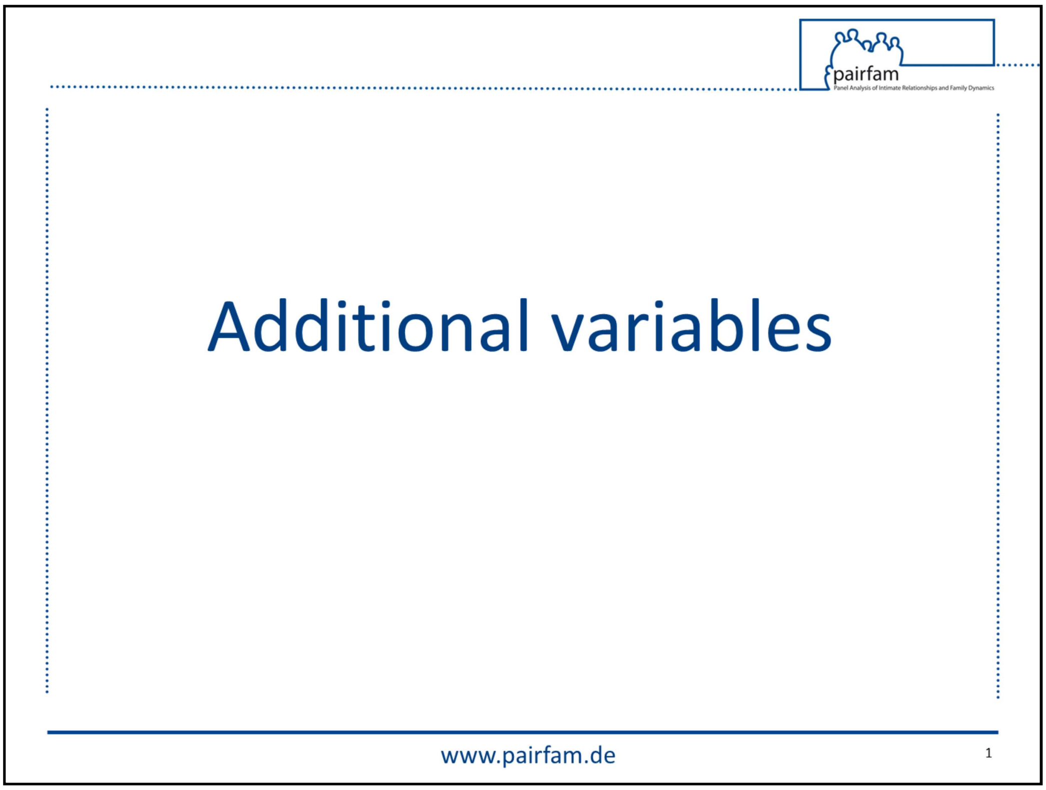


pairfam
Panel Analysis of Intimate Relationships and Family Dynamics

Additional variables

www.pairfam.de

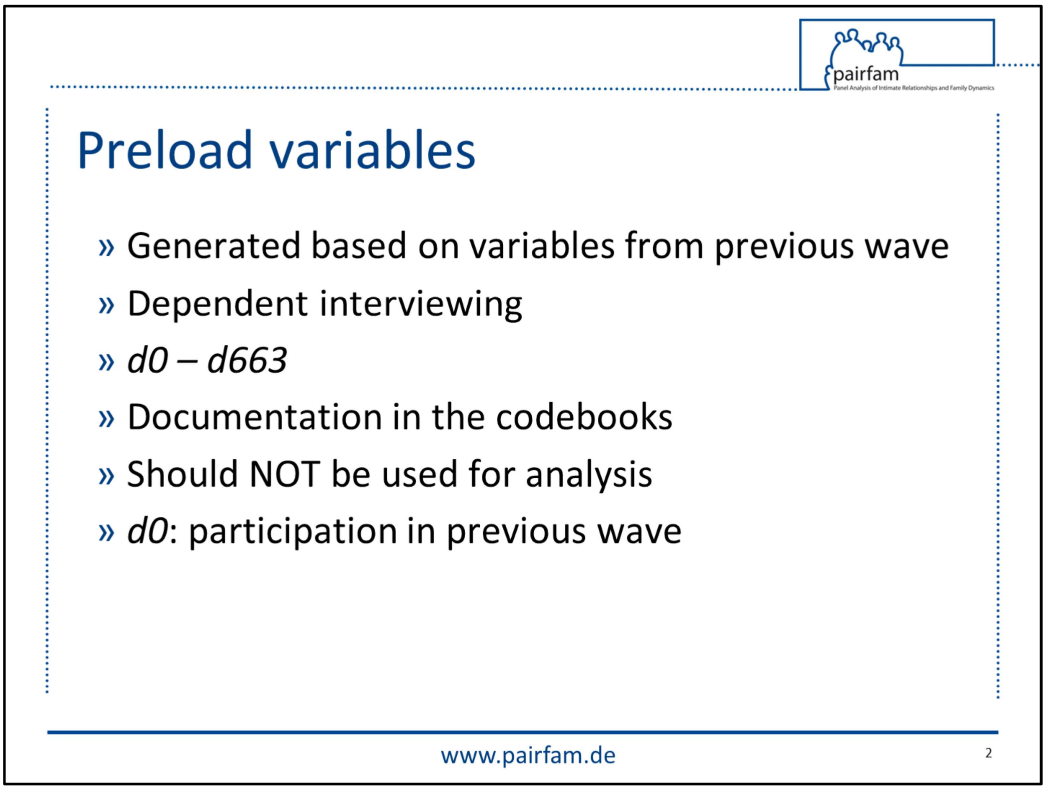
1

pairfam tutorial

7. Additional variables

Kristin Hajek & Madison Garrett

The seventh tutorial explains all other variable types available in addition to the survey variables.



