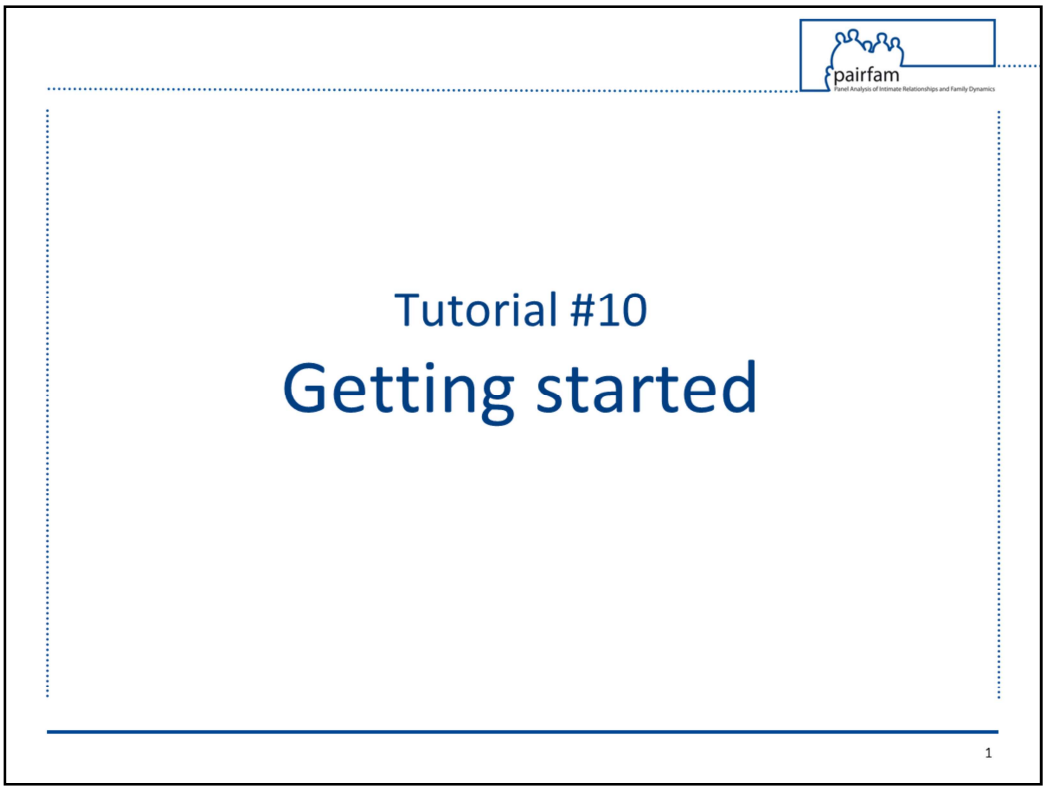
The corresponding video tutorials are available online:
https://www.youtube.com/playlist?list=PL7BcpOtSe5u_zQctYXz4ee79Zc9r4mfnr

