pairfam Data Manual

Release 3.1

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

University of Munich

with the cooperation of Petra Buhr, Stefan Fiedrich, Daniela Klaus, Nadia Lois, Veronika Salzburger,
Markus Schaer, Franziska Schmahl, Carolin Thönnissen, Barbara Wilhelm

December 15, 2012

Contents

1. Introduction

The aim of this manual is to facilitate work with pairfam data sets. The document provides information on the processing of the available wave 1, wave 2 and wave 3 data. If you have questions concerning the data that are not covered by this manual, please contact the pairfam user service at support@pairfam.de.

The manual is structured as follows. Chapter 2 provides an overview of the pairfam survey instruments and the resulting data sets, of person numbers identifying the survey respondents, and of the system of variable names. In subsequent chapters, the processing of each of the provided data sets is described in detail. In chapter 3, we first describe the data editing process for the anchor data of waves 1, 2, and 3. The subsequent sections 3.2 to 3.6 all deal with non-survey variables that have been added to the data set to improve the usability of the data. Constructs that are used frequently have been generated and added to the data set or are available as scales documented in syntax files. These variables are listed and described in section 3.2. Data generated from the original sample are described in section 3.3. Variables generated from the field process are listed in section 3.4. Section 3.5 deals with the computation of weights for the pairfam anchor data and gives advice on how to use these weights. Chapter 3.6 describes the computation of additional data sets, covering the anchors' partner and child relationships, their educational and occupational activities, and information on their dwellings and composition of households. We processed the original data and reformatted them to ease event history and panel data analysis. Chapters 4 to 7 explain the data editing as well as the computation of generated variables and scales for the alteri data sets. The manual closes with a brief outlook on future developments (chapter 8).

1.1 Obtaining the data

Pairfam data release 3.1 contains two data sets from wave 1 (*anchor1* and *partner1*), five data sets from wave 2 (*anchor2*, *partner2*, *parent2*, *child2*, and *parenting2*), five data sets from wave 3 (*anchor3*, *partner3*, *parent3*, *child3*, and *parenting3*), and four generated data sets (*biopart*, *biochild*, *bioact*, and *household*).[1] The files *anchor1*, *anchor2* and *anchor3* have been updated with data release 3.1.[2] Anchor and alteri data files come with English and German labels[3], whereas the generated data sets are labeled in English only.

The data are made available to registered data users only. To become a registered user, interested researchers should follow the instructions on the homepage http://www.pairfam.de. The data are distributed on CD-R via registered mail. Information on data errors detected after the release of the data will be posted on the website. Bug fixes will be available there as well. Data users should visit the homepage from time to time to stay abreast of problems with the data.

[1] We provide the data in Stata and SPSS formats.

[2] These new versions include the changes incorporated in the bug fix files published between release 2.0 and release 3.0 and additional changes of the data release 3.0.

[3] Stata 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 directory (\English or \German).

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 3.3). These data are not included in the scientific use file, however, and can only be analyzed on-site. The requirements and contact information for on-site analyses are posted on the pairfam website.

1.2 Referencing the pairfam project

Receiving credit from data users is of vital interest to the pairfam project. Only with these references we can prove the scientific value of pairfam. Therefore, we ask you to add the following two points to all your publications that are based on pairfam data:

1) For referencing the data and the concept paper please use the following (or a similar) phrase:

Analyses are based on data from the first three waves of the German Family Panel, release 3.1 (Nauck et al. 2012). A detailed description of the pairfam study can be found in Huinink et al. (2011).

The respective entries in the reference section are:

German version: *Nauck, Bernhard; Brüderl, Josef; Huinink, Johannes; Walper, Sabine (2012): Beziehungs- und Familienpanel (pairfam). GESIS Datenarchiv, Köln. ZA5678 Datenfile Version 3.1.0, doi: 10.4232/pairfam.5678.3.1.0*
English version: *Nauck, Bernhard; Brüderl, Josef; Huinink, Johannes; Walper, Sabine (2012): The German Family Panel (pairfam). GESIS Data Archive, Cologne. ZA5678 Data file Version 3.1.0, doi: 10.4232/pairfam.5678.3.1.0*

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

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

2) For acknowledging the project, please include the following (or a similar) phrase:

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

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

Brüderl, Josef; Hajek, Kristin; Huyer-May, Bernadette; Ludwig, Volker; Müller, Bettina; Müller, Ulrike; Passet, Jasmin; Pforr, Klaus; Scholten, Mirte; Schütze, Philipp; Schumann, Nina (2012): pairfam Data Manual. Release 3.1, University of Munich, Technical report.

1.3 Quick Start Files

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

The Quick Starts are available on the pairfam homepage in the "Data / Documentation" section. Currently, examples for getting started and analyzing the wave 1 data are available (see ZIP archive "Getting Started" on the homepage). In addition, we provide commented syntax files to introduce the most common operations for merging separate pairfam data sets from wave 1 and wave 2 to combined multi-waves and/or multi-actor data sets (see ZIP archive "Matching").

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

2. Overview of the data structure

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

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

2.1 Survey instruments and data sets

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

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

The other organizing principle of the data structure is storage by survey instrument, i.e., we produce separate data sets for the anchor, partner, parents, child, and parenting surveys respectively. Data set names consist of an (invariable) radical identifying the survey instrument and a suffix indicating the panel wave ($). The first-wave data release includes the two data sets *anchor1* and *partner1*. The release of wave 2 data contains the data sets: *anchor2*, *partner2*, *parent2*, *child2*, *parenting2*. Accordingly, the following data sets are published for wave 3: *anchor3*, *partner3*, *parent3*, *child3*, *parenting3*. With each release, previously published data sets are updated if necessary[3]

[1] For further information concerning the conceptual framework and design see the pairfam concept paper (Huinink et al. 2011) and the Pairfam Technical Paper No. 01 (Arránz Becker et al. 2012).

[2] There 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.

[3] See table A.1, which documents the changes in the data sets *anchor1* and *partner1* between release 1.0 and release 2.0, tables A.2 and A.4 for changes in the various data sets between release 2.0 and release 3.0, and tables A.3, A.5 and A.6 for changes between release 3.0 and release 3.1.

Table 2.1: Overview of survey instruments and resulting data sets, wave 1 and following

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	parid	parent$
Child survey	2009/10	CAPI	Anchor's child(ren) between 8 and 15 years, living with anchor	cid	child$

2.2 System of person identifiers

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

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

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

Table 2.2: Person identifiers of pairfam respondents

Respondent(s)	Person ID	(Range of) assigned IDs
Anchor	id	3-6 digit household identifier (hhid) $\cdot 1000$
Anchor's partner	pid	wave 1: id + 101
		wave 2 and following:
		id + 101 if partner is still wave 1 partner
		id + 102,...,150 for new partner(s)
Anchor's child(ren)	cid	id + 200,...,299
Anchor's parents	parid	id + 301 for biological (or adoptive) mother,
		id + 302 for biological (or adoptive) father,
		id + 303,305,...,399 for stepmother(s),
		id + 304,306,...,398 for stepfather(s)

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.

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

3. Anchor data

Processing of the anchor data is conducted by the Mannheim pairfam group and will be described in the next section.

3.1 Data editing

This section describes the editing of the anchor data from waves 1, 2, and 3.[1] The main steps in processing the data are:

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

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

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

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

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

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 3.1). Missing values "-1 Don't know", respectively "-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 instead of -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.

Table 3.1: 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

Preload variables for Dependent Interviewing - DI-variables

In order to get more reliable responses in the CAPI of the second and third wave, 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. In wave 2, 270 variables were preloaded to the interview. The preload information of wave 3 is based on wave 2 and wave 1 information depending on the participation in wave 2. In wave 3, 236 preload variables were used.

The preloads or DI-variables are contained in the data sets *anchor2* and *anchor3*. They can be found at the beginning of the data sets. The variables are prefixed by a d- followed by a 1- to 3-digit number.

In the data set *anchor2* the variable names range from *d1* to *d398*. The preload variables of wave 3 extend to *d503*. These variables are all made anonymous if necessary (strings and information on days). Please note that (unlike the other variables in the *anchor$* data sets) these variables did not undergo the usual steps of data editing (e.g., the variables may contain system missings). Thus, the DI-variables should not be used for data analysis. A complete list of all DI-variables is contained in the anchor codebooks of wave 2 and wave 3 in the section "Preface".

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

The Event History Calendar used in wave 2 and wave 3 gathered 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≠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 but not considered problematic.

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 wave 1 to wave 3 were corrected by relocating the children (and all *vark**) falsely positioned in wave 2 and wave 3 according to their original position in wave 1 and wave 2, respectively. 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) 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 interviews[2]. 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 previoius wave seemed to be implausible, they were deleted. Regarding all corrections, the auxiliary variables *hp** which are part of *anchor$* as of wave 3 were adapted accordingly.

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

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

In wave 2 and wave 3, respondents were presented 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 for 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.

While processing the resulting EHC variables of wave 2, we encountered several problems with these data. In wave 1 some interviews apparently contained information on all residences instead of only the current residence. We nonethelesse preloaded this information since we could in most cases not

[2]The 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.8).

[3]Specifically, 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.

decide whether the information actually was wrong. 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.

In wave 3 the data for information 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. We also used information from the retrospective questions (*rtr**), if available, to replace gaps.

For release 2.0 we recommend not using 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. We still plan to release a consistent episode data set combining retrospective and prospective information on mobility in the future. For more detailed analyses of respondents' mobility over time we recommend using this generated data set.

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. Partly it is also used 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 3, 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 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 occured twice but could reasonably only occur once were set to "-5" or were deleted.

Coding open answers

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

Anonymity

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

By these means, all string variables in the data set were finally transformed to numeric variables. Thus the *anchor$* data sets contain no string variables. Valid answers to open questions were recoded to value 1 throughout. The variables affected by the procedures to ensure anonymity are shown in table 3.2 along with the value labels.

Table 3.2: 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
crn12k1i14o,..., crn12k7i14o & crn12k9i14o	Child x: Other, namely:	Other mentioned	2
crn13k1i13o,..., crn13k6i13o	Child care morning child x: Other, own entry	Other mentioned	2 & 3
crn14k1i13o,..., crn14k6i13o	Child care afternoon child x: Other, own entry	Other mentioned	2 & 3
crn18k1o & crn18k2o	Child x: Custody - other, namely:	Other mentioned	2 & 3
crn1k2o	Other place of birth child 2	Other mentioned	2 & 3
d1	Preload: Day of birth (Preload)	Day mentioned	2 & 3
d134,..., d139	City xst place of residence (Preload)	Residence mentioned	2 & 3
d14,..., d23	Name child x (Preload)	Name mentioned	2 & 3
d164,..., d167	Country xst place of residence (Preload)	Country mentioned	2 & 3

continued on next page

[4] For 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 3.2).

continued from previous page

Variable	Variable label	Anonymous value label	Wave
d176,..., d194	Name xst household member [main residence] (Preload)	Name mentioned	2 & 3
d218,..., d226,	Name xst household member [second residence] (Preload)	Name mentioned	2
d267	Other education (Preload)	Yes	2
d274	Other type of employment (Preload)	Yes	2
d282	Current employment activity: open-ended answer (Preload)	Occupation mentioned	2 & 3
d398	Current partner's day of birth (Preload)	Day mentioned	2 & 3
d59,..., d73	Day of birth child x (Preload)	Day mentioned	2 & 3
d8	Name of current partner (Preload)	Name mentioned	2 & 3
dobd	Day of birth	Day mentioned	all
ehc12k1o,..., ehc12k7o	Other parent name child x (EHC)	Name mentioned	2 & 3
ehc14p1i1,..., ehc14p10i1	City place of residence x (EHC)	Residence mentioned	2
ehc14p1i2o,..., ehc14p5i2o	Country place of residence x (EHC)	Country mentioned	2
ehc19i16m1o,..., ehc19i16m32o	Other type of employment, open entry in month x (EHC)	Other mentioned	2 & 3
ehc19i16o	Other type of employment, open entry currently (EHC)	Other mentioned	2 & 3
ehc19i22m1o,..., ehc19i22m32o	Other type of unemployment, open entry in month x (EHC)	Other mentioned	2 & 3
ehc19i22o	Other type of unemployment open entry currently (EHC)	Other mentioned	2 & 3
ehc19i9m1o,..., ehc19i9m32o	Other education open entry in month x (EHC)	Other mentioned	2 & 3
ehc19i9o	Other education open entry currently (EHC)	Other type of training mentioned	2 & 3
ehc1p1n,..., ehc1p5n	Name partner x (EHC)	Name mentioned	2 & 3
ehc20d	Day of birth of partner from previous wave (EHC)	Day mentioned	2 & 3
ehc22p1n,..., ehc22p13n	Name person x in houshold (EHC)	Name mentioned	3
ehc27p1i1,..., ehc27p10i1	City place of residence x (EHC)	Residence mentioned	3
ehc27p1i2o,..., ehc27p5i2o	Country place of residence x (EHC)	Country mentioned	3
ehc7k1n,..., ehc7k10n	Name child x (EHC)	Name mentioned	2 & 3
ehc8k1d,..., ehc8k10d	Day of birth child x (EHC)	Day mentioned	2 & 3
frt13i14o	Reasons against child: Other Reason	Reason mentioned/ Other mentioned	1 & 2

