisfaction (global scale)	psatpart	1,3,5
Partnership: Feelings of competence in the partnership	pcomppart2	2,5
Partnership: Future orientation	pcomfut_apd	1 - 3,5
Partnership: Tolerance of conflicts	pcomctol_apd	1 - 3,5
Partnership: Orientation of reciprocity Self	preciproc_aps	1,3,5
Partnership: Hostile attributions towards anchor (self assessment)	phostattr_aps	3,5
Partnership: Frequency of manifest conflicts	pconfldom_apd pconfldom2_apd	1,5 3
Partnership: Verbal aggression Partner	pverbaggr_apo	all
Partnership: Verbal aggression Self	pverbaggr_aps	all
Partnership: Constructive behavior Partner	pconstrbh_apo	all
Partnership: Constructive behavior Self	pconstrbh_aps	all
Partnership: Withdrawal Partner	pwithdraw_apo	all
Partnership: Withdrawal Self	pwithdraw_aps	all
Partnership: Manipulation Partner	pmanipul_apo	1,3,5
Partnership: Manipulation Self	pmanipul_aps	1,3,5
Partnership: Dyadic coping Partner	pdycop_apo	1,3,5
Partnership: Dyadic coping Self	pdycop_aps	1,3,5
Partnership: Instability of partnership	pinstab_apd	all
Explosiveness and tendency to anger	pexplosive	1,3 - 5
Shyness	pshyness	1,4,5 (only new partners)

continued on next page

continued from previous page

Construct	Variable name	Wave
Emotional autonomy	pemotautn	1,3 - 5
Selfworth	pselfworth	all
Depressiveness	pdepressive	2 - 5
BIG 5: Neuroticism	pneurot	2 - 5 (only new partners)
BIG 5: Extraversion	pextrav	2 - 5 (only new partners)
BIG 5: Agreeableness	pagreeable	2 - 5 (only new partners)
BIG 5: Conscientiousness	pconscient	2 - 5 (only new partners)
BIG 5: Openness	popenness	2 - 5 (only new partners)
Coparenting with the other parent	pcoparent	2,4
Parental Self Efficacy/Competence	pcomperz	2,4
Parenting: Sacrifice in raising children	psacrif_pacs	3,5
Parenting: Recognition / support in education through partner	ppartnersup	3,5
Parenting: Autonomy in the parenting role	pautonom	4
Parenting: Autonomy in the parenting role (short scale)	pautonom2	4
Parenting: Pleasure in the parenting role	ppleasure	4
Parenting goals: Status	ppgoalstatus	2,4,5
Parenting goals: Competence	ppgoalcomp	2,4,5
Parenting goals: Autonomy	ppgoalautn	2,4,5
Intergenerational relationships: Own intimacy within partner-mother dyad (self assessment)	pintim_ams	3 - 5
Intergenerational relationships: Own intimacy within partner-father dyad (self assessment)	pintim_afs	3 - 5
Intergenerational relationships: Conflict within partner-mother dyad	pconfl_amd	3 - 5
Intergenerational relationships: Conflict within partner-father dyad	pconfl_afd	3 - 5

6. Parent data

TNS Infratest compiled a raw data set from the returned parent's PAPI questionnaires. Data editing followed closely the procedures used to clean the anchor data.

6.1 Data editing

This section describes the editing of wave 2, 3, 4, and 5 parent data. The complete data processing was done in Stata and all data editing steps are documented in a Stata do-file. The data editing process comprised two main tasks: the cleaning and debugging of the raw data on the one hand, and the generation of indicators on the other hand. This included labeling variables, variable values, and missing values. Furthermore, this section illustrates how the subject of anonymization was handled, as well as how various checks were performed.

Deletion

The raw data provided by *TNS Infratest* contained 5,039 cases in wave 2. A number of cases were deleted from this original data file either because they were completely empty¹ or because they were suspected to have been derived from duplicate questionnaires or to have been provided by the anchor². Thus, the parent file was reduced by 24 cases, which resulted in a final data file of N=5,015. There were no cases deleted in the third wave. The final data set contains N=3,946 in wave three. In wave four, two cases were deleted³ and the data sets consists of 3,350 entries. In wave five no cases were deleted resulting in a final data file including N=3,546 cases.

Variable and value labels

All variables in the parent data start with the prefix "par-". To facilitate analysis, variables derived from questions that were also included in the anchor survey were labeled with the same variable name (plus the prefix par-). Variables unique to the parent survey had unique names according to our system of variable names.

Every variable in the data set was assigned a label. Variable labels contain a short description of the variable and the position in the parents questionnaire (question number). The values of all variables were labeled according to the parent's codebook.

Missing values

Missings were defined and labeled as in the anchor data set, with one exception. Although the answer category "don't know" (respondent could not answer) was allowed very rarely and thus could not be distinguished precisely from "no answer" (respondent did not answer), code "-2 No answer" was consistently assigned if no box had been checked. This is consistent with the procedure applied for editing the partner data. Table 6.1 shows the missing codes and value labels assigned to the parent's data.

¹N=17; parid = 267932302, 32927301, 144352301, 144352304, 15196302, 233346301, 233346302, 280325304, 337872301, 337872304, 457711301, 457711302, 459740301, 459740302, 632879301, 632879302, 666208301

²N = 7; parid = 33691301, 33691304, 77993302, 257271301, 257271302, 567208301, 567208302

³N = 2; parid = 750419301, 750419302

Code “-9 Invalid multiple answer” was assigned if the respondent had ticked more than the allowed number of boxes, which is rather common in PAPI-questionnaires. The same applies to filter errors and incorrect data entries indicated by missing code “-4” and to inconsistencies between the answers of a respondent that were coded to “-5 Inconsistent value”, provided it was uncovered which value was wrong.

Table 6.1: Missing codes in data set *parent\$*

Value	Label
-1	Don't know
-2	No answer (also: I don't want to answer that, answer refused)
-3	Does not apply (filter)
-4	Filter error / Incorrect entry
-5	Inconsistent value
-6	Unreadable answer
-7	Incomplete data (for generated variables)
-9	Invalid multiple answer

Filter checks

We checked every filter in the parents' data. If a question was skipped by the respondent correctly, the variable was set to “-3 Does not apply” (filter). There are two sources of mistaken filters. First, the respondent may have misread the filter and answered a question by mistake. In this case the affected variable was regarded as “Filter error / Incorrect entry” and set to value “-4”. Second, the respondent entered the answer to a filter question incorrectly, but then continued correctly. If there are indications of this (i.e., a certain number of questions subsequent to a filter question were answered), the affected variables were not regarded as filter errors. Instead, the answers were kept, but the filter question itself was set to “-5 Inconsistent value”.

A complex filter was applied to lead the respondent to one of the CAPI children (which is one selected child of the anchor aged between 8 and 15) as the reference grandchild for the subsequent grandchild module. Accordingly, a complex filter check was applied to uncover whether the respondent had answered with the correct reference child in mind.

In wave 2: Three conditions had to be met: First, the filter questions 16 to 18 had to be correct. Second, the sex of the grandchild indicated in question 20 had to match the sex of the CAPI-child (provided in the anchor's interview). Third, the age of the grandchild indicated in question 21 had to be within the age range of 5 to 18 (which is purposely somewhat broader, as some miscalculation of the respondents is accepted).

In wave 3: Again, three conditions had to be met: First, the filter questions 9 had to be correct. Second, the sex of the grandchild indicated in question 10 had to match the sex of the CAPI-child (provided in the anchor's interview). Third, the age of the grandchild indicated in question 11 had to be within the age-range of 5 to 18 (which is purposely somewhat broader, as some miscalculation of the respondents is accepted).

In wave 4: Again, three conditions had to be met: First, the filter questions 14 had to be correct. Second, the sex of the grandchild indicated in question 15 had to match the sex of the CAPI-child (provided in the anchor's interview). Third, the age of the grandchild indicated in question 16 had to be within the age-range of 5 to 18 (which is purposely somewhat broader, as some miscalculation of the respondents is accepted).

In wave 5: Also in this wave the three conditions had to be met: First, the filter question 11 had to be correct. Second, the sex of the grandchild indicated in question 13 had to match the sex of

the CAPI-child (provided in the anchor's interview). Third, the age of the grandchild indicated in question 14 had to be within the age range of 5 to 18 (which is purposely somewhat broader as some miscalculation of the respondents is accepted).

Checks for consistency

Various checks to identify logically impossible or empirically implausible answers were conducted. The code "-5" was assigned if it was possible to identify which variable under consideration was wrong. The values were left unedited if the inconsistency could not be solved, but were flagged with a generated variable referring to the respective inconsistency. It is recommended to analyze flagged variables with caution. Table A.23 describes the flag variables in detail.

Anonymity

Answers that might threaten our respondents' anonymity are not contained in the data set. This refers to all questions where the respondents provided names. These variables had already been dropped from the data set by *TNS Infratest*.

English data

An English-language data file was produced. English variable and value labels were assigned according to the English parent codebook.

6.2 Generated variables and scales

Table 6.2 displays all generated parent variables along with the relevant paradata and variables meant to facilitate merging the parent data set with other data.

Generated variables

First of all, the file contains a number of paradata variables. The variables *parintm*, *parintd*, and *parinty* indicate the date (month, day, and year) the respondent filled in the questionnaire. In addition, the variable *parlng* in wave 2 through wave 4 indicates the language version of the questionnaire (German, Russian, or Turkish). Starting in wave 5, only a German language version is used. The variables *partype*, *parposition*, *parsexparent*, and *parsexanchor* refer to the relationship between the respondent and the anchor (biological or stepparent, mother or father, daughter or son). The variable *parageanchor* provides the anchor's year of birth taken from the anchor interview, as well as the variable *parcohort*.

In addition, similar to the anchor data, user-friendly parents' data are produced, but only on small scale. On the one hand, based on the information given by the respondent, some socio-economic and demographic indicators were generated that are comparable to the indicators provided in the anchor's file (see do-file *genvars_parents.do*). On the other hand, command-files named *scales_parents\$* are provided (SPSS-syntax and Stata do-files) that allow the computation of a number of scales (see Thönnissen et al. 2014). However, unlike other indicators, these scale indicators are not included in the data file.

Note that in release 4.0 variables *parschool*, *parvocat*, *parcasmin*, *parisced*, and *paryeduc* are based on the generated variables *mschool*, *fschool*, *mvocat*, and *fvocat* of the anchor data sets (for details see chapter 4.2 and do-file *education_parents.do*).

Beside the respondent's identification number (*parid*), the anchor's identification number (*id*), and the CAPI-child's identification number (*cid*) are also provided for merging the data files. As of release

5.0, the anchor's siblings' numbers (*sibidx*) are included as well in all parent data sets (also see chapter 2.2).

Table 6.2: List of generated variables included in data set *parent\$*

Construct	Variable name	Wave
<i>Identification</i>		
Anchor's identification number	id	2 - 5
Respondent's identification number	parid	2 - 5
CAPI-child's identification number	cid	2 - 5
Anchor's siblings' number	sibidx	2 - 5
<i>Paradata</i>		
Respondent's type (based on questionnaire)	partype	2 - 5
Sex of respondent (based on questionnaire)	parseparent	2 - 5
Position of respondent (based on questionnaire)	parposition	2 - 5
Sex of the anchor (based on questionnaire)	parseanchor	2
Wave (based on questionnaire)	wave	2 - 5
Language version (based on questionnaire)	parlng	2 - 4
Date of the interview	parintd, parintm, parinty	2 - 5
Cohort anchor	parcohort	2 - 5
Year of birth anchor based on anchor interview	parageanchor	2 - 5
# of children (corrected): # of children in parsd32; only in case of more children described in parsd14ff, than stated in parsd32: correction; -7 if no valid answer in parsd32 & no children described	parkids	2 - 5
# of grandchildren	pargrandkids	2 - 5
<i>Generated variables</i>		
Age variables (respondent, respondent's parents)	parage parage2 parmage par- mage2 parfage parfage2 parpage parpage2	2 - 5 2
Current primary and secondary activity status	parcasprim, parcassec	2 - 5
Highest school degree	parschool	2 - 5
Highest vocational degree	parvocat	2 - 5
CASMIN classification of educational attainment	parcasmin	2 - 5
ISCED classification of educational attainment	parisced	2 - 5
Years of schooling / vocational qualification	paryeduc	2 - 5
Labor force status	parlfs	2 - 5
Household size main residence	parhhsizemrd	2 & 4
Net equivalence income	parhhincgee	2
Type of children	park1type park2type park3type park4type	2 - 5

continued on next page

continued from previous page

Construct	Variable name	Wave
Marital status	parmarstat	2 & 4
Relationship status	parrelstat	2 & 4
Do parents live in a shared household?	parigr27	2 - 5
Are parents married?	parigr28	2 - 5
State / Bundesland	parbula	3 - 5
Size of community in 7 categories	pargkpol	3 - 5
Settlement structure	parbik	3 - 5
New parent in wave 3	parw3ne	3
New parent in wave 4	parw4ne	4
New parent in wave 5	parw5ne	5
Information about the correct grandchild	parcorgc	3 - 5

Scales

The following scale variables (see table 6.3) can be generated by running the do-file `scales_parents$`, which can be found on the pairfam website and on CD. For further information, refer to the scales manual (Thönnissen et al. 2014) which contains a detailed description of all scales.

