pairfam
Panel Analysis of Intimate Relationships and Family Dynamics

# pairfam Quick Guide

## Release 10.0

Daniel Fuß,

Sabine Keller,

Claudia Schmiedeberg

May 2019

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

Quick Guide of the German Family Panel

This **Quick Guide** gives instructions for handling the complex data of the German Family Panel (pairfam). References to essential documents and information about the composition and structure of the data are intended to facilitate first use. In addition, this guide focuses on the merging of datasets, which is necessary for most analyses.

Words labeled with the ☞ symbol link to our ☞homepage where corresponding documents or further information have been made available. In addition to being included in the scientific use file, all documentation can be downloaded from the pairfam website (www.pairfam.de/en).

Our ☞newsletter gives advice on data handling and informs periodically about current developments, relevant events, and new publications. Please do not hesitate to contact pairfam's ☞user support (☞support@pairfam.de, 0049 89 - 2180 1296).

***Contents***



## *1 Quick Orientation*

The German Family Panel (pairfam) is a multi-disciplinary, longitudinal study for researching partnership and family dynamics in Germany funded by the German Research Foundation (DFG). The survey was first launched in 2008/09 to annually collect survey data from a nationwide, random sample of 12,402 anchor persons. Its "multi actor design" means that interviews are conducted also with the anchors' partners, parents (until wave 8), and children (called "alters"). Surveyed children are included in the anchor survey as new anchor persons themselves ("step-ups") once they reach the age of 16.

The ☞documentation helps in preparing and carrying out analyses:

☞Reference paper: Offers a comprehensive description of the conceptual framework and the design of the German Family Panel by Huinink et al. (2011).

☞Data Manual: Explains the structure of the datasets and provides a detailed description of the data-editing procedures undertaken; contains relevant notes on generated variables and datasets, paradata, the computation and use of weighting factors, data inconsistencies (flag and tag variables), and modifications between the releases.

☞Method reports: Documents the organization and realization of fieldwork ranging from sampling strategies, the use of incentives, and the follow-up concepts to increase response rates as well as procedures for increasing panel stability (in German only). Contents of the methods reports are summarized in ☞ Technical Paper # 01.

☞Technical Papers: Point the users to some important notes and assistances with regard to data preparation and data analysis.

With **Release 10.0**, published summer of 2019, pairfam data from the first ten survey waves were made accessible as a scientific use file for scholarly analyses. For obtaining an overview of the items and constructs of our survey program, you can draw on several resources:

☞Variable lists: Detailed overview of all variables per wave, thematically sorted, with names and labels for each respondent group, as well as notes on modifications between waves; separate descriptions for each respondent group.

☞Codebooks: Complete information on question texts of all available waves; representation of answering options, value labels, filter settings, and ordered tips for interviewers; separate descriptions for each wave and for each respondent group.

☞Scales Manual: Information regarding the creation of index variables and scales on the basis of instruments used in the German Family Panel; provides information regarding the sources of these index variables, their corresponding items, and reliability scores.

## *2 Advice for Data Users*

The processed data are regularly published as a scientific use file approximately one year after the end of the field phase. The period of approximately twelve months between waves means that the data are updated annually. At the moment, Release 10.0, including data from the first ten waves of pairfam is available.

In addition, the scientific use file includes data from the three-wave study ☞DemoDiff, a parallel survey in East Germany targeting anchor persons of the two birth cohorts 1971-73 and 1981-83 together with their partners. DemoDiff started with the first wave one year after the start of the pairfam survey, its third (and last) wave was conducted together with the fourth wave of pairfam. As of wave 5, DemoDiff-respondents were integrated into the pairfam sample, so that DemoDiff and pairfam were merged.

