



pairfam Data Manual

Release 14.1

Josef Brüderl, Rüdiger Edinger, Felicitas Eigenbrodt,
Madison Garrett, Kristin Hajek, Michel Herzig, Renate
Lorenz, Philipp Schütze, Nina Schumann, Katharina
Timmermann

LMU Munich

With contributions from:

Philipp Alt, Svenja Geissler, Tita Gonzalez Avilés, Timo Peter,
Julia Reim, Martin Wetzel

December 2023

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

Cite as:

Brüderl, Josef, Rüdiger Edinger, Felicitas Eigenbrodt, Madison Garrett, Kristin Hajek, Michel Herzig, Renate Lorenz, Philipp Schütze, Nina Schumann & Katharina Timmermann (2023): pairfam Data Manual, Release 14.1. LMU Munich: Technical Report. GESIS Data Archive, Cologne. ZA5678 Data File Version 14.1.0, <https://doi.org/10.4232/pairfam.5678.14.1.0>

Contents

1	Introduction	5
1.1	Obtaining the data	5
1.2	Referencing the pairfam project	5
1.2.1	pairfam Release 14.1	5
1.2.2	Citation of older data versions	6
1.2.3	pairfam COVID-19 data	8
1.3	Quick Guide and Quick Start files	9
2	Overview of the data structure	10
2.1	Actors, survey instruments and data sets	10
2.2	System of sample and person identifiers	11
2.3	Variable names	13
3	pairfam Nuts and Bolts	15
3.1	Changes of survey mode	15
3.2	Selection of (sub-)samples	15
3.2.1	Which samples to use?	15
3.2.2	Changes in <i>cohort</i> variable due to refreshment in wave 11	16
3.3	Anchor data sets	16
3.3.1	Expected number of children in waves 1 and 2 (<i>frt6</i>)	16
3.3.2	Social network data in waves 2 and 4 (<i>net1px-net16px</i>)	17
3.3.3	Mode change (CAPI vs. CASI) after wave 1 (<i>per1_</i>)	17
3.3.4	Dependent interviewing and possible seam effects in the EHC	17
3.4	Alteri data sets	18
3.4.1	Matching grandparents to grandchildren in waves 2 to 7	18
3.4.2	Changes in sample selection for the child interview and questionnaire	18
3.4.3	Change in sample selection for the parent survey between waves 7 and 8	18
3.4.4	Date of interview in partner questionnaire	18
3.5	English or German labels?	18
3.6	Satellite Projects	18
4	Anchor data	20
4.1	Event History Calendar (EHC)	20
4.2	Data editing	21
4.2.1	Variable and value labels	21
4.2.2	Missing values	21
4.2.3	Preload variables for Dependent Interviewing (DI variables)	22
4.2.4	Data cleaning of variables from the Event History Calendar (EHC)	23
4.2.5	Coding open answers	26
4.2.6	Anonymity	26
4.2.7	Value checks	30
4.2.8	Filter checks	30
4.2.9	Consistency checks	31
4.2.10	English-language data	32

4.3	Generated variables and scales	32
4.4	Process-generated variables	59
4.5	Paradata	60
4.6	Weights	62
4.6.1	Design weights	63
4.6.2	Calibrated design weights	64
4.6.3	Best implementation of weights	65
4.7	Generated data sets	66
4.7.1	Anchor-partner episode data - biopart	68
4.7.2	Anchor-child panel/episode data - biochild	74
4.7.3	Anchor-activities episode data - bioact and bioact_rtr	79
4.7.4	Anchor-household panel data - household	83
4.7.5	Anchor-mobility panel data - biomob*	88
4.7.6	Anchor-parent panel data - bioparent	97
4.7.7	Multi-actor panel data - Overview_multi_actor	101
5	Partner data	105
5.1	Data editing	105
5.1.1	Variable and value labels	105
5.1.2	Missing values	105
5.1.3	Open answers	105
5.1.4	Anonymity	106
5.1.5	Value and filter checks	106
5.1.6	Consistency checks	106
5.1.7	English-language data	106
5.2	Generated variables and scales	106
5.3	Changes in answer categories	108
6	Parent data	110
6.1	Data editing	110
6.1.1	Deletion	110
6.1.2	Variable and value labels	110
6.1.3	Missing values	111
6.1.4	Filter checks	111
6.1.5	Consistency checks	112
6.1.6	Anonymity	112
6.1.7	English-language data	113
6.2	Generated variables and scales	113
6.2.1	Generated variables	113
6.2.2	Scales	114
7	Child data	116
7.1	Data editing	116
7.1.1	Variable and value labels	116
7.1.2	Missing values	116
7.1.3	Consistency checks	117
7.1.4	Coding open answers	117
7.1.5	Recoding of SDQ Items	117
7.1.6	Anonymity	118
7.1.7	Coding of cjob	118
7.1.8	Coding of ccor5	118
7.1.9	Value and filter checks	119
7.1.10	English-language data	119
7.2	Generated variables and scales	119
7.3	Sample selection for the child interview	122

7.4	Mode change (CAPI vs. CATI) in waves 12, 13, & 14	122
8	Parenting data	123
8.1	Parenting survey	123
8.1.1	Data editing	124
8.1.2	Generated variables and scales	126
8.2	Parenting u6 partner survey	127
8.2.1	Data editing	127
8.2.2	Generated variables and scales	128
8.3	PAYA survey	128
8.3.1	Data editing	128
8.3.2	Generated variables and scales	129
9	DemoDiff	131
9.1	Introduction	131
9.2	Data structure	131
9.3	Differences in detail	132
9.3.1	Variables excluded from DemoDiff	132
9.3.2	Additional variables in DemoDiff	133
9.4	Codebook	133
10	Step-up data	134
10.1	Data editing of transitional questionnaire	134
10.1.1	Variable and value labels	134
10.1.2	Missing values	135
10.1.3	Anonymity	135
10.1.4	Value checks	135
10.1.5	English-language data	135
10.2	Generated variables and scales	136
10.3	Identifying step-ups in their parents' anchor data	136
11	Refreshment sample	137
11.1	Question program	137
11.2	Retrospective relationship history	137
11.3	Generated variables, data sets, and scales	138
12	Changes in survey mode due to COVID-19	139
13	Changes to wave 14	141
13.1	Mode change in wave 14 and its consequences	141
13.2	Replacement of the Event History Calendar in the anchor survey	142
13.3	Further differences between the web-based and paper-based anchor questionnaires	145
13.4	Data-cleaning process	145
13.5	Further changes in wave 14	145
	Bibliography	146
A	Appendix	149
A.1	DemoDiff-pairfam differences	149
A.2	Changes between releases	155
A.3	Flag and tag variables	196
A.4	Main sample-refreshment sample differences	210
A.5	Additional variables web-based anchor questionnaire W14	211

1. Introduction

The aim of this manual is to facilitate work with pairfam data sets. In the following it is assumed that the reader has some basic knowledge of the pairfam design. If not so please read first the section on the pairfam design in the Technical Paper No. 01 (Brüderl et al. 2023).

1.1 Obtaining the data

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

1.2 Referencing the pairfam project

Receiving credit from data users is of vital interest to the pairfam project. Only with these references can we prove the scientific value of pairfam. Therefore, we kindly ask you to follow these citation rules in your publications:

1.2.1 pairfam Release 14.1

Refer to the study and the data version 14.1 by including the following (or a similar) phrase in your text:

Analyses are based on data from the German Family Panel (pairfam), release 14.1 (Brüderl et al. 2023). A detailed description of the study can be found in Huinink et al. (2011).

Add the respective bibliographic details to your list of references:

Brüderl, Josef, Sonja Drobnič, Karsten Hank, Franz J. Neyer, Sabine Walper, Christof Wolf, Philipp Alt, Irina Bauer, Simon Böhm, Elisabeth Borschel, Christiane Bozoyan, Pablo Christmann, Rüdiger Edinger, Felicitas Eigenbrodt, Madison Garrett, Svenja Geissler, Tita Gonzalez Avilés, Nicolai Gröpler, Tobias Gummer, Kristin Hajek, Michel Herzig, Renate Lorenz, Katharina Lutz, Timo Peter, Richard Preetz, Julia Reim, Barbara Sawatzki, Claudia Schmiedeberg, Philipp Schütze, Nina Schumann, Carolin Thönnissen, Katharina Timmermann, Martin Wetzel. 2023. The German Family Panel (pairfam). GESIS Data Archive, Cologne. ZA5678 Data file Version 14.1.0. doi.org/10.4232/pairfam.5678.14.1.0

Huinink, Johannes, Josef Brüderl, Bernhard Nauck, Sabine Walper, Laura Castiglioni & Michael Feldhaus. 2011. Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual framework and design. Zeitschrift für Familienforschung - Journal of Family Research 23: 77-101. <https://nbn-resolving.org/urn:nbn:de:0168-ss0ar-376463>

The text should also include an acknowledgement like this:

This paper uses data from the German Family Panel pairfam, coordinated by Josef Brüderl, Sonja Drobnič, Karsten Hank, Johannes Huinink, Bernhard Nauck, Franz J. Neyer, and Sabine Walper. The study was funded from 2004 to 2022 as a priority program and long-term project by the German

Research Foundation (DFG).

If you find the Scales Manual or Data Manual helpful, the authors would appreciate a reference:

Thönnissen, Carolin, Julia Reim, Svenja Geissler, Philipp Alt, Barbara Sawatzki, Simon Böhm, and Sabine Walper (2023): pairfam Scales and Instruments Manual, Release 14.1. LMU Munich: Technical report. GESIS Data Archive, Cologne. ZA5678 Data File Version 14.1.0. doi.org/10.4232/pairfam.5678.14.1.0

Brüderl, Josef, Rüdiger Edinger, Felicitas Eigenbrodt, Madison Garrett, Kristin Hajek, Michel Herzig, Renate Lorenz, Philipp Schütze, Nina Schumann & Katharina Timmermann (2023): pairfam Data Manual, Release 14.1. LMU Munich: Technical report. GESIS Data Archive, Cologne. ZA5678 Data File Version 14.1.0. doi.org/10.4232/pairfam.5678.14.1.0

A number of former pairfam colleagues contributed to this data manual. We highly appreciate the work of and would like to thank Simone Braun, Laura Castiglioni, Bernadette Huyer-May, Ulrich Krieger, Volker Ludwig, Katharina Lutz, Bettina Müller, Ulrike Müller, Jasmin Passet-Wittig, Klaus Pforr, and Mirte Scholten.

To cite the Codebooks and other materials without explicitly naming authors, please use "pairfam Group". For the Codebook for example of the anchor person survey of wave 1 (Release 14.1) the following suggested citation results:

pairfam Group (2022): Codebuch Ankerperson, Welle 1 (2008/2009), Release 14.1. GESIS Data Archive, Cologne. ZA5678 Data file Version 14.1.0., <https://doi.org/10.4232/pairfam.5678.14.1.0>

The suggestions for citation of the manuals and codebooks can also be found on the cover pages of the corresponding pdf documents.

1.2.2 Citation of older data versions

Release 14.0:

Brüderl, Josef, Sonja Drobnič, Karsten Hank, Franz J. Neyer, Sabine Walper, Christof Wolf, Philipp Alt, Irina Bauer, Simon Böhm, Elisabeth Borschel, Christiane Bozoyan, Pablo Christmann, Rüdiger Edinger, Felicitas Eigenbrodt, Madison Garrett, Svenja Geissler, Tita Gonzalez Avilés, Nicolai Gröpler, Tobias Gummer, Kristin Hajek, Michel Herzig, Renate Lorenz, Katharina Lutz, Timo Peter, Richard Preetz, Julia Reim, Barbara Sawatzki, Claudia Schmiedeberg, Philipp Schütze, Nina Schumann, Carolin Thönnissen, Katharina Timmermann & Martin Wetzel. 2023. The German Family Panel (pairfam). GESIS Data Archive, Cologne. ZA5678 Data file Version 14.0.0. doi.org/10.4232/pairfam.5678.14.0.0

Release 13.0:

Brüderl, Josef, Sonja Drobnič, Karsten Hank, Franz J. Neyer, Sabine Walper, Philipp Alt, Elisabeth Borschel, Christiane Bozoyan, Madison Garrett, Svenja Geissler, Tita Gonzalez Avilés, Nicolai Gröpler, Kristin Hajek, Michel Herzig, Rüdiger Lenke, Renate Lorenz, Katharina Lutz, Timo Peter, Richard Preetz, Julia Reim, Barbara Sawatzki, Claudia Schmiedeberg, Philipp Schütze, Nina Schumann, Carolin Thönnissen, Katharina Timmermann & Martin Wetzel. 2022. The German Family Panel (pairfam). GESIS Data Archive, Cologne. ZA5678 Data file Version 13.0.0. doi.org/10.4232/pairfam.5678.13.0.0

Release 12.0:

Brüderl, Josef, Sonja Drobnič, Karsten Hank, Franz J. Neyer, Sabine Walper, Philipp Alt, Elisabeth Borschel, Christiane Bozoyan, Madison Garrett, Svenja Geissler, Tita Gonzalez Avilés, Nicolai Gröpler, Kristin Hajek, Michel Herzig, Bernadette Huyer-May, Rüdiger Lenke, Renate Lorenz, Katharina Lutz, Lara Minkus, Timo Peter, Trang Phan, Richard Preetz, Julia Reim, Barbara Sawatzki, Claudia Schmiedeberg, Philipp Schütze, Nina Schumann, Carolin Thönnissen, Katharina Timmermann & Martin Wetzel. 2021. The German Family Panel (pairfam). GESIS Data Archive, Cologne. ZA5678 Data file Version 12.0.0., <https://doi.org/10.4232/pairfam.5678.12.0.0>

Release 11.0:

Brüderl, Josef, Sonja Drobnič, Karsten Hank, Franz J. Neyer, Sabine Walper, Philipp Alt, Christiane Bozoyan, Christine Finn, Renate Frister, Madison Garrett, Tita Gonzalez Avilés, Henriette Greischel, Nicolai Gröpler, Kristin Hajek, Michel Herzig, Bernadette Huyer-May, Rüdiger Lenke, Lara Minkus, Timo Peter, Julia Reim, Claudia Schmiedeberg, Philipp Schütze, Nina Schumann, Carolin Thönnissen, Martin Wetzel & Barbara Wilhelm. 2020. The German Family Panel (pairfam). GESIS Data Archive, Cologne. ZA5678 Data file Version 11.0.0., <https://doi.org/10.4232/pairfam.5678.11.0.0>

Release 10.0:

Brüderl, Josef, Sonja Drobnič, Karsten Hank, Bernhard Nauck, Franz J. Neyer, Sabine Walper, Philipp Alt, Christiane Bozoyan, Petra Buhr, Christine Finn, Madison Garrett, Henriette Greischel, Nicolai Gröpler, Kristin Hajek, Michel Herzig, Bernadette Huyer-May, Rüdiger Lenke, Lara Minkus, Bettina Müller, Timo Peter, Claudia Schmiedeberg, Philipp Schütze, Nina Schumann, Carolin Thönnissen, Martin Wetzel & Barbara Wilhelm. 2019. The German Family Panel (pairfam). GESIS Data Archive, Cologne. ZA5678 Data file Version 10.0.0., <https://doi.org/10.4232/pairfam.5678.10.0.0>

Release 9.1:

Brüderl, Josef, Sonja Drobnič, Karsten Hank, Johannes Huinink, Bernhard Nauck, Franz J. Neyer, Sabine Walper, Philipp Alt, Elisabeth Borschel, Christiane Bozoyan, Petra Buhr, Christine Finn, Madison Garrett, Henriette Greischel, Kristin Hajek, Michel Herzig, Bernadette Huyer-May, Rüdiger Lenke, Bettina Müller, Timo Peter, Claudia Schmiedeberg, Philipp Schütze, Nina Schumann, Carolin Thönnissen, Martin Wetzel & Barbara Wilhelm. 2018. The German Family Panel (pairfam). GESIS Data Archive, Cologne. ZA5678 Data file Version 9.1.0, <https://doi.org/10.4232/pairfam.5678.9.1.0>

Release 9.0:

Brüderl, Josef, Sonja Drobnič, Karsten Hank, Johannes Huinink, Bernhard Nauck, Franz J. Neyer, Sabine Walper, Philipp Alt, Elisabeth Borschel, Christiane Bozoyan, Petra Buhr, Christine Finn, Madison Garrett, Henriette Greischel, Kristin Hajek, Michel Herzig, Bernadette Huyer-May, Rüdiger Lenke, Bettina Müller, Timo Peter, Claudia Schmiedeberg, Philipp Schütze, Nina Schumann, Carolin Thönnissen, Martin Wetzel & Barbara Wilhelm. 2018. The German Family Panel (pairfam). GESIS Data Archive, Cologne. ZA5678 Data file Version 9.0.0., <https://doi.org/10.4232/pairfam.5678.9.0.0>

Release 8.0:

Brüderl, Josef, Karsten Hank, Johannes Huinink, Bernhard Nauck, Franz J. Neyer, Sabine Walper, Philipp Alt, Elisabeth Borschel, Petra Buhr, Laura Castiglioni, Stefan Fiedrich, Christine Finn, Madison Garrett, Kristin Hajek, Michel Herzig, Bernadette Huyer-May, Rüdiger Lenke, Bettina Müller, Timo Peter, Claudia Schmiedeberg, Philipp Schütze, Nina Schumann, Carolin Thönnissen, Martin Wetzel & Barbara Wilhelm. 2017. The German Family Panel (pairfam). GESIS Data Archive,

Release 7.0:

Brüderl, Josef, Karsten Hank, Johannes Huinink, Bernhard Nauck, Franz J. Neyer, Sabine Walper, Philipp Alt, Petra Buhr, Laura Castiglioni, Stefan Fiedrich, Christine Finn, Kristin Hajek, Michel Herzig, Bernadette Huyer-May, Rüdiger Lenke, Bettina Müller, Timo Peter, Veronika Salzburger, Claudia Schmiedeberg, Philipp Schütze, Nina Schumann, Carolin Thönnissen, Martin Wetzel & Barbara Wilhelm. 2016. *The German Family Panel (pairfam)*. GESIS Data Archive, Cologne. ZA5678 Data file Version 7.0.0., <https://doi.org/10.4232/pairfam.5678.7.0.0>

Release 6.0:

Brüderl, Josef, Karsten Hank, Johannes Huinink, Bernhard Nauck, Franz J. Neyer, Sabine Walper, Philipp Alt, Petra Buhr, Laura Castiglioni, Christine Finn, Kristin Hajek, Michel Herzig, Bernadette Huyer-May, Rüdiger Lenke, Bettina Müller, Timo Peter, Veronika Salzburger, Claudia Schmiedeberg, Elisabeth Schubach, Philipp Schütze, Nina Schumann, Carolin Thönnissen & Barbara Wilhelm. 2015. *The German Family Panel (pairfam)*. GESIS Data Archive, Cologne. ZA5678 Data file Version 6.0.0, <https://doi.org/10.4232/pairfam.5678.6.0.0>

Release 5.0:

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

Release 4.0:

Nauck, Bernhard, Josef Brüderl, Johannes Huinink & Sabine Walper. 2013. *The German Family Panel (pairfam)*. GESIS Data Archive, Cologne. ZA5678 Data file Version 4.0.0., <https://doi.org/10.4232/pairfam.5678.4.0.0>

Release 3.1:

Nauck, Bernhard, Josef Brüderl, Johannes Huinink & Sabine Walper. 2012. *The German Family Panel (pairfam)*. GESIS Data Archive, Cologne. ZA5678 Data file Version 3.1.0., <https://doi.org/10.4232/pairfam.5678.3.1.0>

Release 3.0:

Nauck, Bernhard, Josef Brüderl, Johannes Huinink & Sabine Walper. 2012. *The German Family Panel (pairfam)*. GESIS Data Archive, Cologne. ZA5678 Data file Version 3.0.0., <https://doi.org/10.4232/pairfam.5678.3.0.0>

1.2.3 pairfam COVID-19 data

For publication based on the pairfam COVID-19 data please cite the data set as well as the Technical Paper 15 (Walper et al. 2021) as follows:

Release 1.1.0:

Walper, Sabine, Barbara Sawatzki, Philipp Alt, Julia Reim, Claudia Schmiedeberg, Carolin Thönnissen & Martin Wetzel (2021). *The pairfam COVID-19 survey*. GESIS Data Archive, Cologne. ZA59589

Data file Version 1.1.0., <https://doi.org/10.4232/pairfam.5959.1.1.0>

Walper, Sabine, Barbara Sawatzki, Philipp Alt, Julia Reim, Claudia Schmiedeberg, Carolin Thönnissen & Martin Wetzel (2021). *The pairfam COVID-19 survey: Design and instruments. Release 1.1.* LMU Munich: pairfam Technical Paper 15, <https://doi.org/10.5282/ubm/epub.91997>

Release 1.0.0:

Walper, Sabine, Barbara Sawatzki, Philipp Alt, Julia Reim, Claudia Schmiedeberg, Carolin Thönnissen & Martin Wetzel (2020). *The pairfam COVID-19 survey. GESIS Data Archive, Cologne. ZA59589 Data file Version 1.0.0.*, <https://doi.org/10.4232/pairfam.5959.1.0.0>

Walper, Sabine, Barbara Sawatzki, Philipp Alt, Julia Reim, Claudia Schmiedeberg, Carolin Thönnissen & Martin Wetzel (2020). *The pairfam COVID-19 survey: Design and instruments. Release 1.0.* LMU Munich: pairfam Technical Paper 15.

1.3 Quick Guide and Quick Start files

The Quick Guide provides a brief introduction to the most important information and reference points for working with the pairfam data. Furthermore, the pairfam team developed several commented Stata do-files (Quick Starts) for an easy start. The Quick Starts show how common analysis problems can be solved with the pairfam data and how different pairfam datasets can be merged. Adapting the Quick Starts will be an easy way to start with your own analysis.

The following Quick Starts are available in the Scientific Use File in the directory “Syntax files”:

- “Sample Definition”: information on defining the analysis sample,
- “Weighting”: examples of how to use weights in pairfam and DemoDiff,
- “Matching”: introduction of the most common operations for merging separate pairfam data sets to combined multi-waves and/or multi-actor data sets,
- “Missings”: decoding of missing values,
- “Biopart EHA”: information on how to use the generated data set “biopart”,
- “Panel Analysis FE”: demonstration of an exemplary panel analysis.

A description on how to use pairfam data and the Quick Starts for analyzing wave 1 data are also included in the Technical Paper No. 2 (Brüderl et al. 2011).

2. Overview of the data structure

The main goal of the pairfam study is to provide researchers with data for the longitudinal analysis of family processes. To serve this goal, pairfam is designed as a *panel* and a *multi-actor study*. Thus, there are two main tasks for the data analyst: respondents' information over time has to be compiled and information of multiple respondents connected by family ties has to be matched. In this chapter we explain how the pairfam data have been organized to facilitate these two tasks.

2.1 Actors, survey instruments and data sets

The pairfam data are structured according to three principles: by wave, by actor, and by survey instrument. Accordingly, data set names consist of the actor (or topic), a wave suffix (\$), and potentially a suffix indicating the survey mode. Table 2.1 gives an overview of our data structure.

First, information of each wave is stored in separate cross-sectional data sets.¹ To facilitate longitudinal analysis we keep names of variables consistent across waves (see section 2.3). In order to identify the time of the survey, a wave identifier is included in each cross-sectional data set. The variable *wave* displays the number and year of the respective interview.

Second, as a multi-actor-study pairfam interviewed not only the main respondents (called anchors) but also several persons in a certain relation to them (called alteri). Starting from 2008 (wave 1), we conducted interviews with our anchors and (if available) their current partners. As of wave 2, we also collected data from their parents (until wave 8) and their children. In addition, we conducted three surveys on anchors' and their partners' parenting behavior: The parenting survey on children (since wave 2), the parenting survey of adolescents and young adults (PAYA, waves 9-13), and the parenting u6 partner survey (waves 11-13). Finally, respondents of the children's survey (and their current partners) could enter the main anchor survey (so-called stepups, since wave 4).

Third, additional surveys outside the panel structure or deviations from the usual panel design are indicated by suffixes. In wave 8, information on reasons for not having (further) children was collected with computer-assisted recorded interviewing (CARI). This information is also part of the Scientific Use File (as data set *anchor8_cari*, including English labels). A detailed description of the data is included in Technical Paper No. 10 (Gebel et al. 2017). In wave 9, CARI was used again to collect information on how parents deal with their children leaving home (question 225). The anonymized and transcribed answers are available on request.

Fourth, different factorial survey experiments were conducted in waves 10, 11, and 12, which are included in the datasets *partner10_vig*, *anchor10_vig*, *anchor11_vig*, and *anchor12_vig*. For more information see the chapter 3.6 Satellite Projects.

¹There are two main reasons why we did not opt for a long format solution. First, we do not want to force all users to start with a large data set including information for all waves and all variables. There are 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.

In wave 12, the COVID-19 pandemic interrupted the fieldwork of anchor and child interviews. Since face-to-face interviews were no longer possible, we had to change the survey modes to telephone and self-administered paper-and-pencil interviews. In wave 13, respondents and interviewers could choose between face-to-face and telephone interviews. To emphasize the differing survey modes in wave 12 and wave 13, child and anchor data sets each come in two versions: one for the CAPI and one for the CATI interviews. For more information, please see our chapter on the change in survey mode due to COVID-19.

Wave 14 was the last wave before pairfam merged with the Generations and Gender Survey (GGS) to establish FReDA – The German Family Demography Panel Study. In preparation, the data collection mode was adapted to the survey design of the FReDA panel: Most of the pairfam anchor respondents completed the questionnaire online for the first time. For a detailed description of this mode change and its consequences for the question program, see Chapter 13.

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

Table 2.1: Overview of survey instruments and resulting data sets

Survey	Start (and end)	Main sur- vey mode	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 part- ner	id, pid	parenting\$
Parent survey	2009/10 - 2015/16	PAPI	Anchor's biological (or adoptive) mother, bi- ological (or adoptive) father, stepmother, stepfather	mid, fid, smid, sfid	parent\$
Child survey	2009/10	CAPI	Anchor's child(ren) between 8 and 15 years, living with anchor	cid	child\$
PAYA survey	2016/17 - 2020/21	PAPI	Anchor, anchor's part- ner	id, pid	paya\$
Parenting U6 Partner survey	2018/18 - 2020/21	PAPI	Anchor's partner	id, pid	parentingU6- partner\$
Step-up survey	2011/12	CAPI/CASI	Anchor's child(ren) 16 years and older, living with anchor (and part- ners)	id, f_cid	stepup_ \...\$

2.2 System of sample and person identifiers

The pairfam study consists of various anchor samples, which can be identified by three variables: *cohort*, *demodiff*, and *sample*. First, the generated variable *cohort* displays the birth cohort from which our main respondents were drawn. Second, respondents from the integrated DemoDiff study can be identified by the *demodiff* variable. Lastly, pairfam included a refreshment sample from wave 11 (2018/2019). The variable *sample* categorizes respondents from all samples (i.e., pairfam base sample, DemoDiff sample, refreshment sample, step-up sample). For additional information on the

²See the “changes” tables in the Appendix, which document the changes in data sets over the releases.

pairfam samples, please refer to Technical Paper No. 01 (Brüderl et al. 2023).

On an individual level, each respondent was assigned a unique and invariant identifier according to the system described in Table 2.2. A 3-6 digit household number was assigned by *Kantar Public*³ in order to (re-)identify anchors (and step-ups) in each wave (variable *hhid*). The anchor identifier *id* adds three trailing zeros at the end of the household number. The variable *id* is also contained in each alteri data set. This allows easy matching of alteri information to anchor data (and matching any two data sets of the alteri surveys).

Table 2.2: Person identifiers of pairfam respondents

Respondent(s)	Person ID	(Range of) assigned IDs
Anchor	<i>id</i>	3-6 digit household identifier (<i>hhid</i>) · 1000
Anchor's partner	<i>pid</i>	<i>id</i> + 101, ..., 150 (only for current partners)
Anchor's child	<i>cidx</i>	<i>id</i> + 201, ..., 299 (for all children)
	<i>parentidkx</i>	<i>pid</i> of second biological parent of anchor's child(ren)
	<i>parentidstepkx</i>	<i>pid</i> of second biological parent of anchor's stepchild(ren)
Anchor's parents	<i>mid</i>	<i>id</i> + 301 for biological (or adoptive) mother,
	<i>fid</i>	<i>id</i> + 302 for biological (or adoptive) father,
	<i>smid</i>	<i>id</i> + 303, 305, ..., 399 for stepmother(s),
	<i>sfid</i>	<i>id</i> + 304, 306, ..., 398 for stepfather(s)
Anchor's siblings	<i>sibidx</i>	<i>id</i> + 401, ..., 499

The household identifier also forms the basis for the construction of all other person identifiers (see Table 2.2). For example, the identifier of an anchor's first partner (*pid*) adds the trailing number 101 at the end of the household number (only those partners get an *pid* who were in a relationship with the anchor at time of interview). The same *pid* was kept as long as the anchor still was in a relationship with this partner. New partners were numbered consecutively.⁴

Children are identified by the variable *cid*. Since Release 7.0, a child ID is available in the anchor data for all children. Note that *anchor1* and *anchor1_DD* (DemoDiff data) do not contain the variables *cidx*, as no child questionnaire exists in wave 1. If however, there is need for *cidx* in wave 1, the identifiers can be easily generated with the following (Stata) code: `for num 1/10: gen long cidX=id+200+X if !inlist(sd14kXg,-3,6,7)`. In the long format data sets with separate lines for each child (e.g., *child\$* or *biochild*), there exists only the variable *cid*. In the *anchor\$* data sets, the child IDs are reported on one line as *cid1*, *cid2*, This is denoted by the suffix (*x*).

In addition, some anchor variables about children and partners include identifying information in the variable name. Children and partners were numbered consecutively as part of the variable names (*_kx* and *_px*). All of the anchor's children were assigned a number (*k1*, *k2*, *k3*, ...) the first time they were mentioned. The established numbering of an anchor's children remained constant throughout all waves, even in case of decease. In contrast, the position of partners (*p1*, *p2*, *p3*, ...) as part of some variable names followed a rolling system and could be occupied by different partners in different waves. The rolling system works as follows: In wave 1, the current partner was assigned position *p0*. In wave 2, the preloaded partner from wave 1 received the position *p1*, new current partners are assigned to position *p2*, and partners between waves occupy positions *p3*-*p5*. As of wave 3, the auxiliary variable *hpnr* in the *anchor\$* data sets contains the running number of the current partner. Generally, information on partners was documented only for the current relationship status and up until one year after separation. The current partner can always be identified by the partner identifier *pid* which is stable throughout the panel.

³Kantar Public was formerly known as *TNS Infratest*.

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

The variables *parentidkx* contain the identification number of a child's second biological parent (the first is always the anchor). In case a child is an anchor's step child, *parentidstepkx* contain the identification number (*pid*) of the child's biological mother or father. In order to identify this parent as a former partner of the anchor respondent, the variables *parentidkx* and *parentidstepkx* are constant over all waves for each child. Note that these variables can differ from the variable *pid*, as the anchor's relationship to the second parent can result in separation or a new partnership may emerge. *pid* denotes the current partner of the anchor. The computation of these variables is based upon the latest version of the data set *biochild* (for further information see Chapter 4.7). The variables *parentidstepkx* are only available in the *biochild* data set.

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

However, the order of the anchor's siblings in the parent data is not identical to that in the anchor data. In the case of merging parent and anchor data sets to examine the anchor's siblings, siblings can be identified in both data sets by their gender and birth date. A detailed module covering sibling information is included in the anchor data for waves 5, 7, 9, 11, and 13. In wave 5, up to 17 siblings of each anchor respondent were recorded (variables *sib*p1-sib*p17*). If the anchor reported more than 4 siblings, more detailed questions were asked about a random selection of 4 siblings (for example *p1*, *p4*, *p5*, and *p7*). In waves 7, 9, 11, and 13, further questions on the 4 randomly-selected siblings were posed. The order of the siblings does not change within the anchor data, but the additional information on these 4 siblings are stored as *sib*p1*, *sib*p2*, *sib*p3* and *sib*p4*.

2.3 Variable names

The main principle in generating names of variables in the pairfam study is to maintain consistency across waves and across actors. First, consistency across waves is achieved by holding names of variable and values constant across waves. For all actors, questions that are asked repeatedly and that comprise the same concept, always carry the same variable name. In addition, if categories of answer lists or items change between waves, the numbers indicating the categories/items which no longer are valid are left out and new categories/items are added to the next free position of the answer list, i.e. to the end.⁵

Second, variable names are also constant across surveys of different actors. If equivalent information is collected for different actors, the names of the alteri variables correspond to the respective anchor variable. In general, a prefix identifying the alteri is added to the names of anchor variables. Thus, 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. Finally, the prefix *paya-* is used in the parenting survey on young adults.

The names of variables within the anchor data follow a system that consists 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. <i>sat-</i> for "satisfaction", and if necessary
a running number	e.g. <i>sat1</i> , <i>sat2</i> , ..., <i>sat6</i> to distinguish variables within areas.

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

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. <i>sat1i1</i> , <i>sat1i2</i> , 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. <i>doby</i> 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.

3. pairfam Nuts and Bolts

In this chapter we gather tips, tricks, and known problems of the pairfam data sets. Most users (even experienced ones!) will benefit from reading this chapter carefully.

3.1 Changes of survey mode

In March 2020 the COVID-19 pandemic necessitated a change in survey mode among some anchor and child respondents (from CAPI & CASI to CATI & PAPI). Accordingly, the anchor and child data of waves 12 and 13 are each provided in two separate data sets. For more details on the technical and methodological implications of this change, please refer to Chapter 12 of this data manual and to Technical Paper No. 16 (Bozoyan et al. 2021).

Whereas in waves 12 and 13, the interview mode had to be adapted because of the COVID-19 pandemic, in wave 14, the course was set for the integration of the pairfam anchor respondents into FReDA - The German Family Demography Panel Survey - by adapting the interviewer-administered questionnaire to a self-administered mode (CAWI or PAPI). In addition, a subsample was recontacted by an interviewer again to experimentally test whether the mode switch affected the data. Accordingly, the anchor data of wave 14 is provided in three separate data sets. For more details on this change and its consequences, please refer to Chapter 13 of this data manual and see the methods report of wave 14 (Brix et al. 2023) or Technical Paper No. 1 (Brüderl et al. 2023).

In case you want to use waves 12, 13, or 14 data it is a necessity to read these sources!

3.2 Selection of (sub-)samples

A sample identification variable has been added to all data sets as of Release 11.0. The variable *sample* identifies respondents from the different (sub-)samples. The values are “1 pairfam base sample”, “2 DemoDiff sample”, “3 Refreshment sample”, and “4 Step-up sample”. The following section provides some advice on how to deal with the Demodiff sample.

In addition, pairfam respondents are drawn from different cohorts (variable *cohort*). An additional cohort entered the panel survey as part of the refreshment sample starting from wave 11.

3.2.1 Which samples to use?

DemoDiff is a supplementary East German sample that has been integrated into the pairfam base sample data (for details on DemoDiff, see Chapter 9). Thus, if the user is passive, the DemoDiff cases enter the analysis beginning with wave 3. However, including DemoDiff data in a pairfam analysis is not trivial, since the additional sample interferes with the monotonic panel design. In the original pairfam data, all respondents enter at wave 1, and some gradually attrite from the panel. If users append the anchor\$ data sets, users need to be aware that some new respondents enter at wave 3. Further, users should take into account that DemoDiff respondents in wave t did not have the possibility to participate t times in the survey, but only $t-1$ times. Further, users must decide how to merge DemoDiff wave 1 data. Merging it with pairfam wave 1 data creates an artificial gap in

the panel structure (wave 2 information is missing). Merging it with pairfam wave 2 might create incompatibilities, as the contents of DemoDiff wave 1 and pairfam wave 2 differ. Finally, if one intends to do a weighted analysis, special weights must be used (for more details see section 4.6).

To avoid these complications, one can simply delete the DemoDiff data:

```
* Delete Demodiff sample:  
keep if demodiff==0
```

The refreshment sample and the step-up sample impose similar challenges. To avoid complications, one could simply use only the pairfam base sample:

```
* Keep only pairfam base sample  
keep if sample==1
```

This should work, if your interest is in long-term panel analyses of familial processes. However, if your substantive interest is in East-West differences, then you should also use the DemoDiff sample. If you want more power for your analyses you should include the refreshment sample. Including the refreshment sample is also necessary if your substantive interest lies on the youngest cohort (*cohort4*). Finally, if your focus is on the step-ups, you will have to use the step-up sample.

Further, if you want to avoid potential complications from mode changes and pandemic effects in waves 12 and 13 (see Chapter 12), you may restrict your analyses to waves 1 to 11:

```
* Keep only pairfam data not affected by the COVID-19 pandemic  
keep if wave <= 11
```

3.2.2 Changes in *cohort* variable due to refreshment in wave 11

The variable *cohort* indicates the anchor's birth cohort. The pairfam base sample is comprised of three birth cohorts: 1971 to 1973 (value "3"), 1981 to 1983 (value "2"), and 1991 to 1993 (value "1"). In wave 11, a refreshment sample of new anchor respondents entered the panel survey. Some of the respondents were drawn from two of the original cohorts (1981-1983, 1991-1993). In addition, a new, younger birth cohort from 2001 to 2003 was added, which is indicated by value "4". Please note that the order of the *cohort* values is thus not chronological. For more information please see the section on the generated variable *cohort* and the chapter on the refreshment sample.

3.3 Anchor data sets

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

In each wave we asked about the realistically expected number of children (*frt6*). In the first two waves, however, the wording of the question did not clearly state that we wanted to know about *additional* children. Thus, most respondents *with* children were confused, and seemed to have reported the *total* number of realistically expected children. Therefore, the reported number of children on variable *frt6* in waves 1 and 2 is in most cases too high for respondents with children. To avoid further confusion in subsequent waves, we reworded the question for respondents with children as of wave 3. In a first question we asked if the respondent realistically expects having additional children (*frt27*). In a second question, we asked how many additional children the respondent expects (*frt28*). The wording for respondents without children remained unchanged (*frt26*). For more details, see the section "Remark on variable flag_ *frt6* (wave 1 & 2)" in Chapter 4.2.

If you want to use variable *frt6*, we recommend using an adjusted version of it. For further information, see Technical Paper No. 04 (Buhr and Huinink 2014).

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

We collected ego-centered social network data in waves 2 and 4. These data are compromised by large interviewer effects, as most interviewers (presumably) entered no or only a few network persons in order to curtail the amount of questions. For this reason, reported network sizes are unrealistically small. A detailed analysis of the problem can be found in Brüderl et al. (2013).

Since most respondents probably have larger networks than reported, users should not interpret the absolute network size. However, multivariate analyses of effects on network size seem to be valid, as first experiences show. Additionally, analyses considering network person characteristics are still possible. However, interviewer effects should always be accounted for (e.g. by using multilevel models) when analyzing pairfam network data.

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

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

If wave 1 data on personality traits are to be used in analyzes, we suggest correcting them first. One correction method would be normalizing the anchor scores with respect to the partner scores. This method is described in the Technical Paper No. 06 (Sonntag et al. 2015), in Sonntag et al. (2014), as well as Mund et al. (2015). A second method for dealing with the problem would be to include a dummy variable for the first wave in panel regression models. The coefficient of the dummy would then capture the mode effect in wave 1.

3.3.4 Dependent interviewing and possible seam effects in the EHC

The pairfam Event History Calendar (EHC) implements dependent interviewing (DI), i.e. the last available information from the previous interview was displayed in the current interview for the first month of the EHC. This feature provides more accurate information by reducing overreporting at the seam between two consecutive waves.

In wave 3, an experiment was implemented in the education and employment calendar. We randomly selected 1000 wave 2 respondents, for whom we did not display preload information from the previous wave. Based on data from this experiment, Brüderl et al. (2017) found that respondents reported more status changes at the seam between the month of the wave 2 interview and the following month than did the control group. Transition rates were significantly increased at the seam for cases whose preloads had been deleted. However, a seam effect could still be observed when using dependent interviewing. One explanation might be that in contrast to the EHC, which covers partnerships and places of residence, the calendar for educational and occupational activities began with a list of activities that apply to the period since the last interview. Here, preloaded information was displayed only in a second step together with the selected activities. Respondents might thus have misclassified their status in the first step and not corrected this entry once preloaded information was additionally displayed in the following step. Continuing to fill out the calendar with the misclassified status then lead to (incorrect) transitions at the seam (Brüderl et al. 2017).

Data users are thus advised to control for the seam month when conducting an event-history analysis based on EHC activity data.

3.4 Alteri data sets

3.4.1 Matching grandparents to grandchildren in waves 2 to 7

Parent survey respondents were asked to answer several questions concerning one specific grandchild whose name was written on the cover sheet of the PAPI questionnaire. However, many respondents answered the questions in this module although no grandchild was selected.⁶ This has led to an overcoverage of grandchildren in the data. To identify whether or not the information given refers to a CAPI child, the *cid*-identifier in the parent data can be used. Additionally, the variable *parcorgc* indicates whether the child referred to by the grandparent is in fact the relevant anchor's child (see also Chapter 6).

3.4.2 Changes in sample selection for the child interview and questionnaire

Between waves 2 and 8 sample selection for the child interview and the parenting questionnaire changed slightly. For more information see sections 7.3 and 8.1

3.4.3 Change in sample selection for the parent survey between waves 7 and 8

Due to considerably low response rates in the parent survey, this was redesigned as "grandparent" survey for wave 8. The target population changed from "all (living) parents who are in contact with the anchor" in waves 2 to 7 to "parents who are in contact with the anchor and the anchor is living with at least one biological or adopted child in a shared household". Consequently, case numbers dropped between wave 7 (N=2,719) and wave 8 (N=627) as only grandparents were eligible. In wave 8, the question program focused less on anchor-parent relationships and more on focal child-grandparent relationships. If users are interested in analyzing parents in general, they should exclude wave 8 from these analyses and use only the data from waves 2 to 7. If, however, they are interested in analyzing grandparents, data from waves 2 to 8 can be used.

3.4.4 Date of interview in partner questionnaire

In the partner questionnaire, partners were asked to specify the date of filling out the questionnaire themselves. These information concerning the day, month, and year of the interview that are assessed in the PAPI mode were then automatically read and transferred into the variables *pintd*, *pintm*, *pinty*.

Due to incorrect responses from the partners or errors occurring during the reading process, these data may be incorrect for some cases. All users of these data are recommended to check the plausibility of the mentioned variables beforehand.

3.5 English or German labels?

Stata data files contain labels in both languages. 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 in separate folders. Please choose the preferred data from the directory (\English or \German).

3.6 Satellite Projects

The German Family Panel occasionally offered selected researchers the opportunity to approach anchor respondents themselves for additional qualitative and/or quantitative data collection. These Satellite Projects were carried out by the researchers themselves (e.g., planning, data collection, and documentation) and finally verified by the pairfam team. The following projects are part of the Scientific Use File:

⁶The selection process followed the same rules as the selection of the CAPI child (see Chapter 2.2).

- *The Dynamics of Implicit Motives in Intimate Relationships* by Sebastian Pusch, Felix D. Schönbrodt (PI⁷), Caroline Zygar-Hoffmann, and Birk Hagemeyer (PI) (Friedrich Schiller University Jena & Ludwig Maximilian University of Munich). Additional information and the data of this project are located in a separate folder.
- As part of the anchor interview of wave 10, a factorial survey experiment on the distribution of housework and paid work in partnerships was implemented. Respondents were asked to evaluate three experimentally varied hypothetical scenarios (so-called vignettes). The corresponding variables are part of the data set anchor10. In addition, the partners of the anchors were asked to evaluate exactly the same three vignettes as the corresponding anchors (see the data set partner10_vig). Both the anchor and the partner vignettes can be analyzed using the additional decoding data set anchor10_vig. This data set includes all vignette data in long format, i.e., each line represents one vignette. For more information on the vignette implementation and the data please refer to Technical Papers No. 14 (Düval and Auspurg 2020) and No. 20 (Düval and Auspurg 2022).
- In wave 11 the anchor interview included a factorial survey experiment on *infidelity* with three vignettes about hypothetical couples. The corresponding variables are part of the data set anchor11. For analyses they must be combined with the additional decoding data set anchor11_vig via the respondent id which includes the vignette data in long format. For more information about the vignette study, see Technical Paper No. 18 (Bozoyan and Schmiedeberg 2022).
- A third factorial survey experiment was conducted in the anchor interview of wave 12. The vignettes (three vignettes per respondent) were about *parental labor force participation and childcare after returning from parental leave*. The corresponding variables are part of the data set anchor12. For analyses they must be combined with the additional decoding data set anchor12_vig via the respondent id which includes the vignette data in long format. For more information about the vignette study, see Technical Paper No. 19 (Schober et al. 2022).

⁷Principal investigator.

4. Anchor data

Processing of the anchor data will be described in the following sections.

4.1 Event History Calendar (EHC)

Available from wave two to thirteen⁸, one main task of the Event History Calendar (EHC) was to collect the inter-wave information of the respondents on a monthly basis (Brüderl et al. 2017). Each anchor was asked to report what happened since her/his latest survey participation. The EHC gathered information on four life domains: educational and occupational activities, children, partners, and residence. For a better understanding how information is stored in variables, the following table summarizes the case of the partner domain (see codebooks for a full list of EHC variables):

Table 4.1: Structure of EHC partner variables

	Information on...	Partner 1 to 5	Month 1 to 32
	2 (relationship)		
ehc	3 (cohabitation)	p1, p2, p3, p4, p5	m1, m2, ..., m32
	4 (marriage)		
	...		

For example, think of a respondent's relationship status with partner 1 for the different months between the previous and the current interview. This information is stored in the variables `ehc2p1m1` to `ehc2p1m32`. Please note that month 32 always corresponds to the month of the current interview. The months before are "counted backwards" up to the month of the previous interview, which is indicated by the variable `ehcid`. For instance, if a respondent's previous interview was 12 months before the current interview, `ehcid` will be 20 (= maximum month 32-12).⁹ In this case, EHC-variables are only filled from month 20 up to month 32 (e.g., `ehc2p1m20-ehc2p1m32`), whereas all variables concerning the months before are set to missing value "-3". However, there is an exception to this rule: In wave 2, the maximum number of months is 18. Consequently, if a respondent's previous interview was carried out 12 months before, the variable `ehcid` takes value 6 (= maximum month 18-12).

Table 4.2: Example EHC variable

Month	1	2	[...]	18	19	20	21	22	[...]	30	31	32
<i>ehcid</i>			[...]			x			[...]			
<i>ehc2p1m_</i>	-3	-3	[...]	-3	-3	0	0	1	[...]	1	1	1

As EHC data processing is quite complex, the pairfam team recommends using generated variables or bio data sets (4.7) if possible.

⁸In wave 14, the Event History Calendar (EHC) was replaced with the *b*-variables (biography updates). For more information, see Section 13.2

⁹Note that the respondents may differ in their time points of the current and the previous interview. This may lead to different values for the *ehcid* variable.

4.2 Data editing

This section describes the editing of the anchor data. The main steps in processing the data are:

- labeling variables and values,
- defining missing values,
- producing preload information for the CAPI interview,
- cleaning of EHC variables (as of wave 2),
- coding open answers,
- anonymizing the data,
- checking value ranges,
- checking filters,
- checking consistency across answers and across waves; marking data inconsistencies,
- computing user-friendly variables and episode/panel data,
- producing English-language data sets.

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

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

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

4.2.1 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 question number in the CAPI questionnaire. The values of all variables were labeled according to the CAPI questionnaire.

4.2.2 Missing values

For all variables of the anchor data, we defined a set of missing codes, which were applied throughout (see Table 4.3). Negative values are typically missing values; only four generated variables have valid negative values (*kldb2010*, *isco08*, *mcs*, *pcs*).

