

HaSpaD - Data Manual (September 2021)

Schulz, Sonja; Weiß, Bernd; Sterl, Sebastian; Haensch, Anna-Carolina;
Schmid, Lisa; May, Antonia

Veröffentlichungsversion / Published Version

Verzeichnis, Liste, Dokumentation / list

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Schulz, S., Weiß, B., Sterl, S., Haensch, A.-C., Schmid, L., & May, A. (2021). *HaSpaD - Data Manual (September 2021)*. (GESIS Papers, 2021/12). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.75135>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:
<https://creativecommons.org/licenses/by/4.0>

gesis

Leibniz-Institut
für Sozialwissenschaften

GESIS Papers

2021 | 12

HaSpaD - Data Manual

September 2021

*Sonja Schulz, Bernd Weiß,
Sebastian Sterl, Anna-Carolina Haensch,
Lisa Schmid & Antonia May*

Funded by German Research Foundation (DFG)

GESIS Papers 2021|12

HaSpaD - Data Manual

September 2021

*Sonja Schulz, Bernd Weiß, Sebastian Sterl,
Anna-Carolina Haensch, Lisa Schmid &
Antonia May*

Funded by German Research Foundation (DFG)

GESIS – Leibniz Institute for the Social Sciences 2021

GESIS Papers

GESIS – Leibniz Institute for the Social Sciences
Data Archive for the Social Sciences
International Surveys
Unter Sachsenhausen 6-8
50667 Köln

E-Mail: sonja.schulz@gesis.org

Student assistants

Dastan Jasim, Jan Heinicke, Franziska Kern, Marcel Klemm, Lea Kröner & Jan Siepermann

ISSN: 2364-3781 (Online)
Herausgeber,
Druck und Vertrieb: GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln

Table of contents

1	Quickstart	7
2	The Project HaSpaD.....	9
2.1	Overview of all used Survey programs.....	10
2.2	Data acquisition	19
2.3	Data citation.....	19
3	From the original data to the target data set – Instructions for the Stata-code	23
3.1	System requirements	23
3.2	Before running the main file main.do	24
3.3	During the execution of the main file main.do	28
3.4	After the execution of the main file main.do	28
4	Data structure	30
4.1	Identification of survey programs and sub-studies in the HaSpaD data set (data set ID).....	30
4.2	ID variables.....	32
4.3	Biography variables	33
4.4	Flag variables	34
4.5	Missing scheme	35
5	Special characteristics of the HaSpaD data due to the linkage of different data sources.....	37
5.1	Varying definitions of partnerships	37
5.2	Differences in the maximum number of reported partnerships	40
5.3	Varying observation periods for partnerships	42
5.4	Different reference population.....	49
5.5	Dealing with survey weights.....	50
5.6	Imputation of dates in the HaSpaD data set.....	52
5.7	Longitudinal processing of panel surveys without partner ID	53
6	Overview of the harmonized variables	58
6.1	Biography data.....	58
6.1.1	ID variables	58
6.1.2	Biography variables	62
6.1.3	Imputed variables	66
6.1.4	Original variables	70
6.1.5	Additional flag variables	73
6.2	Harmonized additional variables.....	74
7	References	75

A	Appendix	81
A. 1	Acquisition possibilities for the survey programs harmonized in HaSpaD.....	81
A. 2	The HaSpaD program code	86
A. 3	Documentations of the source data.....	91
A. 4	Degree of imputation of the biography variables.....	94
A. 5	Constructing the original variables of the biography data	98
A. 6	Values of the variable problem	106

Table of figures

Figure 1: System requirements	7
Figure 2: Folder structure before running the main-files	25
Figure 3: Instruction how to unpack the original data sets (for the program 7-Zip)	26
Figure 4: Folder structure after unpkacing the source data sets (extract)	27
Figure 5: Folder structure after the execution of the main-file (extract)	29
Figure 6: Procedure for linking equal partnerships.....	54
Figure A 1: Exemplary process of the creation of cumulative partnership biographies in Stata.....	87
Figure A 2: Folder structure after processing the biography data set	88
Figure A 3: Folder structure after processing the additional variables	90

List of tables

<i>Table 1:</i> Overview of the survey programs and sub-studies	11
<i>Table 2:</i> Citation guidelines of the harmonized survey programs	20
<i>Table 3:</i> Packages used in the main-file.....	23
<i>Table 4:</i> Notation of the survey programs in the HaspaD target data set	30
<i>Table 5:</i> Overview of the ID variables.....	32
<i>Table 6:</i> Overview of the biography variables	33
<i>Table 7:</i> Overview of the flag variables containing information about the degree of imputation for the biography variables	34
<i>Table 8:</i> Overview of the flag variables with original values	34
<i>Table 9:</i> Overview of further flag variables	35
<i>Table 10:</i> Missing scheme.....	35
<i>Table 11:</i> Degree of institutionalization per study	37
<i>Table 12:</i> Maximum number of reported partnerships per study.....	41
<i>Table 13:</i> Beginning of observation per study.....	43
<i>Table 14:</i> End of observation per study	45
<i>Table 15:</i> Imputed values of the beginning and end variables for seasonal dates	53
<i>Table 16:</i> Assumptions about plausibility for the identification of identical partnerships reported in wave 1 and 2	55
<i>Table 17:</i> Determination of upper and lower limits according to the combination of the level of imputation	56
<i>Table 18:</i> Example of the determination of the upper and lower limits after the combination of the degree of imputation	57
<i>Table 19:</i> Coding of the months given seasonal information	57

<i>Table 20:</i> ID variable: sid	58
<i>Table 21:</i> ID variable: did	59
<i>Table 22:</i> ID variable: aid	60
<i>Table 23:</i> ID variable: id	60
<i>Table 24:</i> ID variable: pid	61
<i>Table 25:</i> ID variable: pid_o	61
<i>Table 26:</i> ID variable: partner_pointer	61
<i>Table 27:</i> Biography variable: intd	62
<i>Table 28:</i> Biography variable: relbeg	62
<i>Table 29:</i> Biography variable: relend	63
<i>Table 30:</i> Biography variable: cohbeg	63
<i>Table 31:</i> Biography variable: cohend	64
<i>Table 32:</i> Biography variable: marbeg	64
<i>Table 33:</i> Biography variable: marend	65
<i>Table 34:</i> Biography variable: dop	66
<i>Table 35:</i> Imputed variable: flag_relbeg_imp	66
<i>Table 36:</i> Imputed variable: flag_relend_imp	67
<i>Table 37:</i> Imputed variable: flag_cohbeg_imp	67
<i>Table 38:</i> Imputed variable: flag_cohend_imp	68
<i>Table 39:</i> Imputed variable: flag_marbeg_imp	68
<i>Table 40:</i> Imputed Variable: flag_marend_imp	69
<i>Table 41:</i> Imputed Variable: flag_dop_imp	69
<i>Table 42:</i> Original variable: flag_relbeg_ori	70
<i>Table 43:</i> Original variable: flag_relend_ori	70
<i>Table 44:</i> Original variable: flag_cohbeg_ori	71
<i>Table 45:</i> Original variable: flag_cohend_ori	71
<i>Table 46:</i> Original variable: flag_marbeg_ori	71
<i>Table 47:</i> Original variable: flag_marend_ori	72
<i>Table 48:</i> Original variable: flag_dop_ori	72
<i>Table 49:</i> Additional flag variable: flag_beg	73
<i>Table 50:</i> Additional flag variable: problem	73
<i>Table A 1:</i> Acquisition possibilities for the survey programs harmonized in HaSpaD	81
<i>Table A 2:</i> Documentation of the source data	91
<i>Table A 3:</i> Degree of imputation of the biography variables	94
<i>Table A 4:</i> Constructing the original variables of the biography data	98
<i>Table A 5:</i> Values of the variable problem	106

1 Quickstart

The HaSpaD project facilitates the creation of an individualized and harmonized data set of German couple biographies from various source data sets. The harmonization processes are syntax based. This syntax can be selected and compiled via the HaSpaD Harmonization Wizard (<https://haspad.gesis.org/wizard>). This quickstart offers an overview of the necessary system requirements (see *Figure 1*) and serves as a step-by-step guide on how to proceed from the selection of the survey programs and additional variables to the HaSpaD target data set.