## 2.1 Datasets

Because information is stored separately by wave and respondent group, the scientific use file contains several datasets. The linkage of those datasets for longitudinal or dyadic analyses is possible through a standardized key variable (see Sections 0 and 2.4). DemoDiff data are partly included in the scientific use file as separate datasets (DemoDiff waves 1 and 2) and are partly merged with the pairfam datasets (DemoDiff wave 3), as shown in the following table.

| Respondents' datasets | Anchor | Partner | Child | Parent | Parenting | Parenting Adolescents |
|---|---|---|---|---|---|---|
| **Wave 1** **2008/09** **(DemoDiff)** | anchor1 *(N=12,402, V=1,450)* anchor1_DD *(N=1,489, V=1,445)* | partner1 *(N=3,743, V=232)* partner1_DD *(N=684, V=232)* | - | - | - | - |
| **Wave 2** **2009/10** **(DemoDiff)** | anchor2 *(N=9,069, V=3,888)* anchor2_DD *(N=1,173, V=18)* | partner2 *(N=2,687, V=273)* - | child2 *(N=862, V=129)* - | parent2 *(N=5,015, V=359)* - | parenting2 *(N=1,169, V=80)* - | - |
| **Wave 3** **2010/11** | anchor3 *(N=9,074, V=4,291)* | partner3 *(N=2,939, V=280)* | child3 *(N=987, V=148)* | parent3 *(N=3,946, V=220)* | parenting3 *(N=1,408, V=44)* | - |
| **Wave 4** **2011/12** | anchor4 *(N=8,073, V=4,130)* | partner4 *(N=2,731, V=264)* | child4 *(N=1,084, V=171)* | parent4 *(N=3,350, V=276)* | parenting4 *(N=1,576, V=55)* | - |
| **Wave 5** **2012/13** | anchor5 *(N=7,248, V=4,741)* | partner5 *(N=2,529, V=312)* | child5 *(N=1,390, V=188)* | parent5 *(N=3,546, V=242)* | parenting5 *(N=2,022, V=72)* | - |
| **Wave 6** **2013/14** | anchor6 *(N=6,574, V=4,351)* | partner6 *(N=2,357, V=290)* | child6 *(N=1,450, V=209)* | parent6 *(N=3,043, V=378)* | parenting6 *(N=2,251, V=63)* | - |
| **Wave 7** **2014/15** | anchor7 *(N=5,919, V=5,046)* | partner7 *(N=2,170, V=327)* | child7 *(N=1,438, V=221)* | parent7 *(N=2,719, V=264)* | parenting7 *(N=2,889, V=101)* | - |
| **Wave 8** **2015/16** | anchor8 *(N=5,461, V=4,623)* | partner8 *(N=2,051, V=303)* | child8 *(N=1,419, V=251)* | parent8 *(N=627, V=110)* | parenting8 *(N=2,847, V=72)* | - |
| **Wave 9** **2016/17** | anchor9 *(N=5,127, V=5,275)* | partner9 *(N=1,946, V=328)* | child9 *(N=1,401, V=247)* | - | parenting9 *(N=2,919, V=128)* | paya9 *(N=766, V=97)* |
| **Wave 10** **2017/18** | anchor10 *(N=4,750, V=4,670)* | partner10 *(N=1,799, V=283)* | child10 *(N=1,282, V=262)* | - | parenting10 *(N=2,721 V=64)* | paya10 *(N=897, V=92)* |

N = Number of cases, V=Number of Variables

Furthermore, there are several generated datasets with biographical information available across all waves (☞Data Manual, Section 4.6).

| Generated Datasets | Filename | Content |
|---|---|---|
| Biography of activities | bioact<br>bioact_rtr | prospective and retrospective information on anchor's educational and occupational activities |
| Birth biography | biochild | prospective and retrospective information on birth and cohabitation of all anchor's children |
| Biography of mobility | biomob_ | prospective and retrospective information on anchor's mobility (residences, migration history, moving out of the parental house) |
| Biography of parents | bioparent | prospective and retrospective information on anchor's biological, adoptive, and stepparents |
| Biography of relationship | biopart | prospective and retrospective information on anchor's partnerships after the age of 14 (relationships, cohabitation, and marriage history) |
| Household patterns | household | prospective information on anchor's household (waves 1-3): dwelling and household members |
| Multi-actor overview | overview_multi_actor | prospective information on participation of the anchor's alters in the survey |