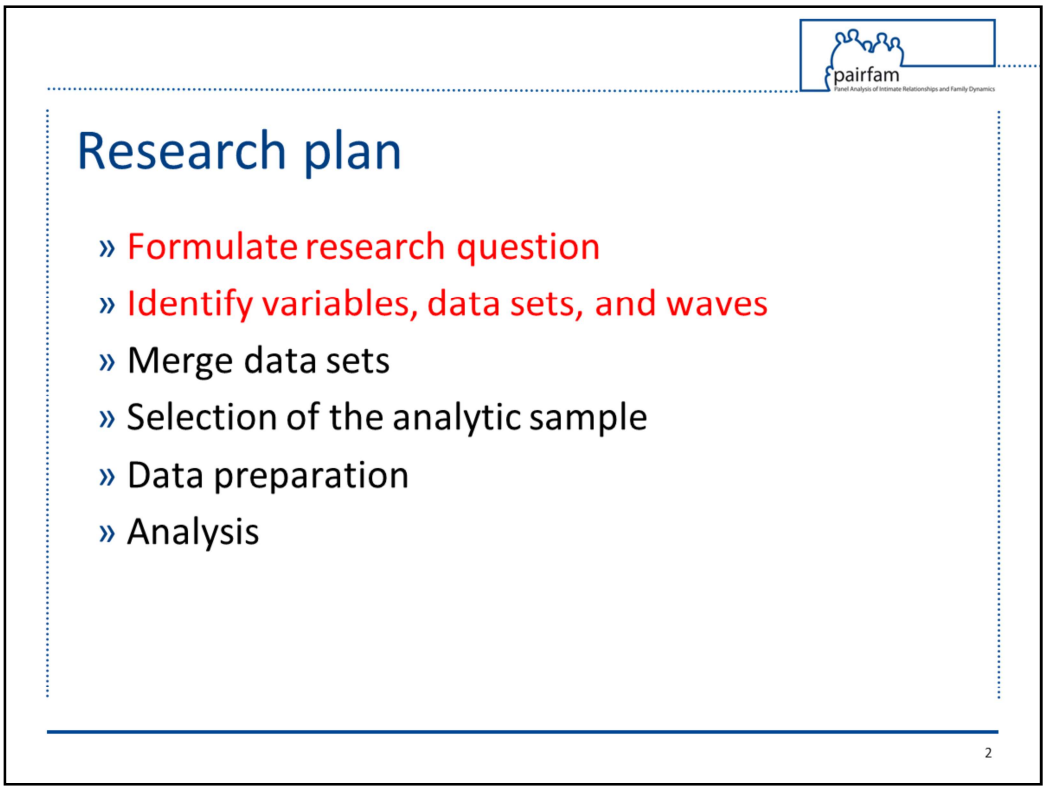


pairfam tutorial

10. Getting started

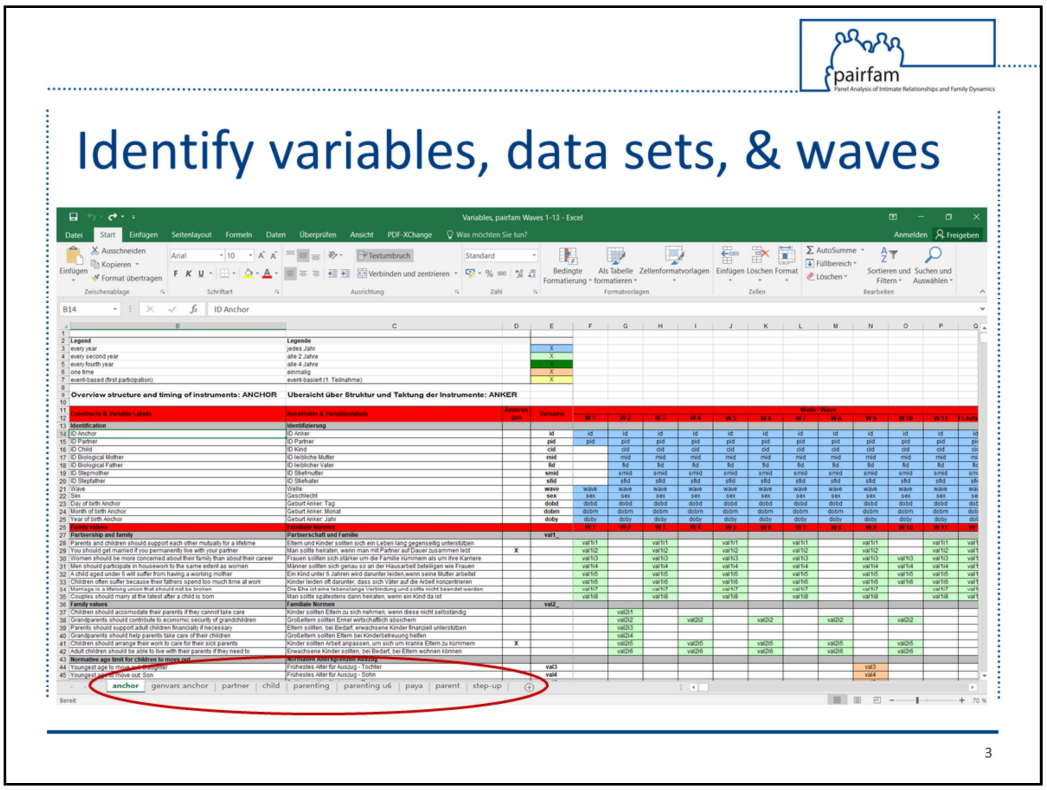
Kristin Hajek & Madison Garrett, May 2022

Now that we have covered the *pairfam* design, samples, data sets, variables and available documentation, you're ready to begin your analysis. This tenth tutorial will help you get started.