Required software	<ul style="list-style-type: none">• Stata 14 SE and newer• Data compression program (e.g. 7-Zip)
Required storage space	<ul style="list-style-type: none">• At least 15 GB
Miscellaneous	<ul style="list-style-type: none">• Internet access (for downloading additional Stata Packages, see section 3.1, Used packages)

Figure 1: System requirements

The following steps must be executed by the user:

1. Selection of the survey programs and additional variables using the HaSpaD Harmonization Wizard at <https://haspad.gesis.org/wizard>.
2. Download of the Zip-folder `haspad.zip` and unzipping the folder. The folder structure must be kept as follows:

```
haspad
|
+-- 1_documentation
|
+-- 2_source
|
+-- 3_harmonization
|
+-- 4_target
|
+-- main.do
|
+-- README.txt
```

When downloading, a date is inserted in the file name of the folder `haspad.zip` to avoid overwriting the folder. Users may want to change the name of the unzipped folder to "haspad". Otherwise, the path to the `haspad` folder in the main file must contain the entire path, including the dates. However, the unzipped file is referred to as "haspad" in the following.

3. Access to the original data sources through GESIS and/or the respective research data centres. In case that the data is issued via ZIP-folders: please do not unpack yet.
4. Saving of the original data sources or the ZIP-folder of the original data sources from step 3 in the directory `~haspad/2_source`.
5. In case that the original data sources were issued via ZIP-folders: Unpacking of the Zip-folder from step 4 in `~haspad/2_source`.
If problems emerge during the unpacking process, please note the suggestions in FAQ on the HaSpaD website.
6. Opening the do-file `main.do` in the folder `haspad`.
7. Insert the file path that indicates where the `haspad` folder has been stored on the user's repository in line 40 of the `main.do`.

Example: unpacked folder `haspad` was stored in `c:/mydata`
> 40 `cd "c:/mydata/haspad"`
8. Execution of the complete do-file `main.do`.
9. The target data set `target.dta` including the selected variables can be opened now via `~haspad/4_target`.

2 The Project HaSpaD

The project “Harmonizing and synthesizing partnership histories (HaSpaD)”, funded by the German Research Foundation (DFG), harmonizes and cumulates information on partnership histories from Germany’s largest survey programs.¹² By pooling data sources, the project pursues three primary goals: (1) By using all relevant German data sets available for secondary analyses, a vast database is provided for social science research syntheses. Said database enables more reliable insights into determinants of relationship events such as mate choice or separation and divorce as it would be possible given a single data source. (2) Based on this data cumulation and the resulting higher number of cases, yet unsolved questions regarding matrimonial and partnership stability across history and life courses, especially regarding the long-term increase of divorce rates and the social change of risk factors for separations, can be investigated. (3) As the set of examples for such vast research syntheses of heterogeneous data sources is quite narrow in the field of the survey based social sciences, the HaSpaD project as a best-practice-model attempts to motivate further activities in this area. The available survey data are heterogeneous in many ways, e.g. in terms of data structure, used weighting schemes or operationalisation of theoretical constructs.

The HaSpaD project offers a web-based tool, which facilitates access to individualized Stata-processing code for the harmonization of selected variables and survey programs. Users can select a variety of variables and survey programs via a web interface (HaSpaD Harmonization Wizard) at <https://haspad.gesis.org/wizard> and subsequently download a ZIP-file containing the adapted metajob (main-file as Stata do-file) as well as processing program-files for the creation of a cumulated data set. The main file, which performs the entire harmonization process, only needs to be slightly adjusted. It is important to note that it is the user’s responsibility to obtain the license agreements for the usage of the data of the particular survey programs (see *chapter 2.2*). This step is necessary, since the HaSpaD project offers no primary data itself but solely Stata-program code for the creation of a harmonized and cumulated data set. All rights for the primary data sources remain with the corresponding data producers.

The data set available for end-users (HaSpaD-target dataset³) includes:

- 1) **Partnership⁴ biographies** including variables on relationship, cohabitation, marriage and death of the partners, date of interview, besides further identification variables and flag-variables.
- 2) **Additional harmonized variables** stemming from the single original data sets, e.g. date of birth, sex, citizenship or level of education.

¹ DFG – project number 316901171 (<https://gepris.dfg.de/gepris/projekt/316901171>)

² Harmonization of measurements means that one or more source variables of an original data set are recoded to a target variable by means of defined transformation rules. While the values of the variables from the source data sources may differ, those of the target variables are identical across each source. Wolf et al (2016, p. 514) describe a harmonization as follows: “The outcomes of this process are thus – in the optimal case – the assembled source variables from different data sources, their mapping or recode to the comparable coding scheme, as well as the newly constructed, detailed harmonized variable.”

³ The following terms are used synonymously: target data set, HaSpaD data.

⁴ A partnership here refers to an intimate relationship with the partners, which can consist of the degrees of institutionalization/institutionalization phases (synonymously: type of partnership, partnership phase, degree of partnership) 1. relationship phase (beginning and end of relationship), 2. cohabitation phase (beginning and end of cohabitation) and 3. marriage phase (beginning and end of marriage). The term "relationship" is therefore only one part of a partnership and will not be used synonymously with the partnership in the following.

The units of analysis for the target dataset are the respondent's partnerships. An instruction for the creation of the target data set can be found in the Quickstart section. Furthermore, an extensive instruction is presented in *chapter 3*.

Researchers who use the HaSpad-data are required to cite all used original data sources as well as the HaSpaD project (see notes on citation in *chapter 2.3*). Citing only the original data sources is not sufficient. Even though the HaSpaD project offers no primary data, its effort to harmonize and cumulate the data sources represents a scientific merit which must be cited. In the following chapter, the survey programs harmonized by the HaSpaD project are introduced.

2.1 Overview of all used Survey programs

In total, nine different survey programs containing longitudinal information on partnership and marriage biographies are included in the project. The population of all survey programs consists of residents within Germany whose data are collected through random samples (or who are recurrently interviewed after they had been picked randomly the first time)⁵. It has to be taken into consideration that certain studies follow a cross-sectional design, e.g. ALLBUS (GESIS 2018), Mannheim Divorce Study (Esser et al. 2018a) and Fertility and Family Survey (Bundesinstitut für Bevölkerungsforschung 2002a), while others follow a panel design, e.g. pairfam (Brüderl et al. 2021a), the Socio-Economic Panel (Liebig et al. 2019) and SHARE (SHARE Waves 1, 2 and 3). A middle ground is occupied by the Family Surveys (DJI 1992a, 2018a, 2018b, 2018c), the Life History Studies (Mayer 1995a, 1995b, 2004a, 2004b, 2014a, 2018a, 2018b, 2018c; Mayer & Kleinhenz 2004a) and the Generations and Gender Survey (GGs Wave 1; GGs Wave 2), in which a part of the respondents of the initial survey are recurrently interviewed in later waves. Furthermore, pairfam, SHARE and the Socio-Economic Panel have a multi-actor-design: As soon as the partners start to live with the respondents (anchor person) in a common household, they are included in the sample and report about partnership histories, too. For the pairfam study, the permission of the anchor person is requested previously to interviewing the partner. Hereby, it does not matter whether the anchor person lives in a cohabitation with the partner or not (Brüderl et al. 2021c, p. 5) presents an overview of the survey programs, their sub-studies and survey design. Moreover, the methodological report for each sub-study is cited, which provides further information about the methodological implementation of the respective study.

⁵ An exception is the Mannheim Divorce Study. Here, a screening procedure was initially carried out within the framework of other, randomly based surveys by the GFM-GETAS Institute. This was followed by a disproportionate stratified sampling based on the stratification characteristics "divorce of first marriage", "sex" and "place of residence" (Esser, Gostomski & Hartmann 2018c).

Table 1: Overview of the survey programs and sub-studies

Study title	Year of survey	Total number of respondents/ total number of respondents with partnership biography	Population	Sample	Survey design	Documentation DOI
Panel Analysis of Intimate Relationships and Family Dynamics (pairfam), Data Release 12.0	2008-2020, annual survey	18,912 / 16,191	Region <i>Main sample</i> Individuals in private households in Germany <i>Refreshment sample</i> Individuals in private households in entire Germany <i>Additional sample DemoDiff</i> Individuals in private households in East Germany Citizenship Germany and foreigners (German speaking) Sex Women and men Age / Birth cohort <i>Main sample</i> Birth cohorts 1971-1973, 1981-1983, 1991-1993 <i>Refreshment sample</i> Birth cohorts 1981-1983, 1991-1993, 2001-2003 <i>Additional sample DemoDiff</i> 1971-73, 1981-83	Number of Samples Three samples: Main sample, additional sample DemoDiff and Refresher sample Sample design Two-stage random sampling	Panel Multi-Actor-Design	Brüderl et al. 2021b DOI: 10.4232/pairfam.5678.12.0.0
ALLBUS/GGSS 1980-2016 (Kumulierte)	1980, 1982, 1984, 1986,	24,786 (Only survey years including the	Region	Number of Samples One sample per survey year	Cumulative Cross-Section	Baumann & Schulz 2018 DOI: 10.4232/1.13029