Preload variables

- » Generated based on variables from previous wave
- » Dependent interviewing
- » *d0 – d663*
- » Documentation in the codebooks
- » Should NOT be used for analysis
- » *d0*: participation in previous wave

www.pairfam.de 2

What are preload variables?

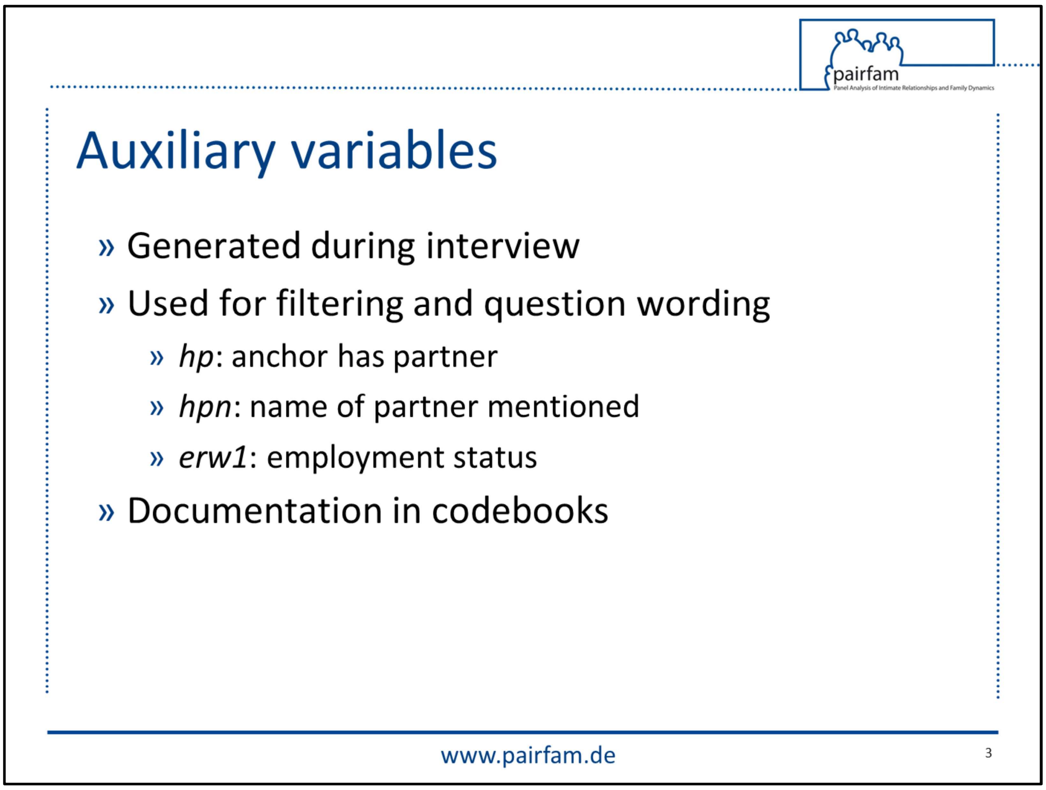
Preload variables are included in the Scientific Use File from wave 2 onward. They were generated based on variables from the previous wave and contain information necessary for dependent interviewing.

The questionnaire in the current wave (more specifically: question wording, filters, etc.) depends, in part, on the articulation of these variables, which are easily recognizable by their form: a lowercase *d* followed by a number.

A detailed list of preload variables can be found in the anchor codebooks for each specific wave.

These variables should ***not*** be used for analysis;

however, the variable d0 can sometimes be useful for data preparation as it can be used to distinguish whether the anchor respondent participated in the previous wave, skipped the last wave, or took part in the survey for the first time in the current wave.



Auxiliary variables

- » Generated during interview
- » Used for filtering and question wording
  - » *hp*: anchor has partner
  - » *hpn*: name of partner mentioned
  - » *erw1*: employment status
- » Documentation in codebooks

What are auxiliary variables?