continued on next page

continued from previous page

Variable	Variable label	Anonymous value label	Wave
hc1p1i1,..., hc1p7i1	Information x-th residence	Residence mentioned	1
hc1p1i2o,..., hc1p4i2o	Country x-th residence	Country mentioned	1
hc8h1p1,..., hc8h1p16 / hc8h1p1,..., hc8h1p12	Main residence: Name person x	Name mentioned	1 & 2
hc8h2p1,..., hc8h2p9 / hc8h2p1,..., hc8h2p7	Second Residence: Name person x	Name mentioned	1 & 2
hcp1i1	Main residence partner	1 Residence mentioned, 7 Same residence as I live	2 & 3
hcp1i1	Main residence partner	Residence mentioned	3
hcp1i2o	Country main residence partner	Country mentioned	2 & 3
hcp6i1	Second residence partner	Residence mentioned	3
hcp6i1o	Country second residence partner	Country mentioned	3
igr1d	Day of birth biological father	Day mentioned	1
igr21d	Day of birth adoptive mother	Day mentioned	2
igr24d	Day of birth adoptive father	Day mentioned	2
igr2d	Day of birth biological mother	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
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
igr83o	Open entry country of birth mother's partner	Other mentioned	3
igr85i12o	Citizenship father's partner: Open entry other country	Other mentioned	3

continued on next page

continued from previous page

Variable	Variable label	Anonymous value label	Wave
igr86o	Open entry country of birth father's partner	Other mentioned	3
int12o	Open entry suggestions or comments on the interview	Mentioned	3
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
netp1n,..., netp30n	Name person x: Complete list name generator	Name mentioned	2
rtr18k1d,..., rtr18k4d	Day of death child x	Day mentioned	1
rtr1p1n,..., rtr1p13n	Name partner x	Name mentioned	1
rtr23h1,..., rtr23h12	Residence x	Residence mentioned	3
rtr24h1o,..., rtr24h8o	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
sd14k1n,..., sd14k11n	Name child x	Name mentioned	1
sd17k1,..., sd17k6	Name other parent child x	Name mentioned	1
sd19k1d,..., sd19k10d	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 & 3
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

continued on next page

continued from previous page

Variable	Variable label	Anonymous value label	Wave
sdp17o	Open entry other level of school education	Other mentioned	3
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
sep5k1o	Other custody of child 1 before separation	Other mentioned	2
sep7k1o,..., sep7k3o	Other: Current custody of child 1-3	Other mentioned	2 & 3

Value checks

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

Filter checks

We checked the filter of every variable as documented in the CAPI questionnaire and assigned the missing code -4 for data errors. On the one hand, a question may have been asked by mistake, or not asked by mistake if the filter was not implemented correctly in the CAPI program. These cases were both regarded as filter errors and the variable then set to value -4.[5] On the other hand, values were classified as data errors if they were incorrectly entered by the interviewer. In these cases, we also assigned value -4. In checking the filters, we proceeded in the order in which questions were posed in the interview, to make sure that ensuing filter errors were detected.

Checks for consistency

As mentioned above, we conducted various checks to identify logically impossible or empirically implausible combinations of values on two or more variables.[6] In some cases it was not possible to resolve the inconsistency by assigning the code -5, because we could not decide which of the variables under consideration was wrong. Consequently, the provided values were left unedited. Instead, a flag variable indicating the respective inconsistency was generated. Furthermore, some flags were produced to explain why the value -5 was assigned to the respective variable (see *flag12*). For each of these variables, code 0 indicates that the respective inconsistency is non-existent and code 1 that there is an inconsistency. The code -3 indicates that the flag variable does not apply to the respective wave. We performed a number of additional checks that did not provide an indication of any inconsistency. Table A.8 describes the flag variables in detail.

[5] If 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 if 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.

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

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.9 lists all of the generated tag variables that are part of the data sets *anchor2* and *anchor3*.

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 3.2 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 is still not solved. Some respondents still reported the total number of children they were thinking of having, still including those already born or conceived.

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

1. Consider using the question on the intention to have a child within the next two years (*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.

In wave 3 we constructed the questionnaire in the way that *ftr6* was divided into three questions. In question *ftr26* respondents without children were asked how many children they will have realistically. Furthermore, to rule out wrong answers a filter variable was integrated (*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 (*ftr28*). Therefore *flag_frt6* was not computed for wave 3.

Remark on variable flag_igb (wave 2)
In 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 version of wave 2 and wave 3 anchor data used to compute the preloads for the next wave. By following the procedures we used to generate these data, users will be able to reconstruct the process.[7] Please note that some generated variables and scales have been modified for release 2.0, release 3.0 and release 3.1. The changes of the generated variables are documented in the appendix, see tables A.1 to A.6. The details on generated variables and scales are given in chapter 3.2, those on processing the generated files *biopart*, *biochild*, *bioact* and *household* in chapter 3.6.

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.

[7] Users should, however, start with the latest version of the wave 1 and wave 2 data (release 3.1) when running the do-files for the respective wave, since there have been minor changes in the data after release 1.0, release 2.0 and release 3.0. These changes are documented in the appendix, see tables A.1 to A.6.

3.2 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 3.3 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 on the release CD and on the project homepage http://www.pairfam.de for each wave separately. Users are invited to adapt the syntax to their special research needs.[8] Before using any of the generated variables we strongly advice users to always check whether the respective generating procedures meet their specific needs. Please note that there is a common missing value "-7 Incomplete data" encompassing the original missing codes -1, -2, -4, -5, and -6.

Furthermore, table 3.16 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 on CD and on the pairfam website. For additional information regarding these scales, please refer to the scales manual (Schmahl et al. 2012).

Table 3.3: List of generated variables included in data set *anchor$*

Construct	Variable name
Generated identifiers sex (anchor, partner, children)	sex_gen, psex_gen, k*sex_gen
Generated identifiers date of birth (anchor, partner, children, mother incl. adoptive mother, father incl. adoptive father, stepmother, stepfather)	dob*_gen, pdob*_gen, k*dob*_gen, mdob*_gen, fdob*_gen, smdob*_gen, sfdob*_gen
Age (anchor, partner, mother incl. adoptive mother, father incl. adoptive father, stepmother, stepfather)	age, page, mage, fage, smage, sfage
Age of children	k1age,..., k10age
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)	nat1, nat2, pnat1, pnat2, pnat3, mnat1, mnat2, fnat1, fnat2, amnat1, amnat2, afnat1, afnat2, smnat1, smnat2, sfnat1, sfnat2
Ethnicity	ethni
Migration status	migstatus
Relationship status	relstat
Marital status (anchor, partner)	marstat, pmarstat
Number of previous partners	np
Number of previous partners with whom anchor cohabitated	ncoh

continued on next page

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

continued from previous page

Construct	Variable name
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
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)	school, pschool
Highest vocational degree attained at time of interview (anchor, partner)	vocat, pvocat
ISCED-97, International Standard Classification of Education, no students (anchor, partner)	isced, pisced
ISCED-97, International Standard Classification of Education, including students (anchor, partner)	isced2, pisced2
CASMIN classification of educational attainment (1999) (anchor, partner)	casmin, pcasmin
Years of schooling / vocational qualification (anchor, partner)	yeduc, pyeduc
KldB classification of occupation	kldb
ISCO classification of occupation	isco
Erikson-Goldthorpe-Portocarero class schema (EGP)	egp
Int. Socio-Economic Index of Occupational Status (ISEI)	isei
Standard Int. Occupational Prestige Scale (SIOPS)	siops
MPS occupational prestige score	mps
Net equivalence income according to GCEE	hhincgcee
Net income (open and estimated information combined) (personal, household)	incnet, hhincnet

continued on next page

continued from previous page

Construct	Variable name
Consumer price index acc. to Nat. stat. off. (baseline 2005)	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
Sex-Ratio	sexratio
Population density	popdens

Generated identifiers sex - sex_gen, psex_gen, k*sex_gen

To solve the problem of asynchronous information on the respondents' sex between the different informational sources[9], we generated "best solution" sex variables for the anchor, the partner, and the children (*sex_gen, psex_gen, k*sex_gen*).

This was done according to the following rules: In general, self-reported sex information was preferred over proxy information. For the self-reported information the most frequently or (if not existent) the newest reported value was preferred. In case of proxy information the newest information was chosen before an older one. Please note that the rules differ from the previous release. If the partner or a child was nonexistent, we assigned the code "-3 Does not apply". If the information was not available to derive the sex, we used the code "-7 Incomplete data". Finally the *_gen Variables were synchronized over the waves. This means, a specific person has the same (best) gender information in all waves. The Stata do-file identifiers.do, which can be found in the syntax folder of the latest wave, contains the syntax used to compute the variables.

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

In order to cope with the problem of asynchronous information on the date of birth between different informational sources[10], we generated best solution variables for the month and the year of birth of the anchor, the partner, the parents (incl. the adoptive parents), the stepparents, and the children (*dob*_gen, pdob*_gen, mdob*_gen, fdob*_gen, smdob*_gen, sfdob*_gen, k*dob*_gen*).

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

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

The variables *age, page, mage, fage, smage, sfage* and *k*age* contain the anchor's, the partner's, the parents' (incl. the adoptive parents'), the stepparents', and the children's age, respectively. These variables were calculated on the basis of the generated date of birth variables (see above). The age values were calculated by subtracting the corresponding generated year of birth from the anchor's 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. 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. Please note that for wave 1, release 2.0 a modified version of this do-file is available (see A.1).

[9] The various data sets of the anchor and the alteri, partially available for different waves.
[10] See footnote above.

Birth cohort - cohort

The variable *cohort* indicates to which of the three birth cohorts 1971-1973, 1981-1983 and 1991-1993 the anchor belongs. The information is derived from both the stated date of birth and the date of birth given by the register data. In cases of a contradiction, *Infratest* recontacted the respondents for clarification. The methods report of wave 1 (Suckow and Schneekloth 2009) states that the gross sample was drawn from the birth cohorts 1971-1973, 1981-1983 and 1991-1993. Due to reporting practices of some municipalities, however, a few respondents in directly adjacent birth cohorts entered the gross sample and also remain in the net sample. These respondents were assigned to the corresponding birth cohorts, e.g., a respondent with year of birth 1970 was assigned to the birth cohort 1971-1973. Since the birth cohort is a time-constant variable it is fixed throughout the waves. The variable was generated by running the do-file cohort.do. Please note that for wave 1, a modified version of this do-file is available (release 2.0, see A.1).

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/father). 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 3.4 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/father 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 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 3 to compute *pcob*. If there is insufficient information in wave 3, information of the previous wave is used, if available. Due to privacy concerns, the syntax which generates these variables cannot be published. Please note that for wave 1, release 2.0 and wave 2, release 3.0 the variables were modified (see tables A.1 and A.4).

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/father). 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 and stepmother/father is reported. Up to three nationalities are stored in variables **nat1*, **nat2* and **nat3*. 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 3.4 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. For wave 1, release 2.0 and wave 2, release 3.0 the variables were slightly modified (see tables A.1 and A.4).

Table 3.4: Values and labels of variables **cob* and **nat*

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

[a] only for country of birth

[b] "Other states" for nationality

[c] only for nationality

Ethnicity and migration status of anchor - ethni, migstatus

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

garded 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. Please note that for wave 2, release 3.0 there is a modified version of this do-file available (see A.4).

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 same previous partner only once although there has been a break or another partner in between. That's 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*, which has been corrected for inconsistencies and is organized in "long" format as of wave 2 (for further information see chapter 3.6). The do-file npart.do includes the syntax to compute these three variables. Please note that for wave 1, release 2.0 and release 3.0 and for wave 2, release 3.0 modified versions of this do-file are available (see A.1, A.2 and A.4).

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 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 above and chapter 3.6). For waves 1 and 2, release 3.0 there are modified versions of this do-file available (see A.2 and A.4).

Anchor homosexual - homosex

The variable *homosex* contains information on whether or not the anchor is homosexual. The information was taken from the anchor interview, combining information of all three waves. The gender information was taken from the generated gender variables (*sex_gen* and *psex_gen*).The anchor was defined as being homosexual if he or she had a same-sex partner and heterosexual if he or she had a

[11] If *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)

partner of the other sex in wave 3. If the anchor had no partner in wave 3 the gender of previous partners was used. If no information concerning a sexual preference was available in wave 3, information from wave 1 and wave 2 (anchor's reported homosexual/heterosexual preference or gender of partner) was added. Where only insufficient information (of neither wave) was available to derive the status, the code "-7 Incomplete data" was given. The syntax used to compute the variable is contained in the Stata do-file homosex.do. Please note that for wave 1, release 2.0 a modified version of this do-file is available (see A.1).