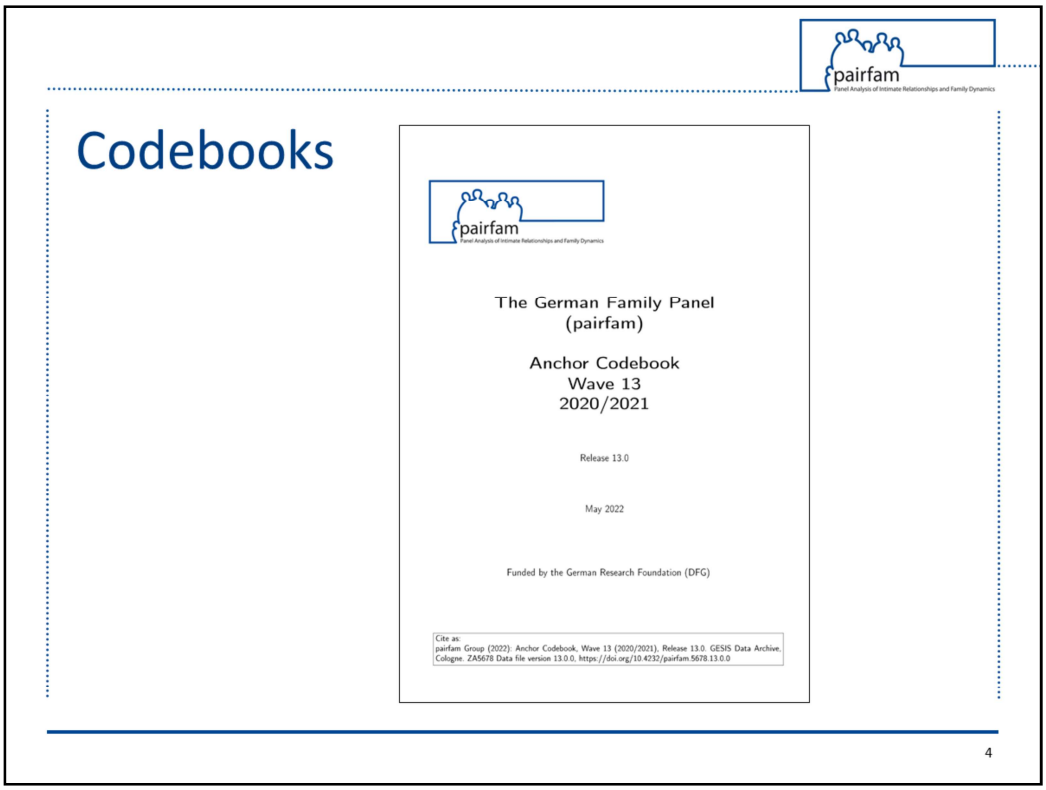


Where to start?

The first step is to formulate your exact research question, which will most likely take place in interaction with the second step: identifying the available variables in the *pairfam* data to analyze your research question. Let's quickly refresh our memory on how to identify variables and data sets.