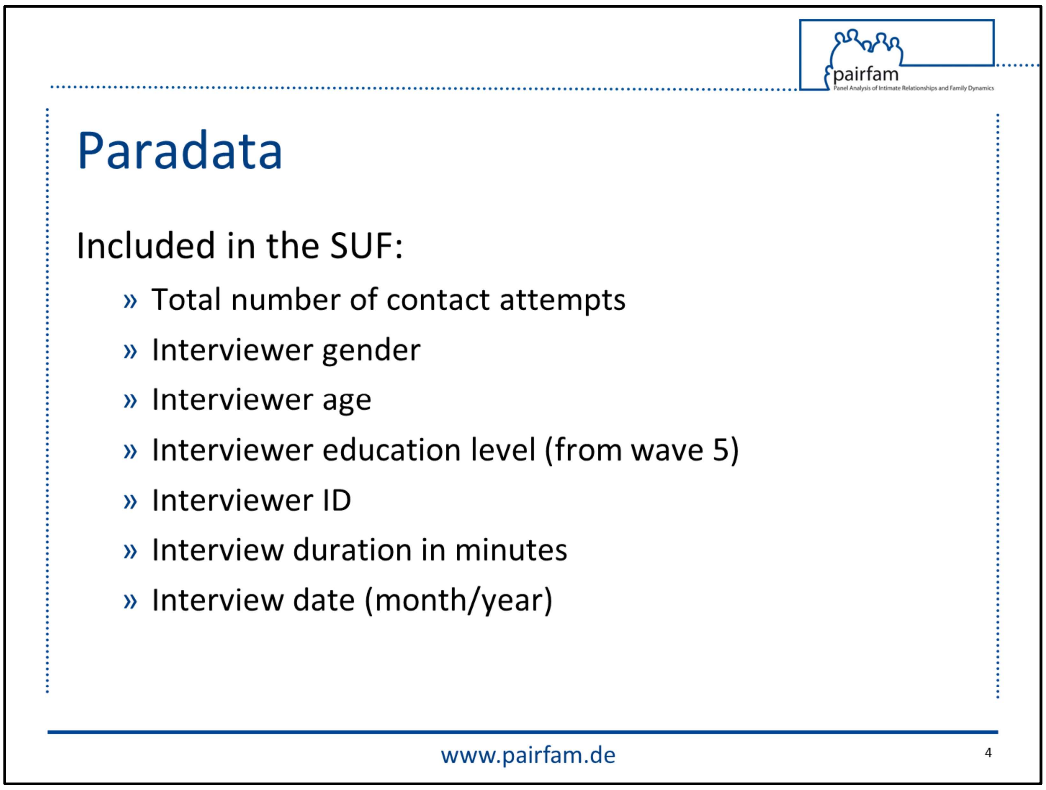
Auxiliary variables were generated during the interview and mainly used to modify filters and question wording.

For example, if a respondent reports a current partner during the survey, they will be posed further questions about this partner.

The auxiliary variable *hpn* stores the name of the current partner and is used as part of the question wording for further questions about this partner. Due to data protection laws, names were eliminated from the data and this variable only states whether a name was mentioned during the interview.

The variable *erw1* reflects whether the anchor indicated at least one occupation at the time of the current interview.

A full list of auxiliary variables is available in the codebooks for each wave.



pairfam
Panel Analysis of Intimate Relationships and Family Dynamics

Paradata

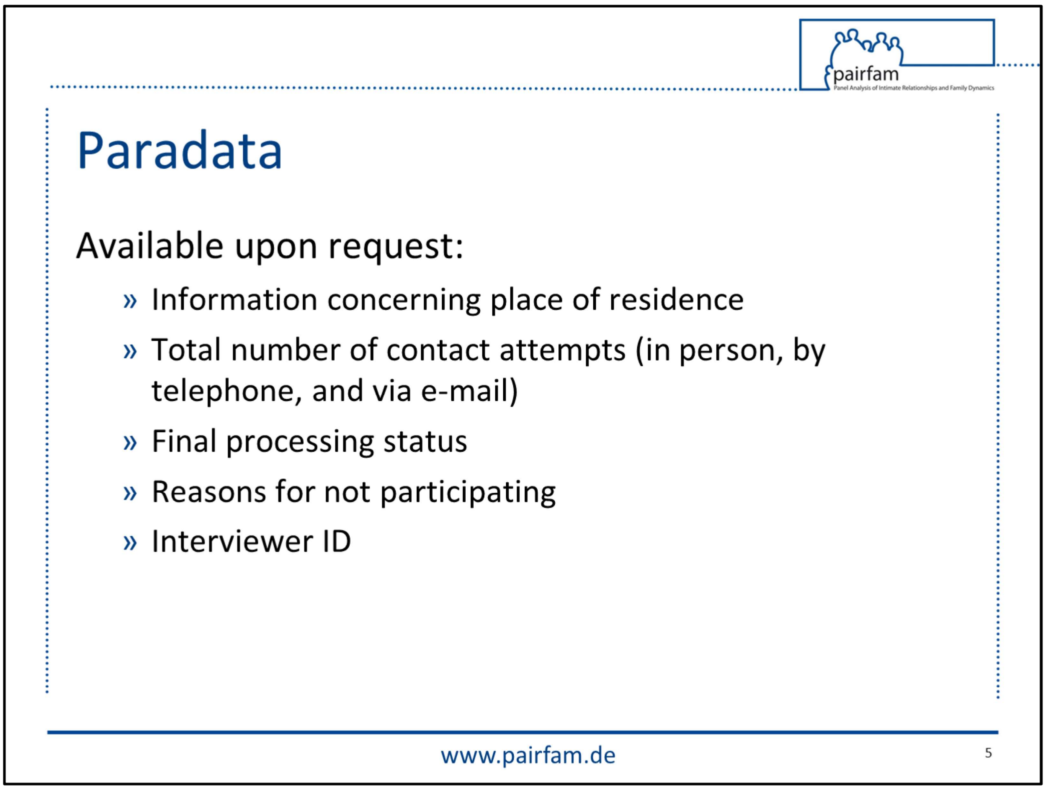
Included in the SUF:

- Total number of contact attempts
- Interviewer gender
- Interviewer age
- Interviewer education level (from wave 5)
- Interviewer ID
- Interview duration in minutes
- Interview date (month/year)

www.pairfam.de 4

What about paradata?