- The missing values “-1 Don’t know” or “-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.
- The 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.

- Errors in the CAPI program, which erroneously guided respondents to the wrong questions in the interview are indicated by the missing code “-4 Filter error / Incorrect entry” as are incorrect data entries by the interviewers. In waves 12 to 14, “-4” was also assigned for PAPI mode, if the respondent incorrectly responded (or incorrectly did not respond) to a question.
- 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 the value “-6 Unreadable answer”.
- For generated data (variables and files), we used the value “-7 Incomplete data” to indicate cases where we lacked the information necessary to compute a valid value. For the generated variables *mcs* and *pcs* the value “-77 Incomplete data” is used.
- For waves 1 and 3, special missing codes indicate differences between the original pairfam questionnaire and DemoDiff. If a question from the original pairfam questionnaire was not part of DemoDiff, the corresponding variable was set to “-10 Not in DemoDiff”. DemoDiff Variables not included in pairfam are indicated by “-11 Not in pairfam”.
- In waves 12 and 13, anchor respondents interviewed via telephone due to COVID-19 who did not return the paper questionnaire (PAPI) received the missing value “-12 Non-response PAPI”.
- Please be aware that the values -10, -11, and -12 are also part of the classification scheme for *kldb2010* and *isco08*.
- System missings are only used for identifiers (e.g. *pid*, *cid*) and preload variables.

To decode all negative values in an anchor data set to system missings, we provide the Quick Start file *Missings.do*.

Table 4.3: 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
-77	Incomplete data (only for variables <i>mcs</i> and <i>pcs</i>)
-10	Not in DemoDiff
-11	Not in <i>pairfam</i>
-12	Non-response PAPI

4.2.3 Preload variables for Dependent Interviewing (DI variables)

In order to get more reliable responses in the CAPI, information from previous waves was preloaded to the CAPI interview and presented to the respondents from the second wave onwards. Generally, preload information was based on cleaned data from the previous wave. For respondents who skipped the previous wave, data from two waves before were used (as of wave 3). Note that respondents were not able to skip more than one wave in a row. Since previously reported information served as a framework for the current wave, the quality of the responses was expected to improve. Further, the preload variables were used for routing respondents through the interview.

The preloads or DI variables are contained in the anchor data sets as of wave 2. They can be found at the beginning of the data sets. The variables are prefixed by a d- followed either by a 1- to 3-digit number or by person identifiers such as pid, smid or sfid.¹⁰ 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 DI variables is contained in the anchor codebooks in the section “Preface”.

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

The Event History Calendar used since wave 2 gathered information on four life domains: educational and occupational activities, children, partners, and residence (for more information about the EHC, see 4.1). 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 (ehc19)*

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.

EHC data cleaning: children (vark)*

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

Children with asynchronous positions between the waves were corrected by relocating the falsely positioned children (and all *vark**) according to their original position in preceding waves. Before any moving of a child to another position it was checked whether this new position was vacant. Subsequently, 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 from a higher to a lower position to close the gaps.

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

We individually inspected all partnership biographies with suspicious entries such as deleted preloads,

¹⁰In the anchor data sets of wave 14, the preload variables are prefixed by a di- followed by numbers from 1 to 150. Only the variables d0 and dpid remain unchanged. For detailed information, see the anchor codebooks for wave 14.

new partners with identical or similar names as the ones of the previous wave, implausible short cohabitation or marriage spells, marriages of anchor respondents belonging to the youngest cohort, and differences between the month of the interview (month 18 in wave 2, month 32 in subsequent waves) 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 preloaded partner to indicate that the partnership with the partner from the previous wave still existed, at least at some time between the interviews.¹¹ In addition, we recoded partners who had been entered as current partners if it was obvious that they were partners from the previous wave. If new partners were entered by mistake or if partners from the previous wave seemed to be implausible, they were deleted. Regarding all corrections, the auxiliary variables *hp** which are part of *anchor\$* as of wave 3 were adapted accordingly.

EHC data cleaning: residence

The EHC residence section collects information on anchors' current residence and mobility between waves. It was also used to filter further questions concerning their current main and secondary household(s) at the time of the interview.

In wave 1, respondents were asked to report their current place(s) of residence. Respondents who reported more than one residence were asked to define their main residence. If more than two places of residence were reported, respondents were also asked to indicate their secondary residence, i.e. apart from their main residence where they spend most of their time.

As of wave 2, respondents were presented with their information from the previous wave as well as monthly information on all place(s) of residence since the previous interview. If respondents indicated still living in the same city (or cities) as in the previous wave, they were asked whether they had moved to another address in the reported city (cities). In case of an overlap of episodes (i.e. at least two consecutive months at the same residences), respondents were asked to indicate their main and secondary residence for each of the overlapping months. In wave 2, the CAPI program automatically placed information regarding the main residence at the time of the interview and the monthly information for this residence to the first position for the respective EHC variables (i.e. *ehc15p1*, whether the anchor currently lives there). If a secondary residence was reported, this information was automatically placed in the second position (*ehc15p2*). Please note that this is not the case as of wave 3. Now, the first position of the respective EHC variables (i.e. *ehc28p1*) relates to the main residence of the previous wave, which is not necessarily the current main residence. As of wave 4, respondents are only asked to name their main residence. After filling in the correct number of months spent at this residence, respondents are asked whether they have a secondary residence or not. Further questions regarding a secondary residence are asked later on, after having finished the EHC module. Thus, secondary residences are no longer included in the EHC data.

¹¹ The variable *tag_idntp* as part of *anchor2* marks some additional cases which have been identified after cleaning the data and generating the preloads for the CAPI interview of wave 3. Consequently, no recoding was done and this tag variable was created instead (see Table A.26).

While processing the resulting EHC variables of wave 2, we encountered several problems with the data. In wave 1, some interviews apparently contained information on all residences instead of simply the current residence. We preloaded this information regardless, as in most cases we were not able to ascertain whether the information was (in)correct. In wave 2, some interviewers appeared to have problems dealing with incorrect 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 was currently a second residence, this was always shifted to the second position *ehc15p2*. Other residences (e.g. former first wave or between-wave households) were shifted to positions three and higher.

However, the CAPI program did not generate an indicator variable for the main and secondary residence(s) from wave 1. Moreover, when more than one current residence was reported, the information regarding which is the main residence was not stored properly, as with information for overlapping episodes. Due to these issues, we analyzed case-by-case observations with a change of residence at the wave 1 seam (after the wave 1 interview month), or at the wave 2 seam (the month of the wave 2 interview, or a difference between the wave 2 interview month and the current status). In addition, we analyzed all interviews with four or more residences mentioned.

As of wave 3, data regarding residence is prepared as follows: First, we checked if residences were indicated in which the anchor had never lived by analyzing all months and current residence. If the residence in question was not preloaded, all respective EHC variables were set to the missing value -4. Furthermore, we merged two or more residences if the cities indicated had the same or similar name (for example München-Aubing and München-Schwabing) and neither of the residences was classified as a secondary residence. We did not do this for Berlin, however, because *ehc27p*i2* contains federal state information (Berlin east or Berlin west) which would be lost. In addition, 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 did not report living at a residence 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 residence in which the anchor had lived in during the previous months. If available, information from retrospective questions (*rtr**) were also used to replace gaps.

As of Release 4.0, four different episode data sets containing information on anchor mobility are available: *biomob_ehc* (residence information collected through the EHC), *biomob_rtr* (retrospective migration history since the age of 18), *biomob_rtr_parents* (moving out of the parental home). For further information, see Chapter 4.7. For more detailed analyses of respondents' mobility over time, we recommend using these generated data sets.

EHC data cleaning: household grid

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

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

In wave 4, household members' dates of birth could not be changed by the respondents. Therefore, the same person was entered again in some instances. This resulted in multiple identical persons. The problem was handled in this way: If similar persons existed, the case was checked in detail. If

the person was the same, we deleted newly entered, not preloaded persons (“-4”). If information on date of birth etc. differed, we used the more recent information. If inconsistencies of date of birth or relationship status existed, the information was set to “-5”. We filled in missing information if additional useful information had been given in previous waves. Relationships which occurred twice but could reasonably only occur once were set to “-5” or were deleted.

From wave 6 and retroactively until wave 3, falsely deleted ex-partners still cohabiting with the anchor were recreated in the household grid. Accordingly these ex-partners are mentioned twice in the anchor data, in partner and household EHC. For ex-partners with an entry in the household grid the relationship to the anchor (*ehc23pX*) was set to “21 Other”.

4.2.5 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).¹² 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.

4.2.6 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 *Kantar Public*. We also deleted the information on exact dates, i.e. the day components, for privacy concerns.

By these means, all string variables in the data set were finally transformed to numeric variables. Thus the *anchor\$* data sets contain no string variables. Valid answers to open questions were recoded to value 1 throughout. The variables affected by the procedures to ensure anonymity are shown in Table 4.4 along with the value labels.

Table 4.4: 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, 11
cla6eXo	Other: Lived with whom?	Other mentioned	2, 11
crn12kxi14o	Child x: Other, namely:	Other mentioned	2, 5, 9, 12
crn13kxi13o	Child care morning child x: Other, own entry	Other mentioned	2 - 13
crn14kxi13o	Child care afternoon child x: Other, own entry	Other mentioned	2 - 13
crn18kxo	Child x: Custody - other, namely:	Other mentioned	2,3,5,7,9,11,13
crn1kxo	Other place of birth child x	Other mentioned	2, 3, 5-13
crn46kxo	Child x: Custody arrangement not biol. child - other, open entry	Other mentioned	5, 7, 9, 11, 13

continued on next page

¹²For information on the country of birth and nationality of the anchor, as well as of his or her partner and parents, we computed new variables where we grouped open answers into additional categories (see Chapter 4.3).

continued from previous page

Variable	Variable label	Anonymous value label	Wave
crn59kxi8o	Waiving parental benefits child x: Other reason	Other mentioned	7, 9
crn61i5o	Reason for lawsuit: Other reason	Other mentioned	7, 9, 11, 13
crn70kxo	Child x: Current type of school, other open entry	Other mentioned	7 - 14
crn91kxi7o	Reason for move: Other specified, open entry	Other mentioned	9 - 13
crn93o	Handling of moving out child x	Other mentioned	9 - 10
d1/di1	Preload: Day of birth (Preload)	Day mentioned	2 - 14
d134-d143	City xst place of residence (Preload)	Residence mentioned	2 - 13
di125	City xst place of residence (Preload)	Residence mentioned	14
d14-d28/di20-di34	Name child x (Preload)	Name mentioned	2 - 14
d164-d173	Country xst place of residence (Preload)	Country mentioned	2 - 13
di127	Country xst place of residence (Preload)	Country mentioned	14
d176-d195	Name xst household member [main residence] (Preload)	Name mentioned	2 - 13
d218-d226	Name xst household member [second residence] (Preload)	Name mentioned	2
d267	Other education (Preload)	Yes	2 - 13
d274	Other type of employment (Preload)	Yes	2 - 13
d282/di142	Current employment activity: open-ended answer (Preload)	Occupation mentioned	2 - 14
d398	Current partner's day of birth (Preload)	Day mentioned	2 - 13
d59-d73	Day of birth child x (Preload)	Day mentioned	2 - 13
d506/di16	Name of marriage partner with whom no relationship at prev. wave (Preload)	First name mentioned	4 - 14
d8/di8	Name of current partner (Preload)	Name mentioned	2 - 14
d611-d614	Name of x-th randomly chosen sibling	Mentioned	7, 9, 11, 13
dobd	Day of birth	Day mentioned	all
ehc12kxo	Other parent name child x (EHC)	Name mentioned	2 - 13
ehc14pxi1	City place of residence x (EHC)	Residence mentioned	2
ehc14pxi2o	Country place of residence x (EHC)	Country mentioned	2
ehc19i16mxo	Other type of employment, open entry in month x (EHC)	Other mentioned	2 - 13
ehc19i16o	Other type of employment, open entry currently (EHC)	Other mentioned	2 - 13
ehc19i22mxo	Other type of unemployment, open entry in month x (EHC)	Other mentioned	2 - 13
ehc19i22o	Other type of unemployment open entry currently (EHC)	Other mentioned	2 - 13
ehc19i9mxo	Other education open entry in month x (EHC)	Other mentioned	2 - 13
ehc19i9o	Other education open entry currently (EHC)	Other type of training mentioned	2 - 13
ehc1pxn	Name partner x (EHC)	Name mentioned	2 - 13
ehc31o	Reason for deletion of partner from prev. wave, open entry	Other mentioned	5 - 13

continued on next page

continued from previous page

Variable	Variable label	Anonymous value label	Wave
ehc32o	Reason for deleting a spouse from prev. wave, open entry	Other mentioned	5 - 13
ehc20d	Day of birth of partner from previous wave (EHC)	Day mentioned	2 - 13
ehc22pxn	Name person x in household (EHC)	Name mentioned	3 - 13
ehc27pxi1	City place of residence x (EHC)	Residence mentioned	3 - 13
ehc27pxi2o	Country place of residence x (EHC)	Country mentioned	3 - 13
ehc7kxn	Name child x (EHC)	Name mentioned	2 - 13
ehc8kxd	Day of birth child x (EHC)	Day mentioned	2 - 13
ehc34kxo	Reason for deleting child x, open entry	Other mentioned	6 - 13
bpa18n	Name new partner	Name mentioned	14
bpa25pxn	Name partner x	Name mentioned	14
bcrn2kxn	Name child x (correction)	Name mentioned	14
bcrn2kxo	Further information on child x (open entry)	mentioned	14
bcrn4kxn	Name (new) child x	Name mentioned	14
bcrn13kxn	Name child x (PAPI version wave 14)	Name mentioned	14
bhc2i1	City place of residence	mentioned	14
bjob1i14o	Since previous wave: Other, not employed, open entry	mentioned	14
bjob2i6o	Currenty: Other type of job, open entry	mentioned	14
ftr13i14o	Reasons against child: Other Reason	Reason mentioned / Other mentioned	1, 2, 4-7, 10-12
hc1pxi1	Information x-th residence	Residence mentioned	1
hc1pxi2o	Country x-th residence	Country mentioned	1
hc8h1px	Main residence: Name person x	Name mentioned	1, 2
hc8h2px	Second Residence: Name person x	Name mentioned	1, 2
hc32i9o	Reasons for moving: Other family-related reasons (open entry)	Mentioned	8 - 13
hc32i17o	Reasons for moving: Other reasons (open entry)	Mentioned	8 - 13
hc27h2i1	Second residence: City place of residence	Residence mentioned	8 - 13
hc27h2i2o	Second residence: Country, open entry	Country mentioned	8 - 13
hcp1i1	Main residence partner	Residence mentioned	2 - 14
hcp1i2o	Country main residence partner	Country mentioned	2 - 14
hcp6i1	Second residence partner	Residence mentioned	3, 5, 7, 9, 11, 13
hcp6i2o	Country second residence partner	Country mentioned	3, 5, 7, 9, 11, 13
igr1d	Day of birth biological mother	Day mentioned	1
igr2d	Day of birth biological father	Day mentioned	1
igr3o	Other country of birth mother	Country mentioned	1, 11
igr4o	Other country of birth father	Country mentioned	1, 11
igr73i12o	Citizenship adoptive mother: Open entry other country	Other mentioned	3, 11
igr74o	Open entry country of birth adoptive mother	Other mentioned	3, 11
igr75o	Open entry (highest) level of school education of mother	Other mentioned	3, 11

continued on next page

continued from previous page

Variable	Variable label	Anonymous value label	Wave
igr77i12o	Citizenship adoptive father: Open entry other country	Other mentioned	3, 11
igr78o	Open entry country of birth adoptive father	Other mentioned	3, 11
igr79o	Open entry (highest) level of school education of father	Other mentioned	3, 11
igr82i12o	Citizenship mother's partner: Open entry other country	Other mentioned	3 - 13
igr83o	Open entry country of birth mother's partner	Other mentioned	3 - 13
igr85i12o	Citizenship father's partner: Open entry other country	Other mentioned	3 - 13
igr86o	Open entry country of birth father's partner	Other mentioned	3 - 13
int12o	Open entry suggestions or comments on the interview	Mentioned	3 - 13
int4i5o	Other persons, namely:	Other person mentioned	1 - 13
int9o	Reason partner won't participate	Reason mentioned	1
int14o	Third person interference in CASI module	Other, namely	10 - 13
int16o	Other reason for discrepancy between EHC residence and contact add.	Other, namely	10 - 13
int17o	Anchor's home: Living environment information available, open entry	Mentioned	11
int18o	Anchor's home: Type of building, open entry	Mentioned	11
job1/job20o	Current occupation	Occupation mentioned	all
mig1i12o	Country citizenship	Citizenship mentioned	1, 11
mig4o	Other country of birth	Country mentioned	1, 11
mig6i12o	Country other citizenship mother	Citizenship mentioned	1, 11
mig7i12o	Country other citizenship father	Citizenship mentioned	1, 11
netpxn	Name person x: Complete list name generator	Name mentioned	2, 4
pa42i4o	Other type of online dating, open entry	Other mentioned	9 - 13
pa44o	Relationship start connected to specific event, open entry	Other mentioned	9
rtr18kxd	Day of death child x	Day mentioned	1
rtr1pxn	Name partner x	Name mentioned	1, 11
rtr23hx	Residence x	Residence mentioned	3
rtr24hxo	Open entry country of residence x	Other mentioned	3
rtr31i9o	Since 18th birthday: Open entry other education	Other mentioned	3
rtr35i7o	After school until W1: Open entry other type of employment	Other mentioned	3
sd14kxn	Name child x	Name mentioned	1
sd17kx	Name other parent child x	Name mentioned	1, 11
sd19kxd	Day of birth child x	Day mentioned	1

continued on next page

continued from previous page

Variable	Variable label	Anonymous value label	Wave
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 - 14
sd4n	Name current partner	Name mentioned	1
sdp10i13o	Partner other type of employment	Occupation mentioned	all
sdp10i14o	Other, not employed	Occupation mentioned	14
sdp10i22o	Partner open entry other education	Other mentioned	3 - 13
sdp17o	Open entry other level of school education	Other mentioned	3 - 14
sdp1d	Day of birth current partner	Day mentioned	1 - 13
sdp2i12o	Country of other citizenship partner	Citizenship mentioned / Other mentioned	all
sdp5	In which country does current partner live	Country mentioned	1 - 13
sdp6o	Other country of birth current partner	Country mentioned	all
sdp25o	Other country of birth mother of partner, open entry	Country mentioned	all
sdp26o	Other country of birth father of partner, open entry	Country mentioned	all
sep5kxo	Other custody of child x before separation	Other mentioned	2
sep7kxo	Other: Current custody of child x	Other mentioned	2, 3
sex21o	Self-assessment: sexual orientation	Mentioned	2, 3
sib2pxn	Name sibling x	Name mentioned	5, 11
sib31px	Name of deceased sibling	Name mentioned	7, 9
sin8i4o	Use of other, open entry	Other mentioned	5 - 13

4.2.7 Value checks

As documented in the anchor questionnaire, a number of value checks were included in the CAPI program and therefore already conducted during the interview. Additionally, we checked for each variable 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.

4.2.8 Filter checks

Filters of every variable were checked as documented in the CAPI questionnaire (in wave 12 and 13 also CATI and PAPI questionnaire; in wave 14 two filter checks were conducted (web-version (CAPI/CAWI) and PAPI questionnaire)), and the missing code -4 was assigned in case of 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 -4.¹³ On the other hand, values were classified as data errors if they were incorrectly entered by the interviewer. In these cases, we also assigned the value -4. In checking the filters, we proceeded in the order in which questions were posed in the interview to ensure that all filter errors would be detected.

¹³If a question is asked despite the filter, the resulting data might be of low quality in many instances, often because the question does not make sense to the respondent. Furthermore, it is often not only hard to determine whether the answers actually are useful, but it also requires considerable effort in terms of data inspection and documentation. Although we might delete useful information in some cases, we nonetheless decided to always recode these cases to -4.

4.2.9 Consistency checks

We conducted various checks to identify logically impossible or empirically implausible combinations of values on two or more variables.¹⁴ Generally, we assigned the missing value -5 for inconsistent cases (see the section on missing value 4.3). However, in some cases it was not possible to mark an inconsistency by assigning -5, because we it was unclear 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. Table A.26 describes the flag variables in detail.

As of wave 2, we computed tag variables to indicate inconsistencies over time, i.e. over waves. These relate to the anchor's, the current partner's, and the children's sex and date of birth. All of these variables were preloaded and should be time-constant. The tag variables identify any inconsistent cases. Furthermore, the indicators whether the anchor or the anchor's partner is the biological parent of a child is tagged if the information from the previous and the current wave are contradictory. Table A.27 lists all of the generated tag variables that are part of the anchor data sets.

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 might be possible to assign plausible values rather than to exclude all inconsistent cases right away. The users themselves have to make this decision. In the case of tagged values marking sex and date of birth, the respective generated identifiers explained in section 4.3 should be used.

Remark on variable flag_frt6 (wave 1 & wave 2)

In wave 1 and 2, 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?" (*frt6*). 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 the previous question 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 as to clarify our intentions. 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, the problem from wave 1 was still not solved. Some respondents still reported the total number of children they were thinking of having.

Thus, In wave 1 and 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.

¹⁴ This happened in addition to a number of checks that had already been implemented in the CAPI program (see anchor codebooks).

When analyzing variable *frt6* with this flagged subset of the sample, please proceed with caution. There are different options for analyzing these data:

1. Consider using the question on the intention to have a child within the next two years (*frt7*) for your analysis instead of *frt6*.
2. Use *frt6* only for childless respondents who are not pregnant (unflagged values). For these cases there should be no problem with this question.

Note that *frt6* is also used to filter some of the following questions. Respondents reporting false values on this variable may also have answered too many or too few of those other questions.

As of wave 3 we constructed the questionnaire in the way that *frt6* was divided into three questions. In question *frt26* respondents without children were asked how many children they will have realistically. Furthermore, to rule out wrong answers a filter variable was integrated (*frt27*). Question *frt27* asked pregnant respondents or respondents who already had children if they think that they will have additional children. Only if they indicated that they will have additional children were they asked how many additional children they plan to have (*frt28*). Therefore *flag_frt6* is not relevant for wave 3 and subsequent waves.

Remark on variable flag_frt (wave 6)

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

Remark on variable flag_igb (as of wave 2)

In the module on intergenerational relations (IGB), we asked respondents about parents whom they had 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). We decided to consider adoptive parents as such only if they lived with the anchor before the age of 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 provide a flag variable indicating these parent combinations. Users have to decide whether or not to use these cases for analysis.

4.2.10 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. All variable labels and value labels were translated according to the wording of the English version of the anchor codebook.

4.3 Generated variables and scales

In order to facilitate data analysis and to enhance comparability of results, the pairfam team produced a number of variables that are of interest to many research projects. Table 4.5 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. Thereby, we try to combine syntaxes for several waves if possible. Users are invited to adapt the syntax to their special

research needs. Before using any of the generated variables we strongly advice users to always check whether the respective generating procedures meet their specific needs. Please note that there is a common missing value “-7 Incomplete data” encompassing the original missing codes -1, -2, -4, -5, and -6. Changes in the syntax for generated variables between different releases are documented in Table A.7 to Table A.24.

Furthermore, Table 4.21 displays all scales for the anchor data from the available waves. These variables are not part of the delivered data sets, but can be generated by users themselves. The corresponding syntax files (available for Stata and SPSS) are provided as part of the Scientific Use File. For additional information regarding these scales, please refer to the scales manual (Thönnissen et al. 2023).

Table 4.5: List of generated variables included in data sets *anchor\$*

Construct	Variable name
Generated identifiers sex (anchor, partner, children)	sex_gen, psex_gen, k*sex_gen
Generated identifiers date of birth (anchor, partner, children, mother incl. adoptive mother, father incl. adoptive father, stepmother, stepfather)	dob*_gen, pdob*_gen, k*dob*_gen, mdob*_gen, fdob*_gen, smdob*_gen, sf-dob*_gen
Age (anchor, partner, mother incl. adoptive mother, father incl. adoptive father, stepmother, stepfather)	age, page, mage, fage, smage, sfage
Age of children	k1age, ..., k10age
Age of anchor's youngest child living with anchor (in months)	ykage
Position of anchor's youngest child	ykid
Age of anchor's youngest CAPI child in months	ykagecapi
Position of anchor's youngest CAPI child	ykidcapi
Birth cohort	cohort
Country of birth (anchor, partner, mother, father, adoptive mother, adoptive father, stepmother, stepfather, mother of partner, father of partner)	cob, pcob, mcob, fcob, amcob, af-cob, smcob, sfcob, pmcob, pfcob
1st/(2nd/3rd/4th) nationality (anchor, partner, mother, father, adoptive mother, adoptive father, stepmother, stepfather)	nat*, pnat*, mnat*, fnat*, amnat*, afnat*, smnat*, sfnat*
Anchor's ethnicity	ethni
Anchor's migration status	migstatus
Partner's ethnicity	pethni
Partner's migration status	pmigstatus
Relationship status	relstat
Marital status (anchor, partner)	marstat, pmarstat
Number of previous partners	np
Number of previous partners with whom anchor cohabited	ncoh
Number of previous marriages	nmar
Months since anchor and current partner got to know each other	meetdur
Duration of current relationship, cohabitation and marriage	reldur, cohabdur, mardur
Anchor's sexual orientation	homosex, homosex_new
Anchor and/or partner infertile	infertile

continued on next page

continued from previous page

Construct	Variable name
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)	fmrdr
Partner lives in household (main residence)	pmrd
Number of children living in household (main residence)	childmrd
Number of others living in household (main residence)	othmrd
Household composition (main residence)	hhcomp
Enrollment in school or vocational qualification at time of interview (anchor, partner)	enrol, penrol
Highest school degree attained at time of interview (anchor, partner, mother, father)	school, pschool, mschool, fschool
Highest vocational degree attained at time of interview (anchor, partner, mother, father)	vocat, pvocat, mvocat, fvocat
ISCED-97, International Standard Classification of Education, no students (anchor, partner, mother, father)	iscd, piscd, miscd, fiscd
ISCED-97, International Standard Classification of Education, including students (anchor, partner)	iscd2, piscd2
CASMIN classification of educational attainment (1999) (anchor, partner, mother, father)	casmin, pcasmin, mcasmin, fcasmin
Years of schooling / vocational qualification (anchor, partner, mother, father)	yeduc, pyeduc, myeduc, fyeduc
KIDB classification of occupation	kldb2010, kldb1992
ISCO classification of occupation	isco08, isco88
Erikson-Goldthorpe-Portocarero class schema (EGP)	egp
Int. Socio-Economic Index of Occupational Status (ISEI)	isei
Standard Int. Occupational Prestige Scale (SIOPS)	siops
MPS occupational prestige score	mps
Net equivalence income according to GCEE	hhincgcee
Net equivalence income according to modified OECD scale	hhincoecd
Number of persons aged under 14 main residence	npu14mr
Number of persons aged 14 and older main residence	npo14mr
Net income (open and estimated information combined) (personal, household)	incnet, hhincnet
Consumer price index acc. to Nat. stat. off. (baseline 2010)	cpi
Current primary and secondary activity status (anchor, partner)	casprim, cassec, pcasprim, pcassec

continued on next page

continued from previous page

Construct	Variable name
Labor force status (anchor, partner)	lfs, plfs
Currently living in East Germany	east
Summary score physical and mental health	pcs, mcs

Generated identifiers sex - sex_gen, psex_gen, k*sex_gen

Due to measurement errors the value of the sex variable can differ between waves (or across the alteri datasets). Therefore, we have generated best solution variables for the anchor, the partner, and the children (*sex_gen*, *psex_gen*, *k*sex_gen*).

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

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

In order to solve the problem of conflicting information from various sources regarding respondents’ date of birth, we have generated best solution variables for the month and the year of birth of the anchor, partner, parents (incl. adoptive parents), stepparents, and children (*dob*_gen*, *pdob*_gen*, *mdob*_gen*, *fdob*_gen*, *smdob*_gen*, *sfdob*_gen*, *k*dob*_gen*).

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

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

The variables *age*, *page*, *mage*, *fage*, *smage*, *sfage*, and *k*age* contain the anchor’s, partner’s, parents’ (incl. adoptive parents’), stepparents’, and children’s age, respectively. These variables are based on the generated date of birth variables (see above). The age values were calculated by subtracting the 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 (*dobm_gen < intm*), the age variable was reduced by 1. If the information on the month of birth was missing, only the year of birth information was used. In cases of ambiguous seasonal information on the month of birth, random values were imputed for the month. If a specific alteri does not exist, the code “-3 Does not apply” was given. If the information necessary to derive the age was not available, the code “-7 Incomplete data” was given. The variables were generated by running the do-file *age.do*.

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

The variables *ykage* and *ykagecapi* indicate the age of the anchor’s youngest child (respectively the youngest CAPI child) living with the anchor. The variables *ykid* and *ykidcapi* contain the position of the youngest child/CAPI child corresponding to the variables *varkX* in the anchor data sets. Please note that there can be new CAPI children in each wave, so the youngest CAPI child might be a different one in the following wave. The variables were generated by running the do-file *age.do*. They

are provided from Release 4.0 on. Please note that *ykagecapi* and *ykidcapi* are not relevant for wave 1. Furthermore, *ykidcapi* is not relevant for wave 2. Since there was only one CAPI child in each household in wave 2, the variable *capikid* already indicates the position of the youngest CAPI child.

Birth cohort - cohort

The variable *cohort* indicates to which of the four birth cohorts 1971-1973, 1981-1983, 1991-1993, and 2001-2003 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, *Kantar Public* recontacted the respondents for clarification. The methods report of wave 1 (Suckow and Schneekloth 2009) states that the gross base 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 base sample and also remain in the net sample. These respondents were assigned to the corresponding birth cohorts, e.g., a respondent with year of birth 1970 was assigned to the birth cohort 1971-1973. Since the birth cohort is a time-constant variable it is fixed throughout the waves except for the former CAPI children. From wave 4 to wave 13, every year some former CAPI children entered the main questionnaire and became anchors themselves (for details, see Chapter 10). Therefore two additional categories were added. We differentiate between former CAPI children that had their first interview (category “0 Former capikid first interview”) and former CAPI children that have participated in the anchor survey before (category “9 Former capikid re-interview”). In wave 11, a refreshment sample of the birth cohorts 1991-1993, 1981-1983, and 2001-2004 was added (for more details, see Chapter 11). The values and value labels are shown in Table 4.6. The variable was generated by running the do-files *cohort*.do*.

Table 4.6: Values and labels of variables *cohort*

Value	Label
0	former capikid first interview
1	1991-1993
2	1981-1983
3	1971-1973
4	2001-2003
9	former capikid re-interview

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

These variables contain information on the country of birth of the anchor, his or her alteri (partner, mother and father, adoptive mother/father, stepmother/stepfather) and the partner’s mother and father. Information on the alteri and the partner’s parents was reported by the anchor. Open answers specifying a country not contained in the corresponding answer list were recoded and grouped into larger geographical or political regions (values “13” to “23”). Table 4.7 lists values and labels for all reported countries and regions.

As country of birth is a time-constant characteristic, the anchor’s as well as his/her parents’ values (*cob*, *mcob*, *fcob*) are based on information gathered in wave 1. For the refreshment sample, information on the anchor’s as well as his/her parents’ country of birth were collected in wave 11. The country of birth of adoptive parents as well as step parents is collected for the first time in wave 3. Accordingly, their values (*amcob*, *afcob*, *smcob*, *sfcob*) were generated for the first time in wave 3 and are carried over to later waves. If a new step mother/step father was reported in the current wave, we use *igr86* and *igr83* from the same wave to compute *smcob/sfcob*. If the partner mentioned in the previous wave is still the current partner, the generated variables *pcob*, *pfcob* and *pmcob* are copied from the previous wave. If the current partner is not the same as in the previous wave, we use *sdp6* and *sdp6o* from current wave to compute *pcob*, *sdp25* and *sdp25o* to compute *pmcob*, and *sdp26*

and *sdp260* to compute *pfcob*. If there is insufficient information in the current wave, information from the previous wave is used if available. If a previous partner or step mother/step father returns to the panel, their country of birth is asked again. For example, if a partner was recorded in waves 2 and 4, but not in wave 3, the anchor is asked again in wave 4 about their partner's country of birth. For some cases, this lead to inconsistencies between waves (i.e., a different country of birth for the same partner in waves 2 and 4), which we left uncorrected. Due to the shortening of the questionnaire in wave 14, no information on step parents are available in this wave. Thus, the variables *smcob/sfcob* are not available for this wave. The country of birth of the partner's parents was collected for the first time in wave 5, and information on the step-up respondents was taken from the *stepup_anchor* data sets. Information on the country of birth of the step-ups' parents was merged via *f_cid* (former child id) from their parents anchor data from the previous wave. Due to privacy concerns, the syntax which generates these variables cannot be published.

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

These variables contain information on the nationality of the anchor and his or her alteri (partner, mother, father, adoptive mother/father, step mother/step father). Information on the alteri was reported by the anchor. From wave 2 onwards, nationality information was gathered only for new partners. Additionally, in wave 3 the nationality of the adoptive mother/father as well as step parents was reported. Multiple nationalities per person are possible, which are stored in the variables **nat1*, **nat2*, etc. The highest number of multiple nationalities reported within a wave determines the number of variables that have been used for storing in variables. For example, in wave 10 mothers had two nationalities at most, while in wave 11 a maximum of four nationalities was reported. This leads to the variables *mnat1* and *mnat2* in wave 10, and additionally *mnat3* and *mnat4* in wave 11. Note that the order of storage does not convey any special meaning; **nat1* always contains the nationality with the lowest value (values are listed in Table 4.7). Therefore, 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"). In wave 14, no information on step parents are available in the anchor data set. Thus, the variables *smnat** and *sfnat** are not available for this wave. The nationality of step-up respondents was taken from the *stepup_transition_anchor* data sets. Information on the step-up parents' nationality was merged via *f_cid* (former child id) from their parents anchor data from the previous wave. As with country of birth, we do not publish the do-file to compute these variables.

Ethnicity and migration status of anchor and partner- *ethni*, *migstatus*, *pethni*, *pmigstatus*

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)"¹⁵, "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", and "2nd generation". As these variables are time-constant, the computation is based on anchor data from wave 1 and the variables are copied for later waves. The same procedure applies to the refreshment sample, however the information is based on the data from wave 11. Ethnicity and migration status are generated in the respective wave for step-up respondents only. As of wave 5, we also computed the partner's ethnicity *pethni* and migration status *pmigstatus* using the country of birth of the partner's parents (generated variables *pmcob* and *pfcob*). The syntax used to compute the variables is contained in Stata do-files *migration*.do* for the respective wave.

¹⁵Note that for step-up respondents and partners, this category could not be generated, as the "Aussiedler" status (*mig3*) was not asked for these cases.

Table 4.7: Values and variable labels for **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

Relationship status - *relstat*

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

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 mainly from the partner questionnaire. If no partner information

was available, information from the anchor interview was considered.¹⁶ The categories are: “never married”, “married/civil union”, “divorced/dissolved civil union” and “widowed/surviving partner in civil union”. Please note that the category “married/civil union” also encompasses married anchors who are separated (unlike the generated variable *relstat*). 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 Stata do-files *marstat*.do*.

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

The variables *np*, *ncoh* and *nmar* cover information on the number of previous, i.e. already terminated partnerships, cohabitations, and marriages. Note that the variable *ncoh* captures the number of previous partners with whom the anchor cohabited, not the total number of cohabitation episodes (i.e. one episode for each previous partner at most). Furthermore, the variable *np* counts the same previous partner only once although there has been a break or another partner in between. That is why the total number of previous partners of the current wave can be smaller than the total number of the wave before. The value “-7 Incomplete data” for *ncoh* and *nmar* indicates that respondents did not answer the relevant questions (“-1” or “-2”). The computation of these variables is based upon the latest version of the data set *biopart* (for further information, see Chapter 4.7). In contrast to previous waves, the data set *biopart* was not generated for wave 14 and thus the variables *np*, *ncoh* and *nmar* could not be generated for this wave either. The do-files *npart*.do* include the syntax to compute these three variables.

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

The variable *meetdur* is a measure of the number of months that have passed since the anchor and his or her current partner got to know each other. The code “-7 Incomplete data” was assigned if the anchor did not provide the date of the first meeting with his or her current partner (“-1” or “-2” for *pa2m* and *pa2y*).¹⁷ 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*. Due to this, it may be possible that the duration of a long-term relationship had been reset even if there was only a short disruption. You have to make up your mind whether this approach is appropriate to your issue or any other strategy should be applied. Please pay attention to the fact that the computation of these variables (see do-files *durpart*.do*) employs variables included in the data set *biopart* (see Chapter 4.7).

Anchor’s sexual orientation - *homosex*, *homosex_new*

These variables contain information on whether or not the anchor is homosexual in a specific wave. Note that these variables are time-varying and can change from wave to wave.

From these variables one might construct a time-constant indicator of homosexuality by, for instance, defining those as homosexual, who are defined as homosexual in at least one wave.

homosex

The variable *homosex* is built from the sex of the current partner, sexual contacts and/or sexual orientation. In wave 1, this variable relies on the sex of the anchor and his/her current partner (using

¹⁶If the anchor stated to be married to his/her current partner, the variables *sd11* (wave 1) or *hpm* (from wave 2 onwards) were used. Starting with wave 10, the variable *sdp30* (relationship status of partner) was taken into account as well.

¹⁷If *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*).

sex_gen and *psex_gen*). If the sex is the same, then *homosex* has the value “1”, otherwise “0”. If there is no current partner, “stated sexual orientation” (“Assuming you would be looking for a partner, would you be you looking for a male or for a female partner”)(*sex2*) is used. In the following waves the value of this variable is defined by the sex of the current partner, resp. the sex of eventual between wave partners (*ehc1pxg*, *x*=1,2,3,4,5). If there is neither a between-wave nor a current partner, the value of *homosex* is carried forward from the last wave. In wave 11, *homosex* relies on information about current partners, previous sexual partners (variable *sex20*) and current sexual orientation (variable *sex21*) for respondents of the refreshment sample.

homosex gets the code “-7 Incomplete data” if an anchor didn’t want to answer *sex2*, if he/she said that he/she is looking for both sexes, or if information from *sex_gen* is missing.

homosex_new

However, there seems to be a problem with the validity of “stated sexual orientation” (*sex2*). About two thirds of those anchors who answer in wave 1 that they would look for a same sex partner start only heterosexual relationships in later waves. Thus, these anchors are probably misclassified as homosexual in wave 1 (and so on, until they begin their first heterosexual relationship). This group comprises almost half of the homosexuals of wave 1, meaning that almost half of the homosexuals in wave 1 are probably misclassified.

Thus, we do not recommend using the *homosex* variable. Instead, beginning with Release 7.0 we provide an alternative definition of homosexuality (*homosex_new*) that is based on sexual orientation in reported partnerships. This variable has the value “-1” if there is no current partner or between wave partner, “0” if the partner is of opposite sex, “1” for gay and “2” for lesbian relationships. In wave 1 also the sex of all previous (before the start of pairfam) partners is checked. In wave 11, the sex of partners mentioned in the retrospective history is taken into account (variable *rtr1pxg*). Value “-7 ” denotes that information from variable *sex_gen* is missing.

The syntax used to compute these variables is contained in the Stata do-files *homosex*.do*

Anchor and/or partner infertile - infertile

The variable *infertile* contains the information that either the anchor and/or the partner is 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 has 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 13 and if the information from previous waves indicated that the anchor or the couple unit was seen to be infertile (only for same couples as in the previous wave respectively), then the code for infertility status was given.

In three cases, the code “-7 Incomplete data” was assigned. First, it was assigned if the question on fertility was not posed (birth cohorts 1991-1993 until wave 6, and 2001-2003) and simultaneously no information on a pregnancy was available. Second, the code “-7” was given if information regarding fertility status was insufficient. Third, value “-7” was given if there was conflicting information about a person, for example if the anchor indicated being fertile but also indicated use of sterilization for contraception. The syntax used to compute the variable is contained in the wave-specific Stata do-file *infertile.do*. While in wave 1, 3, 5, 7, 9, 11 and 13 the anchor as well as the partner were asked about their fertility status or if they use sterilization for contraception, in wave 2, 4, 6, 8, 10, 12 and 14 only

the anchor gave information about his and the partner's fertility status. Therefore the wave-specific do-files differ respectively.

Anchor and/or partner pregnant - pregnant

The variable *pregnant* indicates whether the anchor or his/her partner is currently pregnant. The variable is an indicator at the individual level for female singles and at the couple level for respondents with a partner. The information was taken from the anchor and the partner interviews (if available). A couple was defined as being pregnant if either the anchor or his/her partner explicitly indicated being pregnant, or if the anchor indicated that his/her partner is pregnant. Single female respondents were defined as pregnant if they reported to be so. Additionally, single female respondents who had never had sex were seen to be not pregnant. If a specific single or couple unit did not have the potential to have a pregnancy status (single men and male-male couples), the value "-3 Does not apply" was given. Where information was not sufficient to decide on the status of pregnancy, the code "-7 Incomplete data" was assigned. The syntax used to compute the variable is contained in the Stata do-file *pregnant.do*.

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

These variables contain the number of children of each type listed in Table 4.8. 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-files *nkids*.do* for the syntax used to compute the variables.

Table 4.8: Number of children - **nkids**

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

Type of children - *k*type*

For each child, we generated a variable *k*type* indicating to which category the child could be assigned. In the event that the type of a child could not be found out the code "-7 Incomplete data" was given. If a child did not exist, the code "-3 Does not apply" was assigned. The different types are shown in Table 4.9. The do-files *ktype*.do* generate the variables for each wave.

Second residence - *res2nd*

Variable *res2nd* indicates whether the respondent reported having a second residence. As of wave 4, the information on the domain of residence was not derived from the EHC (as it was the case in waves 1-3). Rather, respondents were asked directly whether they had a second residence or not. The syntax used to compute the variable is contained in Stata do-files *hhszsize.do*. In wave 14, the second residence was not asked for.

Table 4.9: 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

Household size of main residence - *hhsizemrd*

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

Parents living at main residence - *mmrd*, *fmrdrd*

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

Partner living at main residence - *pmrd*

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

Other household members at main residence - *othmrd*

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

Children living at main residence - *childmrd*

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

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

The variable *hhcomp* combines the information on cohabitation with a partner, with at least one parent, with at least one child, and with at least one other household member into one variable. The 16 values of the variable are derived from a cross-classification of these four facts (see Table 4.10). The syntax used to compute the variable is contained in Stata do-files *hhcomp.do*.

Table 4.10: Household composition (partner, kid(s), parent(s), others) - *hhcomp*

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

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

Variables *enrol* and *penrol* capture information on the anchor's and his or her partner's current educational status in the educational system. Both general schooling and vocational qualifications were considered. Values and labels of the variables are listed below (see Table 4.11).

The computation of the variable *enrol* relies on information about the anchor's current occupation, 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 assigned respondents doing "Retraining / Further education" to category "16 Other education". If there is no information at all, previous wave data are used.

If the partner survey provided relevant information in odd-numbered waves, this information was used to construct the variable *penrol*. In even-numbered waves or in the absence of information from the partner, the substitute information given by the anchor was used.

From wave 2 onwards we have more differentiated answers concerning the type of school the partner of the anchor respondent is enrolled in 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. As of wave 3, data is more detailed in respect of enrollment in vocational training. The category "-3 Does not apply" only applies to anchors without partners. Due to the shortened question program and missing information on the type of school in wave 14, the variables *enrol* and *penrol* could not be generated for this wave. The syntax used to compute the variables *enrol* and *penrol* is contained in the Stata do-file *education*.do*, which is provided for each wave.

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

The variables *school* and *pschool* contain information about the highest school degree achieved by the anchor and his or her partner (see Table 4.12). 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

Table 4.11: 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

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 even-numbered waves only proxy information given by the anchor was available to construct the variable *pschool*. In odd-numbered waves the partner is questioned about his/her highest school degree. This direct partner information has priority over the indirect information the anchor gave for generating the variable *pschool*.

Please note that there probably is a certain portion of (not identifiable) respondents who were enrolled in general schooling at the time of interview and who had completed a degree earlier, but nevertheless were classified as students. This is due to the filtering in wave 1, where these respondents were not further asked about their school attainment, and in subsequent waves, 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 became available only in later waves, as they left 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 whereas the codes “3: Lower GDR, POS 8./9.” and “5: Intermediate GDR, POS 10.” were only used once in wave 1 for the anchor, the partner’s highest school degree contains the mentioned codes in all waves.

Information on highest school degree is also provided for respondents’ biological and adoptive parents (variables *mschool/fschool*) in each wave. If parents’ information is available in multiple waves, the newest information is used. In wave 4, only parents who did not participate in wave 2 or wave 3 or who did not indicate their school degree were asked again. Proxy information on parents’ highest school degree given by the anchor in wave 3 is used to compute the variables *mschool* and *fschool* for respondents without direct information from their parents in waves 2 to 8. The variables were computed using the do-file *education_parents.do*. Note that data from the parent survey is only available for waves 2 to 8 for respondents of the pairfam base sample, DemoDiff sample, and the

Table 4.12: Values and labels of variables *school*, *pschool*, *mschool*, *fschool*

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

step-up sample. From Release 9.0 on, information on highest school degree of respondents' parents was no longer updated. However, in wave 11, new information about parents' education was provided for respondents of the refreshment sample (sample=3).

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

The highest vocational degree completed by the anchor and his or her partner is stored in variables *vocat* and *pvocat*, respectively (see Table 4.13). 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 was updated every year if the anchor and his or her partner had 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 partner's highest vocational degree (*pvocat*) was set to "-3" for anchors who are currently not in a relationship. The variables were generated by running the wave-specific do-file education.do.

Information on highest vocational degree is also provided for respondents' biological and adoptive parents (variables *mvocat/fvocat*). If parents' information is available in multiple waves, the newest information is used. In wave 4, only parents who did not participate in wave 2 or wave 3 or who did not indicate their vocational degree were asked again. Proxy information on parents' highest vocational degree given by the anchor in wave 3 is used to complete the variables *mvocat* and *fvocat* for respondents with no direct information from their parents in wave 2 to 8. The variables were computed using the do-file education_parents.do. Note that data from the parent survey is only available for waves 2 to 8 for respondents of the pairfam base sample, DemoDiff sample, and the step-up sample. From Release 9.0 on, information on highest school degree of respondents' parents was no longer updated. However, in wave 11, new information about parents' education was provided for respondents of the refreshment sample (sample=3).

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

The generated variables *isced*, *pisced*, *miscd*, and *fiscd* carry information on anchor's, partner's, and parents' educational attainment classified according to the International Standard Classification of Education (ISCED-97) (UNESCO 2006). We applied a revised version of the scheme adapted to the German institutional context as proposed by Schneider (2008). While the original scheme was

Table 4.13: Values and labels of variables *vocat*, *pvocat*, *mvocat*, *fvocat*

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

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 in school or vocational training were assigned the value “0”, and the ISCED-categories were applied only to persons who had earned an educational certificate and were not currently enrolled. As suggested by Schneider (2008), we introduced a further category “1b” (value “1”) for persons who did not achieve any formal degree. The variables were computed using the do-files *education*.do* (available for each wave) and *education_parents.do*.

The variables *isced* and *pisced* provided the basis for two additional variables (*isced2*, *pisced2*), where afterwards respondents and partners currently enrolled in the German educational system were assigned. Thus it was assumed that those currently enrolled would complete their education and attain the corresponding degree. This can be especially helpful for waves with a high percentage of currently enrolled respondents. Values and labels of the resulting variables are shown in Table 4.14. Please note that it was not possible to generate the variables *isced2* and *pisced2* for wave 14 due to missing information. The syntax used to compute these variables is contained in the Stata do-files *education*.do* (available for each wave) and *education_parents.do*.