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 3 and if the wave 1 or wave 2 information indicated that the anchor or the couple unit was seen to be infertile (only for same couples as in wave 1 or 2 respectively), then the code for infertility status was given.

In three cases, the code "-7 Incomplete data" was assigned. First, value "-7" 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 also contained in the wave-specific Stata do-file infertile.do. While in wave 1 and 3 the anchor as well as the partner were asked about their fertility status or if they use sterilization for contraception, in wave 2 only the anchor gave information about his and the partner's fertility status. Therefore the wave-specific do-files differ respectively. Please note that for wave 1, release 2.0 and 3.0 and wave 2, release 3.0 modified versions of this do-file are available (see A.1, A.2 and A.4).

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. Please note that for wave 1, release 2.0 a modified version of this do-file is available (see A.1).

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 3.5. 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 3.5: Number of children - *nkids**

Variable	Label
nkids	Number of all kids born until time of interview
nkidsbio	Number of all biological kids born until time of interview
nkidsp	Number of all biological kids with partner born until time of interview
nkidsalv	Number of all kids alive
nkidsbioalv	Number of all biological kids alive
nkidspalv	Number of all biological kids with partner alive
nkidsliv	Number of all kids living with anchor
nkidsbioliv	Number of all biological kids living with anchor
nkidspliv	Number of all biological kids with partner living with anchor
pnkidsbioalv	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 3.6. The do-file ktype.do generates the variables for each wave. Please note that for wave 1, release 2.0 a modified version of this do-file is available (see A.1).

Table 3.6: 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 (see do-file household.do which also generates the data set *household*).

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 generated panel data set *household* which combines information on household members given in waves 1 and 2. The variable *hhsize2rd* (household size of the second residence) has been discontinued with pairfam data release 2.0 (see section 3.6). The syntax used to compute the variables is contained in Stata do-file household.do. Do-file hhsize.do which formerly contained the coding is now only used to merge the variables to the anchor data.

Parents living at main residence - mmrd, fmrd

The variables *mmrd* and *fmrd* indicate whether the anchor's mother and father live at the anchor's main residence. The information was derived from the household grid stored in panel data set *household* (see section 3.6). Biological, adoptive, step- and foster parents were counted as parents. The syntax used to compute the variables is contained in Stata do-file household.do. Do-file hhsize.do which formerly contained the coding is now only used to merge the variables to the anchor data.

Other household members at main residence - othmrd

The variable *othmrd* indicates how many household members live at the anchor's main residence. Other household members are all persons other than partner, children, and parents. The information was derived from the household grid stored in panel data set *household* (see section 3.6). The syntax used to compute the variables is contained in Stata do-file household.do. Do-file hhsize.do which formerly contained the coding is now only used to merge the variable to the anchor data.

Children living at main residence - childmrd

The variable *childmrd* indicates how many children lived at the anchor's main residence. Only cohabiting children of the anchor are included, i.e. biological, adopted, step- and foster children. The information was derived from the household grid stored in panel data set *household* (see section 3.6). The syntax used to compute the variables is contained in Stata do-file household.do. Do-file hhsize.do, which formerly contained the coding, is now only used to merge the variable to the anchor data.

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 (see section 3.6 for a list of the values). The syntax used to compute the variables is contained in Stata do-file household.do. Do-file hhcomp.do, which formerly contained the coding, is now only used to merge the variable to the anchor data.

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 3.7).

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 and wave 3 by the partner in the partner survey, this information was used instead. In wave 2, only proxy information given by the anchor was available and was therefore used to construct the variable *penrol*.

Table 3.7: 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

Note that information on partners' enrollment deviates from anchors' 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 qualification" as a default. If the partner gave other, more detailed information, this information was used instead.

In wave 2 and wave 3, 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 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 both variables is contained in the Stata do-file education.do, which is provided for each wave. Please note that we modified the variables for waves 1 and 2, release 3.0 (see A.2 and A.4).

Highest school degree - school, pschool

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

pschool.

Table 3.8: Values and labels of variables *school, pschool*

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

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 and wave 3, 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.

Highest vocational degree - vocat, pvocat

The highest vocational degree completed by the anchor and his or her partner is stored in variables *vocat* and *pvocat*, respectively. In wave 2 and wave 3 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" only applies for anchors without partners. The variables were generated by running the wave-specific do-file education.do.

Table 3.9: Values and labels of variables *vocat, pvocat*

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, isced2, pisced2

The generated variables *isced* and *pisced* carry information on the anchors' and partners' 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 *isced* and *pisced* provided the basis for two additional variables (*isced2*, *pisced2*), where afterwards respondents and partners currently enrolled in the German educational system were included. Thus it was assumed that those currently enrolled would complete their education and attain the corresponding degree. This can be helpful, since the percentage of those currently enrolled was at approximately 30 percent. Values and labels of the resulting variables are shown in table 3.10. The syntax used to compute all four variables is contained in the Stata do-file education.do (available for each wave). Please note that we modified the variables of wave 1, release 2.0 (see A.1).

Table 3.10: Values and labels of variables *isced, pisced*

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

Variables *casmin* and *pcasmin* contain the 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). The complete classification schema - including additional values "-7 Incomplete data" and "0 Currently enrolled" - is shown in table 3.11. The category "-3" only applies for anchors without partners. The syntax used to compute both variables is contained in the Stata do-file education.do (one for each wave).

Table 3.11: Values and labels of variables *casmin, pcasmin*

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

Variables *yeduc* and *pyeduc* are linear measures of the anchor's and his or her partner's 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* and *pyeduc* ranged 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-file education.do (available for each wave).

KldB classification of occupation - kldb

Variable *kldb* contains information according to the 4-digit 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*, which originally carried open answers about the anchor's current occupation. (The original answers have been made anonymous as described in section 3.1).

Variable *flag_isco_kldb*[12], available for wave 1, indicated 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" flags respondents where the most frequent of all possible occupations has been assigned. Value "2" indicates that the occupation requiring the lowest level of qualification has been assigned. Recoding was done by *TNS Infratest* (see Hartmann et al. (2010), Hartmann et al. (2011a) and Hartmann et al. (2011b) for more details).

ISCO classification of occupation - isco

Variable *isco* contains the anchor's occupation classified according to the 4-digit International Standard Classification of Occupations (ISCO) schema. The variable was derived from variable *kldb* (see above). Recoding was done by *TNS Infratest* (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).

When using 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 on CD 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 specified by Ganzeboom and Treiman (1996). The assignment of values is documented in do-file isei.do.

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

[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.

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 (1996). The assignment of scores is documented in do-file siops.do which was slightly modified for wave 2, release 3.0 (see table A.4).

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 *kldb* instead of *isco*. The procedure is described in Frietsch and Wirth (2001). The assignment of values is documented in do-file mps.do.

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

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.

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 the 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 for wave 2. For the household net income the value 5,000 Euros was assigned at the higher end for both waves. The codings for wave 1, wave 2 and wave 3 are shown in table 3.12 and table 3.13. 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 also contained in Stata do-files incnet.do. Please note that for wave 1, release 2.0 a modified version of this do-file is available (see table A.1).

Table 3.12: 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
$(2000, \infty)$	2500	$(3500, \infty)$	5000

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

Table 3.13: Coding scheme for the categorized personal and household net income (wave 2 & 3)

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	$[1000, 2000)$	1125
$[1250, 1500)$	1375	$[1250, 2300)$	1375
$[1500, 1750)$	1625	$[1500, 2800)$	1625
$[1750, 2000)$	1875	$[1750, 3500)$	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
$[4500, \infty)$	5000	$(4500, \infty)$	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 with the baseline year 2005. 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.

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 and wave 3. In wave 2, 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 the anchor this reduction was accomplished by a series of reductions that follow specific dominance rules. The same rules were used for partners.

The following dominance rules were applied:

- If only one activity was stated, it was defined as the primary activity status
- If more than one activity and
  - only employment activities were stated:
    primary activity was defined according to the dominance rule:
    vocat. training > full-time empl. > self-employ. > part-time empl. > marginal empl. > internship > occasionally empl. > other type of job
  - only educational activities were stated[16]:
    primary activity is defined according to the dominance rule:

[15] For partners, we only know if they are involved in general schooling, vocational training, or university.
[16] Due to filtering and question wording partners cannot have multiple education activities.

general secondary school > UCE[17] > college/university > evening school > technical/professional school > pre-vocat. training > vocat. retraining > other education

- only unemployment activities were stated:
  primary activity was defined according to the dominance rule:
  maternity/paternity leave > military service > unemployed > retired > housewife > other, non employed
- both employment and unemployment activities were stated:
  primary activity was defined according to the dominance rule:
  maternity/paternity leave > military service > unemployed > retired > vocat. training > full-time employ. > self-employ. > part-time empl. > marginal empl. > internship > housewife > occasionally empl.
- both unemployment and educational activities were stated:
  primary activity was defined according to the dominance rule:
  maternity/paternity leave > military service > unemployed > retired > general secondary school > UCE > college/university > evening school > technical/professional school > pre-vocat. training > vocat. retraining > housewife
- both employment and educational activities were stated:
  primary activity was defined according to the dominance rule:
  general secondary school > UCE > college/university > evening school > technical/professional school > pre-vocat. training > vocat. retraining > vocat. training > full-time empl. > self-empl. > part-time empl. > marginal empl. > internship > occasionally empl.
- employment, unemployment, and educational activities were stated:
  primary activity was defined according to the dominance rule:
  maternity/paternity leave > military service > unemployed > retired > general secondary school > UCE > college/university > evening school > technical/professional school > pre-vocat. training > vocat. retraining > vocat. training > full-time empl. > self-empl. > part-time empl. > marginal empl. > internship > housewife > occasional empl.

- If more than one activity was stated, the secondary activity was defined according to the same dominance rules and one additional dominance rule:
  if only activities out of "Other education", "Other type of job" and "Other, not employed":
  "Other, not employed" > "Other education" > "Other type of job"

- If at least one employment activity was stated, but neither primary nor secondary activity status, case-by-case decision under consideration of occupation, occupational status, work hours, educational attainment, personal income, and age[18].

After this reduction process, the defined activity statuses were checked for inconsistencies. In table 3.14 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.

[17] University of Cooperative Education ("Berufsakademie")
[18] For partners only occupational status, educational attainment, work hours, and age are considered.

Table 3.14: Inconsistent activity status combinations

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

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

The variables *lfs* and *plfs* contain the anchor's and his or her partner's labor force status. The variables were directly derived from the activity statuses (s.a.) in the following way:
Employment activities are defined as one of the following activities:

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

If only the primary or only the secondary activity status was an employment activity, this was taken for the labor force status. If both activity statuses were employment activities, the primary activity status was used. If both activity statuses were unemployment activities, the primary activity status was used. This derived single activity status was recoded to the labor status. The coding scheme is shown in table 3.15.

Table 3.15: Coding scheme for labor force status

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

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

If there was no partner, the partner's labor force status was coded as "-3". If no information was available to code a status, it was coded as "-7". The syntax used to compute the variables is also contained in Stata do-file lfs.do (available for each wave). Please note that for wave 1, a modified version of this do-file is available (release 2.0, also see table A.1). Furthermore, researchers also need the Stata do-file cas.do, as the labor force status is based on the primary and secondary activity statuses.

Currently living in Eastern Germany - east

To differentiate whether the anchor was currently living in Eastern or Western Germany, we generated the dummy variable *east*. Value "1" indicates that the anchor was living in Eastern Germany, while "0" means that the anchor resided in Western Germany. Note that the computation of this variable is based on the process-generated variable *bula* (see section 3.3). The relevant do-file east.do is available for each wave.

Sex ratio and population density - sexratio, popdens

The variables *sexratio* and *popdens* contain the categorized sex ratios and population density at the municipal level for each anchor. The information was derived from the process-generated 8-digit municipality key variable *gkz* (see chapter 3.3) provided by *TNS Infratest* and the municipality register (*Gemeindeverzeichnis*) of the German National Statistical Office as of 12/31/2008. The municipality keys provided by *Infratest* were updated with the municipality change list (*Gebietsänderungsliste*) of the German National Statistical Office. Due to data privacy regulations the subsequent reconstruction of the anchor's municipality key variable had to be avoided. To guarantee this, the exact figures were grouped to 5%-percentile categories.

Scales

All of the previously described generated variables are included in the anchor data sets. In contrast, scale variables (see table 3.16) 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 on CD and on the pairfam website. For further information, refer to the scales handbook (Schmahl et al. 2012), which contains a detailed description of all scales.