Table 6.3: List of scales included in data set *parent\$*

Construct	Variable name	Wave
Filial Obligation (mean indicator: parval2i1, parval2i5)	parobligationf	2
Grandparental Obligation (mean indicator: parval2i2, parval2i4)	parobligationp	2
Parental Obligation (mean indicator: parval2i3, parval2i6)	parobligationgp	2
Traditional concept of marriage (mean indicator: parval1i2, parval1i7, and parval1i8)	partradmarr	3,5
Satisfaction (mean indicator: parsat1i1, parsat1i2, parsat1i3, parsat1i4)	parsat	5
Readiness to make sacrifices (mean indicator: parcrn32i1, parcrn32i2, and parcrn32i3)	parsacrif_pacs	3,5
Frequency of Joint Activities (mean indicator: parigr73a, parigr74a, parigr75a, parigr76a, and parigr77a)	paractiv_paras	2,4
NRI Partner Conflict	parconfl_parparpd	2,3,5
NRI Partner Intimacy	parintim_parparps	2,3,5
NRI Partner Approval	paradmir_parparpo	2,3,5
NRI Partner Dominance	pardomin_parparpo	2,3,5
NRI Anchor Conflict	parconfl_parad	2 - 5
NRI Anchor Intimacy	parintim_paras	2 - 5
NRI Anchor Approval	paradmir_parao	2,4
NRI Anchor Dominance	pardomin_parao	2,4
Co-parenting with respect to anchor (younger than 21) (mean indicator: parcrn21i1, parcrn21i2, parcrn21i3)	parcoparent	2,4

continued on next page

continued from previous page

Construct	Variable name	Wave
Educational style with respect to anchor (younger than 21): Monitoring (mean indicator: parcr1i2, parcr1i9, parcr1i6, parcr1i12)	parmonitor_paras	2,4
Educational style with respect to anchor (younger than 21): Strict control (mean indicator: parcr1i18, parcr1i19, parcr1i20, parcr1i21)	parstrict_paras	2,4
Educational style with respect to anchor (younger than 21): Emotional Warmth (mean indicator: parcr1i1, parcr1i5, parcr1i14)	parwarmth_paras	2,4
Educational style with respect to anchor (younger than 21): Negative Communication (mean indicator: parcr1i3, parcr1i8, parcr1i11)	parnegcomm_paras	2,4

7. Child data

Editing of the second-, third-, and fourth-wave child data was conducted in line with the procedures to clean the anchor data and the additional alteri data. The complete data processing was done in SPSS.

7.1 Data editing

This section describes the editing of wave 2-5 child data. Data editing includes general information on names of variables, value labels, and missing values. Furthermore, this section illustrates how open answers and issues of depersonalization were handled, as well as how various checks were performed.

Variable and value labels

All variables in the child data start with the prefix “c-”. Every variable in the data set was assigned a label. These variable labels contain the wording of the corresponding question in the child interview. The values of all variables were labeled according to the child codebook.

Missing values

For all variables of the child data, we defined a set of missing codes that were applied throughout (see table 7.1). Missing values “-1 Don’t know” respectively “-2 No answer” have been assigned if the child could not or did not want to answer a question. These two codes are the only missing values also documented in the codebook.

Value “-3 Does not apply” was assigned if a child had not been asked the corresponding question, i.e., if the child was filtered over the question. Errors in the Child-CAPI program that guided children to the wrong questions in the interview were indicated by missing code “-4 Filter error / Incorrect entry”, as are incorrect data entries by the interviewers.

Missing value “-5 Inconsistent value” was assigned if logically impossible or empirically implausible combinations of values on two or more variables were found and if it was clear that the value was wrong. For open answers that were not legible, we assigned value “-6 Unreadable answer”. For generated variables, we used value “-7 Incomplete data” to indicate cases where we lacked the information necessary to compute a valid value.

Table 7.1: Missing codes in data set *child\$*

Value	Label
-1	Don’t know
-2	No answer (also: I don’t want to answer that, answer refused)
-3	Does not apply (filter)
-4	Filter error / Incorrect entry
-5	Inconsistent value
-6	Unreadable answer
-7	Incomplete data

Checks for consistency

In order to detect inconsistencies between the answers of a child, we checked for logically impossible or empirically implausible combinations of values on two or more variables. Only very few inconsistencies were found and it was not possible to resolve the inconsistencies by assigning the code “-5” because we could not determine which of the variables under consideration was wrong. Consequently, the values provided were left unedited. Instead, a flag variable indicating the respective inconsistency was generated. For each of these variables, code “0” indicates that the respective inconsistency is non-existent. Table 7.2 describes the flag variables in detail.

Table 7.2: List of flag variables to identify inconsistencies (*child\$*)

Variable	Label	Value	Value Labels
cflag1	Inconsistency sex of child between statement of interviewer and anchor	0	Non consistency
		1	Inconsistency
cflag2	Inconsistency year of child’s birth between statement of anchor and child	0	No inconsistency
		1	Inconsistency

In addition, we computed tag variables to indicate inconsistencies over time, i.e. over waves. These relate to the children’s sex and date of birth. Table 7.3 lists all of the generated tag variables that are part of the data set *child\$*.

Table 7.3: List of tag variables to identify inconsistencies between waves

Variable	Label	Value	Value Labels	Description
tag_csex	Inconsistency between waves: child’s sex	0	No consistency	Child’s sex in current wave is not child’s sex in previous wave
		1	Inconsistency	
tag_cdobm	Inconsistency between waves: child’s month of birth	0	No inconsistency	Child’s month of birth in current wave is not child’s month of birth in previous wave
		1	Inconsistency	
tag_cdoby	Inconsistency between waves: child’s year of birth	0	No inconsistency	Child’s year of birth in current wave is not child’s year of birth in previous wave
		1	Inconsistency	

Coding open answers

In coding open answers, we adhered to the following procedure. For variable *cedu1ao* (Other school type), which contained information further qualifying the residual category of an answer list, we compared the open answer to the answer list. If appropriate, we recoded the open answer into an existing category, and set the original value to missing (code “-4”). All remaining open answers were coded to a single value indicating merely that an open answer has been provided. The actual string was deleted for privacy reasons.

Anonymity

Answers that might threaten the children’s anonymity were deleted or recoded in the data set. We also deleted the information on exact dates, i.e. the day components, for privacy concerns.

By means of depersonalization, all string variables in the data set were finally transformed to numeric variables. There are thus no string variables contained in the child data set. Valid answers to open questions were recoded to value “1” throughout. The variables affected by this procedure are shown in table 7.4 along with the value labels (<information> mentioned) assigned to valid answers on the anonymous numeric variables.

Table 7.4: List of anonymized variables

Variable	Variable label	Anonymized value label	Wave
cdobd	Child day of birth	Day mentioned	2 - 5
cedu1o	Other schooltype	Other schooltype mentioned	2
cedu1ao	Other schooltype	Other schooltype mentioned	3 - 5
cpcr4	Name anchor	Name mentioned	2 - 5
cpcr5	Name current partner of anchor in same household	Name mentioned	2 - 5
cgp1	Name of anchor's mother	Name mentioned	2 & 4
cgp2	Name of anchor's father	Name mentioned	2 & 4
cpcr13	Name biological parent outside the household	Name mentioned	3 & 5
crom7	Name of romantic partner	Name mentioned	5
ccp2	Stepup's complete name and address	Address received/ Same address as anchor	5

Value and filter checks

We used the same procedures as for the anchor data to check for value ranges and filters. These checks were performed using a SPSS routine.

English data

In order to enable non-German-speakers to use pairfam data, we created the English data set. All variable labels and value labels have been translated according to the wording of the English child codebook.

7.2 Generated variables and scales

In order to facilitate data analysis and to enhance comparability of results, the pairfam staff produced a number of variables that are of interest to many research projects. A list of all generated variables is shown in table 7.5.

The Stata do-file `genvars_child` contains the syntax used to compute some of the variables. Due to data privacy, syntaxes are not available for all of the generated variables.

Table 7.5: List of generated variables included in data set *child\$*

Construct	Variable name	Wave
relstatac	Relationship/kinship between anchor and child	2 - 5
cinty	Child interview year	2 - 5
cintm	Child interview month	2 - 5
cintd	Child interview day	2 - 5
cagey	Child's age in years	2 - 5
cagem	Child's age in months	2 - 5
ctitlea	Child's naming for anchor	2 - 5

continued on next page

continued from previous page

Construct	Variable name	Wave
ctitlep	Child's naming for anchor's partner	2 - 5
ctitleop	Child's naming for the other biological parent outside the household	3, 5

The generated variables are included in wave 2-5 child data. In contrast to these, the scale variables (see table 7.6) are not part of the delivered data set. They can be generated by running the syntax file `scales_child$` which can be found as part of the scientific use file and on the pairfam website. For further information, refer to the scales manual (Thönnissen et al. 2014), which contains a detailed description of all scales.

Table 7.6: List of scales included in syntax file `scales_child`

Construct	Variable name	Wave
Class atmosphere	cclassatmo	2 - 5
Social integration	csocialinteg	2 - 5
Peer rejection	cpeerref	2 - 5
Parental school engagement	cengagement	2,4
Economic deprivation	cecodep	2 - 5
Intimacy child-best friend-relationship	cintim_cbf	3,5
Companionship child-best friend-relationship	ccomp_cbf	3,5
Emotional warmth child-anchor-relationship	cwarmth_cao	2 - 5
Emotional warmth child-partner-relationship	cwarmth_cpo	2 - 5
Emotional warmth child-other parent-relationship	cwarmth_copo	3,5
Strict control child-anchor-relationship	cstrict_cao	2,4
Strict control child-partner-relationship	cstrict_cpo	2,4
Strict control child-other parent-relationship	cstrict_copo	3,5
Monitoring child-anchor-relationship	cmonitor_cao	2 - 5
Monitoring child-partner-relationship	cmonitor_cpo	2 - 5
Fear of love withdrawal child-anchor-relationship	lovewitanx_cas	3 - 5
Fear of love withdrawal child-partner-relationship	lovewitanx_cps	3 - 5
Fear of love withdrawal child-other parent-relationship	lovewitanx_cops	3,5
Intimacy child-anchor-relationship	cintim_cas	2 - 5
Intimacy child-partner-relationship	cintim_cps	2 - 5
Intimacy child-other partner-relationship	cintim_cops	3,5
Conflict child-anchor-relationship	cconfl_cad	2 - 5
Conflict child-partner-relationship	cconfl_cpd	2 - 5
Conflict child-other partner-relationship	cconfl_copd	3,5
Admiration child-anchor-relationship	cadmir_cao	2 - 5
Admiration child-partner-relationship	cadmir_cpo	2 - 5
Admiration child-other parent-relationship	cadmir_copo	3,5
Parental reliability child-anchor-relationship	creliabl_cad	2

continued on next page

continued from previous page

Construct	Variable name	Wave
Parental reliability child-partner-relationship	creliabl_cpd	2
SDQ: Conduct problems	cconduct	2 - 5
SDQ: Emotional symptoms	cemotion	2 - 5
SDQ: Prosocial behavior	cprosoc	2 - 5
SDQ: Hyperactivity	chyper	5
Negative communication in the anchor-child relationship	cnegcomm_cao	4, 5
Negative communication in the partner-of-the-anchor-child relationship	cnegcomm_cpo	4, 5
Intimacy own romantic partner	cintim_crps	5
Conflict own romantic partner	cconfl_crpd	5
Admiration own romantic partner	cadmir_crpo	5
Fear of love withdrawal own romantic partner	clovewitanx_crpd	5
Ambivalence in relationship to own romantic partner	cambiv_crpd	5

8. Parenting data

TNS Infratest compiled a raw data set from the returned PAPI questionnaires. Data editing followed closely the procedures to clean the anchor data.

All anchors with children who consented to the children’s survey for at least one child were asked to fill out the parenting questionnaire, as were their cohabiting partners. Information from anchor and partner is included in one single data set.

8.1 Data editing

This section describes the editing of wave 2-5 parenting data. Data editing includes general information about names of variables, value labels, and missing values. Furthermore, this section illustrates how issues of depersonalization were handled, as well as how value checks were performed.

Variable and value labels

The parenting variables start without any prefix. Values were labeled according to the parenting questionnaire codebook.

Missing values

Missing values were defined and labeled as in the anchor data set. Table 8.1 shows the missing codes and value labels assigned to the parenting data.

Missing values “-1 Don’t know” respectively “-2 No answer” have been assigned if the respondent could not or did not answer a question. In wave 2 the answer category “Don’t know” (respondent could not answer) did not exist and thus could not be distinguished precisely from “No answer” (respondent did not answer). Therefore code “-2 No answer” was consistently assigned if no box had been checked. Code “-9 Invalid multiple answer” was assigned if the respondent had ticked more than the allowed number of boxes, which is rather common in PAPI-questionnaires. As there were no filters and no open answers in the parenting questionnaires of wave 2-4, there were no missing codes “-3” and “-6” necessary. Also, there were no consistency checks possible because we only assessed subjective perceptions. Beginning with wave 5 we have one filtered question (*edu6*). Therefore the value labels “-3 Does not apply” (if a question was skipped by the respondent correctly) and “-5 Inconsistent value” (if logically impossible or empirically implausible combinations of values were found) were assigned.

Table 8.1: Missing codes in data set *parenting*

Value	Label
-1	Don’t know
-2	No answer (also: I don’t want to answer that, answer refused)
-3	Does not apply
-5	Inconsistent value
-9	Invalid multiple answer

Anonymity

The children's names were part of the parenting questionnaire so as to facilitate the attribution of the questionnaires. For privacy concerns, this information was deleted in the data set.

