



pairfam Data Manual

Release 6.0

Revised since Release 5.0, May 2014

Josef Brüderl, Kristin Hajek, Michel Herzig, Bernadette Huyer-May, Rüdiger Lenke, Bettina Müller, Philipp Schütze, Nina Schumann

University of Munich (LMU)

With contributions from:

Philipp Alt, Petra Buhr, Christine Finn, Nadia Lois, Timo Peter, Veronika Salzburger, Elisabeth Schubach, Carolin Thönnissen, Barbara Wilhelm

May 30, 2015

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	3
1.3	Quick Start files	4
2	Overview of the data structure	5
2.1	Survey instruments and data sets	5
2.2	System of person identifiers	6
2.3	Variable names	7
3	pairfam Nuts and Bolts	8
4	Anchor data	10
4.1	Data editing	10
4.2	Generated variables and scales	22
4.3	Process-generated variables	45
4.4	Paradata	47
4.5	Weights	47
4.6	Generated data sets	51
5	Partner data	89
5.1	Data editing	89
5.2	Generated variables and scales	90
6	Parent data	93
6.1	Data editing	93
6.2	Generated variables and scales	95
7	Child data	99
7.1	Data editing	99
7.2	Generated variables and scales	101
8	Parenting data	104
8.1	Data editing	104
8.2	Generated variables and scales	105
9	DemoDiff	106
9.1	Introduction	106
9.2	Data structure	106
9.3	Differences in detail	107
9.4	Codebook	108
10	Step-up data	109
11	Outlook	111

A	Appendix	114
A.1	DemoDiff-pairfam differences	114
A.2	Changes between releases	120
A.3	Information on longitudinal weights	138
A.4	Flag and tag variables	144

1. Introduction

The aim of this manual is to facilitate work with pairfam data sets. If you have questions concerning the data that are not covered by this manual, please contact the pairfam user service at support@pairfam.de.

In the following it is assumed that the reader has some basic knowledge of the pairfam design. If not so please read first the section on the pairfam design in Technical Paper No. 01 (Brüderl et al. 2015).

1.1 Obtaining the data

The data are made available to registered data users only. To become a registered user, interested researchers should follow the instructions on the pairfam data access website. The data are distributed by GESIS.

From wave 4 on, former CAPI children enter the main survey and become anchor respondents themselves. With release 6.0 these data are included in the scientific use file. For details see chapter 10.

Beginning with release 6.0, the new data set *Overview_multi_actor* is available which provides an overview of participation in the anchor, partner, child, parenting and parent survey. For details see chapter 4.6

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 corresponding 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 proper citation to all your publications that are based on pairfam data.

The most important citation rule is that usage of pairfam data should be acknowledged by citing both the reference paper (Huinink et al. 2011) and the dataset (Brüderl et al. 2015). For the details on the rules of pairfam citation, please refer to the citation website.

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; Herzig, Michel; Huyer-May, Bernadette; Lenke, Rüdiger; Müller, Bettina; Schütze, Philipp; Schumann, Nina (2015): pairfam Data Manual. Release 6.0, University of Munich, Technical report.

1.3 Quick Start files

The pairfam team developed several commented Stata do-files (Quick Starts) for an easy start with 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 6.0, will be available soon on the corresponding pairfam website:

- “Sample Definition”: information on defining the analysis sample,
- “Weighting”: 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 EHA”: information on how to use the generated data set “biopart”,
- “Biochild”: information on how to use the generated data set “biochild”,
- “Panel Analysis FE”: demonstration of an exemplary panel analysis using the pairfam data of wave 1 to 6,
- “Definition Homosex”: information on how to define homosexuality in pairfam.

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 technical paper 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*. Thus there are two main tasks for the data analyst: respondents' information over time has to be compiled and information of multiple respondents connected by family ties has to be matched. In this chapter we explain how the pairfam data have been organized to facilitate these two tasks.

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 their children). Furthermore, beginning with wave 2 we have been collecting information on the anchor's and his or her partner's parenting.

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. 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).¹

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 (\$).

With each release, previously published data sets are updated if necessary.²

¹There 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.

²See 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, A.15 for changes between release 3.1 and 4.0., tables A.16, A.17, A.18 and A.19 for changes between release 4.0 and 5.0 and tables A.20, A.21, A.22, A.23 and A.24 for changes between release 5.0 and 6.0.

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 easy matching of alteri information to anchor data (and matching any two 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 an anchor's first partner (*pid*) adds the trailing number 101 to the household number (only those partners get an *pid* who were in a relationship with the anchor at time of interview). The same *pid* will be kept as long as the anchor still is in a relationship with this partner. New partners will be numbered consecutively³.

The child ID is *cid*. Only children who are selected for a CAPI interview get a child ID. In the long format child\$ data sets (one line for each child) there is only the variable *cid*. In the anchor\$ data sets the child IDs are reported on one line as *cid1, cid2, . . .*. This is denoted by the suffix (x). The variable *parentidk* contains the identification number of a child's second biological parent (the first is always the anchor). Note that this variable can differ from the variable *pid*, which denotes the current partner of the anchor. The computation of this variable is based upon the latest version of the data set *biochild* (for further information see chapter 4.6).

The anchor's siblings are also assigned a unique identifier (*sibidx*), which is part of the parent data (where *x* denotes the ordering of the siblings). The parent data include up to three parents per anchor respondent. Each parent provides information about his/her children - the anchor plus his or her siblings. Each sibling is identified by the child's name, gender, year of birth, and status as living or deceased. Based on this information using a record-linkage algorithm (Schnell et al. 2004) siblings are identified over different parents and are assigned the identifier. This is the anchor identifier (*id*) plus a trailing number 401 for the first sibling, 402 for the second, 403 for the third, and so on.

In order to identify the time of the survey, a wave identifier is included in each cross-sectional data set. The variable *wave* is assigned value "1" in data sets of the first wave, value "2" in second wave data sets, and so forth.

³Please note that due to a programming error 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	id + 101, . . . , 150 (only for current partners)
Anchor's child	cidx	id + 201, . . . , 299 (only for CAPI children)
	parentidkx	pid of second biological parent of anchor's child(ren)
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.⁴

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.

⁴See, for example, variables *pa3* or *per1_* in data sets *anchor*\$.

3. pairfam Nuts and Bolts

In this chapter we gather tips, tricks, and known problems of the pairfam data. Most users will benefit from reading this chapter carefully.

How to deal with DemoDiff?

DemoDiff is a supplementary East German sample that has been integrated into the original pairfam sample data (for details on DemoDiff, see chapter 9). Thus, if the user does nothing, the DemoDiff cases enter his/her analyses beginning with wave 3. However, including DemoDiff data in pairfam panel analyses is not trivial. One problem is that DemoDiff “destroys” the monotonic design. In the original pairfam data, all respondents enter at wave 1, and some gradually attrite from the panel. If one simply appends data sets `anchor$`, then all DemoDiff respondents enter at wave 3. When preparing data, users need to be aware that some respondents enter at wave 3. Further, users must realize that DemoDiff respondents in wave t have not participated t times in the survey, but only $t-1$ times. Further, users must decide how to merge DemoDiff wave 1 data. Merging it with pairfam wave 1 data creates an artificial gap in the panel structure (wave 2 information is missing). Merging it with pairfam wave 2 might create incompatibilities as the contents of DemoDiff wave 1 and pairfam wave 2 differ. Finally, if one intends to do weighted analyses, one has to use special weights (for details see section 4.5).

To avoid these complications, one can simply delete the DemoDiff data. Adding this command to each analysis file achieves this:

```
* Keep only original pairfam sample
keep if demodiff==0
```

Expected number of children in waves 1+2 (*frt6*)

In each wave, we ask about the ideal and realistically expected number of children. In the first two waves, however, the wording of the question about the realistically expected number of children (*frt6*) did not clearly state that we wanted to know about *additional* children. On the other side, the question on the ideal number of children referred to the *total* number of children. Most respondents *with* children were confused by this different wording and obviously reported the *total* number of children also on *frt6*. Therefore, the realistically expected number of children in waves 1 and 2 is for respondents with children in most cases too high. As of wave 3, we have reworded the question for respondents with children, asking if the respondent realistically expects having additional children (*frt27*) and in a second question how many additional children the respondent expects (*frt28*). The wording for respondents without children remained unchanged (*frt26*).

For respondents with children, we therefore warn from naive use of *frt6* in waves 1 and 2! If you want to use these data, we recommend using an adjusted version of *frt6*. For further information, see the pairfam Technical Paper No. 04 (Buhr and Huinink 2014) and the section about checks for consistency in chapter 4.1. The ideal number of children (*frt5*) is not concerned.

Social network data in waves 2+4 (*net1px-net16px*)

In waves 2 and 4, ego-centered social network data was collected. This data is compromised by large interviewer effects as presumably most interviewers have curtailed the questions entering no or only a few network persons. Thus, network sizes are unrealistically small in the pairfam dataset. A detailed analysis of the problem can be found in (Brüderl et al. 2013).

Thus, one should obviously not interpret the absolute network size, because most pairfam respondents will have much larger networks than reported in the pairfam data. However, multivariate analyses of effects on network size seem to be valid, as first experiences show. Also analyses considering characteristics of the persons in the networks are possible. However, one should always account for interviewer effects (e.g. by using multilevel models) when analyzing pairfam network data.

Mode change (CAPI vs. CASI) after wave 1 (*per1_*)

In the first wave, questions on personality (self-esteem (self-worth), loneliness, emotional autonomy, shyness, and explosiveness, tendency to anger) were asked in CAPI mode, i.e. the interviewer asked the respondent directly. From wave 2 onward, these questions were asked in the self-interview (CASI) section. As a consequence, values of self-esteem and emotional autonomy are higher in wave 1 than in subsequent waves, and values of loneliness and shyness are smaller in wave 1 than in subsequent waves. This suggests the scales were used differently across the waves, probably due to social desirability.

If wave 1 data on personality are to be used in analyses, we suggest correcting them. One correction method would be normalizing the anchor scores with respect to the partner scores. This method is described in the technical report by Sonntag et al. (2014), as well as Mund et al. (2015, in press). A second method for dealing with the problem would be to include a dummy variable for the first wave in panel regression models. The coefficient of the dummy would then capture the mode effect in wave 1.

Matching of grandparents and grandchildren in waves 2-6

Respondents of the parents' questionnaire are asked to answer some questions about one focal grandchild whose name is written on the cover sheet of the PAPI questionnaire. Despite the explicitly mentioned name, many respondents did not consider the correct grandchild when answering the questions. This can be seen, for instance, from differing entries in the parent and anchor data regarding the grandchild's sex or age. We estimate that approximately one third of the data is about grandchildren other than the designated focal grandchild. The variable *parcorgc* in the parent data indicates if the information about the grandchild mentioned by the grandparent refers to the focal grandchild or to another grandchild (for details see chapter 6).

English or German labels?

Stata data 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).

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. 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.

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 6. In total, 270 (wave 2), 237 (wave 3), 282 (wave 4), 259 (wave 5) and 235 (wave 6) variables were preloaded to the interview.

The preloads or DI-variables are contained in the data sets *anchor2* to *anchor6*. 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 *anchor5* 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 cohab-

itation 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, wave 5 and wave 6) 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 interviews¹. 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.² 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 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

¹The 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.26).

²Specifically, 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.

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).³ 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.

³For 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).

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
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 - 6
crn14kxi13o	Child care afternoon child x: Other, own entry	Other mentioned	2 - 6
crn18kxo	Child x: Custody - other, namely:	Other mentioned	2, 3, 5
crn1kxo	Other place of birth child x	Other mentioned	2, 3, 5, 6
d1	Preload: Day of birth (Preload)	Day mentioned	2 - 6
d134, . . . , d143	City xst place of residence (Preload)	Residence mentioned	2 - 6
d14, . . . , d28	Name child x (Preload)	Name mentioned	2 - 6
d164, . . . , d173	Country xst place of residence (Preload)	Country mentioned	2 - 6
d176, . . . , d195	Name xst household member [main residence] (Preload)	Name mentioned	2 - 6
d218, . . . , d226	Name xst household member [second residence] (Preload)	Name mentioned	2
d267	Other education (Preload)	Yes	2 - 6
d274	Other type of employment (Preload)	Yes	2 - 6
d282	Current employment activity: open-ended answer (Preload)	Occupation mentioned	2 - 6
d398	Current partner's day of birth (Preload)	Day mentioned	2 - 6
d59, . . . , d73	Day of birth child x (Preload)	Day mentioned	2 - 6
d506	Name of marriage partner with whom no relationship at prev. wave (Preload)	First name mentioned	4 - 6
d8	Name of current partner (Preload)	Name mentioned	2 - 6
dobd	Day of birth	Day mentioned	all
ehc12kxo	Other parent name child x (EHC)	Name mentioned	2 - 6
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 - 6
ehc19i16o	Other type of employment, open entry currently (EHC)	Other mentioned	2 - 6
ehc19i22mxo	Other type of unemployment, open entry in month x (EHC)	Other mentioned	2 - 6

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 - 6
ehc19i9mxo	Other education open entry in month x (EHC)	Other mentioned	2 - 6
ehc19i9o	Other education open entry currently (EHC)	Other type of training mentioned	2 - 6
ehc1pxn	Name partner x (EHC)	Name mentioned	2 - 6
ehc20d	Day of birth of partner from previous wave (EHC)	Day mentioned	2 - 6
ehc22pxn	Name person x in household (EHC)	Name mentioned	3 - 6
ehc27pxi1	City place of residence x (EHC)	Residence mentioned	3 - 6
ehc27pxi2o	Country place of residence x (EHC)	Country mentioned	3 - 6
ehc7kxn	Name child x (EHC)	Name mentioned	2 - 6
ehc8kxd	Day of birth child x (EHC)	Day mentioned	2 - 6
frr13i14o	Reasons against child: Other Reason	Reason mentioned/ Other mentioned	1, 2, 4 - 6
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 - 6
hcp1i2o	Country main residence partner	Country mentioned	2 - 6
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
igr74o	Open entry country of birth adoptive mother	Other mentioned	3
igr75o	Open entry (highest) level of school education of mother	Other mentioned	3

continued on next page

continued from previous page

Variable	Variable label	Anonymous value label	Wave
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 - 6
igr83o	Open entry country of birth mother's partner	Other mentioned	3 - 6
igr85i12o	Citizenship father's partner: Open entry other country	Other mentioned	3 - 6
igr86o	Open entry country of birth father's partner	Other mentioned	3 - 6
int12o	Open entry suggestions or comments on the interview	Mentioned	3 - 6
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
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 - 6

continued on next page

continued from previous page

Variable	Variable label	Anonymous value label	Wave
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 - 6
sdp17o	Open entry other level of school education	Other mentioned	3 - 6
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 -4.⁴ 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.