Data concerning step-up anchor respondents is stored separately. The data structure is illustrated in the following table.

| Step-up datasets | Anchor | Partner |
|---|---|---|
| **Wave 4**<br>**2011/12** | stepup_anchor4<br>*(N=50, V=4,131)*<br>stepup_transition_anchor4<br>*(N=43, V=82)* | stepup_partner4<br>*(N=5, V=265)*<br>- |
| **Wave 5**<br>**2012/13** | stepup_anchor5<br>*(N=82, V=4,742)*<br>stepup_transition_anchor6<br>*(N=32, V=75)* | stepup_partner5<br>*(N=8, V=313)*<br>- |
| **Wave 6**<br>**2013/14** | stepup_anchor6<br>*(N=136, V=4,352)*<br>stepup_transition_anchor6<br>*(N=51, V=82)* | stepup_partner6<br>*(N=10, V=291)*<br>- |
| **Wave 7**<br>**2014/15** | stepup_anchor7<br>*(N=190, V=5,117)*<br>stepup_transition_anchor7<br>*(N=47, V=82)* | stepup_partner7<br>*(N=15, V=328)*<br>- |
| **Wave 8**<br>**2015/16** | stepup_anchor8<br>*(N=240, V=4,712)*<br>stepup_transition_anchor8<br>*(N=57, V=82)* | stepup_partner8<br>*(N=19, V=304)*<br>- |
| **Wave 9**<br>**2016/17** | stepup_anchor9+transition.dta<br>*(N=320, V=5,421)* | *stepup_partner9*<br>*(N=37, V=329)* |
| **Wave 10**<br>**2017/18** | stepup_anchor10+transition.dta<br>*(N=410, V=4,834)* | stepup_partner10<br>*(N=47, V=284)* |
| **generated datasets** | stepup_biochild | prospective and retrospective information on birth and cohabitation of all anchor's children |
| | stepup_biopart | prospective and retrospective information on anchor's partnerships after the age of 14 |

## 2.2 Variables

With the intention to make the data as easy to comprehend as possible, the **variable names** (☞Data Manual, Section 2.3) generally follow a logical pattern. Variables that are identical in content and queried in several waves share a common name. Variables in the alters' datasets are distinguished by an appropriate prefix:

- *Variable name sex: anchor = sex / partner = **p**sex / parents = **par**sex / children = **c**sex*

- *Prefix for data from the paya questionnaire is paya*

The **key variables** are integral for working with pairfam data (☞Data Manual, Section 2.2) as they enable an unambiguous allocation of relevant information to the appropriate respondents. Identifiers are indispensable for the aggregation of datasets. Variables are named according to the following pattern:

- *ID-Variables: anchor= id | partner = pid | parents = parid | children= cid(x) | siblings = sibid(x)*
  *mother= mid | father = fid | stepmother = smid | stepfather = sfid*

In addition to the survey data, datasets contain several additional **variables** that are **generated** during data preparation. These variables (☞Data Manual, Section 4.2) capture information often needed by users:

- *examples for generated variables anchor: Birth cohort = cohort*
  *Migration background = ethni, migstatus*
  *Family status, duration of relationship = marstat, reldur*
  *Number of children = nkids*
  *Education = isced, casmin, yeduc*
  *Class schema = egp, isei, kldb2010, isco08, mps*
  *Income = hhincgcee, incnet, hhincnet*

To take the disproportionate stratified sample and systematic nonresponse into account, **weights** are provided (☞Data Manual, Section 4.5) in the anchor datasets. The usage of these weighting factors depends on the type and intention of the analysis being undertaken by the user. In Release 5.0 we provided new weights for merged pairfam and DemoDiff data.