Table 3.16: List of scales included in data set *anchor$*

Construct	Variable name	Wave
Traditional concept of marriage	tradmarr	1 & 3
Value of Partnership - Negative expectations	vopneg/vopneg2	1 / 3
Value of Children: Benefit of stimulation	vocbstim	1 & 2
Value of Children: Costs of comfort	vocccomf	1 & 2
Single: Interest in partnership	partint	all
Single: Desire for partnership	partdes	all
Single: Broad exploration	siexplbr/siexplbr2	1 & 2 / 3
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
Ambivalence: marriage	ambmarr	1 & 3
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	all
Partnership: Anxiety about loss of love	lovewitanx_apd	all
Partnership: Anxiety about being absorbed	enganx_apd	all
Partnership: Autonomy	indep_apd	all
Partnership: Feelings of competence in the partnership	comppart/comppart2	1 / 2 & 3
Partnership: Own partnership satisfaction (global scale)	satpart	1 & 3
Partnership: Instability of partnership	instab_apd	all
Partnership: Readiness to sacrifice Self	sacrif_aps	1 & 3
Partnership: Hostile attribution Self	hostattr_aps	1 & 3
Partnership: Future orientation	comfut_apd	all
Partnership: Tolerance of conflicts	comctol_apd	all
Partnership: Orientation of reciprocity Self	reciproc_aps	1 & 3
Partnership: Frequency of manifest conflicts	confldom_apd	1
	confldom2_apd	3
Partnership: Verbal aggression Partner	verbaggr_apo	all
Partnership: Verbal aggression Self	verbaggr_aps	all
Partnership: Constructive behavior Partner	constrbh_apo	all
Partnership: Constructive behavior Self	constrbh_aps	all
Partnership: Withdrawal Partner	withdraw_apo	all
Partnership: Withdrawal Self	withdraw_aps	all

continued on next page

continued from previous page

Construct	Variable name	Wave
Partnership: Manipulation Partner	manipul_apo	1 & 3
Partnership: Manipulation Self	manipul_aps	1 & 3
Partnership: Dyadic coping Partner	dycop_apo	1 & 3
Partnership: Dyadic coping Self	dycop_aps	1 & 3
Intergenerational Relationships: Conflict with mother	confl_amd	all
Intergenerational Relationships: Conflict with partner of mother	confl_asfd	2 & 3
Intergenerational Relationships: Conflict with father	confl_afd	all
Intergenerational Relationships: Conflict with partner of father	confl_asmd	2 & 3
Intergenerational Relationships: Intimacy mother	intim_ams	all
Intergenerational Relationships: Intimacy partner of mother	intim_asfs	2 & 3
Intergenerational Relationships: Intimacy father	intim_afs	all
Intergenerational Relationships: Intimacy partner of father	intim_asms	2 & 3
Intergenerational Relationships: Admiration mother	admir_amo	2
Intergenerational Relationships: Admiration partner of mother	admir_asfo	2
Intergenerational Relationships: Admiration father	admir_afo	2
Intergenerational Relationships: Admiration partner of father	admir_asmo	2
Intergenerational Relationships: Dominance mother	domin_amo	2
Intergenerational Relationships: Dominance partner of mother	domin_asfo	2
Intergenerational Relationships: Dominance father	domin_afo	2
Intergenerational Relationships: Dominance partner of father	domin_asmo	2
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	sucindivid_amd	2
Intergenerational Relationships cohort1: Successful individuation Father/Partner of mother	sucindivid_afd	2
Intergenerational Relationships cohort1: Fear of love withdrawal Mother/Partner of father	lovewitanx_amd	2
Intergenerational Relationships cohort1: Fear of love withdrawal Father/Partner of mother	lovewitanx_afd	2
Intergenerational Relationships cohort1: Ambivalence Mother/Partner of father	ambiv_amd	2
Intergenerational Relationships cohort1: Ambivalence Father/Partner of mother	ambiv_afd	2
Sexual competence	compsex	2 & 3
Sexual communication	sexcom	2 & 3

continued on next page

continued from previous page

Construct	Variable name	Wave
Explosiveness and tendency to anger	explosiv	1
Shyness	shyness	1
Emotional autonomy	emotautn	1
Self-worth	selfworth	all
Depressiveness	depressiv	2 & 3
Anger	anger	3
Activity	activ	3
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
Economic deprivation anchor	ecodep_a/ecodep2_a	2 / 3
Economic deprivation household	ecodep_hh/ecodep2_hh	2 / 3
Newborn temperament child1	temperc1	2 & 3
Newborn temperament child2	temperc2	2 & 3
Newborn temperament child3	temperc3	2 & 3
Unspecific strain	unspstrain	3
Autonomy in the parenting role (3 Item Scale)	autonoms	2
Pleasure in the parenting role	pleasure	2
Parental Self Efficacy/Competence	comperz	2
Coparenting	coparent	2
Coparenting with ex-partner	coparent_ex	3
Coparenting with ex-partner - capikid1	coparent_opk1	3
Coparenting with ex-partner - capikid2	coparent_opk2	3
Parenting goals: Status	pgoalstatus	2 & 3
Parenting goals: Autonomy	pgoalautn	2 & 3
Parenting goals: Competence	pgoalcomp	2 & 3
Parenting: Partner support	partnersup	3
Parenting: Overprotection	overprotect	3
Parenting: Social support	socialsup	3
Parenting: Readiness to make sacrifices	sacrif_pacs	3

3.3 Process-generated variables

This chapter describes all process-generated variables listed in table 3.17.

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 3.17: 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*). Please note that for wave 1, release 2.0 the variable was modified (see table A.1).

Variable *gkpol* carries information on the size (population) of the community, divided into 7 categories (see table 3.18).

Table 3.18: 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 3.19).

Table 3.19: 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 3.20).

Table 3.20: Microm information availiable

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 Chemnitz pairfam location. For more information on on-site use please refer to the pairfam website.

3.4 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 3.21. 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*). 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 and wave 3 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 3.21: Available paradata in data set *anchor$*

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

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

[20] Duration 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.)

3.5 Weights

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

dweight Design weight to correct for disproportionate sampling across cohorts

psweight Post-stratification weight to correct for systematic nonresponse

dxpsweight Combined design and post-stratification weight.

In addition, we provide the 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 compared to the population. According to German National Statistics Office data[21], 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)

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

[22] The 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.

[23] According 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.

- Marital status x BIK (30 cells + 1 cell for "no answer" to marital status)
- Federal state x gender x having children (64 cells)
- Gender x age (6 cells).

The two distributions that include marital status were not used for the youngest cohort, as this cohort was almost completely unmarried. In the birth cohort 1981-1983 the categories "divorced" and "widowed" were combined to avoid too small cells. Due to data privacy regulations, cell sizes below 5,000 are not provided 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 3.22.

Table 3.22: 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 means 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 3 the weighting factors of temporary dropouts (people who skip at most one wave) refer to wave 1 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*)

[24] Note 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).

[25] The 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.

[26] Weight 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: $E = \left(\left(\sum_{n=1}^{N} g_n\right)^2\right) / \left(N \sum_{n=1}^{N} g_n^2\right)$ with E = effectiveness ($0 \leq E \leq 1$), g_n = weighting factor of the n-th case, N = unweighted sample size.

- probability of contact, given an accepted survey participation (*pcontact*)
- probability of response, given a successful contact (*panswer*).

These probabilities are results of stepwise logistic regressions, which were run separately for each birth cohort. Note that the independent variables of the regression models vary in their categories between each cohort (see table A.7, Appendix). Furthermore to control for temporary dropouts the variable *temp_dropout* is added to the logistic regression models of wave 3. The inverse of the three multiplied probabilities defines the longitudinal weight.

Best practice 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 have to decide whether to analyze only one birth cohort or more. 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. Table 3.23 and table 3.24 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 to treat the post-stratification weight, the combined weight, and the longitudinal weight as a design weight.

Table 3.23: When to use which weight

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

Table 3.24: When to use which weight

		Want to do longitudinal analysis	
		No	Yes
Combine more than one cohort in analysis	No	see table 3.23	psweight*lweight
	Yes	see table 3.23	dxpsweight*lweight