Study title	Year of survey	Total number of respondents/ total number of respondents with partnership biography	Population	Sample	Survey design	Documentation DOI
Allgemeine Bevölkerungsumfrage der Sozialwissenschaften / Cumulated German General Social Survey 1980-2016)	1988, 1991, 2000, 2010	collection of marriage biographies) / 19,498	Private households within West Germany (FRG) up to and including 1990, Unified Germany since 1991 Citizenship German and foreigner (German speaking) Sex Female and male Age / Birth Cohort 18 years and older	Sample Design Multilevel random sample (ADM) 1980-1992/1992 Two-level sample 1994, 1996, since 2000	Single-Actor-Design	
Change and Development of Forms of Family Life in West Germany (Survey of Families)	1988	10,043 / 4,562	Region Private households within West Germany (FRG) Citizenship German Sex Female and male Age / Birth Cohort 18-55 years	Number of Samples Three subsamples Sample Design <i>Subsample 1</i> Three-level random sample (ADM) <i>Subsample 2</i> addresses-Random <i>Subsample 3</i> Follow-up survey of a quota sample from the pretest	Cross-Section, partially Panel Single-Actor-Design	Eschenlohr et al. 1989 DOI: 10.4232/1.2245
Family and Partner Relations in Eastern Germany (Survey of Families)	1991	1,951 / 1,778	Region East Germany Citizenship German Sex Female and male Age / Birth Cohort 18-55 years	Number of Samples One sample Sample Design Simple random sample	Cross-Section Single-Actor-Design	Haas 1991 DOI: 10.4232/1.13196

Study title	Year of survey	Total number of respondents/ total number of respondents with partnership biography	Population	Sample	Survey design	Documentation DOI
Change and Development of Ways of Family Life - 2nd Wave (Survey of Families)	1994	10,994 / 8,215	Region Unified Germany Citizenship German Sex Female and male Age / Birth Cohort 18-61 years	Number of Samples Three subsamples Sample Design <i>Subsample 1</i> Panel West: Follow-up survey of respondents of Family Survey 1.1 <i>Subsample 2</i> Survey Ost – unsystematic random sample of 18-55 years old residents (random route) <i>Subsample 3</i> 18-30 years West: unsystematic random sample of 18-30 years old residents (random route)	Cross-Section, partially follow-up survey Single-Actor-Design	Albers et al. 1994 DOI: 10.4232/1.13197
Change and Development of Families ´ Way of Life - 3rd Wave (Family Survey)	2000	10,318 / 8,862	Region Unified Germany Citizenship German Sex Female and male Age / Birth Cohort 18-67 years	Number of Samples Two subsamples Sample Design <i>Subsample 1</i> Cross-section sample of 18-55 years old residents <i>Subsample 2</i> Panel West: Follow-up survey of panel participants of the Family Survey 1.1 and 2	Cross-Section, partially follow-up survey Single-Actor-Design	Schneekloth 2000 DOI: 10.4232/1.13198

Study title	Year of survey	Total number of respondents/ total number of respondents with partnership biography	Population	Sample	Survey design	Documentation DOI
Mannheim Divorce Study 1996	1996	5,020 / 5,020	Region Unified Germany Citizenship German and foreigner Sex Female and male Age / Birth Cohort 18 years and older	Number of Samples One sample Sample Design Disproportionally stratified sample	Cross-Section Single-Actor-Design	Esser, Gostomski & Hartmann 2018c DOI: 10.4232/1.13056
German Fertility and Family Survey 1992	1992	10,012 / 6,555	Region Unified Germany Citizenship German Sex Female and male Age / Birth Cohort 20-39 years	Number of Samples One sample Sample Design Multilevel random sample (ADM and FGW)	Cross-Section Single-Actor-Design	Pohl 1995 DOI: 10.4232/1.3400
Courses of Life and Social Change: Courses of Life and Welfare Development (Life History Study LV-West I)	1981-1983	2,171 / 2,002	Region West Germany (FRG) Citizenship German Sex Female and male Age / Birth Cohort Birth cohorts 1929-31, 1939-41, 1949-51	Number of Samples One sample Sample Design Multilevel random sample (ADM)	Cross-Section Single-Actor-Design	Mayer et al. 1989 DOI: 10.4232/1.13193
Courses of Life and Social Change: The Between-the-War Cohort in Transition to Retirement (Life	1985-1986	407 / 395	Region Residents in West Germany (FRG) Citizenship German Sex Female and male	Number of Samples One Sample Sample Design Multilevel random sample (ADM)	Cross-Section Single-Actor-Design	Brückner 1993 DOI: 10.4232/1.13194

Study title	Year of survey	Total number of respondents/ total number of respondents with partnership biography	Population	Sample	Survey design	Documentation DOI
History Study LV-West II A - Personal Interview)			Age / Birth Cohort Birth cohort 1919-21			
Courses of Life and Social Change: The Between-the-War Cohort in Transition to Retirement (Life History Study LV-West II T - Telephone Interview)	1987-1988	1,005 / 948	Region West Germany (FRG) Citizenship German Sex Female and male Age / Birth Cohort Birth cohort 1919-21	Number of Samples One sample Sample Design Simple random sample	Cross-Section Single-Actor-Design	Brückner 1993 DOI: 10.4232/1.2647
Courses of Life and Social Change: Access to Occupation in Employment Crisis (Life History Study LV-West III)	1988-1989	2,008 / 1,521	Region West Germany (FRG) Citizenship German Sex Female and male Age / Birth Cohort Birth cohort 1954-56, 1959-61	Number of Samples One sample Sample Design Simple random sample	Cross-Section Single-Actor-Design	Brückner et al. 1995 DOI: 10.4232/1.2648
Courses of Life and Historical Change in East Germany (Life History Study LV DDR)	1991-1992	2,331 / 2,246	Region East Germany (former GDR) Citizenship German Sex Female and male Age / Birth Cohort	Number of Samples One sample Sample Design Simple random sample	Cross-Section Single-Actor-Design	Hess & Smid 1991 DOI: 10.4232/1.2644

Study title	Year of survey	Total number of respondents/ total number of respondents with partnership biography	Population	Sample	Survey design	Documentation DOI
East German Life Courses After Unification (Life History Study LV Ost 71)	1996-1997	610 / 534	Birth cohort 1929-31, 1939-41, 1951-53, 1959-61 Region East Germany (former GDR) Citizenship German Sex Female and male Age / Birth Cohort Birth cohort 1971	Number of Samples One sample Sample Design Simple random sample	Cross-Section, partially Panel Single-Actor-Design	Matthes, Lichtwardt & Mayer 2004 DOI: 10.4232/1.3926
East German Life Courses After Unification (Life History Study LV-Ost Panel) - Birth cohorts 1939-41, 1951-53, 1959-61	1996-1997	1,078 / 995	Region East Germany (former GDR) Citizenship German Sex Female and male Age / Birth Cohort Birth cohort 1939-41, 1951-53, 1959-61	Number of Samples One sample Sample Design Follow-up survey for respondents of Life History Study 4	Follow-up survey for respondents of Life History Study (GDR) Single-Actor-Design	Goedicke, Lichtwardt & Mayer 2004 DOI: 10.4232/1.3925
East German Life Courses After Unification (Life History Study LV-Ost Panel) - Birth cohort 1929-31	1996-1997	329 / 321	Region East Germany (former GDR) Citizenship German Sex Female and male Age / Birth Cohort Birth cohort 1929-31	Number of Samples One sample Sample Design Follow-up survey for respondents of Life History Study 4	Follow-up survey for respondents of Life History Study (GDR) Single-Actor-Design	Goedicke, Lichtwardt & Mayer 2004 DOI: 10.4232/1.3925
Education Training and Occupation: Life Courses of the 1964 and 1971 Birth	1998-1999	2,909 / 2,583	Region West Germany Citizenship German and foreigner	Number of Samples One sample Sample Design	Cross-Section, partially Panel	Hillmert et al. 2004 DOI: 10.4232/1.3927