| Weighting variable | Sample | Description | Application |
|---|---|---|---|
| dweight | *pairfam* | Design weight to correct for disproportionate sampling across cohorts | Stata:<br>svyset [pweight=weightvar]<br>svy: command<br>command [pweight=weightvar], options<br><br>SPSS:<br>`WEIGHT BY weightvar.`<br>`COMMAND`<br>`WEIGHT OFF.` |
| ddweight | *DemoDiff* | | |
| d1weight | *pairfam & DemoDiff* | | |
| psweight | *pairfam* | Post-stratification weight to correct for systematic nonresponse | |
| caweight | *DemoDiff* | | |
| ca1weight | *pairfam & DemoDiff* | | |
| dxpsweight | *pairfam* | Combined design and post-stratification weight | |
| ddcaweight | *DemoDiff* | | |
| d1ca1weight | *pairfam & DemoDiff* | | |
| lweight | *pairfam, DemoDiff, pairfam & DemoDiff* | Longitudinal weight to account for attrition | |

## 2.3 Analyses

Users should know about some pitfalls in the pairfam data. Therefore, you should read section 3 in the ☞Data Manual before starting your analysis. This section provides an overview about „**pairfam Nuts and Bolts**“.

When working with pairfam data, users should display and check frequency distributions for all included variables before starting their analyses. In this way, users can assure that missing values will be included as such in the calculations. Missing or incomplete information is encoded by a uniform set of codes with negative values. Hence, users can decide for themselves which information to treat as **missing** (☞Data Manual). To exclude appropriate cases from the analysis, negative values must be defined as missing beforehand.

| Coding | Meaning | Definition as missing |
|---|---|---|
| -1 | Don’t know | Stata:<br>mvdecode _all, mv(-9/-1)<br>mvdecode _all, mv(-1=.a \...\ -9=.i)<br>mvdecode *varlist*, mv(-1=.a \...\ -9=.i)<br><br>SPSS:<br>`MISSING VALUES ALL (-9 THRU -1).`<br>`MISSING VALUES varlist (-9 THRU -1).`<br>`RECODE varlist (-9 THRU -1 = SYSMIS).` |
| -2 | No answer (also: I don't want to answer, answer refused) | |
| -3 | Does not apply (filter) | |
| -4 | Incorrect input / filter mistake | |
| -5 | Inconsistent value | |
| -6 | Unreadable answer | |
| -7 | Incomplete data | |
| -9 | Invalid multiple answer | |
| -10 | Not in DemoDiff: variable not contained in DemoDiff | |
| -11 | Not in pairfam: variable only contained in DemoDiff | |

For quick access to data analysis, several **syntax files** are available. Syntax files facilitate the understanding of selected analysis methods and help to enhance transparency by documenting all steps by which variables and datasets are generated. All command lines are commented so that they can be used in users’ analysis projects.

Scales: Contains syntax files regarding the creation of index variables and scales on the basis of instruments used in the German Family Panel; available for Stata and SPSS, separated for each respondent group and wave; only available in the scientific use file.

Generated variables: Contains syntax files for the creation of datasets delivered by the scientific use file (see Section 2.1) as well as the generated variables contained in the datasets (see Section 2.2); only available in Stata; only available in the scientific use file.

☞Quick Starts: Commented syntax files that are provided as a service for users in order to facilitate the handling of pairfam data, including suggestions for cross-sectional, longitudinal, and dyadic data matching procedures; for constructing samples, generating index variables, and exercising strategies of analysis as illustrated in the ☞Technical Paper No. 2; only available in Stata; not contained in scientific use file.

### 2.4 Combining datasets

For most analyses, it is necessary to combine data from several datasets. In order to do this, we recommend selecting only relevant variables from the corresponding datasets. Due to the complexity and extent of pairfam data, we recommend that combined datasets contain only the variables required. When combining data, the following decisions must be made depending on the intended type of analysis.

- *target dataset*
  - *actor panel = combining data of one respondent group from multiple waves*
  - *multi-actor cross section = combining data of different respondent group from one wave*
  - *multi-actor panel = combining data of different respondent groups from multiple waves*
- *data format*
  - *WIDE = additional data is added as new variables to the relevant data row*
  - *LONG = additional data is appended as new data rows*
  - *LONG-WIDE = combination of both data formats*
- *handling of nonresponse*
  - *balanced = target dataset only contains completely valid cases*
  - *unbalanced = target dataset contains all cases of matched datasets (e.g. panel dropouts)*

Because there are numerous possibilities for combination, it is important to determine one's analysis methods, goals, and data structure before combining individual datasets. Several examples on the following pages illustrate some typical procedures. Extensively commented Stata syntax can be found on the pairfam website (☞Quick Start).