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 ...  .
```

3.6 Generated data sets

In wave 1, pairfam anchor respondents answered questions on their relationship history since the age of 14 and on their entire fertility history. As of wave 2, an electronic event-history calendar was used to collect retrospective and/or prospective information on the life domains: partnership, children, place(s) of residence, and education/employment activity. The information on these aspects of the individual life history was compiled into four separate, user-friendly data sets to allow convenient analysis of biographical information.

The episode data set *biopart* provides information on the individual relationship, cohabitation and marriage history, including both retrospective and prospective information on a monthly basis. In addition, the episode data set *bioact* contains monthly information on the anchor's educational and occupational activities as of the month of the wave 1 interview.

The data set *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. All four data sets contain information collected in waves 1 to 3.

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 set *bioact* is also "long-long". However the rows are not based on waves, but on activity episodes. If one activity is performed with one or more interruptions, there is one row for each of these separate episodes.

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

If the respondent provided information on the year and only the information on the month was missing, the value of the month was randomly imputed for the data sets *biopart* and *biochild*, 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:

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

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 3.1) is retained for variables that do not provide episodes or dates. The value "-3 Does not apply" was used for all variables in all four data sets.

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

- Person number (ID)
- Date of interview of all 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 wave-specific Stata do-files biopart1-2.do, biopart3.do, biochild1-2.do, biochild3.do, bioact2.do, bioact3.do, household2.do, and household3.do start from data sets *anchor$* and produce the four respective data sets. For some consistency checks, original non-anonymized data (mainly first names given during the interview) were used. The respective commands were commented out to ensure users are able to run these do-files.

Anchor-partner episode data - biopart

The data set *biopart* contains retrospective as well as prospective information on the anchor's partnership, cohabitation, and marriage episodes since the age of 14. There is one row for each partner ("long" format).[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 date[29]. 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 3.25 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*[30] are part of this data set. We generated the variables *respw2* and *respw3* as dummy variables to indicate whether the anchor took part in the interview of the respective wave (no attrition).

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

[28] This is the major difference compared to *biopart* released with wave 1 data, which was organized in "wide" format.

[29] Please 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.

[30] Generated variable from the latest available anchor data set.

by the anchor was retained.

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

To indicate whether a specific partner was the current partner in wave x, a dummy variable *currwx* (with x=number of wave) was included.[31]

For all partners, the variable *sexp* contains the "best" information on the partner's sex (also see chapter 2.2). For (formerly) current partners, the partner's date of birth was known and included (variable *dobp_gen*).[32] The identification number *pid* only exists for (formerly) current partners who (potentially) took part in the partner survey of the respective wave.

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

Please note that it was not possible to re-identify new partners in wave 2 as previous partners mentioned in wave 1. As of wave 3, the variable *pa30* specifies whether the anchor had already had a relationship with the supposedly new partner. We employed this information to detect identical partners for the data set *biopart*.

Table 3.25: List of variables included in data set *biopart*

Variable	Variable label	Values	Value labels
General information			
id	Person number anchor	(see table 2.2)	-
intdatw1	Date of interview wave 1	*date*	-
intdatw2	Date of interview wave 2	-3	Does not apply
		date	-
sex	Sex anchor	1	Male
		2	Female
dob	Date of birth anchor	*date*	-
homosex	Anchor homosexual	-7	Incomplete data
		0	Heterosexual
		1	Homosexual
respw2	Respondent in wave 2	0	No
		1	Yes
Ascending order of relationships			
index	Correct order of relationships:	-3	Does not apply
	Number partner	1	1st partner
		2	2nd partner
		...	...
		13	13th partner

continued on next page

[31]Note 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.

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

continued from previous page

Variable	Variable label	Values	Value labels
(Formerly) current partners			
pid	Person number current partner	(see table 2.2)	-
dobp	Date of birth partner	-7	Incomplete data
		-3	Does not apply
		date	-
Relationship history			
pnow1	Number partner in wave 1	-3	Does not apply
		number	-
pnow2	Number partner in wave 2	-3	Does not apply
		number	-
currw1	Current partner in wave 1	0	No
		1	Yes
currw2	Current partner in wave 2	0	No
		1	Yes
sexp	Sex partner	1	Male
		2	Female
dodp	Date of death partner	-7	Incomplete data
		-3	Does not apply
		date	-
beg	Beginning relationship	-7	Incomplete data
		-3	Does not apply
		date	-
end	End relationship	-99	Ongoing
		-66	Death partner
		-7	Incomplete data
		-3	Does not apply
		date	-
bkbeg	Beginning break k relationship	-7	Incomplete data
		-3	Does not apply
		date	-
bkend	End break k relationship	-7	Incomplete data
		-3	Does not apply
		date	-
Cohabitation history			
cohbeg	Beginning cohabitation	-7	Incomplete data
		-3	Does not apply
		date	-
cohend	End cohabitation	-99	Ongoing
		-66	Death partner
		-7	Incomplete data
		-3	Does not apply
		date	-

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
bkcohbeg	Beginning break k cohabitation	-7	Incomplete data
		-3	Does not apply
		date	-
bkcohend	End break k cohabitation	-7	Incomplete data
		-3	Does not apply
		date	-
Marriage history			
marbeg	Beginning marriage	-7	Incomplete data
		-3	Does not apply
		date	-
marend	End marriage	-99	Ongoing
		-66	Death partner
		-7	Incomplete data
		-3	Does not apply
		date	-
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
biopart flag variables			
biopartflag1	Inconsistency biopart: Marriage earlier than beginning of relationship	0	No inconsistency
		1	Inconsistency
biopartflag2	Inconsistency biopart: Overlapping cohabitation episodes with different partners	0	No inconsistency
		1	Inconsistency
biopartflag3	Inconsistency biopart: Beginning current and end previous marriage	0	No inconsistency
		1	Inconsistency
biopartflag4	Inconsistency biopart: Year of birth partner	0	No inconsistency
		1	Younger than 10 years old

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

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

- Negative durations of relationships, cohabitations, marriages (end before beginning)
- Inconsistencies across breaks in episodes (relationships and cohabitation):
  - subsequent episode (beginning and end) prior to beginning and end of preceding episode
  - subsequent episode (beginning and end) between beginning and end of preceding episode
  - beginning of subsequent episode prior to beginning of preceding episode
  - beginning of subsequent episode prior to end of preceding episode

  - end of subsequent episode prior to beginning of preceding episode
  - end of subsequent episode prior to end of preceding episode
  - identical beginning and end of two episodes

- Beginning of marriage before beginning of relationship (also see *flag5*, table A.8)
- Overlapping episodes of cohabitation with same or different partners (also see *flag6* and *flag7*, table A.8)
- End of previous marriage after beginning of current marriage (also see *flag8*, table A.8)
- Divorce from partner to whom never married (also see *flag10*, table A.8)
- Separation through death/divorce current spouse (also see *flag11*, table A.8)
- Separation before beginning relationship (current partner; also see *flag14*, table A.8)

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

Solution to these inconsistencies:

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

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

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

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

Sources for additional information regarding the respective partners:

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

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

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

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

The questions in the anchor data set that are part of the module "Meeting the current partner" (variables *pa1_-pa3, sdp1-sdp6*) refer to current partners whose value for *currw$* is "0" in the previous wave and "1" in the current wave (currw$: "0" at time t-1 and "1" at time t).

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

For further information on the computation of the data set, please see the Stata do-files biopart1-2.do (data from waves 1 and 2) and biopart3.do (data from wave 3). The do-file biopart1-2.do contains some modifications of the data set *biopart* released with the data from wave 2 (release 2.0) which are documented at the end of the do-file.

Anchor-child panel/episode data - biochild

The data set *biochild* contains retrospective and prospective information on the anchor's children[33] and on episodes of living together with the children in the same household. The data set is in long-long format. This means that there is one row per child and for each child there exists one row per wave (that child was mentioned in). Every wave, *biochild* is updated, i.e., the information on existing children is renewed and new children are added. The current data set contains information of wave 1 to wave 3.

Table 3.26 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 *varkX* in the anchor data sets.[34] The variable *surveykid* is a dummy variable that indicates if a child was interviewed as part of the PAPI child survey. Thus, this variable provides information on the success of merging the anchor files with the data sets *child2* and *child3*, respectively.

For all children, the variable *sexk* contains the "best" information on the children's sex (also see section 3.2).[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 3.6 above.

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

[34] For 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*.

[35] This "best" information was taken from the generated identifiers *k*sex_gen*.

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

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

Similarly to the variables *pnow1* and *pnow2* 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). In wave 1, partners were numbered consecutively; *pno* holds value "0" for the current partner. As of wave 2, the current partner who had already been the current partner in the previous wave gets the number "1", the current new partner the number "2", and partners who had been partners between two interviews the numbers "3", "4", "5" and so forth.

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

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

Table 3.26: List of variables included in data set *biochild*

Variable	Variable label	Values	Value labels
id	Person number anchor	(see table 2.2)	-
intdat	Date of interview (months since january 1900)	*date*	-
wave	Survey year	1	Wave 1: 2008/09
		2	Wave 2: 2009/10
		3	Wave 3: 2010/11
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	Incomplete data
		-3	Does not apply
		0	No
		1	Yes

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
index	Correct order of children (corresponding to date of birth)	-7	Incomplete data
		-3	Does not apply
		1	1st child (oldest)
		...	...
		10	10th child
dobk	Date of birth of child (months since january 1900)	-7	Incomplete data
		-3	Does not apply
		date	-
sexk	Gender of child	1	Male
		2	Female
statusk	Status of child	-7	Incomplete data
		-3	Does not apply
		1	Biological child
		2	Adopted child
		3	Partner's child/stepchild
		4	Foster child
currliv	Cohabitation with child	-7	Incomplete data
		-3	Does not apply
		1	Only with anchor
		2	With anchor and elsewhere
		9	Only alone/flat share
		10	Only with other parent unit
		11	Only with other relative
		12	Children's home
		13	Only elsewhere
currliv_detail	Details on [currliv] category <2. With anchor and elsewhere>	-3	Does not apply
		2	With anchor but also alone/flat share
		3	With anchor but also with other parent unit; mainly with anchor
		4	With anchor but also with other parent unit; namely in equal shares with anchor and with other parent unit
		5	With anchor but also with other parent unit; mainly with other parent unit
		6	With anchor but also with other relative
		7	With anchor but also at children's home
		8	With anchor but also elsewhere

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
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
		date	-
livkend	End living together with child	-99	Ongoing
		-7	Incomplete data
		-3	Does not apply
		date	-
b1livkbeg	Beginning break 1 living together with child	-7	Incomplete data
		-3	Does not apply
		date	-
b1livkxend	End break 1 living together with child	-7	Incomplete data
		-3	Does not apply
		date	-
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
		date	-

Various additional checks have been implemented to identify inconsistent cases.

Checks for episodes of living together with children:

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

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

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

Checks for consistency across wave 1 to wave 3:

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

- Child's status:
  If there are unrealistic differences between the information of wave 1 to wave 3, the newest information is preferred. Such unrealistic differences were changes in status : from or to the status "biological child".

Matching with further information on children:
One general possibility to merge information from the other data sets is to use the anchor's identification variable *id*.

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

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

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

Anchor-activities episode data - bioact

The data set *bioact* contains basic information on all of the anchor's activities in the areas of education and work, starting at the time of the interview of wave 1, i.e. with the release of wave 2 data. It covers the period lasting potentially 32 months between the survey date of wave 1 and wave 3. The data set is provided in "long-long" format, which means that it consists of one row for each education and/or work episode. There can be more than one activity per respondent and several instances of one activity, depending on the number of activities mentioned. With each wave, *bioact* will be updated. Wave 3 data also contains retrospective information about education and work starting at the age of eighteen. This information can be added to the data set, but is not included in *bioact*.

The variable *activity* displays the kind of activity the anchor respondent reported. Table 3.27 displays all possible activities covered by the questionnaire.

The anchor data sets of wave 2 and wave 3 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 has done for each month after the preceding interview wave up to the current interview wave. The idea is to acquire a full overview of what has happened concerning

education and employment during the period of - on average - the last year.

In contrast, the *bioact* data set provides the following information about the duration of each activity: The variables *beg* and *end* indicate the first and last month in which an activity was reported by the respondent. The variables *currw1*, *currw2*, and *currw3* mark the current activity at time of the interview of the respective wave. Additionally, *spell* presents the number of separated episodes per activity. The variable *censor* indicates if and in which way these spells were censored, i.e. if the episode began before or lasted longer than the covered period or if information about the month before or after the episode was missing. For censor details see table 3.27.