Study title	Year of survey	Total number of respondents/ total number of respondents with partnership biography	Population	Sample	Survey design	Documentation DOI
Cohorts in West Germany (Life History Study LV-West 64/71) Early Careers and Starting a Family: Life Courses of the 1971 Birth Cohorts in East and West Germany (Life History Study LV-Panel 71)	2004-2005	1,069 / 971	Sex Female and male Age / Birth Cohort Birth cohort 1964 and 1971 Region East Germany (former GDR) and West Germany Citizenship German Sex Female and male Age / Birth Cohort Birth cohort 1971	Single-level random sample Number of Samples Two subsamples Sample Design <i>Subsample 1</i> Sample East: Follow-up survey for respondents from LV-Ost 71) <i>Subsample 2</i> Sample West Follow-up survey for respondents from LV-West 64/71, Cohort 71	Single-Actor-Design Follow-up survey for LV-Ost 71 and LV-West 64/71, Cohort 71 Single-Actor-Design	Mayer 2014b DOI: 10.4232/1.5099
Gender and Generations Survey (2005) (Subsample Germany)	Main survey 2005	10,017 / 8,168	Region Residents within unified Germany Citizenship German and foreigner Sex Female and male Age / Birth Cohort 18-79 years	Number of Samples One sample Sample Design Three-level random sample (ADM)	Cross-Section, partially Panel Single-Actor-Design	Fokkema et al. 2016 DOI: 10.14301/lcs.v9i4.500
Gender and Generations Survey (2008) (Subsample Germany)	Main survey 2008	3,226 / 2,637	Region Residents within unified Germany Citizenship German and foreigner Sex Female and male	Number of Samples One sample Sample Design Follow-up survey for respondents from	Follow-up survey for GGS (2005) Single-Actor-Design	Fokkema et al. 2016 DOI: 10.14301/lcs.v9i4.500

Study title	Year of survey	Total number of respondents/ total number of respondents with partnership biography	Population	Sample	Survey design	Documentation DOI
The Survey of Health, Ageing and Retirement in Europe (SHARE), SHARE Waves 1, 2 and 3	2008	1,921/ 1,379	Age / Birth Cohort 20-83 years Region Unified Germany Citizenship German and foreigner Sex Female and male Age / Birth Cohort Anchor person 50 years or older; spouse can be younger than 50	Generations and Gender Survey (GGS) 1 Number of Samples One sample per wave Sample Design Multilevel sample	Panel Multi-Actor-Design	Klevmarken, Hesselius & Swensson 2005 DOIs: 10.6103/SHARE.w1.600 , 10.6103/SHARE.w2.600 , 10.6103/SHARE.w3.600
Socio-Economic Panel (SOEP), data from 1984-2018	1984-2018, annual survey	142,308/ 34,410	Region Residents within unified Germany Citizenship German and foreigner Sex Female and male Age / Birth Cohort 17 years or older	Number of Samples Several samples per wave possible (refreshment samples, supplementary samples (e.g. immigrants, high-income stratum) Sample Design Multilevel sample	Panel Multi-Actor-Design	Siegers et al. 2020 DOI: 10.5684/soep-core.v35

2.2 Data acquisition

After the selection of the survey programs and the additional variables using the Harmonization Wizard, user must acquire the required source data sets. Since, data processing routines are only available in Stata, the data sources must be downloaded as Stata-files with the ending `*.dta`. The processing routines have been created with specific versions of the source data sets, thus the whole process relies on **the identical version** of the original data as harmonized by the HaSpaD project.

In order to ensure the sustainable usability, the HaSpaD project only considers data sources that are distributed via research data repositories as Scientific-Use-File or Public-Use-File. In the appendix, *Table A 1* gives an overview of the research data infrastructures and data distribution by survey programs. Further information is available at the respective homepage of the research data center or of the data distributor. For research and acquisition of the data sources the following should be considered:

- Data sets that are distributed via GESIS are only available in most recent versions. In case that the data version used in the HaSpaD project has been updated in the meantime, the required version (see *Table A 1* in the appendix, column “required version”) has to be requested at dataservices@gesis.org.
- The survey programs pairfam, Mannheim Divorce Study and the Life History Studies are classified as access class C (C-studies, effective: 24.11.2019) and therefore, require a fee. A list of prices is available at <https://www.gesis.org/en/services/archiving-and-registering/data-archiving> under “documents/price list”.
- GESIS Cryptshare provides the option to download all offered C-studies together in one meta-ZIP-file (name `cs-transfer.zip`). In consequence, the user first must unpack the meta-ZIP-file to subsequently save and unpack the contained ZIP-files in the folder `~haspad/2_source`. Please note, data of the Socio-Economic Panel (SOEP) are also delivered as `cs-transfer.zip`. In the case of SOEP data, however, only unpack and save the folder `cs-transfer` in `~haspad/2_source`, and do not save the containing data folders within `2_source`.
- The Generations and Gender Survey (GGS) only provides the download of the most recent version. Thus, the data source used by the HaSpaD project should be requested at ggp@nidi.nl, in case that a newer version is provided for the download.
- Likewise, it is important to revise the version for the SHARE-data. At the download section of the SHARE research data center, an item called “Archiv, Go to archived data” below the download matrix of the newest version provides access to older versions.
- Lastly, when requesting the data of the Socio-Economic Panel (SOEP), please contact soepmail@diw.de to request an earlier version.

2.3 Data citation

Publications based on the HaSpaD project must cite the used original data sets as well as the HaSpaD project. Citing only the original data sources is not sufficient. Even though the HaSpaD project offers no primary data, its effort to harmonize and cumulate the data sources represents a scientific merit which must be cited.

It is the responsibility of the user to use the correct citations. The most recent form of citation should be found at the respective homepage of the research data center or data distributor. At the time of the publication of this data manual, all forms of citations listed in *Table 2* were valid.

The HaSpaD project is cited in the following way:

Schulz, Sonja; Weiß, Bernd; Sterl, Sebastian; Haensch, Anna-Carolina; Schmid, Lisa; May, Antonia (2021): *HaSpaD - Harmonizing and synthesizing partnership histories from different research data infrastructures: A model project for linking research data from various infrastructures. Processing files for the creation of a couple biography data set*. Version 1.0.0. GESIS – Leibniz Institute for the Social Sciences, Cologne and Mannheim. <https://doi.org/10.7802/2317>

The HaSpaD data manual is cited in the following way:

Schulz, S., Weiß, B., Sterl, S., Haensch, A.-C., Schmid, L. & May, A. (2021). *HaSpaD – Data Manual, Version 1*. (GESIS Papers, 2021/12). Cologne: GESIS – Leibniz Institute for the Social Sciences. <https://doi.org/10.21241/ss0ar.75135>

Table 2: Citation guidelines of the harmonized survey programs

Survey program	Data citation
pairfam	Brüderl, Josef; Drobnič, Sonja; Hank, Karsten; Neyer, Franz. J.; Walper, Sabine; Alt, Philipp; Borschel, Elisabeth; Bozoyan, Christiane; Garrett, Madison; Geissler, Svenja; Gonzalez Avilés, Tita; Gröpler, Nicolai; Hajek, Kristin; Herzig, Michel; Huyer-May, Bernadette; Lenke, Rüdiger; Lorenz, Renate; Lutz, Katharina; Minkus, Lara; Peter, Timo; Phan, Trang; Preetz, Richard; Reim, Julia; Sawatzki, Barbara; Schmiedeberg, Claudia; Schütze, Philipp; Schumann, Nina; Thönnissen, Carolin; Timmermann, Katharina; Wetzel, Martin (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
ALLBUS	GESIS - Leibniz-Institut für Sozialwissenschaften (2018). ALLBUS/GGSS 1980-2016 (Kumulierte Allgemeine Bevölkerungsumfrage der Sozialwissenschaften / Cumulated German General Social Survey 1980-2016). GESIS Data Archive, Cologne. ZA4586 Data file Version 1.0.0, doi:10.4232/1.13029
Family Surveys	Deutsches Jugendinstitut (DJI), München (1992). Change and Development of Forms of Family Life in West Germany (Survey of Families). GESIS Data Archive, Cologne. ZA2245 Data file Version 1.0.0., doi:10.4232/1.2245 Deutsches Jugendinstitut (DJI), München (2018). Family and Partner Relations in Eastern Germany (Survey of Families). GESIS Data Archive, Cologne. ZA2392 Data file Version 1.1.0, doi:10.4232/1.13196 Deutsches Jugendinstitut (DJI), München (2018). Change and Development of Ways of Family Life - 2nd Wave (Survey of Families). GESIS Data Archive, Cologne. ZA2860 Data file Version 1.1.0, doi:10.4232/1.13197 Deutsches Jugendinstitut (DJI), München (2018). Change and Development of Families` Way of Life - 3rd Wave (Family Survey). GESIS Data Archive, Cologne. ZA3920 Data file Version 1.1.0, doi:10.4232/1.13198
Mannheim Divorce Study 1996	Esser, Hartmut, Gostomski, Christian Babka von, & Hartmann, Josef (2018). Mannheim Divorce Study 1996. GESIS Data Archive, Cologne. ZA3188 Data file Version 2.0.0, doi:10.4232/1.13056