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.⁵ 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

⁴If 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.

⁵This happened in addition to a number of checks that had already been implemented in the CAPI program (see anchor codebooks).

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.26 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.27 lists all of the generated tag variables that are part of the data sets *anchor2*, *anchor3*, *anchor4*, *anchor5* and *anchor6*.

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

When analyzing variable *frt6* 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 (*frt7*) for your analyses instead of *frt6*.
2. Use *frt6* only for childless respondents who are not pregnant (unflagged values). For these cases there should be no problem with this question.

Note that *frt6* 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 *frt6* was divided into three questions. In question *frt26* respondents without children were asked how many children they will have realistically. Furthermore, to rule out wrong answers a filter variable was integrated (*frt27*). Question *frt27* 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 (*frt28*). Therefore *flag_frt6* is not relevant for wave 3 and the following waves.

Remark on variable flag_frt (wave 6)

Due to a filter error in the programming of the general questionnaire, several respondents did not see and answer the questions 260-263 (variables *frt7, frt8, frt9, frt16*). In order to correct this, TNS Infratest conducted a follow-up survey (PAPI) among respondents affected by the filter error. As information on the four relevant variables was not raised identically, *flag_frt* was created to indicate if the information was either given in the general survey or collected in the follow-up survey.

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, wave 5 and wave 6 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.⁶ 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.24. 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.

⁶Users should, however, start with the latest version of the wave 1 to wave 5 data (release 6.0) 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.24.

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 for each wave separately. Users are invited to adapt the syntax to their special research needs.⁷ 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.19.

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. For additional information regarding these scales, please refer to the scales manual (Thönnissen et al. 2015).

Table 4.3: List of generated variables included in data sets *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*
Ethnicity	ethni
Migration status	migstatus
Relationship status	relstat
Marital status (anchor, partner)	marstat, pmarstat

continued on next page

⁷Running the do-files requires Stata version 11.0 or later.

continued from previous page

Construct	Variable name
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)	mmrd
Father lives in household (main residence)	fmrd
Partner lives in household (main residence)	pmrd
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
KIDB 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

continued on next page

continued from previous page

Construct	Variable name
Net equivalence income according to modified OECD scale	hhincoecd
Number of persons aged under 14 main residence	npu14mr
Number of persons aged 14 and older main residence	npo14mr
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

In order to solve the problem of conflicting information from various sources⁸ regarding respondents' sex, we have generated "best solution" sex variables for the anchor, the partner, and the children (*sex_gen*, *psex_gen*, *k*sex_gen*).

The **_gen* variables are synchronized over the waves. Consequently, each respondent has identical (best) sex information in all waves. The best sex information was created according to the following rules: (1) Self-reported sex information was preferred over proxy information. (2) The value stated most often was used. (3) If two values had been stated equally often, the most recent value was preferred. If the partner or a child was nonexistent, we assigned the code "-3 Does not apply". If the information to derive respondents' sex was not available, we used the code "-7 Incomplete data". The Stata do-file *identifiers.do*, which can be found in the syntax folder of the current wave, contains the syntax used to compute these variables. Note that the code -4 was assigned in order to mark a change in respondents' sex over the panel waves.

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

In order to solve the problem of conflicting information from various sources⁹ regarding respondents' date of birth, we have generated best solution variables for the month and the year of birth of the anchor, partner, parents (incl. adoptive parents), stepparents, and 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 gender (see above). The Stata do-file *identifiers.do*, which can be found in the syntax folder of the current wave, contains the syntax used to compute these 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, partner's, parents' (incl. adoptive parents'), stepparents', and 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 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

⁸The various data sets of the anchor and the alteri, (partially) available for different waves.

⁹See footnote above.

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 *varkX* 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 except for the former capikids. From wave 4 on, every year some former capikids enter the main questionnaire and become anchors themselves (for details see chapter 10). Therefore two additional categories were added. We differentiate between former capikids that have their first interview (category “0 Former capikid first interview”) and former capikids that have participated in the anchor survey before (category “9 Former capikid re-interview”). 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 6 we use *igr86* and *igr83* from wave 6 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 6 to compute *pcob*. If there is insufficient information in wave 6, information of the previous wave is used, if available. Due to privacy concerns, the syntax which generates these variables cannot be published.

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*¹⁰. 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. Please note that for wave 3, 4 and 5, release 6.0, some corrections have been performed.

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

¹⁰So far, three nationalities exist for partners only.

non-German background". In addition, the variable *migstatus* reports the anchor's migration status 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*)¹¹. The variables *reldur*, *cohabdur*, and *mardur* document the durations of

¹¹If *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*).

the current relationship, cohabitation, and marriage in months. If the anchor reported more than 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 in a specific wave. Note that this variable is time-varying and can change from wave to wave.

This variable is built from “revealed sexual preference”, i.e. if the sex of the anchor and his/her current partner are the same (using *sex_gen* and *psex_gen*). If there is no current partner, we look whether there has been a partner between the last and the current wave. If the last between-wave partner has the same sex (*ehc1pxg*, $x=1,2,3,4,5$) as the anchor, we classify the anchor as homosexual in the current wave. If there is neither a between-wave nor a current partner, the value of *homosex* is carried forward from the last wave.

Only in wave 1 those anchors without a current partner were asked, whether they are looking for a female or male partner (*sex2*). If an anchor with no current partner in wave 1 answered that he/she is looking for a same-sex partner he/she was classified as homosexual. Thus, these anchor are classified by their “stated sexual preference”. This stated sexual preference is carried forward until a partner (between-wave or current) comes in. Then the definition of homosexuality switches to the concept of revealed sexual preference as explained above.

homosex gets the code “-7 Incomplete data” if an anchor didn’t want to answer *sex2*, or if he/she said that he/she is looking for both sexes. The syntax used to compute the variable is contained in the Stata do-file *homosex.do*.

There seems to be a problem with this definition, because answers to the question *sex2* seem to be not very valid. About two thirds of the stated homosexuals in wave 1 are observed later on in heterosexual partnerships only. Thus, these anchors are probably misclassified as homosexual in wave 1 (and so on, until they begin their first heterosexual relationship). This group comprises almost half of the homosexuals of wave 1, meaning that almost half of the homosexuals in wave 1 are probably misclassified.

Thus, we do not recommend to use this variable. It seems more reasonable to use an alternative definition of homosexuality that rests solely on revealed sexual preference (as described above). The definition of the alternative *homosex* variable is detailed in a Quick Start file available on the corresponding pairfam website (“Definition Homosex”).

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 6 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, 4 and 6 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. See Stata do-file *nkids.do* for the syntax used to compute the variables.

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

Variable	Label
<i>nkids</i>	Number of all kids born until time of interview
<i>nkidsbio</i>	Number of all biological kids born until time of interview
<i>nkidsp</i>	Number of all biological kids with partner born until time of interview
<i>nkidsalv</i>	Number of all kids alive
<i>nkidsbioalv</i>	Number of all biological kids alive
<i>nkidspalv</i>	Number of all biological kids with partner alive
<i>nkidsliv</i>	Number of all kids living with anchor
<i>nkidsbioliv</i>	Number of all biological kids living with anchor
<i>nkidspliv</i>	Number of all biological kids with partner living with anchor
<i>pnkidsbioalv</i>	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. As of wave 4, the information on the domain of residence was not derived from the EHC (as it was the case in waves 1-3). Rather, 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 `hhsizedo`.

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 `hhsizedo`.

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 `hhsizedo`.

Other household members at main residence - `othmrd`

The variable `othmrd` indicates how many other 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 `hhsizedo`.

Children living at main residence - `childmrd`

The variable `childmrd` indicates how many children lived at the anchor’s main residence. Only co-habiting 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 `hhsizedo`.

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*.

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

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, wave 3 and wave 5 by the partner in the partner survey, this information was used instead. In wave 2, wave 4 and wave 6 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 6 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 6 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, wave 4 and wave 6 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 6, 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	Lower, GDR, POS 8./9.
4	Intermediate, Realschulabschluss / mittlere Reife
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 6. 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 6. 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 *iscd*, *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 *iscd* and *pisced* provided the basis for two additional variables (*iscd2*, *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 *iscd*, *pisced*, *miscd*, *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*.

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)

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,
- 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 *job20o* in waves 2 to 6, 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 6, release 6.0, 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 6.0.

Variable *flag_isco08_kldb2010*¹² 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 6, release 6.0. (see Hartmann et al. (2012) for more details). Variable *flag_isco88_kldb1992* based on the former classification is provided in wave 1, release 6.0 (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 6, release 6.0, 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 6.0 (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.

¹²This variable is a renamed copy of variable *beruprob*, which was originally provided by *TNS Infratest*.

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.¹³ 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 6, release 6.0, 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 6, release 6.0, coding of the variable *siops* is based on the new classification ISCO-08.

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*.

Net equivalence income according to the modified OECD equivalence scale and variables for creating equivalence scale weight - *hhincoecd*, *npu14mr*, *npo14mr*

Equivalence income according to the modified OECD equivalence scale assigns a value of 1 to the first household member, 0.5 to each additional adult, and 0.3 to each child below the age of 14. The variable net equivalence income (OECD) (*hhincoecd*) was generated by dividing the household net income (*hhincnet*) by the equivalence scale weight.

Calculation of the equivalence scale weight requires information on the age of all household members. The variables *npu14mr* and *npo14mr* indicate the number of persons besides the anchor aged under and over 14 years in the main residence of the anchor (the anchor respondent is assigned a value of 1

¹³ More 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.

in the equivalence scale weight and thus not counted). For waves 1 and 2, the variables were created based on the age variables of the generated household data set and, starting with wave 3, based on the household matrix and the age variables in the anchor data sets. Some assumptions have been made in order to minimize the number of missing values of the variables *npu14mr* and *npo14mr*, e.g. if the age of the parents, grandparents or parents-in-law is missing in the original data set, it is reasonably assumed that these persons are older than 14. In wave 1, age was not recorded for each household member, and has thus been imputed using the generated household data from wave 2. We assigned a value of 0.5 for household members of anchor persons who did not participate in wave 2, as well as for household members who had moved house since wave 1. Respondents sharing a residence with roommates or housemates, living in dormitories, boarding schools, or similar residences for which the assumption of shared resources within the household is problematic, have been assigned an equivalence scale weight of 1 and are thereby considered single-person households. If the household income was missing, the net equivalence income according to the modified OECD equivalence scale was coded to “-7 Incomplete data”. The syntax used to compute the variable *hhincoced* and the variables *npu14mr* and *npo14mr* is contained in Stata do-file *incnet.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. 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*.¹⁴ 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

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*. Coding of the variable *cpi* is based on the baseline 2010.

¹⁴This 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

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, wave 4 and wave 6 we had only proxy information given by the anchor person. Concerning educational attainment, this information is less rich than for anchors.¹⁵

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

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 stated¹⁶:
 - primary activity is defined according to the dominance rule:
 - general secondary school > UCE¹⁷ > 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:

¹⁵For partners, we only know if they are involved in general schooling, vocational training, or university.

¹⁶Due to filtering and question wording partners cannot have multiple education activities.

¹⁷University of Cooperative Education ("Berufsakademie")

- 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 age¹⁸.

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*.

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.,

¹⁸For 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

- 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.

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 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

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

“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. For further information, refer to the scales manual (Thönnissen et al. 2015), which contains a detailed description of all scales.