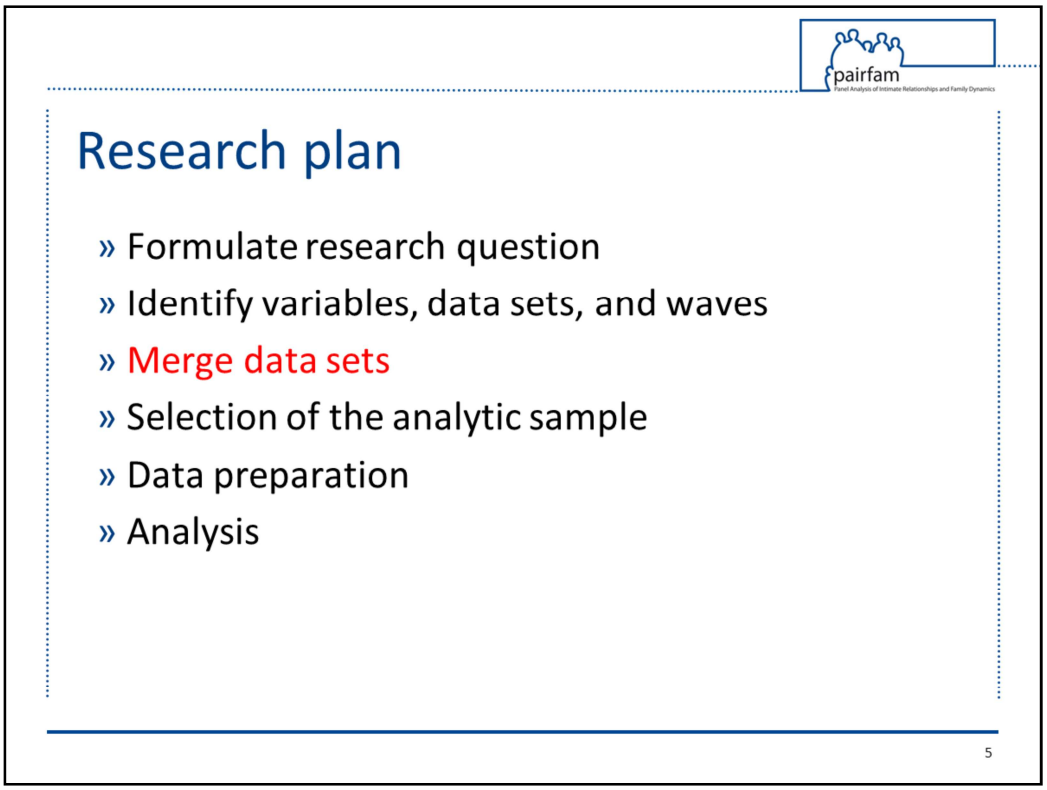


The Question Program summarizes which questions were asked in which waves for each respondent type. Please be aware that the survey variables for the different respondent types are stored on separate Excel sheets. The generated variables included in the anchor data are stored on a separate sheet as well. Once you have found the variables of interest for your research question, take a closer look at them in the codebooks.



The codebooks include complete information on question wording, answer possiblities, coding, filtering, and interviewer instructions. This information will help you further specify your research question, as well as in the data preparation process.

Don‘t forget to check out further documentation materials which I have presented in Tutorial No. 9 and which might help with your research.

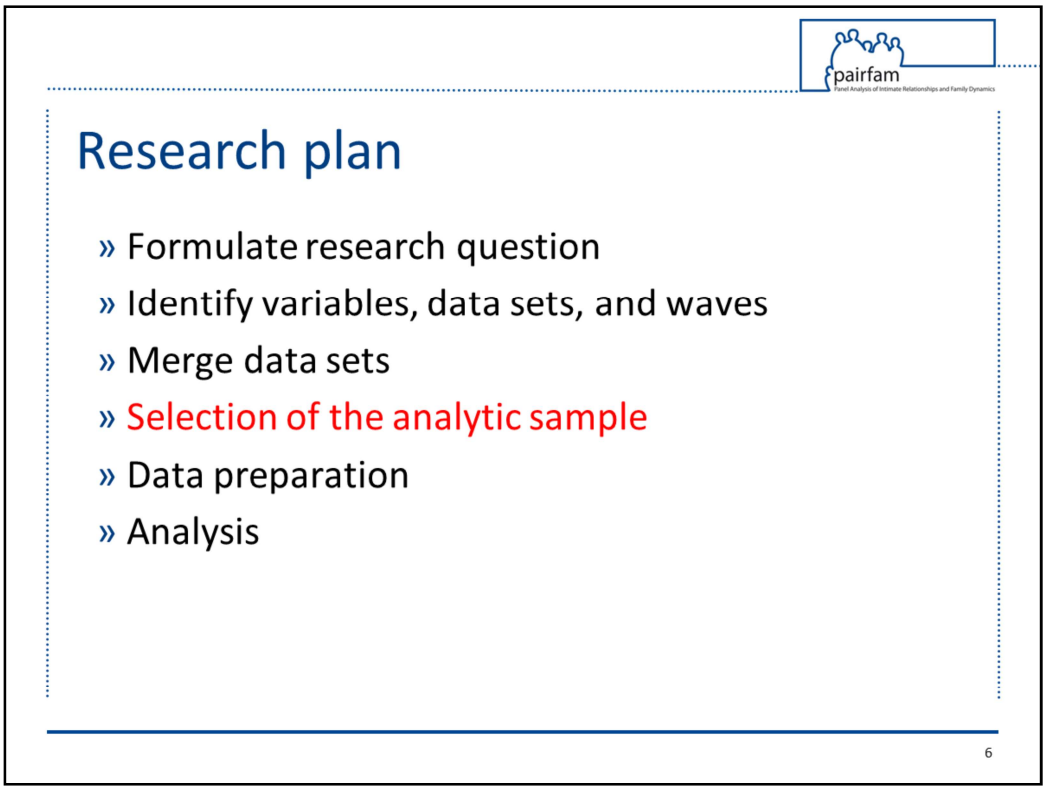


What's next?