Value checks

To check value ranges, we used SPSS routines to identify incorrect entries.

English data

In the final step, we produced an English data set in which variable and value labels have been defined corresponding to the English parenting codebook.

8.2 Generated variables and scales

A major difference to the editing of the anchor data is that we did not produce user-friendly parenting data. Therefore, there are no generated variables included in second wave parenting data. However, it is possible to generate scale variables (see table 8.2) not included in the delivered data set. The file `scales_parenting$` contains the required syntax (in both Stata and SPSS) and is part of the scientific use file and available on the pairfam homepage. For further information, refer to the scales manual (Thönnissen et al. 2014), which contains a detailed description of all scales.

Table 8.2: List of scales included in syntax file `scales_parenting`

Construct	Variable name	Wave
Parent-child-relationship: Readiness to make sacrifices	<code>sacrif_pacs</code>	2
Parent-child-relationship: Hostile attribution	<code>hostattr_pacd</code>	2 - 5
Autonomy in the parenting role	<code>autonom_pacs</code>	2
Autonomy in the parenting role-short scale	<code>autonom2_pacs</code>	2
Pleasure in the parenting role	<code>pleasure_pacs</code>	2
Parent-child-relationship: Emotional warmth	<code>warmth_pacs</code>	2 - 5
Parent-child-relationship: Psychological control	<code>psycontrol_pacs</code>	2,5
Parent-child-relationship: Negative communication	<code>negcomm_pacs</code>	2 - 5
Parent-child-relationship: Monitoring	<code>monitor_pacs</code>	2 - 5
Parent-child-relationship: Strict control	<code>strict_pacs</code>	2,5
Parent-child-relationship: Inconsistent parenting	<code>inconsist_pacs</code>	2 - 5
SDQ: Hyperactivity	<code>hyper_paco</code>	2,5
SDQ: Emotional symptoms	<code>emotion_paco</code>	2 - 5
SDQ: Peer problems	<code>peer_paco</code>	2,5
SDQ: Prosocial behavior	<code>prosoc_paco</code>	2 - 5
SDQ: Conduct problems	<code>conduct_paco</code>	2 - 5
Parent-child-relationship: Intimacy	<code>intim_paco</code>	2 - 5
Parent-child-relationship: Conflict	<code>confl_pacd</code>	2 - 5
Parent-child-relationship: Admiration	<code>admir_pacs</code>	2,4,5

continued on next page

continued from previous page

Construct	Variable name	Wave
Parent-child-relationship: Dominance	domin_paco	2 - 5

9. Outlook

This manual will be updated every year resp. for every new release. The new document will be made available upon the next data release.

Wave 6 will bring new data sets for anchor respondents and partners plus data sets on the parents survey, the child interview, and the parenting survey. Substantively, the rotating modules on intergenerational relationships, family values, fertility, and the Big Five will be part of the anchor interview in wave 6. Additionally, two modules concerning aspects of work-life-balance and critical life events will be implemented for the first time in round 6 of the pairfam project.

Of course, to reflect upon changes in our respondents' lives, the generated variables and data sets, if applicable, will be updated for the next release. Again, the tasks of data production will be divided up and decentralized as has been the case beginning with the second wave.

The pairfam team would like to thank you for using this panel survey in your research. The aim of the project is to foster academic progress in the field of family research. The authors of this manual hope you found the information included here helpful for your analysis. We would be happy to receive your comments on how to improve the data and documentation at: support@pairfam.de.

Bibliography

- Arránz Becker, O., J. Brüderl, P. Buhr, L. Castiglioni, D. Fuß, V. Ludwig, C. Schmiedeberg, J. Schröder, and N. Schumann (2014). The German Family Panel - Study Design and Cumulated Field Report (Waves 1 to 5). Pairfam Technical Paper No. 01.
- Brauns, H. and S. Steinmann (1999). Educational Reform in France, West-Germany, and the United Kingdom. Updating the CASMIN Educational Classification. *ZUMA-Nachrichten* 44(23), 7–44.
- Brüderl, J., V. Ludwig, K. Pforr, and N. Schumann (2011a). Praktische Anwendungsbeispiele zum Umgang mit den pairfam-Daten (Welle 1). In J. Brüderl, L. Castiglioni, and N. Schumann (Eds.), *Partnerschaft, Fertilität und intergenerationale Beziehungen. Ergebnisse der ersten Welle des Beziehungs- und Familienpanels*, pp. 27–45. Würzburg: Ergon Verlag.
- Brüderl, J., V. Ludwig, K. Pforr, and N. Schumann (2011b). Praktische Anwendungsbeispiele zum Umgang mit den pairfam-Daten (Welle 1). Pairfam Technical Paper No. 02.
- Brix, J. and M. Pfister (2014). Gewichtung der Nettostichprobe Welle 1 des Projekts DemoDiff und der Nettostichprobe pairfam gesamt. Technical report, TNS Infratest Sozialforschung, München.
- Brix, J., P. Wich, and U. Schneekloth (2013). Beziehungen und Familienleben in Deutschland (2012/2013) Welle 5. Methods report, TNS Infratest Sozialforschung, München.
- Erikson, R., J. Goldthorpe, and L. Portocarero (1979). Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden. *The British Journal of Sociology* 30(4), 415–441.
- Frietsch, R. and H. Wirth (2001). Die Übertragung der Magnitude-Prestigeskala von Wegener auf die Klassifizierung der Berufe. *ZUMA-Nachrichten* 48.
- Ganzeboom, H. and D. Treiman (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. In J. Hoffmeyer-Zlotnik and C. Wolf (Eds.), *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables*, pp. 159–193. New York: Kluwer Academic Press.
- Ganzeboom, H. and D. Treiman (2010). International Stratification and Mobility File: Conversion Tools.
- Hartmann, J., G. Schütz, and J. Suckow (2010). Die Vercodung der Berufe bei PAIRFAM. Technical report, TNS Infratest Sozialforschung, München.
- Hartmann, J., G. Schütz, and J. Suckow (2011a). Die Vercodung der Berufe bei pairfam (welle 2). Technical report, TNS Infratest Sozialforschung, München.
- Hartmann, J., G. Schütz, and J. Suckow (2011b). Die Vercodung der Berufe bei pairfam (welle 3). Technical report, TNS Infratest Sozialforschung, München.
- Hartmann, J., N. Tschersich, and G. Schütz (2012). Die Vercodung der offenen Angaben zur beruflichen Tätigkeit nach der Klassifikation der Berufe 2010 (KldB2010) und nach der International Standard Classification of Occupations 2008 (ISCO08). Entscheidungsregeln bei nicht eindeutigen Angaben. Technical report, TNS Infratest Sozialforschung, München.

- Huinink, J., J. Brüderl, B. Nauck, S. Walper, L. Castiglioni, and M. Feldhaus (2011). Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual Framework and Design. *Zeitschrift für Familienforschung* 23, 77–101.
- König, W., P. Lüttinger, and W. Müller (1988). A Comparative Analysis of the Development and Structure of Educational Systems. Methodological Foundations and the Construction of a Comparative Educational Scale. CASMIN Working Paper No. 12, University of Mannheim.
- Kreyenfeld, M., J. Huinink, H. Trappe, and R. Walke (2012). DemoDiff: A Dataset for the Study of Family Change in Eastern (and Western) Germany. *Schmollers Jahrbuch* 132, 653–660.
- Schneider, S. (2008). Applying the ISCED-97 to the German educational qualifications. In S. Schneider (Ed.), *The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries*, pp. 76–102. Mannheim Centre for European Social Research, University of Mannheim.
- Schnell, R., T. Bachteler, and S. Bender (2004). A Toolbox for record linkage. *Austrian Journal of Statistics* 33(1-2), 125–133.
- Schnor, C. and S. Bastin (2014). Partners and babies. partnership and fertility histories of the german family panel (pairfam/demodiff, waves 1-3) as an event history data set. Pairfam Technical Paper No. 03.
- Statistisches Bundesamt (2009). Mikrozensus 2008 - Neue Daten zur Kinderlosigkeit in Deutschland. Begleitmaterial zur Pressekonferenz am 29. Juli 2009 in Berlin, Statistisches Bundesamt, Wiesbaden.
- Suckow, J. and U. Schneekloth (2009). Beziehungen und Familienleben in Deutschland (2008/2009) Welle 1. Methods report, TNS Infratest Sozialforschung, München.
- Suckow, J., U. Schneekloth, and P. Wich (2010). Beziehungen und Familienleben in Deutschland (2009/2010) Welle 2. Methods report, TNS Infratest Sozialforschung, München.
- Suckow, J., P. Wich, U. Müller, and U. Schneekloth (2011). Beziehungen und Familienleben in Deutschland (2010/2011) Welle 3. Methods report, TNS Infratest Sozialforschung, München.
- Thönnissen, C., C. Gschwendtner, B. Wilhelm, S. Fiedrich, E.-V. Wendt, and S. Walper (2014). Scales Manual wave 1 to 5. Technical report.
- UNESCO (2006). International Standard Classification of Education: ISCED-97. Technical report, Montreal: UNESCO Institute for Statistics.
- Wegener, B. (1984). Gibt es Sozialprestige? *Zeitschrift für Soziologie* 14(3), 209–235.
- Wich, P., M. Stadler, J. Suckow, and U. Schneekloth (2012). Beziehungen und Familienleben in Deutschland (2011/2012) Welle 4. Methods report, TNS Infratest Sozialforschung, München.

A. Appendix

A.1 DemoDiff-pairfam differences

Table A.1: Variables not included in DemoDiff wave 1, anchor1_DD

Variable	Variable label
co1_	Future: Importance to be successful in job (Question 5)
bce1_	VOP+: Undertake activities with partner (Question 60)
sin1	Interest of potential partner (Question 61)
sin2	Interested in potential partner (Question 62)
sin4_	Dating possibilities (Question 65)
sin5_	Trying to find out sth. about personality (Question 66)
sin6_	Somebody like me always finds a partner (Question 67)
pa1_	Interested in partner (Question 68)
pa4	Partner introduced to parents (Question 77)
pa10_	Feelings regarding common household (Question 83)
pa12_	Feelings regarding marriage (Question 85)
cps2	First request language partner survey (Question 96)
cps4	Handing over partner questionnaire right away (Question 99)
pa16_	Let partner know that I understand him/her (Question 101)
pa17_	Telling partner what you are thinking (Question 102)
pa18_	Partner finds it all right if I pursue own interests (Question 103)
pa19_	I hope relationship lasts for a long time (Question 106)
sat4	Estimation of partner's satisfaction with relationship (Question 105)
pa20_	Serious relationship problem: Prob. alcohol, medication, drugs (Question 107)
pa22xiy	Insulted or abused partner (Question 109)
pa23	Problems with alcohol, medication, or other drugs last year (Question 110)
pa24	Cheating past year (Question 111)
pa25	Arguments using physical force past year (Question 112)
pa28	Suggested a separation/divorce past year (Question 115)
sex1_	Age first sex (Question 116)
sat5	Contraception used past 3 months (Question 120)
lsr1i_	Leisure time (Question 138)
lsr2	Hours watching TV past week (Question 139)

continued on next page

continued from previous page

Variable	Variable label
lsr3	Hours spent on personal Internet use past week (Question 140)
lsr4	Weeks on vacation past year (Question 141)
lsr5_	Leisure with partner: Coffee stores, bars, restaurants (Question 142)
lsr6	Online profile on social network website (Question 143)
lsr7	Visibility of online profile (Question 144)
lsr8	Frequency of visit: Social network sites (Question 145)
hc11h1	Main residence: Monthly expenditures for rented ap./house (Question 157)
hc12h1	Main residence: Monthly expenditures for self-owned ap./house (Question 158)
hc11h2	Second Residence: Monthly expenditures for rented ap./house (Question 169)
hc12h2	Second Residence: Monthly expenditures for self-owned ap./house (Question 170)
igr6_	Year of death biological mother (Question 190)
igr8_	Year of death biological father (Question 192)
igr17_	Frequency: Arguments and fights with biol. mother (Question 202)
igr18_	Frequency: Arguments and fights with biol. father (Question 203)
sd25	Type of school currently attending (Question 208)
sd26	Grade in school (Question 209)
job5_	Owner of company: Anchor (Question 217)
job6_	Employed in own firm: Partner/spouse (Question 218)
job14	How easy to find suitable position? (Question 226)
inc1	Gross income last month (Question 232)
per1_	I often agree with others, even if I'm not sure (Question 253)
hlt2	Number of times feeling low and melancholy past 4 weeks (Question 255)
hlt3	Body height in cm (Question 256)
hlt4	Weight in kg (Question 257)
hlt6	To what extent handicapped? (Question 259)
hlt7	Hours of sleep per night (Question 260)
cps6	Consent partner interview (Question 263)
cps7	Language partner interview (Question 264)
cps8_	Address partner (Question 265)
cps9	Leave partner questionnaire behind or mail (Question 266)
cps10	Send partner questionnaire per mail (Question 267)
cps11	Interviewer pick up partner questionnaire or mail (Question 268)
cps12	Hand out partner address sheet (Question 269)
int3	Attractiveness Anchor (Question 272)
Int8	Time handout partner questionnaire (Question 277)