The following paradata is available as part of the Scientific Use File: total number of contact attempts, interviewer gender, interviewer age, interviewer education level, interviewer ID, interview duration in minutes, and the interview date.



Paradata

Available upon request:

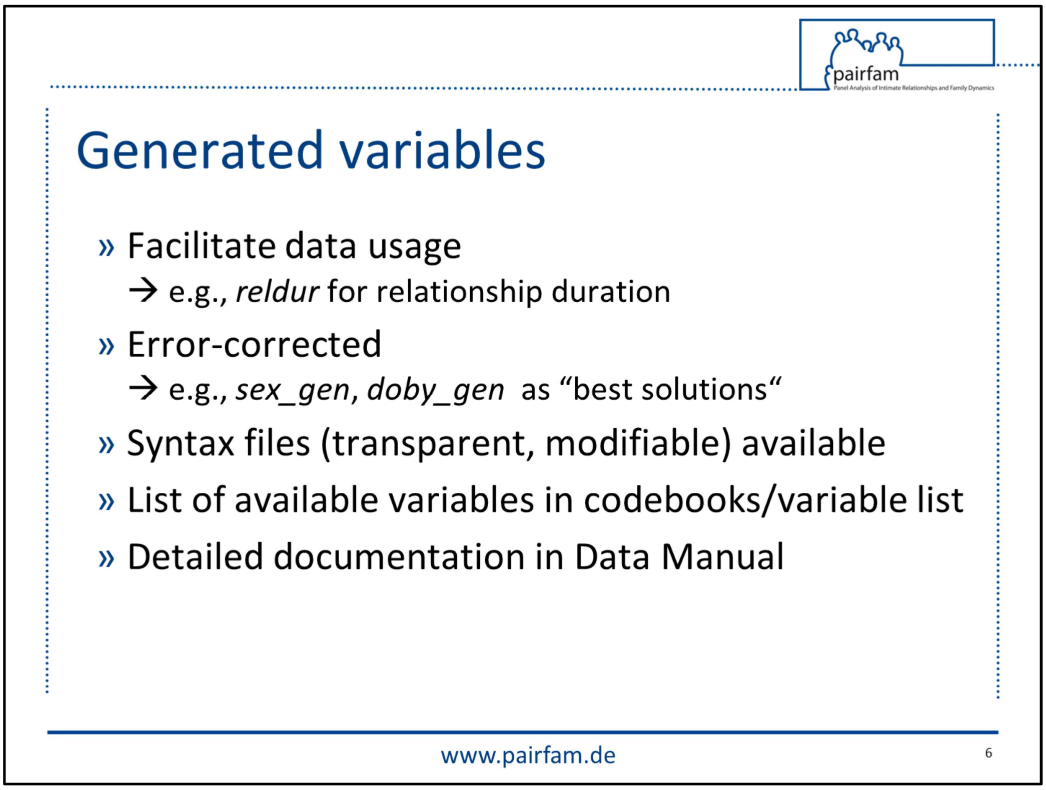
- Information concerning place of residence
- Total number of contact attempts (in person, by telephone, and via e-mail)
- Final processing status
- Reasons for not participating
- Interviewer ID

www.pairfam.de 5

pairfam anchor gross data sets are also available upon request.

These data sets include data for the entire sample population for each wave who have been contacted for an additional interview, i.e. all anchor persons who took part in the previous wave's survey and did not object to being re-interviewed, or who missed out on the previous wave (non-contacts and soft refusals).

The following information on the fieldwork is available: information on the place of residence, total number of contact attempts (in person, by telephone, and via e-mail), a detailed breakdown of the final processing status (including different reasons for non-participation), and interviewer identification numbers.



Generated variables

» Facilitate data usage
→ e.g., *reldur* for relationship duration

» Error-corrected
→ e.g., *sex_gen*, *doby_gen* as “best solutions“

» Syntax files (transparent, modifiable) available

» List of available variables in codebooks/variable list

» Detailed documentation in Data Manual

www.pairfam.de 6

Why does *pairfam* provide generated variables?

Generated variables are indicator variables generated during data processing to facilitate data usage. The generated variable *reldur*, for example, stores the relationship duration with the anchor respondent’s current partner. This information is originally stored in individual Event History Calendar variables for each month of the previous year - that’s a lot of variables across the panel that would need to be transformed to achieve the information included in this variable. As the EHC is very informative, but can also be difficult to use in analyses, we provide generated variables that summarize the key elements of the EHC to make analyses easier for you.