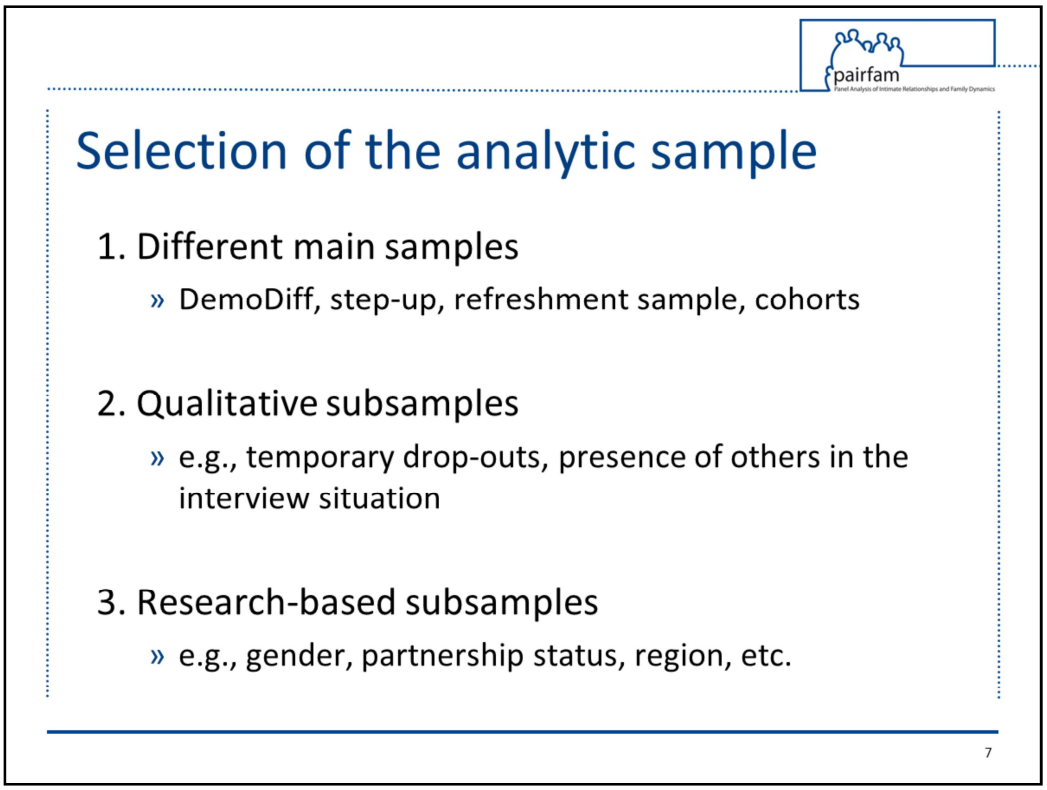
As the survey data is stored as separate cross-sectional data sets and organized by different actors, depending on your research question you will likely need to merge different actor data sets and/or several waves.

How you merge the data sets depends on your planned analysis method. The next tutorial will cover the merging of data sets in more detail.

For further examples on how to merge the different *pairfam* data sets, please consult the Quick Start "Matching".



After you have sucessfully merged the relevant data sets that cover your variables of interest, it‘s time to refine the analytic sample.