A special feature of pairfam data should be noted at this point. Variables with identical content share a common name within the respondent groups throughout all waves. For merging appropriate data in WIDE format, variables need to be renamed first due to problems caused by merging variables

with common names in one dataset. Renaming variables is also necessary when merging information of similar content from different respondent groups (e.g. anchor and partner) in LONG format. In this case, relevant variable names need to be harmonized and adapted.

The following examples show both strategies. Examples and the QuickStart file "Matching.do" are for the statistics program Stata. General commands for SPSS are: `ADD FILES`, `MATCH FILES` and `RENAME VARIABLES`.

***Example 1: Actor panel in WIDE format***

Anchor (Wave 1)

| | id | wave | var1 | var2 | var3 |
|---|---|---|---|---|---|
| 1 | 100000 | 1 | 1 | 2 | 3 |
| 2 | 101000 | 1 | 5 | 4 | 3 |
| 3 | 102000 | 1 | -3 | 3 | -3 |

Anchor (Wave 2)

| | id | wave | var1 | var2 | var3 |
|---|---|---|---|---|---|
| 1 | 100000 | 2 | 2 | 3 | 4 |
| 2 | 101000 | 2 | 4 | 3 | 2 |
| 3 | 102000 | 2 | 2 | -3 | 2 |

Target dataset: Anchor (Waves 1+2)

| | id | wave_01 | var1_01 | var2_01 | var3_01 | wave_02 | var1_02 | var2_02 | var3_02 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 100000 | 1 | 1 | 2 | 3 | 2 | 2 | 3 | 4 |
| 2 | 101000 | 1 | 5 | 4 | 3 | 2 | 4 | 3 | 2 |
| 3 | 102000 | 1 | -3 | 3 | -3 | 2 | 2 | -3 | 2 |

Please note:

- renaming of common variable names is necessary
- no renaming of key variable id → must be identical in both merged datasets

DO-File for Stata:

```
use id wave var1 var2 var3 using anchor1, clear      // selects variables from anchor W1
foreach x in wave var1 var2 var3 {                   // variables for renaming (NOT id)
    rename `x' `x'_01                                // renaming by postfix „_01“
}                                                    // ends the loop
save anchor1_01, replace                             // temporary storage of dataset

use id wave var1 var2 var3 using anchor2, clear      // procedure also applies for anchor W2
foreach x in wave var1 var2 var3 {
    rename `x' `x'_02
}
save anchor2_02, replace

use anchor1_01, clear                                // re-query of temporarily stored dataset anchor W1
merge 1:1 id using anchor2_02                        // adds data from temporarily stored dataset anchor W2
keep if _merge==3                                    // creates a balanced dataset
drop _merge                                          // deletes the proxy variable _merge

erase anchor1_01.dta                                 // deletes temporarily stored datasets
erase anchor2_02.dta
```

### *Example 2: Actor panel in LONG format*

Anchor (Wave 1)

| | id | wave | var1 | var2 | var3 |
|---|---|---|---|---|---|
| 1 | 100000 | 1 | 1 | 2 | 3 |
| 2 | 101000 | 1 | 5 | 4 | 3 |
| 3 | 102000 | 1 | -3 | 3 | -3 |

Anchor (Wave 2)

| | id | wave | var1 | var2 | var3 |
|---|---|---|---|---|---|
| 1 | 100000 | 2 | 2 | 3 | 4 |
| 2 | 101000 | 2 | 4 | 3 | 2 |
| 3 | 102000 | 2 | 2 | -3 | 2 |