Survey program	Data citation
German Fertility and Family Survey 1992	Bundesinstitut für Bevölkerungsforschung, Wiesbaden (2002). German Fertility and Family Survey 1992. GESIS Data Archive, Cologne. ZA3400 Data file Version 1.0.0, doi:10.4232/1.3400
Life History Studies	<p>Mayer, Karl U. (2018). Courses of Life and Social Change: Courses of Life and Welfare Development (Life History Study LV-West I). GESIS Data Archive, Cologne. ZA2645 Data file Version 1.1.0, doi:10.4232/1.13193</p> <p>Mayer, Karl U. (2018). Courses of Life and Social Change: Courses of Life and Welfare Development (Life History Study LV-West I) . GESIS Data Archive, Cologne. ZA2645 Data file Version 1.1.0, doi:10.4232/1.13194</p> <p>Mayer, Karl U. (1995). Courses of Life and Social Change: The Between-the-War Cohort in Transition to Retirement (Life History Study LV-West II T - Telephone Interview) . GESIS Data Archive, Cologne. ZA2647 Data file Version 1.0.0, doi:10.4232/1.2647</p> <p>Mayer, Karl U. (2018). Courses of Life and Social Change: Access to Occupation in Employment Crisis (Life History Study LV-West III). GESIS Data Archive, Cologne. ZA2648 Data file Version 1.1.0, doi:10.4232/1.13195</p> <p>Mayer, Karl U. (1995). Courses of Life and Historical Change in East Germany (Life History Study LV DDR). GESIS Data Archive, Cologne. ZA2644 Data file Version 1.0.0, doi:10.4232/1.2644</p> <p>Mayer, Karl U. (2004). East German Life Courses After Unification (Life History Study LV Ost 71). GESIS Data Archive, Cologne. ZA3926 Data file Version 1.0.0, doi:10.4232/1.3926</p> <p>Mayer, Karl U. (2004). East German Life Courses After Unification (Life History Study LV-Ost Panel). GESIS Data Archive, Cologne. ZA3925 Data file Version 1.0.0, doi:10.4232/1.3925</p> <p>Mayer, Karl U., & Kleinhenz, Gerhard (2004). Education, Training, and Occupation: Life Courses of the 1964 and 1971 Birth Cohorts in West Germany (Life History Study LV-West 64/71). GESIS Data Archive, Cologne. ZA3927 Data file Version 1.0.0, doi:10.4232/1.3927</p> <p>Mayer, Karl U. (2014). Early Careers and Starting a Family: Life Courses of the 1971 Birth Cohorts in East and West Germany (Life History Study LV-Panel 71). GESIS Data Archive, Cologne. ZA5099 Data file Version 1.0.0, doi:10.4232/1.5099</p>
Generations and Gender Survey (GGS)	<p><i>Generations and Gender Survey (2005) (Subsample Germany)</i>. This paper uses data from the GGS Waves 1, and 2 (DOIs: 10.17026/dans-z5z-xn8g, 10.17026/dans-xm6-a262), see Gauthier, A. H. et al. (2018) or visit the GGP website (https://www.ggp-i.org/) for methodological details. Working Version: GGS Wave 1 Version 4.3. – Consolidated identification: GGSW1.Cons.</p> <p><i>Generations and Gender Survey (2008) (Subsample Germany)</i>. This paper uses data from the GGS Waves 1, and 2 (DOIs: 10.17026/dans-z5z-xn8g, 10.17026/dans-xm6-a262), see Gauthier, A. H. et al. (2018) or visit the GGP website (https://www.ggp-i.org/) for methodological details. Working Version: GGS Wave 2 Version 1.3. – Consolidated identification: GGSW2.Cons.</p>
SHARE	<p>The Survey of Health, Ageing and Retirement in Europe (SHARE) /</p> <p><i>SHARE Waves 1, 2 and 3</i>. This paper uses data from SHARE Waves 1, 2 and 3 (DOIs: 10.6103/SHARE.w1.600, 10.6103/SHARE.w2.600, 10.6103/SHARE.w3.600), see Börsch-Supan et al. (2013) for methodological details.(1)</p> <p><i>The SHARE data collection has been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP:</i></p>

Survey program	Data citation
Socio-Economic Panel (SOEP)	<p data-bbox="523 353 1390 568"><i>GA N°227822, SHARE M4: GA N°261982) and Horizon 2020 (SHARE-DEV3: GA N°676536, SERISS: GA N°654221) and by DG Employment, Social Affairs & Inclusion. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged(see www.share-project.org).</i></p> <p data-bbox="523 584 1390 674">(1) Usually this information will be part of the text of a scientific publication already. If this is the case, there is no need to include this information in the acknowledgement once more.</p> <ul data-bbox="571 689 1390 1084" style="list-style-type: none"> <li data-bbox="571 689 1390 779">• Börsch-Supan, A. (2017). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1. Release version: 6.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w1.600 <li data-bbox="571 781 1390 871">• Börsch-Supan, A. (2017). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w2.600 <li data-bbox="571 873 1390 963">• Börsch-Supan, A. (2017). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 3 – <i>SHARELIFE</i>. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w3.600 <li data-bbox="571 965 1390 1084">• Börsch-Supan, A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck, S. Zuber (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). <i>International Journal of Epidemiology</i>. DOI: 10.1093/ije/dyt088 <p data-bbox="523 1133 1390 1308">Stefan Liebig, Jan Goebel, Martin Kroh, Carsten Schröder, Markus Grabka, Jürgen Schupp, Charlotte Bartels, Alexandra Fedorets, Andreas Franken, Jannes Jacobsen, Selin Kara, Peter Krause, Hannes Kröger, Maria Metzging, David Richter, Diana Schacht, Paul Schmelzer, Christian Schmitt, Daniel Schnitzlein, Rainer Siegers, Knut Wenzig, Stefan Zimmermann (2019). Socio-Economic Panel (SOEP), data for years 1984-2018, version 35, SOEP, 2019, doi:10.5684/soep.v35.</p>

3 From the original data to the target data set – Instructions for the Stata-code

The HaSpaD project offers syntax-based harmonization processes. After the selection of the survey programs and the variables at <https://haspad.gesis.org/wizard>, a .zip-file containing the harmonization syntax is provided. This file contains all necessary files including the main-file carrying out the harmonization tasks as well as documentation material and the HaSpaD folder structure. After downloading and unzipping this file, users need to acquire the original source data on their own behalf (see *chapter 2.2*) and save these data sets with their original file designation in the `~haspad/2_source`-folder within the downloaded HaSpaD folder structure. This chapter guides through all necessary steps: from system requirements through necessary preparations prior to the execution of the main-file in Stata to the final target data set.

3.1 System requirements

Used Stata version

For the creation of the single processing code-files, Stata SE 15.1 was used; the HaSpaD-code of the main-file works with any version starting from version 14.0. Older Stata-versions cannot process the Stata-code because some functions (e.g. the transformation of the data set into the unicode-format) are only supported in more recent versions. In case, newer versions are used, the version number can be adapted in the `main.do` (command line 32) to the Stata version used. However, to use versions up to 15.1 is recommended in order to guarantee a smooth execution of the main-file. We strongly recommend using at least Stata/SE.

Used packages

Table 3 contains the Stata-packages (Stata ado-files), which are required for the data processing, as well as their functions and information about the documentation. During the execution of the main-file, these packages, if not already installed, will be downloaded and installed. Hereby, the Stata-packages `markdoc`, `weaver`, `dm31` and the self-written ado-file `list_errorcases` are used. As the packages are downloaded during the execution of the main-file, an internet connection for the time of the execution is required.

Table 3: Packages used in the main-file

Package	Function	Documentation
<code>markdoc</code>	MarkDoc <i>generates</i> a documentation in an HTML-format based on the do-files of the data processing and harmonization.	<p>Documentation MarkDoc:</p> <p>http://www.haghigh.com/statistics/stata-blog/reproducible-research/markdoc.php (accessed 20.09.2019, 23:20h)</p> <p>Haghigh, E. F. (2014). Dynamic documents in Stata: MarkDoc, Ketchup, and Weaver (https://www.stata.com/meeting/italy14/abstracts/materials/it14_haghigh.pdf) (accessed 2019-20-09, 23:20h)</p>

Package	Function	Documentation
weaver	Weaver is required in order to run the package Markdoc.	Documentation weaver Haghish, E. F. (2014). Dynamic documents in Stata: MarkDoc, Ketchup, and Weaver (https://www.stata.com/meeting/italy14/abstracts/materials/it14_haghish.pdf) (accessed 2019-20-09, 23:20h)
dm31	The command <code>rmiss2</code> counts the number of errors in a specified list of variables. Contrary to the Stata-command <code>rmiss</code> , <code>rmiss2</code> accepts string as well as numeric variables.	Documentation <code>rmiss2</code> https://www.stata.com/stb/stb26/dm31/rmiss2.hlp (accessed 2019-20-09, 23:20h)
list_errorcases	The command <code>list_errorcases</code> lists all partnerships of a person even if only one fulfills the conditions (as specified in an <code>if</code> -condition). Furthermore, all partnerships of an anchor person can be marked in an additional option.	Self-written Ado-file

3.2 Before running the main file main.do

To run the harmonization process, some small adjustments need to be executed. Those are described in the following.