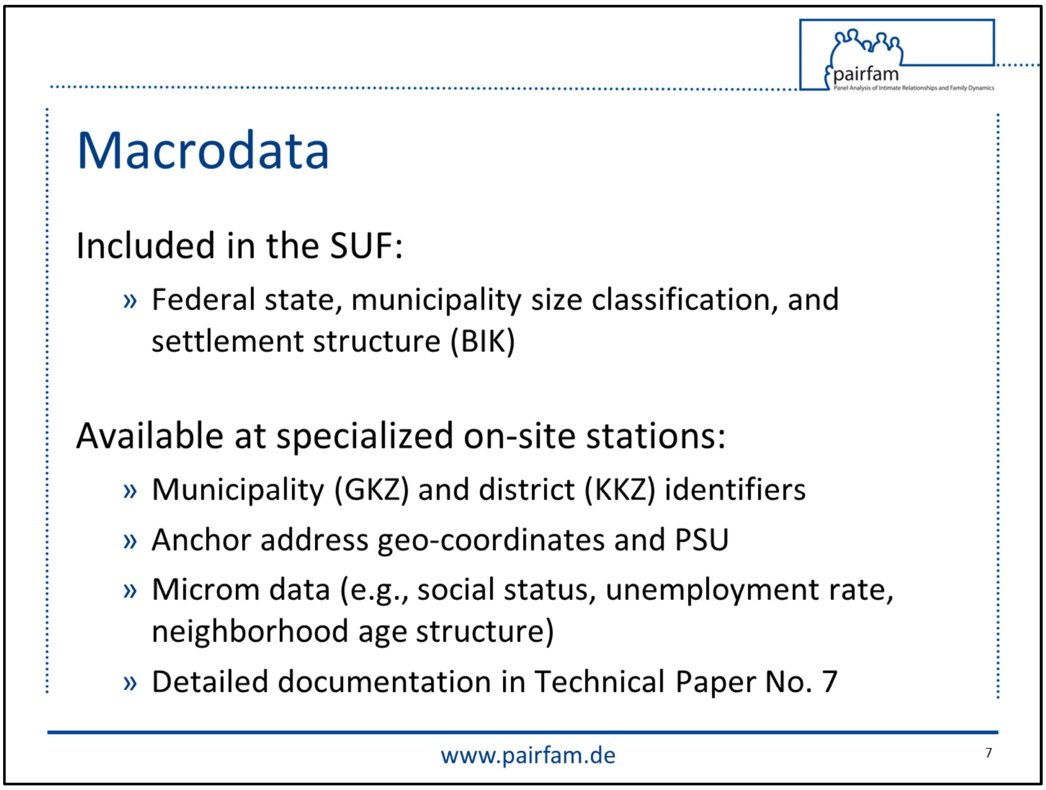
Generated variables are also error-corrected. Gender, birth year, and birth month for anchor and alteri respondents are recognizable by the suffix *_gen* and are “best solutions” over all waves. The original information of each wave is stored in the variables *original_sex* and *original_doby*. For example, if an anchor stated in three waves to be female and in one wave male, we chose the most plausible response based on the assumption of an incorrect entry – here, that the anchor is in fact female – and aligned this information over all waves in the variable *sex_gen*. Two actual changes of gender were communicated throughout the course of the panel to the interviewer. These anchor respondents received the missing code -4 for the variable *sex_gen*.

We highly recommend the usage of generated variables in relevant analyses. The

Stata syntax files for most of the provided generated variables are available as part of the Scientific Use File, making data processing more transparent. Furthermore, you can use the syntax to modify a specific generated variable and adjust it to your research question. Due to data protection measures, the syntax for some generated variables cannot be released. For example, the generation of the variable *isco*, which classifies occupations.

A list of all available variables can be found in the anchor and alteri codebooks as well as in the variable list.

More detailed documentation concerning generated variables is available in the Data Manual.



Macrodata

Included in the SUF:

» Federal state, municipality size classification, and settlement structure (BIK)

Available at specialized on-site stations:

» Municipality (GKZ) and district (KKZ) identifiers

» Anchor address geo-coordinates and PSU

» Microm data (e.g., social status, unemployment rate, neighborhood age structure)

» Detailed documentation in Technical Paper No. 7

www.pairfam.de 7

Which macro variables are available?