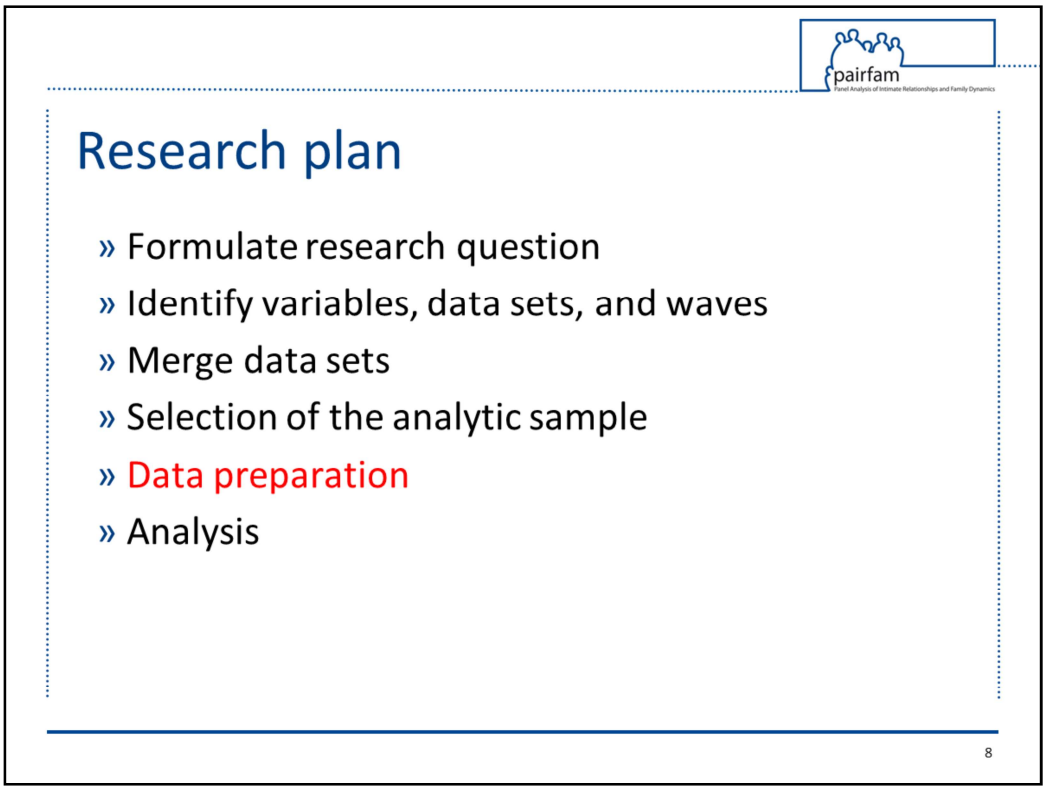


As the anchor data sets are comprised of different subsamples, you will first need to consider which samples you want to include (or exclude) from your analysis. Do you want to include the *DemoDiff*, *step-up*, and *refreshment* samples? Do you plan to run your analysis over all birth cohorts, or maybe separately for each cohort?

Next, you may want to exclude some qualitative subsamples, for example temporary drop-outs: respondents who skipped one wave. Furthermore, it might be relevant whether other people were present for the interview, especially if you are analyzing sensitive questions. For example, the question on satisfaction with one‘s relationship might be biased if the partner was present for the interview.

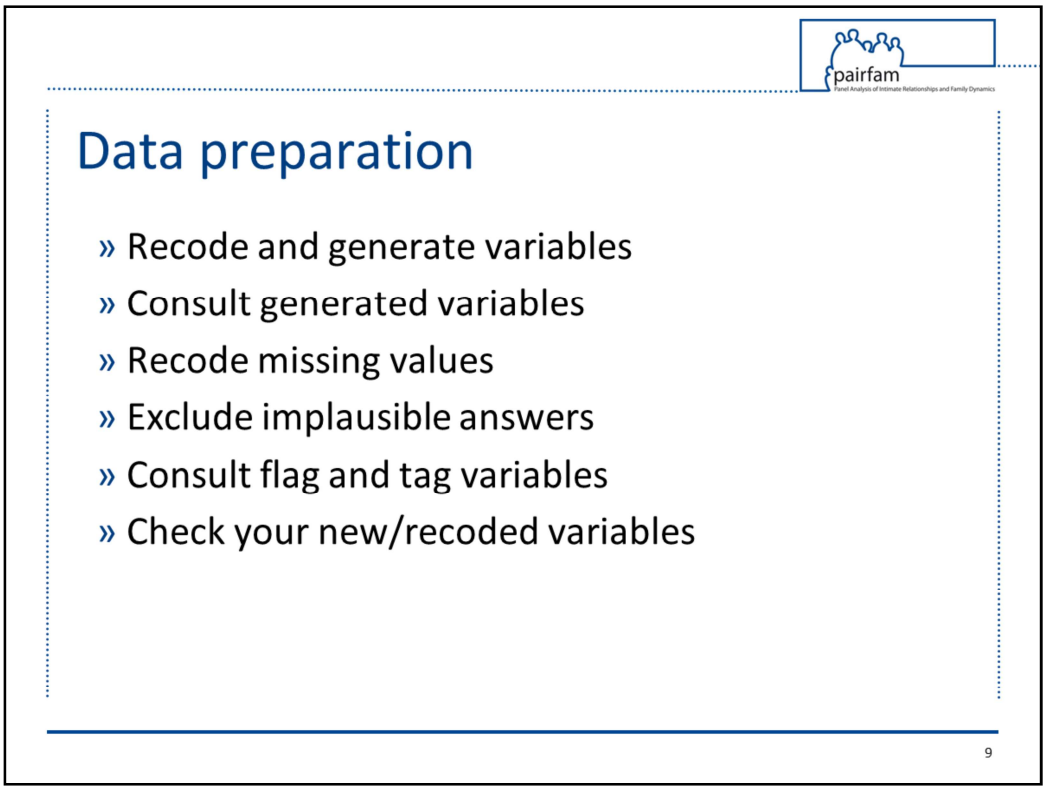
Finally, you may need to restrict your sample further according to your specific research question. Maybe only one gender is of interest, or only respondents in a relationship or living in a certain area.

For further examples on how to restrict the analytic sample, please refer to the Quick Start “Sample Definition“.



Are we there yet?

Almost. Before you can start with your analysis, you need to prepare your data.