Structure of the haspad-folder and required adaptations

After the selection of the requested survey programs and the variables to be harmonized with the HaSpaD-Harmonization Wizard, the archive file `haspad.zip` is available to download and can be saved and unpacked⁶ in the local working directory of choice. It must be made sure that *the content* of the `haspad.zip` file is saved in a directory with the name `haspad`. Some programs for the unpacking of ZIP archives store the unpacked files in a new directory with the name of the ZIP-archive as a default mode, but others do not. As an example, in case of a correct unpacking process the document `README.txt` would be saved in the directory `C:/mydata/haspad/README.txt`. A wrong outcome would be `C:/mydata/README.txt`.

In the following, we explore an example for which the variables `sex` (`sex`) and survey area (`eastwest`) and the survey programs ALLBUS-Cumulation, the Fertility and Family Survey (Germany) and the Mannheim Divorce Study have been selected via the HaSpaD-Harmonization Wizard. In this example, the content of the unpacked ZIP-archive prior to the execution of the main-file looks like the following (*Figure 2*):

⁶ A large selection of data compression programs can be found online and are partly available for free use.

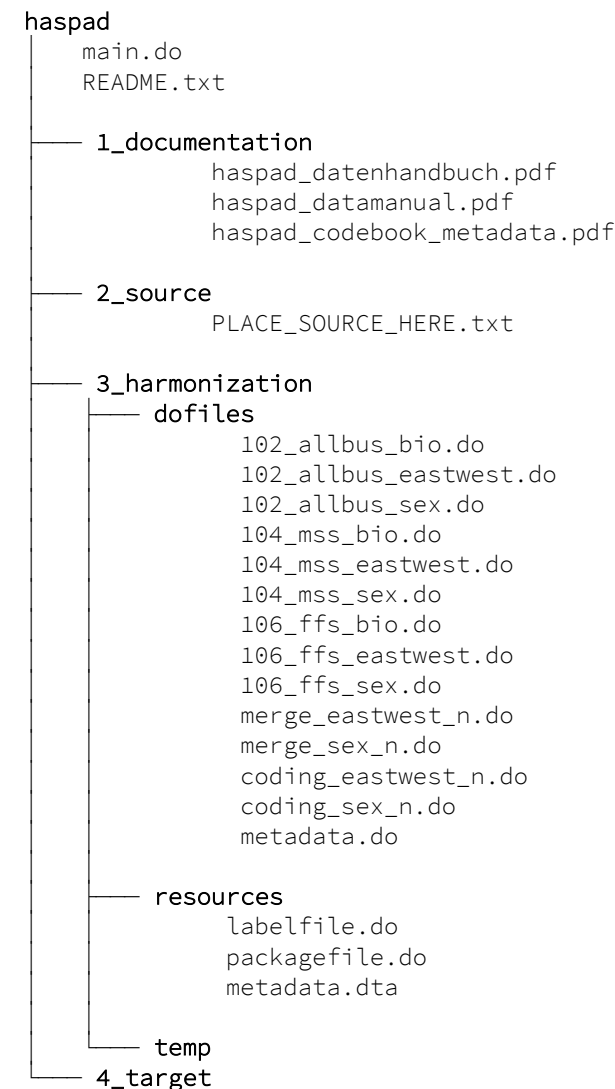


Figure 2: Folder structure before running the main-files

In the folder `haspad` the first folder level consists in a txt-file `README.txt` that includes the folder structure of the `haspad`-folder and the system requirements (see *chapter 1*), a do-file `main.do` as well as four folders `1_documentation`, `2_source`, `3_harmonization` and `4_target`. The folder `1_documentation` contains the HaSpaD data manual in English and German. The folder `3_harmonization` includes the required do-files for the data processing as well as other required Stata-packages. The folder `4_target` is still empty. In a next step, the users must save the acquired data sources in the folder `2_source` and, if necessary, unpack them (if the data distributor provides them as ZIP-archives). The Stata-code for the data processing uses relative paths within the `haspad`-folder. If the directory structure is changed, the program files cannot access the source and target files anymore. Therefore, it is important to ensure that the unpacking process of the source data to be harmonized does *neither change any names of files or folders nor delete or move single files*, so that the *working directories of the provided data remain intact*.

Users must make sure that the data compression program unpacks the ZIP-archive into a new folder with the identical name and folder structure as the ZIP-file. This is the default mode for the program 7-zip (see Figure 3). Figure 3 shows how the ZIP-files for the German General Social Survey ALLBUS-Cumulation 1980-2016 (ZA4586_v1-0-0.dta.zip) and the Mannheim Divorce Study (ZA3188.zip) are stored and unpacked in the folder 2_source. The Fertility and Family Survey (ZA3400.dta) is provided as dta-format and thus, only has to be saved in haspad/2_source. As the screenshot 3 in Figure 3 shows, it is important to keep the red-marked field activated and to keep the pre-configuration “complete path description”, in order to maintain the working directory of all ZIP folders. This ensures that a subfolder is created for each unpacked ZIP folder:

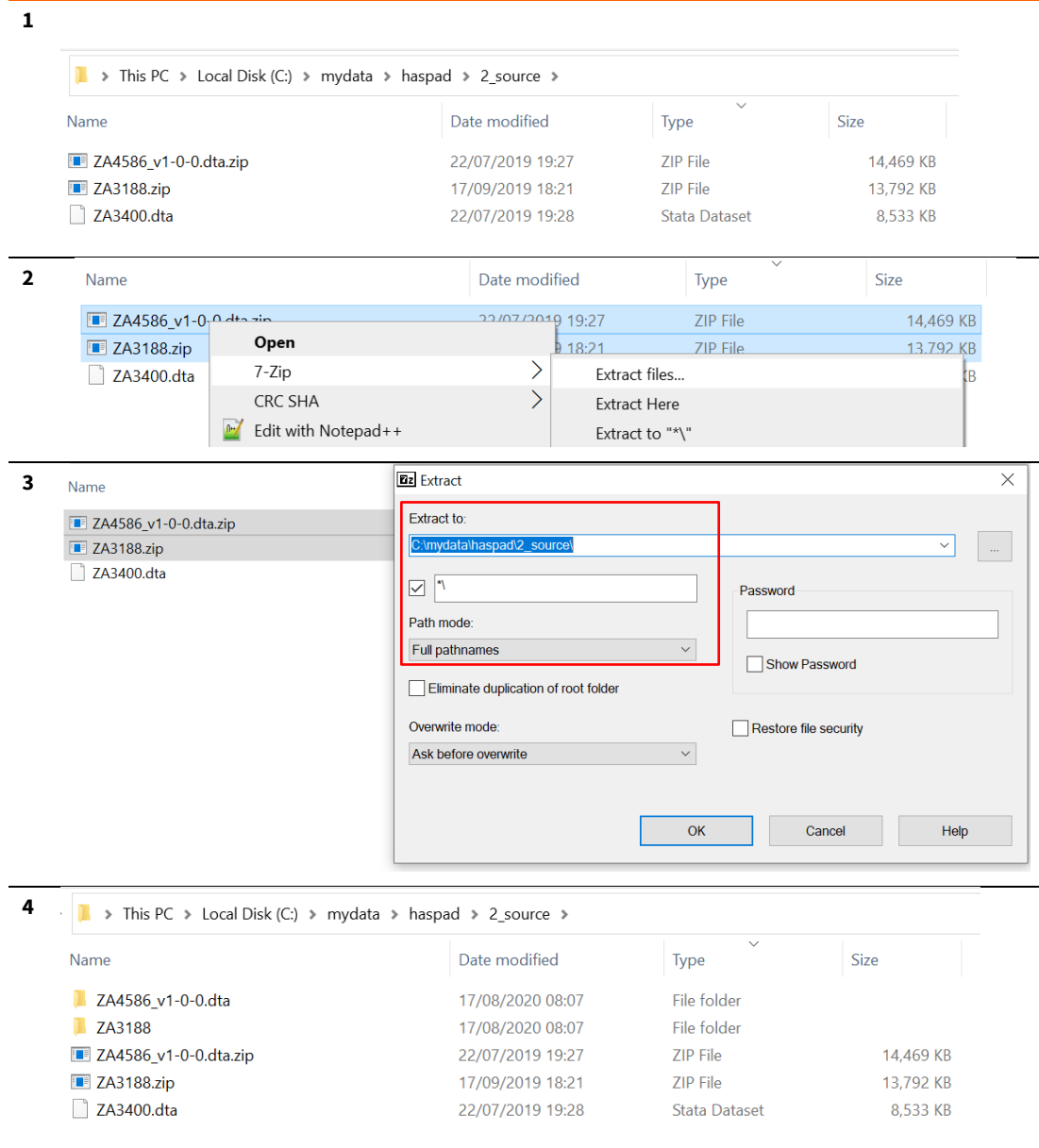


Figure 3: Instruction how to unpack the original data sets (for the program 7-Zip)

Accordingly, for the ALLBUS-Cumulation the correct place to save the Stata-file ZA4586_v1-0-0.dta after the unpacking process is the directory

C:/mydata/haspad/2_source/ZA4586_v1-0-0.dta/ZA4586_v1-0-0.dta.