Table 4.17: List of scales included in syntax file *scales_anchor\$*

Construct	Variable name	Wave
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	1,2,4,5
Value of Children: Costs of comfort	vocccomf	1,2,4,5
Single: Interest in partnership	partint	1 - 3
Single: Desire for partnership	partdes	all
Single: Broad exploration	siexplbr/siexplbr2	1,2/3 - 6
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

continued on next page

continued from previous page

Construct	Variable name	Wave
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	conflom_apd	1,5
	conflom2_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
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 - 6
Intergenerational Relationships: Conflict with father	confl_afd	all
Intergenerational Relationships: Conflict with partner of father	confl_asmd	2 - 6
Intergenerational Relationships: Intimacy mother	intim_ams	all
Intergenerational Relationships: Intimacy partner of mother	intim_asfs	2 - 6
Intergenerational Relationships: Intimacy father	intim_afs	all
Intergenerational Relationships: Intimacy partner of father	intim_asms	2 - 6
Intergenerational Relationships: Admiration mother	admir_amo	2,4,6
Intergenerational Relationships: Admiration partner of mother	admir_asfo	2,4,6
Intergenerational Relationships: Admiration father	admir_afo	2,4,6
Intergenerational Relationships: Admiration partner of father	admir_asmo	2,4,6
Intergenerational Relationships: Dominance mother	domin_amo	2,4,6
Intergenerational Relationships: Dominance partner of mother	domin_asfo	2,4,6
Intergenerational Relationships: Dominance father	domin_afo	2,4,6
Intergenerational Relationships: Dominance partner of father	domin_asmo	2,4,6
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	sucindivd_amd	2
Intergenerational Relationships cohort1: Successful individuation Father/Partner of mother	sucindivd_afd	2

continued on next page

continued from previous page

Construct	Variable name	Wave
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
Emotional autonomy	emotautn	1,5
Self-worth	selfworth	all
Depressiveness	depressive	2 - 6
Anger	anger	3 - 5
Activity	activ/activ2	3/4 - 6
Stress	stress	4 - 6
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 - 6
Newborn temperament child1	temperc1	2 - 6
Newborn temperament child2	temperc2	2 - 6
Newborn temperament child3	temperc3	2 - 6
Newborn temperament child4	temperc4	4 - 6
Newborn temperament child5	temperc5	4 - 6
Newborn temperament child6	temperc6	4 - 6
Newborn temperament child7	temperc7	4 - 6
Unspecific strain	unspstrain	3
Autonomy in the parenting role (3 Item Scale)	autonoms	2
Autonomy in the parenting role (4-Item Scale)	autonom	4,6
Autonomy in the parenting role (2-Item Scale)	autonom2	4,6
Pleasure in the parenting role (newborn module)	pleasure_ newborn	2
Pleasure in the parenting role (children till 15 years)	pleasure	4,6
Parental Self Efficacy/Competence	comperz	2,4,6
Coparenting	coparent	2,4,6
Coparenting with ex-partner	coparent_ ex	3
Coparenting with ex-partner1	coparent_ exp1	5

continued on next page

continued from previous page

Construct	Variable name	Wave
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
Parenting goals: Status	pgoalstatus	2 - 6
Parenting goals: Autonomy	pgoalautn	2 - 6
Parenting goals: Competence	pgoalcomp	2 - 6
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...	5,6
	conduct_aco7	
SDQ Behavior: Hyperactivity childx	hyper_aco1,....,	5,6
	hyper_aco7	
SDQ Behavior: Emotional symptoms childx	emotion_aco1,....,	5,6
	emotion_aco7	
Emotional warmth childx	warmth_acs1,....,	5,6
	warmth_acs1	
Negative communication childx	negcomm_acs1,....,	5,6
	negcomm_acs7	
Inconsistent parenting childx	inconsist_acs1,....,	5,6
	inconsist_acs7	
Work-Family-Conflict: Work impacts family	wif_conflict	6
Work-Family-Conflict: Work impacts family - time	wif_time	6
Work-Family-Conflict: Work impacts family - stress/strain	wif_strain	6
Work-Family-Conflict: Family impacts work	fiw_conflict	6
Work-Family-Conflict: Family impacts work - time	fiw_time	6
Work-Family-Conflict: Family impacts work - stress/strain	fiw_strain	6
Dissolution of boundaries in labour	wf_boundary	6
Parental Relationship: Conflict	confl_mfd	6

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	bula
Size of community in 7 categories	gkpol
Official key of the municipality	gkz
Settlement structure	bik

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 corresponding 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 and wave 6 (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.¹⁹ In this case, we assigned value “-7 Incomplete data”. For wave 2 to 6 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.²⁰ 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 *anchor\$*

Construct	Variable name
Total number of interviewer contacts with respondent	<i>intcont</i>
Interviewer’s sex	<i>intsex</i>
Interviewer’s age	<i>intage</i>
Interviewer’s school degree (waves 5 & 6)	<i>intedu</i>
Interviewer-ID	<i>intid</i>
Duration of CAPI interview in minutes	<i>intdur</i>
Date of interview (month, day, year)	<i>intm</i> , <i>intd</i> , <i>inty</i>

4.5 Weights

Weights are provided to 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

¹⁹Some digits were missing, and it was not possible to recover the information whether it was the hours or the seconds missing.

²⁰Duration 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.).

dxpsweight Combined design and post-stratification weight.

For the DemoDiff sample with release 6.0 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*.

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

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 data²¹, 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 .930.²²

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.²³ 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).

²¹Query at www.destatis.de at 01/20/2010.

²²The 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.

²³According 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.

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.²⁴ The degree to which the factors fit to the distribution of these variables is shown in table 4.23.

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 ²⁵	94.58%	96.05%	96.28%
BIK	98.96%	99.73%	99.82%
Federal 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 ²⁶	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 6 the weighting factors of temporary dropouts (people who skip a maximum of one wave) refer to wave 5 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*).

²⁴Note 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).

²⁵The 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.

²⁶Weight 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.

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.25, 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. Longitudinal weights are also available for DemoDiff waves 3 to 6. Sample information is added to the logistic regression models of the respective waves (see table A.25, 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 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 regarding 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 individual life history as well as information on parents was compiled into separate, user-friendly data sets to allow for 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 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 on 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 only 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 1-6.

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 rather 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 three *biomob* data sets *biomob_ehc*, *biomob_rtr* and *biomob_rtr_parents* is organized in "long" format, with one row for each episode. The data set *bioparent* is also provided in "long" format, with one row for each parent.

As of Release 6.0, there is a new panel data set *Overview_multi_actor*, which provides an overview about participation in the partner, child, parenting and parent surveys.

Please note that the data set *household* only contains information up to and including wave 3. As no conceptual changes to the variable generating process have taken place since wave 3, the do-files from these previous waves still provide instructions on how to enrich later waves data-sets with additional household information.

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()"²⁷.

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", and "-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) were assigned the value "-99 Right-censored, ongoing". In the data sets *bioact* and *bioact_rtr*, information on censoring is given by a separate variable.

If the respondent provided information on the year but not on the month, 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 such 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. 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.

²⁷Consequently, the complete command for year is: `gen year=year(dofm(old variable-720))` and for month: `gen month=month(dofm(old variable-720))`.

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, with one row for each partner ("long" format).²⁸ 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 date²⁹. 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.

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*³⁰ 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 *wavex* (with *x*=number of wave) indicates in which year the interview was conducted.

The variable *partindex* 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; a 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 value "2", and partners who had been 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 *partcurrwx* (with *x*=number of wave) was included.³¹

For all partners, the variable *sexp* contains the "best" information on the partner's sex (also see chapter 2.2). For partners at the time of at least one interview, the partner's date of birth was known and included (variable *dobp*).³² The identification number *pid* only exists for partners at the time of at least one interview 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.

²⁸This is the major difference compared to *biopart* of release 1.0, which was organized in "wide" format.

²⁹Please 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.

³⁰Generated variable "Homosexual union" based on variables *sex* and *sexp*.

³¹Note 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.

³²In case of inconsistencies between different versions of *dobp* in different waves for the same partner, the latest version is employed.

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. This information regarding previous partners was integrated into *biopart* as well.

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

Variable	Variable label	Values	Value labels
<i>General information</i>			
id	Person number anchor		-
demodiff	DemoDiff sample	0 1	pairfam demodiff
cohort	Birth cohort	1 2 3	1991-1993 1981-1983 1971-1973
intdatwx	Date of interview wave x	-10 -3 <i>date</i>	Not in demodiff Does not apply
sex	Sex anchor	1 2	Male Female
dob	Date of birth anchor	<i>date</i>	-
wavex	Survey year: wave x	-10 -3 1 2 3 4 5 6	Not in demodiff Does not apply 2008/09 2009/10 2010/11 2011/12 2012/13 2013/14
respwx	Respondent in wave x	0 1	No Yes
homosex_p	Homosexual union	-7 0 1	Incomplete data No Yes
<i>Ascending order of relationships</i>			
partindex	Correct order of relationships: Number partner	-3 1 2 ... 14	Does not apply 1st partner 2nd partner ... 14th partner
<i>Partners at the time of the previous interview</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>			
pnowx	Number partner in wave x	-3 <i>number</i>	Does not apply -

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
partcurrwx	Current partner in wave x	-10	Not in demodiff
		0	No
		1	Yes
sexp	Sex partner	-7	Incomplete data
		1	Male
		2	Female
dodp	Date of death partner	-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
relbeg	Beginning relationship	-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
relend	End relationship	-99	Ongoing
		-66	Death partner
		-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
bkbeg	Beginning break k relationship	-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
bkend	End break k relationship	-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
<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>	-

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
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 co- habitation	0	No imputation
		1	Only year information
		2	Only season information
imp_cohabend	Imputed date of end cohabitation / end of break of co- habitation	0	No imputation
		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	Inconsistency

Please note that the variables *relend*, *cohend*, and *marend* may be codes as "'-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.26)
 - Overlapping episodes of cohabitation with same or different partners (also see *flag6* and *flag7*, table A.26)
 - End of previous marriage after beginning of current marriage (also see *flag8*, table A.26)
 - Divorce from partner to whom never married (also see *flag10*, table A.26)
 - Separation through death/divorce current spouse (also see *flag11*, table A.26)
 - Separation before beginning relationship (current partner; also see *flag14*, table A.26)

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 *anchor5* as of wave 2. They mark inconsistencies regarding the first meeting with the current partner.

Solution to these inconsistencies:

We have 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 of additional information regarding 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 who were 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, 3, 4 or 5.
3. Partners at the time of previous interviews: Partners who were listed as 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 *partcurrwx* 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 *partcurrw\$* is "0" in the previous wave and "1" in the current wave (*partcurrw\$*: "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 are no longer current partners in the respective wave and who had not died since the previous wave (*partcurrw\$*: "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), *biopart5.do* (data from wave 5) and *biopart6.do* (data from wave 6). 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 children³³ and on episodes of living together with the children in the same household. The data set is in "long-long" format, meaning that there is one row per child and for each child there exists one row per wave (that the child was mentioned in). For 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 waves 1 to 6 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.³⁴ 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 *child6*, respectively.

³³According to the anchor questionnaires, the term "children" refers to biological children, adopted children, children of a partner, or foster children.

³⁴For 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*.

For all children, the variable *sexk* contains the “best” information on the children’s sex (also see section 4.2).³⁵ 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 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*.

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. To identify the second biological parent as former partner, the variable *parentid* is constant over all waves for every child. 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
dob		<i>date</i>	-
intdat	Date of interview (months since january 1900)	-10 <i>date</i>	Not in demodiff
sex	sex anchor	1 2	Male Female
cohort	Birth cohort	1 2 3	1991-1993 1981-1983 1971-1973
wave	Survey year	1 2 3 4 5 6	Wave 1: 2008/09 Wave 2: 2009/10 Wave 3: 2010/11 Wave 4: 2011/12 Wave 5: 2012/13 Wave 6: 2013/14

continued on next page

³⁵This “best” information was taken from the generated identifiers *k*sex_gen*.

continued from previous page

Variable	Variable label	Values	Value labels
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
statusk	Status of child	-7 -3 1 2 3 4	Incomplete data Does not apply Biological child Adopted child Partner's child/stepchild Foster child
currliv	Cohabitation with child	-7 -3 1 2 9 10 11 12 13	Incomplete data Does not apply Only with anchor With anchor and elsewhere Only alone/flat share Only with other parent unit Only with other relative Children's home Only elsewhere
currliv_detail	Details on [currliv] category <2. With anchor and elsewhere>	-3 2 3 4 5 6	Does not apply With anchor but also alone/flat share With anchor but also with other parent unit; mainly with anchor With anchor but also with other parent unit; namely in equal shares with anchor and with other parent unit With anchor but also with other parent unit; mainly with other parent unit With anchor but also with other relative

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
		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)	-
sfid	Person number stepfather	(see table 2.2)	-
livkbeg	Beginning living together with child	-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
livkend	End living together with child	-99	Ongoing
		-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
b1livkbeg	Beginning break 1 living together with child	-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
b1livxend	End break 1 living together with child	-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
livkbirth	Living together since birth with child	-7	Incomplete data
		-3	Does not apply
		0	Not living together since birth
		1	Living together since birth
dodk	Date of death child (months since January 1900)	-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 to live together
- Beginning to live 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 (duplicate name and date of birth), we dropped the second entry. Moreover, we checked whether any second biological parent was identified as inconsistent, and dropped them while generating the data set *biopart* (for more information, see the above section). If this discrepancy occurred, the correct value (in all cases “0 Current partner”) was assigned.