Table A.2: Variables not included in DemoDiff wave 1, partner1_DD

Variable	Variable label
pbce1_	Undertake activities with partner (Question 4)
psat3	Satisfaction with relationship (Question 5)
psat4	Estimation of partner's satisfaction with relationship (Question 6)
ppa19_	I hope relationship lasts for a long time (Question 12)
ppa16_	Anchor let partner know that I understand him/her (Question 13)
ppa22p_	Discussion behaviour (Question 15)

continued on next page

continued from previous page

Variable	Variable label
ppa17_	Handling of differences in relationship (Question 16)
ppa18_	Special situations in relationship (Question 17)
psat5	Satisfaction with sex life (Question 28)
plsr1i_	Leisure time (Question 36)
plsr2	Hours watched TV past week (Question 36)
plsr3	Hours spent on personal Internet use past week (Question 36)
plsr4	Weeks on vacation trips past week (Question 36)
pper1_	Often agree with others, even if I am not sure (Question 37)
phlt2	Number of times feeling low and melancholy past 4 weeks (Question 46)
phlt3	Body height in cm (Question 47)
phlt4	Weight in kg (Question 48)
phlt7	Hours of sleep per night (Question 49)
phlt6	To what extent handicapped? (Question 51)

Table A.3: Variables not included in DemoDiff wave 3, anchor3

Variable	Variable label
hc5h1	Main residence: Ownership of dwelling/house (Question 7)
hc5h2	Second Residence: Ownership of dwelling/house (Question 7)
hc11h1	Main residence: Monthly expenditures for rented dwelling (Question 8)
hc11h2	Second Residence: Monthly expenditures for rented dwelling (Question 8)
hc12h1	Main residence: Monthly expenditures for self-owned dwelling (Question 9)
hc12h2	Second Residence: Monthly expenditures for self-owned dwelling (Question 9)
hc13h1	Main residence: Total living space in square meters (Question 10)
hc13h2	Second Residence: Total living space in square meters (Question 10)
hc14h1	Main residence: Number of rooms (Question 11)
hc14h2	Second Residence: Number of rooms (Question 11)
sd33	Attending same school as in previous wave (Question 13)
sd25	Type of school currently attending (Question 14)
sd26	Grade in school (Question 15)
job5_	Owner of company (Question 23)
job6_	Employed in own firm (Question 24)
bce1_	VOP+: Undertake activities with partner (Question 37)
sin4_	Dating possibilities (Question 42)
sin5_	Trying to find out sth. about personality (Question 43)
pa1	Interested in partner (Question 48)
pa10_	Feelings regarding common household (Question 81)
pa12_	Feelings regarding marriage (Question 83)
pa16_	Let partner know that I understand him/her (Question 91)
pa17_	Telling partner what you are thinking (Question 92)
pa18_	Partner finds it all right if I pursue own interests (Question 93)

continued on next page

continued from previous page

Variable	Variable label
sat4	Estimation of partner's satisfaction with relationship (Question 95)
pa19_	I hope relationship lasts for a long time (Question 96)
pa20_	Serious relationship problem (Question 97)
pa22_	Assessment of partner (Question 99)
pa29	Frequency child witnessed arguments past year (Question 100)
pa23	Problems with alcohol, medication, or other drugs last year (Question 101)
pa24	Extra-marital affair of anchor or partner past year (Question 102)
pa25	Arguments using physical force past year (Question 103)
per1_	I often agree with others, even if I'm not sure (Question 107)
per4_	How did you feel in the last 4 weeks (Question 108)
per2_	How do you feel in general (Question 109)
sex8	Frequency of sexual intercourses last 3 months (Question 128)
sex9	To what extend do the following statements apply to you? (Question 129)
crn19_	Parenting goals (Question 159)
crn1_	Place of birth child x (Question 160)
crn2_	Complications child x: health/life of child was at risk (Question 161)
crn3_	Child x: C-section? (Question 162)
crn4_	Child x: In what week of the pregnancy born (Question 163)
crn5_	Child x: Length at the time of birth in cm (Question 164)
crn6_	Child x: Weight at the time of birth in grams (Question 165)
crn7_	Child x: Type of last checkup (Question 166)
crn37_	Child x: Was your child breast-fed and, if yes, how long (months) (Qu. 167)
crn24_	Child x: How much time did your child need to fall asleep (minutes) (Qu. 168)
crn25_	Child x: How many times per night did your child wake up (Question 169)
crn26_	Child x: How strongly burdened by child's sleeping behavior (Question 170)
crn38_	Child x: Does child set bedtime or do you as parents set bedtime (Qu. 171)
crn39_	Child x: How long altogether does your child sleep at night (Question 172)
crn27_	Child x: How long altogether does your child sleep at night (Question 173)
crn28_	Child x: How often burdened by child's crying or fussiness (Question 174)
crn10_	Child x: My child is . . . (Question 175)
crn11_	I often have no energy (Question 176)
crn29_	Worries concerning own children (Question 177)
crn31_	Child x: Health, generally speaking (Question 179)
crn20_	Role as a parent (Question 183)
ccs1k_	Consent to children's survey child x (Question 189)
ccs2k_	Point of time of children's survey child x (Question 190)
cpas1	Consent to own participation in parenting survey (Question 191)
cpas2	Pick up parenting survey or send by mail (Question 192)
cpas3	Consent to partner's participation in parenting survey (Question 193)
cpas4	Pick up partner's parenting survey or send by mail (Question 194)
lsr1i_	Leisure time (Question 195)

continued on next page

continued from previous page

Variable	Variable label
lsr2_	Hours watched TV past week (Question 196)
lsr3_	Hours spent on personal internet use past week (Question 197)
lsr4_	Weeks on vacation trips past year (Question 198)
lsr5_	Leisure with partner (Question 199)
lsr6_	Online profile on social network website (Question 200)
lsr7_	Visibility of online profile (Question 201)
lsr8_	Frequency of visit: Social network sites (Question 202)
igr73i_	Citizenship adoptive mother (Question 205)
igr74_	Country of birth adoptive mother (Question 206)
igr77i_	Citizenship adoptive father (Question 212)
igr78_	Country of birth adoptive father (Question 213)
igr81	Is mother's new partner same partner as at the time of the last interview (Question 222)
igr82	Citizenship mother's new partner (Question 225)
igr84	Is father's new partner same partner as at the time of the last interview (Question 230)
igr85	Citizenship father's new partner (Question 233)
igr41p2	Timely distance to dwelling of mother's partner (Question 244)
igr41p4	Timely distance to dwelling of father's partner (Question 246)
cprs1p1	Consent to survey mother (Question 251)
cprs2p1	Mother: Name and address (Question 252)
cprs3p1	Language version of questionnaire mother (Question 253)
cprs1p2	Consent to survey mother's partner (Question 254)
cprs2p2	Mother's partner: Name and address (Question 255)
cprs3p2	Language version of questionnaire mother's partner (Question 256)
cprs1p3	Consent to survey father (Question 257)
cprs2p3	Father: Name and address (Question 258)
cprs3p3	Language version of questionnaire father (Question 259)
cprs1p4	Consent to survey father's partner (Question 260)
cprs2p4	Father's partner: Name and address (Question 261)
cprs3p4	Language version of questionnaire father's partner (Question 262)
inc25_	Parents: Financial situation (Question 296)
hlt1	Health status past 4 weeks (Question 301)
hlt3	Body height in cm (Question 302)
hlt4	Weight in kg (Question 303)
hlt7	Hours of sleep at night during working week (Question 304)
int12_	Suggestions or comments regarding the interview (Question 307)

Table A.4: Variables not included in DemoDiff wave 3, partner3

Variable	Variable label
psat4	Estimation of partner's satisfaction with relationship (Question 6)
ppa19_	I hope relationship lasts for a long time (Question 12)
ppa16_	Anchor let partner know that I understand him/her (Question 13)
ppa21p_	Disagree with partner (Question 14)

continued on next page

continued from previous page

Variable	Variable label
ppa22p_	Discussion behaviour (Question 15)
ppa17_	Handling of differences in relationship (Question 16)
ppa18_	Special situations in relationship (Question 17)
pcrn32_	Feeling as a parent: Own role (Question 33)
pcrn20_	Feeling as a parent: Role of partner (Question 33)
plsr1i_	Leisure time (Question 34)
plsr2	Hours watched TV past week (Question 34)
plsr3	Hours spent on personal Internet use past week (Question 34)
plsr4	Weeks on vacation trips past year (Question 34)
pigr42p_	Frequency: You tell your mother/father what you are thinking (Question 40)
pigr45p_	Frequency: You and mother/father annoyed/angry with each other (Question 41)
pigr47p_	Frequency: You and your mother/father disagree and quarrel (Question 42)
pigr49p_	Frequency: You share with mother/father secrets/private feelings (Question 43)
phlt1	Health status past 4 weeks (Question 51)
phlt3	Body height in cm (Question 52)
phlt4	Weight in kg (Question 53)
phlt7	Hours of sleep per night during working week (Question 54)
pper1i2	Sometimes I believe that I'm worthless (Question 61)
pper1i6	I feel lonely (Question 61)
pper2_	How do you feel in general (Question 62)

Table A.5: Additional variables in DemoDiff, wave 1, anchor1_DD

Variable	Variable label
geboz	Federal state of birth of anchor
gebop	Federal state of birth of partner

Table A.6: Additional variables in DemoDiff, wave 3, anchor2_DD + anchor3

Variable	Variable label
frtp1	Partnership status at first birth
frtp2	Coresidence with partner at first birth
frtp3	Marital status at first birth
cla1	In which state (Bundesland) you were born in?
cla2	How many times have you moved since your birth?
cla8	How would you describe your childhood in general?
cla3	Immediately after your birth, have you lived with both of your biological parents?
cla4	With whom did you live immediately after you were born?
cla5_	Has something changed in this situation?

A.2 Changes between releases

Table A.7: Changes in data sets *anchor1* and *partner1*: from release 1.0 to release 2.0

Description of changes	Variables affected
Anchor data	
<i>New generated variables</i>	
New variables (Generated identifiers : anchor's, partner's, children's sex / anchor's, partner's, children's, parents', stepparents' date of birth)	sex_gen, psex_gen, k*sex_gen / doby_gen, dobm_gen, pdoby_gen, pdobm_gen, k*doby_gen, k*dobm_gen, mdoby_gen, mdobm_gen, fdoby_gen, fdobm_gen
New variable (Anchor currently living in East Germany)	east
New variable (Anchor's ethnicity)	ethni
New variable (Anchor's migration status)	migstatus
New variable (Child(ren) living at main residence)	childmrd
<i>Modified generated variables</i>	
Age month-specific (based on month & year of birth); year of anchor's interview used to compute all age variables instead of 2008	age, page, mage, fage, k*age
Federal state Saarland now can be distinguished from Rhineland-Palatinate	bula
Missing values combined to -7	cob, fcob, mcob, pcob, nat1, nat2, pnat1, pnat2, mnat1, mnat2, fnat1, fnat2
Value labels changed to "1991-1993", "1981-1983", "1971-1973"	cohort
Set to -7 instead of 0 if inc13, inc14, inc15, inc16, inc17 is -4	hhincnet
Variable content changed (Household composition)	hhcomp
Computed using the generated identifiers sex_gen and psex_gen	homosex
Set to -3 instead of 0 if inc2, inc3, inc5, inc6 is -3; Set to -7 instead of 0 if inc2, inc3, inc5, inc6 is -4;	incnet
-7 recoded to 1 if couple uses sterilization for contraception; -7 recoded to 0 if female partner with male anchor reports that pregnancy status of couple is positive or anchor or partner is pregnant.	infertile
System missing (.) recoded to -7	intdur
Value labels of categories 4 & 5 interchanged	isced, pisced
Changed condition for existence of children : used variable sd14kxg instead of age; coded to -3 if sd14k1g-sd14k10g is -3,-2,-1,6 or 7	k1type ... k10type
If more than one work activity first act. status overrides 2nd act. status	lfs, plfs
Set to -7 if np is -7 (see above)	ncoh
Set to -7 ("Incomplete data") if just one partner mentioned and this partner's sex is -1 or -2	np

continued on next page

continued from previous page

Description of changes	Variables affected
Set to -3 for male anchor without partner or with male partner; set to 0 for female anchor with female partner or female anchor without partner who states not to be pregnant; set to 1 if partner states that couple is pregnant or female partner directly states to be pregnant	pregnant
<i>Discontinued generated variables</i>	
Variable discontinued (Mother living at main residence)	m2rd
Variable discontinued (Father living at main residence)	f2rd
Variable discontinued (Others living at main residence)	oth2rd
Variable discontinued (Household size second residence)	hysize2rd
<i>Modified labels of main variables</i>	
Some new English variable and value labels (translation edited)	various variables throughout the data set
Variable labels modified (German and English version)	pa16i1-6, pa22pi1-8, pa22ri1-8, frt1, frt2, sdp9i2, pa18i11, col1i3
Variable label modified (German version)	pa17i5
Variable labels modified (only German version: "Anker" instead of "ZP")	sex3, mig3, job5i1, job5i4, job6i3, inc19i2, inc19i3, inc19i4, inc19i5, int1, int2, int3
Value labels modified (only German version: "Anker" instad of "ZP")	cps3, cps3i3, cps8i3
Partner data	
<i>Modified labels</i>	
Some new English variable and value labels (translation edited)	various variables throughout the data set
Variable labels modified (German and English version)	ppa16i1-6, ppa22pi1-8, ppa22ri1-8, pfrt1