Table 4.14: Values and labels of variables *isced*, *pisced*, *misced*, *fiscd*

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

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

Variables *casmin*, *pcasmin*, *mcasmin*, and *fcasmin* contain respondents' educational attainments according to the classification scheme originally set up by the project on Comparative Analysis of Social Mobility in Industrial Nations (CASMIN) (König et al. 1988). We applied the updated version of the original classification (Brauns and Steinmann 1999). Both school and vocational degrees were

considered, i.e., the variables used to assign values of the CASMIN categories were the generated variables *school* and *vocat* (*pschool* and *pvocat* for the partner, *mschool/fschool* and *mvocat/fvocat* for the parents). The complete classification schema is shown in Table 4.15. The syntax used to compute these variables is contained in the Stata do-files *education*.do* (one for each wave) and *education_parents.do*.

Table 4.15: Values and labels of variables *casmin*, *pcasmin*, *mcasmin*, *fcasmin*

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

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

Variables *yeduc*, *pyeduc*, *myeduc*, and *fyeduc* are linear measures of the anchor's, his or her partner's, and parents' educational attainment. Values assigned were derived from the typical duration of educational careers, i.e., values correspond to 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),
- 10 years for other school degrees without further specification.

To years of schooling, we added for vocational qualifications:

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

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

Please note that from wave 9 on, information on educational attainment of respondents’ parents was no longer updated for respondents of the pairfam base sample, DemoDiff sample, and the step-up sample. However, in wave 11, new information about parents’ education was provided for respondents of the refreshment sample (*sample=3*).

KldB classification of occupation - *kldb2010*

Variable *kldb2010* contains information according to the classification of occupations proposed by the German Statistical Office (Klassifikation der Berufe, KldB). The KldB scheme was designed to fit the German occupational system better than the international ISCO scale (see below). The classification is based on variable *job1* in wave 1 and *job20o* in subsequent waves. The original variables carried open answers about the anchor’s current occupation (The original answers have been made anonymous as described in section 4.2). In all waves, the variable *kldb2010* is based on the new 5-digit classification scheme KldB 2010. Responses that do not fit into the classification scheme are assigned negative two-digit values. In the anchor data sets of waves 1 to 3, the variable *kldb1992* based on the former 4-digit classification is additionally provided.

Variable *flag_isco08_kldb2010*¹⁸ indicates that a respondent’s answer was ambiguous and could not be recoded to exactly one KldB (or ISCO) code, but could equally well have been assigned two or more codes. The variable distinguishes between two coding problems. Value “1” indicates that the occupation requiring the lowest level of qualification has been assigned. Value “2” flags respondents where the most frequent of all possible occupations has been assigned. Recoding was done by *Kantar Public*. Variable *flag_isco08_kldb2010* is based on the new classification schema KldB2010 (ISCO-08) (see Hartmann et al. (2012) for more details). The variable *flag_isco08_kldb1992* based on the former classification is provided in wave 1 (see Hartmann et al. (2010), Hartmann et al. (2011a), and Hartmann et al. (2011b) for more details).

ISCO classification of occupation - *isco08*

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

EGP class scheme - *egp*

The variable *egp* is only provided for waves 1 to 3. It 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 scheme 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 the variable *egp*, please take into account that we used proxy information on supervisory functions derived from the occupational status (variable *job2*). At this point, the procedure relies on

¹⁸This variable is a renamed copy of variable *beruprob*, which was originally provided by *Kantar Public*.

quite strong assumptions.¹⁹ The Stata code for computing the variable *egp* is available in the do-file *egp.do*.

ISEI occupational prestige score - *isei*

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

SIOPS occupational prestige score - *siops*

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

MPS occupational prestige score - *mps*

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

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

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

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

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

Calculation of the equivalence scale weight requires information on the age of all household members. The variables *npu14mr* and *npo14mr* indicate the number of persons besides the anchor aged under and over 14 years in the anchor's main residence (the anchor respondent is assigned a value of 1 in the equivalence scale weight and thus is not counted). For waves 1 and 2, the variables were created based on the age variables of the generated household data set and, starting with wave 3, based on the household matrix and the age variables in the anchor data sets. Some assumptions have been

¹⁹ 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.

made in order to minimize the number of missing values of the variables *npu14mr* and *npo14mr*, e.g. if the age of the parents, grandparents or parents-in-law is missing in the original data set, it is reasonably assumed that these persons are older than 14. In wave 1, age was not recorded for each household member, and has thus been imputed using the generated household data from wave 2. We assigned a value of 0.5 for household members of anchor persons who did not participate in wave 2, as well as for household members who had moved house since wave 1. Respondents sharing a residence with roommates or housemates, living in dormitories, boarding schools, or similar residences for which the assumption of shared resources within the household is problematic, have been assigned an equivalence scale weight of 1 and are thereby considered single-person households. Due to missing information about the age of all household members, the variables *hhincoecd*, *npu14mr* and *npo14mr* are not generated for wave 14.

If the household income was missing, the net equivalence income according to the modified OECD equivalence scale was coded to “-7 Incomplete data”. The syntax used to compute the variable *hhincoecd* and the variables *npu14mr* and *npo14mr* is contained in the Stata do-file *incnet.do*.

Personal and household net income - *incnet*, *hhincnet*

The variables *incnet* and *hhincnet* contain the information of both open-ended and categorized answers to questions on personal and household net income. The categorized answers were coded as midpoints of the categories. At the higher end, personal net income was coded 2,500 Euro for wave 1, and 5,000 Euro as of wave 2. For household net income, 5,000 Euro was assigned as a maximum. The codings are shown in Table 4.16 and Table 4.17. If an individual was not gainfully employed (according to answers on employment based on the EHC) the value “-3 Does not apply” was assigned to the variable *incnet*.²⁰ If information was not available to derive the income, the code “-7 Incomplete data” was assigned. The syntax used to compute the variables is contained in the Stata do-file *incnet.do*.

Note that in wave 1 information on household net income was only collected for respondents cohabiting with their partner, children, or additional persons in their main household. For respondents living alone (*hhincnet* coded “-3 Does not apply”), household income can only be approximated based on personal net income information and, if any, social benefits received (variables *inc11* and *inc12ix*).

Further note that beginning of wave 2, respondents were only asked if they received any social benefits (*inc10x*) but were not asked about the specific amount of those social benefits anymore (in wave 1: *inc11* and *inc12x*). The household net income indirectly contains these social benefits, but a detailed decomposition of the amounts of several social benefits is not possible from wave 2 onwards. Also starting of wave 2, alimony payments are part of social benefits (variables *inc10i13*, *inc10i14* and *inc10i15*) whereas in wave 1 these information are missing.

Table 4.16: 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)	625	(1150, 1450)	1300
[750, 900]	875	[1450, 1700]	1575
[900, 1100]	1000	[1700, 2000]	1850
(1100, 1350]	1225	(2000, 2300]	2150
(1350, 1600)	1475	(2300, 2800)	2550
[1600, 2000]	1800	[2800, 3500]	3150
(2000, ∞)	2500	(3500, ∞)	5000

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

personal net income		household net income	
Categories	Codes	Categories	Codes
[0, 250)	125	[0, 250)	125
[250, 500)	375	[250, 500)	375
[500, 750)	625	[500, 750)	625
[750, 1000)	875	[750, 1000)	875
[1000, 1250)	1125	[1000, 1250)	1125
[1250, 1500)	1375	[1250, 1500)	1375
[1500, 1750)	1625	[1500, 1750)	1625
[1750, 2000)	1875	[1750, 2000)	1875
[2000, 2500)	2250	[2000, 2500)	2250
[2500, 3000)	2750	[2500, 3000)	2750
[3000, 3500)	3250	[3000, 3500)	3250
[3500, 4000)	3750	[3500, 4000)	3750
[4000, 4500)	4250	[4000, 4500)	4250
[4500, ∞)	5000	[4500, ∞)	5000

Consumer price index - *cpi*

The variable *cpi* contains the consumer price index on a month-to-month basis according to the German National Statistical Office. The information was taken from a query of www.destatis.de. Note that this variable was constant for all observations. The syntax used to compute the variable is contained in the Stata do-file *cpi1-14.do*. Since Release 11.0, the coding of the variable *cpi* in all waves was based on the baseline year 2015 (previous releases: 2010).

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

The variables *casprim*, *cassec*, *pcasprim*, and *pcassec* contain the anchor's and his or her partner's current primary and secondary activity status. For the anchor, the information was taken from the anchor interview. Information on the partner was taken from the partner questionnaire in odd-numbered waves. In even-numbered waves we only had proxy information given by the anchor. Concerning educational attainment, this information is less rich than for anchors.²¹ The current primary and secondary activity status variables were not generated for wave 14, due to a less extensive differentiation between activities in wave 14.

Multiple possible activities were reduced to the two most important ones. For both anchors and partners this reduction was accomplished by applying the following dominance rules:

- 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-empl. > part-time empl. > marginal empl. > internship > occasionally empl. > other type of job
 - only educational activities were stated:²²
primary activity is defined according to the dominance rule:
general secondary school > UCE²³ > college/university > evening school > technical/professional school > pre-vocat. training > vocat. retraining > other education

²¹For partners, we only know if they are involved in general schooling, vocational training, or university.

²²Due to filtering and question wording partners cannot have multiple education activities.

²³University of Cooperative Education ("Berufsakademie")

- only non-employment 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 non-employment activities were stated:
primary activity was defined according to the dominance rule:
maternity/paternity leave > military service > unemployed > retired > vocat. training > full-time empl. > self-empl. > part-time empl. > marginal empl. > internship > housewife > occasionally empl.
- both non-employment 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, non-employment, 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 were made under consideration of occupation, occupational status, work hours, educational attainment, personal income, and age.²⁴

After this reduction process, the defined activity statuses were checked for inconsistencies. In Table 4.18, inconsistent combinations of the primary and secondary activity status are marked with an X. These cases are marked in the data with the flag variables *flag_cas* and *pflag_cas*.

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

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

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

Employment activities are defined as one of the following activities:

- vocational training,
- full-time employment,

²⁴For partners only occupational status, educational attainment, work hours, and age are considered.

Table 4.18: 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

- 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 assumed to be the labor force status. If both activity statuses were employment activities, the primary activity status was used. If both activity statuses were non-employment activities, the primary activity status was used. This derived single activity status was recoded to the labor status. The coding scheme is shown in Table 4.19.

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

If there was no partner, the partner’s labor force status was coded as “-3”. If no information was available to code a status, it was coded as “-7”. The syntax used to compute the variables is contained in the Stata do-file lfs.do (available for wave 1-13, not generated for wave 14). Furthermore,

Table 4.19: 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

researchers also need the Stata do-file *cas.do*, as the labor force status is based on the primary and secondary activity statuses.

Currently living in Eastern Germany - east

To differentiate whether the anchor was currently living in Eastern or Western Germany, we generated the dummy variable *east*. Value “1” indicates that the anchor was living in Eastern Germany, while “0” means that the anchor resided in Western Germany. Note that the computation of this variable is based on the process-generated variable *bula* (see section 4.4). The relevant do-file *east1-14.do* is applicable to all waves.

Indices of physical and mental health - *pcs*, *mcs*

From waves 7 to 13, pairfam applied more detailed health measures, namely the SF12. This is a shortened version of the previously used SF36 index (Ware et al. 2001, 2002) which evaluates anchor health status using 12 instead of 36 items, but nevertheless encompasses all eight health subscales: physical functioning (2 items), role physical (2), bodily pain (1), general health (1), vitality (1), social functioning (1), role emotional (2), and mental health (2). Four of the eight subscales consist of one item and four were built from two items. The SF12 was included in every second wave (waves 7, 9, 11 and 13), whereas the other waves include the standard pairfam health items. Following the procedures used by the German SOEP (Nübling et al. 2006) exactly, we generated from these 12 items two indices in wave 7: one for physical health (*pcs*) and one for mental health (*mcs*). The indices are only computed for those anchors who provided valid values for all 12 items. Each item was recoded to a scale ranging from 0 to 100. For the two-item subscales, the mean value was used. Afterwards, all eight health subscales were z-transformed with a mean of 0 and standard deviation of 1. Factor analysis (PCA, varimax rotation) confirmed that the eight subscales load on two factors: the four physical subscales load on the physical factor and the four mental subscales load on the mental factor.

The results of the factor analysis are the basis for the weighting of the eight subscales. Using the factor loadings, we calculated the factor scores from the eight subscales and generated the indices *pcs* and *mcs*. Analogous to the German SOEP (Nübling et al. 2006), the loadings of the subscales on the non-primary factor were used for the generation as well. This means, that the four subscales for mental health were also used for the generation of the physical health indices *pcs* (however, with a lesser weight) and vice versa. In a final step, *pcs* and *mcs* were both transformed to norm-based scores (mean = 50, SD = 10). Thus, a value of 50 indicates average health. In waves 9, 11, and 13, means and standard deviations of wave 7 were used as baseline for the z-transformation and factor score coefficients of wave 7 were used for the calculation of the aggregate scores physical health (*pcs*)

Table 4.20: Rotated factor loadings and factor score coefficients of 8 subscales on two main factors

	Factor loadings		Factor score coefficients	
	Factor 1	Factor 2	Factor 1	Factor 2
Physical fitness (factor, 2 vars)	0.7597	0.0859	0.39401	-0.16617
Role physical (factor, 2 vars)	0.7856	0.2883	0.35067	-0.05573
Bodily pain (1 var)	0.8059	0.1302	0.40680	-0.15349
General health (inverted, 1 var)	0.6416	0.3646	0.24961	0.02972
Vitality (inverted, 1 var)	0.0659	0.6068	-0.13632	0.33445
Social functioning (1 var)	0.3063	0.7125	-0.03404	0.32764
Role emotional (factor, 2 vars)	0.2028	0.7820	-0.11083	0.39756
Mental health (factor, 2 vars)	0.1851	0.7551	-0.11292	0.38694

and mental health (*mcs*). Missing values were recoded as "-77 Incomplete data". The syntax used to compute these variables is contained in the Stata do-files *health*.do*.

Scales

All of the previously described generated variables are included in the anchor data sets. In contrast, scale variables (see Table 4.21) are not part of the delivered data set. They can be generated by running the syntax files *scales_anchor\$* (available in Stata and SPSS format), which are available for each wave as part of the Scientific Use File.

In wave 11, a refreshment sample of new anchor respondents was drawn. To provide the same scales for the base and the refreshment sample, some scales already available for the base sample had to be created for the refreshment sample in wave 11 (see the following table). Additionally, some scales were generated exclusively for the refreshment sample (see Table 11.1). All scales for both samples are part of the syntax files *scales_anchor\$*. For further information, please refer to the Scales Manual which contains a detailed description of all scales.

Table 4.21: List of scales included in syntax file *scales_anchor\$*

Construct	Variable name	Wave
Traditional concept of marriage	tradmarr	1,3,5,7,9,11,13
Value of Partnership - Negative expectations	vopneg/vopneg2	1/3,5,7,9,11
Personality of ideal partner - Neuroticism	neurot_p	13
Personality of ideal partner - Extraversion	extrav_p	13
Personality of ideal partner - Agreeableness	agreeable_p	13
Personality of ideal partner - Conscientiousness	conscient_p	13
Personality of ideal partner - Openness	openness_p	13
Value of Children: Benefit of stimulation	vocbstim	1,2,4,6,8,10,12
Value of Children: Costs of comfort	vocccomf	1,2,4,6,8,10,12
Single: Interest in partnership	partint	1 - 3,7 - 13
Single: Desire for partnership	partdes	all
Single: Broad exploration	siexplbr/siexplbr2	1,2/3 - 14

continued on next page

continued from previous page

Construct	Variable name	Wave
Single: In-depth exploration	siexplde	1 - 13
Single: Mating confidence	chanpm	1 - 13
Getting to know each other: In-depth exploration	npexplde	1 - 13
Ambivalence: moving in together	ambcoh	1,3,5,7,9,11,13
Ambivalence: marriage	ambmarr	1,3,5,7,9,11,13
Partnership: Conflict	confl_apd	all
Partnership: Intimacy	intim_aps	all
Partnership: Admiration	admir_apo	all
Partnership: Dominance	domin_apo	all
Partnership: Ambivalence	ambiva_apd	7 - 13
Partnership: Emotional ambivalence	ambiv_apd	1 - 3,5,7,9,11,13
Partnership: Anxiety about loss of love	lovewitanx_apd	1 - 3,5,7,9,11,13
Partnership: Anxiety about being absorbed	enganx_apd	1 - 3,5,7,9,11,13
Partnership: Autonomy	indep_apd	1 - 3,5,7,9,11,13
Partnership: Feelings of competence in partnership	comppart/comppart2	1/2,3,5,7,9,11,13
Partnership: Own partnership satisfaction	satpart	1,3,5,7,9,11,13
Partnership: Instability of partnership	instab_apd	1 - 13
Partnership: Readiness to sacrifice Self	sacrif_aps	1,3,5,7,9,11,13
Partnership: Hostile attribution Self	hostattr_aps	1,3,5,7,9,11,13
Partnership: Future orientation	comfut_apd	1 - 3,5,7,9,11,13
Partnership: Tolerance of conflicts	comctol_apd	1 - 3,5,7,9,11,13
Partnership: Orientation of reciprocity Self	reciproc_aps	1,3,5,7,9,11,13
Partnership: Frequency of manifest conflicts	confl_om_apd	1,5,7,9,11,13
	confl_om2_apd	3
Partnership: Verbal aggression Partner	verbaggr_apo	1 - 13
Partnership: Verbal aggression Self	verbaggr_aps	all
Partnership: Constructive behavior Partner	constrbh_apo	1 - 13
Partnership: Constructive behavior Self	constrbh_aps	all
Partnership: Withdrawal Partner	withdraw_apo	1 - 13
Partnership: Withdrawal Self	withdraw_aps	all
Partnership: Manipulation Partner	manipul_apo	1,3,5,7,9,11,13
Partnership: Manipulation Self	manipul_aps	1,3,5,7,9,11,13
Partnership: Dyadic coping Partner	dycop_apo	1,3,5,7,9,11,13
Partnership: Dyadic coping Self	dycop_aps	1,3,5,7,9,11,13
Interagen. Relationships: Conflict with mother	confl_amd	1 - 13
Interagen. Relationships: Conflict with partner of mother	confl_asfd	2 - 13
Interagen. Relationships: Conflict with partner's mother	confl_pamd	8
Interagen. Relationships: Conflict with father	confl_afd	1 - 13
Interagen. Relationships: Conflict with partner of father	confl_asmd	2 - 13
Interagen. Relationships: Conflict with partner's father	confl_pafd	8
Interagen. Relationships: Intimacy mother	intim_ams	1 - 13
Interagen. Relationships: Intimacy partner of mother	intim_asfs	2 - 13
Interagen. Relationships: Intimacy partner's mother	intim_pams	8
Interagen. Relationships: Intimacy father	intim_afs	1 - 13
Interagen. Relationships: Intimacy partner of father	intim_asms	2 - 13
Interagen. Relationships: Intimacy partner's father	intim_pafs	8
Interagen. Relationships: Admiration mother	admir_amo	2,4,6,8,10,12
		11 (refreshment)
Interagen. Relationships: Admiration partner of mother	admir_asfo	2,4,6,8,10,12
		11 (refreshment)
Interagen. Relationships: Admiration partner's mother	admir_pamo	8

continued on next page

continued from previous page

Construct	Variable name	Wave
Intergen. Relationships: Admiration father	admir_afo	2,4,6,8,10,12 11 (refreshment)
Intergen. Relationships: Admiration partner of father	admir_asmo	2,4,6,8,10,12 11 (refreshment)
Intergen. Relationships: Admiration partner's father	admir_pafo	8
Intergen. Relationships: Dominance mother	domin_amo	2,4,6,8,10,12 11 (refreshment)
Intergen. Relationships: Dominance partner of mother	domin_asfo	2,4,6,8,10,12 11 (refreshment)
Intergen. Relationships: Dominance partner's mother	domin_pamo	8
Intergen. Relationships: Dominance father	domin_afo	2,4,6,8,10,12 11 (refreshment)
Intergen. Relationships: Dominance partner of father	domin_asmo	2,4,6,8,10,12 11 (refreshment)
Intergen. Relationships: Dominance partner's father	domin_pafo	8
Intergen. Relationships cohort1: Negative Communication Mother/Partner of father	negcomm_amo	2 11 (refreshment)
Intergen. Relationships cohort1: Negative Communication Father/Partner of mother	negcomm_afo	2 11 (refreshment)
Intergen. Relationships cohort1: Successful individuation Mother/Partner of father	sucindiv_amd	2 11 (refreshment)
Intergen. Relationships cohort1: Successful individuation Father/Partner of mother	sucindiv_afd	2 11 (refreshment)
Intergen. Relationships cohort1: Fear of love withdrawal Mother/Partner of father	lovewitanx_amd	2 11 (refreshment)
Intergen. Relationships cohort1: Fear of love withdrawal Father/Partner of mother	lovewitanx_afd	2 11 (refreshment)
Intergen. Relationships cohort1: Ambivalence Mother/Partner of father	ambiv_amd	2 11 (refreshment)
Intergen. Relationships: Ambivalence Mother	ambiv2_amd	7 - 11
Intergen. Relationships: Ambivalence Partner of mother	ambiv2_asfd	7 - 11
Intergen. Relationships: Ambivalence Partner's mother	ambiv2_pamd	8
Intergen. Relationships cohort1: Ambivalence Father/Partner of mother	ambiv_afd	2 11 (refreshment)
Intergen. Relationships: Ambivalence Father	ambiv2_afd	7 - 11
Intergen. Relationships: Ambivalence Partner of father	ambiv2_asmd	7 - 11
Intergen. Relationships: Ambivalence Partner's father	ambiv2_pafd	8
Sexual competence	compsex	2,3,5,7,9,11,13
Sexual communication	sexcom	2,3,5,7,9,11,13
Explosiveness and tendency to anger	explosive	1,5,9,13
Shyness	shyness	1,5,9,13
Emotional autonomy	emotautn	1,5,9,13
Self-esteem	selfesteem	all
Depressiveness	depressive/depressive2	2 - 14
Anger	anger	3 - 5, 13
Activity	activ/activ2	3/4 - 13
Stress	stress	4 - 13
Anxiety	anxiety	4,5,13
Prevention	prevent	4
Promotion	promot	4
Goal Engagement - primary selective control	goaleng_psc	8,10,12

continued on next page

continued from previous page

Construct	Variable name	Wave
Goal Engagement - secondary selective control	goaleng_ssc	8,10
Goal Disengagement	goaldiseng	8,10,12
Goal Reengagement	goalreeng	8,10,12
Self-Protection	selfprotect	8,10,12
BIG 5: Neuroticism	neurot	2,6,10 11 (refreshment)
BIG 5: Extraversion	extrav	2,6,10 11 (refreshment)
BIG 5: Agreeableness	agreeable	2,6,10 11 (refreshment)
BIG 5: Conscientiousness	conscient	2,6,10 11 (refreshment)
BIG 5: Openness	openness	2,6,10 11 (refreshment)
Dark Triad: Machiavellianism	machiav	9,11,13
Dark Triad: Psychopathy	psychop	9,11,13
Dark Triad: Narcissism/Rivalry	riv	9,11,13
Dark Triad: Narcissism/Admiration	adm	9,11,13
Economic deprivation parents	ecodep_par	2
	ecodep2_par	3,4
Economic deprivation anchor	ecodep_a	2
	ecodep2_a	3,4
Economic deprivation household	ecodep_hh	2
	ecodep2_hh	3 - 14
Newborn temperament child1	temperc1	2 - 14
Newborn temperament child2	temperc2	2 - 14
Newborn temperament child3	temperc3	2 - 14
Newborn temperament child4	temperc4	2 - 14
Newborn temperament child5	temperc5	2 - 14
Newborn temperament child6	temperc6	2 - 14
Newborn temperament child7	temperc7	2,4 - 14
Unspecific strain	unspstrain	3,7 - 14
Autonomy in the parenting role (3 Item Scale)	autonoms	2
Autonomy in the parenting role (4-Item Scale)	autonom	4,6,8 - 14
Autonomy in the parenting role (2-Item Scale)	autonom2	4,6,8 - 14
Pleasure in the parenting role (newborn module)	pleasure_newb	2
Pleasure in the parenting role (children till 15 years)	pleasure	4,6,8 - 14
Parental Self Efficacy/Competence	comperz	2,4,6,8 - 14
Co-parenting	coparent	2,4,6,8,10,12,14
Co-parenting with ex-partner	coparent_ex	3
Co-parenting with ex-partner1	coparent_exp1	5,7
Co-parenting with ex-partner2	coparent_exp2	5,7
Co-parenting with ex-partner3	coparent_exp3	5,7
Co-parenting with ex-partner - (capi)kid1	coparent_opk1 ^a	3,5,7 - 9,11,13
Co-parenting with ex-partner - (capi)kid2	coparent_opk2	3,5,7 - 9,11,13
Co-parenting with ex-partner - (capi)kid3	coparent_opk3	5,7 - 9,11,13
Co-parenting with ex-partner - (capi)kid4	coparent_opk4	5,7 - 9,11,13
Co-parenting with ex-partner - (capi)kid5	coparent_opk5	7 - 9,11,13
Co-parenting with ex-partner - (capi)kid6	coparent_opk6	5,7 - 9,11,13
Parenting goals: Status	pgoalstatus	2 - 13

continued on next page

^aUp to wave 7 only for capikids and from wave 8 onwards child specific.

continued from previous page

Construct	Variable name	Wave
Parenting goals: Autonomy	pgoalautn	2 - 13
Parenting goals: Competence	pgoalcomp	2 - 13
Parenting: Partner support	partnersup	3,5,7,9,11,13
Parenting: Overprotection	overprotect	3,5,7,9,11,13
Parenting: Social support	socialsup	3,5,7,9,11,13
Parenting: Readiness to make sacrifices	sacrif_pacs	3,5,7,9,11,13
Demands on parenting: Perfectionism	pstandperfect	5,7,9,11,13
Demands on parenting: Pragmatism	pstandpragmat	5,7,9,11,13
SDQ Behavior: Conduct problems childx	conduct_aco1...	5 - 13
	conduct_aco7	
SDQ Behavior: Hyperactivity childx	hyper_aco1,...,	5 - 13
	hyper_aco7	
SDQ Behavior: Emotional symptoms childx	emotion_aco1,...,	5 - 13
	emotion_aco7	
Emotional warmth childx	warmth_acs1,...,	5 - 14
	warmth_acs7	
Negative communication childx	negcomm_acs1,...,	5 - 14
	negcomm_acs7	
Inconsistent parenting childx	inconsist_acs1,...,	5 - 14
	inconsist_acs7	
Sibling Relationship: Intimacy to siblingx	intim_assp1,...,	5,7,9,11,13
	intim_assp13	
Sibling Relationship: Conflict with siblingx	confl_asdp1,...,	5,7,9,11,13
	confl_asdp13	
Work-Family-Conflict: Work impacts family	wif_conflict	6,8,10,12
Work-Family-Conflict: Work impacts family - time	wif_time	6,8,10,12
Work-Family-Conflict: Work impacts family - stress/strain	wif_strain	6,8,10,12
Work-Family-Conflict: Family impacts work	fiw_conflict	6,8,10,12
Work-Family-Conflict: Family impacts work - time	fiw_time	6,8,10,12
Work-Family-Conflict: Family impacts work - stress/strain	fiw_strain	6,8,10,12
Dissolution of boundaries in labour	wf_boundary	6,8,10,12,14
Parental Relationship: Conflict	confl_mfd	6,8
Family Values: Trust	trust	10 - 13
Hostile attributions towards the child	hostattr_acd	13
Problematic use of cellphones	probcell	13

4.4 Process-generated variables

This chapter describes the process-generated variables listed in Table 4.22. Variables *bula*, *gkpol* and *bik* contain information about the respondent's place of residence, i.e. the address where *Kantar Public* contacted participants (*Meldeadresse*).

Table 4.22: List of process-generated variables

Construct	Variable name
State / Bundesland	bula
Size of community in 7 categories	gkpol
Settlement structure	bik

These variables are included in the data set *anchor\$*. Variable *bula* contains information on the state (*Bundesland*). Variable *gkpol* carries information on the size (population) of the community, divided into 7 categories (see Table 4.23). Variable *bik* contains information on the settlement structure (*Siedlungsstruktur*, BIK-Typ), divided into 10 categories (see Table 4.24) and is based on the classification of BIK Aschpurwis + Behrens GmbH. This information was updated annually based on the current population statistics provided by the State Statistical Offices. BIK regions thus correspond to the current territory information.

Table 4.23: Values and labels of variable *gkpol*

1	< 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

Table 4.24: 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

4.5 Paradata

We added information on the survey process to the anchor data sets for all valid and completed interviews. The available variables are listed in Table 4.25. For all addresses contacted, the total number of contacts was reported by the interviewers (variable *intcont*). Gender and age of interviewers are on file (variables *intsex* and *intage*). Interviewers' school degree is available from wave 5 onwards (variable *intedu*). Please note that response categories for this variable changed between waves 5 and 6. Only wave 5 contains the category "1 left school w/o degree" whereas waves 6 to 13 contain the categories "4 left university w/o degree", "5 University degree" and "6 other school degree". 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*). Due to the self-administered survey mode, no interviewer-related variables are contained in the data set *anchor14_cawi* and *anchor14_papi*.

We also included the length of the interview in minutes (variable *intdur*). In wave 1 the relevant variables for computing the total length were provided by the survey institute *Kantar Public*. 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.²⁵ In this case, we assigned

²⁵Some digits were missing, and it was not possible to recover the information whether it was the hours or the seconds missing.

the value “-7 Incomplete data”. From wave 2 onwards 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.²⁶ Therefore, we could not differentiate between inconsistent values and incomplete data and assigned the value “-7 Incomplete data” in both cases. Please note that no information on *intdur* is available for anchors who used the PAPI version in wave 14.

For waves 12 and 13, we added a new variable on the interview date of a specific group of respondents: Due to the COVID-19 pandemic, we had to change survey modes from CAPI/CASI to CATI/PAPI for some anchors (for more information, see Chapter 12). The variable *intdat_papi* specifies the calendar week in which Kantar received the paper questionnaire from CATI/PAPI respondents.

Table 4.25: 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's school degree (since wave 5)	intedu
Interviewer ID	intid
Duration of CAPI/CATI interview in minutes	intdur
Date of interview (month, day, year)	intm, intd, inty
Calendar week of PAPI questionnaire reception (wave 12 & 13)	intdat_papi
Interview mode (waves 12-14)	intmode

For waves 12 to 14 the variable *intmode* has been integrated into the anchor data sets as well. It provides information on the respective survey mode (see Table 4.26 for details). Please note that values 1 and 2 are only valid for waves 12 and 13 while values 3-6 only exist in wave 14.

Table 4.26: Values and labels of variable *intmode*

1	CAPI & CASI (as in waves 1-11)
2	CATI & PAPI
3	CATI only
4	CAPI only
5	PAPI only
6	CAWI only

Moreover, the *Paradata* directory in the Scientific Use File contains three additional types of data sets: the cumulative gross anchor data set, wave-specific anchor timestamps and duration, as well as metadata on the web-based survey of wave 14.

First, the anchor gross data set *Paradata_Fieldwork_W1-W14* combines all waves and includes all cases contacted for an additional interview in wave *t* (i.e., anchor respondents who participated in wave *t-1* as well as non-contacts and soft refusals from wave *t-1*). Respondents who have withdrawn permission to be re-contacted are not included. The following fieldwork information is provided: number of contact attempts (personal, by phone, and by email), interviewer identification numbers, information on the place of residence, final processing status (including reasons for non-participation), and some fieldwork information specific for wave 14.

Second, the data sets *Timestamps+Duration_WX* contain wave-specific information on the interview day, timestamps, and duration of individual modules for waves 1-14. Please note that these are raw

²⁶Duration 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.).

data sets that have not been reviewed or processed. Variables are labeled in German only. Variable names and labels can differ between waves.

Third, the data set *Meta_data_W14* includes technical information on the web-based participation process in wave 14. The following variables are included for respondents who took part via CAWI or CAPI mode: browser ID, browser name and version, operating system, device, brand, model, and output mode.

4.6 Weights

In preparation for Release 12.0, the pairfam team made substantial modifications to the weighting procedure used for all waves of the pairfam data in collaboration with the GESIS Team on Survey Statistics²⁷. Users interested in the concept of the previous weighting approach and its respective weights should refer to data and documentation for releases up to Release 11.0. These are available via the GESIS data archive.

Changes made affect both the design weight ($dweight_w$), which corrects disproportionate sampling across cohorts and the combination of multiple selection “frames” including DemoDiff and the wave 11 refreshment sample (for more details, see Chapter 9 and 11), as well as the calibrated design weight ($cdweight_w$), which calibrates the design weights to reference characteristics, thereby correcting both baseline and longitudinal survey non-response. As using the respective design weights is a precondition of applying calibration weights, pairfam provides ready-to-use combinations of calibration and design weights: calibrated design weights. All weights are centered for each wave to a mean of 1. Step-up respondents are not part of the random sample selection process; therefore, their data is not weighted (see Chapter 10).

In addition to the design weights of the pairfam base sample ($dweight_w$) and the design weights of both the pairfam base sample and the DemoDiff sample ($d1weight_w$), two design weights address the inclusion of the refreshment sample from wave 11: one allows for analyses of the full data set ($d2weight_w$; a combination of the pairfam base sample, DemoDiff, and the wave 11 refreshment sample), while the other is a design weight for the refreshment sample only ($d3weight_w$).

A calibrated design weight is provided for each of the above-mentioned study populations. These weights provide factors to adjust the observed data to characteristics of the general population. This weighting step has three advantages: First, it ensures that the weighted data more closely represent the population of interest in central characteristics and size. Second, selective non-response can be managed by assigning observations with characteristics of higher selectivity a higher analysis weight, tackling both cross-sectional survey participation bias and longitudinal panel attrition bias for the following waves. Third, a correction of cohort-specific non-response aiming to represent actual cohort sizes in the population of interest can be integrated.

This section is an overview of all important information regarding weights in the pairfam data. The Technical Paper No. 17 (Wetzel et al. (2021)) includes further information, in particular concerning key weight characteristics, the changing influence of the calibration variables over the panel, and insights into the quality of the weights. For practical guidance in applying weights in Stata, please refer to the Quick Start file “Weighting”.

Please note that as of Release 12.0, the variables *panswer*, *ppanel*, and *pcontact* are no longer included in the data.

²⁷In particular, we would like to thank Dr. Bruch, Dr. Felderer, and Dr. Sand of GESIS for their consulting and concrete support in implementing the new weighting strategy. All remaining issues with the weighting are solely attributable to the pairfam team.

4.6.1 Design weights

Design weights have been defined to achieve the following:

1. Correct for different cohort inclusion probabilities, i.e. under- or over-representation in the gross base sample as compared to the population of interest
2. Integrate the DemoDiff sample into the pairfam sample by addressing the shares of respondents living in Eastern and Western Germany
3. Address the issues arising from drawing a refreshment sample with a different sample population.

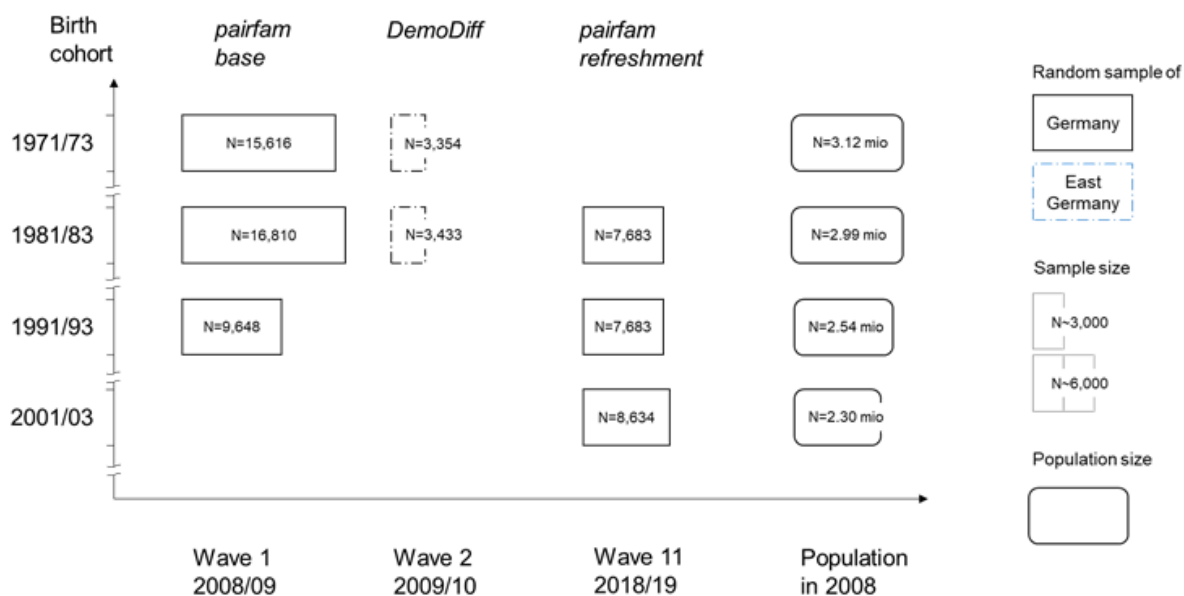


Figure 4.1: Gross pairfam samples relative to population size

Figure 4.1 illustrates these challenges by presenting gross sample and population sizes. Although population sizes for each cohort vary only slightly, at wave 1 the youngest cohort was targeted with a considerably smaller gross sample due to the fact that response rates were expected to be higher than in the older cohorts (Huinink et al. (2011)). This led to relatively similar sample sizes over cohorts in the pairfam base sample. An additional sample of people living in the eastern part of Germany (DemoDiff) was integrated post-hoc into the pairfam study. If both samples would be analyzed unweighted, (1) a combination of these two data sets would over-represent residents of Eastern Germany and (2) the youngest cohort would be under-represented as this cohort was not sampled in DemoDiff. In the wave 11 refreshment sample, no new cases were drawn from the oldest cohort (1970-73), but a new, younger cohort (2000-03) was integrated into the pairfam sample. If, for instance, an unweighted analysis of the wave 11 survey would be pursued over all cohorts, (3) the oldest cohort would be under- and the youngest over-represented.

Accordingly, inclusion probabilities must be addressed for analyses including all samples into one data set. This can be done by identifying the different “frames” and applying a composite estimator in which weights of multiple frames are combined in the ratio of their respective net sample sizes. (Brick et al. (2011); Lohr and Rao (2000); Sand (2018)). Adapted to the pairfam survey, three samples have been drawn from three different frames:

- F1 pairfam base frame: Representative of the German population of three age cohorts (1971-73, 1981-83, 1991-93)
- F2 DemoDiff frame: Representative of the population living in Eastern Germany in two age cohorts (1971-73, 1981-83)

F3 pairfam refreshment frame: Representative of the German population of three age cohorts (1981-83, 1991-93, 2001-03)

As the analysis sample depends on the research question at hand, four design weights are provided based on the time of sample selection:

1. $dweight_w$ Initial pairfam design weights adjusting only the pairfam base gross sample to the German population for the three initial cohorts
2. $d1weight_w$ Combined pairfam base sample and DemoDiff sample design weights adjusting both samples to the German population for the three initial cohorts
3. $d2weight_w$ Combined pairfam base, DemoDiff, and refreshment sample design weights adjusting all gross samples to the German population for all four cohorts
4. $d3weight_w$ Refreshment sample design weights adjusting the wave 11 refreshment gross sample to the German population for three cohorts

All design weights are standardized stepwise for each cohort i (and at each observation t) as follows, leading to a mean $\bar{w}_t = 1$ and a sum of all weights of the sample size ($\sum w = N_{T,NS}$).

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

with $N_{i,P}$ 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.

4.6.2 Calibrated design weights

Note that ready-to-use combinations of calibration and design weights are provided for all samples, as using the respective design weights is a precondition of applying calibrated design weights. Accordingly, users do not need to combine these further.

Each of the calibrated design weights aims to adjust the data to the target population and simultaneously control for baseline survey participation and panel attrition bias. To achieve this, an iterative proportional fitting (IPF) approach was used, applying the ipfranking package (Kolenikov (2014, 2019)) for Stata to successively identify higher or lower weights until an optimal adjustment to the reference data (Mikrozensus) was achieved.

The following calibration variables from the Mikrozensus were used to define the reference characteristics:

1. Gender: male, female
2. Federal state (Bundesland): 14 categories²⁸
3. Education level: no or primary education (Hauptschule), lower secondary (Realschule), higher secondary (Abitur), still in education
4. Migration background: none vs. first generation or second generation
5. Settlement structure: 8 categories²⁹

²⁸ Germany has 16 federal states. To avoid small case numbers in cells, Saarland has been combined with Rheinland Pfalz and Bremen with Hamburg.

²⁹ The original "753er systematic" of the BIK has 10-categories. To avoid small case numbers in cells, <2.000 and 2.000 to <5.000, as well as 50.000 to <100.000 rural and urban were collapsed to a joined category.

6. Family status: single, married, widowed/divorced³⁰

7. Child(ren) living in household: none, one, two or more³⁰

All cohorts were considered simultaneously for calibration, which included cross-combinations of cohort with each above-mentioned characteristic. In doing so, calibration weights address both the characteristics in each cohort and the overall ratio of the cohorts to each other. Weights were constructed considering the respective design weights, which resulted in four different calibrated design weights based on four analytical samples and their respective design weights:

1. $cdweight_w$ Calibrated design weight adjusting the pairfam base sample to central characteristics of the German population for the three initial cohorts
2. $cd1weight_w$ Calibrated design weight adjusting the combined pairfam base and DemoDiff samples to central characteristics of the German population for the three initial cohorts³¹
3. $cd2weight_w$ Calibrated design weight adjusting the combined pairfam base sample, DemoDiff, and refreshment sample to central characteristics of the German population for all four cohorts
4. $cd3weight_w$ Calibrated design weight adjusting the pairfam wave 11 refreshment sample to central characteristics of the German population for three cohorts

For each cohort i , the calibrated design weights were calibrated at the net sample size $N_{i,NS}$ of each wave w , resulting in a mean of the weight $\bar{w}_i = 1$ at each wave.

$$\hat{w}_i = w_i * \frac{N_{i,NS}}{\sum w_i}$$

4.6.3 Best implementation of weights

“To weight or not to weight?” is not a question which can be addressed here adequately (see Bollen et al. (2016); Gelman (2007); Solon et al. (2015)). The survey design of the full pairfam sample is complex in terms of sample selection (i.e., by cohort, East/West, and base/refreshment) and attrition patterns (Brüderl et al. (2023)). Accordingly, not addressing these complexities in an unweighted approach carries a larger risk of mis-specification than the lower efficiency and lower statistical power often associated with using weights (Bollen et al. (2016)). However, if only singular samples will be applied (e.g., pairfam base only), controlling for cohort differences and variables that might predict selective drop-out could also be an appropriate approach.

If analyzing all samples (pairfam base sample, DemoDiff, pairfam wave 11 refreshment), using calibrated design weights is highly recommended. All weights are part of each anchor data set, allowing users to apply weights for each sample in each wave. The following commands are recommended for weighting data with Stata using the survey commands:

```
* Insert this before relevant command lines
svyset [pweight=cd2weight]
* Commands
svy: tabulate wave sex
svy, subpop (if cohort==1): tabulate wave sex
```

The applied calibrated design weight ($cd2weight_w$) is applicable for the use of the full data set. For singular sample analyses (e.g., refreshment sample only), please use the other calibrated design weights

³⁰ Due to the distribution of this variable, it has not been considered for wave-cohort combinations in which the youngest cohort was under 21. This age limit is based on characteristics in the population (Mikrozensus) indicating increasing diversity starting at that age. Therefore, family status and number of children in the household are not included for cohort 1 (1991-93) before wave 7 (2014), and never for cohort 4 (2001-03).

³¹ Note that DemoDiff contributes information for only the two oldest cohorts.

cdweight_w, cd1weight_w, and cd3weight_w, respectively.

More detailed examples are provided in the Quick Start file "Weighting".

In particular, in (descriptive) analyses in which it is difficult to control for the stratification by cohort and over-sampling of individuals living in Eastern Germany, design weights are necessary. They correct for differential selection probabilities (i.e., population to gross sample), but not for non-response (i.e., gross to net sample). As non-response bias is often an issue, we recommend the use of the calibrated design weights, which address issues regarding systematic non-response for the above-mentioned calibration variables. Nevertheless, design weights are provided – in particular for advanced users who may want to model a non-response correction themselves.

Please keep in mind that as the pairfam study observes distinct birth cohorts that could differ significantly in behavior/characteristics, estimating point estimates (e.g., mean levels) over multiple cohorts might not be an adequate approach for most research questions. Often, presenting cohort-specific results is a more informative choice. In both cases, using the included weights is appropriate.

If only particular sub-samples of the pairfam study are of interest (e.g., cohort 1991-93, residents of Eastern Germany, men, parents), weights are still appropriate. Finally, even if you are not using weights in your analyses, we recommend additional analyses with calibrated design weights as a robustness check. If results are similar, you probably have no problem with design/nonresponse/attrition bias. If results differ, further thinking is warranted. Please keep in mind that the design weights do not control for unequal cohort sizes, neither in the net sample nor in the population (whereas calibration weights do).

4.7 Generated data sets

In wave 1, respondents of the pairfam base sample (and in wave 11 respondents of the refreshment sample) answered questions regarding their relationship history from age 14 and their complete fertility history. As of wave 2, an electronic Event History Calendar (EHC) was implemented in each wave to collect retrospective and prospective information on the life domains of partnership, children, place(s) of residence, and education and employment activity (see also 4.1). The information on these individual life history domains, as well as information on parents, were compiled into separate user-friendly data sets to allow for a convenient analysis of biographical information (*biopart*, *stepup_biopart*, *biochild*, *stepup_biochild*, *biomob_ehc*, and *bioact*). As of Release 5.0, the data sets include previous DemoDiff respondents as well. For Release 14.0 and 14.1, the episode data sets were not supplemented with the information from wave 14 due to the different question program (only errors in the versions of Release 13.0 have been corrected). The panel data set *Overview_multi_actor* has been updated for Release 14.0.

The episode data set *biopart* provides information on individual relationships, cohabitation and marriage history, including both retrospective and prospective information on a monthly basis. In addition, the episode data set *bioact* contains monthly information from the month of the respondent's first interview, covering educational and occupational activities. The data set *bioact_rtr* covers retrospective information collected in wave 3 on education and occupation from the age of eighteen.

The data set *biochild* is a panel data set which additionally includes retrospective episode data collected in wave 1. This data set covers fertility biographies, information on (biological, step, adoptive, and foster) children, and cohabitation episodes with children. In contrast, the data set *household* is a pure panel data set which only contains information pertaining to residence(s) and household members at the time of the anchor interview.

Three panel data sets contain information on the anchor respondent's mobility: *biomob_ehc* (residence information collected through the EHC),³² *biomob_rtr* (retrospective mobility history from the age of 18), and *biomob_rtr_parents* (retrospective information on moving out of the parental house).

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

The data set *biopart* is organized in "long" form, i.e. it contains one row for each partner. In comparison, *biochild* and *household* contain one row for each child or household for each wave. Thus, their format is "long-long". The data sets *bioact* and *bioact_rtr* are also "long-long", however, the rows are not based on waves, but rather on activity episodes. If one activity takes place with one or more interruptions, the data include one row for each episode separately. Each of the three *biomob* data sets *biomob_ehc*, *biomob_rtr* and *biomob_rtr_parents* is organized in "long" format with one row for each episode. The data set *bioparent* is also provided in "long" format, with one row for each parent.