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

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

Table 3.27 displays the full set of variables which are part of this data set, in contrast to table 3.28, which shows all relevant variables included in anchor data. The variables can be easily merged with variables from anchor data by using the key variable *id*.

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

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

The do-files bioact2.do and bioact3.do document in detail how the data set *bioact* was computed.

Table 3.27: List of variables included in data set *bioact*

Variable	Variable label	Values	Value labels
General information			
id	Person number anchor	(see table 2.2)	-
intdatw1	Date of interview wave 1	*date*	-
intdatw2	Date of interview wave 2	*date*	-
intdatw3	Date of interview wave 3	*date*	-
dob	Date of birth	*date*	-
Activity history			
activity	Type of activity	1	Evening school, working on a school leaving certificate for adults
		2	Vocational training (apprenticeship, business school etc.)
		3	Vocational retraining / further education
		4	University of cooperative education (also "Berufsakademie")
		5	University of applied sciences, college, university
		6	General secondary school (first education)
		7	Pre-vocational training
		8	Technical/professional school
		9	Other education
		10	Full-time employment
		11	Self-employment
		12	Part-time employment (also multiple part-time jobs)
		13	Internships, traineeship, including unpaid work
		14	Marginal part-time employment, mini-job, "Ein-Euro-Job"
		15	Occasional or irregular employment
		16	Other type of employment, namely
		17	Maternity or paternity leave or other leave of absence for childcare
		18	Military service, alternative civilian service, voluntary social service year
		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
spell	Counter - Number of spells per activity	*number*	-

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
beg	Beginning of activity in month	*date*	-
end	End of activity in month	*date*	-
censor	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
currw1	Current activity in wave 1	0	No
		1	Yes
currw2	Current activity in wave 2	0	No
		1	Yes
currw3	Current activity in wave 3	0	No
		1	Yes
respw2	Respondent in wave 2	0	No
		1	Yes
respw3	Respondent in wave 3	0	No
		1	Yes

Table 3.28: List of available variables covering education and work in anchor data set

Variable	Description	Connection
sd32i*	Attained educational certificate/voc. qualification last year	Between waves
rtr31i*	Education experience after 18th birthday until first interview date	Before first interview
sd33	Attending same school as in previous wave	Currently
sd25	Type of school currently attending	Currently
sd26	Grade in school	Currently
rtr35i*	Work experience after 18th birthday until first interview date	Before first interview
job19	Same occupation as in previous wave	Currently
job20o	Current occupation	Currently
job21	Same occupational status as in previous wave	Currently
job2	Current occupational status	Currently
job3-job14	Additional information on current job(s)	Currently
inc 2 / inc21	Net income (earnings) last month	Currently

Anchor-household panel data - household

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

Table 3.29 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 a different way from wave 2 we generated variables for the differnt relationships. In wave 3 we distinguish between characteristics of partner (p*), children (c*), parents (pa*), and other household members (other*). Variables *p*_rel* contain the information of cohabitation with partners. We assigned value 1 to variable *p1_rel* if the anchor reported that he/she lives still (or again) with the wave 1 or wave 2 partner. We assigned value 1 to variable *p2_rel* if the anchor cohabited with a new partner.

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

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

For all persons for which we had the information (partners, children, parents), we used the generated variables on the birth month and year (**doby_gen*, **dobm_gen*, **sex_gen*) to fill in the respective variables. If this information was not available for some household members, we used the information gathered in the household grid instead.[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*. Since wave 3 we also matched the person number of the Anchor's child(ren) (*cid*) to variables *c*_id*. Note that this id is only available for so-called "capi-childen". Additionally, we generated variables *c*_point*, which include information about the position of the childen in the event history calendar. Moreover we assigned valid values on sex, date of birth, and person number forwards and backwards for household member present in both waves.

Table 3.29: List of variables included in data set *household*

Variable	Variable label	Values	Value labels
id	Person number anchor	(see table 2.2)	-
wave	Survey year	1	Wave 1: 2008/09
		2	Wave 2: 2009/10
		3	Wave 3: 2010/11
pid	Person number partner	(see table 2.2)	-
mid	Person number mother	(see table 2.2)	-
fid	Person number father	(see table 2.2)	-
smid	Person number stepmother	(see table 2.2)	-

continued on next page

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

continued from previous page

Variable	Variable label	Values	Value labels
sfid	Person number stepfather	(see table 2.2)	-
cid	Person number capi child	(see table 2.2)	-
intm	Month of interview	(see section 3.4)	-
inty	Year of interview	(see section 3.4)	-
sex_gen	Generated sex anchor	1	Male
		2	Female
doby_gen	Generated year of birth anchor	*see dob*_gen*	-
dobm_gen	Generated month of birth anchor	*see dob*_gen*	-
hhsizemrd	Number of hh members main residence	*see hhsizemrd*	-
pmrd	Partner lives in household main residence	-7	Incomplete data
		0	Partner does not live at main residence
		1	Partner lives at main residence
mmrd	Mother lives in household main residence	-7	Incomplete data
		0	Mother does not live at main residence
		1	Mother lives at main residence
fmrd	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	Incomplete data
		0...10	-
othmrd	Number of other hh members main residence	-7	Incomplete data
		0...11	-
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
		5	Other type of household

continued on next page

continued from previous page

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

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
hhincgcee	Net equivalence income (GCEE)	*see hhincgcee*	-
pX_rel	Anchor's relation to partner X	-3	Does not apply
		1	partner
pX_sex	Sex of partner X	-7	Incomplete data
		-3	Does not apply
		1	Male
		2	Female
pX_byear	Year of birth of partner X	-7	Incomplete data
		-3	Does not apply
pX_bmonth	Month of birth of partner X	-7	Incomplete data
		-3	Does not apply
pX_id	Person number of partner X	-7	Incomplete data
		-3	Does not apply
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
		9	foster father

continued on next page

continued from previous page

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

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

After completing the editing of the information on household members, we generated aggregate household characteristics. The data can be merged with information from other data sets by using the person number of the anchor, or the person number of household members (partner, child, parents) who are potentially respondents of the respective multi-actor surveys.

The data set will be updated with each new wave.

4. Partner data

TNS Infratest compiled a raw data set from the returned PAPI questionnaires. Data preparation was performed by the Bremen pairfam team following closely the procedures for cleaning the anchor data.

4.1 Data editing

This section describes the editing of wave 1, wave 2, and wave 3 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 4.1 shows the missing codes and value labels assigned to the partner data.

Table 4.1: 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. For variable *psd27o*, included in the wave 1 & 3 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.

4.2 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 wave2 (*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 3.2).

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 4.2) are not part of the delivered data set. They can be generated by running the syntax file scales_partner$ which is available for each wave on the pairfam website and on CD. For further information, refer to the scales handbook (Schmahl et al. 2012) which contains a detailed description of all scales.

Table 4.2: List of scales included in data sets *partner$*

Construct	Variable name	Wave
Traditional concept of marriage	ptradmarr	1 & 3
Value of Partnership: Negative expectations	pvopneg	1 & 3
Value of Children: Benefit of stimulation	pvocbstim	1 & 2
Value of Children: Costs of comfort	pvocccomf	1 & 2
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
Partnership: Anxiety about loss of love	plovewitanx_apd	1 & 3
Partnership: Anxiety about being absorbed	penganx_apd	1 & 3
Partnership: Independence	pindep_apd	all

continued on next page

continued from previous page

Construct	Variable name	Wave
Partnership: Own partnership satisfaction (global scale)	psatpart	1 & 3
Partnership: Feelings of competence in the partnership	pcomppart2	2
Partnership: Future orientation	pcomfut_apd	all
Partnership: Tolerance of conflicts	pcomctol_apd	all
Partnership: Orientation of reciprocity Self	preciproc_aps	1 & 3
Partnership: hostile attributions towards anchor (self assessment)	phostattr_aps	3
Partnership: Frequency of manifest conflicts	pconfldom_apd	1
	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
Partnership: Manipulation Self	pmanipul_aps	1 & 3
Partnership: Dyadic coping Partner	pdycop_apo	1 & 3
Partnership: Dyadic coping Self	pdycop_aps	1 & 3
Partnership: Instability of partnership	pinstab_apd	all
Explosiveness and tendency to anger	pexplosiv	1 & 3
Shyness	pshyness	1
Emotional autonomy	pemotautn	1 & 3
Selfworth	pselfworth	all
Depressiveness	pdepressive	2 & 3
BIG 5: Neuroticism	pneurot	2 & 3 (only new partners)
BIG 5: Extraversion	pextrav	2 & 3 (only new partners)
BIG 5: Agreeableness	pagreeable	2 & 3 (only new partners)
BIG 5: Conscientiousness	pconscient	2 & 3 (only new partners)
BIG 5: Openness	popenness	2 & 3 (only new partners)
Coparenting with the other parent	pcoparent	2
Parental Self Efficacy/Competence	pcomperz	2

continued on next page

continued from previous page

Construct	Variable name	Wave
Partnership: sacrifice in raising children	psacrif_pacs	3
Partnership: recognition / support in education through partner	ppartnersup	3
Partnership: Own intimacy within partner-mother dyad (self assessment)	pintim_ams	3
Partnership: Own intimacy within partner-father dyad (self assessment)	pintim_afs	3
Parntership: Conflict within partner-mother dyad	pconfl_amd	3
Parntership: Conflict within partner-father dyad	pconfl_afd	3

5. Parent data

TNS Infratest compiled a raw data set from the returned parent's PAPI questionnaires. Data editing was performed by the Chemnitz pairfam group following closely the procedures used to clean the anchor data.

5.1 Data editing

This section describes the editing of wave 2 and 3 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 empty[1] or because they were suspected to have been derived from duplicate questionnaires or to have been provided by the anchor[2]. Thus, parent's file was reduced by 24 cases, which resulted in a final data file of N=5,015. There were no cases deleted in the third wave. The final data set contains N=3,946 entries.

Variable and value labels

All variables in the parent data start with the prefix "par-". To facilitate analysis, variables derived from questions that were also included in the anchor survey were labeled with the same variable name (plus the prefix par-). Variables unique to the parent survey had unique names according to our system of variable names.
Every variable in the data set was assigned a label. Variable labels contain a short description of the variable and the position in the parents questionnaire (question number). The values of all variables were labeled according to the parent's codebook.

Missing values

Missings were defined and labeled as in the anchor data set, with one exception. Although the answer category "don't know" (respondent could not answer) was allowed very rarely and thus could not be distinguished precisely from "no answer" (respondent did not answer), code "-2 No answer" was consistently assigned if no box had been checked. This is consistent with the procedure applied for editing the partner data. Table 5.1 shows the missing codes and value labels assigned to the parent's 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

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

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

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 5.1: 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 parent's 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 the CAPI-child (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 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).

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.10 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.

5.2 Generated variables and scales

Table 5.2 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* indicates the language version of the questionnaire (German, Russian, or Turkish). The variables *partype*, *parposition*, *parsexparent*, and *parsexanchor* refer to the relationship between the respondent and the anchor (biological or stepparent, mother or father, daughter or son). The variable *parageanchor* provides the anchor's year of birth taken from the anchor interview, as well as the variable *parcohort*.

In addition, similar to the anchor data, user-friendly parent's 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 Schmahl et al. 2012). However, unlike other indicators, these scale-indicators are not included in the data-file.

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.

Table 5.2: List of generated variables included in data set *parent$*

Construct	Variable name	Wave
Identification		
Anchor's identification number	id	2 & 3
Respondent's identification number	parid	2 & 3
CAPI-child's identification number	cid	2 & 3
Paradata		
Respondent's type (based on questionnaire)	partype	2 & 3
Sex of respondent (based on questionnaire)	parsexparent	2 & 3
Position of respondent (based on questionnaire)	parposition	2 & 3
Sex of the anchor (based on questionnaire)	parsexanchor	2
Wave (based on questionnaire)	wave	2 & 3
Language version (based on questionnaire)	parlng	2 & 3
Date of the interview	parintd, parintm, parinty	2 & 3
Cohort anchor	parcohort	2 & 3
Year of birth anchor based on anchor interview	parageanchor	2 & 3
# of children (corrected): # of children in parsd32; only in case of more children described in parsd14ff, than stated in parsd32: correction; -7 if no valid answer in parsd32 & no children described	parkids	2 & 3

continued on next page

continued from previous page

Construct	Variable name	Wave
# of grandchildren	pargrandkids	2 & 3
Generated variables		
Age variables (respondent, respondent's parents)	parage parage2 parmage parmage2 parfage parfage2	2 & 3
Age variables (respondent's partner)	parpage parpage2	2
Current primary and secondary activity status	parcasprim, parcassec	2 & 3
Highest school degree	parschool	2 & 3
Highest vocational degree	parvocat	2 & 3
CASMIN classification of educational attainment	parcasmin	2 & 3
ISCED classification of educational attainment	parisced	2 & 3
Years in education, schooling and vocational qualification	paryeduc	2 & 3
Labor force status	parlfs	2 & 3
Household size main residence	parhhsizemrd	2
Net equivalence income	parhhincgcee	2
Type of children	park1type park2type park3type park4type	2
Marital status	parmarstat	2
Partner status	parrelstat	2
Do parents live in a shared household?	parigr27	2 & 3
Are parents married?	parigr28	2 & 3
State /Bundesland	parbula	3
Size of community in 7 categories	pargkpol	3
Settlement structure	parbik	3
New parent in wave 3	parw3ne	3
Information about the correct grandchild	parcorgc	3

Scales

The following scale variables (see table 5.3) can be generated by running the do-file scales_parents$, which can be found on the pairfam website and on CD. For further information, refer to the scales manual (Schmahl et al. 2012) which contains a detailed description of all scales.

Table 5.3: List of scales included in data set *parent$*

Construct	Variable name	Wave
Filial Obligation (mean indicator: parval2i1, parval2i5)	parobligationf	2