The following directory would yield errors: C:/mydata/haspad/ZA4586_v1-0-0.dta.

Figure 4 informs about the haspad-folder structure after unpacking the original data sets.

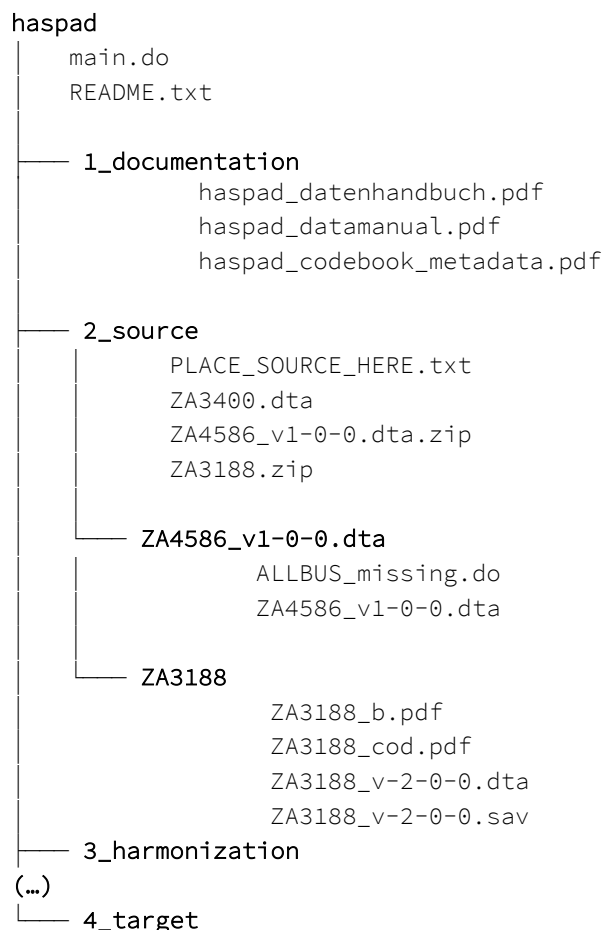


Figure 4: Folder structure after unpkacing the source data sets (extract)

Necessary adaption of the main file main.do

After the selection of the variables via the Harmonization Wizard, the HaSpaD-download and the acquisition, correct storage and unpacking of the original data, the main-file `main.do` is opened and adapted. *Line 40*, which informs about the folder path for the folder containing the unpacked haspad folder, needs to be adjusted to the path where the haspad-folder has been saved. For instance, if the folder `haspad` is stored in `C:/mydata`, then *line 40* consists in:

```

*****
> 40 cd "c:/mydata/haspad"
*****
  
```

3.3 During the execution of the main file `main.do`

Depending on the computer capacity and the number of selected variables and survey programs, the execution of the main-file `main.do` might take between 30 seconds (in case only one survey program without additional variables has been selected) and approx. 15 minutes (if the user selects all survey programs and additional variables). All required packages are loaded and installed, the processing do-files are called and executed, the data sets are cumulated and labelled, and the respective documentations and the target data set are saved. For interested users, *appendix A. 2* describes all single steps of the data processing in more detail.

3.4 After the execution of the main file `main.do`

The HaSpaD-folder now contains several new files and folders (see *Figure 5*):

- The documentation of the data processing in an HTML-format at `haspad/1_documentation/harmonization_targets/` and `haspad/1_documentation/preparation_couplebio/` respectively.
- The HaSpaD-data set `target.dta` (ready to be analyzed) at `haspad/4_target/`.

The HTML-files help to understand the processing of the couple biographies (`preparation_couplebio`) and the harmonized additional variables (`harmonization_targets`). The target data set `target.dta` from our example consists in 33 variables: 31 variables of the couple biography data set and the two additional variables `sex` (`sex`) and survey area (`eastwest`). In addition - after adapting the Stata version number and the file path - various metadata at study level (sample, survey design, survey years, etc.) can be merged to the previously created HaSpaD target data set using do-File `metadata.do` under `haspad/3_harmonization/dofiles`.

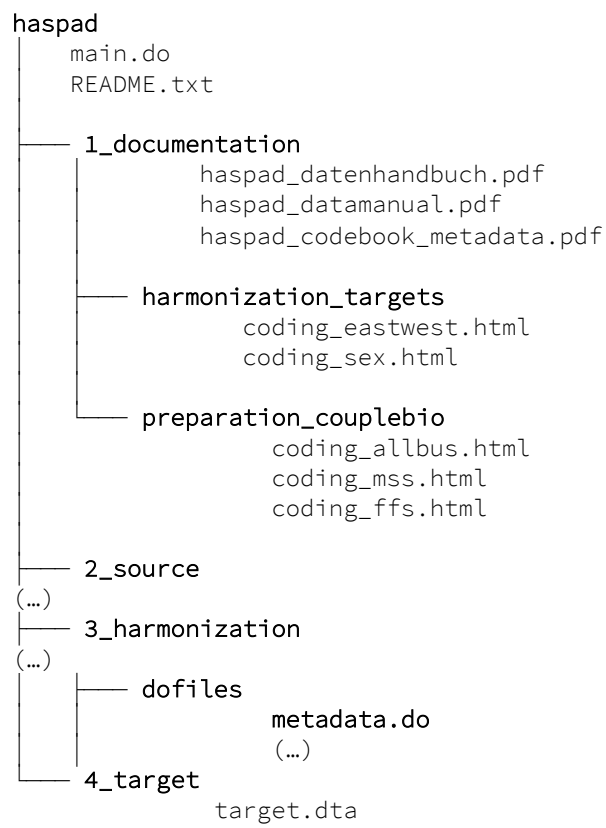


Figure 5: Folder structure after the execution of the main-file (extract)

4 Data structure

In the HaSpaD project the target data set consists in harmonized biography data and additional harmonized variables. The data set is in a long format. This implies that each row contains information about one partnership. Individuals are identifiable by the anchor ID `a_id`. Respondents from the original data sets who do not report any couple biography data are not included in the HaSpaD data set. The biography data set covers 31 variables, divided into three sections: ID variables, biography variables and flag variables (see *Table 5*).

4.1 Identification of survey programs and sub-studies in the HaSpaD data set (data set ID)

The survey programs used in the HaSpaD project often consist of several sub-studies. To indicate the origin of each couple biography, an ID is assigned to each study as well as each sub-study. Based on the source data set, the sub-study ID is assigned. Therefore, in the following the sub-study ID is also addressed as data set ID. If partnerships are stored in the same original study in different files (for instance because they are based on different questionnaires), different sub-study IDs or data set IDs are assigned to the original data (even when they are cited by the same DOI; this occurs for the Life History Study). In some cases, a 0 was added to the data set-ID (e.g. 40 for the Life History Study GLHS GDR or 50 for the Life History Study GLHS-East 71). Otherwise, overlaps could emerge in the HaSpaD-target data set while creating the person identification number (anchor ID). This problem is further explained in the section about the creation of the anchor-ID (`a_id`). *Table 4* helps to understand the classification scheme behind the study ID and data set IDs.

At this point it should be already highlighted that for partnerships which have been updated in a longitudinal format (and thus, originate from various source data sets), the data set-ID was assigned by concatenating the data set IDs of the source data sets. If a partnership, which had been initially reported in the Life History Study GLHS-East 71 (`d_id = 50`), is updated with information from the Life History Study GLHS-Panel 71 (`d_id = 80`), then the data set ID “5080” is assigned. This means that the data set IDs of the Life History Study GLHS-East 71 (“50”) and of the Life History Study GLHS-Panel 71 (“80”) have been concatenated to “5080”.