Checks for consistency across waves 1 to 6:

- Child's date of death (*dodk*):
If death occurred in wave *t* but the information from wave *t+1* does not reflect this, then wave *t+1* information is preferred to that of wave *t*. If two different dates of death have been recorded, the newest information (wave *t+1*) is preferred.
- Child's status:
If there are unrealistic differences between the information in waves 1 to 6, the newest information is preferred. Such unrealistic differences were changes in status, e.g. 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 add 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 the variable *parentid*. This variable exists only for a quick 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 (waves 1 and 2) of the data were modified and how the checks to identify inconsistencies in the retrospective sections were generated. The do-files *biochild3.do* through *biochild6.do* show the preparation of the wave 3 - 6 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 from the time of the wave 1 interview, i.e. with the release of wave 2 data. It contains data from *pairfam* as well as *DemoDiff*. It covers the period between the survey data of the first and the current wave. The data set is provided in “long-long” format, meaning 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 from 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 each month after the preceding wave up to the current interview. 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 *actcurrwx* (with x =number of wave) mark the activities reported in the interviews of each respective wave. Additionally, *spell* presents the number of separated episodes per activity. The variable *actcensor* 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. The dummy variables *respwx* indicate whether the anchor took part in the respective interview (identification variable for temporary dropouts).

The structures of the two data sets *bioact* and *bioact_rtr* are almost identical. The variables carrying the activity information in the *bioact_rtr* data set are marked with the suffix '*_rtr*'. 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 *actflag1_rtr* (beginning) and *actflag2_rtr* (ending). When the respondent answered "none of the above" to the questions 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's sake, 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. These variables can be easily merged with variables from anchor data by using the key variable *id*.

Concerning the variables *ehc19i23* and *ehc19i23m** from the anchor data set (which were used to generate the *bioact* data) additional alterations needed to take place. 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 as the variables *ehc19i23* and *ehc19i23m** in anchor data set. These variables were constructed to act as "gap fillers" for months in which 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 these variables work 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.

The do-files *bioact2.do* through *bioact6.do* document in detail how the data set *bioact* was computed. The do-file *bioact_rtr* documents how the data set *bioact_rtr* 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>			
id	Person number anchor	(see table 2.2)	-
demodiff	DemoDiff sample	0 1	pairfam demodiff
cohort	Birth cohort	1 2 3	1991-1993 1981-1983 1971-1973
dob	Date of birth	<i>date</i>	-
sex	Sex anchor	1 2	Male Female
wavex	Survey year wave x	-3 1 2 3 4 5 6	Does not apply 2008/09 2009/10 2010/11 2011/12 2012/13 2013/14
<i>Activity history</i>			
activity_rtr	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 education University of cooperative education (also "Berufsakademie") University of applied sciences, college, university General secondary school (first education) Pre-vocational training Technical/professional school Other education Full-time employment Self-employment Part-time employment (also multiple 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	No education since	0	Not mentioned
(bioact_rtr only)	18th birthday and wave 1	1	Mentioned
rtr35i14	No occupation between 18th birthday and wave 1	0	Not mentioned
(bioact_rtr only)		1	Mentioned
actflag1_rtr	Ambiguous information on beginning	-3	Does not apply
(bioact_rtr only)		21	Original information on the month
		24	Original information on the month
		27	Original information on the month
		30	Original information on the month
		32	Original information on the month
actflag2_rtr	Ambiguous information on ending	-3	Does not apply
(bioact_rtr only)		21	Original information on the month
		24	Original information on the month
		27	Original information on the month
		30	Original information on the month
		32	Original information on the month
spell_rtr	Counter - Number of spells per activity	<i>number</i>	-
actbeg_rtr	Beginning of activity in month	<i>date</i>	-
actend_rtr	End of activity in month	<i>date</i>	-
actcensor_rtr	Indicator for censored spells	-3	No censoring (only for ehc19i23*)
		0	Uncensored
		1	Left-censored
		2	Right-censored, end of episode is missing
		3	Right-censored, ongoing episode
		4	Combination of 1 & 2
		5	Combination of 1 & 3
		6	left censored, beginning is missing
		7	Combination of 3 & 6

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
actcurrwx	Current activity in wave x	0	No
		1	Yes
respwx	Respondent in wave x	0	No
		1	Yes

Table 4.29: List of available variables covering education and occupation 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, 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. As we have only used 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*. The generated household data set was not continued after wave 3 because it was too time consuming to prove if the household members were identical over the waves. Furthermore, the household matrix was improved starting with wave 3 so that it is easier for users to generate the composition of the household and identify changes over the waves.

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 have distinguished between characteristics of partners (*p**), children (*c**), parents (*pa**), and other household members (*other**). Variables *p*_rel* contain information on cohabitation with partners. We assigned the value of 1 to variable *p1_rel* if the anchor reported that he/she lives still or is living again with the partner from waves 1 or 2. We assigned value 1 to variable *p2_rel* if the anchor cohabited with a new partner.

The anchor's children were assigned either the value of 10 (biological children) or 11 (step/adopted/foster children) to *c*_rel* if the anchor reported living 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 the variables

*ehc21p**, *ehc22p**, and *ehc25p*h1*.

For all persons (partners, children, parents) for which we had the respective information, 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.³⁶

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*. As of wave 3 we also matched the person number of the anchor's child(ren) (*cid*) to the 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 members present in both waves.

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

Variable	Variable label	Values	Value labels
<i>id</i>	Person number anchor	(see table 2.2)	-
<i>wave</i>	Survey year	1 2 3	Wave 1: 2008/09 Wave 2: 2009/10 Wave 3: 2010/11
<i>pid</i>	Person number partner	(see table 2.2)	-
<i>mid</i>	Person number mother	(see table 2.2)	-
<i>fid</i>	Person number father	(see table 2.2)	-
<i>smid</i>	Person number stepmother	(see table 2.2)	-
<i>sfid</i>	Person number stepfather	(see table 2.2)	-
<i>cid</i>	Person number capic child	(see table 2.2)	-
<i>intm</i>	Month of interview	(see section 4.4)	-
<i>inty</i>	Year of interview	(see section 4.4)	-
<i>sex_gen</i>	Generated sex anchor	1 2	Male Female
<i>doby_gen</i>	Generated year of birth anchor	<i>see doby*_gen</i>	-
<i>dobm_gen</i>	Generated month of birth anchor	<i>see doby*_gen</i>	-
<i>hhsizemrd</i>	Number of hh members main residence	<i>see hhsizemrd</i>	-
<i>pmrd</i>	Partner lives in household main residence	-7 0 1	Incomplete data Partner does not live at main residence Partner lives at main residence

continued on next page

³⁶Information 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 to the anchor person.

continued from previous page

Variable	Variable label	Values	Value labels
mmrd	Mother lives in household main residence	-7	Incomplete data
		0	Mother does not live at main residence
		1	Mother lives at main residence
fmrdr	Father lives in household main residence	-7	Incomplete data
		0	Father does not live at main residence
		1	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	Incomplete data
		1	Own household
		2	Parental household (father/mother/step/foster parents)
		3	Shared dwelling with roommates/housemates
		4	Dormitory, student dormitory, boarding school, or similar
owner	Home-ownership	5	Other type of household
		-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		
nights	Nights spent at residence	7	joint property with partner
		8	Other
		-7	Incomplete data
rent	Monthly rent for dwelling (euros)	-3	Does not apply
		1...6	-
		-7	Incomplete data
expown	Monthly expenditures for self-owned apartment / house (euros)	-3	Does not apply
		-7	Incomplete data

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
dwsiz	Size of dwelling (square meters)	-7 -3	Incomplete data Does not apply
rooms	Number of rooms of dwelling	-7 -3	Incomplete data Does not apply
res2nd	Respondent has 2nd residence	-7 0 1	Incomplete data No 2nd residence 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 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	Incomplete data <i>Partner-Kids-parent-other</i> w - w - w - w w - w - w -w/o w - w - w/o -w w - w - w/o - w/o w - w/o - w - w w - w/o - w - w/o w - w/o - w/o - w w - w/o - w/o - w/o w/o - w - w - w w/o - w - w - w/o w/o - w - w/o - w w/o - w - w/o - w/o w/o - w/o - w - w w/o - w/o - w - w/o w/o - w/o - w/o - w w/o - w/o - w/o - w/o
hhincgcee	Net equivalence income (GCEE)	<i>see 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

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
cX_rel	Anchor's relation to child X	-3	Does not apply
		10	biological child
		11	step/adopt./foster child
cX_sex	Sex of child X	-7	Incomplete data
		-3	Does not apply
		1	Male
		2	Female
cX_byear	Year of birth of child X	-7	Incomplete data
		-3	Does not apply
cX_bmonth	Month of birth of child X	-7	Incomplete data
		-3	Does not apply
cX_id	Person number of child X	-7	Incomplete data
		-3	Does not apply
paX_rel	Anchor's relation to parent X	-3	Does not apply
		2	biological mother
		3	biological father
		4	adoptive mother
		5	adoptive father
		6	stepmother / father's partner
		7	stepfather / mother's partner
		8	foster mother
paX_sex	Sex of parent X	9	foster father
		-7	Incomplete data
		-3	Does not apply
		1	Male
paX_byear	Year of birth of parent X	2	Female
		-7	Incomplete data
paX_bmonth	Month of birth of parent X	-3	Does not apply
		-7	Incomplete data
paX_id	Person number of parent X	-3	Does not apply
		-7	Incomplete data
otherX_rel	Anchor's relation to partner X	-3	Does not apply
		-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		

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
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 household members throughout the waves, we used the person number for identification of partners and parents and the position in the event-history calendar for identification of children. These positions remain constant in each wave (unlike the position in the household grid). To identify the other persons in the household, we have 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 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 date of birth. If neither variable was available, we decided individually. Please note that in wave 3, household members' sex is unknown. Thus, 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 these cases as "new" household members. Later, we used the variables *other*_point* to match information from wave 3 to the former waves. Consistency checks showed that matching did not function properly if names had been 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 anchor ID, or the household members' ID (partner, child, parents) who are potentially respondents of the respective multi-actor surveys. Please note that the data set *household* does not encompass information beyond wave 3.

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, has been gathered through the Event History Calendar (EHC). The respondents were 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 their second and all further residences. As of wave 4, they were only asked for information concerning their main residences. Thus, information on the second residence can only be found in the data sets *anchor2* and *anchor3*. The data sets *biomob_ehc* and *biomob_ehc_moves* have been generated from this data gathered via the EHC in wave 2 - 6.

In wave 3, respondents answered questions on their migration history from 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 of these data sets contain information on migration and residential changes, the data has to be partitioned into more than one data set as the retrospective information does not include second residences and a residence at the parental house does not mean automatically that no spatial mobility occurred.

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 has been added: BIK-Classification *resbik* and migration distance *resdis* (the corresponding variables in the *biomob_rtr* data set are provided with the suffix “_rtr”). 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 *resbik* equates to the variable *bik* as 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 *resdis* 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 locations. The number 6370 equates to the radius of the Earth in kilometers. The calculated migration distances are rounded accurately to kilometers. If a place or residence is not detectable, *resdis* is coded to -99. If the calculation is not possible the code -98 is assigned. If there is a break between two episodes at a second residence, *resdis* is coded to -97. If there is no migration history, *resdis* is coded to “-3 does not apply”. If the respondent has not answered the questions or didn’t know the place of residence, *resdis* is coded to “-2 No answer” or “-1 Don’t know”.

Biomob_ehc

The data set *biomob_ehc* contains information on residences as collected through the EHC in pairfam and DemoDiff (see table 4.31). 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 *resbeg* and *resend*) are stored in the data set. The variable end is coded to “97 ongoing” for the respondents’ last episode.

To determine if an anchor person has already lived at a new place of residence, an identification number for the mentioned place of residence was generated.

The variables *index_mr* and *index_sr* 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 *resflag1* is coded to 1. The same problem appears with some second residences so the variable *resflag2* 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 *resflag3* is coded to 1.

To improve the quality of the data, the episodes marked *resflag1*=1 were inspected. If these episodes endured only one month and were followed by a non-missing episode they were added to the following episode. The latest information was assumed as better and therefore correct. If the place of residence of an ongoing episode was marked *resflag1*=1 the information provided through *ehc28px* “Currently living in dwelling in previous wave” was used to correct the episodes. In these cases the variable *resflag1* is still coded to 1.