Grandparental Obligation (mean indicator: parval2i2, parval2i4)	parobligationp	2
Parental Obligation (mean indicator: parval2i3, parval2i6)	parobligationgp	2
Traditional concept of marriage (mean indicator: parval1i2, parval1i7, and parval1i8)	partradmarr	3
Readiness to make sacrifices (mean indicator: parcrn32i1, parcrn32i2, and parcrn32i3)	parsacrif_pacs	3

continued on next page

continued from previous page

Construct	Variable name	Wave
Frequency of Joint Activities (mean indicator: parigr73a, parigr74a, parigr75a, parigr76a, and parigr77a)	paractiv_paras	2
NRI Partner Conflict	parconfl_parparpd	2 & 3
NRI Partner Intimacy	parintim_parparps	2 & 3
NRI Partner Approval	paradmir_parparpo	2 & 3
NRI Partner Dominance	pardomin_parparpo	2 & 3
NRI Anchor Conflict	parconfl_parad	2 & 3
NRI Anchor Intimacy	parintim_paras	2 & 3
NRI Anchor Approval	paradmir_parao	2
NRI Anchor Dominance	pardomin_parao	2
Co-parenting with respect to anchor (younger than 21) (mean indicator: parcrn21i1, parcrn21i2, parcrn21i3)	parcoparent	2
Educational style with respect to anchor (younger than 21): Monitoring (mean indicator: parcr1i2, parcr1i9, parcr1i6, parcr1i12)	parmonitor_paras	2
Educational style with respect to anchor (younger than 21): Strict control (mean indicator: parcr1i18, parcr1i19, parcr1i20, parcr1i21)	parstrict_paras	2
Educational style with respect to anchor (younger than 21): Emotional Warmth (mean indicator: parcr1i1, parcr1i5, parcr1i14)	parwarmth_paras	2
Educational style with respect to anchor (younger than 21): Negative Communication (mean indicator: parcr1i3, parcr1i8, parcr1i11)	parnegcomm_paras	2

6. Child data

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

6.1 Data editing

This section describes the editing of wave 2 & 3 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 6.1). 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.

Table 6.1: 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

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 ech of these variables, code "0" indicates that the respective inconsistency is non-existant. Table 6.2 describes the flag variables in detail.

Table 6.2: List of flag variables to identify inconsistencies *child$*

Variable	Lable	Value	Value Labels
cflag1	Inconsistency sex of child between statement of interviewer and statement of anchor	0	Non consistency
		1	Inconsistency
cflag2	Inconsistency year of child's birth between statement of anchor and statement of 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 6.3 lists all of the generated tag variables that are part of the data set *child3*.

Table 6.3: List of flag variables to identify inconsistencies *child3*

Variable	Label	Value	Value Labels	Description
tag_csex	Inconsistency between waves: child's sex	0	No consistency	Child's sex in wave 3 is not child's sex in wave 2
		1	Inconsistency	
tag_cdobm	Inconsistency between waves: child's month of birth	0	No inconsistency	Child's month of birth in wave 3 is not child's month of birth in wave 2
		1	Inconsistency	
tag_cdoby	Inconsistency between waves: child's year of birth	0	No inconsistency	Child's year of birth in wave 3 is not child's year of birth in wave 2
		1	Inconsistency	

Coding open answers

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

Anonymity

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

By means of depersonalization, all string variables in the data set were finally transformed to numeric variables. There are thus no string variables contained in the child data set. Valid answers to open questions were recoded to value "1" throughout. The variables affected by this procedure are shown in table 6.4 along with the value labels (<information> mentioned) assigned to valid answers on the anonymous numeric variables.

Table 6.4: List of anonymized variables

Variable	Variable label	Anonymized value label	Wave
cdobd	Child day of birth	Day mentioned	2 & 3
cedu1o	Other schooltype	Other schooltype mentioned	2
cedu1ao	Other schooltype	Other schooltype mentioned	3
cpcr4	Name anchor	Name mentioned	2 & 3
cpcr5	Name current partner of anchor in same household	Name mentioned	2 & 3
cgp1	Name of anchor's mother	Name mentioned	2
cgp2	Name of anchor's father	Name mentioned	2
cpcr13	Name biological parent outside the household	Name mentioned	3

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.

6.2 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 6.5.

The file genvars_child (available in both Stata and SPSS) 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 6.5: List of generated variables included in data set *child$*

Construct	Variable name	Wave
relstatac	Relationship/kinship between anchor and child	2 & 3
cinty	Child interview year	2 & 3
cintm	Child interview month	2 & 3
cintd	Child interview day	2 & 3
cagey	Child's age in years	2 & 3
cagem	Child's age in months	2 & 3
cgp1	Anonymized: Name of anchor's mother - mentioned	2
cgp2	Anonymized: Name of anchor's father - mentioned	2
ctitlea	Child's naming for anchor	2 & 3
ctitlep	Child's naming for anchor's partner	2 & 3
ctitleop	Child's naming for the other biological parent outside the household	3

The generated variables are included in second- and third-wave child data. In contrast to these, the scale variables (see table 6.6) are not part of the delivered data set. They can be generated by running the syntax file scales_child$ which can be found on CD and on the pairfam website. For further information, refer to the scales manual (Schmahl et al. 2012), which contains a detailed description of all scales.

Table 6.6: List of scales included in data set *child$*

Construct	Variable name	Wave
Class atmosphere	cclassatmo	2 & 3
Social integration	csocialinteg	2 & 3
Peer rejection	cpeerref	2 & 3
Parental school engagement	cengagement	2
Economic deprivation	cecodep	2 & 3
Intimacy child-best friend-relationship	cintim_cbf	3
Companionship child-best friend-relationship	ccomp_cbf	3
Emotional warmth child-anchor-relationship	cwarmth_cao	2 & 3
Emotional warmth child-partner-relationship	cwarmth_cpo	2 & 3
Emotional warmth child-other parent-relationship	cwarmth_copo	3
Strict control child-anchor-relationship	cstrict_cao	2
Strict control child-partner-relationship	cstrict_cpo	2

continued on next page

continued from previous page

Construct	Variable name	Wave
Strict control child-other parent-relationship	cstrict_copo	3
Monitoring child-anchor-relationship	cmonitor_cao	2 & 3
Monitoring child-partner-relationship	cmonitor_cpo	2 & 3
Fear of love withdrawal child-anchor-relationship	lovewitanx_cas	3
Fear of love withdrawal child-partner-relationship	lovewitanx_cps	3
Fear of love withdrawal child-other parent-relationship	lovewitanx_cops	3
Intimacy child-anchor-relationship	cintim_cas	2 & 3
Intimacy child-partner-relationship	cintim_cps	2 & 3
Intimacy child-other partner-relationship	cintim_cops	3
Conflict child-anchor-relationship	cconfl_cad	2 & 3
Conflict child-partner-relationship	cconfl_cpd	2 & 3
Conflict child-other partner-relationship	cconfl_copd	2 & 3
Admiration child-anchor-relationship	cadmir_cao	2 & 3
Admiration child-partner-relationship	cadmir_cpo	2 & 3
Admiration child-other parent-relationship	cadmir_copo	3
Parental reliability child-anchor-relationship	creliabl_cad	2
Parental reliability child-partner-relationship	creliabl_cpd	2
SDQ: Conduct problems	cconduct	2 & 3
SDQ: Emotional symptoms	cemotion	2 & 3
SDQ: Prosocial behavior	cprosoc	2 & 3

7. Parenting data

TNS Infratest compiled a raw data set from the returned PAPI questionnaires. Data editing was performed by the Munich pairfam group following closely the procedures to clean the anchor data.

All anchors with children who took part in the child-interview 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.

7.1 Data editing

This section describes the editing of wave 2 & 3 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

The definition of missings was not as complex as in the anchor data set. As there were no filters in the parenting questionnaire and no open answers, there were no missing codes "-3" and "-6". There was no possibility to indicate whether the respondent could not or did not want to answer a question. Therefore, there was no missing value "-1", only the missing value "-2 No answer". Furthermore, there were no consistency checks possible in the parenting questionnaire because we only assessed subjective perceptions. Therefore the missing code "-5" was not used. As there were no computed variables or scales, the missing code "-7 Incomplete data" also was not necessary. Hence "-2 No answer" and "-4 Incorrect entry" are the only applied missing codes in the parenting data.

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.

7.2 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 7.1) not included in the delivered data set. The file scales_parenting$ contains the required syntax (in both Stata and SPSS) and can be found on CD and on the pairfam website. For further information, refer to the scales manual (Schmahl et al. 2012), which contains a detailed description of all scales.

Table 7.1: List of scales included in data set *parenting$*

Construct	Variable name	Wave
Parent-child-relationship: Readiness to make sacrifices	sacrif_pacs	2
Parent-child-relationship: Hostile attribution	hostattr_pacd	2 & 3
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 & 3
Parent-child-relationship: Psychological control	psycontrol_pacs	2
Parent-child-relationship: Negative communication	negcomm_pacs	2 & 3
Parent-child-relationship: Monitoring	monitor_pacs	2 & 3
Parent-child-relationship: Strict control	strict_pacs	2
Parent-child-relationship: Inconsistent parenting	inconsist_pacs	2 & 3
SDQ: Hyperactivity	hyper_paco	2
SDQ: Emotional symptoms	emotion_paco	2 & 3
SDQ: Peer problems	peer_paco	2
SDQ: Prosocial behavior	prosoc_paco	2 & 3
SDQ: Conduct problems	conduct_paco	2 & 3
Parent-child-relationship: Intimacy	intim_paco	2 & 3
Parent-child-relationship: Conflict	confl_pacd	2 & 3
Parent-child-relationship: Admiration	admir_pacs	2
Parent-child-relationship: Dominance	domin_paco	2 & 3

8. Outlook

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

Wave 4 will bring new data sets for anchor respondents and partners plus data sets on the parents survey, the child interview, and the parenting survey. As a new feature, anchor's children who formerly participated in the child survey and turned 16 or 17 years old in the mean time will become anchor respondents themselves. There will be no retrospective modules as part of the anchor questionnaire any longer. Substantively, the rotating modules on intergenerational relationships, fertility, and social networks will be part of the anchor interview in wave 4.

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

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

Bibliography

Arránz Becker, O., J. Brüderl, P. Buhr, L. Castiglioni, D. Fuß, V. Ludwig, J. Schröder, and N. Schumann (2012). The German Family Panel - Study Design and Cumulated Field Report (Waves 1 to 3). Pairfam Technical Paper No. 01.

Brauns, H. and S. Steinmann (1999). Educational Reform in France, West-Germany, and the United Kingdom. Updating the CASMIN Educational Classification. *ZUMA-Nachrichten 44*(23).

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.

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 (1996). Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of occupations. *Social Science Research 25*.

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.

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.

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.

Schmahl, F., B. Wilhelm, S. Fiedrich, E.-V. Wendt, C. Thönnissen, and S. Walper (2012). Scales Manual wave 1 to 3. Technical report.

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.

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.

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

Table A.1: Changes in data sets *anchor1* and *partner1*: from release 1.0 to release 2.0

Description of changes	Variables affected
Anchor data	
New generated variables	
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
Modified generated variables	
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

continued on next page

continued from previous page

Description of changes	Variables affected
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
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
Discontinued generated variables	
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)	hhsize2rd
Modified labels of main variables	
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, co1i3
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	
Modified labels	
Some new English variable and value labels (translation edited)	various variables throughout the data set
Variable labels modified (German and English version)	ppa16i1-6, ppa22pi1-8, ppa22ri1-8, pfrt1

Table A.2: Changes in data set *anchor1*: from release 2.0 to release 3.0

Description of changes	Variables affected
Anchor data	
Modified generated variables	
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 *biopart*	np, ncoh
Value change (3 to 4 and vice versa) and rename of value label "POS 8./9."	school, pschool
Modified tag and flag variables	
Additional couples of anchor respondents identified	flag26
Modified values of main variables	
Value of variable sat2 is modified: -1 is set to 8; -2 is set to 9	sat2
Modified labels of main variables	
Variable labels modified (German version)	frt11v1i*, pa16*
Variable labels modified (English version)	job18*
Variable label modified (German and English version)	frt2
Value labels modified (German version)	sd13, rtr11, rtr15p*

Table A.3: Changes in data set *anchor1* and *partner1*: from release 3.0 to release 3.1

Description of changes	Variables affected
Anchor data	
Modified generated variables	
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
Modified labels of main variables	
Variable labels modified (German and English version): Question number in variable label corrected	frt11v2i1-int10i2
Partner data	
Modified values of main variables	
Recoding of values to avoid discontinuity compared to later waves	psd10
Recoding of missing values: "-1" was mistakenly included in category "-2"	psat1*, psat3

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

Description of changes	Variables affected
Anchor data	
Modified generated variables	
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 frt1 and frt2 variable infertile of wave 1 is used for computation if no valid information available in wave 2	infertile
Modified and new value: value 16 also includes "Vocational retraining / further education"; new value 7 "Kolleg, 2. Bildungsweg"	enrol
Value change (3 to 4 and vice versa) and rename of value label "POS 8./9."	school, pschool
Modified values: -7, 2, 3	marstat
Some values modified according to generated variables marstat and relstat	cohabdur, mardur
Slightly modified due to corrections of data set *biopart*	ncoh, nmar
Value labels modified	sex_gen, psex_gen, k*sex_gen, dobm_gen, siops
Modified tag and flag variables	
Additional couples of anchor respondents identified	flag26
Value label -3 recoded to 0 (theoretically possible)	flag8, flag16
Value labels slightly modified	flag26, flag_frt6
Variable label modified	flag_frt6, flag20, flag25
Modified values of main variables	
96 is set to 6; 97 is set to 7	frt24i1-8
Value labels of categories 1 & 0 interchanged	ehc18p1-4
Modified labels of main variables	
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*

continued on next page

continued from previous page

Description of changes	Variables affected
Parent data	
Modified generated variables	
English variable label added	pargkpol, parbik, parbula
Modified labels of main variables	
Value labels modified (English version)	parigr85a, pargc12, parigr103, parinc23
Child data	
Modified generated variables	
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
Modified labels	
The value was added to the value label	all variables of this data set
Parenting data	
Discontinued variables	
Variables deleted	anchor, partner

Table A.5: Changes in data set *anchor2*, *child2*, *parenting2*: from release 3.0 to release 3.1

Description of changes	Variables affected
Anchor data	
Modified generated variables	
Values of some cases corrected (see generated identifiers)	age, page, mage, fage, smage, sfage, k*age, homosex, hhcomp, lweight, ppanel, pcontact, panswer
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
Variable labels modified (English version)	*nat1, *nat2, *cob, infertile, *enrol, *school, *vocat, *casmin, *isced, *isced2, *yeduc, siops, *sex_gen, *doby_gen, *dobm_gen
Value labels modified (English version)	k*dobm_gen, k*doby_gen, k*sex_gen
Modified tag and flag variables	
Value labels modified (English version)	flag1, flag5, flag16, flag17, flag_cas, pflag_cas
Modified labels of main variables	