As of Release 6.0, the new panel data set *Overview_multi_actor* provides an overview of participation of (secondary) respondents in the partner, child, parenting and parent surveys.

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

In order to make duration calculations easier, dates within the generated data sets (with the exception of *household*) are stored in a numerical variable which combines both month and year. The value of this variable represents the number of months that have passed since January of 1900. We chose this date as a reference point in order to avoid negative values due to dates previous to January 1960, the baseline date in Stata. The following formula was used to calculate date values: $((Year\ of\ respective\ date - 1900) * 12 + Month\ of\ respective\ date - 1)$. In order to reconvert this information into the original month and year, a new variable must be generated by subtracting 720 (60*12 months; difference between 1900 and the baseline 1960 in Stata). Subsequently, this variable must be transformed into format "%td" by using the "dofm" function: "gen new variable=dofm(old variable)". Finally, two variables containing the respective year and month can be generated using the functions "year()" and "month()".³³

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

If respondents provided information on the year, but not the month, the month value was randomly imputed (except for *household*), taking into account both potential lower and upper boundaries. The imputation were made in the data sets *bioact_rtr*, *biochild*, *biomob_rtr*, *biomob_rtr_parents*, *bioparent* and *biopart*. Cases with imputed values are indicated by flag variables in the respective data sets. Please note that in cases of ambiguous seasonal information regarding the month but valid year information for date variables, random values within the following bounds were imputed for the month:

³²Please note: Until Release 11.0, *biomob_ehc* included only moves between municipalities. Since Release 12.0, moves within the same municipality have now also been integrated into it. Therefore, *biomob_ehc_moves*, the former data set including only moves within the same municipality, has become obsolete.

³³Consequently, 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 random imputation of month information, the value of the last available wave was retained. The standard missing value definition (see Table 4.3) is retained for variables which do not provide episodes or date information. The value “-3 Does not apply” was used for all such variables in all data sets.

For the seam of consecutive waves there are generally several entries surveyed by the EHC, which could deliver inconsistent information. For example, in the case of activities, there is the information “Full-time employment in month 32 (EHC)” (variable *ehc19i10m32*), “Full-time employment currently (EHC)” (variable *ehc19i10*), and in the following wave the preloads and differing entries in the EHC for the time of the previous interview. To meet potential inconsistencies in the generated data sets the most recent information has been used. This means, information on the months given in the EHC (e.g. variable *ehc19i10m32*), is replaced by the current information (e.g. variable *ehc19i10*) and old information on the month of the previous interview is replaced by the more recent information given in the following interview.

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

- person number (ID), date of interview of all (relevant) waves, sex, birth cohort, sample indicator and date of birth

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

The Stata do-files to produce the respective data sets start from the *anchor\$* data sets, and are provided as part of the Scientific Use File.

4.7.1 Anchor-partner episode data - biopart

The data set *biopart* (and *stepup_biopart* for the step-up anchor respondents) contains retrospective and prospective information regarding anchor’s partnership, cohabitation, and marriage episodes from the age of 14 over all available waves, with one row for each partner (“long” format). With each new wave, *biopart* is updated, meaning information on existing partners is potentially renewed and/or new partners are added. Table 4.27 displays the complete list of variables in this data set. Please note that only those anchors are included in this dataset who had reported at least one partnership.

For each partner the beginning of the very first (variables *relbeg* and *cohbeg*), as well as the end (variables *relend* and *cohend*) of the last relationship or cohabitation episode were coded as the overall beginning and end of the respective episode. If the anchor reported more than two episodes with the same partner, the beginning (variable *bkbeg* and *bkcohbeg*) and end (variables *bkend* and *bkcohend*) of each break are also stored in the data, sorted in ascending order according to the beginning date.³⁴ In addition, we include the beginning (variable *marbeg*) and end (variable *marend*) dates of each marriage and the type of marriage ceremony (variable *marcer*). As only one marriage episode per partner has thus far been recorded, there is no information on breaks between marriage episodes.

Please note that the variables *relend*, *cohend*, and *marend* may be coded as “-99 Ongoing” even if respondents did not participate in the most current wave, and we cannot ensure the correctness of

³⁴Please 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.

this information (right-censoring). If they did participate in the most recent interview but there is some degree of ambiguity with respect to the end of the episode, the value “-7 Incomplete data” is assigned. For a few cases of the refreshment sample, there has been contradicting information on the current marital status in wave 11. According to *sd10*, those respondents reported to be married, but no current marriage is mentioned in the EHC for the corresponding partners. In these cases *marend* is coded “-7 Incomplete data”.

In addition to the date of birth (variable *dob*, based on the generated identifiers *doby_gen* and *dobm_gen*), the variables *sex* (based on the generated identifier *sex_gen*) and *homosex_p*³⁵ are included. We also generated the dummy variables *respwx* (with x=number of wave) to indicate whether the anchor took part in the respective wave (no attrition). The variable *wavex* (with x=number of wave) indicates in which year the respective interview was conducted.

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

The original serial number of each respective wave (value “x” of variables “varpx” in the anchor data sets) is included as well with the variables *pnwx*. This facilitates the use of additional information from the anchor data sets. In wave 1, partners were numbered consecutively, with the current partner assigned “0”. In wave 2, the partner who was also the reported partner in wave 1 received “1”, new partners the value “2”, and partners between waves values “3” to “5”. As of wave 3, the auxiliary variable *hpnr* in the *anchor\$* data sets contains the running number of each current partner.

In order to indicate whether a specific partner was the reported current partner in wave x, a dummy variable *partcurrwx* (with x=number of wave) was included.³⁶ This variable assumes value “1” only if the anchor identified the respective partner as his/her current partner at the time of the interview in wave x. In a few cases, a change of the partner status has occurred at the seam of two waves, which lead to the generation of *relbeg*, indicating a partnership at the time of the interview but *partcurrwx* as “0”.

For all partners, the variable *sexp* contains the best information on the partner’s sex (see section 2.2). Partner’s date of birth *dobp* was included when reported at the time of at least one interview (variable *dobp*).³⁷ The identification number *pid* only exists for partners who (potentially) took part in the partner survey in at least one wave; i.e. only partners whose partnership existed at the time of an interview (current partners) were assigned a *pid*.

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

Please note that it was not possible to reidentify new partners in wave 2 as previous partners reported in wave 1. As of wave 3, the variable *pa30* specifies whether the anchor had already a relationship with the supposedly new partner. This information was used to detect identical partners, using names and dates of birth, for the data set *biopart*. Beginning with wave 4 the variables *pa31*, *pa32*, *pa33m*, and *pa33y* in the anchor data sets indicate whether and, if so, how and when a marriage reported in a previous wave ended. This information regarding previous partners was integrated into *biopart* as well.

³⁵ Generated variable “Homosexual union” based on variables *sex* and *sexp*.

³⁶ 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.

³⁷ In case of inconsistencies between different versions of in different waves for the same partner, the latest version is employed.

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

Variable	Variable label	Values	Value labels
<i>General information</i>			
id	Person number anchor		-
demodiff	DemoDiff sample	0	pairfam
		1	demodiff
sample	sample indicator	1	pairfam base sample
		2	DemoDiff sample
		3	Refreshment sample
cohort	Birth cohort	1	1991-1993
		2	1981-1983
		3	1971-1973
		4	2001-2003
intmodew12	Interview mode due to Covid-19 pandemic (Wave 12)	-3	Does not apply
		1	CATI & CASI
		2	CATI & PAPI
intmodew13	Interview mode due to Covid-19 pandemic (Wave 13)	-3	Does not apply
		1	CATI & CASI
		2	CATI & PAPI
intdatwx	Date of interview wave x	-10	Not in demodiff
		-3	Does not apply
		<i>date</i>	
sex	Sex anchor	-7	Incomplete data
		1	Male
		2	Female
dob	Date of birth anchor	<i>date</i>	-
wavex	Survey year: wave x	-10	Not in demodiff
		-3	Does not apply
		1	2008/2009
		2	2009/2010
		3	2010/2011
		4	2011/2012
		5	2012/2013
		6	2013/2014
		7	2014/2015
		8	2015/2016
		9	2016/2017
		10	2017/2018
		11	2018/2019
		12	2019/2020
		13	2020/2021
respwx	Respondent in wave x	0	No
		1	Yes
<i>Ascending order of relationships</i>			
partindex	Correct order of relationships:	-3	Does not apply
	Number partner	1	1st partner
		2	2nd partner
	
		14	14th partner

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
<i>Partners at the time of the previous interview</i>			
pid	Person number current partner	(Tab. 2.2)	
dobp	Date of birth partner	-7 -3 <i>date</i>	Incomplete data Does not apply -
<i>Relationship history</i>			
pnowx	Number partner in wave x	-3 <i>number</i>	Does not apply -
partcurrwx	Current partner in wave x	-10 0 1	Not in demodiff No Yes
sexp	Sex partner	-7 1 2	Incomplete data Male Female
homosex__p	Homosexual union	-7 0 1	Incomplete data Heterosexual union Homosexual union
dodp	Date of death partner	-7 -3 <i>date</i>	Incomplete data Does not apply -
relbeg	Beginning relationship	-7 <i>date</i>	Incomplete data
relend	End relationship	-99 -66 -7 -3 <i>date</i>	Ongoing Death partner Incomplete Data Does not apply -
bkbeg	Beginning break k relationship	-7 -3 <i>date</i>	Incomplete data Does not apply -
bkend	End break k relationship	-7 -3 <i>date</i>	Incomplete data Does not apply -
<i>Cohabitation history</i>			
cohbeg	Beginning cohabitation	-7 -3 <i>date</i>	Incomplete data Does not apply -
cohend	End cohabitation	-99 -66 -7 -3 <i>date</i>	Ongoing Death partner Incomplete data Does not apply -
bkcohbeg	Beginning break k cohabitation	-7 -3 <i>date</i>	Incomplete data Does not apply -
bkcohend	End break k cohabitation	-7 -3 <i>date</i>	Incomplete data Does not apply -
<i>Marriage history</i>			
marbeg	Beginning marriage	-7 -3	Incomplete data Does not apply

continued on next page

continued from previous page

continued from previous page

Variable	Variable label	Values	Value labels
marend	End marriage	<i>date</i>	-
		-99	Ongoing
		-66	Death partner
		-7	Incomplete data
		-3	Does not apply
marcer	Marriage ceremony	<i>date</i>	-
		-7	Incomplete data
		-3	Does not apply
		1	Only a civil ceremony
		2	Civil and religious ceremony
		3	Only a religious ceremony
<i>Imputations</i>			
imp_unionbeg	Imputed date of beginning union / end of break of union	0	No imputation
		1	Only year information
		2	Only season information
imp_unionend	Imputed date of end union / end of break of union	0	No imputation
		1	Only year information
		2	Only season information
imp_cohabbeg	Imputed date of beginning cohabitation / end of break of cohabitation	0	No imputation
		1	Only year information
		2	Only season information
imp_cohabend	Imputed date of end cohabitation / end of break of cohabitation	0	No imputation
		1	Only year information
		2	Only season information
imp_marbeg	Imputed date of beginning marriage	0	No imputation
		1	Only year information
		2	Only season information
imp_marend	Imputed date of end	0	No imputation
		1	Only year information
		2	Only season information
<i>biopart flag variables</i>			
biopartflag1	Inconsistency biopart: Marriage earlier than beginning of relationship	0	No inconsistency
		1	Inconsistency
biopartflag2	Inconsistency biopart: Overlapping cohabitation episodes with different partners	0	No inconsistency
		1	Inconsistency
biopartflag3	Inconsistency biopart: Beginning current and end previous marriage	0	No inconsistency
		1	Inconsistency
biopartflag4	Inconsistency biopart: Year of birth partner	0	No inconsistency
		1	Younger than 10 years old
		2	Year of birth after beginning of relationship with anchor

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

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

- beginning of subsequent episode prior to end of preceding episode
 - end of subsequent episode prior to beginning of preceding episode
 - end of subsequent episode prior to end of preceding episode
 - identical beginning and end of two episodes
- Beginning of marriage before beginning of relationship (also see *flag5*, Table A.26)
 - Overlapping episodes of cohabitation with same or different partners (also see *flag6* and *flag7*, Table A.26)
 - End of previous marriage after beginning of current marriage (also see *flag8*, Table A.26)
 - Divorce from partner to whom never married (also see *flag10*, Table A.26)
 - Separation through death/divorce current spouse (also see *flag11*, Table A.26)
 - Separation before beginning relationship (current partner; also see *flag14*, Table A.26)

With respect to information collected in wave 1 (retrospective partnership history and prospective information), we also checked whether the beginning of the episode (relationship/cohabitation/marriage) was prior to the first meeting of the anchor with 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 sets *anchor\$* from wave 2. These variables mark inconsistencies regarding the first meeting and beginning of a partnership episode.

Solution to these inconsistencies:

We have attempted to eliminate inconsistencies as far as possible. In addition to sorting 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 modification seemed plausible and necessary. Therefore, new variables (*imp_unionbeg*, *imp_unionend*, *imp_cohabbeg*, *imp_cohabend*, *imp_marbeg* and *imp_marend*) are included in the *biopart* data set to mark these imputations for the beginning and end date information. If two or more episodes had identical dates, one of them was dropped. Moreover, episodes completely contained within another concerning the same partner were also eliminated. Any episode breaks coded with “-1” or “-2” for both the beginning and ending dates were also dropped. The variables *biopartflag1*, *biopartflag2*, *biopartflag3*, and *biopartflag4* were generated to mark these inconsistencies.

Partners and corresponding information about cohabitation and marriage episodes were dropped if:

- the current partner was mistakenly provided as previous partner as well,
- the name of the previous partner equaled name current partner, and the relationship beginning is identical or almost identical (often the case if end previous relationship = date interview or “-1/-2”).

Sources of additional information about respective partners:

As of wave 2, three types of partners have been differentiated:

1. Retrospective partners: Partners whom the anchor respondent mentioned as part of the retrospective partner history but who were not reported as current partners by respondents of the pairfam base sample and the DemoDiff sample in wave 1 or by respondents of the refreshment sample in wave 11.
2. In-between-waves partners: Partners who were listed as previous partners in the course of the EHC, which captures the time between the previous and the current wave, but who were not reported as current partners in wave 2 to 13.

3. Partners at the time of previous interviews: Partners who were reported as current partners at the time of at least one interview.

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

The variable *partcurrwx* provides the wave number to which questions in the anchor data sets regarding the current partner or current relationship refer. Information from the partner survey (PAPI questionnaire) is saved in the respective *partner\$* data set for each wave, and can be merged using the identification variable *id*. If applicable, the data set *parentingx* includes information from the partner's parenting survey which can also be matched using *id*.

Questions in the anchor data sets from the module "Meeting the current partner" (variables *pa1_-pa3*, *sdp1-sdp21*) refer to current partners whose value for *partcurrw\$* is "0" in the previous wave and "1" in the current wave (*partcurrw\$*: "0" at time t-1 and "1" at time t).

The module "Separation from the ex-partner" (variables with the prefix "sep" from wave 2) covers partners who were reported current partners in the previous wave, but were not reported current partners in the respective wave and who had not passed away since the previous wave (*partcurrwx*: "1" at time t-1 and "0" at time t).

For further information regarding the computation of this data set, please see the Stata do-files *biopart1-2.do* (data from waves 1 and 2) and *biopart3.do* up to *biopart13.do* (data from wave 3 up to wave 13).

4.7.2 Anchor-child panel/episode data - biochild

The data set *biochild* (and *stepup_biochild* for the step-up anchor respondents) contains retrospective and prospective information on the anchor's children³⁸ and episodes of cohabitation with these children in the same household. This data set is in "long-long" format, with one row per child, and for each child one row per wave in which the respective child was reported. The data set is updated in each wave, meaning the information on existing children is renewed and if necessary, new children are added. The current data *biochild* set contains information from wave 1 to 13 of the pairfam base sample, DemoDiff, and refreshment respondents and *stepup_biochild* contains information from wave 4 to 13 of the step-up sample. Table 4.28 depicts the variables included in this data set.

In order to identify the children, the variable *number* contains the value X on the X-th child, corresponding to the variables *varX* in the *anchor\$* data sets.³⁹ Since Release 7.0, for all children a child ID (variable *cid*) is available in the data set. The dummy variable *surveykid* indicates whether a child was interviewed as part of the CAPI child survey. Thus, this variable provides information on the success of merging the *anchor\$* files with *child2* to *child13*, respectively.

For all children, the variable *sexk* contains the best information on the children's sex (see Section 4.3).⁴⁰ The variable *dobk* depicts the best information on the child's date of birth, provided as number of months passed since January 1900. In case of seasonal or missing information for the month but a non-missing year of birth, the month information was randomly imputed. For further details on this principle and the generation of *dobk*, see Section 4.7 above.

³⁸ According to the anchor questionnaires, the term "children" refers to biological children, adopted children, children of a partner, or foster children.

³⁹ 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*.

⁴⁰ 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 child age (variable *dobk*). The oldest child was labeled “1 1st child”, the second oldest “2 2nd child”, and so forth. In the case of a missing date of birth, the value “-7 Incomplete data” was assigned, and the order was built according to the remaining information of the other children, if existent.

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

The anchor persons were asked whether one of the mentioned partners was also the second parent of the biological child and, if so, which partner. The variable *pnowx* contains the original serial number for that partner in the respective wave (value X of variables *varpX* in the anchor data sets). If it was not one of the mentioned partners, the variable *pnowx* is coded with “97 Another person”. In contrast to the variables *pnowx* in the data set *biopart*, *pno* contains not all, but only the serial numbers of partners who are also parents.

The variable *parentid* contains the identification number (*pid*) of the child’s second biological parent (not anchor). In case a child is an anchor’s step child, *parentidstepk* contains the identification number (*pid*) of the child’s biological mother or father. In order to identify this parent as a former partner of the anchor respondent, the variables *parentid* and *parentidstepk* are constant over all waves for each child. Note that these variables can differ from the variable *pid*, as the anchor’s relationship to the second parent can result in separation or a new partnership may emerge. In some cases, information on the identity of the second biological partner varied between waves and the *parentid* had to be adjusted accordingly. The variable *flag_parentid* marks these inconsistencies. In the first wave the retrospective partnership history was asked, so sometimes a *pid* is assigned, but *parentid* is missing. This is the case, when a former partner is the second biological parent and the anchor person was in a new relationship at the time of the interview of wave 1. A *pid* in the anchor data sets was only assigned to current partners.

Please note: Only when a biological child was added by the anchor in the EHC he or she was asked who the second parent was (see *ehc12k*). If a stepchild was newly added by the anchor person, we assumed that it was the biological child of the anchor’s current partner at that time and therefore there is a small risk that in some cases the partners identified by *parentidstepk* are not the biological parents.

Concerning retrospective information collected in wave 1 (pairfam base sample and DemoDiff) and wave 11 (refreshment sample), note that the variables *livk*beg*, *livk*end*, *b1livk*beg*, *b1livk*end*, and *livk*birth* contain retrospective information only from two waves: wave 1, for the respondents of the pairfam base or the DemoDiff sample, or wave 11, for the respondents of the refreshment sample. The term “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.). In case of ambiguous or unknown information on the month random values were imputed. The variables *imp_livkbeg* and *imp_livkend* are included to mark these cases.

Table 4.28: List of variables included in data set *biochild*

Variable	Variable label	Values	Value labels
id	Person number anchor	(Tab. 2.2)	-
sample	Sample indicator	1	pairfam base sample

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
		2	DemoDiff sample
		3	Refreshment sample
dob	Date of birth anchor	<i>date</i>	-
intdat	Date of interview (months since January 1900)	<i>date</i>	-
sex	Sex anchor	1	Male
		2	Female
cohort	Birth cohort	1	1991-1993
		2	1981-1983
		3	1971-1973
		4	2001-2003
intmode_w12	Interview mode due to Covid-19 pandemic (wave 12)	-3	Does not apply
		1	CAPI & CASI
		2	CATI & PAPI
intmode_w13	Interview mode due to Covid-19 pandemic (wave 13)	-3	Does not apply
		1	CAPI & CASI
		2	CATI & PAPI
wave	Survey year	1	2008/09
		2	2009/2010
		3	2010/2011
		4	2011/2012
		5	2012/2013
		6	2013/2014
		7	2014/2015
		8	2015/2016
		9	2016/2017
		10	2017/2018
		11	2018/2019
		12	2019/2020
		13	2020/2021
number	Pointer on Xth child (corresponding varkX in anchor data)	1...12	-
cid	Person number CAPI-child	(Tab. 2.2)	-
surveykid	Indicator: child was asked in CAPI child survey	-7	Incomplete data
		-3	Does not apply
		0	No
		1	Yes
index	Correct order of children (corresponding to date of birth)	-7	Incomplete data
		-3	Does not apply
		1	1st child (oldest)
	
		12	12th child
dobk	Date of birth of child (months since January 1900)	-7	Incomplete data
		<i>date</i>	-
sexk	Gender of child	-7	Incomplete data
		1	Male
		2	Female
statusk	Status of child	-7	Incomplete data
		1	Biological child
		2	Adopted child
		3	Partner's child /stepchild

continued on next page

continued from previous page

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

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
	with child	<i>date</i>	-
b1livkxend	End break 1 living together with child	-3	Does not apply
		<i>date</i>	-
livkbirth	Living together since birth with child	-7	Incomplete data
		-3	Does not apply
		0	Not living together since birth
		1	Living together since birth
imp_livkbeg	Imputed date of beginning living together	-3	Does not apply
		0	No imputation
		1	Only year information
		2	Only season information
imp_livkend	Imputed date of end living together	-3	Does not apply
		0	No imputation
		1	Only year information
		2	Only season information
dodk	Date of death child (months since January 1900)	-7	Incomplete data
		-3	Does not apply
		<i>date</i>	-
flag_parentid	Changing information about second biological parent	0	No change
		1	parentid changed to missing
		2	parentid changed

Various additional checks have been implemented to identify inconsistent cases.

Checks for episodes of living together with children:

- Date of child's birth after beginning to live together
- Beginning to live together after breaks in living together
- Beginning to live together 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
- Identical end of living together and date of interview.

As no such inconsistencies were found, no changes were made.

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

Consistency checks across wave 1 to 13:

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

- Child status: If there are unrealistic differences between the information in waves 1 to 13, the newest information is preferred. Such unrealistic differences were changes in status, e.g. from or to the status “biological child”.

Matching with further information on children:

One general possibility to merge information from the other data sets is to use the anchor’s identification variable *id*. By using the anchor’s parents’ identification variables (*mid*, *fid*, *smid*, and *sfid*), the data set *parent\$* can be matched, and information reported by the grandparents of the child can be added. Note that this only provides information on the CAPI children. To add information on the parents themselves, 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 must be temporarily renamed to *pid*. Otherwise, the matching process will fail, as the other data sets do not contain the variable *parentid*. This variable exists only for a quick differentiation between the general partner identification number (*pid*) and identification number of the second biological parent (*parentid*).

The do-file *biochild1-2.do* documents how the first parts (waves 1 and 2) of the data were modified and how inconsistency checks in the retrospective sections were generated. The do-files *biochild3.do* through *biochild13.do* show the preparation of the wave 3 to 13 data, respectively, as well as the combination of data from all waves which finally results in the data set *biochild*.

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

The data set *bioact* contains basic information on all anchor activities in the areas of education and employment, starting from the time of the wave 1 interview. It includes respondents of the pairfam base sample, DemoDiff, and refreshment sample. It covers the period between the survey data of the first and the current wave. The data set is provided in “long-long” format, meaning it consists of one row for each education and/or employment episode. More than one activity per respondent and several instances of one activity are possible, depending on the number of mentioned activities. Wave 3 data also contain retrospective information on education and employment starting from the age of eighteen. This information can be found in *bioact_rtr*.

Starting with wave 2, the anchor data sets contain variables that store information about education and employment separately for each month (Event history calendar). For the activity calendar, respondents were asked to indicate what they had been doing each month after the preceding wave up to the current interview. The idea is to acquire a full overview of what has happened concerning education and employment during the period of - on average - the last year.

In the *bioact* data, the variable *activity* displays the kind of activity the anchor respondent reported. Table 4.29 displays all possible activities covered by the questionnaire. In addition, the *bioact* data set provides the following information on the duration of each activity: The variables *actbeg* and *actend* indicate the start and the end month of an episode reported by the respondent. The variable *actcurrwx* (with x=number of wave) marks the activity episodes which were current at the time of the interview of the respective wave. Additionally, *actspell* presents the ascending number of each episode per activity. The variable *actcensor* indicates if and in which way the episodes were censored, i.e. if an episode began before or lasted longer than the period covered by the EHC, or if information on the month right before or after the episode were missing or unknown. For more details, see Table 4.29.

Furthermore, *bioact* and *bioact_rtr* data include the anchor’s date of birth (*dob*) and the interview dates of the current and preceding waves for identification purposes. The dummy variables *respwx* indicate whether the anchor took part in the respective interview (identification variable for temporary dropouts).

The structures of the two data sets *bioact* and *bioact_rtr* are almost identical. The variables carrying the activity information in the *bioact_rtr* data set are marked with the suffix “_rtr”. The retrospective information in wave 3 was not collected with the EHC, thus, there is ambiguous seasonal information for the beginning and end of episodes. In such cases, random values were imputed. The original ambiguous seasonal information is provided with the variables *actflag1_rtr* (beginning) and *actflag2_rtr* (ending). If respondents answered “none of the above” to questions on education or employment, that information can be found in the variables *rtr31i10* (education) and *rtr35i14* (occupation), as no further data regarding the beginning or end were collected in these cases.

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

Table 4.29 displays the full set of variables which are included in these data sets, in contrast to Table 4.30, which shows all relevant variables included in the anchor data. These variables can be easily merged with variables from anchor data by using the key variable *id*.

Concerning the variables *ehc19i23* and *ehc19i23m** from the anchor data set (which were used to generate the *bioact* data) additional alterations needed to take place. If respondents have gaps in their calendar and clicked the “finish”-button, a pop-up appeared and they were 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 the variables *ehc19i23* and *ehc19i23m** in the anchor data. These variables were constructed to act as “gap fillers” for months in which respondents could not recall what they had done.

In the progress of data editing, *ehc19i23* and *ehc19i23m** were recoded to “1” for all cases with gaps in the activity calendar if no activity applied for a given month. In contrast, the variables were recoded “0” if at least one activity was mentioned per month. Thus, in the final data these variables work as “gap fillers” as originally intended. However, we recommend considering *ehc19i23* and *ehc19i23m** when using monthly activity information contained in anchor data and not the *bioact* data.

The do-file *bioact2-13.do* documents in detail how the data set *bioact* was computed. The do-file *bioact_rtr.do* documents how the data set *bioact_rtr* was computed.

Table 4.29: List of variables included in data sets *bioact* and *bioact_rtr*

Variable	Variable label	Values	Value labels
<i>General information</i>			
<i>id</i>	Person number anchor	(see Table 2.2)	-
<i>sample</i>	Sample indicator	1	pairfam base sample
		2	DemoDiff sample
		3	Refreshment sample
<i>intmodew12</i>	Interview mode due to Covid-19 pandemic (wave 12)	-3	Does not apply
		1	CAP1 & CAS1 (as in waves 1-11)
		2	CAT1 & PAP1
<i>intmodew13</i>	Interview mode due to Covid-19 pandemic (wave 13)	-3	Does not apply
		1	CAP1 & CAS1 (as in waves 1-11)
		2	CAT1 & PAP1
<i>cohort</i>	Birth cohort	1	1991-1993

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
		2	1981-1983
		3	1971-1973
		4	2001-2003
dob	Date of birth	<i>date</i>	-
sex	Sex anchor	-4	Filter error / Incorrect entry
		1	Male
		2	Female
wavex	Survey year wave x	-3	Does not apply
		1	2008/2009
		2	2009/2010
		3	2010/2011
		4	2011/2012
		5	2012/2013
		6	2013/2014
		7	2014/2015
		8	2015/2016
		9	2016/2017
		10	2017/2018
		11	2018/2019
		12	2019/2020
		13	2020/2021
intdat	Date of interview	-10	Not in DemoDiff
		-3	Does not apply
		<i>date</i>	-
<i>Activity history</i>			
activity(_rtr)	Type of activity	1	General secondary school (first education)
		2	Evening school, working on a school leaving certificate for adults
		3	Vocational training (apprenticeship, business school etc.)
		4	Vocational retraining / further education
		5	University of cooperative education (also "Berufsakademie")
		6	University of applied sciences, college, university
		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

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
		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
rtr31i10	No education since	0	Not mentioned
(bioact_rtr only)	18th birthday and wave 1	1	Mentioned
rtr35i14	No occupation between 18th birthday and wave 1	0	Not mentioned
(bioact_rtr only)		1	Mentioned
actspell(_rtr)	Counter - Number of spells per activity	<i>number</i>	-
actbeg(_rtr)	Beginning of activity in month	<i>date</i>	-
actend(_rtr)	End of activity in month	97 <i>date</i>	ongoing -
actcensor(_rtr)	Indicator for censored spells	-3 0 1 2	No censoring (only for ehc19i23*) Uncensored Left-censored Right-censored, end of episode is unknown
		3	Right-censored, ongoing episode
		4	Combination of 1 & 2
		5	Combination of 1 & 3
		6	Left-censored, beginning is unknown
		7	Combination of 3 & 6
actcurrwx	Current activity in wave x	-10 -3 0 1	not in demodiff Does not apply No Yes
respwx	Respondent in wave x	-10 0 1	Not in DemoDiff No Yes
actflag1_rtr (bioact_rtr only)	Ambiguous information on beginning	-3 21 24 27 30 32	Does not apply Original information on the month Original information on the month Original information on the month Original information on the month Original information on the month
actflag2_rtr (bioact_rtr only)	Ambiguous information on ending	-3 21 24 27	Does not apply Original information on the month Original information on the month Original information on the month

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
		30	Original information on the month
		32	Original information on the month

Table 4.30: List of available variables covering education and occupation in anchor data set

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

4.7.4 Anchor-household panel data - household

The data set *household* contains information on the anchor's main residence, household members, and household income. The data set uses information from the anchor data sets over all waves and combines them into a panel data set with one row per household per wave. As we have thus far only used information on the main residence, there is only one household per wave, and therefore one row for each wave in which the anchor participated in the data. Data processing is documented in the Stata do-files *household2.do* and *household3.do*. The generated household data set was not continued after wave 3 as it was too time consuming to prove if household members were identical over waves. Furthermore, the household matrix was improved starting with wave 3 so that it is easier for users to generate the composition of the household and identify changes over waves.

Table 4.31 shows the variables contained in the data set. For each household member mentioned, we generated variables indicating the relationship of the anchor to that household member, their sex, and date of birth (year and month). In contrast to wave 2, we generated variables for different relationships. In wave 3 we distinguished between characteristics of partners (*p**), children (*c**), parents (*pa**), and other household members (*other**). Variables *p*_rel* contain information on cohabitation with partners. We assigned the value "1" to variable *p1_rel* if the anchor reported that he/she still lives or is living together again with the partner from waves 1 or 2. We assigned the value "1" to variable *p2_rel* if the anchor cohabited with a new partner.

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

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

For all persons (partners, children, parents) for which we had the respective information, we used the generated variables for 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

information gathered in the household grid instead.⁴¹

Furthermore, we matched the person number of the current partner to variables p^*_id and the person number of biological, adopted, or stepparents (mid , fid , $smid$, $sfid$) to the variables pa^*_id from the anchor3 data set. As of wave 3 we also matched the person number of the anchor's child(ren) (cid) to the variables c^*_id . Note that cid is only available for so-called "CAPI children". Additionally, we generated variables c^*_point , which include information about the position of the children in EHC. Moreover, we assigned valid values for sex, date of birth, and person number for household members present in both waves.

Table 4.31: 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: 2010/2011
pid	Person number partner	(see Table 2.2)	-
mid	Person number mother	(see Table 2.2)	-
fid	Person number father	(see Table 2.2)	-
smid	Person number stepmother	(see Table 2.2)	-
sfid	Person number stepfather	(see Table 2.2)	-
cid	Person number capi child	(see Table 2.2)	-
intm	Month of interview	(see Section 4.5)	-
inty	Year of interview	(see Section 4.5)	-
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
fmrdr	Father lives in household main residence	-7	Incomplete data
		0	Father does not live at main residence

continued on next page

⁴¹Information on sex was collected only in wave 1, information on dates of birth only in wave 2. For mothers and fathers, however, we could simply infer sex from the relationship to the anchor person.

continued from previous page

Variable	Variable label	Values	Value labels
childmrd	Number of children in household main residence	1	Father lives at main residence
		-7	Incomplete data
othmrd	Number of other hh members main residence	0...10	-
		-7	Incomplete data
dwtype	Type of household (own, parental, shared, dorm., other)	0...10	-
		-7	Incomplete data
owner	Home-ownership	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
		-7	Incomplete data
		-3	Does not apply
		1	sublease
		2	rented apartment/house
		3	own exclusive property
		4	property of parent(s)
		5	property of another person
		6	property of partner
		7	joint property with partner
nights	Nights spent at residence	8	Other
		-7	Incomplete data
		-3	Does not apply
rent	Monthly rent for dwelling (euros)	1...6	-
		-7	Incomplete data
expown	Monthly expenditures for self-owned apartment / house (euros)	-3	Does not apply
		-7	Incomplete data
dwsize	Size of dwelling (square meters)	-3	Does not apply
		-7	Incomplete data
rooms	Number of rooms of dwelling	-3	Does not apply
		-7	Incomplete data
res2nd	Respondent has 2nd residence	-3	Does not apply
		-7	Incomplete data
hhincnet	Household net income (open and estimated information combined)	0	No 2nd residence
		1	2nd residence
		see <i>hhincgcee</i>	-

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
hhcomp	Household Composition (partner, kid(s), parent(s), others)	-7	Incomplete data
		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
hhincgcee	Net equivalence income (GCEE)	see <i>hhincgcee</i>	-
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

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
paX_sex	Sex of parent X	6	stepmother / father's partner
		7	stepfather / mother's partner
		8	foster mother
		9	foster father
		-7	Incomplete data
		-3	Does not apply
		1	Male
paX_byear	Year of birth of parent X	2	Female
		-7	Incomplete data
		-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
otherX_byear	Year of birth of partner X	2	Female
		-7	Incomplete data
otherX_bmonth	Month of birth of partner X	-3	Does not apply
		-7	Incomplete data
otherX_id	Person number of partner X	-3	Does not apply
		-7	Incomplete data
		-3	Does not apply

Regarding the identification of household members throughout waves, we used the person number for identification of partners and parents, and the position in the EHC for the identification of children. These positions remain constant in each wave, unlike the position in the household grid. In order to identify other household members, we have generated the variables *other*_point*, which include information about the position in the household grid in the first two waves. Using these household members' first names, we were able to detect different household members throughout the waves and assign the respective value to the variables *other*_point*. If the first name was not available, we used their date of birth. If neither variable was available, we decided on a case-by-case basis. Please note that in wave 3, household members' sex is unknown. Thus, in some cases only the relationship to the anchor is available. We accepted the possibility of potential mistakes (wrong sex) because we wanted to avoid classifying these cases as "new" household members. Later, we used the variables *other*_point* to match information from wave 3 to the former waves. Consistency checks showed

that matching did not function properly if names had been modified. Again, we decided individually how to classify these cases.

After completing the editing process for information on household members, we generated aggregate household characteristics. The data can be merged with information from other data sets by using either the anchor ID or the household members' ID (partner, child, parents), who are potential respondents of the respective multi-actor surveys. Please note that the data set *household* does not encompass information beyond wave 3.

4.7.5 Anchor-mobility panel data - biomob*

In addition to information on educational and occupational activities, partners, and the household grid, information on the current main, secondary, and further residences, as well as former residences, has been gathered through the Event History Calendar (EHC). Respondents were asked to provide information on all residences in which they have lived or are still living since the last interview. During the interviews of the second and third wave, respondents were asked to give additional information on their secondary and all further residences. As of wave 4, they were only asked for information concerning their main residence. Thus, information on secondary and any further residences can only be found in the data sets *anchor2* and *anchor3*. The data set *biomob_ehc* has been generated from this data gathered via the EHC in waves 2 to 13.

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

Although all of these data sets contain information on migration and residential changes, the data must be partitioned into more than one data set as the retrospective information does not include secondary residences and a residence coinciding with the parental home does not necessarily imply that no spatial mobility occurred.

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

In the two data sets *biomob_ehc* and *biomob_rtr*, new data have been added: BIK classification *resbik* and migration distance *resdis* (the corresponding variables in the *biomob_rtr* data set are provided with the suffix "_rtr"). Information on places of residence is synchronized with the "Gemeindeverzeichnis" published by the Federal Statistical Office of Germany in 2011). By doing so, the BIK classification and geo coordinates of the geographical centers of the municipalities where respondents were living have been attached. The migration distance *resdis* has been calculated based on these coordinates. The coordinates are not included in the Scientific Use File.

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

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

Φ_A and Φ_B represent the latitudes of the two consecutive places of residence; λ_A and λ_B represent the corresponding longitudes of the two locations. The number 6370 equates to the radius of the Earth in kilometers. The calculated migration distances are rounded accurately to kilometers. In some cases the respondent gave a location name which described several places in the same federal

state or the given location name did not indicate a certain geographical location. In these cases a migration distance could not be calculated and therefore *resdis* was coded “-99 untraceable place”. If the antecedent location was unknown the *resdis* is coded “-98 No calculation possible”. If there is no migration history, *resdis* is coded to “-3 does not apply”.

For the first three waves of DemoDiff there is no information available about the places of residence. For this reason, no migration distances are provided for the first three waves in these cases.

Please note: Until Release 11.0, *biomob_ehc* included only moves between municipalities. Since Release 12.0, moves within the same municipality have now also been integrated into it. Therefore, *biomob_ehc_moves*, the former data set including only moves within the same municipality, has become obsolete.

Biomob_ehc

The data set *biomob_ehc* contains information on residences as collected through the EHC in pairfam and DemoDiff (see Table 4.32). The EHC was first used in the second wave of pairfam and covers the time between the previous and current interview. This data set is provided in long format, i.e. one row for every episode at each residence. For waves 2 and 3, information on secondary and further residences is also included. Since wave 4, no data on secondary or further residences was collected. Until Release 11.0, *biomob_ehc* included only moves between municipalities. Since Release 12.0, moves within the same municipality have now also been integrated into it. Information on third and fourth residences is now also included.

For each episode, both the beginning and end (variables *resbeg* and *resend*) are stored. The variable *resend* is coded to “97 ongoing” for the last reported episode. The variable *rexcurrewx* (with x=number of wave) indicate the residences which were current at the time of the interview of the respective wave. Whether the anchor person has specified one or more places of residence can be determined from the variable *residplace*. To determine if an anchor person has previously lived at a reported new place of residence, an identification number (variable *resnumber*) for each municipality, in which the reported place of residence is located, is generated. This identification is also assigned if a respondent has given place names which cannot be allocated to a certain municipality. The variable *resland* contains information on the federal state, or whether a place of residence is located abroad.

The variable *resid* marks a residence as main, secondary residence or further residence. The variables *index_mr* and *index_sr* indicate the ascending order of episodes at the main and the secondary residences, separately. In some cases it is not possible to determine the main residence based on the information delivered by the anchor. Here, the variable *resflag1* is coded to “1 Yes”. The same issue appears with some secondary residences, in which cases the variable *resflag2* is also coded to “1 Yes”. If possible to indicate a main or secondary residence for a certain period of time, both variables are coded to “0 No”. Some cases involve only incorrect entries, for which *resflag3* is coded to “1 Yes”.

If there have been residential changes, then the distance between the two residences, that is, between the municipalities, is found in *resdis*. Interviewers were advised not to enter a new place of residence into the EHC if respondents moved within the same municipality. Questions on such moves were posed only when a person indicated that he or she had been living continuously in the same place since the last interview. The variables *within_wx* indicate whether the anchor moved within a municipality. In some cases, respondents mentioned a district of a town already mentioned, causing the interviewers to register this district of the same town as a new place of residence in the EHC. Those case are indicated by a migration distance of zero kilometers.

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

Variable	Variable label	Values	Value labels
id	Person number anchor	(see Table 2.2)	-
sample	Sample indicator	1	pairfam base sample
		2	DemoDiff sample
		3	Refreshment sample
cohort	Birth cohort	1	1991-1993
		2	1981-1983
		3	1971-1973
		4	2001-2003
intmodew12	Interview mode due to Covid-19 pandemic (Wave 12)	-3	Does not apply
		1	CATI & CASI (as in wave 1-11)
		2	CATI & PAPI
intmodew13	Interview mode due to Covid-19 pandemic (Wave 13)	-3	Does not apply
		1	CATI & CASI (as in wave 1-11)
		2	CATI & PAPI
dob	Date of birth	<i>date</i>	-
sex	Sex anchor	-4	Filter error / Incorrect entry
		1	Male
		2	Female
wavex	Survey year wave x	-3	Does not apply
		1	2008/2009
		2	2009/2010
		3	2010/2011
		4	2011/2012
		5	2012/2013
		6	2013/2014
		7	2014/2015
		8	2015/2016
		9	2016/2017
		10	2017/2018
		11	2018/2019
		12	2019/2020
		13	2020/2021
respwx	Respondent in wave x	-10	Not in demodiff
		0	No
		1	Yes
intdatwx	Date of interview wave x	-10	Not in DemoDiff
		-3	Does not apply
		<i>date</i>	-
residplace	City mentioned	-10	Not in demodiff
		-7	Incomplete data
		0	No answer
		1	Place mentioned
		2	Several places/Main residence unknown
resbeg	Beginning of episode	<i>date</i>	-
resend	End of episode	97	ongoing
		<i>date</i>	-

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
rescensor	Censor	0	Uncensored
		1	Left censored, beginning of episode first interview
		3	Right censored; ongoing episode
		5	Combination of 1 & 3
		9	Right censored, no secondary residences after wave 3
resnumber	Identification number city	-7	Incomplete data
		<i>number</i>	-
resid	Main or secondary residence	1	Main residence
		2	Secondary residence
		3	Third residence
		4	Fourth residence
resland	Federal state	-99	Place unknown
		-2	No answer
		-1	Don't know
		0	Berlin (West)
		1	Schleswig-Holstein
		2	Hamburg
		3	Niedersachsen (Lower Saxony)
		4	Bremen
		5	Nordrhein-Westfalen (North Rhine-Westfalia)
		6	Hessen (Hesse)
		7	Rheinland-Pfalz (Rhineland-Palatinate)
		8	Baden-Württemberg
		9	Bayern (Bavaria)
		10	Berlin (East)
		11	Brandenburg
		12	Mecklenburg-Vorpommern (Mecklenburg-Western Pomerania)
		13	Sachsen (Saxony)
		14	Sachse-Anhalt (Saxony-Anhalt)
		15	Thüringen (Thuringia)
		16	Saarland
		17	Another country
index_mr	Index main residence	-3	Does not apply
		1	1st main residence
	
index_sr	Index secondary residence	36	36th main residence
		-3	Does not apply
		1	1st second residence
rescurrwx	Current place of residence in wave x
		15	15th second residence
		-10	Not in demodiff
		-3	Does not apply
		0	No
		1	Yes

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
resdis	Distance between consecutive residences	-99 -98 -96 -3 -2 <i>number</i>	Untraceable place No calculation possible Moved abroad Does not apply No answer -
resbik	BIK classification	-3 0 1 2 3 4 5 6 7 8 9	Does not apply City Center - population 500,000+ Periphery - population 500,000+ City Center - population 100,000-500,000 Periphery - population 100,000-500,000 City Center - population 50,000-100,000 Periphery - population 50,000-100,000 Region - population 20,000-50,000 Region - population 5,000-20,000 Region - population 2,000-5,000 Region - population < 2,000
within_w2	Moved within city (ehc18) wave 2	-10 -3 0 1	Not in DemoDiff Does not apply No Yes
within_wx	Moved within city (ehc29) wave x	-3 0 1	Does not apply No Yes
resflag1	Main residence unknown	0 1	No Yes
resflag2	Secondary residence unknown	0 1	No Yes
resflag3	Filter error/Incorrect entry	0 1	No Yes

Biomob_rtr

The data set *biomob_rtr* contains information on all anchor main residences from the age of 18 up to the first date of interview. Respondents were asked about former residences (before the date of the interview of wave 1) in chronological order, the date on which they began living there, and the date they moved away. This data was gathered during the wave 3 interview. The *biomob_rtr* data set is provided in “long” format, so that every episode of each person is represented by one row.

Table 4.33 displays all variables which are part of this data set. The variables *resbeg_rtr* and *resend_rtr* represent the beginning and end of an episode at a specific place of residence. By definition, the month of the anchor's 18th birthday is the beginning of the first episode ($resbeg_rtr = dob + 216$). The variable *resend_rtr* is coded “97 ongoing” for the last reported episode. If only ambiguous seasonal information was available, the beginning was coded to a random value within the bounds

mentioned above. If there was a complete missing (“-1 don’t know” or “-2 no answer”) for the month but the year was reported, the month values were coded randomly between 1 and 12. If such random information for the month was generated, the original ambiguous seasonal information was saved in the variable *resflag1_rtr*. If the information on the month was coded to “-1 don’t know” or “-2 no answer”, the original information was stored in *resflag2_rtr*.

The variable *rescensor_rtr* indicates if and in which way episodes at residence were censored. To determine if an anchor person has lived multiple times at the same place, an identification number (variable *resnumber_rtr*) for each municipality is generated. This identification number is also assigned if a respondent has given place names which cannot be allocated to a municipality. The variable *resindex_rtr* indicates the ascending order of episodes. *resland_rtr* gives information on the federal state or whether a respondent has lived abroad. If there have been residential changes, then the distance between the two residences, that is, between the municipalities, is found in *resdis_rtr*.

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

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

Variable	Variable label	Values	Value labels
id	Person number anchor	(see Table 2.2)	-
sample	Sample indicator	1	pairfam base sample
		2	DemoDiff sample
cohort	Birth cohort	1	1991-1993
		2	1981-1983
		3	1971-1973
dob	Date of birth	<i>date</i>	-
sex_gen	Generated sex anchor	-7	Incomplete data
		1	Male
		2	Female
wavex	Survey year wave x	-3	Does not apply
		1	2008/09
		2	2009/2010
		3	2010/2011
respwx	Respondent in wave x	-10	Not in demodiff
		0	No
		1	Yes
intdatwx	Date of interview	-3	Does not apply
	wave x	<i>date</i>	-
resplace_rtr	City mentioned	0	No place mentioned
		1	Place mentioned
		2	Demodiff
resbeg_rtr	Beginning of episode	-7	Incomplete data
		<i>date</i>	-
resend_rtr	End of episode	-7	Incomplete data
		97	ongoing
		<i>date</i>	-
rescensor_rtr	Censor	0	Uncensored
		1	Left censored, beginning of episode first interview

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
resland_rtr	Federal state	3	Right censored; ongoing episode
		5	Combination of 1 and 3
		-3	Does not apply
		-2	No answer
		-1	Don't know
		0	Berlin (West)
		1	Schleswig-Holstein
		2	Hamburg
		3	Niedersachsen (Lower Saxony)
		4	Bremen
		5	Nordrhein-Westfalen (North Rhine-Westfalia)
		6	Hessen (Hesse)
		7	Rheinland-Pfalz (Rhineland-Palatinate)
		8	Baden-Württemberg
		9	Bayern (Bavaria)
		10	Berlin (East)
		11	Brandenburg
		12	Mecklenburg-Vorpommern (Mecklenburg-Western Pomerania)
		13	Sachsen (Saxony)
		14	Sachsen-Anhalt (Saxony-Anhalt)
resindex_mr	Number of main residences	15	Thüringen (Thuringia)
		16	Saarland
		17	Another country
		-3	Does not apply
		1	1st main residence
resdis_rtr	Distance between consecutive residences
		12	12th main residence
		-99	Untraceable place
		-98	No calculation possible
		-7	Incomplete data
resbik_rtr	BIK classification	-3	Does not apply
		<i>number</i>	-
		-3	Does not apply
		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

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
		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
resnumber_rtr	Identification number city	-99	Place unknown
		-10	Not in demodiff
		<i>number</i>	-
resflag1_rtr	Inaccurate information on month	-3	Does not apply
		21	21 Original information on the month
		24	24 Original information on the month
		27	27 Original information on the month
		30	Original information on the month
resflag2_rtr	Unknown information on month	-3	Does not apply
		-2	-2 Original information on the month
		-1	Original information on the month

Biomob_rtr_parents

The data set *biomob_rtr_parents* contains retrospective information on the residency episodes within the parental household. Respondents were asked to provide the date of their first move out of the parental home and all following moves out of the parental home, as well as all following moves back into the parental home. This data set is also available in “long” format, with each episode represented by one row.