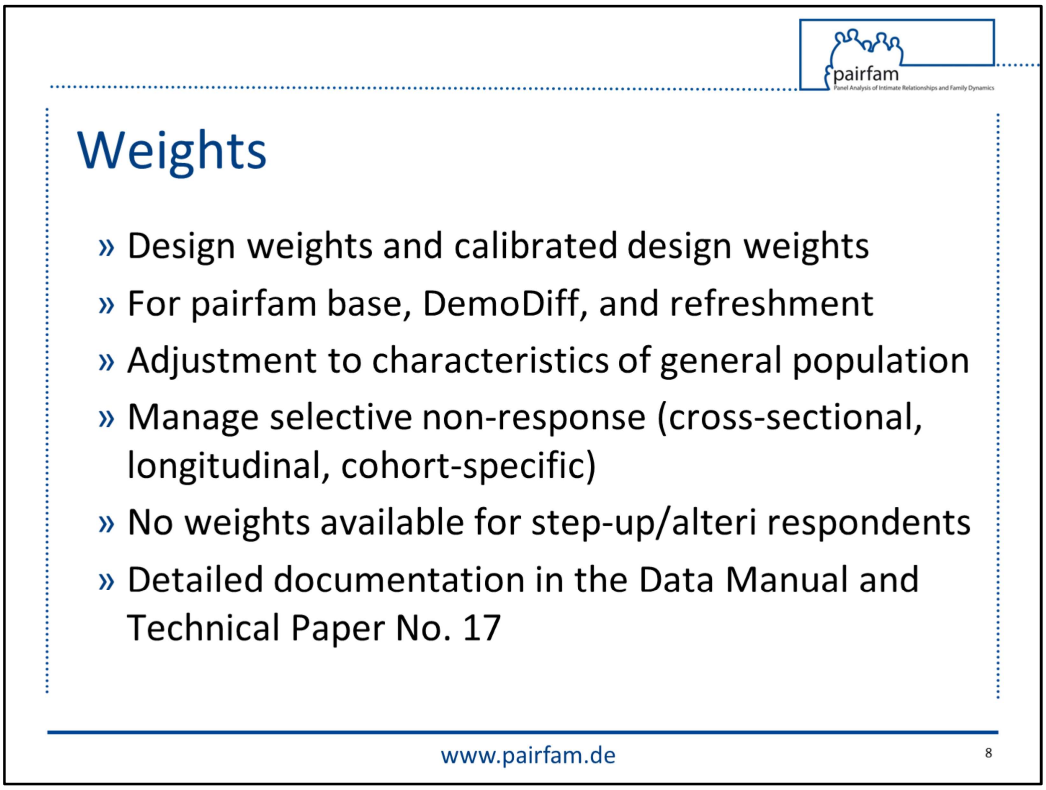
Macrodata are retrospective context attributes concerning the anchor respondent’s place of residence. They allow for analyses of contextual conditions by linking microdata from the *pairfam* survey with external macrodata via the following variables:

Included in the Scientific Use File are federal state, municipality size at the anchor’s main residence, and the settlement structure at both the anchor’s and parents’ main residences.

Municipality, district identifiers, and geo-coordinates of the anchor’s main residence, as well as information on primary sampling units are available for analysis at specialized on-site working stations.

Furthermore, microm regional indicators for waves 1 to 5 including social status, unemployment rate, neighborhood age structure, population density, proportion of foreigners, residential fluctuation, or social milieu at the place of residence are available.

A full list of the microm indicators can be found in Technical Paper No. 7.



pairfam
Panel Analysis of Intimate Relationships and Family Dynamics

Weights

- » Design weights and calibrated design weights
- » For pairfam base, DemoDiff, and refreshment
- » Adjustment to characteristics of general population
- » Manage selective non-response (cross-sectional, longitudinal, cohort-specific)
- » No weights available for step-up/alteri respondents
- » Detailed documentation in the Data Manual and Technical Paper No. 17

www.pairfam.de 8

Which weights are provided?

Design weights as well as calibrated design weights are available for the *pairfam* base, *DemoDiff*, and refreshment sample anchor respondents.

Weights provide factors to adjust the observed data to characteristics of the general population and manage selective non-response by assigning observations with characteristics of higher selectivity a higher analysis weight.

Therefore, both cross-sectional survey participation bias and longitudinal panel attrition bias for the following waves can be tackled.

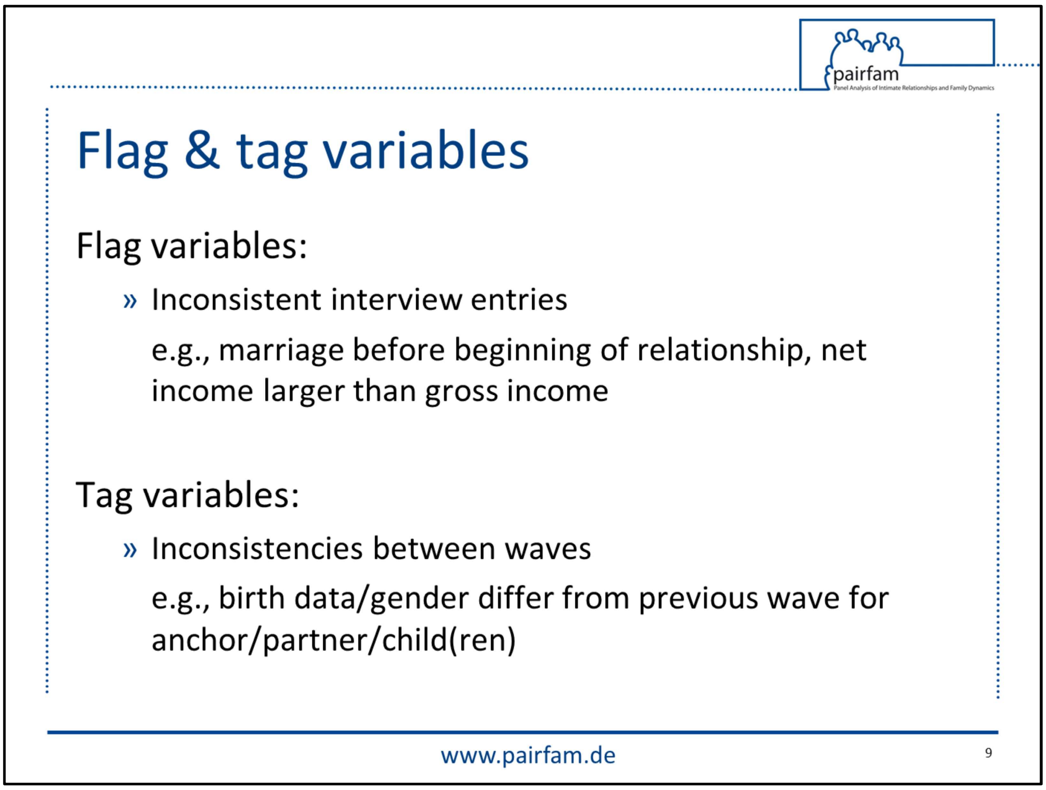
Third, a correction of cohort-specific non-response aiming to represent actual cohort sizes in the population of interest can be integrated.