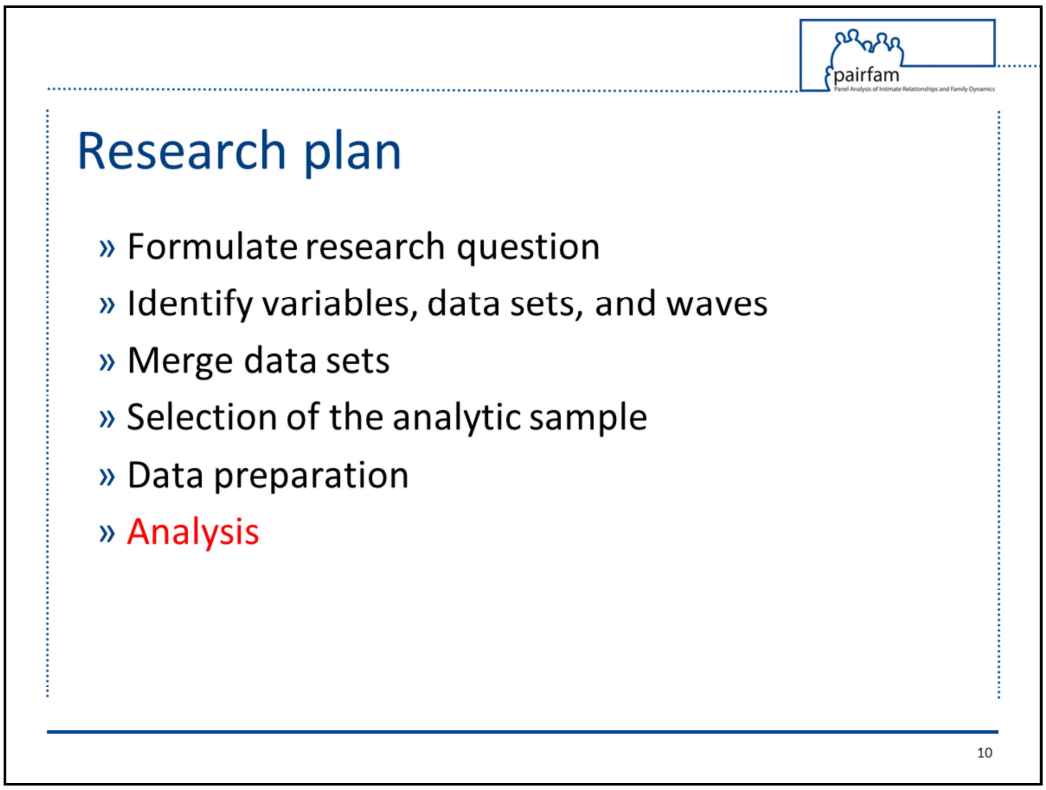
In other words, you likely need to recode some variables and/or generate new ones.

Don‘t forget to consult the list of generated variables, where you might find that we have already constructed a suitable variable for your research question. You can also use our syntax file to adapt one of the generated variables to your specific research question.

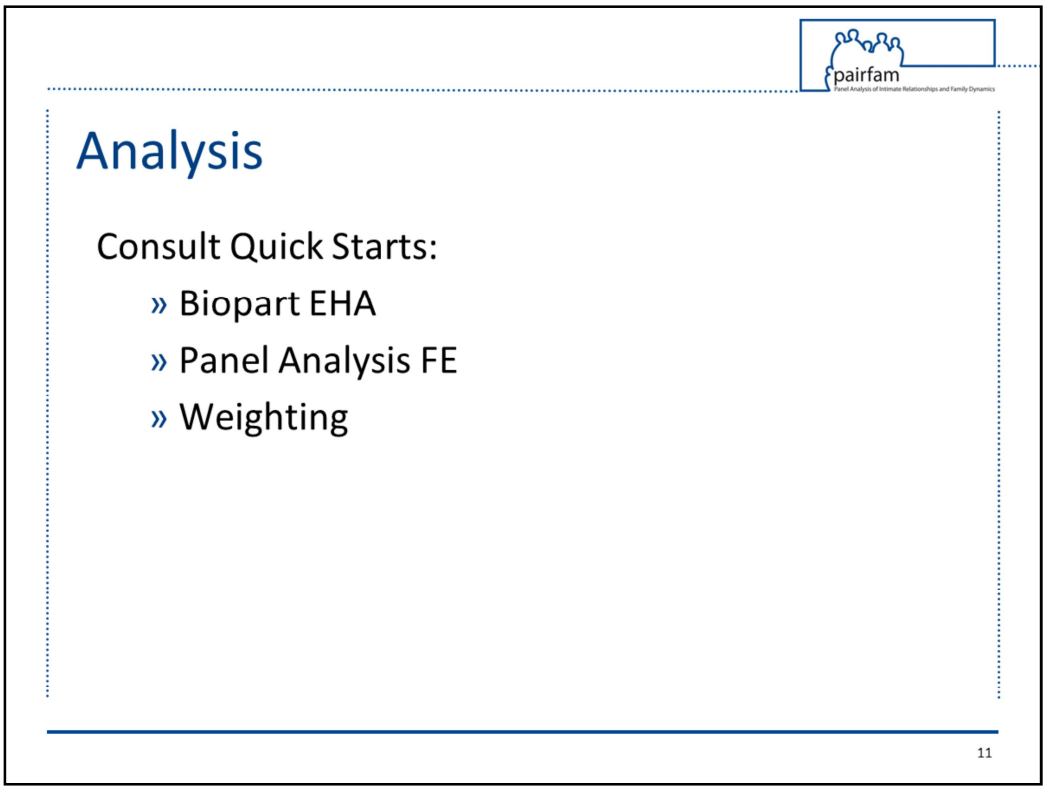
Furthermore, please be aware that missing values in the *pairfam* data are typically coded with negative numbers, and that you may need to recode these for your analysis. The Quick Start “Missings“ can help you with that.