Variable labels modified (German and English version)	frt11v1*
Variable labels modified (English version)	crn19i1
Value labels modified (German and English version)	sin3*, sin4*, sin5*, sin6*, pa1*, frt10*, crn10*, crn11*, crn20*, cpas5
Value labels modified (German version)	cps8i3
Value labels modified (English version)	pa3, cprs2, netp*n, inc25, inc26, inc27, capikid, he3, hm2, hsv2, hv2, hsm2
Child data	
Modified labels of main variables	
Variable labels modified (German version)	cedu1, cedu5i4, cpcr8i8, clsr1i1, clsr1i14
Variable labels modified (English version)	cedu8i2, cpcr8i1, cgp1, l391s

continued on next page

continued from previous page

Description of changes	Variables affected
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	
Changes in case number	
Two empty cases deleted from data set	(id=494001000, id=609745000)
Modified values of main variables	
Replacing 6 missing person numbers partner	pid
Modified labels of main variables	
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

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

Description of changes	Variables affected
Anchor data	
Modified generated variables	
Variables school & 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); changes in other variables due to changes of school & pschool	school, pschool, casmin, pcasmin, isced, pisced, isced2, pisced2, yeduc, pyeduc
Modified tag and flag variables	
Value labels modified (English version)	flag_cas, pflag_cas
Modified labels of main variables	
Variable labels modified (English version)	job18h, job18m
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	
Modified values of main variables	
Recoding: 1 to -3 if no inconsistency possible since child new in wave 3	tag_csex, tag_cdobm, tag_cdoby
Modified labels of main variables	
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	
Modified values of main variables	
Change of response code: 6 is set to 8	parigr53a-parigr59a, parigr78a-parigr79a, parigr80a-parigr81a, parigr60a-parigr66a
New category "no partner" added	parpa17i1-parpa17i8
Modified labels of main variables	
Assignment of variable labels to variables adjusted (German and English version)	parval1ix

continued on next page

continued from previous page

Description of changes	Variables affected
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
Modified filter of main variables	
Variables asked irrespective of existing contact to parent (German and English version)	parigr40p1, parigr40p3, parigr41p1, parigr41p3
Parenting data	
Modified labels of main variables	
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.7: List of regressors for logit estimations of longitudinal weight

Independent variables and reference category	Definition		
	Cohort 1	**Cohort 2**	**Cohort 3**
int5	*Interference of other persons in the interview*	*Interference of other persons in the interview*	*Interference of other persons in the interview*
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
lfs	*Labor force status*	*Labor force status*	*Labor force status*
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
marstat	*Marital status*	*Marital status*	*Marital status*
ehe1		never married	never married
ehe2 (ref.)		married	married
ehe3	widowed, divorced or incomplete data	widowed, divorced or incomplete data	

continued on next page

Table A.7, continued from previous page

Independent variables and reference category	Definition		
	Cohort 1	**Cohort 2**	**Cohort 3**
incnet	*Personal net income*	*Personal net income*	*Personal net income*
pinc1	incomplete data	incomplete data	incomplete data
pinc2 (ref.)	no income	no income	no income
pinc3	1 Euro or more income	1 - 986 Euro income	1 - 986 Euro income
pinc4		987 - 1500 Euro income	987 - 1500 Euro income
pinc5		1501 Euro or more income	1501 - 2500 Euro income
pinc6			2501 Euro or more income
hhsizemrd	*Household size main residence*	*Household size main residence*	*Household size main residence*
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
hc5h1	*Main Residence: Ownership*	*Main Residence: Ownership*	*Main Residence: Ownership*
house	Ownership	Ownership	Ownership
house (ref.)	for rent, sublease, others, no answer	for rent, sublease, others, no answer	for rent, sublease, others, no answer
school	*Highest School Degree*	*Highest School Degree*	*Highest School Degree*
edu1	(ref.) currently enrolled	currently enrolled, without degree, other school degree, missing	currently enrolled, without degree or other school degree
edu2	without degree or other school degree, Missings	lower, Volks-/Hauptschulabschluss	lower, Volks-/Hauptschulabschluss

continued on next page

Table A.7, continued from previous page

Independent variables and reference category	Definition		
	Cohort 1	**Cohort 2**	**Cohort 3**
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
telgesamt	*Frequency of telephone contacts*	*Frequency of telephone contacts*	*Frequency of telephone contacts*
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
persgesamt	*Frequency of personal contacts*	*Frequency of personal contacts*	*Frequency of personal contacts*
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
bik	*Settlement Structure*	*Settlement Structure*	*Settlement Structure*
biki1 (ref.)	City Center - population 500.000 or more	City Center - population 500.000 or more	City Center - population 500.000 or more
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

continued on next page

Table A.7, continued from previous page

Independent variables and reference category	Definition		
	Cohort 1	**Cohort 2**	**Cohort 3**
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
bula	*Federal State*	*Federal State*	*Federal State*
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
bundl10	Mecklenburg-Vorpommern	Mecklenburg-Vorpommern	Mecklenburg-Vorpommern
bundl11	Sachsen (Saxony)	Sachsen (Saxony)	Sachsen (Saxony)

continued on next page

Table A.7, continued from previous page

Independent variables and reference category	Definition		
	Cohort 1	**Cohort 2**	**Cohort 3**
bundl12	Sachsen-Anhalt (Saxony-Anhalt)	Sachsen-Anhalt (Saxony-Anhalt)	Sachsen-Anhalt (Saxony-Anhalt)
bundl13	Thüringen (Thuringia)	Thüringen (Thuringia)	Thüringen (Thuringia)
sat6	*Life satisfaction*	*Life satisfaction*	*Life satisfaction*
lsat1	Life satisfaction: 0 - 6 and no answer	Life satisfaction: 0 - 5 and no answer	Life satisfaction:: 0 - 5 and no answer
lsat2	Life satisfaction: 7	Life satisfaction: 6	Life satisfaction: 6
lsat3	Life satisfaction: 8	Life satisfaction: 7	Life satisfaction:: 7
lsat4	Life satisfaction: 9	Life satisfaction: 8	Life satisfaction: 8
lsat5	(ref.) Life satisfaction: 10	Life satisfaction: 9	Life satisfaction: 9
lsat6		(ref.) Life satisfaction: 10	(ref.) Life satisfaction: 10
hl1t	*Health status past 4 weeks*	*Health status past 4 weeks*	*Health status past 4 weeks*
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"
age	*age*	*age*	*age*
age1	15 years		
age2	16 years		
age3 (ref.)	17 years		
age4		25 years	
age5		26 years	

continued on next page

Table A.7, continued from previous page

Independent variables and reference category	Definition		
	Cohort 1	**Cohort 2**	**Cohort 3**
age6 (ref.)		27 and 28 years	
age8			34 years
age9			36 years
age10 (ref.)			37 and 38 years
nat1	*Nationality*	*Nationality*	*Nationality*
german	German Nationality	German Nationality	German Nationality
sex	*Gender*	*Gender*	*Gender*
sex1	Male	Male	Male
temp_dropout			*temporary drop-out*

Table A.8: List of flag variables to identify inconsistencies (*anchor$*)

Variable	Label	Values	Value Labels	Description
flag1	Inconsistency biological child and sex of the other parent	0	No inconsistency	
		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
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant

continued on next page

Table A.8, continued from previous page

Variable	Label	Values	Value Labels	Description
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_1z_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	
		"xz"	Incons. partner x episode z and curr. cohab.	End previous cohabitation episode with previous partner after beginning of current cohabitation episode with current partner

continued on next page

Table A.8, 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	 *see value labels*

continued on next page

Table A.8, continued from previous page

Variable	Label	Values	Value Labels	Description
		4	Younger than 14 years old but with completed tertiary education	
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 child and anchor	0	No inconsistency	
		"x"	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
flag17	Inconsistency personal net and household income	0	No inconsistency	
		1	Inconsistency	Net personal income larger than net household income

continued on next page

Table A.8, 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 psex partner data)
		2	Inconsistency, female partner	Partner male according to anchor, female according to partner (see variable psex 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
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
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
flag23	Inconsistency home size & number of rooms	0	No inconsistency	
		1	Inconsistency HH1	Home size<=10qm and number of rooms>2 or Home size<=20qm and number of rooms>4

continued on next page

Table A.8, continued from previous page

Variable	Label	Values	Value Labels	Description
flag24	Inconsistency overnight stays in main and second residence	0 1	No inconsistency Inconsistency	Stay in residence one and two almost every night
flag25	Inconsistency biological parents in household who should be dead according to preload	0 1 2 3 4	No inconsistency HH1: biological mother is dead HH1: biological father is dead HH2: biological mother is dead HH2: biological father is dead	Person mentioned although this person is dead (acc. to preload)
flag26	Relationship between two anchor persons	0 "x"	No inconsistency Couple "x"	Two anchor persons have a relationship
flag_cas	Inconsistency current activity status	0 "a_1a_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_1a_2"	No inconsistency Inconsistency activ. status a_1 & a_2	Multiple answers (a_1 & a_2) for current activity status not plausible

continued on next page

Table A.8, 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
flag_frt6	Probably inconsistent data in frt6, values are to 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
flag_igb	Inconsistent combination: biological and adoptive parents	0	No error	
		1	Programming error	Having both biological and adoptive parents inconsistent according to codebook
flag_isco _kldb	Problem coding isco and kldb	0	No inconsistency	
		1	Answer ambiguous, most frequent occ. assigned	Classification of occupations problematic *(also see chapter 3.2)*
		2	Answer ambiguous, lowest-level occ. assigned	

Table A.9: 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	
		1	Inconsistency	Anchor's sex in current wave is not anchor's sex in preload
tag_dob	Inconsistency date of birth anchor between waves	0	No inconsistency	
		1	Inconsistency: both month & year	
		2	Inconsistency: month	Anchor's date of birth is not anchor's date of birth in preload
		3	Inconsistency: year	
tag_identk"x"	Inconsistency identity child "x" between waves	0	No inconsistency	
		1	Inconsistency	Child "x" in previous wave is not the same child as in current wave
tag_sexk"x"	Inconsistency sex child"x" between waves	0	No inconsistency	
		1	Inconsistency: sex child	Child "x" changed sex
		2	Child deleted	Child deleted
tag_dobk"x"	Inconsistency date of birth child "x" between waves	0	No inconsistency	
		1	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
tag_biok"x"	Inconsistency status child "x" in anchor data between waves	0	No inconsistency	
		1	Inconsistency: status child	Status of child"x" is not equal to status of child"x" in preload
		2	Child deleted	Child"x" deleted
tag_biokp"x"	Inconsistency biological parent partner child "x" between waves	0	No inconsistency	
		1	Inconsistency: biological parent partner	Status partner (biological parent) varies between waves

continued on next page

Table A.9, 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.10: List of flag variables to identify inconsistencies (*parent$*)

Variable	Label	Values	Value Labels	Description
parflag1	Number of children	0	No inconsistency	
		1	Number of children in question 2 and question 3 do not match	parsd32 (question 2) ≠ number of children indicated in childbiography (question 3) OR (parsd32 = -5,-2)
parflag2	Anchor information	0	No inconsistency	
		1	Anchor's year of birth in question 3 does not match anchor interview	(parsd19k1y ≠ doby (from anchor's interview))
		2	Anchor's status in question 3 does not match anchor interview	(parsd15k1 ≠ type)
parflag3	Age 2nd child	0	No inconsistency	
		1	Respondent <13 or >50/70 when bio-childbirth	(50<(parsd19k2y - pardoby)<13) if parsd15k2 = 1,2 & parsex = 2 (70<(parsd19k2y - pardoby)<13) if parsd15k2 = 1,2 & parsex = 1
parflag4	Age 3rd child	0	No inconsistency	
		1	Respondent <13 or >50/70 when bio-childbirth	*analogous to parflag3*
parflag5	Age 4th child	0	No inconsistency	
		1	Respondent <13 or >50/70 when bio-childbirth	*analogous to parflag3*
parflag6	Partner status	0	No inconsistency	
		1	No partner in question 12 but partner in question 48	(parcrn21i1 = 8 & parsd3 =1,2)

continued on next page

Table A.10, 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	No inconsistency	
		1	Ambiguous # of grandkids	pargc2i6=1 (no grandchildren) and at least (one valid value in pargc2i1-pargc2i5 or missing in pargc2i1- pargc2i5) for correction see variable grandparkids
parflag8	Parent's age	0	No inconsistency	
		1	Parent <12 years old at respondent's birth or	*see value labels*
		2	adoptive/stepparent Living parent 100+ years old	
parflag9	Support by non-existent people	0	No inconsistency	
		1	Mother	Mother not alive (parigr22 = 2)
		2	Father	Father not alive (parigr25 = 2)
		3	Children other than Anchor	Only one child (parsd32 = 1)
		4	Partner	No partner (parsd3 = 3)
		5	Siblings	No siblings (parnet17i2 = 0 & parnet17i1 = 0)
		6	Son-/daughter-in-law	No children with partner (parsd35k1 & parsd35k2 & parsd35k3 & parsd35k4 ≠ 1)
		7	Grandchildren	No grandchildren (pargc2i6 = 0)
parflag10	Household composition	0	No inconsistency	
		1	Children in hh (question 3) but single-hh (question 47)	(parhh = 0 & (parigr41k1 = 1 OR parigr41k2 = 1 OR parigr41k3 = 1 OR parigr41k4 = 1))

continued on next page

Table A.10, continued from previous page

Variable	Label	Values	Value Labels	Description
		2	Partner in hh (question 48) but single-hh (question 47)	(parhh = 0 & parsd3 = 1)
parflag11	Leaving home	0	No inconsistency	
		1	Left home before birth	(parigr9y < pardoby)
parflag12	Educational degree	0	No inconsistency	
		1	Multiple answer: highest degree kept	*see value labels*
		2	Multiple answer incl. other	
parflag13	Occupational degree	0	No inconsistency	
		1	No but also valid occupational degree	Any valid occupational degree provided & (parsd29i8 = 1)
parflag14	Mother's partnership status	0	No inconsistency	
		1	Anchor's answer doesn't match mother's answer	papa17i* != igr30 (from Anchor's interview)
parflag15	Father's partnership status	0	No inconsistency	
		1	Anchor's answer doesn't match father's answer	papa17i* != igr35 (from Anchor's interview)
parflag_cas	Current activity	0	No inconsistency	
		1019	Unemployed but full-time	
		1219	Unemployed but part-time	Multiple answers for current activity status that are not plausible
		1021	Fulltime but occupational/disability	
		1721	Marternal oder paternal leave or other/Retired, occupational disability	

Table A.11: Canges in codebooks

Codebook	Release/Date of Change	Description of Changes/Variables affected
Anchor 1	Release 2.0	Improved English translation (no revision with regard to content)
Partner 1	Release 2.0	Improved English translation (no revision with regard to content)
Partner 1	Release 3.1	psd19kxm, psd19kxy, psd14kxg, psd15kx: Correction of variable names psd10: Change in response categories
Anchor 2	Release 3.0	inc10, pa3: Item numbers changed Output of EHC-Screen H6: ehc18pX instead of ehc14pxi1
Partner 2	Release 2.0 2011/11/21	Major corrections in English version, (nearly) all variables revised
Child 2	Release 2.0 2011/11/21	cpcr4, cpcr5: filter changed
Parent 2	Release 3.0	Question numbers changed
Parent 3	Release 3.1	parigr41kx, parigr45a, parigr47a, parigr80a, pargc5g, parigr101ix, parigr12ix: Minor changes in variable labels parigr53a-59a, parigr60a-66a, parigr78a-79a, parigr80a-81a: Minor change in formulation of a response category and response code (6 -> 8) parigr103, parsd23i13, parsd23i21: Minor change in formulation of response categories parigr40p1, parigr40p3, parigr41p1, parigr41p3: Filter changed parpa17ix: Missing response category added
Parenting 3	Release 3.1	pcr3: Change in formulation of response categories