All variables contained in this data set are shown in Table 4.34. The variables *presbeg* and *presend* indicate the first and the last month of an episode at the parental household. The beginning of the first episode has been defined as the respondent’s date of birth. We assume that respondents had lived in their parents’ home since birth. If there was a data modification due to ambiguous seasonal information, the original information is stored in the variable *presflag1*. In the case of missing information on the month, the original information is stored in the variable *presflag2*. Analogous to this procedure, the original ambiguous seasonal information for the end of an episode is stored in the variable *presflag3* and the original missing variable is represented in *presflag4*. Additionally, *presindex* presents the number of different episodes at the parental household.

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

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

Variable	Variable label	Values	Value labels
id	Person number anchor	(see Table 2.2)	-
sample	Sample indicator	1	pairfam base sample
		2	DemoDiff sample
cohort	Birth cohort	1	1991-1993
		2	1981-1983
		3	1971-1973
dob	Date of birth	<i>date</i>	-
sex	Sex anchor	-4	Filter error / Incorrect entry
		1	Male
		2	Female
wavex	Survey year wave x	-10	Not in DemoDiff
		-3	Does not apply
		1	2008/2009
		2	2009/2010
		3	2010/2011
intdatwx	Date of interview wave x	-10	Does not apply
		-3	Does not apply
		<i>date</i>	-
presindex	Number of episode at parental household	1	First episode at parental household
	
		5	5th Episode at parental household
presbeg	Beginning of episode at parental household	-4	Filter error/incorrect entry
		-2	No answer
		-1	Don't know
		<i>date</i>	-
presend	End of episode at parental household	-4	Filter error/incorrect entry
		-2	No answer
		-1	Don't know
		97	Ongoing
		<i>date</i>	-
presflag1	Ambiguous information on month (Begin)	-3	Does not apply
		-2	Original information on the month
		-1	Original information on the month
		21	Original information on the month
		24	Original information on the month
		27	Original information on the month
		30	Original information on the month
		32	Original information on the month

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
presflag2	Ambiguous information on month (End)	-3	Does not apply
		-2	Original information on the month
		-1	Original information on the month
		21	Original information on the month
		24	Original information on the month
		27	Original information on the month
		30	Original information on the month
presflag3	Unknown information on month (Begin)	32	Original information on the month
		-3	Does not apply
		-2	Original information on the month
presflag4	Unknown information on month (End)	-1	Original information on the month
		-3	Does not apply
		-2	Original information on the month
		-1	Original information on the month

4.7.6 Anchor-parent panel data - bioparent

The file *bioparent* contains retrospective and prospective information on the anchor's biological, adoptive, or stepparents. The data set is provided in "long" format. This means that for each parent there is one row.⁴² Also, biological and adoptive parents who died before the first wave are included. *Bioparent* is updated every wave by renewing the information on existing parents and/or by adding information on new (step)parents. The available data are based on the survey waves 1 to 13.

In addition to the anchor's ID (*id*), the data contains the identification numbers of the (step) parents (*mid*, *fid*, *sfid* and *smid*). *parid* identifies those parents who took part in the parent survey (in any wave). This variable can be used for matching *bioparent* with specific waves of the parent survey. Additionally, the variable *surveypar* was included in order to identify parents who participated in the parents' survey in the respective wave. To differentiate between biological/adoptive and social parents, the variable *partype* has been generated to indicate the sex and the status of the parent, i.e. whether the parent is an adoptive, biological, or social mother or father. If a stepparent drops out of the survey due to separation or death and the biological father or mother begins a new partnership, then the new stepparent will be consecutively numbered (e.g. "1st stepfather", "2nd stepfather").

The variables *pardob* and *pardod* indicate the date of birth and, if relevant, the date of death of the respective parent. These variables include the number of months passed since January 1900. In cases when seasonal information of birth (or death) is available but month of birth (or death) is not, the monthly information was randomly imputed.

⁴²This is the major difference as compared to *bioparent* of Release 5.0 and below, which was organized in "long-long" format.

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

The variables *partnerparwx*, *samepartnerwx*, and *cohabpartnerwx* indicate whether the anchor’s adoptive or biological parent has a new partner (if this parent is separated from the other anchor’s biological or adoptive parent), whether this new partner is the same partner as in the previous wave, and whether the parent is living together with this new partner in the same household. Information on the beginning of the new partnership is given in the variable *begpartnerx*.

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

If only ambiguous seasonal information were available, the beginning was coded to a random value within the bounds mentioned above. If there was a complete missing (“-1 don’t know” or “-2 no answer”) for the month but the year was reported, the month values were coded randomly between 1 and 12. If random information for the month was generated, the original information was stored for the respective variable in *flag_pardob* for the date of birth and *flag_pardod* for the date of death.

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

Variable	Variable label	Values	Value labels
id	Person number anchor	(see Table 2.2)	-
sample	Sample indicator	1 2 3	pairfam base sample Demodiff sample Refreshment sample
parid	Person number parent (parent survey)	(see Table 2.2)	
intmodew12	Interview mode due to Covid-19 pandemic (wave 12)	-3 1 2	Does not apply CAPI & CASI (as in waves 1-11) CATI & PAPI
intmodew13	Interview mode due to Covid-19 pandemic (wave 13)	-3 1 2	Does not apply CAPI & CASI (as in waves 1-11) CATI & PAPI
mid	Person number mother	(see Table 2.2)	
fid	Person number father	(see Table 2.2)	
smid	Person number step- mother	(see Table 2.2)	
sfid	Person number stepfa- ther	(see Table 2.2)	
cohort	Birth cohort	1 2	1991-1993 1981-1983

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
		3	1971-1973
		4	2001-2003
dob	Date of birth anchor	<i>date</i>	-
sex	Sex anchor	-7	Incomplete data
		1	Male
		2	Female
surveyparwx	Participation in parents' survey wave x	-10	Not in DemoDiff
		0	No
		1	Yes
sf_index	No. of stepfather	-3	Does not apply
		1	1st step father
	
		6	6th step father
sm_index	No. of stepmother	-3	Does not apply
		1	1st step mother
	
		5	5th step mother
partype	Type of parent	1	Biological mother
		2	Biological father
		3	Adoptive mother
		4	Adoptive father
		5	First stepmother
		6	First stepfather
		7	Second stepmother
		8	Second stepfather
		9	Third stepmother
		10	Third stepfather
		11	Fourth stepmother
		12	Fourth stepfather
		13	Fifth stepmother
		14	Fifth stepfather
		15	Sixth stepfather
pardob	Parent: Date of birth (months since January 1900)	-7	Incomplete data
		<i>date</i>	
pardod	Parent: Date of death (months since January 1900)	-97	step parent
		-7	Incomplete data
		-3	Does not apply
		<i>date</i>	
wave	Survey year: wave x	-10	Not in demodiff
		-3	Does not apply
		1	2008/2009
		2	2009/2010
		3	2010/2011
		4	2011/2012
		5	2012/2013
		6	2013/2014
		7	2014/2015
		8	2015/2016
		9	2016/2017
		10	2017/2018

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
		11	2018/2019
		12	2019/2020
		13	2020/2021
respwx	Respondent in wave x	0	No
		1	Yes
intdatwx	Date of interview wave x	-10	Not in DemoDiff
		-3	Does not apply
		<i>date</i>	-
marriedparwx	Parent is married with other biological parent in wave x	-10	Not in demodiff
		-7	Incomplete data
		-3	Does not apply
		1	Yes
		2	No
cohabparwx	Parent is living together with other bio./adop. parent in one household in wave x	-10	Not in DemoDiff
		-7	Incomplete data
		-3	Does not apply
		1	Yes
		2	No
marstatparwx	Marital status of parent in wave x	-10	Not in DemoDiff
		-7	Incomplete data
		-3	Does not apply
		1	Single
		2	Married
		3	Civil union
		4	Divorced or dissolved civil union
		5	Widowed or surviving partner in a civil union
partnerparwx	Parent has a new partner in wave x	-10	Not in demodiff
		-7	Incomplete data
		-3	Does not apply
		1	Yes, with other bio. parent
		2	Yes
		3	No
cohabpartnerx	Parent is living together with the new partner in wave x	-10	Not in demodiff
		-7	Incomplete data
		-3	Does not apply
		1	Yes
		2	No
livanchorwx	Parent is living together with anchor in wave x	-10	Not in demodiff
		-7	Incomplete data
		-3	Does not apply
		0	Does not live at main residence
		1	Lives at main residence
contactwx	Contact of parent to anchor wave x	-10	Not in DemoDiff
		-7	Incomplete data
		-3	Does not apply
		1	Daily

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
		2	Several times per week
		3	Once per week
		4	1-3 times per month
		5	Several times per year
		6	Less often
		7	Never
		10	Never had contact
samepartnerwx	New partner of the	-10	Not in demodiff
	parent is the same	-7	Incomplete data
	partner as previous wave	-3	Does not apply
	to wave x	1	Yes
		2	No
begnocontact	Beginning of no	-7	Incomplete data
	contact of parents to	-3	Does not apply
	anchor	<i>number</i>	
	(months since January		
	1900)		
begpartnerwx	Duration of new	-10	Not in demodiff
	partnership wave x	-7	Incomplete data
	(categorized in years)	-3	Does not apply
		1	Less than five years ago
		2	5-10 years ago
		3	10-15 years ago
		4	More than 15 years ago
pardob_flag	Original information on	-7	Incomplete data
	month of birth	-3	Does not apply
		21	21 Original information
		24	24 Original information
		27	27 Original information
		30	30 Original information
		32	32 Original information
pardod_flag	Original information on	-7	Incomplete data
	month of death	-3	Does not apply
		21	21 Original information
		24	24 Original information
		27	27 Original information
		30	30 Original information
		32	32 Original information

4.7.7 Multi-actor panel data - Overview_multi_actor

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

Furthermore, the partner's identification number (*pid*) is included in this data set. Based on this

person identification number and the variable *respartner*, users can easily find out how many same partners were surveyed during the panel.

The variable *childX* was generated in order to indicate if child x exists and was selected as a CAPI-child. The variable *reschildX* is a dummy variable that documents if a child was interviewed as part of the CAPI child survey. As already described in section “Anchor-child panel/episode data - biochild”, the variable *parentidkX* contains the identification number of childX’s second biological parent. By using this identification number, stepfamilies are easily identified. If *pid* and *parentidkX* differ, the current partner is not the second biological parent. If children have different values in *parentidkX*, they likely have two different biological parents. Please note that *parentidkX* as well as *pid* are only available for partners listed as current or previous partners in the course of the EHC and not for retrospective partners.

The variables *resparchildX_a* and *resparchildX_p* reflect whether the anchor and/or their cohabiting partner have filled out the parenting questionnaire. There are a few anchors and partners who fill out the questionnaire by mistake, and these cases were set to the missing value “-4”. The selection of the children relevant for the parenting survey has been slightly modified between waves.

The variables *resfather*, *resmother*, *resstepfather*, and *resstepmother* provide information on participation in the parent survey (waves 2–7), respectively in the grandparent survey (wave 8).

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

Variable	Variable label	Values	Value labels
id	Person number anchor	(see Table 2.2)	-
wave	Survey year	1	Wave 1: 2008/09
		2	Wave 2: 2009/10
		3	Wave 3: 2010/11
		4	Wave 4: 2011/12
		5	Wave 5: 2012/13
		6	Wave 6: 2013/14
		7	Wave 7: 2014/15
		8	Wave 8: 2015/16
		9	Wave 9: 2016/17
		10	Wave 10: 2017/18
		11	Wave 11: 2018/19
		12	Wave 12: 2019/20
		13	Wave 13: 2020/21
		13	Wave 14: 2021/22
cohort	Birth cohort	1	1991-1993
		2	1981-1983
		3	1971-1973
		4	2001-2003
demodiff	DemoDiff sample	0	pairfam
		1	DemoDiff
sample	Sample indicator	1	pairfam base sample
		2	DemoDiff sample
		3	Refreshment sample
sex_gen	Generated sex anchor	see sex_gen	-
pid	Person number partner	(see Table 2.2)	-

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
respartner	Participation in partner survey	-3 0 1	Does not apply No Yes
childx	Child x existing	-3 1 2	Does not apply Child exists, but is no capi-child Child exists, selected as capi-child
cidx	Child x: Person number	(see Table 2.2)	-
reschildx	Child x: Participation in children's survey	-10 -3 0 1	Not in DemoDiff Does not apply No Yes
parentidx	Child x: Person number second biological parent	see variable pid	-
resparchildx_a	Child x: Participation parenting survey - anchor	-10 -4 -3 0 1	Not in DemoDiff PAPI-Qu. was filled out by mistake Does not apply No Yes
resparchildx_p	Child x: Participation parenting survey - partner	-10 -4 -3 0 1	Not in DemoDiff PAPI-Qu. was filled out by mistake Does not apply No Yes
kxdoby_gen	Child x: Generated year of birth	see dob*_gen	-
kxdobm_gen	Child x: Generated month of birth	see dob*_gen	-
resfather	Participation in parent survey: Father	-10 -3 0 1	Not in DemoDiff Does not apply No Yes
resmother	Participation in parent survey: Mother	-10 -3 0 1	Not in DemoDiff Does not apply No Yes
resstepfather	Participation in parent survey: Stepfather	-10 -3 0 1	Not in DemoDiff Does not apply No Yes

continued on next page

continued from previous page

Variable	Variable label	Values	Value labels
resstepmother	Participation in parent survey: Stepmother	-10	Not in DemoDiff
		-3	Does not apply
		0	No
		1	Yes

5. Partner data

Kantar Public compiled a raw data set from the returned partner PAPI questionnaires. Data preparation followed closely the procedures for editing the anchor data.

5.1 Data editing

This section describes the editing of wave 1-14 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.

5.1.1 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.

5.1.2 Missing values

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

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

Value	Label
-1	Don't know
-2	No answer (also: I don't want to answer that, answer refused)
-3	Does not apply
-4	Filter error / Incorrect entry / Unreadable answer
-9	Invalid multiple answer

5.1.3 Open answers

Only two string variables had to be recoded in the partner data in wave 1; one had to be recoded in waves 2, 3, 5, 7, 9, 11 & 13 and none in waves 4, 6, 8, 10, 12 and 14. For variable *psd27o* (included in wave 1, 3, 5, 7, 9, 11, & 13 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.

5.1.4 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”).

5.1.5 Value and filter checks

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

5.1.6 Consistency checks

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

5.1.7 English-language data

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

5.2 Generated variables and scales

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

The exceptions for wave 2 are:

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

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

The Stata do-file *genvars_partner2.do* contains the syntax used to compute the variables for wave 2.

Note that to produce some of the generated variables included in the anchor data set, we used information given by the partner (see Chapter 4.3).

We refrained from constructing weights for this sample.

The generated variables described above are included in second-wave partner data. In contrast, the scale variables (see Table 5.2) are not part of the delivered data set. They can be generated by running the syntax file *scales_partner\$* which is available for each wave as part of the Scientific Use File. For further information, refer to the scales manual (Thönnissen et al. 2023) which contains a detailed description of all scales.

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

Construct	Variable name	Wave
Traditional concept of marriage	ptradmarr	1,3,5,7,9,11,13
Value of Partnership: Negative expectations	pvpneg/pvpneg2	1,3,5,7,9,11
Value of Children: Benefit of stimulation	pvocbstim	1,2,4,6,8,10
Value of Children: Costs of comfort	pvocccomf	1,2,4,6,8,10
Partnership: Conflict	pconfl_apd	all
Partnership: Intimacy	pintim_aps	all
Partnership: Esteem	padmir_apo	all
Partnership: Dominance	pdomin_apo	all
Partnership: Ambivalence	pambiva_apd	7,9 - 14
Partnership: Emotional ambivalence	pambiv_apd	1,3,5,7,9,11,13
Partnership: Anxiety about loss of love	plovewitanx_apd	1,3,5,7,9,11,13
Partnership: Anxiety about being absorbed	penganx_apd	1,3,5,7,9,11,13
Partnership: Independence	pindep_apd	1-3,5,7,9,11,13
Partnership: Own partnership satisfaction (global scale)	psatpart	1,3,5,7,9,11,13
Partnership: Feelings of competence in the partnership	pcomppart2	2,5,7,9,11,13
Partnership: Future orientation	pcomfut_apd	1 - 3,5,7,9,11,13
Partnership: Tolerance of conflicts	pcomctol_apd	1 - 3,5,7,9,11,13
Partnership: Orientation of reciprocity Self	preciproc_aps	1,3,5,7,9,11,13
Partnership: Hostile attributions towards anchor (self assessment)	phostattr_aps	3,5,7,9,11,13
Partnership: Frequency of manifest conflicts	pconfl_aps	1,5,7,9,11,13
	pconfl_aps2	3
Partnership: Verbal aggression Partner	pverbaggr_apo	1 - 13
Partnership: Verbal aggression Self	pverbaggr_aps	all
Partnership: Constructive behavior Partner	pconstrbh_apo	1 - 13
Partnership: Constructive behavior Self	pconstrbh_aps	all
Partnership: Withdrawal Partner	pwithdraw_apo	1 - 13
Partnership: Withdrawal Self	pwithdraw_aps	all
Partnership: Manipulation Partner	pmanipul_apo	1,3,5,7,9,11,13
Partnership: Manipulation Self	pmanipul_aps	1,3,5,7,9,11,13
Partnership: Dyadic coping Partner	pdycop_apo	1,3,5,7,9,11,13
Partnership: Dyadic coping Self	pdycop_aps	1,3,5,7,9,11,13
Partnership: Instability of partnership	pinstab_apd	1-13
Explosiveness and tendency to anger	pexplosive	1,3,5; 4 and from 6 all (only new partners)
Shyness	pshyness	1,3,5; 4 and from 6 all (only new partners)
Emotional autonomy	pemotautn	1,3,5; 4 and from 6 all (only new partners)
Self-esteem	pselfesteem	1-6,8,10,12,13,14 all; 7,9,11 (only new partners)
Depressiveness	pdepressive/pdepressive2	all
Narcissism: Rivalry	priv	13

continued on next page

continued from previous page

Construct	Variable name	Wave
Narcissism: Admiration	padm	13
BIG 5: Neuroticism	pneurot	2,12 all; 3-14 (only new partners)
BIG 5: Extraversion	pextrav	2,12 all; 3-14 (only new partners)
BIG 5: Agreeableness	pagreeable	2,12 all; 3-14 (only new partners)
BIG 5: Conscientiousness	pconscient	2,12 all; 3-14 (only new partners)
BIG 5: Openness	popenness	2,12 all; 3-14 (only new partners)
Co-parenting with the other parent	pcoparent	2,4,6,8,10,12
Parental Self Efficacy/Competence	pcomperz	2,4,6,8; from 10 all
Parenting: Sacrifice in raising children	psacrif_pacs	3,5,7,9,11-13
Parenting: Overprotectiveness	poverprotect	11,13
Parenting: Recognition / support in education through partner	ppartnersup	3,5,7,9,11-13
Parenting: Unspecific strain	punspstrain	11
Parenting: Autonomy in the parenting role	pautonom	4,6,8; from 10 all
Parenting: Autonomy in the parenting role (short scale)	pautonom2	4,6,8; from 10 all
Parenting: Pleasure in the parenting role	ppleasure	4,6,8; from 10 to 13
Parenting goals: Status	ppgoalstatus	2; 4-10,12 (only new partners)
Parenting goals: Competence	ppgoalcomp	2; 4-10,12 (only new partners)
Parenting goals: Autonomy	ppgoalautn	2; 4-10,12 (only new partners)
Intergenerational relationships: Own intimacy within partner-mother dyad (self assessment)	pintim_ams	3-13
Intergenerational relationships: Own intimacy within partner-father dyad (self assessment)	pintim_afs	3-13
Intergenerational relationships: Conflict within partner-mother dyad	pconfl_amd	3-13
Intergenerational relationships: Conflict within partner-father dyad	pconfl_afd	3-13
Intergenerational relationships: Ambivalence with mother	pambiv_amd	7-11
Intergenerational relationships: Ambivalence with father	pambiv_afd	7-11
Work-Family-Conflict: Work impacts family	pwif_conflict	6,8,10,12
Work-Family-Conflict: Work impacts family - time	pwif_time	6,8,10,12
Work-Family-Conflict: Work impacts family - stress/strain	pwif_strain	6,8,10,12
Work-Family-Conflict: Family impacts work	pfiw_conflict	6,8,10,12
Work-Family-Conflict: Family impacts work - time	pfiw_time	6,8,10,12
Work-Family-Conflict: Family impacts work - stress/strain	pfiw_strain	6,8,10,12
Problematic use of cellphones	pprobcell	13

5.3 Changes in answer categories

In wave 11, the variables "phlt8" and "phlt9" from the section "Addictive behaviors" were permanently moved to the section for new partners. As the remaining variables from this block ("phlt10-14"), are still presented to all partners in every odd-numbered wave, a change in answer categories

for "phlt10" - "Do you now smoke cigarettes, a pipe, cigars, or cigarillos?" was necessary. Instead of the two answer categories "1 Yes" and "2 No", we now assess "1 Yes", "2 No, currently not, but in the past", and "3 No, I have never smoked regularly".

6. Parent data

The parent questionnaire was part of the pairfam study from wave 2 to wave 8. From wave 2 to wave 7, in the course of the anchor CAPI, residential addresses of the relevant parents were collected for sending them the questionnaires by mail. However, due to response rates below 30 percent in wave 7 (see Brüderl et al. (2023)), the instrument was tremendously shortened in wave 8 with a strong focus on the relationship of the parent with their grandchildren. Furthermore, anchor respondents had the chance to give the questionnaire personally to their parents additionally to the postal transmission of the questionnaire. The target population changed from “all (living) parents who are in contact with the anchor” in waves 2 to 7 to “parents who are in contact with the anchor and the anchor is living with at least one biological or adopted child in a shared household”. Consequently, case numbers dropped between wave 7 (N=2,719) and wave 8 (N=627) as only grandparents were eligible. However, as these changes did not lead to a significant increase of the response rates, the parent survey was conducted in wave 8 for the last time.

Kantar Public compiled a raw data set from the returned parent’s PAPI questionnaires. Data editing followed closely the procedures used to clean the anchor data.

6.1 Data editing

This section describes the editing of waves 2 to 8 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.

6.1.1 Deletion

The raw data provided by *Kantar Public* contained 5,039 cases in wave 2. A number of cases were deleted from this original data file either because they were completely empty⁴³ or because they were suspected to have been derived from duplicate questionnaires or to have been provided by the anchor.⁴⁴ Thus, the parent file was reduced by 24 cases, which results in a final data file of N=5,015. There were no cases deleted in wave 3. The final data set contains N=3,946 in wave three. In wave 4, two cases were deleted⁴⁵ and the data sets consists of 3,350 entries. In wave 5 no cases were deleted resulting in a final data file including N=3,546 cases. Again, there were no cases deleted in wave 6 (final N=3,043), wave 7 (N=2,719) and in wave 8 (N=627).

6.1.2 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

⁴³N=17; parid = 267932302, 32927301, 144352301, 144352304, 15196302, 233346301, 233346302, 280325304, 337872301, 337872304, 457711301, 457711302, 459740301, 459740302, 632879301, 632879302, 666208301

⁴⁴N=7; parid = 33691301, 33691304, 77993302, 257271301, 257271302, 567208301, 567208302

⁴⁵N=2; parid = 750419301, 750419302

(plus the prefix par-). Variables unique to the parent survey had unique names according to our system of variable names.

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

6.1.3 Missing values

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

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

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

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

6.1.4 Filter checks

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

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

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

the respondents is accepted).

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

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

In wave 5: Three conditions had also to be met in this wave: First, the filter question 11 had to be correct. Second, the sex of the grandchild indicated in question 13 had to match the sex of the CAPI-child (provided in the anchor's interview). Third, the age of the grandchild indicated in question 14 had to be within the age range of 5 to 18 (which is purposely somewhat broader as some miscalculation of the respondents is accepted).

In wave 6: Again, the three conditions had to be met in this wave: First, the filter question 11 had to be correct. Second, the sex of the grandchild indicated in question 14 had to match the sex of CAPI-child (provided in the anchor's interview). Third, the age of the grandchild indicated in question 15 had to be within the age-range of 5 to 18 (which is purposely somewhat broader, as some miscalculation of the respondents is accepted).

Also in wave 7 three conditions had to be met in this wave: First, the filter question 11 had to be correct. Second, the sex of the grandchild indicated in question 12 had to match the sex of CAPI-child (provided in the anchor's interview). Third, the age of the grandchild indicated in question 13 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 8, the procedure could be simplified due to the focus of the interview on the CAPI-child. As first question, respondents were asked for the name of their child (anchor) and their grandchild (CAPI-child) mentioned at the frontpage of the PAPI-interview. *Kantar Public* delivered an indicator showing whether or not these names match. If they did not match, they compared the mentioned name to the names of the other children in the household.

6.1.5 Consistency checks

Various checks to identify logically impossible or empirically implausible answers were conducted. The code "-5" was assigned if it was possible to identify which variable under consideration was wrong. The values were left unedited if the inconsistency could not be solved, but were flagged with a generated variable referring to the respective inconsistency. It is recommended to analyze flagged variables with caution. Table A.28 describes the flag variables in detail.

6.1.6 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 *Kantar Public*.

6.1.7 English-language data

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

6.2 Generated variables and scales

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

6.2.1 Generated variables

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

In addition, similar to the anchor data, user-friendly parents' data were 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 data (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. However, unlike other indicators, these scale indicators are not included in the data files. For further information, refer to the scales manual (Thönnissen et al. 2023) which contains a detailed description of all scales.

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. The anchor's siblings' numbers (*sibidx*) are included as well in all parent data sets (see also Chapter 2.2).

In wave 7, two sub-scales of the SF12 health indicator are delivered within the parent data set: physical health composite scale (*parpcs*) and mental health composite scale (*parmcs*). For the construction, the same procedure was followed as in the anchor data.

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

Construct	Variable name	Wave
<i>Identification</i>		
Anchor's identification number	<i>id</i>	2 - 8
Respondent's identification number	<i>parid</i>	2 - 8
CAPI-child's identification number	<i>cid</i>	2 - 8
Anchor's siblings' number	<i>sibidx</i>	2 - 8
<i>Paradata</i>		
Respondent's type (based on questionnaire)	<i>partype</i>	2 - 8
Sex of respondent (based on questionnaire)	<i>parseparent</i>	2 - 8
Position of respondent (based on questionnaire)	<i>parposition</i>	2 - 8
Sex of the anchor (based on questionnaire)	<i>parseanchor</i>	2
Wave (based on questionnaire)	<i>wave</i>	2 - 8

continued on next page

continued from previous page

Construct	Variable name	Wave
Language version (based on questionnaire)	parlng	2 - 4
Date of the interview	parintd, parintm, parinty	2 - 8
Cohort anchor	parcohort	2 - 8
Year of birth anchor based on anchor interview	pargeanchor	2 - 8
# 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 - 7
# of grandchildren	pargrandkids	2 - 8
<i>Generated variables</i>		
Age variable: respondent	parage parage2	2 - 8
Other age variables (respondent's parents)	parmage parmage2 parpage parpage2	2 - 7
	parpage parpage2	2
Current primary and secondary activity status	parcasprim, parcassec	2 - 8
Highest school degree	parschool	2 - 8
Highest vocational degree	parvocat	2 - 8
CASMIN classification of educational attainment	parcasmin	2 - 8
ISCED classification of educational attainment	pariscd	2 - 8
Years of schooling / vocational qualification	paryeduc	2 - 7
Labor force status	parlfs	2 - 8
Household size main residence	parhhsizemrd	2,4,6
Net equivalence income	parhhincgcee	2
Physical / mental health composite scale (SF12)	parpcs, parmcs	7
Type of children	park1type, park2type, park3type, park4type	2 - 7
Marital status	parmarstat	2,4,6
Relationship status	parrelstat	2,4,6,8
Do parents live in a shared household?	parigr27	2 - 7
Are parents married?	parigr28	2 - 7
State / Bundesland	parbula	3 - 8
Size of community in 7 categories	pargkpol	3 - 8
Settlement structure	parbik	3 - 8
New parent in wave 3	parw3ne	3
New parent in wave 4	parw4ne	4
New parent in wave 5	parw5ne	5
New parent in wave 6	parw6ne	6
New parent in wave 7	parw7ne	7
New parent in wave 8	parw8ne	8
Information about the correct grandchild	parcorgc	3 - 8

6.2.2 Scales

The following scale variables (see Table 6.3) can be generated by running the do-file scales_parents\$, which are part of the Scientific Use File. For further information, refer to the scales manual (Thönissen et al. 2023) which contains a detailed description of all scales.

Table 6.3: List of scales included in syntax file *scales_parent\$*

Construct	Variable name	Wave
Filial Obligation (mean indicator: parval2i1, parval2i5)	parobligationf	2
Grandparental Obligation (mean indicator: parval2i2, parval2i4)	parobligationp	2
Parental Obligation (mean indicator: parval2i3, parval2i6)	parobligationgp	2
Autonomy in Grandparental Role 4-Items (mean indicator: pargrcn11i7, pargrcn11i8(r), pargrcn11i9(r), pargrcn11i10)	pargcautonom	8
Autonomy in Grandparental Role 2-Items (mean indicator: pargrcn11i7, pargrcn11i10)	pargcautonom2	8
Pleasure in Grandparental Role 2-Items (mean indicator: pargrcn11i8, pargrcn11i9)	pargcpleasure	8
Traditional concept of marriage (mean indicator: parval1i2, parval1i7, and parval1i8)	partradmarr	3,5,7
Satisfaction (mean indicator: parsat1i1, parsat1i2, parsat1i3, parsat1i4)	parsat	5,7
Readiness to make sacrifices (mean indicator: parcrn32i1, parcrn32i2, and parcrn32i3)	parsacrif_pacs	3,5,7
Frequency of Joint Activities (mean indicator: parigr73a, parigr74a, parigr75a, parigr76a, and parigr77a)	paractiv_paras	2,4,6
NRI Partner Conflict	parconfl_parparpd	2,3,5,7
NRI Partner Intimacy	parintim_parparps	2,3,5,7
NRI Partner Approval	paradmir_parparpo	2,3,5,7
NRI Partner Dominance	pardomin_parparpo	2,3,5,7
NRI Anchor Conflict	parconfl_parad	2 - 7
NRI Anchor Intimacy	parintim_paras	2 - 7
NRI Anchor Approval	paradmir_parao	2,4,6
NRI Anchor Dominance	pardomin_parao	2,4,6
NRI Anchor Ambivalence	parambiv_parad	7,8
Co-parenting with respect to anchor (younger than 21) (mean indicator: parcrn21i1, parcrn21i2, parcrn21i3)	parcoparent	2,4
Educational style with respect to anchor (younger than 21): Monitoring (mean indicator: parcr1i2, parcr1i9, parcr1i6, parcr1i12)	parmonitor_paras	2,4
Educational style with respect to anchor (younger than 21): Strict control (mean indicator: parcr1i18, parcr1i19, parcr1i20, parcr1i21)	parstrict_paras	2,4
Educational style with respect to anchor (younger than 21): Emotional Warmth (mean indicator: parcr1i1, parcr1i5, parcr1i14)	parwarmth_paras	2,4
Educational style with respect to anchor (younger than 21): Negative Communication (mean indicator: parcr1i3, parcr1i8, parcr1i11)	parnegcomm_paras	2,4

7. Child data

Editing of the child data from waves 2-14 was conducted in line with the procedures to clean the anchor data and the additional alteri data. The complete data processing was done in SPSS.

7.1 Data editing

This section describes the editing of wave 2-14 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.

7.1.1 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.

7.1.2 Missing values

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

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

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

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

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

7.1.3 Consistency checks

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

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

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

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

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

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

7.1.4 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.

7.1.5 Recoding of SDQ Items

In order to match the item numbers of the child SDQ and the parenting SDQ, we recoded the items for the child SDQ starting with wave 6, retroactively until wave 2, where the SDQ was introduced. The changes were also implemented in the syntax file *scales* for every *scale_child\$* and the scales manual (Thönnissen et al. 2023).

7.1.6 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" or the according classification system throughout. The variables affected by this procedure are shown in Table 7.4 along with the value labels (<information> mentioned) assigned to valid answers on the anonymous numeric variables.

Table 7.4: List of anonymized variables

Variable	Variable label	Anonymized value label	Wave
cdobd	Child day of birth	Day mentioned	2 - 14
cedu1o	Other schooltype	Other schooltype mentioned	2
cedu1ao	Other schooltype	Other schooltype mentioned	3 - 14
cpcr4	Name anchor	Name mentioned	2 - 13
cpcr4a	Name biological mother in same household	Name mentioned	14
cpcr4b	Name stepmother in same household	Name mentioned	14
cpcr4c	Name foster mother in same household	Name mentioned	14
cpcr4d	Name adoptive mother in same household	Name mentioned	14
cpcr5	Name current partner of anchor in same household	Name mentioned	2 - 13
cpcr5a	Name biological father in same household	Name mentioned	14
cpcr5b	Name stepfather in same household	Name mentioned	14
cpcr5c	Name foster father in same household	Name mentioned	14
cpcr5d	Name adoptive father in same household	Name mentioned	14
cgp1	Name of anchor's mother	Name mentioned	2,4,6,8
cgp2	Name of anchor's father	Name mentioned	2,4,6,8
cpcr13	Name biological parent outside the household	Name mentioned	3,5,7,9
cpcr13a	Name biological father outside the household	Name mentioned	14
cpcr13b	Name biological mother outside the household	Name mentioned	14
crom7	Name of romantic partner	Name mentioned	5 - 14
ccp2	Step-up's complete name and address	Address received/ Same address as anchor	4 - 6
ccp3	Step-up's complete address	Address received/ Same address as anchor	7 - 13
cbf2	Name of best friend	Name mentioned	7,9,11,13,14
cjob	Description of job	ISCO08	9,10,11,12,14
ccor5	Positive aspects of lockdown	Change in assigned aspect	12

7.1.7 Coding of cjob

In waves 9-12 and 14 variable *cjob* contains the child's desired occupation classified according to the 4-digit International Standard Classification of Occupations (ISCO-08) schema. Recoding of the open answers was done on the basis of the rules set by *Kantar Public* (see Hartmann et al. (2012) for details).

7.1.8 Coding of ccor5

In wave 12 variable *ccor5* contains the child's subjective feeling of positive aspects of the first German COVID-19 lockdown. Open answers were coded by the pairfam team into self-developed categories: no change (0), more time for hobbies (1), more time for friends (2), more time for family (3), more

time for self-reflection (4), more time to relax (5), less school (6), less pollution (7), better learning environment (8).

7.1.9 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.

7.1.10 English-language data

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

7.2 Generated variables and scales

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

The Stata do-files `genvars_child.do` contain the syntax used to compute some of the variables for each wave respectively. Due to data privacy, syntaxes are not available for all of the generated variables.

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

Construct	Variable name	Wave
relstatac	Relationship/kinship between anchor and child	2 - 13
cinty	Child interview year	2 - 14
cintm	Child interview month	2 - 14
cintd	Child interview day	2 - 14
cagey	Child's age in years	2 - 14
cagem	Child's age in months	2 - 14
ctitlea	Child's naming for anchor	2 - 13
ctitlep	Child's naming for anchor's partner	2 - 13
ctitleop	Child's naming for the other biological parent outside the household	3,5,7,9,11,13
ctitlem	Child's naming for mother in household	14
ctitlef	Child's naming for father in household	14
ctitleopm	Child's naming for mother outside the household	14
ctitleopf	Child's naming for father outside the household	14
cintmode	Child's interview mode	12 & 14

The generated variables are included in wave 2-14 child data. In contrast to these, the scale variables (see Table 7.6) are not part of the delivered data set. They can be generated by running the syntax file `scales_child$` which can be found as part of the Scientific Use File. For further information, refer to the scales manual (Thönnissen et al. 2023) which contains a detailed description of all scales.

Table 7.6: List of scales included in syntax file scales_ *child*\$

Construct	Variable name	Wave
Class atmosphere	cclassatmo	2 - 14
Social integration	csocialinteg	2 - 14
Peer rejection	cpeerref	2 - 14
Parental school engagement	cengagement	2,4 & 6
Economic deprivation	cecodep	2 - 14
Intimacy child-best friend-relationship	cintim_cbf	3,5,7,9,11,13,14
Companionship child-best friend-relationship	ccomp_cbf	3,5,7,9,11,13,14
Emotional autonomy child-best friend-relationship	cauto_cbf	7,9,11,13,14
Emotional warmth child-anchor-relationship	cwarmth_cao	2 - 13
Emotional warmth child-partner-relationship	cwarmth_cpo	2 - 13
Emotional warmth child-mother-relationship	cwarmth_cmo	14
Emotional warmth child-father-relationship	cwarmth_cfo	14
Emotional warmth child-other parent-relationship	cwarmth_copo	3,5,7,9,11,13
Strict control child-anchor-relationship	cstrict_cao	2,4,6,8
Strict control child-partner-relationship	cstrict_cpo	2,4,6,8
Strict control child-other parent-relationship	cstrict_copo	3,5
Monitoring child-anchor-relationship	cmonitor_cao	2 - 13
Monitoring child-partner-relationship	cmonitor_cpo	2 - 13
Monitoring child-mother-relationship	cmonitor_cmo	14
Monitoring child-father-relationship	cmonitor_cfo	14
Negative communication in the anchor-child relationship	cnegcomm_cao	4 - 13
Negative communication in the partner-of-the-anchor-child relationship	cnegcomm_cpo	4 - 13
Negative communication in the anchor-child relationship	cnegcomm_cao	4 - 13
Negative communication in the mother-child relationship	cnegcomm_cmo	14
Negative communication in the father-child relationship	cnegcomm_cfo	14
Inconsistent discipline in the anchor-child relationship	cincondis_cao	9 - 13
Inconsistent discipline in the partner-of-the-anchor-child relationship	cincondis_cpo	9 - 13
Inconsistent discipline in the mother-child relationship	cincondis_cmo	14
Inconsistent discipline in the father-child relationship	cincondis_cfo	14
Inconsistent discipline in the other parent-child relationship	cincondis_copo	9,11,13
Rules in the anchor-child relationship	crules_cao	9 - 13
Rules in the partner-of-the-anchor-child relationship	crules_cpo	9 - 13
Rules in the mother-child relationship	crules_cmo	14
Rules in the father-child relationship	crules_cfo	14
Fear of love withdrawal child-anchor-relationship	lovewitanx_cas	3 - 13

continued on next page

continued from previous page

Construct	Variable name	Wave
Fear of love withdrawal child-partner-relationship	lovewitanx_cps	3 - 13
Fear of love withdrawal child-mother-relationship	lovewitanx_cms	14
Fear of love withdrawal child-father-relationship	lovewitanx_cfs	14
Fear of love withdrawal child-other parent-relationship	lovewitanx_cops	3,5,7,9,11,13
Intimacy child-anchor-relationship	cintim_cas	2 - 13
Intimacy child-partner-relationship	cintim_cps	2 - 13
Intimacy child-mother-relationship	cintim_cms	14
Intimacy child-father-relationship	cintim_cfs	14
Intimacy child-other partner-relationship	cintim_cops	3,5,7,9,11,13
Conflict child-anchor-relationship	cconfl_cad	2 - 13
Conflict child-partner-relationship	cconfl_cpd	2 - 13
Conflict child-mother-relationship	cconfl_cmd	14
Conflict child-father-relationship	cconfl_cfd	14
Conflict child-other partner-relationship	cconfl_copd	3,5,7,9,11,13
Admiration child-anchor-relationship	cadmir_cao	2 - 13
Admiration child-partner-relationship	cadmir_cpo	2 - 13
Admiration child-mother-relationship	cadmir_cmo	14
Admiration child-father-relationship	cadmir_cfo	14
Admiration child-other parent-relationship	cadmir_copo	3,5,7,9,11,13
Parental reliability child-anchor-relationship	creliabl_cad	2
Parental reliability child-partner-relationship	creliabl_cpd	2
SDQ: Conduct problems	cconduct	2 - 14
SDQ: Emotional symptoms	cemotion	2 - 14
SDQ: Prosocial behavior	cprosoc	2 - 14
SDQ: Hyperactivity	chyper	5 - 14
SDQ: Peer problems	cpeerprob	8 - 14
Intimacy own romantic partner	cintim_crps	5 - 14
Conflict own romantic partner	cconfl_crpd	5 - 14
Admiration own romantic partner	cadmir_crpo	5 - 14
Fear of love withdrawal own romantic partner	clovewitanx_crpd	5 - 14
Ambivalence in relationship to own romantic partner	cambiv_crpd	5 - 14
Child's perception of interparental conflict: frequency	ccpicf_capd	6 - 14
Child's perception of interparental conflict: resolution	ccpicr_capd	6 - 14
Coalition pressure	ccoal	7,9,11,13
Self-esteem	cselfesteeem	8 - 14
Sibling affection	csibaff	8,10,12
Sibling hostility	csibhos	8,10,12
Sibling rivalry	csibriv	8,10,12
Sibling conflict resolution	csibcon	8,10,12
BIG 5: Neuroticism	cneurot	10,12,14
BIG 5: Extraversion	cextrav	10,12,14
BIG 5: Agreeableness	cagreeable	10,12,14
BIG 5: Conscientiousness	cconscient	10,12,14
BIG 5: Openness	copenness	10,12
Narcissism: Rivalry	cnarriv	13
Narcissism: Admiration	cnarad	13

7.3 Sample selection for the child interview

In wave 2 the CAPI-child was first assessed as the youngest child living in the household that was between 8 and 15 years of age. If this condition held true for more than one child, the child was chosen from the group of eligible children who were mentioned by the respondent first.

Starting with wave 3, selection of the child of the respondent as a CAPI-child was made on the following criteria:

If there were children in the previous wave, who were eligible for the child interview, the attempt was made again to interview them. Additionally, children who were younger than the first (and thus oldest) child previously selected for the CAPI survey and who were now 8 years or older were also asked to be interviewed.

If no children were eligible to be a CAPI-child in the previous wave because there were no children between the ages of 8 and 15 in the household but in the current wave there now was such an eligible child, the youngest of these was interviewed as a CAPI-child. Again, if this condition held true for more than one child, the child was chosen from the group of eligible children who were mentioned by the respondent first. The selection of the CAPI-child for child interview is documented in the anchor codebook of each wave.

7.4 Mode change (CAPI vs. CATI) in waves 12, 13, & 14

In wave 12, 144 out of 1146 children were interviewed with the CATI instead of the CAPI due to the COVID-19 pandemic. The standard CAPI had been slightly modified for the telephone interview and the CASI section was completely removed. Furthermore, only the SDQ subscales *Emotional symptoms* and *Hyperactivity* were part of the CATI. Instead, a section on corona was implemented which assessed the children's worries and concerns in relation to the COVID-19 pandemic. The generated variable *cintmode* indicates the mode of the interview, starting with wave 12. For further information on the specific items of the corona module, refer to the scales manual (Thönnissen et al. 2023).

In wave 13, 599 out of 1027 children were interviewed with the CATI instead of the CAPI due to the COVID-19 pandemic. The standard CAPI had again been slightly modified for the telephone interview and the CASI section was completely removed. Furthermore, only the SDQ subscales *Emotional symptoms* and *Hyperactivity* were part of the CATI again. The section on corona was implemented in an abbreviated form assessing children's worries and concerns in relation to the COVID-19 pandemic.

In wave 14, 507 out of 632 children were interviewed via CATI instead of the CAPI. Questions that were posed in the CASI section of the CAPI interview were also included in the CATI interview of this wave. All SDQ subscales were further part of both the CAPI and the CATI. Additionally, preloads from the previous waves were no longer available in wave 14. Therefore, questions related to the children's parents were no longer split by anchor/partner of the anchor. Instead, these questions were posed regarding biological mother/stepmother/adoptive mother/foster mother and biological father/stepfather/adoptive father/foster father living in the same household. Questions regarding external parents were posed regarding biological mother/stepmother/adoptive mother/foster mother and biological father/stepfather/adoptive father/foster father not living in the same household as the child.

For further information on the mode changes from CAPI to CATI in wave 12 to wave 14, refer to the methods report of the respective wave and to the Technical Paper No. 01 (Brüderl et al. 2023).

8. Parenting data

The pairfam study collected three different types of parenting data, all of those have been PAPI questionnaires. First, the *parenting survey* has been part of the pairfam study since wave 2. It contains information from anchors and their partners pertaining to their relationships and interactions with their children up to the age of 15. Starting with wave 9, pairfam introduced a second parenting survey called “*Parenting Adolescents and Young Adults*” (PAYA), which surveyed anchors and their partners regarding their relationship with their adolescent and young adult offspring aged 16 and older. This information complements young people’s perspectives (as assessed for step-up respondents, see Chapter 10) by including their parents’ perspective (both anchors and their partners). Third, starting in wave 11, the *parenting u6 partner survey* was introduced to assess partners’ attitudes and behavior on parenting of children under the age of 6 years to complement the anchor survey modules on children under the age of 6 years with dyadic information for anchors’ partners.

The three parenting surveys followed different sample selection strategies. For an overview, see Table 8.1. Starting with wave 2, the *parenting survey* was presented to anchors and their partners with a child or children between 8 and 15 years of age and that was selected for the child interview. In order to capture information from parents not living together with their child, from wave 6 onwards the questionnaire was additionally handed out to anchors with a biological or adopted child between 8 and 15 years of age that lived exclusively outside the household, but was in contact with the respondent. From wave 7 onwards the questionnaire was also presented to anchors and their partners with a younger child or children (i.e. children between 6 and 7 years of age) and, therefore, followed up on the measurement of child development and parenting for 0 to 5 year old children as captured in the anchor CAPI. Subsequently, in wave 10, the age limit for children not living in the household was lowered as well.

Starting in wave 9, the *PAYA survey* was presented to anchors and their partners if they were living in the same household or were in contact with a child age 15 and older.

In wave 11, the *parenting u6 partner survey* was introduced for partners only who were living in one household with the anchor and a child or children under the age of 6 years.

For all three types of parenting data, more details on the selection criteria are documented at the respective places of handing out the PAPI in the anchor codebook. After the interviews, *Kantar Public* compiled raw data sets from the returned PAPI questionnaires. Data editing closely followed the procedures for editing of the anchor data. For the parenting survey and PAYA, information from both anchor and partner respondents is included in one single data set.

8.1 Parenting survey

The parenting survey has been part of the pairfam study since wave 2. It contains information from anchors and their partners pertaining to their relationships and interactions with their children up to the age of 15. All anchors with children selected for the parenting survey (see Table 8.1) were asked to fill out the parenting questionnaire, as were their cohabiting partners.