Table A.8: Changes in data set *anchor1*: from release 2.0 to release 3.0

Description of changes	Variables affected
Anchor data	
<i>Modified generated variables</i>	
Self-report of partner is used for computation of variable as well as the anchor's report, self-report dominates indirect report	infertile
Modified and new value: value 16 also includes "Vocational retraining / further education"; new value 7 "Kolleg, 2. Bildungsweg"	enrol, penrol
Some values modified according to generated variables marstat and relstat	meetdur, reldur, cohabdur, mardur
Slightly modified due to corrections of data set <i>biopart</i>	np, ncoh
Value change (3 to 4 and vice versa) and rename of value label "POS 8./9."	school, pschool
<i>Modified tag and flag variables</i>	
Additional couples of anchor respondents identified	flag26
<i>Modified values of main variables</i>	
Value of variable sat2 is modified: -1 is set to 8; -2 is set to 9	sat2
<i>Modified labels of main variables</i>	
Variable labels modified (German version)	ftr11v1i*, pa16*
Variable labels modified (English version)	job18*
Variable label modified (German and English version)	ftr2
Value labels modified (German version)	sd13, rtr11, rtr15p*

Table A.9: Changes in data set *anchor2*, *parent2*, *child2*, *parenting2*: from release 2.0 to release 3.0

Description of changes	Variables affected
Anchor data	
<i>Modified generated variables</i>	
Use information of wave 1 if no valid information available in wave 2; variable only for new partners	pcob
Nationalities of new partners were stored in downward order, now in ascending order; variables only for new partners	pnat1, pnat2
Partner's self-report is used for computation of variable as well as the anchor's report, self-report dominates indirect report; instead of ftr1 and ftr2 variable infertile of wave 1 is used for computation if no valid information available in wave 2	infertile
Modified and new value: value 16 also includes "Vocational retraining / further education"; new value 7 "Kolleg, 2. Bildungsweg"	enrol
Value change (3 to 4 and vice versa) and rename of value label "POS 8./9."	school, pschool
Modified values: -7, 2, 3	marstat
Some values modified according to generated variables marstat and relstat	cohabdur, mardur
Slightly modified due to corrections of data set <i>biopart</i>	ncoh, nmar
Value labels modified	sex_gen, psex_gen, k*sex_gen, dobm_gen, siops

continued on next page

continued from previous page

Description of changes	Variables affected
<i>Modified tag and flag variables</i>	
Additional couples of anchor respondents identified	flag26
Value label -3 recoded to 0 (theoretically possible)	flag8, flag16
Value labels slightly modified	flag26, flag_frt6
Variable label modified	flag_frt6, flag20, flag25
<i>Modified values of main variables</i>	
96 is set to 6; 97 is set to 7	frt24i1-8
Value labels of categories 1 & 0 interchanged	ehc18p1-4
<i>Modified labels of main variables</i>	
Variable label modified (German version)	ehc19i13m1-18, ehc19i13, d175, cpas3, ehc13k1-7, cla5e1-10, frt11v1i*, crn19i1, sdp9i1, sdp9i6, sdp9i7
Variable labels modified (English version)	inc10i12-inc10i16, sdp9i6, sdp9i7
Value labels modified (German version)	ehc5p1-5
Value labels modified (English version)	cps5, pa3, cprs2p*, inc27*
Value labels modified (German and English version)	sex9i*, sep1i*, per1*, inc25*, inc26*
Parent data	
<i>Modified generated variables</i>	
English variable label added	pargkpol, parbik, parbula
<i>Modified labels of main variables</i>	
Value labels modified (English version)	parigr85a, pargc12, parigr103, parinc23
Child data	
<i>Modified generated variables</i>	
Modified value labels: '5 stepfather' to '5 stepmother', '6 stepmother' to '6 stepfather', '7 foster father' to '7 foster mother', '8 foster mother' to '8 foster father'	relstatac
<i>Modified labels</i>	
The value was added to the value label	all variables of this data set
Parenting data	
<i>Discontinued variables</i>	
Variables deleted	anchor, partner

Table A.10: Changes in data set *anchor1* and *partner1*: from release 3.0 to release 3.1

Description of changes	Variables affected
Anchor data	
<i>Modified generated variables</i>	
Recoding of implausible values for variables school & pschool: “3 lower GDR, POS 8./9.” to “2 lower, Volks-/Hauptschulabschluss” and “5 intermediate GDR, POS 10.” to “4 intermediate, Realschulabschluss / mittlere Reife” since cohort 1 & 2 respondents and partners born after 1980 (POS 8./9.) or 1979 (POS 10.) could not have earned a degree from a polytechnic secondary school (only existed in the GDR); changes in other variables due to changes of school & pschool	school, pschool, casmin, pcasmin, isced, pisced, yeduc, pyeduc
<i>Modified labels of main variables</i>	
Variable labels modified (German and English version): Question number in variable label corrected	ftr11v2i1-int10i2
Partner data	
<i>Modified values of main variables</i>	
Recoding of values to avoid discontinuity compared to later waves	psd10
Recoding of missing values: “-1” was mistakenly included in category “-2”	psat1*, psat3

Table A.11: Changes in data set *anchor2*, *child2*, *parenting2*: from release 3.0 to release 3.1

Description of changes	Variables affected
Anchor data	
<i>Modified generated variables</i>	
Variables school & pschool changed due to corrections in anchor1 and some recodings (some values 3 recoded to 4) plus recoding of implausible values for new partners: “3 lower GDR, POS 8./9.” to “2 lower, Volks-/Hauptschulabschluss” and “5 intermediate GDR, POS 10.” to “4 intermediate, Realschulabschluss / mittlere Reife” since partners born after 1980 (POS 8./9.) or 1979 (POS 10.) could not have earned a degree from a polytechnic secondary school (only existed in the GDR); changes in other variables due to changes of school & pschool	school, pschool, casmin, pcasmin, isced, pisced, isced2, pisced2, yeduc, pyeduc
Values of some cases corrected (see generated identifiers)	age, page, mage, fage, smage, sfage, k*age, homosex, hhcomp, lweight, ppanel, pcontact, panswer
Variable labels modified (English version)	*nat1, *nat2, *cob, infertile, *enrol, *school, *vocat, *casmin, *iscd, *iscd2, *yeduc, siops, *sex_gen, *doby_gen, *dobm_gen

continued on next page

continued from previous page

Description of changes	Variables affected
Value labels modified (English version)	k*dobm_gen, k*doby_gen, k*sex_gen
<i>Modified tag and flag variables</i>	
Value labels modified (English version)	flag1, flag5, flag16, flag17, flag_cas, pflag_cas
<i>Modified labels of main variables</i>	
Variable labels modified (German and English version)	ftr11v1*
Variable labels modified (English version)	crn19i1
Value labels modified (German and English version)	sin3*, sin4*, sin5*, sin6*, pa1*, ftr10*, crn10*, crn11*, crn20*, cpas5
Value labels modified (German version)	cps8i3
Value labels modified (English version)	pa3, cprs2, netp*n, inc25, inc26, inc27, capikid, he3, hm2, hsv2, hv2, hsm2
Child data	
<i>Modified labels of main variables</i>	
Variable labels modified (German version)	cedu1, cedu5i4, cpcr8i8, clsr1i1, clsr1i14
Variable labels modified (English version)	cedu8i2, cpcr8i1, cgp1, l391s
Correction of value labels for missing values (German and English version; only in Stata)	all variables of this data set
Value labels modified (German version)	csex, cdobm, cedu1, l391s
Value labels modified (English version)	cpcr5, cdobm, l391s
Parenting data	
<i>Changes in case number</i>	
Two empty cases deleted from data set	(id=494001000, id=609745000)
<i>Modified values of main variables</i>	
Replacing 6 missing person numbers partner	pid
<i>Modified labels of main variables</i>	
Correction of value labels for missing values (German and English version; only in Stata)	all variables of this data set
Value labels modified (German version)	pcr3i1-pcr3i8

Table A.12: Changes in data set *anchor3*, *child3*, *parent3*, *parenting3*: from release 3.0 to release 3.1

Description of changes	Variables affected
Anchor data	
<i>Modified tag and flag variables</i>	
Value labels modified (English version)	flag_cas, pflag_cas
<i>Modified labels of main variables</i>	
Variable labels modified (English version)	job18h, job18m

continued on next page

continued from previous page

Description of changes	Variables affected
Value labels modified (German and English version)	pa11, pa14*, pa17*, pa18*, pa19*
Value labels modified (English version)	rtr26*m, rtr30*m, rtr33*m, rtr37*m, rtr41*m, rtr45m, he3, hm2, hsv2, hv2, hsm2
Child data	
<i>Modified values of main variables</i>	
Recoding: 1 to -3 if no inconsistency possible since child new in wave 3	tag_csex, tag_cdobm, tag_cdoby
<i>Modified labels of main variables</i>	
Variable labels slightly modified (German version)	cedu7i1, csdq1i5, csdq1i10
Variable labels modified (English version)	cedu1ao, cedu13i4, cpcr4, cpcr5, cpcr13, cint1i1, cint1i2, cint1i3, cint1i4, cint1i5, cint2, l391s
Parent data	
<i>Modified values of main variables</i>	
Change of response code: 6 is set to 8	parigr53a-parigr59a, parigr78a-parigr79a, parigr80a-parigr81a, parigr60a-parigr66a
New category “no partner” added	parpa17i1-parpa17i8
<i>Modified labels of main variables</i>	
Assignment of variable labels to variables adjusted (German and English version)	parval1ix
Variable labels slightly modified (German and English version)	parigr45a, parigr47a, parsd23i13, parsd23i21
Change in wording of value label from “Trifft nicht zu” to “Kein Bedarf” (German version)	parigr53a-parigr59a, parigr78a-parigr79a, parigr80a-parigr81a, parigr60a-parigr66a
Value labels slightly modified (German and English version)	parigr103
<i>Modified filter of main variables</i>	
Variables asked irrespective of existing contact to parent (German and English version)	parigr40p1, parigr40p3, parigr41p1, parigr41p3
Parenting data	
<i>Modified labels of main variables</i>	
Value labels modified (German and English version)	pcr3i1-pcr3i8

Please note: For release 3.1, two format changes have been implemented in Stata:

- All variables indicating person numbers (such as “id”, “pid” etc.) have been reformatted to avoid abbreviations. This does not include any recodings of the data.
- As of wave 2, some value label lists have been renamed to avoid problems when merging data sets of different waves.

Table A.13: Changes in data set *anchor1*: from release 3.1 to release 4.0

Description of changes	Variables affected
Anchor data	
<i>Modified values of main variables</i>	
Correction of information on children for id=165520000, 308441000 & 689470000	sd14k*-sd21k*, rtr16k*-rtr21k*, sd22k*
<i>New generated variables</i>	
New variables (Parents' educational status)	mschool, fschool, mcasmin, fcasmin, misced, fisced, myeduc, fyeduc
New variables (Parents' vocational status)	mvocat, fvocat
New variables (Youngest child living with anchor)	ykage, ykid
New variable (ISCED classification of educational attainment, incl. students)	isced2
New variable (Flag variable KldB and ISCO)	flag_isco08_kldb2010
New variable (Flag variable EHC)	flag_ehc
New variable (Flag variable intergenerational relations)	flag_igb
<i>Modified generated variables</i>	
New classification schema KldB 2010	kldb2010
New classification schema ISCO-08	isco08, isei, siops
New coding in cases of ambiguous seasonal information on the month and valid answers for the year, random values for month imputed	age, page, mage, fage, k*age
Some modifications for three cases due to correction of information on children (see above)	nkids*, k*type, hhsizemrd
<i>Discontinued generated variables</i>	
Variable discontinued (sex ratio)	sexratio
Variable discontinued (population density)	popdens
<i>Modified tag and flag variables</i>	
Value "1 Inconsistency" instead of values indicating episodes to gain equivalence across waves	flag7
Variables added to achieve consistency across waves (-3 for all cases)	flag20-flag25, flag_ehc, flag_igb
Variable label modified (German and English version)	flag_isco88_kldb1992
Variable and value labels modified (German and English version)	*flag_cas, flag_frt6
Value labels modified (German and English version)	flag1-flag18
<i>Modified labels of generated variables</i>	
Variable label of variable isco based on former classification schema ISCO-88 modified (German and English version)	isco88
Variable label of variable kldb based on former classification schema modified (German and English version)	kldb1992
Value labels modified (German and English version)	kldb2010, isco08
Variable and value labels modified (German and English version)	gkpol, bik, bula

Table A.14: Changes in data set *anchor2*, *child2*, *parenting2*, *parent2*: from release 3.1 to release 4.0