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
cohort	Birth cohort	1 2 3	1991-1993 1981-1983 1971-1973
dob	Date of birth	<i>date</i>	-
sex	Sex anchor	1 2	Male Female
wavex	Survey year wave x	-3 1 2 3 4 5 6	Does not apply 2008/09 2009/10 2010/11 2011/12 2012/13 2013/14
respwx	Respondent in wave x	-10 0 1	Not in demodiff No Yes
intdatwx	Date of interview wave x	-10 -3 <i>date</i>	Not in demodiff Does not apply
residplace	City mentioned	0 1 2 3	No place mentioned Place mentioned demodiff Multiple places (<i>resflag1</i>)
resbeg	Beginning of episode	<i>date</i>	-
resend	End of episode	97 <i>date</i>	ongoing
rescensor	Censor	0 1 3 5 9	Uncensored Left censored, beginning of episode first interview Right censored; ongoing episode Combination of 1 & 3 Right censored, no secondary residences after wave 3

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
resnumber	Identification number city	-99	Place unknown
		-10	Not in demodiff
		<i>number</i>	-
resid	Main or secondary residence	1	Main residence
		2	Secondary residence
resland	Federal state (Bundesland)	-3	Does not apply
		-2	No answer
		-1	Don't know
		0	Berlin (West)
		1	Schleswig-Holstein
		2	Hamburg
		3	Niedersachsen (Lower Saxony)
		4	Bremen
		5	Nordrhein-Westfalen (North Rhine-Westfalia)
		6	Hessen (Hesse)
		7	Rheinland-Pfalz (Rhineland-Palatinate)
		8	Baden-Württemberg
		9	Bayern (Bavaria)
		10	Berlin (East)
		11	Brandenburg
		12	Mecklenburg-Vorpommern (Mecklenburg-Western Pomerania)
		13	Sachsen (Saxony)
14	Sachsen-Anhalt (Saxony-Anhalt)		
15	Thüringen (Thuringia)		
16	Saarland		
17	Another country		

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
index_mr	Index main residence	-3	Does not apply
		1	1st main residence
	
index_sr	Index secondary residence	36	36th main residence
		-3	Does not apply
		1	1st second residence
rescurrwx	Current place of residence in wave x
		15	15th second residence
		-3	Does not apply
resdis	Distance between consecutive residences	0	No
		1	Yes
		-99	Untraceable place
		-98	No calculation possible
		-97	Secondary residence with break
resflag1	Main residence unknown	-3	Does not apply
		-2	No answer
		<i>number</i>	-
		0	No
		1	Yes
resflag2	Secondary residence unknown	0	No
		1	Yes
resflag3	Filter error/Incorrect entry	0	No
		1	Yes

Biomob_ehc_moves

The EHC 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. 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 as the EHC was intended to only cover "real" migration, these changes 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 moves.

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

The variable *mig0_w1_w2_mr* contains information about how many of the moves described above occurred for the main residence between the interview date of waves 1 and 2. If the respondent did not participate in wave 2 but did participate in wave 3, the variable *mig0_w1_w3_mr* indicates how many moves occurred between waves 1 and 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_mr*, *within_w1_w3_mr*, *within_w2_w3_mr* 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. Respondents were asked only if they moved within a city, not how often.

For further information on 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
<i>General information</i>			
id	Person number anchor	(see table 2.2)	-
demodiff	DemoDiff sample	0 1	pairfam demodiff
cohort	Birth cohort	1 2 3	1991-1993 1981-1983 1971-1973
dob	Date of birth	<i>date</i>	-
sex	Sex anchor	1 2	Male Female
intdatwx	Date of interview wave x	-3 <i>date</i>	Does not apply
mig0_w1_w2_mr	No. of moves (resdis=0) w1 to w2: Main residence	-3 <i>number</i>	Does not apply
mig0_w1_w3_mr	No. of moves (resdis=0) w1 to w3: Main residence	-3 <i>number</i>	Does not apply
mig0_w2_w3_mr	No. of moves (resdis=0) w2 to w3: Main residence	-3 <i>number</i>	Does not apply
mig0_w2_w4_mr	No. of moves (resdis=0) w2 to w4: Main residence	-3 <i>number</i>	Does not apply
mig0_w3_w4_mr	No. of moves (resdis=0) w3 to w4: Main residence	-3 <i>number</i>	Does not apply
mig0_w3_w5_mr	No. of moves (resdis=0) w3 to w5: Main residence	-3 <i>number</i>	Does not apply
mig0_w4_w5_mr	No. of moves (resdis=0) w4 to w5: Main residence	-3 <i>number</i>	Does not apply
mig0_w5_w6_mr	No. of moves (resdis=0) w5 to w6: Main residence	-3 <i>number</i>	Does not apply
mig0_w4_w6_mr	No. of moves (resdis=0) w4 to w6: Main residence	-3 <i>number</i>	Does not apply
mig0_w1_w2_sr	No. of moves (resdis=0) w1 to w2: 2nd residence	-3 <i>number</i>	Does not apply
mig0_w1_w3_sr	No. of moves (resdis=0) w1 to w3: 2nd residence	-3 <i>number</i>	Does not apply
mig0_w2_w3_sr	No. of moves (resdis=0) w2 to w3: 2nd residence	-3 <i>number</i>	Does not apply
within_w1_w2_mr	Within city (ehc18) w1 to w2: Main residence	-3 0 1	Does not apply No Yes
within_w1_w3_mr	Within city (ehc29) w1 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_w3_mr	Within city (ehc29) w2 to w3:	-3	Does not apply
	Main residence	0	No
		1	Yes
within_w2_w4_mr	Within city (ehc29) w2 to w4:	-3	Does not apply
	Main residence	0	No
		1	Yes
within_w3_w4_mr	Within city (ehc29) w3 to w4:	-3	Does not apply
	Main residence	0	No
		1	Yes
within_w3_w5_mr	Within city (ehc29) w3 to w5:	-3	Does not apply
	Main residence	0	No
		1	Yes
within_w4_w5_mr	Within city (ehc29) w4 to w5:	-3	Does not apply
	Main residence	0	No
		1	Yes
within_w4_w6_mr	Within city (ehc29) w4 to w6:	-3	Does not apply
	Main residence	0	No
		1	Yes
within_w5_w6_mr	Within city (ehc29) w5 to w6:	-3	Does not apply
	Main residence	0	No
		1	Yes
within_w1_w2_sr	Within city (ehc18) w1 to w2:	-3	Does not apply
	2nd residence	0	No
		1	Yes
within_w1_w3_sr	Within city (ehc29) w1 to w3:	-3	Does not apply
	2nd residence	0	No
		1	Yes
within_w2_w3_sr	Within city (ehc29) w2 to w3:	-3	Does not apply
	2nd residence	0	No
		1	Yes

Biomob_rtr

The data set *biomob_rtr* contains information on all anchor 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 began living there, and about the date they moved away. This data was gathered during the interviews of wave 3. The *biomob_rtr* data set is provided in "long" format; every episode of a person is represented by one row.

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

The variables *resbeg_rtr* and *resend_rtr* 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 (*resbeg_rtr* = *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 and 12. If such random information for the month was generated, the original ambiguous seasonal information was saved in the variable *resflag1_rtr*. If the information on the month was coded to " -1 don't know" or " -2 no answer", the original information was stored in *resflag2_rtr*.

Geo-coordinates were collected for places of residence outside the borders of Germany. If there was information on the country but not 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 migration distances.

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

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

Variable	Variable label	Values	Value labels
id	Person number anchor	(see table 2.2)	-
demodiff	DemoDiff sample	0 1	pairfam demodiff
cohort	Birth cohort	1 2 3	1991-1993 1981-1983 1971-1973
dob	Date of birth	<i>date</i>	-
sex	Sex anchor	1 2	Male Female
wavex	Survey year wave x	-3 1 2 3 4 5 6	Does not apply 2008/09 2009/10 2010/11 2011/12 2012/13 2013/14
respwx	Respondent in wave x	-10 0 1	Not in demodiff No Yes
intdatwx	Date of interview wave x	-10 -3 <i>date</i>	Not in demodiff Does not apply
resplace_rtr	City mentioned	0 1 2	No place mentioned Place mentioned Demodiff
resbeg_rtr	Beginning of episode	<i>date</i>	-
resend_rtr	End of episode	97 <i>date</i>	ongoing
rescensor_rtr	Censor	0 1 3 5	Uncensored Left censored, beginning of episode first interview Right censored; ongoing episode Combination of 1 and 3
resnumber_rtr	Identification number city	-99 -10 <i>number</i>	Place unknown Not in demodiff -

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
resland_rtr	Federal state (Bundesland)	-3	Does not apply
		-2	No answer
		-1	Don't know
		0	Berlin (West)
		1	Schleswig-Holstein
		2	Hamburg
		3	Niedersachsen (Lower Saxony)
		4	Bremen
		5	Nordrhein-Westfalen (North Rhine-Westfalia)
		6	Hessen (Hesse)
		7	Rheinland-Pfalz (Rhineland-Palatinate)
		8	Baden-Württemberg
		9	Bayern (Bavaria)
		10	Berlin (East)
		11	Brandenburg
		12	Mecklenburg-Vorpommern (Mecklenburg-Western Pomerania)
		resindex_mr	Number of main residences
1	1st main residence		
...	...		
12	12th main residence		
resdis_rtr	Distance between consecutive residences	-99	Untraceable place
		-98	No calculation possible
		-3	Does not apply
		-2	No answer
		-1	Don't know
resflag1_rtr	Main residence unknown	<i>number</i>	-
		-3	Does not apply
		21	21 Original information on the month
		24	24 Original information on the month
		27	27 Original information on the month
	30	30 Original information on the month	

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
resflag2_rtr	Secondary residence	-3	Does not apply
	unknown	-2	-2 Original information on the month
		-1	-1 Original information on the month

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, as well as all following moves back into the parental house. This data set also is in "long" format. Every episode is represented by one 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 due to ambiguous seasonal information, the original information is stored in the variable *presflag1*. In the case of missing information on the month, the original information is stored in the variable *presflag2*. Analogous to this procedure, the original ambiguous seasonal information for the end of an episode is stored in the variable *presflag3* and the original missing variable is represented in *presflag4*.

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)	-
demodiff	DemoDiff sample	0	pairfam
		1	demodiff
cohort	Birth cohort	1	1991-1993
		2	1981-1983
		3	1971-1973
dob	Date of birth	<i>date</i>	-
sex	Sex anchor	1	Male
		2	Female
wavex	Survey year wave x	-3	Not in demodiff
		1	Does not apply
		2	2008/09
		3	2009/10
		4	2010/11
		5	2011/12
		6	2012/13
	2013/14		

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
intdatwx	Date of interview wave x	-10 -3 <i>date</i>	Does not apply
presindex	Number of episode at parental household	1 ... 5	First episode at parental household 5th Episode at parental household
presbeg	Beginning of episode at parental household	-4 -2 -1 <i>date</i>	Filter error/incorrect entry No answer Don't know -
presend	End of episode at parental household	-4 -2 -1 <i>date</i>	Filter error/incorrect entry No answer Don't know -
presflag1	Ambiguous information on month (Begin)	-3 -2 -1 21 24 27 30 32	Does not apply -2 Original information on the month -1 Original information on the month 21 Original information on the month 24 Original information on the month 27 Original information on the month 30 Original information on the month 32 Original information on the month
presflag2	Ambiguous information on month (End)	-3 -2 -1 21 24 27 30 32	Does not apply -2 Original information on the month -1 Original information on the month 21 Original information on the month 24 Original information on the month 27 Original information on the month 30 Original information on the month 32 Original information on the month

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
presflag3	Unknown information on month (Begin)	-3	Does not apply
		-2	-2 Original information on the month
		-1	-1 Original information on the month
presflag4	Unknown information on month (End)	-3	Does not apply
		-2	-2 Original information on the month
		-1	-1 Original information on the 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" format. This means that for each parent there is one row.³⁷ Also, biological and adoptive parents who died before the first wave are included. *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 6.