The design weight corrects disproportionate sampling across cohorts and the combination of multiple selection frames including *DemoDiff* and the wave 11 refreshment sample.

The calibrated design weight calibrates the design weights to reference characteristics, thereby correcting both baseline and longitudinal survey non-response.

No weights are available for *step-up* respondents or alteri respondents.

A detailed documentation of the weighting factors can be found in the Data Manual and in Technical Paper No. 17.



Flag & tag variables

Flag variables:

- » Inconsistent interview entries

  e.g., marriage before beginning of relationship, net income larger than gross income

Tag variables:

- » Inconsistencies between waves

  e.g., birth data/gender differ from previous wave for anchor/partner/child(ren)

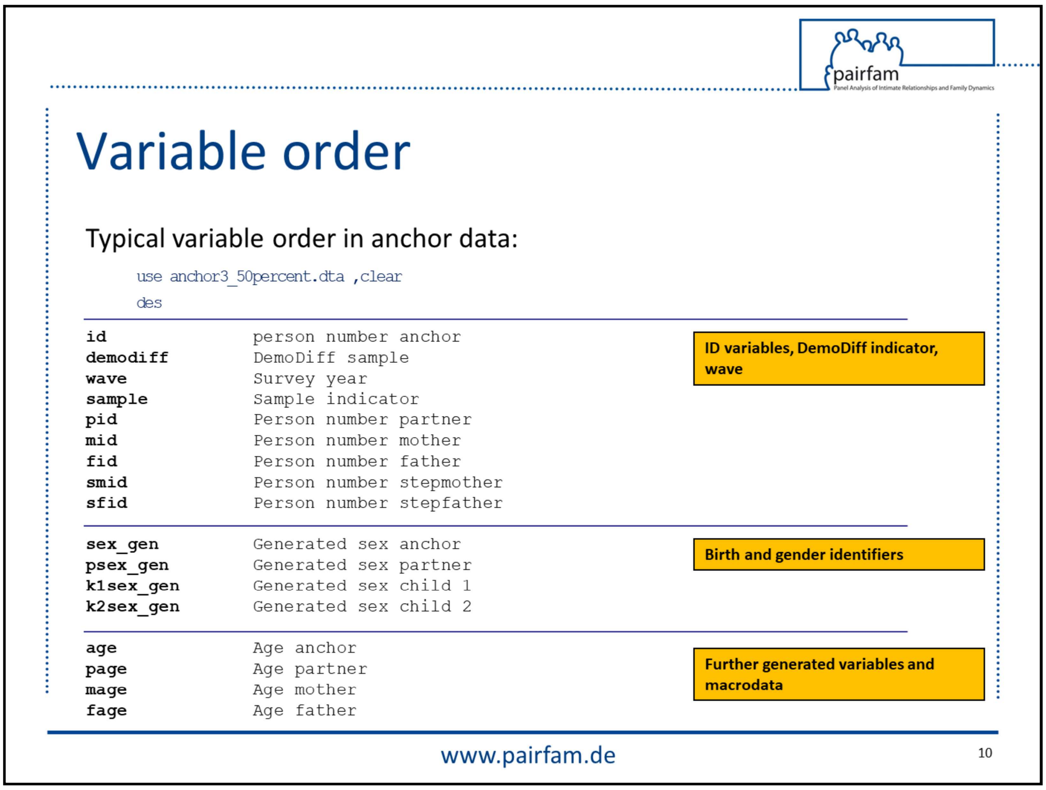
www.pairfam.de 9

What are flag and tag variables?

Flag variables mark inconsistent interview entries in one wave, for example if a marriage reportedly began before the relationship with the same partner, or if a respondent’s net income is higher than the gross income.

Tag variables, on the other hand, mark inconsistencies between waves, for example if the birth data or gender differs from previous waves for anchor, partners, or children.

A full list of flag and tag variables is available in the Data Manual.