You may also want to exclude implausible answers (such as working over 100 hours a week). The flag and tag variables can point you to some inconsistent entries.

Don‘t forget to check and tabulate your new and recoded variables!

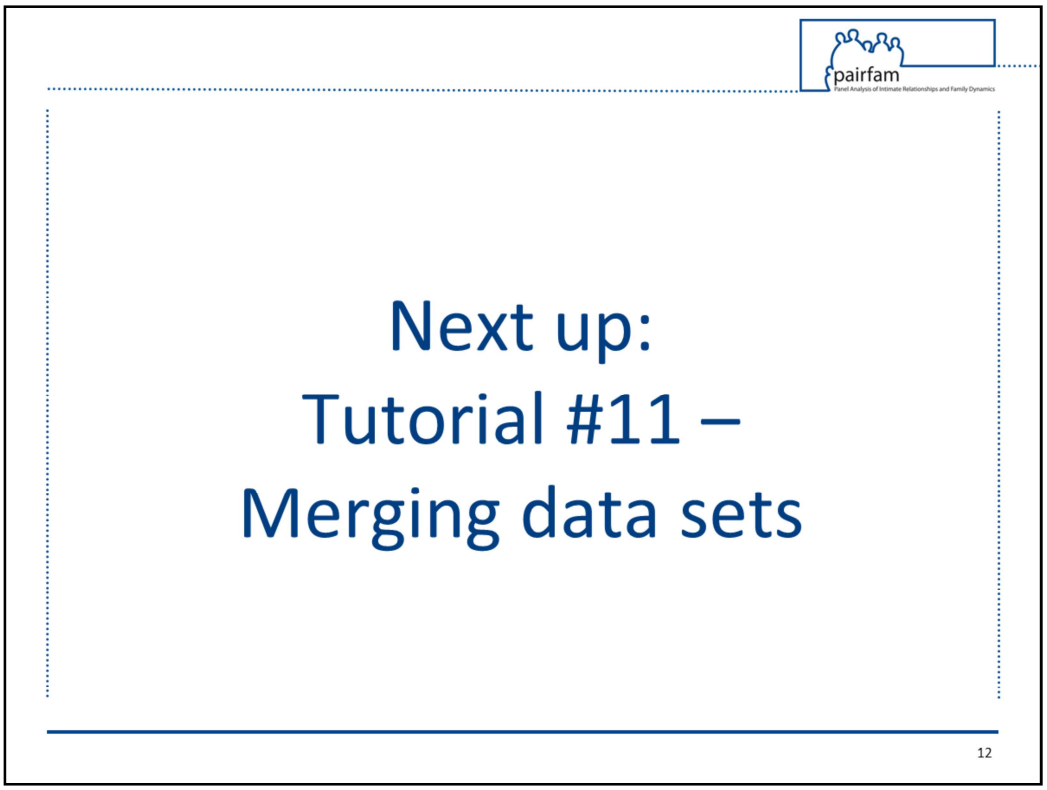


Now you can (finally) begin with your analysis.



We have compiled some examples of analysis methods to help you get started. The Quick Start “Biopart EHA“ shows an example of an event history analysis, and the Quick Start “Panel Analysis FE“ demonstrates fixed effects panel analysis.

Don‘t forget to consult the Quick Start “Weighting“ to help you decide whether to implement the weights provided in the pairfam data.



This concludes the tenth tutorial.

The next tutorial goes into more detail on how to merge different *pairfam* data sets.