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)	-
demodiff	DemoDiff sample	0	pairfam
		1	demodiff
cohort	Birth cohort	1	1991-1993
		2	1981-1983
		3	1971-1973
intdatwx	Date of interview wave x	-10	Not in demodiff
		-3	Does not apply
sex	Sex anchor	<i>date</i>	-
		1	Male
dob	Date of birth anchor	2	Female
		<i>date</i>	-
wave	Survey year: wave x	-10	Not in demodiff
		-3	Does not apply
		1	2008/09
		2	2009/10
		3	2010/11
		4	2011/12
		5	2012/13
6	2013/14		

continued on next page

³⁷This is the major difference compared to bioparent of release 5.0 and below, which was organized in long long format.

continued from previous page

Variable	Variable label	Values	Value labels
respwx	Respondent in wave x	0	No
		1	Yes
sf_index	No. of stepfather	-3	Does not apply
		1	1st step father
	
		4	4th step father
sm_index	No. of stepmother	-3	Does not apply
		1	1st step mother
	
		5	5th step mother
partype	Type of parent	1	Biological mother
		2	Biological father
		3	Adoptive mother
		4	Adoptive father
		5	First stepmother
		6	First stepfather
		7	Second stepmother
		8	Second stepfather
		9	Third stepmother
		10	Third stepfather
		11	Fourth stepmother
		12	Fourth stepfather
		13	Fifth stepmother
pardob	Parent: Date of birth (months since January 1900)	-7	Incomplete data
pardod	Parent: Date of death (months since January 1900)	<i>date</i>	step parent
		-97	Incomplete data
		-7	Does not apply
		-3	Does not apply
marriedparwx	Parent is married with other biological parent in wave x	<i>date</i>	-
		-10	Not in demodiff
		-4	Incorrect entry/filter error
		-3	Does not apply
		-2	No answer
		-1	Don't know
		1	Yes
2	No		

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
cohabparwx	Parent is living together with other bio./adop. parent in one household in wave x	-10	Not in demodiff
		-4	Incorrect entry/filter error
		-3	Does not apply
		-2	No answer
		-1	Don't know
		1	Yes
		2	No
marstatparwx	Marital status of parent in wave x	-10	Not in demodiff
		-4	Incorrect entry/filter error
		-3	Does not apply
		-2	No answer
		-1	Don't know
		1	Single
		2	Married
		3	Civil union
		4	Divorced or dissolved civil union
5	Widowed or surviving partner in a civil union		
partnerparwx	Parent has a new partner in wave x	-10	Not in demodiff
		-4	Incorrect entry/filter error
		-3	Does not apply
		-2	No answer
		-1	Don't know
		1	Yes, with other bio. parent
		2	Yes
3	No		
cohabpartnerw	Parent is living together with the new partner in wave x	-10	Not in demodiff
		-3	Does not apply
		-2	No answer
		-1	Don't know
		1	Yes
		2	No
livanchorwx	Parent is living together with anchor in wave x	-10	Not in demodiff
		-7	Incomplete data
		-3	Does not apply
		0	Does not live at main residence
		1	Lives at main residence
contactwx	Contact of parent to anchor wave x	-10	Not in demodiff
		-4	Filter error/Incorrect entry
		-3	Does not apply
		-2	No answer
		-1	Don't know
		1	Daily
		2	Several times per week
		3	Once per week
		4	1-3 times per month
		5	Several times per year
6	Less often		
7	Never		
10	Never had contact		

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
samepartnerwx	New partner of the parent is the same partner as previous wave to wave x	-10	Not in demodiff
		-3	Does not apply
		-2	No answer
		-1	Don't know
		1	Yes
		2	No
begnocontact	Beginning of no contact of parents to anchor (months since January 1900)	-7	Incomplete data
		-3	Does not apply
		number	-
begpartnerwx	Duration of new partnership (categorized in years) wave x	-10	Not in demodiff
		-3	Does not apply
		-2	No answer
		-1	Don't know
		1	Less than five years ago
		2	5-10 years ago
		3	10-15 years ago
4	More than 15 years ago		

In addition to the anchor's ID (*id*), the data set contains the identification numbers for the parents and the step parents. *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. To differentiate between biological/adoptive and social parents the variable *partype* has been generated.

The variable *partype* 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).

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

The variables *partnerparwx*, *samepartnerwx*, and *cohabpartnerwx* 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 *begpartnerwx*.

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.

Multi-actor panel data - Overview_multi_actor

The data set *Overview_multi_actor* is a panel data set which provides an overview of participation in the anchor, partner, child, parenting, and parent survey. Each of the variables *respartner*, *reschildX*, *resparchildX_a*, *resparchildX_p*, *resfather*, *resmother*, *resstepfather*, and *resstepmother* is a dummy variable that indicates if partner, child, father, mother, etc. participate in the respective survey. If a relevant alteri does not exist or is not relevant for the respective survey, the variable assigned the value -3 "Does not apply". Please note that DemoDiff only surveyed anchor respondents and their partners between waves 1 and 4. If a multi-actor survey was not conducted in DemoDiff, the dummy variable was set to -10 "Not in DemoDiff". Apart from that, the data set is not available for former CAPI children.

Furthermore, the partner's identification number (*pid*) is included in this data set. Based on this person number and the variable *respartner*, users can easily find out how many same partners were surveyed during the panel.

The variable *childX* was generated to indicate if child x exists and was selected as a CAPI-child. The variable *reschildX* is a dummy variable that documents if a child was interviewed as part of the CAPI child survey. As already described in section "Anchor-child panel/episode data - biochild" the variable *parentidkX* contains the identification number of child's X second biological parent. By using this identification number, stepfamilies are easily to identify. If *pid* and *parentidkX* differ from each other that means that current partner is not the second biological parent. If children have different values in *parentidkX*, this is a hint that the children have two different biological parents. Please note that *parentidkX* as well as *pid* is only available for partners listed as current or previous partners in the course of the event-history calendar (EHC) and not for retrospective partners.

The variables *resparchildX_a* and *resparchildX_p* contain the information if anchor and/or their cohabiting partner fill out the parenting questionnaire. There are a few anchors and partners who fill out the questionnaire by mistake. These cases was set to -4. Please note that the selection of the children relevant for the parenting survey was slightly modified between the waves.

The variables *resfather*, *resmother*, *resstepfather*, and *resstepmother* provides information on participation in parent survey.

Table 4.36: List of variables included in data set *Overview_multi_actor*

Variable	Variable label	Values	Value labels
id	Person number anchor	(see table 2.2)	-
wave	Survey year	1 2 3 4 5 6	Wave 1: 2008/09 Wave 2: 2009/10 Wave 3: 2010/11 Wave 4: 2011/12 Wave 5: 2012/13 Wave 6: 2013/14
pid	Person number partner	(see table 2.2)	-
respartner	Participation in partner survey	-3 0 1	Does not apply No Yes
childx	Child x existing	-3 1 2	Does not apply Child exists, but is no capi-child Child exists, selected as capi-child
cidx	Child x: Person number	(see table 2.2)	-
reschildx	Child x: Participation in children's survey	-10 -3 0 1	Not in DemoDiff Does not apply No Yes
parentidx	Child x: Person number second biological parent	see variable pid	-
resparchildx_a	Child x: Participation parenting survey - anchor	-10 -4 -3 0 1	Not in DemoDiff PAPI-Qu. was filled out by mistake Does not apply No Yes
resparchildx_p	Child x: Participation parenting survey - partner	-10 -4 -3 0 1	Not in DemoDiff PAPI-Qu. was filled out by mistake Does not apply No Yes

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
kxdoby_gen	Child x: Generated year of birth	<i>see dob*_gen</i>	-
kxdobm_gen	Child x: Generated month of birth	<i>see dob*_gen</i>	-
resfather	Participation in parent survey: Father	-10 -3 0 1	Not in DemoDiff Does not apply No Yes
resmother	Participation in parent survey: Mother	-10 -3 0 1	Not in DemoDiff Does not apply No Yes
resstepfather	Participation in parent survey: Stepfather	-10 -3 0 1	Not in DemoDiff Does not apply No Yes
resstepmother	Participation in parent survey: Stepmother	-10 -3 0 1	Not in DemoDiff Does not apply No Yes
demodiff	DemoDiff sample	0 1	pairfam demodiff
cohort	Birth cohort	1 2 3	1991-1993 1981-1983 1971-1973
sex_gen	Generated sex anchor	<i>see sex_gen</i>	-

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 1-6 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 and 6. 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 *pfrt13i1*, . . . , *pfrt13i13* 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 as part of the scientific use file. For further information, refer to the scales manual (Thönnissen et al. 2015) which contains a detailed description of all scales.

Table 5.2: List of scales included in syntax file scales_ *partner*\$

Construct	Variable name	Wave
Traditional concept of marriage	ptradmarr	1,3,5
Value of Partnership: Negative expectations	pvopneg/pvopneg2	1/3,5
Value of Children: Benefit of stimulation	pvocbstim	1,2,4,6
Value of Children: Costs of comfort	pvocccomf	1,2,4,6
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	1,5
	pconfldom2_apd	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 - 6
Shyness	pshyness	1,4 - 6 (only new partners)
Emotional autonomy	pemotautn	1,3 - 6
Selfworth	pselfworth	all
Depressiveness	pdepressive	2 - 6

continued on next page

continued from previous page

Construct	Variable name	Wave
BIG 5: Neuroticism	pneurot	2 - 6 (only new partners)
BIG 5: Extraversion	pextrav	2 - 6 (only new partners)
BIG 5: Agreeableness	pagreeable	2 - 6 (only new partners)
BIG 5: Conscientiousness	pconscient	2 - 6 (only new partners)
BIG 5: Openness	popenness	2 - 6 (only new partners)
Coparenting with the other parent	pcoparent	2,4,6
Parental Self Efficacy/Competence	pcomperz	2,4,6
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	4,6
Parenting: Autonomy in the parenting role (short scale)	pautonom2	4,6
Parenting: Pleasure in the parenting role	ppleasure	4,6
Parenting goals: Status	ppgoalstatus	2,4 - 6
Parenting goals: Competence	ppgoalcomp	2,4 - 6
Parenting goals: Autonomy	ppgoalautn	2,4 - 6
Intergenerational relationships: Own intimacy within partner-mother dyad (self assessment)	pintim_ams	3 - 6
Intergenerational relationships: Own intimacy within partner-father dyad (self assessment)	pintim_afs	3 - 6
Intergenerational relationships: Conflict within partner-mother dyad	pconfl_amd	3 - 6
Intergenerational relationships: Conflict within partner-father dyad	pconfl_afd	3 - 6
Work-Family-Conflict: Work impacts family	wif_conflict	6
Work-Family-Conflict: Work impacts family - time	wif_time	6
Work-Family-Conflict: Work impacts family - stress/strain	wif_strain	6
Work-Family-Conflict: Family impacts work	fiw_conflict	6
Work-Family-Conflict: Family impacts work - time	fiw_time	6
Work-Family-Conflict: Family impacts work - stress/strain	fiw_strain	6

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, 5, and 6 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 results 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. Again, there were no cases deleted in the sixth wave. The final data set contains N=3,043 cases in wave six.

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.

A label was assigned to every variable in the data set. 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

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

data.

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: Three conditions had also to be met in this wave: 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).

In wave 6: Again, the three conditions had to be met in this wave: First, the filter question 11 had to be correct. Second, the sex of the grandchild indicated in question 14 had to match the sex of CAPI-child (provided in the anchor's interview). Third, the age of the grandchild indicated in question 15 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.28 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*, *parseparent*, and *parseanchor* 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. 2015). However, unlike other indicators, these scale indicators are not included in

the data file.

Note that in release 6.0 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 6.0, 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	<i>id</i>	2 - 6
Respondent's identification number	<i>parid</i>	2 - 6
CAPI-child's identification number	<i>cid</i>	2 - 6
Anchor's siblings' number	<i>sibidx</i>	2 - 6
<i>Paradata</i>		
Respondent's type (based on questionnaire)	<i>partype</i>	2 - 6
Sex of respondent (based on questionnaire)	<i>parseparent</i>	2 - 6
Position of respondent (based on questionnaire)	<i>parposition</i>	2 - 6
Sex of the anchor (based on questionnaire)	<i>parseanchor</i>	2
Wave (based on questionnaire)	<i>wave</i>	2 - 6
Language version (based on questionnaire)	<i>parlng</i>	2 - 4
Date of the interview	<i>parintd</i> , <i>parintm</i> , <i>parinty</i>	2 - 6
Cohort anchor	<i>parcohort</i>	2 - 6
Year of birth anchor based on anchor interview	<i>parageanchor</i>	2 - 6
# of children (corrected): # of children in <i>parsd32</i> ; only in case of more children described in <i>parsd14ff</i> , than stated in <i>parsd32</i> : correction; -7 if no valid answer in <i>parsd32</i> & no children described	<i>parkids</i>	2 - 6
# of grandchildren	<i>pargrandkids</i>	2 - 6
<i>Generated variables</i>		
Age variables (respondent, respondent's parents)	<i>parage</i> <i>parage2</i> <i>parmage</i> <i>parmage2</i> <i>parpage</i> <i>parpage2</i>	2 - 6 2
Current primary and secondary activity status	<i>parcasprim</i> , <i>parcassec</i>	2 - 6
Highest school degree	<i>parschool</i>	2 - 6
Highest vocational degree	<i>parvocat</i>	2 - 6
CASMIN classification of educational attainment	<i>parcasmin</i>	2 - 6
ISCED classification of educational attainment	<i>parisced</i>	2 - 6

continued on next page

continued from previous page

Construct	Variable name	Wave
Years of schooling / vocational qualification	paryeduc	2 - 6
Labor force status	parlfs	2 - 6
Household size main residence	parhhsizemrd	2,4,6
Net equivalence income	parhhincgcee	2
Type of children	park1type park2type park3type park4type	2 - 6
Marital status	parmarstat	2,4,6
Relationship status	parrelstat	2,4,6
Do parents live in a shared household?	parigr27	2 - 6
Are parents married?	parigr28	2 - 6
State / Bundesland	parbula	3 - 6
Size of community in 7 categories	pargkpol	3 - 6
Settlement structure	parbik	3 - 6
New parent in wave 3	parw3ne	3
New parent in wave 4	parw4ne	4
New parent in wave 5	parw5ne	5
New parent in wave 6	parw6ne	6
Information about the correct grandchild	parcorgc	3 - 6

Scales

The following scale variables (see table 6.3) can be generated by running the do-file `scales_parents$`, which are part of the scientific use file (SUF). For further information, refer to the scales manual (Thönnissen et al. 2015) which contains a detailed description of all scales.