Description of changes	Variables affected
Anchor data	
<i>Modified values of main variables</i>	
Correction of information on children for id=689470000 (also see changes in wave 1)	ehc7k*-ehc13k*, hc15k*, sep4k*-sep10k*, crn1k*-crn10k*, crn12k*-crn18k*, ccs1k*, ccs2k*
<i>New generated variables</i>	
New variables (Parents' educational status)	mschool, fschool, mcasmin, fcasmin, misced, fised, myeduc, fyeduc
New variables (Parents' vocational status)	mvocat, fvocat
New variable (Flag variable KldB and ISCO)	flag_isco08_kldb2010
New variables (Youngest child/capi-child living with anchor)	ykage, ykid, ykagecapi
<i>Modified generated variables</i>	
Recoding of values greater than 5 to 5	lweight
New classification schema KldB 2010	kldb2010
New classification schema ISCO-08	isco08, isei, siops
New coding in cases of ambiguous seasonal information on the month and valid answers for the year, random values for month imputed	age, page, mage, fage, smage, sfage, k*age
Some changes from "1 Never married" to "3 Divorced/dissolved civil union" due to marriage at some point in the EHC	marstat
Some corrections due to marriage at some point in the EHC	relstat
Some changes of cases where the information from the partner questionnaire was not available	pnkidsbioalv
Some modifications due to correction of information on children in wave 1 (see above)	k*sex_gen, k*doby_gen, k*dobm_gen, k*age, nkids, k*type
<i>Discontinued generated variables</i>	
Variable discontinued (sex ratio)	sexratio
Variable discontinued (population density)	popdens
<i>Modified tag and flag variables</i>	
Value 0 recoded to -3	flag16
Variable labels modified (German and English version)	flag8, flag_frt6
Value labels modified (German and English version)	flag6-flag15, flag18
<i>Modified labels of generated variables</i>	
Variable label of variable isco based on former classification schema ISCO-88 modified (German and English version)	isco88
Variable label of variable kldb based on former classification schema modified (German version)	kldb1992

continued on next page

continued from previous page

Description of changes	Variables affected
Value labels modified (German and English version)	kldb2010, isco08
Value labels modified (German and English version)	bik, bula
Value labels modified (English version)	intsex
Parent data	
<i>New main variables</i>	
New unique identifier of anchor's siblings 1 to 3	sibid1, sibid2, sibid3
<i>Modified main variables</i>	
Missing values set to system missing (.) instead of -3	cid
Some values modified due to new order of child-information (child 1 to 4) reported by parents	parsd14k1g-parsd14k4g, parsd15k1-parsd15k4, parsd19k1y-parsd19k4y, parsd33k2-parsd33k4, parsd34k2-parsd34k4, parsd35k2-parsd35k4, parsd36k2-parsd36k4, parigr39k2-parigr39k4, parigr40k2-parigr40k4, parigr41k2-parigr41k4, parflag1-parflag5, parkids
<i>Modified generated variables</i>	
Variables based on information from waves 2 to 3 of parent data and wave 3 of anchor data instead of wave-specific information from parent data; only for biological and adoptive parents	parschool, parvocat, parcasmin, parisced, paryeduc
<i>Modified labels of generated variables</i>	
Variable labels slightly modified (German and English version)	parbula, parbik, pargkpo, parage, parage2, parpage, parpage2, parmage, par- mage2, parfage, parfage2, parcasprim, parcassec, parflag_cas, parhhincnet, parhhsizemrd, parhhincgee, park1type, park2type, park3type, park4type, parlfs, parmarstat, parrel- stat, parigr27, parigr28

continued on next page

continued from previous page

Description of changes	Variables affected
Child data	
<i>Modified labels of main variables</i>	
Variable labels slightly modified (German and English version)	id, pid, cdobd, cedu1o, cedi1ao, cpcr4, cpcr5, cpcr13, cgp1, cgp2
Value labels slightly modified (German and English version)	l391s, sex, cdobd, cedu1o, cedi1ao, cpcr4, cpcr5, cpcr13, cgp1, cgp2
Parenting data	
<i>Changes in case number</i>	
Two empty cases deleted from data set	(id=494001000, id=609745000)
<i>Modified labels of main variables</i>	
Variable label slightly modified (English Version)	cid

Table A.15: Changes in data set *anchor3*, *child3*, *parenting3*, *parent3*: from release 3.1 to release 4.0

Description of changes	Variables affected
Anchor data	
<i>Modified main variables</i>	
One correction of pid (id=241451000)	pid
<i>New generated variables</i>	
New variables (Parents' educational status)	mschool, fschool, mcasmin, fcasmin, misced, fisced, myeduc, fyeduc
New variables (Parents' vocational status)	mvocat, fvocat
New variables (Youngest child living with anchor)	ykage, ykid, ykagecapi, ykidcapi
New variable (Flag variable KldB and ISCO)	flag_isco08_kldb2010
New variable (Tag inconsistent identity current partner between waves)	tag_identp
<i>Modified generated variables</i>	
Variable pschool changed due to corrections in anchor1 & anchor2 plus recoding of implausible values for new partners: "3 lower GDR, POS 8./9." to "2 lower, Volks-/Hauptschulabschluss" and "5 intermediate GDR, POS 10." to "4 intermediate, Realschulabschluss / mittlere Reife" since partners born after 1980 (POS 8./9.) or 1979 (POS 10.) could not have earned a degree from a polytechnic secondary school (only existed in the GDR) plus correction for partners without school degree who are not currently in school; changes in other variables due to changes of pschool	pschool, pcasmin, pisced, pisced2, pyeduc

continued on next page

continued from previous page

Description of changes	Variables affected
Correction of coding infertile	infertile
Recoding of values greater than 5 to 5	lweight
New classification schema KldB 2010	kldb2010
New classification schema ISCO-08	isco08, isei, siops
New coding in cases of ambiguous seasonal information on the month and valid answers for the year, random values for month imputed	age, page, mage, fage, smage, sfage, k*age
Some changes from "1 Never married" to "3 Divorced/dissolved civil union" due to marriage at some point in the EHC	marstat
Some corrections due to marriage at some point in the EHC	relstat
<i>Discontinued generated variables</i>	
Variable discontinued (sex ratio)	sexratio
Variable discontinued (population density)	popdens
<i>Modified tag and flag variables</i>	
Variables added to achieve consistency across waves (-3 for all cases)	flag_ehc, flag_frt6
<i>Modified labels of generated variables</i>	
Variable label of variable isco based on former classification schema ISCO-88 modified (German and English version)	isco88
Variable label of variable kldb based on former classification schema modified (German version)	kldb1992
Value labels modified (German and English version)	kldb2010, isco08
Value labels modified (English version)	intsex
Parent data	
<i>Modified values of main variables</i>	
Variables parsd29i1-parasd29i8 set to 0 and 1 instead of -2 due to corrections by <i>TNS Infratest</i>	parsd29i1-parasd29i8
Set to -3 instead of -4	parsd37, parsd38, parsd17i1, parsd17i2, parigr9y
<i>Modified generated variables</i>	
Variables based on information from waves 2 to 3 of parent data and wave 3 of anchor data instead of wave-specific information from parent data; only for biological and adoptive parents	parschool, parvocat, parcasmin, parisced, paryeduc
Child data	
<i>Modified labels of main variables</i>	
Variable labels slightly modified (German and English version)	id, pid, cdobd, cedu1o, cedi1ao, cpcr4, cpcr5, cpcr13, cgp1, cgp2
Value labels slightly modified (German and English version)	l391s, sex, cdobd, cedu1o, cedi1ao, cpcr4, cpcr5, cpcr13, cgp1, cgp2
Parenting data	
<i>Modified labels of main variables</i>	
Variable label slightly modified (English version)	cid

Table A.16: Changes in data set *anchor1*: from release 4.0 to release 5.0

Description of changes	Variables affected
Anchor data	
<i>Modified variable names of main variables</i>	
New variable names original_sex original_dobm original_doby	sex, dobm, doby
<i>Modified main variables</i>	
Household grid correction, cases set to -5, inconsistent value	hc9h1p2-hc9h1p5, hc9h2p4
<i>New generated variables</i>	
New weighting variables	d1weight, ca1weight, d1ca1weight
<i>Modified generated variables</i>	
New baseline year 2010	cpi
Correction of coding number of partner's biological children	pnkidsbioalv
Minor corrections for different variables	pregnant, mschool, mvocat, fvocat, mcasmin, misced, myeduc, fyeduc
<i>Modified tag and flag variables</i>	
Additional couples of anchor respondents identified	flag26

Table A.17: Changes in data set *anchor2*, *parent2*, *parenting2*: from release 4.0 to release 5.0

Description of changes	Variables affected
Anchor data	
<i>Modified variable names of main variables</i>	
New variable names original_sex original_dobm original_doby	sex, dobm, doby
<i>Modified main variables</i>	
Minor correction, inconsistency set on -5	hc9h1p2
<i>New generated variables</i>	
New weighting variables	d1weight, ca1weight, d1ca1weight
<i>Modified generated variables</i>	
New baseline year 2010	cpi
Correction of coding number of partner's biological children	pnkidsbioalv
Minor corrections for different variables	pregnant, mschool, mvocat, fvocat, mcasmin, misced, myeduc, fyeduc
<i>Modified tag and flag variables</i>	
Additional couples of anchor respondents identified	flag26
<i>Modified value labels</i>	
Changed english label <i>liste180_en</i> (0 "Child is alive" to 0 "Child is dead" and 1 "Child is dead" to 1 "Child is alive")	d327, d328, d329, d330, d331, d332, d333, d334, d335, d336

continued on next page

continued from previous page

Description of changes	Variables affected
Parent data	
<i>Modified generated variables</i>	
Correction of sibid for 497 siblings	sibid1, sibid2, sibid3
New variables for identified siblings	sibid4, sibid5, sibid6, sibid7, sibid8
Parenting data	
<i>Modified values of main variables</i>	
Set to 0, 1 and 2 instead of 1, 2 and 3 in all parenting data sets due to data harmonization with the child data set	sdqpi1,...,sdqpi25

Table A.18: Changes in data set *anchor3*, *parent3*, *partner3*, *parenting3*: from release 4.0 to release 5.0

Description of changes	Variables affected
Anchor data	
<i>Modified variable names of main variables</i>	
New variable names original_sex original_dobm original_doby	sex, dobm, doby
<i>Modified generated variables</i>	
Modified value	ehc23p2
New baseline year 2010	cpi
Correction of coding number of partner's biological children	pnkidsbioalv
Correction of coding type of children	k*type
Minor corrections for different variables	pregnant, smcob, sfcob, mschool, fschool, mvocat, fvocat, mcasmin, misced, myeduc, fyeduc
<i>Modified tag and flag variables</i>	
Additional couples of anchor respondents identified	flag26
<i>Modified value labels</i>	
Changed english label <i>liste180_en</i> (0 "Child is alive" to 0 "Child is dead" and 1 "Child is dead" to 1 "Child is alive")	d327, d328, d329, d330, d331, d332, d333, d334, d335, d336
Parent data	
<i>Modified values of main variables</i>	
Categories "no answer" and "does not apply" are now (release 5.0) differentiated	parigr41p1, parigr41p3
Partner data	
<i>Modified main variables</i>	
Minor corrections in assigning of missing values	psd100i*, psd101i*
<i>Modified labels of main variables</i>	
English value label added	pdobd
Parenting data	

continued on next page

continued from previous page

Description of changes	Variables affected
<i>Modified values of main variables</i>	
Set to 0, 1 and 2 instead of 1, 2 and 3 in all parenting data sets due to data harmonization with the child data set	sdqpi1,...,sdqpi25

Table A.19: Changes in data set *anchor4*, *partner4*, *parenting4*: from release 4.0 to release 5.0

Description of changes	Variables affected
Anchor data	
<i>Modified variable names of main variables</i>	
New variable names original_sex original_dobm original_doby	sex, dobm, doby
<i>Modified main variables</i>	
1 repositioned child	cid2, cid3, and all corresponding child variables *k2* and *k3*
<i>New generated variables</i>	
New weighting variables	d1weight, calweight, d1calweight
<i>Modified generated variables</i>	
New baseline year 2010	cpi
Correction of coding number of partner's biological children	pnkidsbioalv
Correction of coding type of children	k*type
Correction of coding variables	smcob, sfcob
Recoding of missing values	isei, siops
Recoding of implausible values for new partners; changes in other variables due to changes of pschool	pschool, pcasmin, pised, pised2, pyeduc
Minor corrections for different variables	mschool, fschool, mvocat, fvocat, mcasmin, misced, myeduc, fyeduc
<i>Modified tag and flag variables</i>	
Additional couples of anchor respondents identified	flag17, flag26
<i>Modified value labels</i>	
Changed english label <i>liste180_en</i> (0 "Child is alive" to 0 "Child is dead" and 1 "Child is dead" to 1 "Child is alive")	d327, d328, d329, d330, d331, d332, d333, d334, d335, d336
Partner data	
<i>Modified labels of main variables</i>	
English value label added	pdobd
Parenting data	
<i>Modified values of main variables</i>	
Set to 0, 1 and 2 instead of 1, 2 and 3 in all parenting data sets due to data harmonization with the child data set	sdqpi1,...,sdqpi25

