



## **pairfam Quick Guide**

**Release 12.0**

pairfam Group

May 2021

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

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 use. In addition, this guide focuses on the merging of data sets, 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 (<https://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](mailto:support@pairfam.de)).

## **Contents**

1	Quick Orientation .....	1
2	Advice for Data Users .....	2
2.1	Data sets .....	3
2.2	Variables .....	5
2.3	Analyses .....	6
2.4	Combining data sets .....	7
2.5	Teaching pairfam .....	12
3	Publications .....	12
3.1	Literature Database .....	12
3.2	Citation .....	12

### **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 data sets and provides a detailed description of the data-editing procedures undertaken; contains relevant notes on generated variables and data sets, 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 12.0**, published in summer of 2021, pairfam data from the first twelve 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 12.0, including data from the first twelve 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. In wave 11, a refreshment sample of approximately 5,000 respondents was added to strengthen the birth cohorts 1981-83 and 1991-93, as well as add a new, younger cohort (2001-03). Information about number of cases in the various ☞ [samples](#) is given on the pairfam website.

As a separate scientific use file, data from the pairfam COVID-19 survey are available as of December 2020.

## 2.1 Data sets

Because information is stored separately by wave and respondent group, the scientific use file contains several data sets. The linkage of those data sets for longitudinal or dyadic analyses is possible through a standardized key variable (see Sections 2.2 and 2.4). DemoDiff data are partly included in the scientific use file as separate data sets (DemoDiff waves 1 and 2) and are partly merged with the pairfam data sets (DemoDiff wave 3), as shown in the following table. Data from the refreshment sample are included in all wave 11 data sets. The various samples (pairfam original sample, DemoDiff, and refreshment sample) can be identified by the variable *sample*.

Anker	Partner	Child	Parents	Parenting	Parenting Adolescents	Parenting U6 (Partner)
<b>Wave 1 (2008/2009)</b>						
anchor1	partner1	-	-	-	-	-
anchor1_DD	partner1_DD	-	-	-	-	-
<b>Wave 2 (2009/2010)</b>						
anchor2	partner2	child2	parent2	parenting2	-	-
anchor2_DD	-	-	-	-	-	-
<b>Wave 3 (2010/2011)</b>						
anchor3	partner3	child3	parent3	parenting3	-	-
<b>Wave 4 (2011/2012)</b>						
anchor4	partner4	child4	parent4	parenting4	-	-
<b>Wave 5 (2012/2013)</b>						
anchor5	partner5	child5	parent5	parenting5	-	-
<b>Wave 6 (2013/2014)</b>						
anchor6	partner6	child6	parent6	parenting6	-	-
<b>Wave 7 (2014/2015)</b>						
anchor7	partner7	child7	parent7	parenting7	-	-
<b>Wave 8 (2015/2016)</b>						
anchor8	partner8	child8	parent8	parenting8	-	-
<b>Wave 9 (2016/2017)</b>						
anchor9	partner9	child9	-	parenting9	paya9	-
<b>Wave 10 (2017/2018)</b>						
anchor1	partner10	child10	-	parenting10	paya10	-
<b>Wave 11 (2018/2019)</b>						
anchor11	partner11	child11	-	parenting11	paya11	parenting-U6partner11
<b>Wave 12 (2019/2020)</b>						
anchor12_CAPI	partner12	child12_capi	-	parenting12	paya12	parenting-U6partner12
anchor12_CATI		child12_cati				

Furthermore, there are several generated data sets with biographical information available across all waves ([Data Manual](#), Section 4.6).