Table 6.3: List of scales included in syntax file `scales_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	2,4,6
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

continued on next page

continued from previous page

Construct	Variable name	Wave
NRI Anchor Conflict	parconfl_parad	2 - 6
NRI Anchor Intimacy	parintim_paras	2 - 6
NRI Anchor Approval	paradmir_parao	2,4,6
NRI Anchor Dominance	pardomin_parao	2,4,6
Co-parenting with respect to anchor (younger than 21) (mean indicator: parcrn21i1, parcrn21i2, parcrn21i3)	parcoparent	2,4
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 child data from waves 2-6 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 2-6 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	Lable	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.

Recoding of SDQ Items

In order to match the item numbers of the child SDQ and the parenting SDQ, we recoded the items for the child SDQ starting with wave 6, retroactively until wave 2, where the SDQ was introduced. The changes were also implemented in the syntax file *scales* for every *scale_child\$* and the scales manual (Thönnissen et al. 2015).

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 - 6
cedu1o	Other schooltype	Other schooltype mentioned	2
cedu1ao	Other schooltype	Other schooltype mentioned	3 - 6
cpcr4	Name anchor	Name mentioned	2 - 6
cpcr5	Name current partner of anchor in same household	Name mentioned	2 - 6
cgp1	Name of anchor's mother	Name mentioned	2,4 & 6
cgp2	Name of anchor's father	Name mentioned	2,4 & 6
cpcr13	Name biological parent outside the household	Name mentioned	3 & 5
crom7	Name of romantic partner	Name mentioned	5,6
stepupa1	Step-up's complete name and address	Address received/ Same address as anchor	5,6

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 - 6
cinty	Child interview year	2 - 6
cintm	Child interview month	2 - 6
cintd	Child interview day	2 - 6
cagey	Child's age in years	2 - 6
cagem	Child's age in months	2 - 6
ctitlea	Child's naming for anchor	2 - 6
ctitlep	Child's naming for anchor's partner	2 - 6
ctitleop	Child's naming for the other biological parent outside the household	3, 5

The generated variables are included in wave 2-6 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. For further information, refer to the scales manual (Thönnissen et al. 2015), 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 - 6
Social integration	csocialinteg	2 - 6
Peer rejection	cpeerref	2 - 6
Parental school engagement	engagement	2,4 & 6
Economic deprivation	cecodep	2 - 6
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 - 6
Emotional warmth child-partner-relationship	cwarmth_cpo	2 - 6
Emotional warmth child-other parent-relationship	cwarmth_copo	3,5
Strict control child-anchor-relationship	cstrict_cao	2,4 & 6
Strict control child-partner-relationship	cstrict_cpo	2,4 & 6
Strict control child-other parent-relationship	cstrict_copo	3,5
Monitoring child-anchor-relationship	cmonitor_cao	2 - 6
Monitoring child-partner-relationship	cmonitor_cpo	2 - 6
Fear of love withdrawal child-anchor-relationship	lovewitanx_cas	3 - 6
Fear of love withdrawal child-partner-relationship	lovewitanx_cps	3 - 6
Fear of love withdrawal child-other parent-relationship	lovewitanx_cops	3,5
Intimacy child-anchor-relationship	cintim_cas	2 - 6
Intimacy child-partner-relationship	cintim_cps	2 - 6
Intimacy child-other partner-relationship	cintim_cops	3,5
Conflict child-anchor-relationship	cconfl_cad	2 - 6
Conflict child-partner-relationship	cconfl_cpd	2 - 6

continued on next page

continued from previous page

Construct	Variable name	Wave
Conflict child-other partner-relationship	cconfl_copd	3,5
Admiration child-anchor-relationship	cadmir_cao	2 - 6
Admiration child-partner-relationship	cadmir_cpo	2 - 6
Admiration child-other parent-relationship	cadmir_copo	3,5
Parental reliability child-anchor-relationship	creliabl_cad	2
Parental reliability child-partner-relationship	creliabl_cpd	2
SDQ: Conduct problems	cconduct	2 - 6
SDQ: Emotional symptoms	cemotion	2 - 6
SDQ: Prosocial behavior	cprosoc	2 - 6
SDQ: Hyperactivity	chyper	5,6
Negative communication in the anchor-child relationship	cnegcomm_cao	4 - 6
Negative communication in the partner-of-the-anchor-child relationship	cnegcomm_cpo	4 - 6
Intimacy own romantic partner	cintim_crps	5,6
Conflict own romantic partner	cconfl_crpd	5,6
Admiration own romantic partner	cadmir_crpo	5,6
Fear of love withdrawal own romantic partner	clovewitanx_crpd	5,6
Ambivalence in relationship to own romantic partner	cambiv_crpd	5,6
Child's perception of interparental conflict	ccpic_capd	6

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 2-6 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. As from wave 6, the missing category “-1 Don't know” no longer exists. 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. For further information, refer to the scales manual (Thönnissen et al. 2015), 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	sacrif_pacs	2
Parent-child-relationship: Hostile attribution	hostattr_pacd	2 - 6
Autonomy in the parenting role	autonom_pacs	2
Autonomy in the parenting role-short scale	autonom2_pacs	2
Pleasure in the parenting role	pleasure_pacs	2
Parent-child-relationship: Emotional warmth	warmth_pacs	2 - 6
Parent-child-relationship: Psychological control	psycontrol_pacs	2,5
Parent-child-relationship: Negative communication	negcomm_pacs	2 - 6
Parent-child-relationship: Monitoring	monitor_pacs	2 - 6
Parent-child-relationship: Strict control	strict_pacs	2,5
Parent-child-relationship: Inconsistent parenting	inconsist_pacs	2 - 6
SDQ: Hyperactivity	hyper_paco	2,5
SDQ: Emotional symptoms	emotion_paco	2 - 6
SDQ: Peer problems	peer_paco	2,5
SDQ: Prosocial behavior	prosoc_paco	2 - 6
SDQ: Conduct problems	conduct_paco	2 - 6
Parent-child-relationship: Intimacy	intim_paco	2 - 6
Parent-child-relationship: Conflict	confl_pacd	2 - 6
Parent-child-relationship: Admiration	admir_pacs	2,4 - 6
Parent-child-relationship: Dominance	domin_paco	2 - 6

9. DemoDiff

9.1 Introduction

In 2009 the Max Planck Institute for Demographic Research (MPIDR) initiated and funded DemoDiff (Demographic Differences in Life Course Dynamics in Eastern and Western Germany), a panel study closely following the design of pairfam (Kreyenfeld et al. 2012). DemoDiff started parallel with pairfam's wave 2 and was conducted by the MPIDR for three waves. Beginning with wave 5, DemoDiff has been fully integrated in pairfam.

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

- DemoDiff only sampled respondents residing in Eastern Germany (excluding West Berlin) at the time of first interview.
- It only sampled the cohorts 1971-1973 and 1981-1983.
- In its three waves it only surveyed anchor respondents and their partners, not their children or parents.

Further, there are differences concerning questionnaire content:

- DemoDiff's wave 1 questionnaire was a shortened version of pairfam's wave 1 questionnaire. Mainly psychological scales were dropped resulting in a 15 minute shorter interview duration.
- DemoDiff's wave 2 questionnaire was based on pairfam's wave 3 questionnaire (again dropping mainly psychological scales). This has been done in order to synchronize the two studies. However, some modules of pairfam's wave 2 questionnaire were included (e.g., childhood history).
- DemoDiff's wave 3 questionnaire was identical with pairfam's wave 4 questionnaire (except for the modules on consent for the parents' and children's interviews).

At the beginning of the field period of wave 5, respondents of the DemoDiff sample received a letter announcing the change from DemoDiff to pairfam. From this wave onward, DemoDiff respondents are treated as "normal" pairfam respondents.

This merger has two major implications for pairfam: First, the two older cohorts have been restocked, leading to higher Ns. Second, respondents of the two older cohorts living in Eastern Germany in 2009 are overrepresented in the sample.

9.2 Data structure

The DemoDiff data has been integrated into the original pairfam data. As of pairfam wave 3, both surveys operate in parallel. Therefore, DemoDiff waves 2 and 3 are fully integrated into the anchor and partner data sets (of pairfam waves 3 and 4!). However, DemoDiff wave 1 took place one year later than pairfam wave 1. Therefore, DemoDiff wave 1 data are stored in separate files anchor1_DD and partner1_DD. Thus, users of DemoDiff data must decide whether he/she wants to match DemoDiff wave 1 data to pairfam wave 1 or wave 2. Another exception is the childhood history section: This was collected in DemoDiff wave 2, but was part of pairfam's wave 2, and so has been stored in a

Table 9.1: Number of DemoDiff interviews

DemoDiff pairfam	Wave 1	Wave 2	Wave 3		
	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
	2009/10	2010/11	2011/12	2012/13	2013/14
Anchor	1,489	1,173	1,074	987	878
Partner	684	578	550	490	435
Child	/	/	/	233	240
Parent	/	/	/	371	305
Parenting	/	/	/	349	382
Data sets	anchor1_DD partner1_DD	anchor2_DD anchor3 partner3	anchor4 partner4	anchor5 partner5 child5 parent5 parenting5	anchor6 partner6 child6 parent6 parenting6

separate data set named anchor2_DD. Again it is left to the user, how to match these data with the original pairfam data.

Table 9.1 gives an overview of the number of respondents for each wave of DemoDiff.

Obviously, including DemoDiff data in pairfam panel analyses is not trivial. One problem is that DemoDiff “destroys” the monotonic design. In the original pairfam data, all respondents enter at wave 1, and some gradually attrite from the panel. If one appends data sets anchor\$, then all DemoDiff respondents enter at wave 3. Thus, when preparing data, users must be aware that some respondents enter at wave 3. Further, users must realize that DemoDiff respondents in wave t have not participated t times in the survey, but only t-1 times. Further, users need to decide how to merge DemoDiff wave 1 data. Merging it with pairfam wave 1 data creates an artificial gap in the panel structure (wave 2 information is missing). Merging it with pairfam wave 2 might create incompatibilities as the contents of DemoDiff wave 1 and pairfam wave 2 differ. Finally, if one intends to do weighted analyses, one has to use special weights (for details see section 4.5).

To avoid these complications, one can simply delete the DemoDiff data. Adding this command to each analysis file achieves this:

```
* Keep only original pairfam sample
keep if demodiff==0
```

9.3 Differences in detail

Some differences between the pairfam and the DemoDiff sets exist before pairfam 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” indicates this.

Variables that were dropped from DemoDiff

In order to adhere to the maximum interview time agreed upon with *TNS 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 DemoDiff waves 1 and 2. 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 2, 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.

9.4 Codebook

For wave 1 and wave 2, separate codebooks (for anchor and partner, respectively) are available for pairfam and DemoDiff. Questions that had been eliminated from the original pairfam-questionnaire have been crossed out. For all other waves the differences are marked in the combined codebooks.

10. Step-up data

Beginning with wave 4, respondents of the child interview who had reached the eligible age of 15 were asked to take part in the main anchor interview, including the partner interview. In addition, these step-up anchor respondents were asked to answer an additional, transitional PAPI questionnaire for the wave in which they first entered the anchor questionnaire. This transitional questionnaire was optional, and respondents were asked to fill it out only once. With release 6.0 these data are now available for analysis. Data and documentation of step-up respondents can be found in a separate step-up file.

The following step-up data sets are published for each wave: *stepup_anchor*\$, *stepup_partner*\$, *stepup_transition*\$ along with the generated data sets *stepup_biopart* and *stepup_biochild*. Each step-up data set includes the variables *entry* and *cohort*. The variable *entry* indicates the wave in which the former CAPI children entered the anchor survey. With the variable *cohort* we differentiate between former CAPI children that have their first interview (category “0 Former capikid first interview”) and former CAPI children that have participated in the anchor survey before (category “9 Former capikid re-interview”).

Data editing of the anchor, partner, and episode step-up data follows the same rules as for the main study. For details, please refer to chapters 4 and 5 in this manual. Data editing of the transitional questionnaire is described in the following section.

Data editing of transitional questionnaire

TNS Infratest has compiled a raw data set from the returned PAPI questionnaires. The data processing has been done in SPSS and all data editing steps are documented in an SPSS syntax file. Data editing includes general information about variable names, value labels, and missing values. Furthermore, this section illustrates the anonymization of the data.

Variable and value labels

Values were labeled according to the step-up transition codebook. To facilitate the analysis, variables derived from questions that were included in the anchor survey were labeled identically. The 10-stage response format of the variables *co1_**, however, ranges from 1=Not important at all to 10=Very important, and is therefore not consistent to the 11-stage response format of the corresponding variables *co1_** in the anchor data (0= Not important at all to 10= Very important). Similarly, the response format of the variable *cla8* has 10 stages (1= Very dissatisfied to 10=Very satisfied), while the corresponding variable in the anchor data *cla8* ranges from 0=Very dissatisfied to 10=Very satisfied.

Missing values

For all variables of the transition step-up data we defined a set of missing codes, which were applied throughout the data (table 10.1).