A.3 Information on longitudinal weights

Table A.20: List of regressors for logit estimations of longitudinal weight

Independent variables and reference category	Definition		
	Cohort 1	Cohort 2	Cohort 3
<i>int5</i>	<i>Interference of other persons in the interview</i>	<i>Interference of other persons in the interview</i>	<i>Interference of other persons in the interview</i>
inter1	missings	missings	missings
inter2 (ref.)	no interference	no interference	no interference
inter3	inferences frequently and sometimes	inferences frequently and sometimes	inferences frequently and sometimes
<i>lfs</i>	<i>Labor force status</i>	<i>Labor force status</i>	<i>Labor force status</i>
work	employed: vocational training, full-time employment, part-time employment, marginal employment, self-employed	employed: vocational training, full-time employment, part-time employment, marginal employment, self-employed	employed: vocational training, full-time employment, part-time employment, marginal employment, self-employed
work (ref.)	unemployed: education, parental leave, housewife/man, unemployed , military service, pensions, others, missings	unemployed: education, parental leave, housewife/man, unemployed , military service, pensions, others, missings	unemployed: education, parental leave, housewife/man, unemployed , military service, pensions, others, missings
<i>marstat</i>	<i>Marital status</i>	<i>Marital status</i>	<i>Marital status</i>
ehe1		never married	never married
ehe2 (ref.)		married	married

continued on next page

Table A.20, continued from previous page

Independent variables and reference category	Definition		
	Cohort 1	Cohort 2	Cohort 3
ehe3		widowed, divorced or incomplete data	widowed, divorced or incomplete data
<i>incnet</i>	<i>Personal net income</i>	<i>Personal net income</i>	<i>Personal net income</i>
pinc1	incomplete data	incomplete data	incomplete data
pinc2 (ref.)	no income	no income	no income
pinc3	1 Euro or more income	1 - 986 Euro income	1 - 986 Euro income
pinc4		987 - 1500 Euro income	987 - 1500 Euro income
pinc5		1501 Euro or more income	1501 - 2500 Euro income
pinc6			2501 Euro or more income
<i>hhsizemrd</i>	<i>Household size main residence</i>	<i>Household size main residence</i>	<i>Household size main residence</i>
hhp1	household with 1 or 2 persons	household with 1 or 2 persons	household with 1 or 2 persons
hhp2	household with 3 persons	household with 3 persons	household with 3 persons
hhp3	household with 4 persons	household with 4 persons	household with 4 persons
hhp4 (ref.)	household with 5 or more persons	household with 5 or more persons	household with 5 or more persons
<i>hc5h1</i>	<i>Main Residence: Ownership</i>	<i>Main Residence: Ownership</i>	<i>Main Residence: Ownership</i>
house	Ownership	Ownership	Ownership
house (ref.)	for rent, sublease, others, no answer	for rent, sublease, others, no answer	for rent, sublease, others, no answer
<i>school</i>	<i>Highest School Degree</i>	<i>Highest School Degree</i>	<i>Highest School Degree</i>
edu1	(ref.) currently enrolled	currently enrolled, without degree, other school degree, missing	currently enrolled, without degree or other school degree

continued on next page

Table A.20, continued from previous page

Independent variables and reference category	Definition		
	Cohort 1	Cohort 2	Cohort 3
edu2	without degree or other school degree, Missings	lower, Volks-/Hauptschulabschluss	lower, Volks-/Hauptschulabschluss
edu3	lower, Volks-/Hauptschulabschluss	intermediate and comparable	intermediate and comparable
edu4	intermediate or higher school degree	upper, Fachhochschulreife	upper, Fachhochschulreife
edu5		(ref.) upper, Allgemeine Hochschulreife	(ref.) upper, Allgemeine Hochschulreife
<i>telgesamt</i>	<i>Frequency of telephone contacts</i>	<i>Frequency of telephone contacts</i>	<i>Frequency of telephone contacts</i>
telcon1	No telephone contacts	No telephone contacts	No telephone contacts
telcon2	1 telephone contact	1 telephone contact	1 telephone contact
telcon3	2 telephone contacts	2 telephone contacts	2 telephone contacts
telcon4	3 telephone contacts	3 telephone contacts	3 telephone contacts
telcon5	(ref.) 4 or more telephone contacts	4 telephone contacts	4 telephone contacts
telcon6		(ref.) 5 or more telephone contacts	(ref.) 5 or more telephone contacts
<i>persgesamt</i>	<i>Frequency of personal contacts</i>	<i>Frequency of personal contacts</i>	<i>Frequency of personal contacts</i>
perscon1	No personal contacts	No personal contacts	No personal contacts
perscon2	1 personal contact	1 personal contact	1 personal contact
perscon3	2 personal contacts	2 personal contacts	2 personal contacts
perscon4	3 personal contacts	3 personal contacts	3 personal contacts
perscon5	(ref.) 4 or more personal contacts	4 personal contacts	4 personal contacts
perscon6		(ref.) 5 or more personal contacts	(ref.) 5 or more personal contacts
<i>bik</i>	<i>Settlement Structure</i>	<i>Settlement Structure</i>	<i>Settlement Structure</i>
biki1 (ref.)	City Center - population 500.000 or more	City Center - population 500.000 or more	City Center - population 500.000 or more

continued on next page

Table A.20, continued from previous page

Independent variables and reference category	Definition		
	Cohort 1	Cohort 2	Cohort 3
biki2	Periphery - population 500.000 or more	Periphery - population 500.000 or more	Periphery - population 500.000 or more
biki3	City Center - population 100.000 - 500.000	City Center - population 100.000 - 500.000	City Center - population 100.000 - 500.000
biki4	Periphery - population 100.000 - 500.000	Periphery - population 100.000 - 500.000	Periphery - population 100.000 - 500.000
biki5	City Center - population 50.000 - 100.000	City Center - population 50.000 - 100.000	City Center - population 50.000 - 100.000
	Periphery - population 50.000 - 100.000	Periphery - population 50.000 - 100.000	Periphery - population 50.000 - 100.000
biki6	Region - population 20.000 - 50.000	Region - population 20.000 - 50.000	Region - population 20.000 - 50.000
biki7	Region - population 5.000 - 20.000	Region - population 5.000 - 20.000	Region - population 5.000 - 20.000
biki8	Region - population 2.000 - 5.000	Region - population 2.000 - 5.000	Region - population 2.000 - 5.000
	Region - population less than 2.000	Region - population less than 2.000	Region - population less than 2.000
<i>bula</i>	<i>Federal State</i>	<i>Federal State</i>	<i>Federal State</i>
bundl1 (ref.)	Stadtstaaten: Berlin, Hamburg, Bremen, Berlin (East)	Stadtstaaten: Berlin, Hamburg, Bremen, Berlin (East)	Stadtstaaten: Berlin, Hamburg, Bremen, Berlin (East)
bundl2	Schleswig-Holstein	Schleswig-Holstein	Schleswig-Holstein
bundl3	Niedersachsen (Lower Saxony)	Niedersachsen (Lower Saxony)	Niedersachsen (Lower Saxony)
bundl4	Nordrhein-Westfalen (North Rhine-Westphalia)	Nordrhein-Westfalen (North Rhine-Westphalia)	Nordrhein-Westfalen (North Rhine-Westphalia)
bundl5	Hessen (Hesse)	Hessen (Hesse)	Hessen (Hesse)
bundl6	Rheinland-Pfalz (Rhineland-Palatinate)	Rheinland-Pfalz (Rhineland-Palatinate)	Rheinland-Pfalz (Rhineland-Palatinate)
bundl7	Baden-Württemberg	Baden-Württemberg	Baden-Württemberg
bundl8	Bayern (Bavaria)	Bayern (Bavaria)	Bayern (Bavaria)
bundl9	Brandenburg	Brandenburg	Brandenburg

continued on next page

Table A.20, continued from previous page

Independent variables and reference category	Definition		
	Cohort 1	Cohort 2	Cohort 3
bundl10	Mecklenburg-Vorpommern	Mecklenburg-Vorpommern	Mecklenburg-Vorpommern
bundl11	Sachsen (Saxony)	Sachsen (Saxony)	Sachsen (Saxony)
bundl12	Sachsen-Anhalt (Saxony-Anhalt)	Sachsen-Anhalt (Saxony-Anhalt)	Sachsen-Anhalt (Saxony-Anhalt)
bundl13	Thüringen (Thuringia)	Thüringen (Thuringia)	Thüringen (Thuringia)
<i>sat6</i>	<i>Life satisfaction</i>	<i>Life satisfaction</i>	<i>Life satisfaction</i>
lsat1	Life satisfaction: 0 - 6 and no answer	Life satisfaction: 0 - 5 and no answer	Life satisfaction: 0 - 6 and no answer
lsat2	Life satisfaction: 7	Life satisfaction: 6	Life satisfaction: 7
lsat3	Life satisfaction: 8	Life satisfaction: 7	Life satisfaction: 8
lsat4	Life satisfaction: 9	Life satisfaction: 8	Life satisfaction: 9
lsat5	(ref.) Life satisfaction: 10	Life satisfaction: 9	(ref.) Life satisfaction: 10
lsat6		(ref.) Life satisfaction: 10	
<i>h11t</i>	<i>Health status past 4 weeks</i>	<i>Health status past 4 weeks</i>	<i>Health status past 4 weeks</i>
well1	Health status "bad", "not so good" and no answer	Health status "bad", "not so good" and no answer	Health status "bad", "not so good" and no answer
well2	Health status "satisfactory"	Health status "satisfactory"	Health status "satisfactory"
well3	Health status "good"	Health status "good"	Health status "good"
well4 (ref.)	Health status "very good"	Health status "very good"	Health status "very good"
<i>age</i>	<i>age</i>	<i>age</i>	<i>age</i>
age1	17 and 18 years		
age2	19 years		
age3 (ref.)	20 and 21 years		

continued on next page

Table A.20, continued from previous page

Independent variables and reference category	Definition		
	Cohort 1	Cohort 2	Cohort 3
age4		27 and 28 years	
age5		29 years	
age6 (ref.)		30 and 31 years	
age8			37 and 38 years
age9			39 years
age10 (ref.)			40 and 41 years
<i>nat1</i>	<i>Nationality</i>	<i>Nationality</i>	<i>Nationality</i>
german	German Nationality	German Nationality	German Nationality
<i>sex</i>	<i>Gender</i>	<i>Gender</i>	<i>Gender</i>
sex1	Male	Male	Male
<i>DemoDiff sample</i>		<i>DemoDiff sample</i>	<i>DemoDiff sample</i>
demodiff		demodiff	demodiff
<i>temp_dropout</i>	<i>temporary drop-out</i>	<i>temporary drop-out</i>	<i>temporary drop-out</i>

A.4 Flag and tag variables

Table A.21: List of flag variables to identify inconsistencies (*anchor\$*)

Variable	Label	Values	Value Labels	Description
flag1	Inconsistency	0	No inconsistency	
	biological child and sex of the other parent	1	Inconsistency	Biological child existing and sex of second biological parent = anchor's sex
flag2	Inconsistency partner and household grid	0	No inconsistency	
		11 / 12	Partner in hh1/2, no current relationship	Partner living in household 1/2 (see hh grid), but according to question on relationship status currently no partnership
		21 / 22	Partner in hh1/2, no current cohabitation	Partner living in household 1/2 (see hh grid), but according to question on cohabitation status currently no cohabitation
		31 / 32	Several partners in hh1/2	More than one partner in household 1/2 (see hh grid)
		41 / 42	Name partner in hh1/2 unequal to sd4n	Name partner in household 1/2 (see hh grid) unequal to name current partner
		51 / 52	No answer name partner in hh1/2	Partner's name in household grid not mentioned
		-3	Does not apply	Inconsistency irrelevant in this wave
flag3	Inconsistency biological child and household grid	0	No inconsistency	
		11 / 12	Name child in hh1/2 unequal to sd14kxn	Name child according to household grid unequal to name child reported before
		21 / 22	Additional child(ren) in hh1/2	More children in household according to household grid than reported before

continued on next page

Table A.21, continued from previous page

Variable	Label	Values	Value Labels	Description
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag4	Inconsistency partner's child and household grid	0	No inconsistency	
		11 / 12	Name child in hh1/2 unequal to sd14kxn	Name child according to household grid unequal to name child reported before
		21 / 22	Additional child(ren) in hh1/2	More children in household according to household grid than reported before
		31 / 32	Biological child is stepchild in hh1/2	Status of children different in household grid than reported before
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag5	Inconsistency marriage before beginning of relationship (current partner)	0	No inconsistency	
		1	Inconsistency	Marriage before (first) beginning of relationship
flag6	Inconsistency beginning current and end previous cohabitation (current partner)	0	No inconsistency	
		"z"	Incons. episode z and current cohab.	End previous cohabitation episode with current partner after beginning of current cohabitation episode with this partner
		"z ₁ z ₂ "	Incons. episodes z ₁ & z ₂ and current cohab.	End of two previous cohabitation episodes with current partner after beginning of current cohabitation episode with this partner
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag7	Inconsistency beginning current and end previous cohabitation (different partners)	0	No inconsistency	
		1	Inconsistency	End previous cohabitation episode with previous partner after beginning of current cohabitation episode with current partner

continued on next page

Table A.21, continued from previous page

Variable	Label	Values	Value Labels	Description
flag8	Inconsistency beginning current and end previous marriage (different partners)	0 "x"	No inconsistency Inconsistency partner x	End last marriage before beginning current marriage
flag9	Inconsistency divorced/widowed and no partner before current relationship	0 1 -3	No inconsistency Inconsistency Does not apply	Current marital status divorced/separated civil union or widowed and "no partner before current relationship/never had partner" Inconsistency not possible in this wave; flag irrelevant
flag10	Inconsistency divorce from a partner to whom never married	0 "x" -3	No inconsistency Inconsistency partner x Does not apply	End of previous relationship through divorce, but never married to this partner Inconsistency not possible in this wave; flag irrelevant
flag11	Inconsistency separation through death/divorce current spouse	0 "x" -3	No inconsistency Inconsistency partner x Does not apply	Divorced from current spouse or current spouse dead Inconsistency not possible in this wave; flag irrelevant
flag12	Inconsistency year of birth current partner	0 1 2 3	No inconsistency Younger than 10 years old Year of birth after beginning of relationship Younger than 12 years old at birth of biological child	<i>see value labels</i>