Target dataset: Anchor (Waves 1+2)

| | id | wave | var1 | var2 | var3 |
|---|---|---|---|---|---|
| 1 | 100000 | 1 | 1 | 2 | 3 |
| 2 | 100000 | 2 | 2 | 3 | 4 |
| 3 | 101000 | 1 | 5 | 4 | 3 |
| 4 | 101000 | 2 | 4 | 3 | 2 |
| 5 | 102000 | 1 | -3 | 3 | -3 |
| 6 | 102000 | 2 | 2 | -3 | 2 |

Please note:

- not necessary to rename common variables
- same procedure for matching data from DemoDiff

DO-File for Stata:

```
use id wave var1 var2 var3 using anchor1, clear        // selects variables from anchor W1
append using anchor2, keep (id wave var1 var2 var3)    // adds selected variables from anchor W2
sort id wave                                           // sorts data by id and wave
```

### *Example 3: Multi actor cross section with anchor and partner in WIDE format*

Anchor (Wave 1)

| | id | wave | var1 | var2 | var3 |
|---|---|---|---|---|---|
| 1 | 100000 | 1 | 1 | 2 | 3 |
| 2 | 101000 | 1 | 5 | 4 | 3 |
| 3 | 102000 | 1 | -3 | 3 | -3 |

Partner (Wave 1)

| | id | pid | wave | pvar1 | pvar2 | pvar3 |
|---|---|---|---|---|---|---|
| 1 | 100000 | 100101 | 1 | 2 | 3 | 4 |
| 2 | 101000 | 101101 | 1 | 4 | 3 | 2 |
| 3 | 102000 | 102101 | 1 | 2 | -3 | 2 |

Target dataset: Anchor + Partner (Wave 1)

| | id | wave | var1 | var2 | var3 | pid | pvar1 | pvar2 | pvar3 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 100000 | 1 | 1 | 2 | 3 | 100101 | 2 | 3 | 4 |
| 2 | 101000 | 1 | 5 | 4 | 3 | 101101 | 4 | 3 | 2 |
| 3 | 102000 | 1 | -3 | 3 | -3 | 102101 | 2 | -3 | 2 |

Please note:

- not necessary to rename dissimilarly named variables

DO-File for Stata:

```
use id wave var1 var2 var3 using anchor1, clear        // selection of variables from anchor W1
merge 1:1 id using partner1, keepusing (pid pvar1 pvar2 pvar3)   // adds selected variables from partner W1
keep if _merge==3                                      // creates a balanced dataset
drop _merge                                            // deletes proxy variable _merge
```

### *Example 4: Multi actor cross section with anchor and partner in LONG format*

Anchor (Wave 1)

| | id | pid | wave | var1 | var2 |
|---|---|---|---|---|---|
| 1 | 100000 | 100101 | 1 | 1 | 2 |
| 2 | 101000 | 101101 | 1 | 5 | 4 |
| 3 | 102000 | 102101 | 1 | -3 | 3 |

Partner (Wave 1)

| | id | pid | wave | pvar1 | pvar2 |
|---|---|---|---|---|---|
| 1 | 100000 | 100101 | 1 | 2 | 3 |
| 2 | 101000 | 101101 | 1 | 4 | 3 |
| 3 | 102000 | 102101 | 1 | 2 | -3 |

Target dataset: Anchor + Partner (Wave 1)

| | id | pid | anker | wave | var1 | var2 |
|---|---|---|---|---|---|---|
| 1 | 100000 | 100101 | 1 | 1 | 1 | 2 |
| 2 | 100000 | 100101 | 0 | 1 | 2 | 3 |
| 3 | 101000 | 101101 | 1 | 1 | 5 | 4 |
| 4 | 101000 | 101101 | 0 | 1 | 4 | 3 |
| 5 | 102000 | 102101 | 1 | 1 | -3 | 3 |
| 6 | 102000 | 102101 | 0 | 1 | 2 | -3 |

Please note:

- dissimilarly named variables from partner data need to be adapted to anchor data
- no renaming of key variable id → must be identical in both combined datasets
- generation of an origin variable to identify whether data originates from anchor or partner