Table 8.1: Target groups for the three parenting surveys

Target group	Wave										
	2	...	6	7	8	9	10	11	12	13	14
Anchor and/or partner is selected to answer the parenting survey with regards to child x if the following conditions are met:											
Child x is a CAPI-child	X	X	X	X	X	X	X	X	X	X	X
Child x is a biological or adopted child, was born between 8 and 15 years before the start of fieldwork (and alive), lives exclusively outside the household, and is in contact with respondent			X	X	X	X	X	X	X	X	X
Child x is 6 or 7 years old and lives with the respondent in the same HH				X	X	X	X	X	X	X	X
Child x is a biological or adopted child, was born between 6 and 7 years before the start of fieldwork (and alive), lives exclusively outside the household, and is in contact with respondent							X	X	X	X	X
Anchor and/or partner is selected to answer the PAYA survey with regards to child x if the following conditions are met:											
Child x is a biological or adopted child, was born at least 16 years before the start of fieldwork (and alive), lives outside the household and is in contact with respondent or lives with the respondent in the same HH						X	X	X	X	X	
Partner is selected to answer the parenting u6 partner survey with regards to child x if the following conditions are met:											
Child x is a biological or adopted child, was born at max 5 years before the start of fieldwork (and alive) and partner lives in the same HH as anchor and child x								X	X	X	

8.1.1 Data editing

This section describes the editing of waves 2-14 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 respective parenting questionnaire codebook.

Missing values

Missing values were defined and labeled as in the anchor data set. Table 8.2 shows the missing codes and value labels assigned to the parenting data.

The following changes in defining missing values in the parenting data should be noticed: In waves 3 to 5, missing values “-1 Don’t know” and “-2 No answer” have been assigned if the respondent could not or did not answer a question. In wave 2 and as from wave 6, the missing category “-1 Don’t know” does not exist/no longer exists. Therefore code “-2 No answer” was consistently assigned if no box had been checked. As there were no filters and no open answers in the parenting questionnaires of waves 2 to 4, there were no missing codes “-3 Does not apply” and “-6 Unreadable answer” necessary. Beginning with wave 5, we included filtered questions. Therefore the value labels “-3 Does not apply” (if a question was skipped by the respondent correctly) were assigned. As we only assess subjective perceptions in the parenting questionnaires (in contrast to other survey parts) there were no consistency checks necessary and hence no “-5 Inconsistent value” (if logically impossible or empirically implausible combinations of values were found) assigned. For the first time in wave 7, respondents were able to give an open answer (*edu8i12o*). Open answers that were not legible were assigned the value “-6 Unreadable answer”, those that matched with existing categories of the answer list were coded to value “-4 Incorrect entry” (as described below). Code “-9 Invalid multiple answer” indicates that the respondent had ticked more than the allowed number of boxes.

Table 8.2: Missing codes in data sets *parenting\$*

Value	Label
-1	Don’t know
-2	No answer (also: I don’t want to answer that, answer refused)
-3	Does not apply
-4	Incorrect entry
-6	Unreadable answer
-9	Invalid multiple answer

Consistency checks

For filtered questions in the parenting data (beginning with wave 7) on school-related contents checks to identify implausible answers were conducted. In some cases information was filled in although anchor information (variable *crn13kxi12*) suggests that the child was not enrolled in school. Consequently, the provided values were left unedited but flag variables were generated to identify inconsistent answers. Flag variables were produced for the variables *edu5i_*, *edu7i_*, *edu10i_*, *edu11*, *edu12* and *edu13i_*. For each of these variables, code “0” indicates that the respective inconsistency is non-existent and code “1” that there is an inconsistency.

In wave 14, parental burnout was assessed for the first time. For parents with more than one child, the variables *pbi_* only needed to be answered once for one of their children. However, in some cases the parents provided different values for different children. Consequently, the provided values were left unedited but flag variables were generated to identify inconsistent answers between different children in the same family. For each of these variables, code “0” indicates that the respective inconsistency is non-existent, code “1” that there is an inconsistency for the anchor data, code “2” that there is an inconsistency for the partner data and code “3” that there is an inconsistency for both the anchor and the partner data. In cases without inconsistencies, parents either provided the same value for every single child in their family or only provided a value for one child. In the later case, all other children got assigned the same value.

Table A.29 describes the flag variables in detail.

Coding open answers

In the parenting data, open answers of variable *edu8i12o* (other extra-curricular activities), which contain information further qualifying the residual category of the answer list, were compared to the answer list. If possible, we recoded open answers into existing categories and set the original answer

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.

In wave 14, open answers of variable *hltc5* (“What do parents and children need most considering the COVID-19 crisis?”) were coded by the pairfam team into self-developed categories. The actual string was deleted for privacy reasons.

Anonymity

The children’s names were part of the parenting questionnaires 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 routines to identify incorrect entries. Routines were coded in SPSS.

English-language data

In the final step, we produced English-language data sets in which variable and value labels have been defined corresponding to the English codebooks.

8.1.2 Generated variables and scales

A major difference to the editing of the anchor data is that we did not produce generated variables. However, it is possible to generate scales (see Table 8.3) which are not directly included in the delivered data set but syntax files called *scales_parenting\$* are available (both in Stata and SPSS) and are part of the general distribution. For further information, refer to the scales manual (Thönnissen et al. 2023) which contains a detailed description of all scales.

Table 8.3: List of scales included in syntax file *scales_parenting\$*

Construct	Variable name	Wave
Parent-child-relationship: Readiness to make sacrifices	sacrif_pacs	2
Parent-child-relationship: Hostile attribution	hostattr_pacd	2 - 14
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 - 14
Parent-child-relationship: Psychological control	psycontrol_pacs	2,5
Parent-child-relationship: Negative communication	negcomm_pacs	2 - 14
Parent-child-relationship: Monitoring	monitor_pacs	2 - 14
Parent-child-relationship: Strict control	strict_pacs	2,5
Parent-child-relationship: Inconsistent parenting	inconsist_pacs	2 - 14
Parent-child-relationship: Rules	rules_pacs	9 - 14
SDQ: Hyperactivity	hyper_paco	2,5,7,9,11,13,14
SDQ: Emotional symptoms	emotion_paco	2 - 14
SDQ: Peer problems	peer_paco	2,5,7,9,11,13,14
SDQ: Prosocial behavior	prosoc_paco	2 - 14
SDQ: Conduct problems	conduct_paco	2 - 14
NRI Parent-child-relationship: Intimacy	intim_paco	2 - 14

continued on next page

continued from previous page

Construct	Variable name	Wave
NRI Parent-child-relationship: Conflict	confl_pacd	2 - 14
NRI Parent-child-relationship: Admiration	admir_pacs	2,4 - 14
NRI Parent-child-relationship: Dominance	domin_paco	2 - 14
Parental Burnout	pburnout	14

8.2 Parenting u6 partner survey

Starting with wave 11, we introduced an additional parenting survey for partners to assess dyadic information on children under the age of 6 years as addressed by the module for babies and toddlers and the module for 3 to 5 year old children in the anchor survey.

All partners with children selected for the parenting u6 survey (see Table 8.1) were asked to fill out the parenting u6 questionnaire. Please note that there are two data sets: *parentingU6partner\$* and *stepup_parentingU6partner\$*. Information on all selected children is included in one single data set.

8.2.1 Data editing

This section describes the editing of waves 11-13 parenting u6 partner 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 u6 partner variables start with the prefix “p”. Values were labeled according to the respective questionnaire codebook.

Missing values

Missing values were defined and labeled as in the anchor data set. Table 8.4 shows the missing codes and value labels assigned to the parenting u6 partner data.

Table 8.4: Missing codes in data sets *parentingU6partner\$*

Value	Label
-2	No answer (also: I don't want to answer that, answer refused)
-3	Does not apply
-4	Incorrect entry
-9	Invalid multiple answer

Consistency checks

For filtered questions in the parenting u6 partner data content checks to identify inconsistent answers were conducted. In some cases information was filled in although the child's age did not correspond to the age range required by the filter. Consequently, the provided values were left unedited but flag variables were generated to identify inconsistent answers. Flag variables were produced for the variables *pcrn43i_*, *pcrn44i_* and *pcrn100i_*. For each of these flag variables, code “0” indicates that the respective inconsistency is non-existent and code “1” that there is an inconsistency. Table A.30 describes the flag variables in detail.

Anonymity

The children's names were part of the parenting u6 partner questionnaires 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 routines to identify incorrect entries. Routines were coded in SPSS.

English-language data

In the final step, we produced English-language data sets in which variable and value labels have been defined corresponding to the English codebooks.

8.2.2 Generated variables and scales

A major difference to the editing of the anchor data is that we did not produce generated variables. However, it is possible to generate scales (see Table 8.5) which are not directly included in the delivered data set but syntax files called `scales_parentingU6partner$` are part of the general distribution. For further information, refer to the scales manual (Thönnissen et al. 2023) which contains a detailed description of all scales.

Table 8.5: List of scales included in syntax file `scales_parentingU6partner$`

Construct	Variable name	Wave
Parent-child-relationship: Hostile attribution	hostattr_pcd	11 - 13
Parent-child-relationship: Emotional warmth	warmth_pcs	11 - 13
Parent-child-relationship: Negative communication	negcomm_pcs	11 - 13
Parent-child-relationship: Inconsistent parenting	inconsist_pcs	11 - 13
Child temperament: Child temperament	temperc_pco	11 - 13
SDQ: Hyperactivity	hyper_pco	11 - 13
SDQ: Emotional symptoms	emotion_pco	11 - 13
SDQ: Conduct problems	conduct_pco	11 - 13

8.3 PAYA survey

Starting with wave 9, we introduced a second parenting survey called “Parenting Adolescents and Young Adults” (PAYA), which surveyed anchors and their partners regarding their relationship with their adolescent and young adult offspring aged 16 and older. This information complements young people's perspectives (as assessed for step-up respondents, see Chapter 10) by including their parents' perspective (both anchors and their partners). The PAYA survey was conducted as PAPI questionnaires addressing all anchors with children (for detailed instructions see the anchor's codebook) and their cohabiting partners (see Table 8.1). *Kantar Public* compiled a raw data set from the returned PAPI questionnaires which includes information from both anchor and partner respondents in one single data set.

8.3.1 Data editing

This section describes the editing of waves 9-13 PAYA data. The data editing process followed the procedures for anchor data as close as possible. 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 PAYA variables start with the prefix “paya”. Values were labeled according to the respective questionnaire codebook.

Missing values

Missing values were defined and labeled as in the anchor data set. Table 8.6 shows the missing codes and value labels assigned to the PAYA data.

Table 8.6: Missing codes in data sets *paya\$*

Value	Label
-1	Don't know
-2	No answer (also: I don't want to answer that, answer refused)
-3	Does not apply
-4	Incorrect entry
-5	Inconsistent value
-6	Unreadable answer
-9	Invalid multiple answer

Anonymity

The childrens' names were part of the PAYA questionnaires so as to facilitate the attribution of the questionnaires. For privacy concerns, this information was deleted during data preparation.

Value checks

To check value ranges, we used routines to identify incorrect entries. Routines were coded in Stata.

English-language data

In the final step, we produced English-language data sets in which variable and value labels have been defined corresponding to the English codebook.

8.3.2 Generated variables and scales

We produced a limited number of generated variables (see Table 8.7). In addition, it is possible to generate scales (see Table 8.8) which are not directly included in the delivered data set but syntax files called *scales_paya\$* are available (both in Stata and SPSS) and are part of the general distribution. For further information, refer to the scales manual (Thönnissen et al. 2023) which contains a detailed description of all scales.

Table 8.7: List of generated variables included in data set *paya\$*

Construct	Variable name	Wave
Enrollment in school or vocational qualification of child	enrolc	9 - 13
Highest school degree of child	schoolc	9 - 13
Highest vocational degree of child	vocatc	9 - 13
CASMIN classification of education of child	casminc	9 - 13

continued on next page

continued from previous page

Construct	Variable name	Wave
ISCED classification of education of child	iscedc	9 - 13
Years of schooling or vocational training of child	yeducc	9 - 13

Table 8.8: List of scales included in syntax file scales_*paya*\$

Construct	Variable name	Wave
NRI Parent-child-relationship: Intimacy (self)	payaintim_acs	9 - 13
NRI Parent-child-relationship: Intimacy (child)	payaintim_aco	9 - 13
NRI Parent-child-relationship: Conflict	payaconfl_acd	9 - 13
NRI Parent-child-relationship: Ambivalence	payaambiv_acd	9 - 13
Conflict style: Withdrawal (self)	payawithdraw_acs	9 - 13
Conflict style: Verbal Aggression (self)	payaverbaggr_acs	9 - 13
Conflict style: Constructive behavior (self)	payaconstrbh_acs	9 - 13
Conflict style: Withdrawal (child)	payawithdraw_aco	9 - 13
Conflict style: Verbal Aggression (child)	payaverbaggr_aco	9 - 13
Conflict style: Constructive behavior (child)	payaconstrbh_aco	9 - 13
Parental role: Overinvolvement	payaoverinv_acd	9 - 13
Parental role: Overprotection	payaoverpro_acd	9 - 13
Parental role: Autonomy support	payaauto_acd	9 - 13
Parental role: Career related support	payacsupp_acd	9 - 13
Parental role: Transition management beliefs	payacbelief_acd	9 - 13

9. DemoDiff

9.1 Introduction

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

The main design differences between DemoDiff and pairfam are:

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

Further, there are differences concerning questionnaire content:

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

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

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

9.2 Data structure

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

Table 9.1: Number of DemoDiff interviews

DemoDiff pairfam	Wave 1 Wave 2 2009/10	Wave 2 Wave 3 2010/11	Wave 3 Wave 4 2011/12	Wave 5 Wave 6 2012/13	W6	W7	W8	W9	...
Anchor	1,489	1,173	1,074	987	878	800	734	704	
Partner	684	578	550	490	435	378	322	320	
Child	—	—	—	233	240	239	229	253	
Parent	—	—	—	371	305	283	121	—	
Parenting	—	—	—	349	382	507	488	572	
Data sets	anchor1_DD partner1_DD	anchor2_DD anchor3 partner3	anchor4 partner4	anchor\$ partner\$ child\$ parent\$ parenting\$					

data set named anchor2_DD. Again it is left to the user, how to match these data with the original pairfam data. Table 9.1 gives an overview of the number of respondents for each wave of DemoDiff. Note that the sample selection of the parent survey changed in wave 8 (for details, see Chapter 3).

Obviously, including DemoDiff data in pairfam panel analyzes is not trivial. One problem is that DemoDiff “destroys” the monotonic design. In the original pairfam data, all respondents enter at wave 1, and some gradually attrite from the panel. If one appends data sets anchor\$, then all DemoDiff respondents enter at wave 3. Thus, when preparing data, users must be aware that some respondents enter at wave 3. Further, users must realize that DemoDiff respondents in wave t have not participated t times in the survey, but only $t-1$ times. Further, users need to decide how to merge DemoDiff wave 1 data. Merging it with pairfam wave 1 data creates an artificial gap in the panel structure (wave 2 information is missing). Merging it with pairfam wave 2 might create incompatibilities as the contents of DemoDiff wave 1 and pairfam wave 2 differ. Finally, if one intends to do weighted analyzes, one has to use special weights (for details see section 4.6).

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

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

9.3 Differences in detail

Some differences between the pairfam and the DemoDiff data sets exist before pairfam wave 5. When a question from the original pairfam questionnaire was not part of DemoDiff, the corresponding variable was set to -10 “Not in DemoDiff”. If a variable was not included in pairfam, but in DemoDiff, -11 “Not in pairfam” indicates this.

9.3.1 Variables excluded from DemoDiff

In order to adhere to the maximum interview time agreed upon with *Kantar Public*, some questions had to be eliminated from DemoDiff. Table A.1 to Table A.4 show the list of the variables that were deleted in DemoDiff waves 1 and 2. Note that these tables include both content-specific differences and missing variables caused by the differences in the multi-actor approach.

9.3.2 Additional variables in DemoDiff

There are a few additional variables included in DemoDiff which were not included in pairfam (see Table A.5 and Table A.6). In wave 1, this applies to regional information regarding the anchor person and the partner's place of birth. In wave 2, this applies to information on partnership status at first childbirth, retrospective activity and residential history, and the childhood history that was saved in anchor2_DD, as mentioned above.

9.4 Codebook

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

10. Step-up data

From wave 4 to wave 13, respondents of the children's survey who had reached the eligible age of 15 were asked to take part in the main anchor interview, including the partner interview. In addition, these "step-up" anchor respondents were asked to answer an additional, transitional PAPI questionnaire for the wave in which they first completed the main anchor questionnaire. This transitional questionnaire was optional, and respondents were asked to fill it out only once. In wave 14, no new step-up respondents were included in the anchor survey, but instead were invited to the child survey again. Data and documentation of the step-up respondents can be found in a separate step-up file as part of the Scientific Use File.

For waves 4-8 two separate step-up anchor data sets exist: *stepup_anchor\$* and *stepup_transition\$*. From wave 9 to wave 13, the transitional questionnaire was included in the CAPI interview. Therefore, information from the main anchor interview and the transitional interview is contained in one step-up data set (*stepup_anchor\$+transition*) and one step-up codebook. In addition, the partner data sets *stepup_partner\$* are available from wave 4 on as well as the generated data sets *stepup_biopart* and *stepup_biochild*. For wave 11 to wave 13, there is also a data set *stepup_parentingU6partner\$*.

Each step-up data set includes the variables *entry* and *cohort*. The variable *entry* indicates the wave in which the former CAPI children entered the anchor survey. With the variable *cohort* we differentiate between former CAPI children that have their first interview (category "0 Former capikid first interview") and former CAPI children that have participated in the anchor survey before (category "9 Former capikid re-interview"). Beginning with wave 7, two modules ("risk-taking" and "social media") were introduced, that only had to be answered by step-ups in the main anchor interview. Within the social media module, the scales "fear of missing out" and "group norms" were introduced in wave 9. Beginning with wave 8, step-ups' SDQ was also measured.

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

10.1 Data editing of transitional questionnaire

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

10.1.1 Variable and value labels

Values were labeled according to the step-up transition codebook. To facilitate the analysis, variables derived from questions that were included in the anchor survey were labeled identically. The 10-stage response format of the variables *co1_**, however, ranged from 1=Not important at all to 10=Very important, and was therefore not consistent to the 11-stage response format of the corresponding

variables `col_*` in the anchor data (0= Not important at all to 10= Very important) until wave 6. Since wave 6, the response formats have been harmonized. Similarly, the response format of the variable `cla8` had 10 stages (1= Very dissatisfied to 10=Very satisfied) until wave 6, while the corresponding variable in the anchor data `cla8` ranged from 0=Very dissatisfied to 10=Very satisfied. Since wave 6, these response formats have also been harmonized.

10.1.2 Missing values

For all variables of the transition step-up data we defined a set of missing codes, which were applied throughout the data (Table 13.1).

Table 10.1: Missing codes in data set *stepup*

Value	Label
-1	Don't know
-2	No answer (also: I don't want to answer that, answer refused)
-3	Does not apply
-4	Filter error / Incorrect entry
-5	Inconsistent value
-6	Unreadable answer
-97	Have no experience up to now

10.1.3 Anonymity

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

Table 10.2: List of variables made anonymous

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

10.1.4 Value checks

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

10.1.5 English-language data

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

10.2 Generated variables and scales

There are no generated variables included in the step-up transition data. Generated variables for the anchor data of this group are provided in the anchor step-up data sets. There are three step-up-specific scales (see Table 10.3), which are not part of the delivered data set. They can be generated by running the syntax file `scales_step-up$` which is available for waves 10 to 13 as part of the Scientific Use File.

Table 10.3: List of scales included in syntax file `scales_stepup_anchor$`

Construct	Variable name	Wave
<i>step-up data</i>		
Parental role: Overinvolvement	overinv_hp	10 - 13
Parental role: Autonomy support	autonomysup_hp	10 - 13
Parental role: Helicopter parenting	helicop	10 - 13

10.3 Identifying step-ups in their parents' anchor data

Merging of step-up data with their parents' anchor data or with their former child data is possible via the variable `f_cid` which represents their former child id (cid).

11. Refreshment sample

At the start of wave 11 a refreshment sample of new anchor respondents was drawn. In addition to a batch of respondents from a new, younger birth cohort (2001-2003), the two youngest cohorts from the pairfam base sample (1981-1983, 1991-1993) were strengthened. The oldest base sample cohort (1971-1973) was not refreshed. For further information on the composition of the refreshment sample, please read the corresponding section on the pairfam design in Technical Paper No. 01 (Brüderl et al. 2023).

The question program for the refreshment sample in wave 11 was, to a large extent, based on the regular questionnaire for the base sample. Only few questions were excluded due to, for example, a repeated participation requirement. In addition, new anchor respondents in wave 11 were posed several questions answered by the base sample in previous waves. For a more precise overview of the modified question program, please see Chapter 11.1.

Data editing for the refreshment sample followed the same rules as for the main study. Please refer to Chapter 4 and Chapter 5 for more details. Data from refreshment sample respondents is included in the anchor data sets. To distinguish between samples, the variable *sample* has been generated. For further information on this variable, see Chapter 2.2.

11.1 Question program

With few exceptions, the refreshment sample was posed the same CAPI question program as the base sample in wave 11. Questions concerning educational and vocational attainment since the previous wave and some questions from the separation module were deleted due to the requirement of repeated participation. Moreover, the modules concerning reasons for moving, pornography consumption, reasons against having (further) children, and parenting goals were omitted. For an overview of all excluded questions, refer to Table A.31.

Furthermore, some questions were only posed to the refreshment sample (see Table A.32). This includes questions asked only once to base sample respondents, but also recurring questions that the refreshment sample were posed outside of the normal sequence.

From wave 12 onwards, there were no differences in the question program for anchor respondents from the base and refreshment samples.

11.2 Retrospective relationship history

As for wave 1 respondents, refreshment sample respondents were posed the (similar) retrospective relationship history module in wave 11.⁴⁶ In contrast to wave 1, the wave 11 questionnaire included the Event History Calendar (EHC) which usually spans the period since the date of the last interview.

⁴⁶Retrospective information on previous relationship and cohabitation episodes with the current partner, education and employment, children (all originally asked in wave 1), own childhood history (wave 2), and residential biography (wave 3) were collected as well. Substantially, these modules are identical to the questions asked to the base sample years ago. Only filters had to be adapted and depending on question, time before EHC is considered instead of time before interview. Please have a look at the codebooks for more details.

Because there has been no preceding interview for the refreshment sample, January 2017 has been selected as starting point. Therefore, the relationship history module had to be modified to cover only the period before January 2017 and filters had to be adapted. All relationship episodes which took place before the EHC time frame were collected. This included previous relationship episodes both with partners not mentioned in the EHC and with partners mentioned in the EHC but who are not current partners.⁴⁷ Information on the name/gender of partner 1 to 5 were transferred into the retrospective module and stored in the variables *rtr1pxn/rtr1pxg* if partner X was mentioned in the EHC (*ehc1pxn!=3*). Accordingly, *x>=6* refers to partners not mentioned in the EHC. Relationship beginning (*rtr2pxezb_*) and end (*rtr2pxeze_*) were noted for all relationship episodes. Furthermore, additional episodes (*rtr3pxez*) were recorded. In addition, for ex-partners mentioned at the beginning of the EHC, relationship beginning (*rtr2pxe0b_*) was recorded. The variable *rtr1pxehc* indicates that a relationship existed at the beginning of the EHC (*rtr1pxehc=1* reflects that an ex-partner has been stated in the EHC).

An analogous approach was used for previous cohabitation periods. The variable *rtr6px* reflects a cohabitation period with partner x. Partners 1 to 5 are recorded in the EHC. Cohabitation beginning (*rtr7pxezb_*) and end (*rtr7pxeze_*) were noted for all cohabitation episodes. Furthermore, additional episodes (*rtr8pxez*) were recorded.

Retrospective information on marriages were recorded for current marriages not mentioned in the EHC, as well as previous marriages.

11.3 Generated variables, data sets, and scales

Generated variables for the refreshment sample are provided in the anchor data sets. Please refer to Chapter 4.3 for more details.

The refreshment sample is part of the episode data sets *bioact*, *biochild*, *biomob_ehc*, *bioparent*, and *biopart*. These user-friendly data sets allow for a convenient analysis of biographical information. For more details on these data sets, please refer to Chapter 4.7.

Scales for this sample are part of the anchor scales syntaxes. To provide the same scales for the base and refreshment samples, some scales already available for the base sample had to be added for the refreshment sample in wave 11.⁴⁸ Some scales are provided exclusively for the refreshment sample in wave 11 (see the following Table 11.1).

Table 11.1: Scales for refreshment sample only; included in syntax file *scales_anchor\$*

Construct	Variable name	Wave
SDQ-Behavior of anchors below 18: Conduct problems	conduct	11
SDQ-Behavior of anchors below 18: Emotional symptoms	emotion	11
SDQ-Behavior of anchors below 18: Prosocial behavior	prosoc	11
SDQ-Behavior of anchors below 18: Hyperactivity	hyper	11
SDQ-Behavior of anchors below 18: Behavioral issues with peers	peerprob	11

⁴⁷Information on previous relationship and cohabitation episodes with the current partner were posed in a separate module.

⁴⁸For these scales, please refer to Table 4.21.

12. Changes in survey mode due to COVID-19

In the spring of 2020, the COVID-19 pandemic interrupted the fieldwork of wave 12. As face-to-face interviews were no longer possible, the pairfam team switched the survey mode of the anchor and child interviews. Up to March 2020, Kantar Public followed standard computer-assisted personal interview (CAPI) procedures with both anchor and child respondents. After a brief fieldwork stop, the anchor and child survey was continued as a computer-assisted telephone interview (CATI) in late April 2020. The sensitive section of the anchor questionnaire, previously assessed in a computer-assisted self-interview (CASI), was substituted by a paper-and-pencil interview (PAPI). For more details on wave 12 survey design, fieldwork, and response rates, see Technical Paper No. 01 (Brüderl et al. 2023). In wave 13, respondents and interviewer could choose between CAPI/CASI and CATI/PAPI interviews.

As questions and filters were slightly adapted to the new survey modes, separate anchor and child codebooks for all interview modes are provided in the Scientific Use File. In the data, the new CATI/PAPI filtering of the anchor questionnaire was aligned to the more restrictive CAPI/CASI filters for better comparability. For instance, if a question in the standard CASI section was directed to a specific age group but the corresponding PAPI section had a broader filter, the responses of PAPI respondents not in this age group were recoded to “-4 Filter error / Incorrect entry”.

The switch in survey modes required several adjustments in data processing and delivery. Most importantly, anchor and child data of waves 12 to 14 are stored in separate data sets according to the main survey mode (*anchor\$_capi*, *anchor\$_cati*, *child\$_capi*, *child\$_cati*). The new variable *intmode* (*cintmode* for child data) additionally indicates interview mode (1=“CAPI & CASI”, 2=“CATI & PAPI”). Further, the new variable *intdat_papi* specifies the calendar week in which Kantar Public received the paper questionnaire from CATI/PAPI respondents. Consequently, these respondents have separate interview date variables for the CATI (*inty*, *intm*, *intd*) and PAPI (*intdat_papi*) surveys, as they may differ. Furthermore, a new missing code was generated for CATI respondents who did not return a completed paper questionnaire: In these cases, all variables from the PAPI questionnaire were coded with “-12 Non-response PAPI”.

Table 12.1: Overview of W12 survey modes and data structure

		before COVID-19	during COVID-19
Anchor 12	<i>Survey mode</i>	CAPI	CATI
		CASI	PAPI
	<i>Data set</i>	<i>anchor12_capi</i>	<i>anchor12_cati</i>
Child 12	<i>Survey mode</i>	CAPI	CATI
	<i>Data set</i>	<i>child12_capi</i>	<i>child12_cati</i>

The generated variables for waves 12 and 13 were computed with a combined data set of all anchor respondents. Therefore, users wishing to run the Stata do-files for generated variables must append the separate data sets *anchor\$_capi* and *anchor\$_cati* beforehand.

Using wave 12 data: The data before the pandemic (**_capi*) are structurally equivalent to previous pairfam waves. However, the data collected through the pandemic (**_cati*) have to be analyzed with care. The reason is that they are affected by both a mode effect (switch to CATI) and a period effect (COVID-19). One has to control for these, for instance by including *intmode* as a control variable. Do not interpret the effects of *intmode* as COVID-19 effects! These effects mix the mode and period effect. Further information and recommendations regarding data quality and handling of the wave 12 anchor/child data can be found in Technical Paper No. 16 (Bozoyan et al. 2021). For those interested in research questions specifically related to COVID-19, the supplemental pairfam COVID-19 survey is also available.

Using wave 13 data: In wave 13, respondents and the interviewer could choose the interview mode (CAPI/CASI or CATI/PAPI). In contrast to wave 12, there is no coincidence of mode and period effects. However, users should be aware of potential selection biases when analyzing the combined data set of all anchor respondents.

13. Changes to wave 14

Wave 14 marked the end of the official funding period of pairfam by the German Research Foundation (DFG). However, the surveying of the previous participants has not ended, but rather is continuing in the context of the newly-launched survey FReDA – The German Family Demography Panel Survey. In wave 14, the anchor persons were asked to consent to the linkage of their previous responses to the new study, FReDA, and to the transfer of their address data to GESIS – Leibniz Institute for the Social Sciences so that they could be contacted in the following survey year.

The transition to FReDA had certain consequences for wave 14 of pairfam. First, a further mode change took place for the respondents. Whereas the mode changes in the two preceding waves (waves 12 and 13) were due to the COVID-19 pandemic (see Chapter 12), the mode change in wave 14 was made in anticipation of the interview modes in the subsequent FReDA survey. Most of the pairfam anchor respondents completed the questionnaire online for the first time, which corresponds to the main survey mode in FReDA. For a detailed overview of this mode change and its consequences, see Section 13.1 below.

Second, the Event History Calendar (EHC) was no longer part of the anchor survey. Instead, the respective information on relationships (*bpa_*), jobs (*bjob_*), household composition (*bhc_*), and children (*bcrn_*) was collected by means of regular survey questions. For more information on this, see Section 13.2.

Third, although data editing for wave 14 followed the same rules as for the previous waves, certain parts had to be adapted, for example, because of the mode change and the replacement of the EHC. See Section 13.4. for more details.

13.1 Mode change in wave 14 and its consequences

Whereas in waves 12 and 13, the interview mode had to be adapted because of the COVID-19 pandemic, in wave 14, the course was set for the integration of the pairfam anchor respondents into FReDA by adapting the interviewer-administered questionnaire to a self-administered mode - either as a computer-assisted web interview (CAWI) or a paper-and-pencil interview (PAPI). In addition, a subsample was recontacted by an interviewer again to experimentally test whether the mode switch affected the data. Before the start of the field period, the anchor persons were randomly sorted into two unequally sized groups: a smaller face-to-face (f2f) group and a larger self-administered group comprising the majority of the anchor persons.

The anchors in the f2f group were interviewed by an interviewer who used the computer-assisted personal interview (CAPI) version of the web-based survey. The respondents were given the option to postpone the interview for some time if they had misgivings about personal contact during the COVID-19 pandemic. If this still did not result in an interview, they could, in exceptional cases, complete the self-administered paper-based or web-based version of the questionnaire. In a few cases, a computer-assisted telephone interview (CATI) was conducted.

In the self-administered group, CAWI was initially the default mode. However, in the second reminder letter, anchors respondents were also given the option to complete an enclosed paper-based (PAPI)

version of the questionnaire.

The anchor data are stored as three distinct data sets representing the three survey modes: *anchor14_capi* (CAPI plus 3 CATI interviews), *anchor14_cawi*, and *anchor14_papi*. For detailed information, see the methods report (Brix et al. 2023) or Technical Paper No. 1 (Brüderl et al. 2023).

From waves 1 to 13, the interviews were conducted either face-to-face (CAPI) or by telephone (CATI) via a questionnaire program on the interviewer's service laptop.⁴⁹ Thus, neither the web-based nor the paper-based version of the questionnaire in wave 14 could be as complex as the questionnaire program used in waves 1-13.⁵⁰ As a result, the Event History Calendar (EHC) was not used for the anchor survey in either mode, and the paper-based questionnaire differed from the online version to some extent. This not only affected the positioning of questions in the PAPI questionnaire, but also led to the omission, adaptation, or addition of questions in that version. For more information, see Sections 13.2 and 13.3 below.

The mode change had consequences not only for the anchor survey. In contrast to the previous waves, the partner questionnaires in wave 14 were dispatched exclusively by the survey provider, Kantar, together with a 5 Euro unconditional incentive rather than handed over by the interviewer. Furthermore, most of the child surveys were conducted via telephone.⁵¹

13.2 Replacement of the Event History Calendar in the anchor survey

Although the Event History Calendar (EHC) was an inherent part of the previous pairfam anchor surveys, it could not be realized in wave 14. As FReDA's web questionnaire is designed primarily for mobile devices, a complex graphical representation like the EHC is not feasible. Therefore, the questions about life events that were part of the former EHC were replaced with the "b-variables" (biography updates). This resulted in a loss of complexity because in waves 1 to 13, anchors had the possibility to select each month separately for episodes of relationships, jobs, and places of residence, whereas in wave 14, they could enter only the start and end dates. Moreover, the b-variables do not represent the entire content of the EHC. Deviations from the EHC for each category are discussed separately below.

bpa_ (relationship)

The previous division into partners 1 to 5 was replaced with the division into pre-wave partner (*bpa1-bpa9*), pre-wave marriage (*bpa10-bpa16*), current partner (*bpa17-bpa23*), and further relationships (*bpa24p_-bpa29p_*). Each partner's category has its own variable set.⁵²

There are some differences between the web-based and the paper-based versions. Because preloads could not be used in the PAPI questionnaire, anchor persons were first asked whether they had had a partner at the time of the last interview (*bpa30*), before answering the same questions on the pre-wave partner (*bpa1-bpa9*) as in the web-based interview.

The questions concerning a pre-wave marriage (*bpa10-bpa16*) were not part of the PAPI questionnaire, whereas questions on the current partner (*bpa17-bpa23*) were identical to those in the web-based questionnaire.

⁴⁹With the exception of the section of the wave 12 and wave 13 questionnaires devoted to sensitive questions, which was completed via a PAPI questionnaire.

⁵⁰Could be completed on different devices.

⁵¹For a more precise overview on the mode change in the child survey, see section 7.4.

⁵²For detailed information, see the Codebook for wave 14.

Not all questions on further relationships were part of the PAPI questionnaire. PAPI respondents were asked how many partners they had had since the last interview (*bpa24*⁵³). Then, in contrast to the web-based questionnaire, they were asked only to enter the partner's gender (*bpa25pXg*), birth month (*bpa25pXm*), and birth year (*bpa25pXy*).⁵⁴ Moreover, the start month (*bpa26pXbm*) and year (*bpa26pXby*) of the relationship as well as its end month (*bpa26pXem*) and year (*bpa26pXey*) had to be specified.

bjob_ (job)

The categories known from the job EHC were adapted for wave 14. Categories *i1* "General secondary school," *i2* "Evening school, adult education," and *i8* "Technical/professional school" were combined into one category. The same was done with categories *i3* "Vocational training" and *i7* "Pre-vocational training" as well as with categories *i13* "Internship, traineeship, voluntary service" and *i18* "Military service, alternative civilian service, voluntary social service year." Furthermore, there was only one category for university education, whereas there were two in previous waves.⁵⁵ Category *i4* "Vocational retraining/further education" was omitted completely in wave 14. Thus, the categories in wave 14 were as follows:

Table 13.1: *bjob1i_*

Variable	Label
<i>i1</i>	Full-time employment (not self-employed)
<i>i2</i>	Self-employed
<i>i3</i>	Part-time employment (not self-employed)
<i>i4</i>	Marginal part-time or irregular employment, mini-job, "Ein-Euro-Job"
<i>i5</i>	Vocational training, probationary year, military service, or similar
<i>i6</i>	Other type of employment, namely: " " (<i>bjob1i6o</i>)
<i>i7</i>	Unemployed, seeking employment
<i>i8</i>	School, evening school, university of applied sciences, or similar
<i>i9</i>	Degree course (university, university of applied sciences, dual study course)
<i>i10</i>	Vocational training, pre-vocational training
<i>i11</i>	Parental leave
<i>i12</i>	Housewife / househusband
<i>i13</i>	Retirement, occupational disability
<i>i14</i>	Other, no employment: " " (<i>bjob1i14o</i>)

The collection of data on job information differed between the PAPI and the web-based versions of the questionnaire in wave 14. Whereas the PAPI version covered only current activities (*bjob2i1-bjob2i14*) and their start month (*bjob2i1m-bjob2i14m*) and year (*bjob2i1y-bjob2i14y*), the web-based version also collected information on further episodes. The variable *bjob1iX* (*X* = order number of activity) states whether an anchor engaged in a certain activity or not. If an activity was mentioned, the start month (*bjob2iXb1m*) and year (*bjob2iXb1y*) as well as the end month (*bjob2iXe1m*) and year (*bjob2iXe1y*) could be specified in more detail. The variable *bjob2iXb* indicates whether an activity started before the previous wave (value 1) or not. If there was more than one episode for certain activities, the variable *bjob2a* has the value 1. Each further episode (*Y*) of an activity is listed in the variable *bjob2iXbY*. If there was a further episode, the start month (*bjob2iXbYm*) and year (*bjob2iXbYy*) as well as the end month (*bjob2iXeYm*) and year (*bjob2iXeYy*) could be specified. The variable *bjob2iXbY* indicates whether a further activity started before the previous wave (value 1) or not. If an activity was a current activity, the variable *bjob2iX* has the value 1.⁵⁶

⁵³Not the same variable as *bpa24p_*.

⁵⁴They did not have to enter their partner's name.

⁵⁵"*i7* University of cooperative education" and *i6* "University of applied sciences, college, university"

⁵⁶This accounts for the first and any further episodes.

***bhc_* (household composition)**

The variable set *bhc_* combines the entries for place of residence and household composition.⁵⁷ In wave 14, anchors did not have the possibility to enter a second place of residence. Rather, all questions focused on the main place of residence. The anchors were asked whether they still lived at the same place of residence (*bhc1*).⁵⁸ If they did not, they had to enter their new place of residence (*bhc2i1*), the corresponding federal state (*bhc2i2*), and the date on which they had moved to the new place of residence (*bhc3_*).

Afterwards, the anchors answered some questions on the household composition, but in a shortened version compared with the EHC. They were asked whether they lived alone at that place of residence (*bhc4*); how many people lived there, including themselves, their partners, and their children (*bhc5*); how many people in the household were younger than 15 (*bhc6*); and whether their biological mother/father or their stepparents were living in that household too (*bhc7i1-bhc7i5*).⁵⁹ If no (step)parents were living in that household, *bhc7i5* has the value 1.

***bcrn_* (children)**

In the EHC, all information on children was stored in the same variables from wave to wave, and was updated if new children were mentioned. In the self-administered version of the questionnaire in wave 14, by contrast, information on children mentioned in previous waves was stored only in the preload variables (*di_*),⁶⁰ unless the anchor person corrected this information. If this was the case, *bcrn1* (entries for child/children from previous wave are correct) equals 2 (= not correct) and *bcrn1akX* (X = xth child) equals 1. Corrected child information is stored in the variable set *bcrn2_*.⁶¹

The anchor could also add new children. In this case, the variable *bcrn3* equals 1 (one new child) or 2 (X new children).⁶² Information on children is stored in the variables *bcrn4kX_* (name, sex, and date of birth of child X), *bcrn5kX* (status of child X), and *bcrn6kX* (other biological parent of child X).

For all children (whether entered in previous waves or newly mentioned) the anchor was asked whether child X lived with them (*bcrn7kX*) or lived elsewhere (*bcrn8kX*), and to what extent child X lived with the anchor or the other parent (*bcrn9kX*).

In contrast to the web-based interview, anchor persons who completed the PAPI questionnaire were asked only whether they had children (*bcrn12*), and not whether they had additional children compared with the last wave. In the questions that followed, they could make entries only for their three youngest children. They were asked to enter the child's name (*bcrn13kXn*), gender (*bcrn13kXg*), birth month (*bcrn13kXm*), birth year (*bcrn13kXy*), status (*bcrn5kX*⁶³), and whether the child lived in the household of the anchor person (*bcrn11kX*⁶⁴).

The variable *bcrn6kX* ("Who is the other biological parent of the child?") is the only *bcrn*-variable in the newborn module of the web-based questionnaire that is also part of the PAPI questionnaire.

⁵⁷In the EHC, these variables were more or less in separate sections.

⁵⁸Only for anchors who completed the online version.

⁵⁹Only part of the web-based version.

⁶⁰For an overview, see the *di*-variables table in the codebook for wave 14.

⁶¹For more details, see the anchor codebook for wave 14.

⁶²The open entry is stored in the variable *bcrn3o*.

⁶³Same question as in the web-based questionnaire.

⁶⁴Although it appears to be same as the variable *bcrn7kX* in the web-based questionnaire, the categories differ significantly from the web-based version.

13.3 Further differences between the web-based and paper-based anchor questionnaires

Beside the differences in the *b*-variables used in the web-based and paper-based questionnaire, there are more differences to look at. Some questions were only asked in the web-based questionnaire. For an overview, please see Table A.32.

Moreover, few questions are part of the PAPI questionnaire only. Apart from the variable *sd10* “Marital Status”, the variables *job36* “Current occupational status”⁶⁵, *sex22* “Anchor or partner expecting a child”⁶⁶, *frt36* “(Additional) children realistic”⁶⁷, and *crn102* “Satisfaction with child care situation (all children younger than 14)”⁶⁸ are modifications or combinations of certain variables of the web-based questionnaire.

13.4 Data-cleaning process

For the data-cleaning procedures, the web-based (self-administered and *f2f*) and the paper-based data were combined, prepared together, and then separated again. Although the general data-preparation process was almost identical to that in previous waves, there were some noteworthy differences.

The EHC routines were the basis for the preparation of the *b*-variables. Whereas some parts could be adapted, most had to be written from the scratch due to the new data-generating process.

As the web-based and paper-based versions of the questionnaire differed - especially in terms of the filtering of questions - two separate filter checks were conducted.

In addition, child (*cid*) and partner IDs (*pid*) for the paper-based questionnaire were created based on the respective identification variables from wave 13 (or wave 12 in the case of temporary dropouts in wave 13) and information from wave 14.

Last but not least, child-specific questions (*bcrn_*/*crn_*) had to be integrated into the “old” EHC structure. In the web-based version of the questionnaire, there were some questions only for children born after the previous interview. As these children were not integrated into the “old” EHC structure in the raw interview data, they had to be sorted after the last child mentioned in the previous wave. In the paper-based version of the questionnaire, anchors could enter information only on the three youngest children. These were compared with children mentioned in previous waves and, if identified, they were sorted into the correct position. Newly mentioned/born children were sorted after the last child mentioned in the previous wave.

13.5 Further changes in wave 14

In contrast to previous waves, the parenting survey of adolescents and young children (PAYA) and the additional parenting survey for partners with children under the age of 6 years (parenting u6 partner survey) were not continued in wave 14. In addition, respondents of the child survey who had reached the age of 15 were still surveyed as part of the child survey. Thus, there were no new “step-up” anchor respondents and no separate “step-up” questionnaire in wave 14. Moreover, the generated data sets (see Section 4.7) were not supplemented with the information from wave 14 (only errors in the versions of Release 13.0 have been corrected; see Appendix).

⁶⁵Combination of *job30* to *job35*.

⁶⁶Combination of *sex3* and *sex4*.

⁶⁷Combination of *frt26* and *frt27*.

⁶⁸Not child-specific version of *crn15kX*.

Bibliography

- Bollen, K. A., P. P. Biemer, A. F. Karr, S. Tueller, and M. E. Berzofsky (2016). Are Survey Weights Needed? A Review of Diagnostic Tests in Regression Analysis. *Annual Review of Statistics and Its Application* 3(1), 375–392. <https://doi.org/10.1146/annurev-statistics-011516-012958>.
- Bozoyan, C., M. Herzig, R. Preetz, and C. Schmiedeberg (2021). Effects of survey design modifications due to the covid-19 pandemic in pairfam wave 12. pairfam Technical Paper No. 16, doi: <https://doi.org/10.5282/ubm/epub.91998>.
- Bozoyan, C. and C. Schmiedeberg (2022). The Factorial Survey Experiment on “Infidelity Norms and Attitudes” in the German Family Panel (pairfam). pairfam Technical Paper No. 18, doi: <https://doi.org/10.5282/ubm/epub.92000>.
- Brauns, H. and S. Steinmann (1999). Educational Reform in France, West-Germany, and the United Kingdom. Updating the CASMIN Educational Classification. *ZUMA-Nachrichten* 44(23), 7–44.
- Brick, J. M., I. F. Cervantes, S. Lee, and G. Norman (2011). Nonsampling errors in dual frame telephone surveys. *Survey Methodology* 37(1), 1–12.
- Brüderl, J., L. Castiglioni, V. Ludwig, K. Pforr, and C. Schmiedeberg (2017). Collecting Event History Data with a Panel Survey: Combining an Electronic Event History Data with a Panel Survey. *methods, data, analyses* 11(1), 45–66. doi: <https://doi.org/10.12758/mda.2016.013>.
- Brüderl, J., B. Huyer-May, and C. Schmiedeberg (2013). Interviewer Behavior and the Quality of Social Network Data. In P. Winker, R. Port, and N. Menold (Eds.), *Interviewers’ Deviations in Surveys. Impact, Reasons, Detection and Prevention*, pp. 147–160. Peter Lang, Frankfurt.
- Brüderl, J., V. Ludwig, K. Pforr, and N. Schumann (2011). Praktische Anwendungsbeispiele zum Umgang mit den pairfam-Daten (Welle 1). pairfam Technical Paper No. 02, doi: <https://doi.org/10.5282/ubm/epub.91984>.
- Brüderl, J., C. Schmiedeberg, L. Castiglioni, O. Arránz Becker, P. Buhr, D. Fuß, V. Ludwig, J. Schröder, and N. Schumann (2023). The German Family Panel - Study Design and Cumulated Field Report (Waves 1 to 14). pairfam Technical Paper No. 01, doi: <https://doi.org/10.5282/ubm/epub.104749>.
- Buhr, P. and J. Huinink (2014). Adjustment of variable frt6. pairfam Technical Paper No. 04, doi: <https://doi.org/10.5282/ubm/epub.91986>.
- Düval, S. and K. Auspurg (2020). The Factorial Survey Experiment on "Distribution of Work in Partnerships" in the German Family Panel (pairfam). pairfam Technical Paper No. 14, doi: <https://doi.org/10.5282/ubm/epub.91996>.
- Düval, S. and K. Auspurg (2022). The Factorial Survey Experiment on "Distribution of Work in Partnerships" in the German Family Panel (pairfam). Content and Technical Information on the Partner Vignettes. pairfam Technical Paper No. 20. doi: <https://doi.org/10.5282/ubm/epub.92002>.
- 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. doi: <https://doi.org/10.2307/589632>.