Generated data sets	Filename	Content
Biography of activities	bioact 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 data sets	Anchor	Partner	Parenting U6 (Partner)
<b>Wave 4</b> <b>2011/12</b>	stepup_anchor4 stepup_transition_anchor4	stepup_partner4 -	
<b>Wave 5</b> <b>2012/13</b>	stepup_anchor5 stepup_transition_anchor5	stepup_partner5 -	
<b>Wave 6</b> <b>2013/14</b>	stepup_anchor6 stepup_transition_anchor6	stepup_partner6 -	
<b>Wave 7</b> <b>2014/15</b>	stepup_anchor7 stepup_transition_anchor7	stepup_partner7 -	
<b>Wave 8</b> <b>2015/16</b>	stepup_anchor8 stepup_transition_anchor8	stepup_partner8 -	
<b>Wave 9</b> <b>2016/17</b>	stepup_anchor9+transition.dta -	stepup_partner9 -	
<b>Wave 10</b> <b>2017/18</b>	stepup_anchor10+transition.dta -	stepup_partner10 -	
<b>Wave 11</b> <b>2018/19</b>	stepup_anchor11+transition.dta -	stepup_partner11 -	stepup_parentingU6partner11
<b>Wave 12</b> <b>2019/20</b>	stepup_anchor12_capi+transition.dta stepup_anchor12_cati+transition.dta -	stepup_partner12 -	stepup_parentingU6partner12
<b>generated data sets</b>	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' data sets are distinguished by an appropriate prefix:

- *Variable name sex: anchor = sex / partner = psex / parents = parsex / children = csex*
- *Prefix for data from the paya questionnaire is paya, for data from the U6 parenting questionnaire for partners it is p.*

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 data sets. 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, data sets contain several additional **variables** that are **generated** during data preparation. These variables ([Data Manual](#), Section 4.3) 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 4.6) in the anchor data sets. The usage of these weighting factors depends on the type and intention of the analysis being undertaken by the user. As of Release 12.0 we provide new weights which enable weighted analyses for all subsamples (pairfam base, DemoDiff, refreshment sample).

Weighting variable	Sample	Description	Application
dweight	<i>base sample</i>	Design weight: Correct for disproportional size of gross sample of the cohorts	Stata: svyset [pweight=weightvar] svy: command  command [pweight=weightvar], options
d1weight	<i>base &amp; DemoDiff</i>		
d2weight	<i>base &amp; DemoDiff &amp; refreshment</i>		
d3weight	<i>refreshment</i>		
cdweight	<i>base sample</i>	Calibrated design weight: Adjust the data to the target population & control for selective baseline survey participation and panel attrition; = Combination of design and calibration weight	SPSS: WEIGHT BY weight- var. COMMAND WEIGHT OFF.
cd1weight	<i>base &amp; DemoDiff</i>		
cd2weight	<i>base &amp; DemoDiff &amp; refreshment</i>		
cd3weight	<i>refreshment</i>		

## 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 (also see [Quick Start](#) “Missings.do”).

Coding	Meaning	Definition as missing
-1	Don't know	Stata: mvdecode _all, mv(-12/-1)  mvdecode _all, mv(-1=.a \...\ -12=.k) mvdecode varlist, mv(-1=.a \...\ -12=.k)  SPSS: MISSING VALUES ALL (-12 THRU -1). MISSING VALUES varlist (-12 THRU -1). RECODE varlist (-12 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	
-12	Non-response PAPI	

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 data sets 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 data sets delivered by the scientific use file (see Section 2.1) as well as the generated variables contained in the data sets (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, defining missings, using weighting variables, as well as examples of event-history and fixed-effects panel data analysis; only available in Stata; not contained in scientific use file.

## 2.4 Combining data sets

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

- target data set
  - *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 data set only contains completely valid cases*
  - *unbalanced = target data set contains all cases of matched data sets (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 data sets. 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 data set. 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 Quick Start 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)						Anchor (Wave 2)					
	id	wave	var1	var2	var3		id	wave	var1	var2	var3
1	100000	1	1	2	3	1	100000	2	2	3	4
2	101000	1	5	4	3	2	101000	2	4	3	2
3	102000	1	-3	3	-3	3	102000	2	2	-3	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 data sets

## 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 data set

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 data set anchor W1
merge 1:1 id using anchor2_02 // adds data from temporarily stored data set anchor W2
keep if _merge==3 // creates a balanced data set
drop _merge // deletes the proxy variable _merge

erase anchor1_01.dta // deletes temporarily stored data sets
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 data set: 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 data set: 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 data set
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 data set: 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 data sets
- 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 data set: 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 data set 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 data sets should be used in order to protect respondent privacy. To make the reduced version of data sets accessible to students a separate [form for the teaching version](#) must be completed and sent to the user support. In the scientific use file, we provide the syntax required to produce this reduced version from the original data by generating new data sets 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](mailto: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 data set [\(Brüderl et al., 2021\)](#). Note that the citation of the data set 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](#).