DO-File for Stata:

```
use id pid wave pvar1 pvar2 using partner1, clear      // selects variables from partner W1
foreach x in var1 var2 {                                // variables for renaming (NOT id)
    rename p`x' `x'                                     // renaming = adaption to anchor W1
}                                                       // ends the loop

append using anchor1, keep (id pid wave var1 var2) generate (anchor)   // adds anchor data + origin variable
order id pid anker wave var1 var2                       // orders variables
gsort id anchor                                         // sorts by id and origin variable
```

### *Example 5: Multi actor cross section with anchor and parents in LONG-WIDE format*

Anchor (Wave 2)

| | id | wave | var1 | var2 |
|---|---|---|---|---|
| 1 | 100000 | 2 | 1 | 2 |
| 2 | 101000 | 2 | 5 | 4 |
| 3 | 102000 | 2 | -3 | 3 |
| 4 | 103000 | 2 | 5 | 5 |

Parents (Wave 2 - LONG)

| | id | parid | wave | parvar3 |
|---|---|---|---|---|
| 1 | 100000 | 100301 | 2 | 1 |
| 2 | 100000 | 100302 | 2 | 2 |
| 3 | 100000 | 100304 | 2 | 3 |
| 4 | 101000 | 101301 | 2 | 4 |
| 5 | 101000 | 101302 | 2 | 5 |
| 6 | 102000 | 102303 | 2 | -3 |

Target dataset: Anchor + Parents (Wave 2)

| | id | wave | var1 | var2 | parid | parvar3 |
|---|---|---|---|---|---|---|
| 1 | 100000 | 2 | 1 | 2 | 100301 | 1 |
| 2 | 100000 | 2 | 1 | 2 | 100302 | 2 |
| 3 | 100000 | 2 | 1 | 2 | 100304 | 3 |
| 4 | 101000 | 2 | 5 | 4 | 101301 | 4 |
| 5 | 101000 | 2 | 5 | 4 | 101302 | 5 |
| 6 | 102000 | 2 | -3 | 3 | 102303 | -3 |
| 7 | 103000 | 2 | 5 | 5 | . | . |

Please note:

- not necessary to rename dissimilarly named variables
- creates an unbalanced target dataset which contains information about anchors without parental information (e.g., case id=1030000)

DO-File for Stata:

```
use id wave var1 var2 using anchor2, clear          // selects variables from anchor W2
merge 1:m id using parent2, keepusing (parid parvar3)  // adds variables from parents W2
drop _merge                                         // deletes proxy variable _merge
sort id wave                                        // sorts by id and parent ID
```

### 2.5 Teaching pairfam

The pairfam data can be used for teaching purposes. If students need access to the data, reduced datasets should be used in order to protect respondent privacy. The reduced version of datasets can be made accessible to students without completing the ☞distribution form. In the scientific use file, we provide the syntax required to produce this reduced version from the original data by generating new datasets that contain only 50% of cases.

## *3 Publications*

The German Family Panel is an infrastructure project that provides extensive processed data for scientific analyses. The value of the survey is seen above all in the number and quality of publications based on its data.

### 3.1 Literature Database

A compilation of published works based on data from the pairfam survey can be found on our website. Overviews of all ☞ pairfam publications are updated regularly. We very much rely on users self-reporting their new publications and working papers: **Please inform us on all your publications based on data of the German Family Panel per email** (☞support@pairfam.de) The registered publications are taken as a basis for our ☞Best Paper Award.

☞ Bibliography: The database of pairfam literature includes bibliographical information and abstracts for all known publications related to the German Family Panel.

☞ Student papers: Student-authored final theses and papers based on data from the German Family Panel are registered by pairfam user support.

### 3.2 Citation

We kindly request you to mention our survey within your own publications. The use of pairfam data in your work should be acknowledged by citing both the reference paper ☞(Huinink et al., 2011) and the dataset ☞(Brüderl et al., 2019). Note that the citation of the dataset is specific to the Release used as year of publication, list of authors, and doi change across Releases. For more details on the rules of pairfam citation, please see the ☞pairfam website.