- Frietsch, R. and H. Wirth (2001). Die Übertragung der Magnitude-Prestigeskala von Wegener auf die Klassifizierung der Berufe. *ZUMA-Nachrichten* 48(3), 139–163.
- Ganzeboom, H. and D. Treiman (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. In J. Hoffmeyer-Zlotnik and C. Wolf (Eds.), *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables*, pp. 159–193. New York: Kluwer Academic Press.
- Ganzeboom, H. and D. Treiman (2010). International Stratification and Mobility File: Conversion Tools.
- Gebel, L., N. Böcker, V. Melchior, S. Köster, and P. Buhr (2017). Quantitative und qualitative erfassung der gründe keine (weiteren) kinder zu bekommen in pairfam. pairfam Technical Paper No. 10, doi: <https://doi.org/10.5282/ubm/epub.91992>.
- Gelman, A. (2007). Struggles with Survey Weighting and Regression Modeling. *Statistical Science* 22(2), 153–164. <https://doi.org/10.1214/088342306000000691>.
- Hartmann, J., G. Schütz, and J. Suckow (2010). Die Vercodung der Berufe bei PAIRFAM. Technical report, TNS Infratest Sozialforschung, München.
- Hartmann, J., G. Schütz, and J. Suckow (2011a). Die Vercodung der Berufe bei pairfam (welle 2). Technical report, TNS Infratest Sozialforschung, München.
- Hartmann, J., G. Schütz, and J. Suckow (2011b). Die Vercodung der Berufe bei pairfam (welle 3). Technical report, TNS Infratest Sozialforschung, München.
- Hartmann, J., N. Tschersich, and G. Schütz (2012). Die Vercodung der offenen Angaben zur beruflichen Tätigkeit nach der Klassifikation der Berufe 2010 (KldB2010) und nach der International Standard Classification of Occupations 2008 (ISCO08). Entscheidungsregeln bei nicht eindeutigen Angaben. Technical report, TNS Infratest Sozialforschung, München.
- Huinink, J., J. Brüderl, B. Nauck, S. Walper, L. Castiglioni, and M. Feldhaus (2011). Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual Framework and Design. *Zeitschrift für Familienforschung* 23, 77–101. doi: <https://doi.org/10.20377/jfr-235>.
- Kolenikov, S. (2014). Calibrating survey data using iterative proportional fitting (raking). *The Stata Journal: Promoting communications on statistics and Stata* 14(1), 22–59. <https://doi.org/10.1177/1536867X1401400104>.
- Kolenikov, S. (2019). Updates to the ipfraking ecosystem. *The Stata Journal: Promoting communications on statistics and Stata* 19(1), 143–184. <https://doi.org/10.1177/1536867X19830912>.
- König, W., P. Lüttinger, and W. Müller (1988). A Comparative Analysis of the Development and Structure of Educational Systems. Methodological Foundations and the Construction of a Comparative Educational Scale. CASMIN Working Paper No. 12, University of Mannheim.
- Kreyenfeld, M., J. Huinink, H. Trappe, and R. Walke (2012). DemoDiff: A Dataset for the Study of Family Change in Eastern (and Western) Germany. *Schmollers Jahrbuch* 132, 653–660. doi: <https://doi.org/10.3790/schm.132.4.653>.
- Lohr, S. L. and J. N. K. Rao (2000). Inference from Dual Frame Surveys. *Journal of the American Statistical Association* 95(449), 271–280. <https://doi.org/10.2307/2669544>.
- Mund, M., C. Finn, B. Hagemeyer, J. Zimmermann, and F. J. Neyer (2015). The Dynamics of Self-Esteem in Partner Relationships. *European Journal of Personality* 29(2), 235–249. doi: <https://doi.org/10.1002/per.1984>.
- Nübling, M., H. H. Andersen, and A. Mühlbacher (2006). Entwicklung eines Verfahrens zur Berechnung der körperlichen und psychischen Summenskalen auf Basis der SOEP-Version des SF12 (Algorithmus). Technical report, DIW Data Documentation, Berlin.

- Sand, M. (2018). *Gewichtungsverfahren in Dual-Frame-Telefonerhebungen bei Device-Specific Nonresponse*. Ph. D. thesis, GESIS - Leibnitz-Institut für Sozialwissenschaften, Köln. <https://doi.org/10.21241/ssoar.60293>.
- Schneider, S. (2008). Applying the ISCED-97 to the German educational qualifications. In S. Schneider (Ed.), *The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries*, pp. 76–102. Mannheim Centre for European Social Research, University of Mannheim.
- Schnell, R., T. Bachteler, and S. Bender (2004). A Toolbox for record linkage. *Austrian Journal of Statistics* 33(1-2), 125–133. doi: <https://doi.org/10.17713/ajs.v33i1&2.434>.
- Schober, P. S., S. Büchau, M. Philipp, C. K. Spiess, C. Bozoyan, and C. Schmiedeberg (2022). The Factorial Survey Experiment on “Normative Judgements of Work-Care Arrangements” in the German Family Panel. pairfam Technical Paper No. 19, doi: <https://doi.org/10.5282/ubm/epub.92001>.
- Solon, G., S. J. Haider, and J. M. Wooldridge (2015). What Are We Weighting For? *Journal of Human Resources* 50(March), 301–316. doi: <https://doi.org/10.3368/jhr.50.2.301>.
- Sonntag, M., M. Mund, E. Schubach, and F. J. Neyer (2014). Report of a method effect in the assessment of personality traits in pairfam. Technical report, Lehrstuhl für Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik, Friedrich-Schiller-Universität Jena, Jena. doi: <https://doi.org/10.5282/ubm/epub.91987>.
- Sonntag, M., F. Neyer, and E. Schubach (2015). Measurement invariance of the pairfam scales. pairfam Technical Paper No. 06, doi: <https://doi.org/10.5282/ubm/epub.91988>.
- Suckow, J. and U. Schneekloth (2009). Beziehungen und Familienleben in Deutschland (2008/2009) Welle 1. Methods report, TNS Infratest Sozialforschung, München.
- Thönnissen, C., J. Reim, S. Geissler, P. Alt, B. Sawatzki, S. Böehm, and S. Walper (2023). pairfam Scales and Instruments Manual, Release 14.1. LMU Munich: Technical Report. GESIS Data Archive, Cologne. ZA5678 Data file Version 14.1.0, doi: <https://doi.org/10.4232/pairfam.5678.14.1.0>.
- UNESCO (2006). International Standard Classification of Education: ISCED-97. Technical report, Montreal: UNESCO Institute for Statistics.
- Walper, S., B. Sawatzki, P. Alt, J. Reim, C. Schmiedeberg, C. Thönnissen, and M. Wetzel (2021). The pairfam COVID-19 survey: Design and instruments. Release 1.1. pairfam Technical Paper No. 15, doi: <https://doi.org/10.5282/ubm/epub.91997>.
- Ware, J. E., M. Kosinski, and J. E. Dewey (2001). *How to Score Version Two of the SF-36 Health Survey (Standard and acute forms)*. Qualitymetric Inc. (3rd edition), Lincoln, RI.
- Ware, J. E., M. Kosinski, D. M. Turner-Bowker, and B. Gandek (2002). *How to Score Version 2 of the SF12 Health Survey (with a supplement documenting version 1)*. Qualitymetric Inc., Lincoln, RI.
- Wegener, B. (1984). Gibt es Sozialprestige? *Zeitschrift für Soziologie* 14(3), 209–235. doi: <https://doi.org/10.1515/zfsoz-1985-0303>.
- Wetzel, M., N. Schumann, and C. Schmiedeberg (2021). New weights for the pairfam anchor data. pairfam Technical Paper No. 17, doi: <https://doi.org/10.5282/ubm/epub.91999>.

A. Appendix

A.1 DemoDiff-pairfam differences

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

Variable	Variable label
co1_	Future: Importance to be successful in job (Question 5)
bce1_	VOP+: Undertake activities with partner (Question 60)
sin1_	Interest of potential partner (Question 61)
sin2_	Interested in potential partner (Question 62)
sin4_	Dating possibilities (Question 65)
sin5_	Trying to find out sth. about personality (Question 66)
sin6_	Somebody like me always finds a partner (Question 67)
pa1_	Interested in partner (Question 68)
pa4_	Partner introduced to parents (Question 77)
pa10_	Feelings regarding common household (Question 83)
pa12_	Feelings regarding marriage (Question 85)
cps2_	First request language partner survey (Question 96)
cps4_	Handing over partner questionnaire right away (Question 99)
pa16_	Let partner know that I understand him/her (Question 101)
pa17_	Telling partner what you are thinking (Question 102)
pa18_	Partner finds it all right if I pursue own interests (Question 103)
pa19_	I hope relationship lasts for a long time (Question 106)
sat4_	Estimation of partner's satisfaction with relationship (Question 105)
pa20_	Serious relationship problem: Prob. alcohol, medication, drugs (Question 107)
pa22xiy	Insulted or abused partner (Question 109)
pa23_	Problems with alcohol, medication, or other drugs last year (Question 110)
pa24_	Cheating past year (Question 111)
pa25_	Arguments using physical force past year (Question 112)
pa28_	Suggested a separation/divorce past year (Question 115)
sex1_	Age first sex (Question 116)
sat5_	Contraception used past 3 months (Question 120)
lsr1i_	Leisure time (Question 138)
lsr2_	Hours watching TV past week (Question 139)
lsr3_	Hours spent on personal Internet use past week (Question 140)
lsr4_	Weeks on vacation past year (Question 141)
lsr5_	Leisure with partner: Coffee stores, bars, restaurants (Question 142)
lsr6_	Online profile on social network website (Question 143)
lsr7_	Visibility of online profile (Question 144)
lsr8_	Frequency of visit: Social network sites (Question 145)
hc11h1	Main residence: Monthly expenditures for rented ap./house (Question 157)

continued on next page

continued from previous page

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

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

Variable	Variable label
pbce1_	Undertake activities with partner (Question 4)
psat3	Satisfaction with relationship (Question 5)
psat4	Estimation of partner's satisfaction with relationship (Question 6)
ppa19_	I hope relationship lasts for a long time (Question 12)
ppa16_	Anchor let partner know that I understand him/her (Question 13)
ppa22p_	Discussion behaviour (Question 15)
ppa17_	Handling of differences in relationship (Question 16)
ppa18_	Special situations in relationship (Question 17)
psat5	Satisfaction with sex life (Question 28)
plsr1i_	Leisure time (Question 36)
plsr2	Hours watched TV past week (Question 36)
plsr3	Hours spent on personal Internet use past week (Question 36)
plsr4	Weeks on vacation trips past week (Question 36)
pper1_	Often agree with others, even if I am not sure (Question 37)
phlt2	Number of times feeling low and melancholy past 4 weeks (Question 46)
phlt3	Body height in cm (Question 47)

continued on next page

continued from previous page

Variable	Variable label
phlt4	Weight in kg (Question 48)
phlt7	Hours of sleep per night (Question 49)
phlt6	To what extent handicapped? (Question 51)

Table A.3: Variables not included in DemoDiff wave 2, anchor3

Variable	Variable label
hc5h1	Main residence: Ownership of dwelling/house (Question 7)
hc5h2	Second Residence: Ownership of dwelling/house (Question 7)
hc11h1	Main residence: Monthly expenditures for rented dwelling (Question 8)
hc11h2	Second Residence: Monthly expenditures for rented dwelling (Question 8)
hc12h1	Main residence: Monthly expenditures for self-owned dwelling (Question 9)
hc12h2	Second Residence: Monthly expenditures for self-owned dwelling (Question 9)
hc13h1	Main residence: Total living space in square meters (Question 10)
hc13h2	Second Residence: Total living space in square meters (Question 10)
hc14h1	Main residence: Number of rooms (Question 11)
hc14h2	Second Residence: Number of rooms (Question 11)
sd33	Attending same school as in previous wave (Question 13)
sd25	Type of school currently attending (Question 14)
sd26	Grade in school (Question 15)
job5_	Owner of company (Question 23)
job6_	Employed in own firm (Question 24)
bce1_	VOP+: Undertake activities with partner (Question 37)
sin4_	Dating possibilities (Question 42)
sin5_	Trying to find out sth. about personality (Question 43)
pa1_	Interested in partner (Question 48)
pa10_	Feelings regarding common household (Question 81)
pa12_	Feelings regarding marriage (Question 83)
pa16_	Let partner know that I understand him/her (Question 91)
pa17_	Telling partner what you are thinking (Question 92)
pa18_	Partner finds it all right if I pursue own interests (Question 93)
sat4	Estimation of partner's satisfaction with relationship (Question 95)
pa19_	I hope relationship lasts for a long time (Question 96)
pa20_	Serious relationship problem (Question 97)
pa22_	Assessment of partner (Question 99)
pa29	Frequency child witnessed arguments past year (Question 100)
pa23	Problems with alcohol, medication, or other drugs last year (Question 101)
pa24	Extra-marital affair of anchor or partner past year (Question 102)
pa25	Arguments using physical force past year (Question 103)
per1_	I often agree with others, even if I'm not sure (Question 107)
per4_	How did you feel in the last 4 weeks (Question 108)
per2_	How do you feel in general (Question 109)
sex8	Frequency of sexual intercourses last 3 months (Question 128)
sex9	To what extend do the following statements apply to you? (Question 129)
crn19_	Parenting goals (Question 159)
crn1_	Place of birth child x (Question 160)
crn2_	Complications child x: health/life of child was at risk (Question 161)
crn3_	Child x: C-section? (Question 162)

continued on next page

continued from previous page

Variable	Variable label
crn4_	Child x: In what week of the pregnancy born (Question 163)
crn5_	Child x: Length at the time of birth in cm (Question 164)
crn6_	Child x: Weight at the time of birth in grams (Question 165)
crn7_	Child x: Type of last checkup (Question 166)
crn37_	Child x: Was your child breast-fed and, if yes, how long (months) (Qu. 167)
crn24_	Child x: How much time did your child need to fall asleep (minutes) (Qu. 168)
crn25_	Child x: How many times per night did your child wake up (Question 169)
crn26_	Child x: How strongly burdened by child's sleeping behavior (Question 170)
crn38_	Child x: Does child set bedtime or do you as parents set bedtime (Qu. 171)
crn39_	Child x: How long altogether does your child sleep at night (Question 172)
crn27_	Child x: How long altogether does your child sleep at night (Question 173)
crn28_	Child x: How often burdened by child's crying or fussiness (Question 174)
crn10_	Child x: My child is $\frac{1}{2}$ (Question 175)
crn11_	I often have no energy (Question 176)
crn29_	Worries concerning own children (Question 177)
crn31_	Child x: Health, generally speaking (Question 179)
crn20_	Role as a parent (Question 183)
ccs1k_	Consent to children's survey child x (Question 189)
ccs2k_	Point of time of children's survey child x (Question 190)
cpas1	Consent to own participation in parenting survey (Question 191)
cpas2	Pick up parenting survey or send by mail (Question 192)
cpas3	Consent to partner's participation in parenting survey (Question 193)
cpas4	Pick up partner's parenting survey or send by mail (Question 194)
lsr1i_	Leisure time (Question 195)
lsr2_	Hours watched TV past week (Question 196)
lsr3_	Hours spent on personal internet use past week (Question 197)
lsr4_	Weeks on vacation trips past year (Question 198)
lsr5_	Leisure with partner (Question 199)
lsr6_	Online profile on social network website (Question 200)
lsr7_	Visibility of online profile (Question 201)
lsr8_	Frequency of visit: Social network sites (Question 202)
igr73i_	Citizenship adoptive mother (Question 205)
igr74	Country of birth adoptive mother (Question 206)
igr77i_	Citizenship adoptive father (Question 212)
igr78	Country of birth adoptive father (Question 213)
igr81	Is mother's new partner same partner as at the time of the last interview (Question 222)
igr82	Citizenship mother's new partner (Question 225)
igr84	Is father's new partner same partner as at the time of the last interview (Question 230)
igr85	Citizenship father's new partner (Question 233)
igr41p2	Timely distance to dwelling of mother's partner (Question 244)
igr41p4	Timely distance to dwelling of father's partner (Question 246)
cprs1p1	Consent to survey mother (Question 251)
cprs2p1	Mother: Name and address (Question 252)
cprs3p1	Language version of questionnaire mother (Question 253)
cprs1p2	Consent to survey mother's partner (Question 254)
cprs2p2	Mother's partner: Name and address (Question 255)
cprs3p2	Language version of questionnaire mother's partner (Question 256)
cprs1p3	Consent to survey father (Question 257)

continued on next page

continued from previous page

Variable	Variable label
cprs2p3	Father: Name and address (Question 258)
cprs3p3	Language version of questionnaire father (Question 259)
cprs1p4	Consent to survey father's partner (Question 260)
cprs2p4	Father's partner: Name and address (Question 261)
cprs3p4	Language version of questionnaire father's partner (Question 262)
inc25_	Parents: Financial situation (Question 296)
hlt1	Health status past 4 weeks (Question 301)
hlt3	Body height in cm (Question 302)
hlt4	Weight in kg (Question 303)
hlt7	Hours of sleep at night during working week (Question 304)
int12_	Suggestions or comments regarding the interview (Question 307)

Table A.4: Variables not included in DemoDiff wave 2, partner3

Variable	Variable label
psat4	Estimation of partner's satisfaction with relationship (Question 6)
ppa19_	I hope relationship lasts for a long time (Question 12)
ppa16_	Anchor let partner know that I understand him/her (Question 13)
ppa21p_	Disagree with partner (Question 14)
ppa22p_	Discussion behaviour (Question 15)
ppa17_	Handling of differences in relationship (Question 16)
ppa18_	Special situations in relationship (Question 17)
pcrn32_	Feeling as a parent: Own role (Question 33)
pcrn20_	Feeling as a parent: Role of partner (Question 33)
plsr1i_	Leisure time (Question 34)
plsr2	Hours watched TV past week (Question 34)
plsr3	Hours spent on personal Internet use past week (Question 34)
plsr4	Weeks on vacation trips past year (Question 34)
pigr42p_	Frequency: You tell your mother/father what you are thinking (Question 40)
pigr45p_	Frequency: You and mother/father annoyed/angry with each other (Question 41)
pigr47p_	Frequency: You and your mother/father disagree and quarrel (Question 42)
pigr49p_	Frequency: You share with mother/father secrets/private feelings (Question 43)
phlt1	Health status past 4 weeks (Question 51)
phlt3	Body height in cm (Question 52)
phlt4	Weight in kg (Question 53)
phlt7	Hours of sleep per night during working week (Question 54)
pper1i2	Sometimes I believe that I'm worthless (Question 61)
pper1i6	I feel lonely (Question 61)
pper2_	How do you feel in general (Question 62)

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

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

Table A.6: Additional variables in DemoDiff, wave 2, anchor2_DD + anchor3

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

A.2 Changes between releases

Changes in *anchor* data sets

Table A.7: Changes in data set *anchor1*

Variables affected	Description of changes	Changed in release
<i>New generated variables</i>		
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 generated variables (identifiers: anchor's, partner's, children's sex \ anchor's, partner's, children's, parents', stepparents' date of birth)	2.0
mschool, fschool, mcasmin, fcasmin, miscd, fiscd, myeduc, fyeduc	New variables (Parents' educational status)	4.0
mvocat, fvocat	New variables (Parents' vocational status)	4.0
ykage, ykid	New variables (Youngest child living with anchor)	4.0
iscd2	New variable (ISCED classification of educational attainment, incl. students)	4.0
flag_isco08_kldb2010	New variable (Flag variable KIdB and ISCO)	4.0
flag_ehc	New variable (Flag variable EHC)	4.0
flag_igb	New variable (Flag variable intergenerational relations)	4.0
d1weight, ca1weight, d1ca1weight	New weighting variables	5.0
hhincoecd, npu14mr, npo14mr	New equivalence income variable and information on household composition necessary for equivalence scale weight	6.0
parentidkx	Person number second biological parent of anchor's child(ren) were included from the data set biochild	6.0
homosex_new	New variable (Anchor's revealed sexual orientation)	7.0
<i>Modified main variables</i>		
all corresponding child variables	2 repositioned children	7.0
continued on next page		

Table *anchor1* continued

Variables affected	Description of changes	Changed in release
<i>Modified generated variables</i>		
east	New variable (Anchor currently living in East Germany)	2.0
ethni	New variable (Anchor's ethnicity)	2.0
migstatus	New variable (Anchor's migration status)	2.0
childmrd	New variable (Child(ren) living at main residence)	2.0
age, page, mage, fage, k*age	Age month-specific (based on month & year of birth); year of anchor's interview used to compute all age variables instead of 2008	2.0
bula	Federal state Saarland now can be distinguished from Rhineland	2.0
cob, fcob, mcob, pcob, nat1, nat2, pnat1, pnat2, mnat1, mnat2, fnat1, fnat2	Missing values combined to -7	2.0
cohort	Value labels changed to "1991-1993", "1981-1983", "1971-1973"	2.0
hhincnet	Set to -7 instead of 0 if inc13, inc14, inc15, inc16, inc17 is -4	2.0
hhcomp	Variable content changed (Household composition)	2.0
homosex	Computed using the generated identifiers sex_gen and psex_gen	2.0
incnet	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	2.0
infertile	-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	2.0
intdur	System missing (.) recoded to -7	2.0
iscd, pscd	Value labels of categories 4 & 5 interchanged	2.0
k1type ... k10type	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	2.0
lfs, plfs	If more than one work activity first act. status overrides 2nd act. status	2.0
ncoh	Set to -7 if np is -7 (see above)	2.0
np	Set to -7 ("Incomplete data") if just one partner mentioned and this partner's sex is -1 or -2	2.0

continued on next page

Table *anchor1* continued

Variables affected	Description of changes	Changed in release
pregnant	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	2.0
infertile	Self-report of partner is used for computation of variable as well as the anchor's report, self-report dominates indirect report	3.0
enrol, penrol	Modified and new value: value 16 also includes "Vocational retraining / further education"; new value 7 "Kolleg, 2. Bildungsweg"	3.0
meetdur, reldur, cohabdur, mardur	Some values modified according to generated variables marstat and relstat	3.0
np, ncoh	Slightly modified due to corrections of data set <i>biopart</i>	3.0
school, pschool	Value change (3 to 4 and vice versa) and rename of value label "POS 8./9."	3.0
school, pschool, casmin, pcasmin, isced, pisced, yeduc, pyeduc	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	3.1
kldb2010	New classification schema KldB 2010	4.0
isco08, isei, siops	New classification schema ISCO-08	4.0
age, page, mage, fage, k*age	New coding in cases of ambiguous seasonal information on the month and valid answers for the year, random values for month impute	4.0
nkids*, k*type, hhsizemrd	Some modifications for three cases due to correction of information on children (see above)	4.0
cpi	New baseline year 2010	5.0
pnkidsbioalv	Correction of coding number of partner's biological children	5.0
pregnant, mschool, mvocat, fvocat, mcasmin, misced, myeduc, fyeduc	Minor corrections for different variables	5.0
yeduc, pyeduc, myeduc, fyeduc	Minor corrections for different variables	6.0
mschool, fschool	Correction of value labels: "3 lower GDR, POS 8./9." to "3 intermediate, Realschulabschluss / mittlere Reife" and "4 intermediate, Realschulabschluss / mittlere Reife" to "4 lower GDR, POS 8./9."	7.0

continued on next page

Table *anchor1* continued

Variables affected	Description of changes	Changed in release
isced, isced2, pisced	Correction of value labels: “4 upper secondary education general (3b)” to “4 upper secondary education vocational (3b)” and “5 upper secondary education vocational (3a)” to “5 upper secondary education general (3a)”	7.0
npo14mr, npu14mr, hhincoecd	Correction of number of persons aged under 14/ aged 14 and older in main residence	9.0
pcasprim, pcassec, plfs	Correction of coding if no information available: “-3 Does not apply” set to “-7 Incomplete data” for anchor respondents with partner	10.0
reldur	Set to “-7 Incomplete data” for some cases - relationship break started before intw1 and end of break unknown	10.0
mardur	Correction from “-7 Incomplete data” to “-3 Does not apply” for anchor respondents which weren't married	10.0
cpi	New baseline year 2015	11.0
ethni	Some cases were changed from “-7 Incomplete data” to “2 Ethnic-German Immigrant (Aussiedler)”	11.0
mschool, fschool	Recoding of value 3 to 4 and value 4 to 3 to make coding consistent with school and pschool	11.0
parentidkX	Minor corrections of parentidkX	11.0
<i>Discontinued generated variables</i>		
m2rd	Variable discontinued (Mother living at main residence)	2.0
f2rd	Variable discontinued (Father living at main residence)	2.0
oth2rd	Variable discontinued (Others living at main residence)	2.0
hhsz2rd	Variable discontinued (Household size second residence)	2.0
sexratio	Variable discontinued (sex ratio)	4.0
popdens	Variable discontinued (population density)	4.0
<i>Modified tag and flag variables</i>		
flag26	Additional couples of anchor respondents identified	3.0
flag7	Value “1 Inconsistency” instead of values indicating episodes to gain equivalence across waves	4.0

continued on next page

Table *anchor1* continued

Variables affected	Description of changes	Changed in release
flag20-flag25, flag_ehc, flag_igb	Variables added to achieve consistency across waves (-3 for all cases)	4.0
flag_isco88_kldb1992	Variable label modified (German and English version)	4.0
flag_cas, flag_frt6	Variable and value labels modified (German and English version)	4.0
flag1-flag18	Value labels modified (German and English version)	4.0
flag26	Additional couples of anchor respondents identified	5.0
flag27	Variable added to achieve consistency across waves (-3 for all cases)	7.0
<i>Modified variable names of main variables</i>		
sex, dobm, doby	New variable names original_sex original_dobm original_doby	5.0
<i>Modified labels of main variables</i>		
pa16i1-6, pa22pi1-8, pa22ri1-8, frt1, frt2, sdp9i2, pa18i11, col13	Variable labels modified (German and English version)	2.0
pa17i5	Variable label modified (German version)	2.0
sex3, mig3, job5i1, job5i4, job6i3, inc19i2, inc19i3, inc19i4, inc19i5, int1, int2, int3	Variable labels modified (only German version: "Anker" instead of "ZP")	2.0
cps3, cps3i3, cps8i3	Value labels modified (only German version: "Anker" instead of "ZP")	2.0
frt11v1i*, pa16*	Variable labels modified (German version)	3.0
job18*	Variable labels modified (English version)	3.0
frt2	Variable label modified (German and English version)	3.0
sd13, rtr11, rtr15p*	Value labels modified (German version)	3.0
frt11v2i1-int10i2	Variable labels modified (German and English version): Question number in variable label corrected	3.1
<i>Modified labels of generated variables</i>		
isco88	Variable label of variable isco based on former classification schema ISCO-88 modified (German and English version)	4.0
kldb1992	Variable label of variable kldb based on former classification schema modified (German and English version)	4.0
kldb2010, isco08	Value labels modified (German and English version)	4.0
gkpol, bik, bula	Variable and value labels modified (German and English version)	4.0
continued on next page		

Table *anchor1* continued

Variables affected	Description of changes	Changed in release
<i>Modified values of main variables</i>		
sat2	Value of variable sat2 is modified: -1 is set to 8; -2 is set to 9	3.0
sd14k*-sd21k*, rtr16k*-rtr21k*, sd22k*	Correction of information on children for id=165520000, 308441000 & 689470000	4.0

Table A.8: Changes in data set *anchor2*

Variables affected	Description of changes	Changed in release
<i>New generated variables</i>		
mschool, fschool, mcasmin, fcasmin, miscd, fiscd, myeduc, fyeduc	New variables (Parents' educational status)	4.0
mvocat, fvocat	New variables (Parents' vocational status)	4.0
flag_isco08_kldb2010	New variable (Flag variable KldB and ISCO)	4.0
ykage, ykid, ykagecapi	New variables (Youngest child/capi-child living with anchor)	4.0
d1weight, calweight, d1calweight	New weighting variables	5.0
hhincoecd, npu14mr, npo14mr	New equivalence income variable and information on household composition necessary for equivalence scale weight	6.0
parentidkx	Person number second biological parent of anchor's child(ren) were included from the data set biochild	6.0
homosex_new	New variable (Anchor's revealed sexual orientation)	7.0
<i>Modified main variables</i>		
all corresponding child variables	2 deleted and 1 repositioned child	7.0
ehc12kX	Correction to 10 (other person) for some cases because other parent is not part of EHC anymore	8.0
<i>Modified generated variables</i>		
pcob	Use information of wave 1 if no valid information available in wave 2; variable only for new partners	3.0

continued on next page

Table *anchor2* continued

Variables affected	Description of changes	Changed in release
pnat1, pnat2	Nationalities of new partners were stored in downward order, now in ascending order; variables only for new partners	3.0
infertile	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	3.0
enrol	Modified and new value: value 16 also includes "Vocational retraining / further education"; new value 7 "Kolleg, 2. Bildungsweg"	3.0
school, pschool	Value change (3 to 4 and vice versa) and rename of value label "POS 8./9."	3.0
marstat	Modified values: -7, 2, 3	3.0
cohabdur, mardur	Some values modified according to generated variables marstat and relstat	3.0
ncoh, nmar	Slightly modified due to corrections of data set <i>biopart</i>	3.0
sex_gen, psex_gen, k*sex_gen, dobm_gen, siops	Value labels modified	3.0
school, pschool, casmin, pcasmin, isced, pised, isced2, pised2, yeduc, pyeduc	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	3.1
age, page, mage, fage, smage, sfage, k*age, homosex, hhcomp, lweight, ppanel, pcontact, panswer	Values of some cases corrected (see generated identifiers)	3.1
nat1, *nat2, *cob, infertile, *enrol, *school, *vocat, *casmin, *iscd, *iscd2, *yeduc, siops, *sex_gen, *doby_gen, *dobm_gen	Variable labels modified (English version)	3.1

continued on next page

Table *anchor2* continued

Variables affected	Description of changes	Changed in release
k*dobm_gen, k*doby_gen, k*sex_gen	Value labels modified (English version)	3.1
lweight	Recoding of values greater than 5 to 5	4.0
klldb2010	New classification schema KldB 2010	4.0
isco08, isei, siops	New classification schema ISCO-08	4.0
age, page, mage, fage, smage, sfage, k*age	New coding in cases of ambiguous seasonal information on the month and valid answers for the year, random values for month imputed	4.0
marstat	Some changes from “1 Never married” to “3 Divorced/dissolved civil union” due to marriage at some point in the EHC	4.0
relstat	Some corrections due to marriage at some point in the EHC	4.0
pnkidsbioalv	Some changes of cases where the information from the partner questionnaire was not available	4.0
k*sex_gen, k*doby_gen, k*dobm_gen, k*age, nkids, k*type	Some modifications due to correction of information on children in wave 1 (see above)	4.0
cpi	New baseline year 2010	5.0
pnkidsbioalv	Correction of coding number of partner’s biological children	5.0
mschool, mvocat, fvocat, mcasmin, misced, myeduc, fyeduc	Minor corrections for different variables	5.0
yeduc, pyeduc, myeduc, fyeduc	Minor corrections for different variables	6.0
mschool, fschool	Correction of value labels: “3 lower GDR, POS 8./9.” to “3 intermediate, Realschulabschluss / mittlere Reife” and “4 intermediate, Realschulabschluss / mittlere Reife” to “4 lower GDR, POS 8./9.”	7.0
pcasprim, pcassec, plfs	Correction of coding if no information available: “-3 Does not apply” set to “-7 Incomplete data” for anchor respondents with partner	10.0
reldur, cohabdur	Minor corrections for reldur and cohabdur	10.0
cpi	New baseline year 2015	11.0
ethni	Some cases were changed from “-7 Incomplete data” to “2 Ethnic-German Immigrant (Aussiedler)”	11.0
mschool, fschool	Recoding of value 3 to 4 and value 4 to 3 to make coding consistent with school and pschool	11.0

continued on next page

Table *anchor2* continued

Variables affected	Description of changes	Changed in release
parentidkX	Minor corrections of parentidkX	11.0
<i>Discontinued generated variables</i>		
sexratio	Variable discontinued (sex ratio)	4.0
popdens	Variable discontinued (population density)	4.0
<i>Modified tag and flag variables</i>		
flag26	Additional couples of anchor respondents identified	3.0
flag8, flag16	Value label -3 recoded to 0 (theoretically possible)	3.0
flag26, flag_frt6	Value labels slightly modified	3.0
flag_frt6, flag20, flag25	Variable label modified	3.0
flag1, flag5, flag16, flag17, flag_cas, pflag_cas	Value labels modified (English version)	3.1
flag16	Value 0 recoded to -3	4.0
flag8, flag_frt6	Variable labels modified (German and English version)	4.0
flag6-flag15, flag18	Value labels modified (German and English version)	4.0
flag26	Additional couples of anchor respondents identified	5.0
flag27	Variable added to achieve consistency across waves (-3 for all cases)	7.0
<i>Modified variable names of main variables</i>		
sex, dobm, doby	New variable names original_sex original_dobm original_doby	5.0
<i>Modified labels of main variables</i>		
ehc19i13m1-18, ehc19i13, d175, cpas3, ehc13k1-7, cla5e1-10, frt11v1i*, crn19i1, sdp9i1, sdp9i6, sdp9i7	Variable label modified (German version)	3.0
inc10i12-inc10i16, sdp9i6, sdp9i7	Variable labels modified (English version)	3.0
ehc5p1-5	Value labels modified (German version)	3.0
cps5, pa3, cprs2p*, inc27*	Value labels modified (English version)	3.0
sex9i*, sep1i*, per1*, inc25*, inc26*	Value labels modified (German and English version)	3.0
frt11v1*	Variable labels modified (German and English version)	3.1
crn19i1	Variable labels modified (English version)	3.1
sin3*, sin4*, sin5*, sin6*, pa1*, frt10*, crn10*, crn11*, crn20*, cpas5	Value labels modified (German and English version)	3.1
continued on next page		

Table *anchor2* continued

Variables affected	Description of changes	Changed in release
cps8i3	Value labels modified (German version)	3.1
pa3, cprs2, netp*n, inc25, inc26, inc27, capikid, he3, hm2, hsv2, hv2, hsm2	Value labels modified (English version)	3.1
<i>Modified labels of generated variables</i>		
isco88	Variable label of variable isco based on former classification schema ISCO-88 modified (German and English version)	4.0
kldb1992	Variable label of variable kldb based on former classification schema modified (German version)	4.0
kldb2010, isco08	Value labels modified (German and English version)	4.0
bik, bula	Value labels modified (German and English version)	4.0
intsex	Value labels modified (English version)	4.0
wavedist	Variable label added (English version)	14.1
<i>Modified value labels</i>		
d327, d328, d329, d330, d331, d332, d333, d334, d335, d336	Changed english label <i>liste180_en</i> (0 "Child is alive" to 0 "Child is dead" and 1 "Child is dead" to 1 "Child is alive")	5.0
<i>Modified values of main variables</i>		
frt24i1-8	96 is set to 6; 97 is set to 7	3.0
ehc18p1-4	Value labels of categories 1 & 0 interchanged	3.0
ehc7k*-ehc13k*, hc15k*, sep4k*-sep10k*, crn1k*-crn10k*, crn12k*-crn18k*, ccs1k*, ccs2k*	Correction of information on children for id=689470000 (also see changes in wave 1)	4.0

Table A.9: Changes in data set *anchor3*

Variables affected	Description of changes	Changed in release
<i>New generated variables</i>		
continued on next page		

Table *anchor3* continued

Variables affected	Description of changes	Changed in release
mschool, fschool, mcasmin, fcasmin, misced, fised, myeduc, fyeduc	New variables (Parents' educational status)	4.0
mvocat, fvocat	New variables (Parents' vocational status)	4.0
ykage, ykid, ykagecapi, ykidcapi	New variables (Youngest child living with anchor)	4.0
flag_isco08_kldb2010	New variable (Flag variable KIdB and ISCO)	4.0
tag_idntp	New variable (Tag inconsistent identity current partner between waves)	4.0
hhincoecd, npu14mr, npo14mr	New equivalence income variable and information on household composition necessary for equivalence scale weight	6.0
parentidkx	Person number second biological parent of anchor's child(ren) were included from the data set biochild	6.0
homosex_new	New variable (Anchor's revealed sexual orientation)	7.0
<i>Modified main variables</i>		
pid	One correction of pid (id=241451000)	4.0
ehc23p2	Modified value	5.0
ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pXh1-ehc25pXh3	Corrections because cohabiting ex-partner was deleted in household grid	6.0
pid, ehc12kX, pa30, pa13, crn17kX, crn23kX, crn35k1iX, crn35k2iX, crn36kX	Correction of information on partnership for id=67350000, 100983000, 369237000, 606712000, 680227000 & 704733000	7.0
all corresponding child variables	1 deleted child	7.0
ehc12k2	Correction to 10 (other person) because other parent is not part of EHC anymore	8.0
ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pXh1	Deletion of current partner from household grid	9.0
ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pXh1	Deletion of mistakenly created children from household grid	9.0
ehc3p1m32 ehc3p1 ehc26h1	Creation of missing cohabitation entry for 4 partners	9.0
all corresponding child variables *k3* and *k1*	1 repositioned child	9.0
<i>Modified generated variables</i>		
pschool, pcasmin, pised, pised2, pyeduc	Variable pschool changed due to corrections in anchor1 & anchor2 plus recoding of implausible values for new partners: "3 lower GDR, POS 8./9." to "2 lower, Volks-/Hauptschulabschluss" and "5 intermediate GDR, POS 10." to "4 intermediate, Realschulabschluss / mittlere Reife" since partners born after 1980 (POS 8./9.) or 1979 (POS 10.) could not have earned a degree from a polytechnic secondary school (only existed in the GDR) plus correction for partners without school degree who are not currently in school; changes in other variables due to changes of pschool	4.0

Table *anchor3* continued

Variables affected	Description of changes	Changed in release
infertile	Correction of coding infertile	4.0
lweight	Recoding of values greater than 5 to 5	4.0
klldb2010	New classification schema KldB 2010	4.0
isco08, isei, siops	New classification schema ISCO-08	4.0
age, page, mage, fage, smage, sfage, k*age	New coding in cases of ambiguous seasonal information on the month and valid answers for the year, random values for month imputed	4.0
marstat	Some changes from "1 Never married" to "3 Divorced/dissolved civil union" due to marriage at some point in the EHC	4.0
relstat	Some corrections due to marriage at some point in the EHC	4.0
cpi	New baseline year 2010	5.0
smcob, sfcob	Correction of coding variables of DemoDiff sample	5.0
mschool, fschool, mvocat, fvocat, mcasmin, misced, myeduc, fyeduc	Minor corrections for different variables	5.0
yeduc, pyeduc, myeduc, fyeduc	Minor corrections for different variables	6.0
pmrdr, mmrd, fmrdr, childmrd, othmrd, hhcomp, hhsizemrd	Corrections for variables regarding household size and household composition due to changes in household grid	6.0
mschool, fschool	Correction of value labels: "3 lower GDR, POS 8./9." to "3 intermediate, Realschulabschluss / mittlere Reife" and "4 intermediate, Realschulabschluss / mittlere Reife" to "4 lower GDR, POS 8./9.'	7.0
np, ncoh, nmar, meetdur, mardur, nkidspalv, nkidspliv, kXtype	Some modifications for six cases due to correction of information on partnership (see above)	7.0
pnat1, kXtype	Minor corrections for different variables	7.0
mardur	Minor corrections for mardur	8.0
cohabdur	Minor corrections for cohabdur	9.0
pcasprim, pcassec, plfs	Correction of coding if no information available: "-3 Does not apply" set to "-7 Incomplete data" for anchor respondents with partner	10.0
cohabdur	Minor correction for cohabdur	10.0
infertile	Set to "-7 Incomplete data" for 1 case	10.0
cpi	New baseline year 2015	11.0

continued on next page

Table *anchor3* continued

Variables affected	Description of changes	Changed in release
ethni	Some cases were changed from “-7 Incomplete data” to “2 Ethnic-German Immigrant (Aussiedler)”	11.0
mschool, fschool	Recoding of value 3 to 4 and value 4 to 3 to make coding consistent with school and pschool	11.0
relstat, marstat	Minor corrections for relstat and marstat	11.0
parentidkX	Minor corrections of parentidkX	11.0
<i>Discontinued generated variables</i>		
sexratio	Variable discontinued (sex ratio)	4.0
popdens	Variable discontinued (population density)	4.0
<i>Modified tag and flag variables</i>		
flag_cas, pflag_cas	Value labels modified (English version)	3.1
flag_ehc, flag_frt6	Variables added to achieve consistency across waves (-3 for all cases)	4.0
flag26	Additional couples of anchor respondents identified	5.0
flag27	Variable added to achieve consistency across waves (-3 for all cases)	7.0
<i>Modified variable names of main variables</i>		
sex, dobm, doby	New variable names original_sex original_dobm original_doby	5.0
<i>Modified labels of main variables</i>		
job18h, job18m	Variable labels modified (English version)	3.1
pa11, pa14*, pa17*, pa18*, pa19*	Value labels modified (German and English version)	3.1
rtr26*m, rtr30*m, rtr33*m, rtr37*m, rtr41*m, rtr45m, he3, hm2, hsv2, hv2, hsm2	Value labels modified (English version)	3.1
<i>Modified labels of generated variables</i>		
isco88	Variable label of variable isco based on former classification schema ISCO-88 modified (German and English version)	4.0
kldb1992	Variable label of variable kldb based on former classification schema modified (German version)	4.0
kldb2010, isco08	Value labels modified (German and English version)	4.0
intsex	Value labels modified (English version)	4.0
continued on next page		

Table *anchor3* continued

Variables affected	Description of changes	Changed in release
wavedist	Variable label added (English version)	14.1
<i>Modified value labels</i>		
d327, d328, d329, d330, d331, d332, d333, d334, d335, d336	Changed english label <i>liste180_en</i> (0 “Child is alive” to 0 “Child is dead” and 1 “Child is dead” to 1 “Child is alive”)	5.0

Table A.10: Changes in data set *anchor4*

Variables affected	Description of changes	Changed in release
<i>New generated variables</i>		
d1weight, calweight, d1calweight	New weighting variables	5.0
hhincoecd, npu14mr, npo14mr	New equivalence income variable and information on household composition necessary for equivalence scale weight	6.0
parentidkx	Person number second biological parent of anchor's child(ren) were included from the data set biochild	6.0
homosex_new	New variable (Anchor's revealed sexual orientation)	7.0
<i>Modified main variables</i>		
cid2, cid3, and all corresponding child variables *k2* and *k3*	1 repositioned child	5.0
ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pX	Corrections because cohabiting ex-partner was deleted in household grid	6.0
ehc26	Corrections current partner lives in main residence	6.0
pid, ehc12kX, pa30, pa13, crn17kX	Correction of information on partnership for id=67350000, 97393000, 100983000, 369237000, 390748000, 606712000, 680227000, 704733000 & 748663000	7.0
all corresponding child variables	3 deleted children	7.0
ehc12kX	Correction to 10 (other person) for 2 cases because other parent is not part of EHC anymore	8.0

continued on next page

Table *anchor4* continued

Variables affected	Description of changes	Changed in release
ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pX	Deletion of current partner from household grid	9.0
ehc22p11n, ehc23p11, ehc24p11m, ehc24p11y, ehc25p11	Deletion of 1 mistakenly created child from household grid	9.0
ehc31	Recoding of value -4 to -2	9.0
ehc12k1, ehc12k2, crn17k1, crn17k2	Corrections because current partner is parent of both children for id=586215000	9.1
<i>Modified generated variables</i>		
cpi	New baseline year 2010	5.0
smcob, sfcob, pnkidsbioalv, k*type	Correction of coding variables	5.0
isei	Recoding of missing values	5.0
siops	Recoding of missing values	5.0
pschool, pcasmin, pisced, pisced2, pyeduc	Recoding of implausible values for new partners; changes in other variables due to changes of pschool	5.0
mschool, fschool, mvocat, fvocat, mcasmin, misced, myeduc, fyeduc	Minor corrections for different variables	5.0
lweight, ppanel, pcontact, panswer	Modifications of weighting variables	5.0
yeduc, pyeduc, myeduc, fyeduc	Minor corrections for different variables	6.0
lweight, ppanel, pcontact, panswer	Minor corrections of weighting variables	6.0
pmrd, mmrd, fmrdr, childmrd, othmrd, hhcomp, hhsizemrd	Minor corrections for different variables regarding household size and household composition due to changes in household grid	6.0
mschool, fschool	Correction of value labels: "3 lower GDR, POS 8./9." to "3 intermediate, Realschulabschluss / mittlere Reife" and "4 intermediate, Realschulabschluss / mittlere Reife" to "4 lower GDR, POS 8./9."	7.0
np, ncoh, nmar, meetdur, mardur, nkidsp, nkidspalv, nkidspliv, kXtype	Some modifications for nine cases due to correction of information on partnership (see above)	7.0
np, nmar, mardur	Minor corrections for different variable	7.0
vocat	Step-up anchor respondents set to "-3 Does not apply"	8.0
cob, nat1, nat2	Additional information was merged from the Step-up transition data set	8.0
mcob, fcob, mnat1, mnat2, fnat1, fnat2	Additional information on the Step-ups' parents was merged from their parents anchor data	8.0

continued on next page

Table *anchor4* continued

Variables affected	Description of changes	Changed in release
ethni, migstatus	Additional information on the Step-ups available because of additional information on country of birth	8.0
mardur	Minor correction for mardur	8.0
pcasprim, pcassec, plfs	Correction of coding if no information available: “-3 Does not apply” set to “-7 Incomplete data” for anchor respondents with partner	10.0
reldur, marstat, mardur	Minor corrections for reldur, marstat and mardur	10.0
cpi	New baseline year 2015	11.0
ethni	Some cases were changed from “-7 Incomplete data” to “2 Ethnic-German Immigrant (Aussiedler)”	11.0
migstatus	Minor changes for the Step-ups	11.0
mschool, fschool	Recoding of value 3 to 4 and value 4 to 3 to make coding consistent with school and pschool	11.0
marstat	Minor corrections for marstat	11.0
parentidkX	Minor corrections of parentidkX	11.0
<i>Modified tag and flag variables</i>		
flag17, flag26	Additional couples of anchor respondents identified	5.0
flag27	Variable added to achieve consistency across waves (-3 for all cases)	7.0
<i>Modified variable names of main variables</i>		
sex, dobm, doby	New variable names original_sex original_dobm original_doby	5.0
<i>Modified labels of generated variables</i>		
wavedist	Variable label added (English version)	14.1