Variable order

Typical variable order in anchor data:

```
use anchor3_50percent.dta ,clear
des
```

```
id              person number anchor
demodiff        DemoDiff sample
wave            Survey year
sample          Sample indicator
pid             Person number partner
mid             Person number mother
fid             Person number father
smid            Person number stepmother
sfid            Person number stepfather

sex_gen         Generated sex anchor
psex_gen        Generated sex partner
k1sex_gen       Generated sex child 1
k2sex_gen       Generated sex child 2

age             Age anchor
page            Age partner
mage            Age mother
fage            Age father
```

ID variables, DemoDiff indicator, wave

Birth and gender identifiers

Further generated variables and macrodata

www.pairfam.de

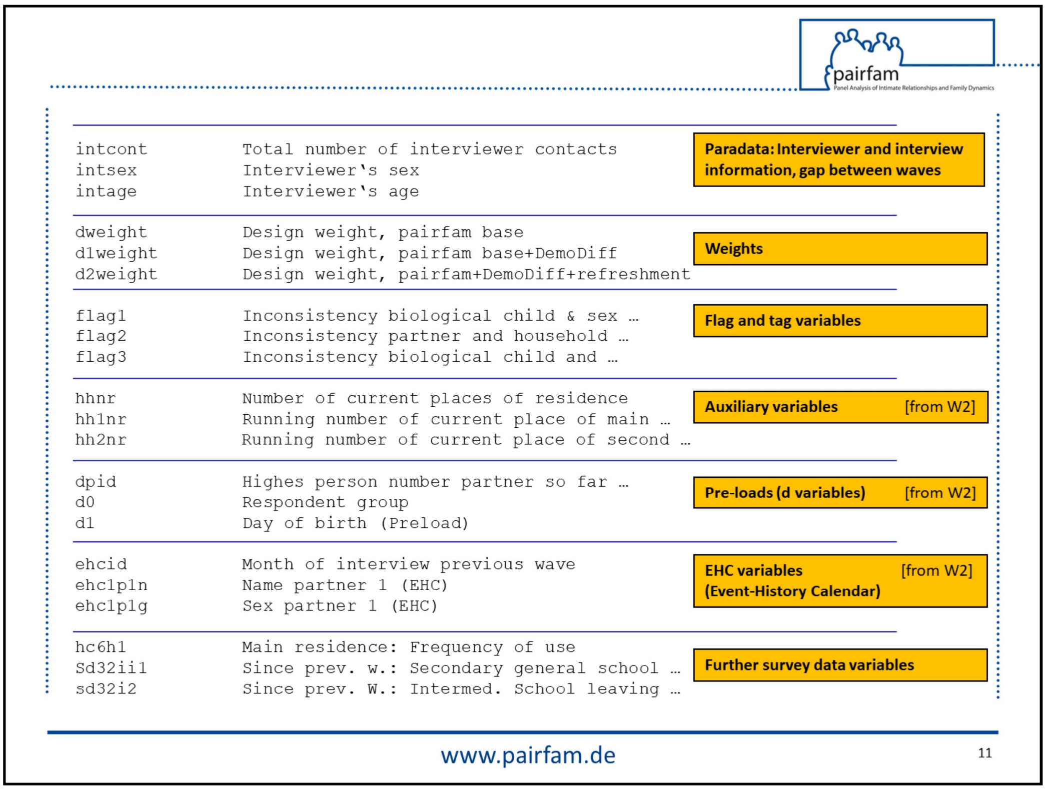
10

Is there a specific order to the variables?

All anchor data sets have the same variable order. Note that this table does not include the full list of variables, but only an excerpt of each variable type.

Identification variables are listed first, followed by sample indicators and the survey year.

Next are the generated birth and gender identifiers, followed by further generated variables, for example age, and macro variables.



pairfam
Panel Analysis of Intimate Relationships and Family Dynamics

intcont	Total number of interviewer contacts	**Paradata: Interviewer and interview information, gap between waves**
intsex	Interviewer's sex	
intage	Interviewer's age	
dweight	Design weight, pairfam base	**Weights**
d1weight	Design weight, pairfam base+DemoDiff	
d2weight	Design weight, pairfam+DemoDiff+refreshment	
flag1	Inconsistency biological child & sex …	**Flag and tag variables**
flag2	Inconsistency partner and household …	
flag3	Inconsistency biological child and …	
hhnr	Number of current places of residence	**Auxiliary variables** [from W2]
hh1nr	Running number of current place of main …	
hh2nr	Running number of current place of second …	
dpid	Highes person number partner so far …	**Pre-loads (d variables)** [from W2]
d0	Respondent group	
d1	Day of birth (Preload)	
ehcid	Month of interview previous wave	**EHC variables (Event-History Calendar)** [from W2]
ehc1p1n	Name partner 1 (EHC)	
ehc1p1g	Sex partner 1 (EHC)	
hc6h1	Main residence: Frequency of use	**Further survey data variables**
Sd32ii1	Since prev. w.: Secondary general school …	
sd32i2	Since prev. W.: Intermed. School leaving …	

www.pairfam.de 11

Afterwards, paradata are listed, for example the number of interviewer contacts and interviewer's gender and age.

The weighting variables then follow along with the flag and tag variables.

Next are the auxiliary and preload variables.

Last but not least – in fact the majority of variables – the Event History Calendar variables followed by the rest of survey data variables.

Next up: Tutorial #8 – Missing values and filter missings

www.pairfam.de

12

This concludes the description of available variable types in the *pairfam* data.

The next tutorial will cover missing values and filter missings.