continued on next page

Table A.21, continued from previous page

Variable	Label	Values	Value Labels	Description
		4	Younger than 14 years old and completed vocat. training/univ. degree	
flag13	Inconsistency year of birth parents	0	No inconsistency	
		1	Inconsistency year of birth mother	Age difference mother and anchor less than 12 years
		2	Inconsistency year of birth father	Age difference father and anchor less than 12 years
		3	Inconsistency year of birth mother and father	Age differences mother and anchor, and father and anchor less than 12 years
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag14	Inconsistency separation before beginning of relationship (current partner)	0	No inconsistency	
		1	Inconsistency	Separation episode(s) with current partner before first beginning of relationship with current partner
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag15	Inconsistency year of birth of child and anchor	0 "x"	No inconsistency Inconsistency child x	Age difference biological child and anchor less than 12 years
flag16	Inconsistency gross and net personal income	0	No inconsistency	
		1	Inconsistency	Net income larger than gross income
		-3	Does not apply	
flag17	Inconsistency personal net and household income	0 1	No inconsistency Inconsistency	Net personal income larger than net household income

continued on next page

Table A.21, continued from previous page

Variable	Label	Values	Value Labels	Description
flag18	Inconsistency sex of partner	0	No inconsistency	
		1	Inconsistency, male partner	Partner female according to anchor, male according to partner (see variable <i>psex</i> partner data)
		2	Inconsistency, female partner	Partner male according to anchor, female according to partner (see variable <i>psex</i> partner data)
flag20	Inconsistency beginning of relationship before getting to know (current partner)	0	No inconsistency	
		1	Inconsistency	First month of relationship earlier than date of getting to know each other
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag21	Inconsistency cohabitation before getting to know each other (current partner)	0	No inconsistency	
		1	Inconsistency	First month of cohabitation earlier than date of getting to know each other
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag22	Inconsistency marriage before getting to know each other (current partner)	0	No inconsistency	
		1	Inconsistency	First month of marriage earlier than date of getting to know each other
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag23	Inconsistency home size & number of rooms	0	No inconsistency	
		1	Inconsistency HH1	Home size \leq 10qm and number of rooms $>$ 2 or Home size \leq 20qm and number of rooms $>$ 4
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant

continued on next page

Table A.21, continued from previous page

Variable	Label	Values	Value Labels	Description
flag24	Inconsistency overnight stays in main and second residence	0	No inconsistency	Stay in residence one and two almost every night
		1	Inconsistency	
		-3	Does not apply	
flag25	Inconsistency biological parents in household who should be dead (acc. preload)	0	No inconsistency	Person mentioned although this person is dead (acc. to preload)
		1	HH1: biological mother is dead	
		2	HH1: biological father is dead	
		3	HH2: biological mother is dead	
		4	HH2: biological father is dead	
-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant		
flag26	Relationship between two anchor persons	0 "x"	No inconsistency Couple "x"	Two anchor persons have a relationship
flag27	Age sibling at birth of sibling's child	0 1	No inconsistency Inconsistency	Sibling of anchor younger than 12 years old at birth of sibling's biolog. child
flag_cas	Inconsistency current activity status	0 "a ₁ a ₂ "	No inconsistency Inconsistency activ. status a ₁ & a ₂	Multiple answers (a ₁ & a ₂) for current activity status not plausible
pflag_cas	Inconsistency current activity status, partner	0 "a ₁ a ₂ "	No inconsistency Inconsistency activ. status a ₁ & a ₂	Multiple answers (a ₁ & a ₂) for current activity status not plausible

continued on next page

Table A.21, continued from previous page

Variable	Label	Values	Value Labels	Description
flag_ehc	1st (and 2nd) month not recorded	0	No inconsistency	
		1	Int. month W1 missing	Distance between waves 17 months; update of preload month (i.e. month 1) missing
		2	Int. month W1 and next month missing	Distance between waves 18 months, update of preload month (i.e. month 1) and following month missing
		3	Ehc missing completely	Event-history calendar missing due to programming error
		4	Int. month W1 missing	Update of preload month missing, but not due to wave distance (some other mistake)
		5	Invalid months before int. month W1	Event-history calendar started one month too early
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag_frt6	Probably inconsistent data in frt6, values are too high	0	No inconsistency	
		1	Potentially inconsistent	Respondents who are expecting a child and regard additional children as realistic or respondents who are not expecting a child, but already have children and regard at least as many children as realistic as they already have
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag_igb	Inconsistent combination: biological and adoptive parents	0	No inconsistency	
		1	Inconsistency	Having both biological and adoptive parents inconsistent according to codebook
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant

continued on next page

Table A.21, continued from previous page

Variable	Label	Values	Value Labels	Description
flag_isco88 _kldb1992	Coding problem kldb and isco	0	No problem	Classification of occupations problematic (also see chapter 4.2)
		1	Qualification unclear, low- est coded	
		2	Activity not unequivocally assignable	
flag_isco08 _kldb2010	Coding problem kldb and isco	-7	Incomplete data	Classification of occupations problematic (also see chapter 4.2)
		-3	Does not apply	
		0	No problem	
		1	Qualification unclear, low- est coded	
		2	Activity not unequivocally assignable	

Table A.22: List of tag variables to identify inconsistencies with preloads (*anchor\$*)

Variable	Label	Values	Value Labels	Description
tag_sex	Inconsistency sex anchor between waves	0	No inconsistency	Anchor's sex in current wave is not anchor's sex in preload
		1	Inconsistency	
tag_dob	Inconsistency date of birth anchor between waves	0	No inconsistency	Anchor's date of birth is not anchor's date of birth in preload
		1	Inconsistency: both month & year	
		2	Inconsistency: month	
tag_idenk"x"	Inconsistency identity child "x" between waves	0	No inconsistency	Child "x" in previous wave is not the same child as in current wave
		1	Inconsistency	
tag_sexk"x"	Inconsistency sex child "x" between waves	0	No inconsistency	Child "x" changed sex Child deleted
		1	Inconsistency: sex child	
tag_dobk"x"	Inconsistency date of birth child "x" between waves	0	No inconsistency	Year of birth previous wave is not equal to year of birth current wave Month/range of birth previous wave is not equal to month/range of birth current wave
		1	Inconsistency	
tag_biok"x"	Inconsistency status child "x" in anchor data between waves	0	No inconsistency	Status of child "x" is not equal to status of child "x" in preload Child "x" deleted
		1	Inconsistency: status child	
tag_biokp"x"	Inconsistency biological parent partner child "x" between waves	0	No inconsistency	Status partner (biological parent) varies between waves
		1	Inconsistency: biological parent partner	

continued on next page

Table A.22, continued from previous page

Variable	Label	Values	Value Labels	Description
tag_biokp"x"	Inconsistency biological parent partner child "x" between waves	2	Child deleted	Child "x" deleted
tag_identp	Inconsistency identity current partner between waves	0 1	No inconsistency Inconsistency	Current partner of current wave is partner 1 (same partner as in previous wave) instead of new partner 2
tag_dobp	Inconsistency date of birth partner between waves	0 1	No inconsistency Inconsistency	Date of birth partner previous wave is not equal to date of birth partner current wave
tag_sexp	Inconsistency sex partner between waves	0 1	No inconsistency Inconsistency	Sex partner previous wave is not equal to sex partner current wave

Table A.23: List of flag variables to identify inconsistencies (*parent\$*)

Variable	Label	Values	Value Labels	Description
parflag1	Number of children	0	No inconsistency	parsd32 \neq number of children indicated in childbiography OR (parsd32 = -5,-2)
		1	Discordantly information about number of children	
parflag2	Anchor information	0	No inconsistency	(parsd19k1y \neq doby (from anchor's interview))
		1	Anchor's year of birth in question 3 does not match anchor interview	
		2	Anchor's status in question 3 does not match anchor interview	
parflag3	Age 2nd child	0	No inconsistency	(50 < (parsd19k2y - pardoby) < 13) if parsd15k2 = 1,2 & parsex = 2 (70 < (parsd19k2y - pardoby) < 13) if parsd15k2 = 1,2 & parsex = 1
		1	Respondent <13 or >50/70 when bio-childbirth	
parflag4	Age 3rd child	0	No inconsistency	<i>analogous to parflag3</i>
		1	Respondent <13 or >50/70 when bio-childbirth	
parflag5	Age 4th child	0	No inconsistency	<i>analogous to parflag3</i>
		1	Respondent <13 or >50/70 when bio-childbirth	
parflag6	Partner status	0	No inconsistency	(parcrn21i1 = 8 & parsd3 = 1,2)
		1	No partner in question 12 but partner in question 48	

continued on next page

Table A.23, continued from previous page

Variable	Label	Values	Value Labels	Description
		2	Partner in question 12 but no partner in question 48	(parcrn21i1 = 1,2,3,4,5 & parsd3 = 3) OR (parcrn21i2 = 1,2,3,4,5 & parsd3 = 3) OR (parcrn21i3 = 1,2,3,4,5 & parsd3 = 3)
parflag7	Number of grandchildren	0 1	No inconsistency Ambiguous # of grandkids	Coding from wave 2 through wave 4: pargc2i6=1 (no grandchildren) and at least (one valid value in pargc2i1-pargc2i5 or missing in pargc2i1- pargc2i5) for correction see variable grandparkids Wave 5: at least one missing in pargc2i1- pargc2i5
parflag8	Parent's age	0 1 2	No inconsistency Parent <12 years old at respondent's birth or adoptive/stepparent Living parent 100+ years old	<i>see value labels</i>
parflag9	Support by non-existent people	0 1 2 3 4 5 6 7	No inconsistency Mother Father Children other than Anchor Partner Siblings Son-/daughter-in-law Grandchildren	Mother not alive (parigr22 = 2) Father not alive (parigr25 = 2) Only one child (parsd32 = 1) No partner (parsd3 = 3) No siblings (parnet17i2 = 0 & parnet17i1 = 0) No children with partner (parsd35k1 & parsd35k2 & parsd35k3 & parsd35k4 ≠ 1) No grandchildren (pargc2i6 = 0)

continued on next page

Table A.23, continued from previous page

Variable	Label	Values	Value Labels	Description
parflag10	Household composition	0	No inconsistency	(parhh = 0 & (parigr41k1 = 1 OR parigr41k2 = 1 OR parigr41k3 = 1 OR parigr41k4 = 1)) (parhh = 0 & parsd3 = 1)
		1	Children in hh (question 3) but single-hh (question 47)	
		2	Partner in hh (question 48) but single-hh (question 47)	
parflag11	Leaving home	0	No inconsistency	(parigr9y < pardoby)
		1	Left home before birth	
parflag12	Educational degree	0	No inconsistency	<i>see value labels</i>
		1	Multiple answer: highest degree kept	
		2	Multiple answer incl. other	
parflag13	Occupational degree	0	No inconsistency	Any valid occupational degree provided & (parsd29i8 = 1)
		1	No but also valid occupational degree	
parflag14	Mother's partnership status	0	No inconsistency	papa17i* != igr30 (from Anchor's interview)
		1	Anchor's answer doesn't match mother's answer	
parflag15	Father's partnership status	0	No inconsistency	papa17i* != igr35 (from Anchor's interview)
		1	Anchor's answer doesn't match father's answer	
parflag_cas	Current activity	0	No inconsistency	Multiple answers for current activity status that are not plausible
		1019	Unemployed but full-time	
		1219	Unemployed but part-time	

continued on next page

Table A.23, continued from previous page

Variable	Label	Values	Value Labels	Description
		1021	Fulltime but occupational/disability	
		1721	Maternal oder paternal leave or other/Retired, occupational disability	
		0	No inconsistency	
parflag_lfs1	Labor force status	1	Employed but no hours of work mentioned	<i>see value labels</i>

A.5 Changes in codebooks

Table A.24: Changes in codebooks

Codebook	Release/Date of Change	Description of Changes/Variables affected
Anchor 1	Release 2.0	Improved English translation (no revision with regard to content)
Partner 1	Release 2.0	Improved English translation (no revision with regard to content)
Partner 1	Release 3.1	psd19kxm, psd19kxy, psd14kxg, psd15kx: Correction of variable names psd10: Change in response categories
Anchor 2	Release 3.0	inc10, pa3: Item numbers changed Output of EHC-Screen H6: ehc18pX instead of ehc14pxi1
Partner 2	Release 2.0 2011/11/21	Major corrections in English version, (nearly) all variables revised
Child 2	Release 2.0 2011/11/21	cpcr4, cpcr5: filter changed
Child 2	Release 5.0	Improved English translation; improved filter expressions
Parent 2	Release 3.0	Question numbers changed
Parent 3	Release 3.1 Release 5.0	parigr41kx, parigr45a, parigr47a, parigr80a, pargc5g, parigr101ix, parigr12ix: Minor changes in variable labels parigr53a-59a, parigr60a-66a, parigr78a-79a, parigr80a-81a: Minor change in formulation of a response category and response code (6 -> 8) parigr103, parsd23i13, parsd23i21: Minor change in formulation of response categories parigr40p1, parigr40p3, parigr41p1, parigr41p3: Filter changed parpa17ix: Missing response category added Categories "no answer" and "does not apply" of variables are now differentiated.
Parenting 3	Release 3.1	pcr3: Change in formulation of response categories
Partner 3	Release 5.0	pjob2: Correction of answer categories in english version pinc21: Correction of answer categories in english version psex5, psex6, psex7, psat5, pftr1: Correction of literal error in german and english version of response category "I don't want to answer that. / No answer"
Child 3	Release 5.0	Improved English translation; improved filter expressions
Partner 4	Release 5.0	psex5, psex7, psat5: Correction of literal error in english version of response category "I don't want to answer that. / No answer"
Child 4	Release 5.0	Improved English translation; improved filter expressions