Table A.11: Changes in data set *anchor5*

Variables affected	Description of changes	Changed in release
<i>New generated variables</i>		
hhincoecd, npu14mr, npo14mr	New equivalence income variable and information on household composition necessary for equivalence scale weight	6.0
parentidkx	Person number second biological parent of anchor's child(ren) were included from the data set biochild	6.0
homosex_new	New variable (Anchor's revealed sexual orientation)	7.0
pmcob, pfcob	New variables (Country of birth of the partner's parents)	8.0
pethni, pmigstatus	New variables (Partner's ethnicity and migration status)	9.0
<i>Modified main variables</i>		
crn33px, crn34px, crn35kx, crn36kx	Minor filter corrections for different variables	6.0
ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pX	Corrections because cohabiting ex-partner was deleted in household grid	6.0
ehc26	Corrections current partner lives in main residence	6.0
pid, ehc12kX, pa30, pa13, crn17kX, crn23kX, crn18kX, crn48k1iX-crn48k3iX, crn35k1iX, crn35k2iX, crn36kX	Correction of information on partnership for id=67350000, 97393000, 158925000, 369237000, 606712000, 680227000, 704733000, 748663000 & 864337000	7.0
ehc33pX	Corrections because cohabiting ex-partner was deleted in household grid; further corrections	7.0
ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pX, ehc33pX	Deletion of current partner from household grid	9.0
ehc22p11n, ehc23p11, ehc24p11m, ehc24p11y, ehc25p11, ehc33p11	Deletion of 1 mistakenly created child from household grid	9.0
sib3p4, sibXp3	Deletion of sibling 3 and correction of sex of sibling 4 for one case because of anchor/interviewer notes wave 9	9.0
pa11, crn18kX, ehc4p1, ehc4p1mX	Correction of 1 case because of anchor/interviewer notes wave 9	9.0
ehc31	Recoding of value -4 to -2	9.0

continued on next page

Table *anchor5* continued

Variables affected	Description of changes	Changed in release
ehc12k1, ehc12k2, crn17k1, crn17k2, crn23k1, crn35k1i1-2, crn35k2i1-2, crn36k1, crn18k1, crn18k2	Corrections because ex-partner is parent of both children for id=586215000	9.1
sib2p1	Correction of 3 cases from 7 to -4	11.0
sib2pX	Recoding of falsely “-3 Does not apply” coded cases to “7 no sibling”	11.0
<i>Modified variable names of main variables</i>		
crn45k1iX-crn45k10iX	New variable names crn45ak1iX-crn45ak10iX due to change in value ranges in waves 6 and the following waves	9.0
<i>Modified generated variables</i>		
yeduc, pyeduc, myeduc, fyeduc	Minor corrections for different variables	6.0
lweight, ppanel, pcontact, panswer	Minor corrections of weighting variables	6.0
pmr, mmrd, fmr, childmrd, othmrd, hhcomp, hhsizemrd	Minor corrections for different variables regarding household size and household composition due to changes in household grid	6.0
mschool, fschool	Correction of value labels: “3 lower GDR, POS 8./9.” to “3 intermediate, Realschulabschluss / mittlere Reife” and “4 intermediate, Realschulabschluss / mittlere Reife” to “4 lower GDR, POS 8./9.”	7.0
np, ncoh, nmar, meetdur, mardur, nkidsp, nkidspalv, nkidspliv, kXtype	Some modifications for nine cases due to correction of information on partnership (see above)	7.0
pnat1, np, nmar, cohabdur, mardur	Minor corrections for different variables	7.0
vocat	Step-up anchor respondents set to “-3 Does not apply”	8.0
cob, nat1, nat2	Additional information was merged from the Step-up transition data set	8.0
mcob, fcob, mnat1, mnat2, fnat1, fnat2	Additional information on the Step-ups’ parents was merged from their parents anchor data	8.0
ethni, migstatus	Additional information on the Step-ups available because of additional information on country of birth	8.0
relstat, marstat, cohabdur	Minor corrections for marstat, relstat and cohabdur	8.0
meetdur, reldur	Minor corrections for meetdur and reldur	9.0
pfcob	Minor corrections for pfcob	9.0
continued on next page		

Table *anchor5* continued

Variables affected	Description of changes	Changed in release
hpm	Correction of 1 case because of anchor/interviewer notes wave 9	9.0
pcasprim, pcassec, plfs	Correction of coding if no information available: “-3 Does not apply” set to “-7 Incomplete data” for anchor respondents with partner	10.0
reldur, marstat, mardur	Minor corrections for reldur, marstat and mardur	10.0
cpi	New baseline year 2015	11.0
ethni, pmigstatus	Minor corrections for ethni and pmigstatus	11.0
migstatus	Minor changes for the Step-ups	11.0
mschool, fschool	Recoding of value 3 to 4 and value 4 to 3 to make coding consistent with school and pschool	11.0
marstat	Minor corrections for marstat	11.0
parentidkX	Minor corrections of parentidkX	11.0
<i>Modified labels of generated variables</i>		
wavedist	Variable label added (English version)	14.1

Table A.12: Changes in data set *anchor6*

Variables affected	Description of changes	Changed in release
<i>New generated variables</i>		
homosex_new	New variable (Anchor's revealed sexual orientation)	7.0
pmcob, pfcob	New variables (Country of birth of the partner's parents)	8.0
pethni, pmigstatus	New variables (Partner's ethnicity and migration status)	9.0
<i>Modified main variables</i>		
pid, ehc12kX, pa13, crn17kX, crn49kX, crn52k1iX-crn52k3iX	Correction of information on partnership for id=67350000, 97393000, 158925000, 606712000, 680227000, 704733000, 748663000 & 864337000	7.0
ehc33pX	Corrections because cohabiting ex-partner was deleted in household grid; further corrections	7.0
ehc12kX	Correction to 10 (other person) for some cases because other parent is not part of EHC anymore	8.0
ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pX, ehc33pX	Deletion of current partner from household grid	9.0
ehc3p1m32 ehc3p1 ehc26	Creation of missing cohabitation entry for 2 partners	9.0
pa11, pa34iX, ehc4p1, ehc4p1mX	Correction of 1 case because of anchor/interviewer notes wave 9	9.0
ehc31	Recoding of value -4 to -2	9.0
ehc12k1, ehc12k2, crn17k1, crn17k2, crn49k1, crn49k2	Corrections because ex-partner is parent of both children for id=586215000	9.1
per3i1-per3i21	Minor changes in English labels	10.0
<i>Modified generated variables</i>		
np, ncoh, nmar, meetdur, mardur, nkidsp, nkidspalv, nkidspliv, kXtype	Some modifications for eight cases due to correction of information on partnership (see above)	7.0
pnat1, marstat, np, nmar, meetdur, mardur	Minor corrections for different variables	7.0
vocat	Step-up anchor respondents set to "-3 Does not apply"	8.0
cob, nat1, nat2	Additional information was merged from the Step-up transition data set	8.0
continued on next page		

Table *anchor6* continued

Variables affected	Description of changes	Changed in release
mcob, fcob, mnat1, mnat2, fnat1, fnat2	Additional information on the Step-ups' parents was merged from their parents anchor data	8.0
ethni, migstatus	Additional information on the Step-ups available because of additional information on country of birth	8.0
relstat, marstat, meetdur	Minor corrections for marstat, relstat and meetdur	8.0
meetdur, reldur, cohabdur	Minor corrections for meetdur, reldur and cohabdur	9.0
hpm	Correction of 1 case because of anchor/interviewer notes wave 9	9.0
pfcob	Correction of 1 case	9.0
pcasprim, pcassec, plfs	Correction of coding if no information available: "-3 Does not apply" set to "-7 Incomplete data" for anchor respondents with partner	10.0
reldur, marstat, mardur	Minor corrections for reldur, marstat and mardur	10.0
cpi	New baseline year 2015	11.0
ethni, pethni, migstatus, pmigstatus	Minor changes for ethni, pethni, migstatus and pmigstatus	11.0
mschool, fschool	Recoding of value 3 to 4 and value 4 to 3 to make coding consistent with school and pschool	11.0
relstat, marstat	Minor corrections for relstat and marstat	11.0
pcob	Correction of the partner's country of birth in 2 cases	11.0
parentidkX	Minor corrections of parentidkX	11.0
intedu	Correction of erroneous response categories	13.0
<i>Modified labels of generated variables</i>		
wavedist	Variable label added (English version)	14.1

Table A.13: Changes in data set *anchor7*

Variables affected	Description of changes	Changed in release
<i>New generated variables</i>		
pmcob, pfcob	New variables (Country of birth of the partner's parents)	8.0
pethni, pmigstatus	New variables (Partner's ethnicity and migration status)	9.0
<i>Modified main variables</i>		
ehc12kX	Correction to 10 (other person) for some cases because other parent is not part of EHC anymore	8.0
ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pX, ehc33pX	Deletion of current partner from household grid	9.0
sib19p3	Deletion of sibling 3 for one case because of anchor/interviewer notes wave 9	9.0
ehc31, ehc32	Recoding of value -4 to -2	9.0
crn45kX	Adaption of German value labels to a previous change in value ranges	10.0
d611, d612, d613, d614	Some cases were changed from "1 Mentioned" to "-1 Don't know"	11.0
sib19pX	Some cases were changed from "-4 Filter error/Incorrect entry" to "-3 Does not apply"	11.0
<i>Modified generated variables</i>		
ykagecapi, ykidcapi	Corrections due to modified CAPI-child identifier	8.0
vocat	Step-up anchor respondents set to "-3 Does not apply"	8.0
cob, nat1, nat2	Additional information was merged from the Step-up transition data set	8.0
mcob, fcob, mnat1, mnat2, fnat1, fnat2	Additional information on the Step-ups' parents was merged from their parents anchor data	8.0
ethni, migstatus	Additional information on the Step-ups available because of additional information on country of birth	8.0
relstat, marstat, meetdur, mardur	Minor corrections for different variables	8.0
meetdur, reldur, cohabdur	Minor corrections for meetdur, reldur and cohabdur	9.0
pfcob	Correction of 1 case	9.0
pcasprim, pcassec, plfs	Correction of coding if no information available: "-3 Does not apply" set to "-7 Incomplete data" for anchor respondents with partner	10.0

continued on next page

Table *anchor7* continued

Variables affected	Description of changes	Changed in release
ehc12k1, crn17k1, crn49k1	Minor corrections for 1 case	10.0
cpi	New baseline year 2015	11.0
ethni, pethni, migstatus, pmigstatus	Minor changes for ethni, pethni, migstatus and pmigstatus	11.0
mschool, fschool	Recoding of value 3 to 4 and value 4 to 3 to make coding consistent with school and pschool	11.0
mcs, pcs	Value was changed from “-7 Incomplete data” to “-77 Incomplete data” for mcs and pcs	11.0
relstat, marstat	Minor corrections for relstat and marstat	11.0
pcob, pmcob, pfcob	Correction for 1 case	11.0
parentidkX	Minor corrections of parentidkX	11.0
intedu	Correction of erroneous response categories	13.0
<i>Modified labels of generated variables</i>		
wavedist	Variable label added (English version)	14.1

Table A.14: Changes in data set *anchor8*

Variables affected	Description of changes	Changed in release
<i>New generated variables</i>		
pethni, pmigstatus	New variables (Partner’s ethnicity and migration status)	9.0
<i>Modified main variables</i>		
ehc22pXn, ehc23pX, ehc24pXm, ehc24pXy, ehc25pX, ehc33pX	Deletion of current partner from household grid	9.0
all corresponding child variables	1 deleted child	9.0
lsr11iX become lsr11iX_2	Renaming of variables due to differences in response categories between waves 7 & 8	9.0

continued on next page

Table *anchor8* continued

Variables affected	Description of changes	Changed in release
pa11, pa34iX, ehc4p1, ehc4p1mX	Correction of 1 case because of anchor/interviewer notes wave 9	9.0
ehc19i5, ehc19i5mX, ehc19i6, ehc19i6mX	Correction of 1 case because of anchor/interviewer notes wave 9	9.0
ehc19i14mX, ehc19i14, ehc19i3mX, ehc19i3, ehc19i19mX, ehc19i10, job22	Correction of 1 case because of anchor/interviewer notes wave 9	9.0
ehc31	Recoding of value -4 to -2	9.0
crn45kX	Adaption of German value labels to a previous change in value ranges	10.0
fsk3-fsk5	Unnecessary information, variables dropped	10.0
<i>Modified generated variables</i>		
meetdur, reldur	Minor corrections for meetdur and reldur	9.0
nat1, nat2	Minor corrections for Step-up respondents	9.0
pfcob	Correction of one case	9.0
hpm	Correction of 1 case because of anchor/interviewer notes wave 9	9.0
pcasprim, pcassec, plfs	Correction of coding if no information available: “-3 Does not apply” set to “-7 Incomplete data” for anchor respondents with partner	10.0
relstat, marstat, mardur, homosex_new	Minor corrections for different variables	10.0
cpi	New baseline year 2015	11.0
ethni, pethni, migstatus, pmigstatus	Minor changes for ethni, pethni, migstatus and pmigstatus	11.0
mschool, fschool	Recoding of value 3 to 4 and value 4 to 3 to make coding consistent with school and pschool	11.0
relstat, marstat	Minor corrections for relstat and marstat	11.0
parentidkX	Minor corrections of parentidkX	11.0
intedu	Correction of erroneous response categories	13.0
<i>Modified labels of generated variables</i>		
wavedist	Variable label added (English version)	14.1

Table A.15: Changes in data set *anchor9*

Variables affected	Description of changes	Changed in release
<i>Modified main variables</i>		
pa3	Corrections due to a coding error; Changes also affect the data set <i>stepup_anchor9+transition</i>	9.1
mig1i12o	Correction of 1 case in data set <i>stepup_anchor9+transition</i>	9.1
crn45kX	Adaption of German value labels to a previous change in value ranges	10.0
d611, d612, d613, d614	Some cases were changed from "1 Mentioned" to "-1 Don't know"	11.0
sib19pX	Some cases were changed from "-4 Filter error/Incorrect entry" to "-3 Does not apply"	11.0
<i>Modified generated variables</i>		
pcasprim, pcassec, plfs	Correction of coding if no information available: "-3 Does not apply" set to "-7 Incomplete data" for anchor respondents with partner	10.0
mardur	Minor correction for 1 case	10.0
cpi	New baseline year 2015	11.0
ethni, pethni, migstatus, pmigstatus	Minor changes for ethni, pethni, migstatus and pmigstatus	11.0
mschool, fschool	Recoding of value 3 to 4 and value 4 to 3 to make coding consistent with school and pschool	11.0
mcs, pcs	Value was changed from "-7 Incomplete data" to "-77 Incomplete data" for mcs and pcs	11.0
infertile	Correction for one case	11.0
relstat, marstat	Minor corrections for relstat and marstat	11.0
parentidkX	Minor corrections of parentidkX	11.0
intedu	Correction of erroneous response categories	13.0
<i>Modified labels of generated variables</i>		
wavedist	Variable label added (English version)	14.1

Table A.16: Changes in data set *anchor10*

Variables affected	Description of changes	Changed in release
<i>Modified generated variables</i>		
cpi	New baseline year 2015	11.0
ethni, pethni, migstatus, pmigstatus	Minor changes for ethni, pethni, migstatus and pmigstatus	11.0
mschool, fschool	Recoding of value 3 to 4 and value 4 to 3 to make coding consistent with school and pschool	11.0
nkids, nkidsalv	Correction for one case	11.0
relstat, marstat	Minor corrections for relstat and marstat	11.0
meetdur, reldur	Minor corrections for meetdur and reldur	11.0
np, ncoh, nmar	Minor corrections for np, ncoh and nmar	11.0
parentidkX	Minor corrections of parentidkX	11.0
intedu	Correction of erroneous response categories	13.0
<i>Modified labels of generated variables</i>		
wavedist	Variable label added (English version)	14.1

Table A.17: Changes in data set *anchor11*

Variables affected	Description of changes	Changed in release
<i>Modified generated variables</i>		
intedu	Correction of erroneous response categories	13.0
<i>Modified labels of generated variables</i>		
wavedist	Variable label added (English version)	14.1
<i>Modified labels of main variables</i>		
sdp31iX, sdp37iX	Value labels modified (English version)	14.1

Table A.18: Changes in data set *anchor12*

Variables affected	Description of changes	Changed in release
<i>Modified main variables</i>		
pid, hpmt, hp0nr, hp0n, hp0g, hpbm, hpby	Correction of information on partnership for id=814196000,	14.0
<i>Modified generated variables</i>		
intedu	Correction of erroneous response categories	13.0
np, meetdur	Modifications for 2 cases due to correction of information on partnership	14.0
<i>Modified labels of generated variables</i>		
wavedist	Variable label added (English version)	14.1
<i>Modified labels of main variables</i>		
sdp31iX, sdp37iX	Value labels modified (English version)	14.1

Table A.19: Changes in data set *anchor13*

Variables affected	Description of changes	Changed in release
<i>Modified generated variables</i>		
pnat1, pnat2, amnat1, afnat1, smnat1, smnat2, sfnat1, sfnat2	Correction of incorrectly generated nationality of anchor, partner, adoptive mother/father and step mother/father for few cases	14.0
np, meetdur	Minor corrections for meetdur and reldur	14.0
<i>Modified labels of generated variables</i>		
wavedist	Variable label added (English version)	14.1
<i>Modified labels of main variables</i>		
sdp31iX, sdp37iX	Value labels modified (English version)	14.1

Table A.20: Changes in data set *anchor14*

Variables affected	Description of changes	Changed in release
continued on next page		

Table *anchor13* continued

Variables affected	Description of changes	Changed in release
<i>Modified main variables</i>		
int23	Added to PAPI data, was deleted erroneously	14.1
<i>Modified generated variables</i>		
smcob, sfcob	Both variables dropped, because both are empty	14.1
intsex, intage, intedu, intid	Dropped for CAPI and PAPI data, because interview without interviewer	14.1
<i>Modified labels of generated variables</i>		
meetdur, reldur, cohabdur, mardur, intage, wavedist, k*age	Minor variable and value label corrections (German and English version)	14.1
<i>Modified labels of main variables</i>		
bpa*, bjob*, sdp*, crn*, inc*, etc.	Minor variable and value label corrections (German and English version)	14.1

Changes in *partner* data sets

Table A.21: Changes in *partner* data sets

Variables affected	Description of changes	Changed in release
Partner 1		
<i>Modified labels</i>		
various variables throughout the data set	Some new English variable and value labels (translation edited)	2.0
ppa16i1-6, ppa22pi1-8, ppa22ri1-8, pftr1	Variable labels modified (German and English version)	2.0
<i>Modified values of main variables</i>		
psd10	Recoding of values to avoid discontinuity compared to later waves	3.1
psat1*, psat3	Recoding of missing values: “-1” was mistakenly included in category “-2”	3.1
Partner 2		
<i>Changes in case number</i>		
id=307529000	Deletion of one case because partner falsely answered partner survey (no current partner in <i>anchor2</i> data set)	10.0
Partner 3		
<i>Modified main variables</i>		
psd100i*, psd101i*	Minor corrections in assigning of missing values	5.0
pid	Correction of pid (id=67350100, 100983100, 369237100, 704733100)	7.0
<i>Changes in case number</i>		
id=918071000	Deletion of one case because ex-partner falsely answered partner survey (no current partner in <i>anchor3</i> data set)	6.0
<i>Modified labels of main variables</i>		
pdobd	English value label added	5.0
Partner 4		
<i>Modified main variables</i>		
pid	Correction of pid (id=97393100, 157821100, 704733100, 812026100, 848665100, 875056100, 899491100, 907272100, 908049100, 918071100)	7.0

continued on next page

Table *partner* continued

Variables affected	Description of changes	Changed in release
<i>Modified labels of main variables</i>		
pdobd	English value label added	5.0
<i>Changes in case number</i>		
id=864337000	Deletion of one case because ex-partner falsely answered partner survey (no current partner in <i>anchor4</i> data set)	7.0
Partner 5		
<i>Modified main variables</i>		
pid	Correction of pid (id=67350100, 97393100, 157821100, 369237100, 704733100, 864337100)	7.0
Partner 6		
<i>Modified main variables</i>		
pid	Correction of pid (id=97393100, 704733100, 864337100)	7.0
Partner 7		
<i>Modified main variables</i>		
pid	Correction of pid (id=293598100)	8.0
Partner 8		
<i>Modified generated variables</i>		
pinty	Corrections of interview year for some cases	9.1
Partner 9		
<i>Modified generated variables</i>		
pinty	Corrections of interview year for some cases; Changes also affect the data set <i>stepup_partner9</i>	9.1
Partner 11		
<i>Modified answer categories</i>		
phlt10	Adaptation of answer categories because phlt8 is now only answered by new partners in every wave <i>stepup_partner11</i>	11.0
Partner 12		
<i>Modified main variables</i>		
continued on next page		

Table *partner* continued

Variables affected	Description of changes	Changed in release
pcrn21i*, pcrn20i*, pcrn11i*, pcrn32i*	Incorrect filter correction fixed, overwritten values restored	14.0

Changes in *parent* data sets

Table A.22: Changes in *parent* data sets

Variables affected	Description of changes	Changed in release
Parent 2		
<i>New main variables</i>		
sibid1, sibid2, sibid3	New unique identifier of anchor's siblings 1 to 3	4.0
<i>Modified main variables</i>		
cid	Missing values set to system missing (.) instead of -3	4.0
parsd14k1g-parsd14k4g, parsd15k1- parsd15k4, parsd19k1y-parsd19k4y, parsd33k2-parsd33k4, parsd34k2- parsd34k4, parsd35k2-parsd35k4, parsd36k2-parsd36k4, parigr39k2- parigr39k4, parigr40k2-parigr40k4, parigr41k2-parigr41k4, parflag1-parflag5, parkids	Some values modified due to new order of child-information (child 1 to 4) reported by parents	4.0
<i>Modified generated variables</i>		
pargkpol, parbik, parbula	English variable label added	3.0
parschool, parvocat, parcasmin, parisced, paryeduc	Variables based on information from waves 2 to 3 of parent data and wave 3 of anchor data instead of wave-specific information from parent data; only for biological and adoptive parents	4.0
sibid1, sibid2, sibid3	Correction of sibid for 497 siblings	5.0
sibid4, sibid5, sibid6, sibid7, sibid8	New variables for identified siblings	5.0
parigr29, parigr34	Coding of wave 2 variables were harmonized with upcoming waves: (5-1) (6-4) (7-5) (1-6) (2-7) (3-8) (4-9)	7.0
parigr28	Correction of number of married couples in wave 2	7.0
kXtype	Minor corrections	7.0
continued on next page		

Table *parent* continued

Variables affected	Description of changes	Changed in release
<i>Modified labels of main variables</i>		
parigr85a, pargc12, parigr103, parinc23	Value labels modified (English version)	3.0
parigr28	Minor correction	7.0
<i>Modified labels of generated variables</i>		
parbula, parbik, pargkpo, parage, parage2, parpage, parpage2, parmage, parmage2, parfage, parfage2, parcasprim, parcassec, parflag_cas, parhhincnet, parhhsizemrd, parhhincgcee, park1type, park2type, park3type, park4type, parlfs, parmarstat, parrelstat, parigr27, parigr28	Variable labels slightly modified (German and English version)	4.0
Parent 3		
<i>Modified main variables</i>		
cid, parcorgc	Correction of cids pointing to focal child which have been re-identified from previous waves (no. of changes: 37); Changes in cid affected also parcorgc	9.0
<i>Modified generated variables</i>		
parschool, parvocat, parcasmin, parisced, paryeduc	Variables based on information from waves 2 to 3 of parent data and wave 3 of anchor data instead of wave-specific information from parent data; only for biological and adoptive parents	4.0
<i>Modified filter of main variables</i>		
parigr40p1, parigr40p3, parigr41p1, parigr41p3	Variables asked irrespective of existing contact to parent (German and English version)	3.1
<i>Modified labels of main variables</i>		
parval1ix	Assignment of variable labels to variables adjusted (German and English version)	3.1
parigr45a, parigr47a, parsd23i13, parsd23i21	Variable labels slightly modified (German and English version)	3.1
parigr53a-parigr59a, parigr78a-parigr79a, parigr80a-parigr81a, parigr60a-parigr66a	Change in wording of value label from “Trifft nicht zu” to “Kein Bedarf” (German version)	3.1
continued on next page		

Table *parent* continued

Variables affected	Description of changes	Changed in release
parigr103	Value labels slightly modified (German and English version)	3.1
<i>Modified values of main variables</i>		
parigr53a-parigr59a, parigr78a-parigr79a, parigr80a-parigr81a, parigr60a-parigr66a	Change of response code: 6 is set to 8	3.1
parpa17i1-parpa17i8	New category “no partner” added	3.1
parsd29i1-parsd29i8	Variables parsd29i1-parsd29i8 set to 0 and 1 instead of -2 due to corrections by TNS Infratest	4.0
parsd37, parsd38, parsd17i1, parsd17i2, parigr9y	Set to -3 instead of -4	4.0
parigr41p1, parigr41p3	Categories “no answer” and “does not apply” are now (release 5.0) differentiated	5.0
Parent 4		
<i>Modified main variables</i>		
parid	Correction of parid (id=127889301, 127889302, 127889304, 127889305)	7.0
cid, parcorgc	Correction of cids pointing to focal childs which have been re-identified from previous waves (no. of changes: 48); Changes in cid affected also parcorgc	9.0
Parent 5		
<i>Modified main variables</i>		
cid	Correction of cid in correspondence to anchor data (cid=718933203)	7.0
cid, parcorgc	Correction of cids pointing to focal childs which have been re-identified from previous waves (no. of changes: 81); Changes in cid affected also parcorgc	9.0
Parent 6		
<i>Modified main variables</i>		
parid	Correction of parid (id=12482301, 12482302, 12482303, 170670403, 506324301, 506324302, 506324303)	7.0
cid, parcorgc	Correction of cids pointing to focal childs which have been re-identified from previous waves (no. of changes: 115); Changes in cid affected also parcorgc	9.0
Parent 7		
<i>Modified main variables</i>		
parid	Correction of parid (id=12482301, 12482302, 12482303, 170670403, 506324301, 506324302, 506324303)	7.0

continued on next page

Table *parent* continued

Variables affected	Description of changes	Changed in release
cid, parcorgc	Correction of cids pointing to focal childs which have been re-identified from previous waves (no. of changes: 137); Changes in cid affected also parcorgc	9.0
Parent 8		
<i>Modified main variables</i>		
parid	Correction of parid (id=12482301, 12482302, 12482303, 170670403, 506324301, 506324302, 506324303)	7.0
cid, parcorgc	Correction of cids pointing to focal childs which have been re-identified from previous waves (no. of changes: 103); Changes in cid affected also parcorgc	9.0

Changes in *parenting* data sets

Table A.23: Changes in *parenting* data sets

Variables affected	Description of changes	Changed in release
Parenting 2		
<i>Discontinued variables</i>		
anchor, partner	Variables deleted	3.0
<i>Modified labels of main variables</i>		
all variables of this data set	Correction of value labels for missing values (German and English version; only in Stata)	3.1
pcr3i1,...,pcr3i8	Value labels modified (German version)	3.1
cid	Variable label slightly modified (English Version)	4.0
<i>Modified values of main variables</i>		
pid	Replacing 6 missing person numbers partner	3.1
sdqpi1,...,sdqpi25	Set to 0, 1 and 2 instead of 1, 2 and 3 in all parenting data sets due to data harmonization with the child data set	5.0
<i>Changes in case number</i>		
(id=494001000, id=609745000)	Two empty cases deleted from data set	3.1
(id=494001000, id=609745000)	Two empty cases deleted from data set	4.0
Parenting 3		
<i>Modified labels of main variables</i>		
pcr3i1,...,pcr3i8	Value labels modified (German and English version)	3.1
cid	Variable label slightly modified (English version)	4.0
<i>Modified values of main variables</i>		
sdqpi1,...,sdqpi25	Set to 0, 1 and 2 instead of 1, 2 and 3 in all parenting data sets due to data harmonization with the child data set	5.0
Parenting 4		
<i>Modified values of main variables</i>		
continued on next page		

Table *parent* continued

Variables affected	Description of changes	Changed in release
sdqpi1,...,sdqpi25	Set to 0, 1 and 2 instead of 1, 2 and 3 in all parenting data sets due to data harmonization with the child data set	5.0
Parenting 7		
<i>Modified main variables</i>		
edu10i9	Variable name changed from edu10i9 to edu10i7	9.0
edu7_, edu9_, edu10_	Correction in assigning of missing and inconsistent values	9.0
<i>New generated variables</i>		
flag_edu7i1,...,flag_edu7i5	New variables (Flag variables edu7_)	9.0
flag_edu10i1,...,flag_edu10i7	New variables (Flag variables edu10_)	9.0
Parenting 8		
<i>Modified main variables</i>		
edu5_, edu11, edu12	Correction in assigning of missing and inconsistent values	9.0
<i>New generated variables</i>		
flag_edu5i1,...,flag_edu5i4	New variables (Flag variables edu5_)	9.0
flag_edu11	New variable (Flag variable edu11)	9.0
flag_edu12	New variable (Flag variable edu12)	9.0
Parenting 9		
<i>New generated variables</i>		
flag_edu13i1,...,flag_edu13i10	New variables (Flag variables edu13_)	9.0

Changes in *child* data sets

Table A.24: Changes in *child* data sets

Variables affected	Description of changes	Changed in release
child 2		
<i>Modified generated variables</i>		
relstatac	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'	3.0
<i>Modified variable names of main variables</i>		
csdq1i1-csdq1i3, csdq1i5-csdq1i20	Harmonization of SDQ item numbers with SDQ item numbers in parenting data set	6.0
<i>Modified labels</i>		
all variables of this data set	The value was added to the value label	3.0
<i>Modified labels of main variables</i>		
cedu1, cedu5i4, cpcr8i8, clsr1i1, clsr1i14	Variable labels modified (German version)	3.1
cedu8i2, cpcr8i1, cgp1, l391s	Variable labels modified (English version)	3.1
all variables of this data set	Correction of value labels for missing values (German and English version; only in Stata)	3.1
csex, cdobm, cedu1, l391s	Value labels modified (German version)	3.1
cpcr5, cdobm, l391s	Value labels modified (English version)	3.1
id, pid, cdoibd, cedu1o, cedi1ao, cpcr4, cpcr5, cpcr13, cgp1, cgp2	Variable labels slightly modified (German and English version)	4.0
l391s, sex, cdoibd, cedu1o, cedi1ao, cpcr4, cpcr5, cpcr13, cgp1, cgp2	Value labels slightly modified (German and English version)	4.0
child 3		
<i>Modified variable names of main variables</i>		
csdq1i1-csdq1i3, csdq1i5-csdq1i20	Harmonization of SDQ item numbers with SDQ item numbers in parenting data set	6.0
continued on next page		

Table *child* continued

Variables affected	Description of changes	Changed in release
<i>Modified labels of main variables</i>		
cedu7i1, csdq1i5, csdq1i10	Variable labels slightly modified (German version)	3.1
cedu1ao, cedu13i4, cpcr4, cpcr5, cpcr13, cint1i1, cint1i2, cint1i3, cint1i4, cint1i5, cint2, l391s	Variable labels modified (English version)	3.1
id, pid, cdobd, cedu1o, cedi1ao, cpcr4, cpcr5, cpcr13, cgp1, cgp2	Variable labels slightly modified (German and English version)	4.0
l391s, sex, cdobd, cedu1o, cedi1ao, cpcr4, cpcr5, cpcr13, cgp1, cgp2	Value labels slightly modified (German and English version)	4.0
<i>Modified values of main variables</i>		
tag_csex, tag_cdobm, tag_cdoby	Recoding: 1 to -3 if no inconsistency possible since child new in wave 3	3.1
child 4		
<i>Modified variable names of main variables</i>		
csdq1i1-csdq1i3, csdq1i5-csdq1i20	Harmonization of SDQ item numbers with SDQ item numbers in parenting data set	6.0
child 5		
<i>Modified variable names of main variables</i>		
csdq1i1-csdq1i3, csdq1i5-csdq1i20	Harmonization of SDQ item numbers with SDQ item numbers in parenting data set	6.0
<i>Modified labels of main variables</i>		
crom9	English variable label slightly modified	6.0

Changes in data set *anchor1_DD*

Table A.25: Changes in *anchor1_DD*

Variables affected	Description of changes	Changed in release
<i>New generated variables</i>		
iscd2	New variable (ISCED classification of educational attainment, incl. students)	6.0
<i>Modified generated variables</i>		
school, pschool, casmin, pcasmin, isced, pised, isced2, pised2, yeduc, pyeduc	Variables school & pschool changed due to corrections in anchor1_DD and some recodings (some values 3 recoded to 4) plus recoding of implausible values for new partners: "3 lower GDR, POS 8./9." to "2 lower, Volks-/Hauptschulabschluss" and "5 intermediate GDR, POS 10." to "4 intermediate, Realschulabschluss / mittlere Reife" since partners born after 1980 (POS 8./9.) or 1979 (POS 10.) could not have earned a degree from a polytechnic secondary school (only existed in the GDR); changes in other variables due to changes of school & pschool	6.0
pnkidsbioalv	Correction of coding variable	6.0
mschool, fschool	Correction of value labels: "3 lower GDR, POS 8./9." to "3 intermediate, Realschulabschluss / mittlere Reife" and "4 intermediate, Realschulabschluss / mittlere Reife" to "4 lower GDR, POS 8./9."	7.0
iscd, iscd2, pised	Correction of value labels: "4 upper secondary education general (3b)" to "4 upper secondary education vocational (3b)" and "5 upper secondary education vocational (3a)" to "5 upper secondary education general (3a)"	7.0
npo14mr, npu14mr, hhincoecd	Correction of number of persons aged under 14/ aged 14 and older in main residence	9.0
siops	Correction of assignment of ISCO scores: Variable was erroneously based on ISCO-88 instead of ISCO-08	9.1
pcasprim, pcassec, plfs	Correction of coding if no information available: "-3 Does not apply" set to "-7 Incomplete data" for anchor respondents with partner	9.1
continued on next page		

Table *demodiff* continued

Variables affected	Description of changes	Changed in release
cp	New baseline year 2015	11.0
mschool, fschool	Recoding of value 3 to 4 and value 4 to 3 to make coding consistent with school and pschool	11.0
pmarstat	Correction of coding pmarstat	11.0
<i>Modified tag and flag variables</i>		
flag27	Variable added to achieve consistency across waves (-3 for all cases)	7.0

A.3 Flag and tag variables

Table A.26: 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
flag3	Inconsistency biological child and household grid	-3	Does not apply	Inconsistency irrelevant in this wave
		0	No inconsistency	
		11 / 12	Name child in hh1/2 unequal to sd14kxn	Name child according to household grid unequal to name child reported before
		21 / 22	Additional child(ren) in hh1/2	More children in household according to household grid than reported before

continued on next page

continued from previous page

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

continued on next page

continued from previous page

Variable	Label	Values	Value Labels	Description
		1	Inconsistency	End previous cohabitation episode with previous partner after beginning of current cohabitation episode with current partner
flag8	Inconsistency beginning current and end previous marriage (different partners)	0	No inconsistency	
		"x"	Inconsistency partner x	End last marriage before beginning current marriage
flag9	Inconsistency divorced/widowed and no partner before current relationship	0	No inconsistency	
		1	Inconsistency	Current marital status divorced/separated civil union or widowed and "no partner before current relationship/never had partner"
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag10	Inconsistency divorce from a partner to whom never married	0	No inconsistency	
		"x"	Inconsistency partner x	End of previous relationship through divorce, but never married to this partner
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag11	Inconsistency separation through death/divorce current spouse	0	No inconsistency	
		"x"	Inconsistency partner x	Divorced from current spouse or current spouse dead
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag12	Inconsistency year of birth current partner	0	No inconsistency	

continued on next page

continued from previous page

Variable	Label	Values	Value Labels	Description
		1	Younger than 10 years old	
		2	Year of birth after beginning of relationship	
		3	Younger than 12 years old at birth of biological child	<i>see value labels</i>
		4	Younger than 14 years old and completed vocat. training/univ. degree	
flag13	Inconsistency year of birth parents	0	No inconsistency	
		1	Inconsistency year of birth mother	Age difference mother and anchor less than 12 years
		2	Inconsistency year of birth father	Age difference father and anchor less than 12 years
		3	Inconsistency year of birth mother and father	Age differences mother and anchor, and father and anchor less than 12 years
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag14	Inconsistency separation before beginning of relationship (current partner)	0	No inconsistency	
		1	Inconsistency	Separation episode(s) with current partner before first beginning of relationship with current partner
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag15	Inconsistency year of birth of child and anchor	0	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	

continued on next page

continued from previous page

Variable	Label	Values	Value Labels	Description
		1	Inconsistency	Net income larger than gross income
		-3	Does not apply	
flag17	Inconsistency personal net and household income	0	No inconsistency	
		1	Inconsistency	Net personal income larger than net household income
flag18	Inconsistency sex of partner	0	No inconsistency	
		1	Inconsistency, male partner	Partner female according to anchor, male according to partner (see variable <i>psex</i> partner data)
		2	Inconsistency, female partner	Partner male according to anchor, female according to partner (see variable <i>psex</i> partner data)
flag20	Inconsistency beginning of relationship before getting to know (current partner)	0	No inconsistency	
		1	Inconsistency	First month of relationship earlier than date of getting to know each other
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag21	Inconsistency cohabitation before getting to know each other (current partner)	0	No inconsistency	
		1	Inconsistency	First month of cohabitation earlier than date of getting to know each other
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag22	Inconsistency marriage before getting to know each other (current partner)	0	No inconsistency	

continued on next page

continued from previous page

Variable	Label	Values	Value Labels	Description
flag23	Inconsistency home size & number of rooms	1	Inconsistency	First month of marriage earlier than date of getting to know each other
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
		0	No inconsistency	
		1	Inconsistency HH1	Home size \leq 10qm and number of rooms $>$ 2 or Home size \leq 20qm and number of rooms $>$ 4
flag24	Inconsistency overnight stays in main and second residence	-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
		0	No inconsistency	
		1	Inconsistency	Stay in residence one and two almost every night
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag25	Inconsistency biological parents in household who should be dead (acc. preload)	0	No inconsistency	
		1	HH1: biological mother is dead	Person mentioned although this person is dead (acc. to preload)
		2	HH1: biological father is dead	
		3	HH2: biological mother is dead	
		4	HH2: biological father is dead	
		-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
flag26	Relationship between two anchor persons	0	No inconsistency	
		"x"	Couple "x"	Two anchor persons have a relationship
flag27	Age sibling at birth of sibling's child	0	No inconsistency	
		1	Inconsistency	Sibling of anchor younger than 12 years old at birth of sibling's biolog. child

continued on next page

continued from previous page

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

continued on next page

continued from previous page

Variable	Label	Values	Value Labels	Description
flag_frt	Survey method for questions frt7/8/9/16 (filter error)	-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
		0	General questionnaire	
		1	Follow-up questionnaire	Indicates if information on frt7/8/9/16 was collected via follow-up survey (PAPI); due to a filter error in the main questionnaire
flag_igb	Inconsistent combination: biological and adoptive parents	-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
		0	No inconsistency	
		1	Inconsistency	Having both biological and adoptive parents inconsistent according to codebook
flag_isco88 _kldb1992	Coding problem kldb and isco	-3	Does not apply	Inconsistency not possible in this wave; flag irrelevant
		0	No problem	
		1	Qualification unclear, lowest coded	Classification of occupations problematic (<i>also see chapter 4.3</i>)
flag_isco08 _kldb2010	Coding problem kldb and isco	2	Activity not unequivocally assignable	
		-7	Incomplete data	
		-3	Does not apply	
		0	No problem	
		1	Qualification unclear, lowest coded	Classification of occupations problematic (<i>also see chapter 4.3</i>)
		2	Activity not unequivocally assignable	

Table A.27: List of tag variables to identify inconsistencies with preloads (*anchor\$*)

Variable	Label	Values	Value Labels	Description
tag_sex	Inconsistency sex	0	No inconsistency	Anchor's sex in current wave is not anchor's sex in preload
	anchor between waves	1	Inconsistency	
tag_dob	Inconsistency date of birth anchor between waves	0	No inconsistency	Anchor's date of birth is not anchor's date of birth in preload
		1	Inconsistency: both month & year	
		2	Inconsistency: month	
tag_idenk"x"	Inconsistency identity child "x" between waves	3	Inconsistency: year	Child "x" in previous wave is not the same child as in current wave
		0	No inconsistency	
tag_sexk"x"	Inconsistency sex child "x" between waves	1	Inconsistency	Child "x" changed sex Child deleted
		0	No inconsistency	
		2	Inconsistency: sex child	
tag_dobk"x"	Inconsistency date of birth child "x" between waves	0	No inconsistency	Year of birth previous wave is not equal to year of birth current wave Month/range of birth previous wave is not equal to month/range of birth current wave
		1	Inconsistency	
tag_biok"x"	Inconsistency status child "x" in anchor data between waves	0	No inconsistency	Status of child "x" is not equal to status of child "x" in preload Child "x" deleted
		1	Inconsistency: status child	
		2	Child deleted	
tag_biokp"x"	Inconsistency biological parent partner child "x" between waves	0	No inconsistency	Status partner (biological parent) varies between waves
		1	Inconsistency: biological parent partner	
tag_biokp"x"	Inconsistency biological parent partner child "x" between waves	2	Child deleted	Child "x" deleted

continued on next page

continued from previous page

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

Table A.28: List of flag variables to identify inconsistencies (*parent\$*)

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

continued on next page

continued from previous page

Variable	Label	Values	Value Labels	Description
parflag7	Number of grandchildren	0	No inconsistency	Coding from wave 2 through wave 4: pargc2i6=1 (no grandchildren) and at least (one valid value in pargc2i1-pargc2i5 or missing in pargc2i1- pargc2i5) for correction see variable grandparkids Wave 5: at least one missing in pargc2i1- pargc2i5
		1	Ambiguous # of grandkids	
parflag8	Parent's age	0	No inconsistency	<i>see value labels</i>
		1	Parent <12 years old at respondent's birth or	
		2	adoptive/stepparent Living parent 100+ years old	
parflag9	Support by non-existent people	0	No inconsistency	Mother not alive (parigr22 = 2) Father not alive (parigr25 = 2) Only one child (parsd32 = 1)
		1	Mother	
		2	Father	
		3	Children other than Anchor	No partner (parsd3 = 3) No siblings (parnet17i2 = 0 & parnet17i1 = 0) No children with partner (parsd35k1 & parsd35k2 & parsd35k3 & parsd35k4 ≠ 1) No grandchildren (pargc2i6 = 0)
		4	Partner	
		5	Siblings	
		6	Son-/daughter-in-law	
parflag10	Household composition	7	Grandchildren	(parhh = 0 & (parigr41k1 = 1 OR parigr41k2 = 1 OR parigr41k3 = 1 OR parigr41k4 = 1)) (parhh = 0 & parsd3 = 1)
		0	No inconsistency	
		1	Children in hh (question 3) but single-hh (question 47)	
parflag11	Leaving home	2	Partner in hh (question 48) but single-hh (question 47)	(parigr9y < pardoby)
		0	No inconsistency	
		1	Left home before birth	

continued on next page

continued from previous page

Variable	Label	Values	Value Labels	Description
parflag12	Educational degree	0	No inconsistency	<i>see value labels</i>
		1	Multiple answer: highest degree kept	
		2	Multiple answer incl. other	
parflag13	Occupational degree	0	No inconsistency	Any valid occupational degree provided & (parsd29i8 = 1)
		1	No but also valid occupational degree	
parflag14	Mother's partnership status	0	No inconsistency	papa17i* != igr30 (from Anchor's interview)
		1	Anchor's answer doesn't match mother's answer	
parflag15	Father's partnership status	0	No inconsistency	papa17i* != igr35 (from Anchor's interview)
		1	Anchor's answer doesn't match father's answer	
parflag_cas	Current activity	0	No inconsistency	Multiple answers for current activity status that are not plausible
		1019	Unemployed but full-time	
		1021	Early retirement but full-time	
		1219	Unemployed but part-time	
		1721	Maternal or paternal leave or other/Retired, occupational disability	
parflag_lfs1	Labor force status	0	No inconsistency	<i>see value labels</i>
		1	Employed but no hours of work men- tioned	

Table A.29: List of flag variables to identify inconsistencies (*parenting\$*)

Variable	Label	Values	Value Labels	Description
flag_edu5i1,...,flag_edu5i4				
flag_edu7i1,...,flag_edu7i5	Inconsistency child	0	No inconsistency	Filled in although anchor information (crn13kxi12)
flag_edu10i1,...,flag_edu10i7	not enrolled in school	1	Inconsistency	says not enrolled in school
flag_edu11	based on anchor			
flag_edu12	information			
flag_edu13i1,...,flag_edu13i10				

Table A.30: List of flag variables to identify inconsistencies (*parentingU6partner\$*)

Variable	Label	Values	Value Labels	Description
flag_pcrn43i1,...,flag_pcrn43i4	Inconsistency child			
flag_pcrn44i1,...,flag_pcrn44i10	younger than 3 years	0	No Inconsistency	Filled in although anchor information (kxdoby_gen)
flag_pcrn100i1,...,flag_pcrn100i3	old based on anchor	1	Inconsistency	says child is younger than 3 years
	information			

A.4 Main sample-refreshment sample differences

Table A.31: Variables not posed to refreshment sample in wave 11, anchor11

Module	Variables
Reason for move	hc32_
Educational and vocational attainment since previous wave	sd32_
Separation from ex-partner (CASI) - Separation Module	sep1_, sep2, sep3, sep4, sep12-16, sep17_, sep18-20
Pornography consumption	sex14-16
Child rearing goals	crn19_

Table A.32: Additional variables for refreshment sample in wave 11, anchor11

Module	Variables
Importance of different domains of life (future)	co1_
Partnership episodes, Cohabitation episodes, Marital status (Current partnership (status quo))	sd5ezbm/y, sd6ezem/y, sd6ez, sd7ez, sd8ezbm/y, sd8ezem/y, sd10, sd12m/y
Children (status quo) - for each Child x (x=1 to 10)	sd16-18kx, rtr16kxez, rtr17kxezbm/y, rtr17kxezem/y, rtr19kx, rtr20kxm/y, rtr21kx
Education and employment	sd27/28, sd29_
Retrospective Event History Calendar (Retro-EHC)	rtr1pxn/g, rtr2pxezbm/y, rtr2pxezem/y, rtr3pxez, rtr4px, rtr5, rtr6px, rtr7pxezbm/y, rtr7pxezem/y, rtr8pxez, rtr9, rtr10m/y, rtr11, rtr12, rtr13px, rtr14pxbm/y, rtr14pxem/y, rtr15px
Childhood history (until 18th birthday)	cla1-4, cla4o, cla5ex, cla6ex, cla6exo, cla7ex, cla8
Personality (CASI)	per3_, sdq1_
Risk Taking	rtr1_, rtr2_, rtr3_
Critical Life Events (CLE)	cle1_
Perception of parental parenting style (<21 years) (CASI)	par1_, par2_, par3_, par4_
Sexuality and contraception (CASI)	sex1_, sex20, sex21
Addictive behavior (CASI)	hlt8, hlt9
Origin Anchor (migration background)	mig1_, mig2m/y, mig3, mig4, mig4o, mig5m/y
Intergenerational relationships	igr1d/m/y, igr100, igr2d/m/y, igr101, igr21m/y, igr106, igr24m/y, igr107, igr3, igr3o, igr4, igr4o, igr74, igr74o, igr78, igr78o, mig6_, mig7_, igr73_, igr77_, igr87, igr92, igr93, igr93y, igr94, igr94y, igr95-97, igr97y, igr32, igr37

continued on next page

continued from previous page

Module	Variables
Education and occupation of parents	igr75, igr75o, igr76_, igr79, igr79o, igr80_, igr102-105
NRI, Ambivalence	igr43p_, igr44p_, igr46p_, igr48p_
Detailed sibling module	sib1, sib2px, sib3px, sib4pxy, sib5px, sib6pxy, sib7px, sib8px, sib10px, sib11px
Previous places of residence, school attendance, and occupations	rtr27m/y, rtr28hx, rtr29hxbm/y, rtr30hxem/y
Religiosity	sd30, sd31, sd36
Health and social background during own childhood	cla9_, cla10, cla11, cla13_
Health and life satisfaction	hlt5, hlt6
Interview assessment by the interviewer	int1-3, int17-22

A.5 Additional variables web-based anchor questionnaire W14

Table A.33: Additional variables for anchor respondents of the web version in wave 14

Variable	Label
job19	Same occupation as in previous wave
job21	Same occupational status as in previous wave
job30	Current occupational status (broad categories)
job31	Current occupational status: white-collar occupations
job32	Current occupational status: blue-collar occupations
job33	Current occupational status: public administration
Job34	Current occupational status: self-employed
Job35	Current occupational status: farmers
sdp27	Partner's biological siblings
sdp28	Partner's half or step siblings
sdp10i14o	Partner: Other type of non-employment, open entry
sdp29	Partner: previous marriage/partnership
sdp30	Partner: relationship status
sex3	Anchor expecting a child
sex4	Partner expecting a child
sex6_	Contraception: ...
ftr26	Realistic number of children (if no children so far)
ftr27	Realistic having additional children
crn17k_	Child x: How often does the other parent see the child
crn49k_	Child x: Travel time to other parent
crn65k_	Child x: How many nights does child x spend at your residence?
crn67k_	Child x: How many nights does child x spend at his/her other parent's residence
int11	How did you like the interview
int12	Suggestions or comments regarding the interview