Table 10.1: Missing codes in data set *stepup*

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
-97	Have no experience up to now

Anonymity

Answers that might threaten respondents' anonymity were deleted or recoded as such in the data set. By means of anonymization, all string variables in the data set have been finally transformed to numeric variables. There are thus no string variables contained in the transitional step-up data set. Valid answers to open questions were recoded to value "1" throughout. The variables affected by anonymization are shown in table 10.2 along with the corresponding value labels (<information> mentioned) assigned to valid answers on the anonymized numeric variables.

Table 10.2: List of variables made anonymous

Variable	Variable label	Anonymous value label
rtr1p1n	Name partner 1	Name mentioned
rtr1p2n	Name partner 2	Name mentioned
rtr1p3n	Name partner 3	Name mentioned
rtr1p4n	Name partner 4	Name mentioned
mig1i12o	Country Citizenship	Citizenship mentioned
mig4o	Other country of birth	Country mentioned
cla4o	With whom did you live immediately after your birth?	Other mentioned
cla6e1o	Other: Lived with whom?	Other mentioned
cla6e2o	Other: Lived with whom?	Other mentioned
cla6e3o	Other: Lived with whom?	Other mentioned
cla6e4o	Other: Lived with whom?	Other mentioned

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 step-up codebook.

Generated variables and scales

There are no generated variables included in the step-up transition data. Generated variables for the anchor data of this group are provided in the anchor step-up data sets.

11. Outlook

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

Wave 7 will bring new data sets for anchor respondents and partners, as well as data sets from the parents survey, child interview, and parenting survey. Substantively, modules on respondents' current partnership, leisure activities, and intergenerational relationships will be part of the anchor interview in wave 7. Additionally, the siblings module, as well as questions on addictive behavior asked in wave 5, will be repeated for the first time in wave 7. Modules on parental leave and parental allowance ("Elterngeld"), as well as ambivalence measures and questions on health during childhood and current health status will be implemented for the first time in wave 7 of the pairfam project.

Of course, to reflect upon changes in 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 and decentralized as has been the case since the second wave.

The pairfam team would like to thank you for using this panel survey for your research. The aim of this project is to foster academic progress in the field of family research. The authors of this manual hope that you have 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

- 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., K. Hank, J. Huinink, B. Nauck, F. J. Neyer, S. Walper, P. Alt, P. Buhr, L. Castiglioni, C. Finn, K. Hajek, M. Herzig, B. Huyer-May, R. Lenke, B. Müller, T. Peter, V. Salzburger, C. Schmiedeberg, E. Schubach, P. Schütze, N. Schumann, C. Thönnissen, and B. Wilhelm (2015). The German Family Panel (pairfam). GESIS Data Archive, Cologne. ZA5678 Data file Version 6.0.0, DOI: 10.4232/pairfam.5678.6.0.0.
- Brüderl, J., B. Huyer-May, and C. Schmiedeberg (2013). Interviewer Behavior and the Quality of Social Network Data. In P. Winker, R. Port, and N. Menold (Eds.), *Interviewers' Deviations in Surveys. Impact, Reasons, Detection and Prevention*, pp. 147–160. Peter Lang, Frankfurt.
- 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.
- Brüderl, J., C. Schmiedeberg, L. Castiglioni, O. Arránz Becker, P. Buhr, D. Fuß, V. Ludwig, J. Schröder, and N. Schumann (2015). The German Family Panel - Study Design and Cumulated Field Report (Waves 1 to 6). Pairfam Technical Paper No. 01.
- Buhr, P. and J. Huinink (2014). Adjustment of variable frt6. Pairfam Technical Paper No. 04.
- 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.
- Mund, M., C. Finn, B. Hagemeyer, J. Zimmermann, and F. J. Neyer (2015). The dynamics of self-esteem in partner relationships. *European Journal of Personality*, in press.
- 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.
- Sonntag, M., M. Mund, E. Schubach, and F. J. Neyer (2014). Report of a method effect in the assessment of personality traits in pairfam. Technical report, Lehrstuhl für Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik, Friedrich-Schiller-Universität Jena, Jena.
- 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.
- Thönnissen, C., B. Wilhelm, S. Fiedrich, P. Alt, and S. Walper (2015). Scales Manual Waves 1 to 6. 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.

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)
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)

continued on next page

continued from previous page

Variable	Variable label
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)
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 2, 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)
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)

continued on next page

continued from previous page

Variable	Variable label
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)
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)

continued on next page

continued from previous page

Variable	Variable label
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 2, 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)
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 2, 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

Changes from release 1.0 to 2.0

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

Changes from release 2.0 to 3.0

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
continued on next page	

continued from previous page

Description of changes	Variables affected
Slightly modified due to corrections of data set <i>biopart</i> Value labels modified	ncoh, nmar sex_gen, psex_gen, k*sex_gen, dobm_gen, siops
<i>Modified tag and flag variables</i>	
Additional couples of anchor respondents identified Value label -3 recoded to 0 (theoretically possible) Value labels slightly modified Variable label modified	flag26 flag8, flag16 flag26, flag_frt6 flag_frt6, flag20, flag25
<i>Modified values of main variables</i>	
96 is set to 6; 97 is set to 7 Value labels of categories 1 & 0 interchanged	frt24i1-8 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	frt11v2i1-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

Changes from release 3.0 to 3.1

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, cpsr2, 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

Changes from release 3.1 to 4.0

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

continued on next page

continued from previous page

Description of changes	Variables affected
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 (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, parmage2, parfage, parfage2, parcasprim, parcassec, parflag_cas, parhhincnet, parhhsizemrd, parhhincgee, park1type, park2type, park3type, park4type, parlfs, parmarstat, parrelstat, 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

Changes from release 4.0 to 5.0

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>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>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	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
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

continued on next page

continued from previous page

Description of changes	Variables affected
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
<i>Modified generated variables</i>	
New baseline year 2010	cpi
Correction of coding variables of DemoDiff sample	smcob, sfcob
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	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	
<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 main variables</i>	
1 repositioned child	cid2, cid3, and all corresponding child variables *k2* and *k3*
<i>Modified variable names of main variables</i>	
New variable names original__sex original__dobm original__doby	sex, dobm, doby
<i>New generated variables</i>	
New weighting variables	d1weight, ca1weight, d1ca1weight
<i>Modified generated variables</i>	
New baseline year 2010	cpi
Correction of coding variables	smcob, sfcob, pnkidsbioalv, k*type
Recoding of missing values	isei
Recoding of missing values	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
Modifications of weighting variables	lweight, ppanel, pcontact, panswer
<i>Modified tag and flag variables</i>	
Additional couples of anchor respondents identified	flag17, flag26
<i>Modified variable names of main variables</i>	
New variable names original__sex original__dobm original__doby	sex, dobm, doby
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

Changes from release 5.0 to 6.0

Table A.20: Changes in data sets *anchor1*, *anchor1_DD*: from release 5.0 to release 6.0

Description of changes	Variables affected
Anchor data	
<i>New generated variables</i>	
New equivalence income variable and information on household composition necessary for equivalence scale weight	hhincoecd, npu14mr, npo14mr
continued on next page	

continued from previous page

Description of changes	Variables affected
Person number second biological parent of anchor's child(ren) were included from the data set biochild	parentidkx
<i>Modified generated variables</i>	
Minor corrections for different variables	yeduc, pyeduc, myeduc, fye- duc
Anchor DemoDiff data	
<i>New generated variables</i>	
New variable (ISCED classification of educational attainment, incl. students)	isced2
<i>Modified generated variables</i>	
Variables school & pschool changed due to corrections in anchor1_DD 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
Correction of coding variable	pnkidsbioalv

Table A.21: Changes in data set *anchor2, child2*: from release 5.0 to release 6.0

Description of changes	Variables affected
Anchor data	
<i>New generated variables</i>	
New equivalence income variable and information on household composition necessary for equivalence scale weight	hhincoecd, npu14mr, npo14mr
Person number second biological parent of anchor's child(ren) were included from the data set biochild	parentidkx
<i>Modified generated variables</i>	
Minor corrections for different variables	yeduc, pyeduc, myeduc, fye- duc
Child data	
<i>Modified variable names of main variables</i>	
Harmonization of SDQ item numbers with SDQ item numbers in parenting data set	csdq1i1-csdq1i3, csdq1i5- csdq1i20

Table A.22: Changes in data set *anchor3, child3, partner3*: from release 5.0 to release 6.0

Description of changes	Variables affected
Anchor data	
<i>Modified main variables</i>	
Corrections because cohabiting ex-partner was deleted in household grid	ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pXh1-ehc25pXh3

continued on next page

continued from previous page

Description of changes	Variables affected
<i>New generated variables</i>	
New equivalence income variable and information on household composition necessary for equivalence scale weight	hhincoecd, npu14mr, npo14mr
Person number second biological parent of anchor's child(ren) were included from the data set biochild	parentidkx
<i>Modified generated variables</i>	
Minor corrections for different variables	yeduc, pyeduc, myeduc, fyeduc
Corrections for variables regarding household size and household composition due to changes in household grid	pmrd, mmrd, fmr, childmrd, othmrd, hhcomp, hhsizemrd
Child data	
<i>Modified variable names of main variables</i>	
Harmonization of SDQ item numbers with SDQ item numbers in parenting data set	csdq1i1-csdq1i3, csdq1i5-csdq1i20
Partner data	
<i>Changes in case number</i>	
Deletion of one case because ex-partner falsely answered partner survey (no current partner in <i>anchor3</i> data set)	id=918071000

Table A.23: Changes in data set *anchor4*, *child4*: from release 5.0 to release 6.0

Description of changes	Variables affected
Anchor data	
<i>Modified main variables</i>	
Corrections because cohabiting ex-partner was deleted in household grid	ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pX
Corrections current partner lives in main residence	ehc26
<i>Modified variable names of main variables</i>	
<i>New generated variables</i>	
New equivalence income variable and information on household composition necessary for equivalence scale weight	hhincoecd, npu14mr, npo14mr
Person number second biological parent of anchor's child(ren) were included from the data set biochild	parentidkx
<i>Modified generated variables</i>	
Minor corrections for different variables	yeduc, pyeduc, myeduc, fyeduc
Minor corrections of weighting variables	lweight, ppanel, pcontact, panswer
Minor corrections for different variables regarding household size and household composition due to changes in household grid	pmrd, mmrd, fmr, childmrd, othmrd, hhcomp, hhsizemrd

continued on next page

continued from previous page

Description of changes	Variables affected
Child data	
<i>Modified variable names of main variables</i>	
Harmonization of SDQ item numbers with SDQ item numbers in parenting data set	csdq1i1-csdq1i3, csdq1i5-csdq1i20

Table A.24: Changes in data set *anchor5, child5*: from release 5.0 to release 6.0

Description of changes	Variables affected
Anchor data	
<i>Modified main variables</i>	
Minor filter corrections for different variables	crn33px, crn34px, crn35kx, crn36kx
Corrections because cohabiting ex-partner was deleted in household grid	ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pX
Corrections current partner lives in main residence	ehc26
<i>New generated variables</i>	
New equivalence income variable and information on household composition necessary for equivalence scale weight	hhincoecd, npu14mr,
Person number second biological parent of anchor's child(ren) were included from the data set biochild	npo14mr parentidkx
<i>Modified generated variables</i>	
Minor corrections for different variables	yeduc, pyeduc, myeduc, fyeduc
Minor corrections of weighting variables	lweight, ppanel, pcontact, panswer
Minor corrections for different variables regarding household size and household composition due to changes in household grid	pmrd, mmrd, fmrdr, childmrd, othmrd, hh-comp, hhsizemrd
Child data	
<i>Modified variable names of main variables</i>	
Harmonization of SDQ item numbers with SDQ item numbers in parenting data set	csdq1i1-csdq1i3, csdq1i5-csdq1i20
<i>Modified labels of main variables</i>	
English variable label slightly modified	crom9

A.3 Information on longitudinal weights

Table A.25: 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

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 -995 Euro income	1 - 995 Euro income	1 - 995 Euro income
pinc4	996 Euro or more income	996 - 1500 Euro income	996 - 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

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

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

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>hl1t</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	18 and 19 years		
age2	20 years		
age3 (ref.)	21 and 22 years		

continued on next page

continued from previous page

Independent variables and reference category	Definition		
	Cohort 1	Cohort 2	Cohort 3
age4		28 and 29 years	
age5		30 years	
age6 (ref.)		31 and 32 years	
age8			38 and 39 years
age9			40 years
age10 (ref.)			41 and 42 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.26: 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

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

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

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

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

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

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_frt	Survey method for questions frt7/8/9/16 (filter error)	0	General questionnaire	
		1	Follow-up questionnaire	Indicates if information on frt7/8/9/16 was collected via follow-up survey (PAPI); due to a filter error in the main questionnaire
		-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

continued on next page

continued from previous page

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

Table A.27: 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

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.28: 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 does not match anchor interview	
		2	Anchor's status 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

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

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	
		1021	Early retirement but full-time	

continued on next page

continued from previous page

Variable	Label	Values	Value Labels	Description
		1219	Unemployed but part-time	
		1721	Maternal or 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>