Table 4: Notation of the survey programs in the HaspaD target data set

Study ID / Data set ID / Acronym	Study title	Name of data set including couple biography Data sets linked to the couple biography data set	Number of partnerships
101 / 1 / pairfam	pairfam Data Release 12.0	biopart	36,994
102 / 1 / allbus	Cumulated German General Social Survey (ALLBUS) – Cumulation 1980 – 2016	ZA4586_v1-0-0	21,536
103 / 11 / fs	Change and Development of Family Life Forms in West Germany (Survey of Families)	ZA2245_a	6,781
103 / 12 / fs	Family and Partner Relations in Eastern Germany (Survey of Families)	ZA2392_Partner	2,584

Study ID / Data set ID / Acronym	Study title	Name of data set including couple biography Data sets linked to the couple biography data set	Number of partnerships
103 / 2 / fs	Change and Development of Ways of Family Life - 2nd Wave (Survey of Families)	ZA2860_PartnerBiographie_NoMissing	11,838
103 / 3 / fs	Change and Development of Families` Way of Life - 3rd Wave (Family Survey)	ZA3920_Partnerbiographie	12,879
104 / 1 / mss	Mannheim Divorce Study 1996	ZA3188_v2-0-0	5,020
106 / 1 / ffs	German Fertility and Family Survey 1992	ZA3400	7,905
107 / 1 / glhs	Courses of Life and Social Change: Courses of Life and Welfare Development (Life History Study GLHS-West I)	REC1	2,129
107 / 21 / glhs	Courses of Life and Social Change: The Between-the-War Cohort in Transition to Retirement (Life History Study GLHS-West II A - Personal Interview)	FP, PA	463
107 / 22 / glhs	Courses of Life and Social Change: The Between-the-War Cohort in Transition to Retirement (Life History Study GLHS-West II T - Telephone Interview)	FP, PA	1,128
107 / 3 / glhs	Courses of Life and Social Change: Access to Occupation in Employment Crisis (Life History Study GLHS-West III)	FP, PA, VB, GT	1,633
107 / 40 / glhs	Courses of Life and Historical Change in East Germany (Life History Study GLHS GDR)	PARTNER, PARTQUER, KONTAKT, pa3456p	1,432
107 / 50 / glhs	East German Life Courses After Unification (Life History Study GLHS-East 71)	fp7p, pa7, public_caseid	796
107 / 61 / glhs	East German Life Courses After Unification (Life History Study GLHS-East Panel) - Birth cohort 1939-41, 1951-53, 1959-61	fpp456p, pa3456p, fpq456p	1,001 (of which 942 are updated in a longitudinal format)
107 / 62 / glhs	East German Life Courses After Unification (Life History Study GLHS-East Panel) - Birth cohort 1929-31	fa3p, pa3456p	275 (of which 267 are updated in a longitudinal format))
107 / 70 / glhs	Education Training and Occupation: Life Courses of the 1964 and 1971 Birth Cohorts in West Germany (Life History Study GLHS-West 64/71)	rec16sp, rec1, public_caseid	3,635

Study ID / Data set ID / Acronym	Study title	Name of data set including couple biography Data sets linked to the couple biography data set	Number of partnerships
107 / 80 / glhs	Early Careers and Starting a Family: Life Courses of the 1971 Birth Cohorts in East and West Germany (Life History Study GLHS-Panel 71)	fpl , public_caseid, pa	1,189 (of which 395 are updated in a longitudinal format)
108 / 1 / ggs	Gender and Generations Survey (2005), Wave 1 (Subsample Germany)	GGG_Wave1_Germany_V.4.3 , Germany_Wave2_V.1.3	7,463
108 / 2 / ggs	Gender and Generations Survey (2008), Wave 2 (Subsample Germany)	Germany_Wave2_V.1.3	2,681 (of which 2,223 are updated in a longitudinal format)
109 / 1 / share	The Survey of Health, Ageing and Retirement in Europe (SHARE); Wave 1, 2, 3 (Subsample Germany)	sharew3_rel6-0-0_rp , sharew3_rel6-0-0_cv_r, sharew3_rel6-0-0_st	2,483
110 / 1 / soep	Socio-Economic Panel (SOEP), Years 1984-2018	biomarsy , pgen	38,446
Total number of partnerships			170,291

4.2 ID variables

Table 5: Overview of the ID variables

Variable name	Variable label
sid	Study id
did	Dataset id
aid	Anchor id
id	Original anchor id
pid	Partner id
pid_o	Original partner id
partner_pointer	Partner pointer

In total, the target data set covers seven different identification variables, which are important for the correct identification of the survey program (study ID, *sid*) and sub-study (data set ID, *did*) of origin for each partnership. The anchor ID (*aid*) and, respectively, partner ID (*pid*) serve to link the biography data with the additional variables selected by the user.

The `aid` results from the concatenation of information on the original study (`sid + did`) and the identification features within the original study (`id`).

The `pid` was constructed in a similar way: Here, the original study (`sid + did`) and on the respective partnership were linked. For original studies that provide a partner ID (`pairfam`, `SHARE`), the respective partnership is marked with the original partner ID (`pid_o`). The `pid` thus results from the concatenation of `sid + did + pid_o`. In studies in which there is no independent identification feature for a partnership (e.g. `ALLBUS`, life history studies, `SOEP (biomarsy)`), a partnership is identified via the `aid (sid + did + id)` concatenated with the variable `partner_pointer`. This variable `partner_pointer` results from the position of the original information on a partnership in the original data set and thus necessarily follows its own logic in each original study (e.g. first, second, third spell; first, second, third set of variables on partnerships). If users want to apply more information to the target data set than the additional variables provided, we recommend that they familiarize themselves with the system of the variable `partner_pointer` using the HTML documentation on data processing and variable construction.

4.3 Biography variables

Table 6: Overview of the biography variables

Variable name	Variable label
<code>intd</code>	Date of interview (in century months*)
<code>relbeg</code>	Beginning relationship (in century months)
<code>relend</code>	End relationship (in century months)
<code>cohbeg</code>	Beginning cohabitation (in century months)
<code>cohend</code>	End cohabitation (in century months)
<code>marbeg</code>	Beginning marriage (in century months)
<code>marend</code>	End marriage (in century months)
<code>dop</code>	Date of partner's death (in century months)

* A value of 1 corresponds to January 1900.

The biography variables contain information about the beginning and end of relationships, cohabitations and marriages for all partnerships reported by the anchor person and are denoted in century months since 1900 (the value 1 corresponds to January 1900, see Table 6). The date of interview of the last survey of the anchor person can be used to identify a right censoring for the duration of partnerships. If the partner of the anchor person deceased, the date of death is indicated by the variable `dop`.

4.4 Flag variables

The flag variables contain meta-information about single biography variables.

Imputed variables

Table 7: Overview of the flag variables containing information about the degree of imputation for the biography variables

Variable name	Variable label
flag_relbeg_imp	Imputed date of beginning relationship
flag_relend_imp	Imputed date of end relationship
flag_cohbeg_imp	Imputed date of beginning cohabitation
flag_cohend_imp	Imputed date of end cohabitation
flag_marbeg_imp	Imputed date of beginning marriage
flag_marend_imp	Imputed date of end marriage
flag_dop_imp	Imputed date of partner's death

The imputation variables indicate whether the biography variables are provided on a monthly, seasonal or annual base (see *Table 7*; see *chapter 5.6* for further explanation). Hereby, it makes no difference if the respective sub-study only allows for annual reports or if the respondent could only make vague indications due to a lack of memory.

Original variables

Table 8: Overview of the flag variables with original values

Variable name	Variable label
flag_relbeg_ori	Original date of beginning relationship
flag_relend_ori	Original date of end relationship
flag_cohbeg_ori	Original date of beginning cohabitation
flag_cohend_ori	Original date of end cohabitation
flag_marbeg_ori	Original date of beginning marriage
flag_marend_ori	Original date of end marriage
flag_dop_ori	Original date of partner's death

The original variables contain the non-harmonized values from the original data source for the respective biography variable and ensure transparency for the coding of the source variables. Since the original variables are based mainly on information regarding the calendar year (p.r.n. the month), they offer a quick overview of the plausibility of the statements of the respondents (see *Table 8*, see *chapter 6.1.4* for further explanation).

Further flag variables (flag_beg, problem)

Table 9: Overview of further flag variables

Variable name	Variable label
flag_beg	Beginning partnership biography
problem	Flag used to mark problem cases

Some sub-studies only provide information about cohabitating partnerships, others only about marriages. Only in a few sub-studies partnerships can be analyzed from the beginning to a possible divorce. Therefore, the variable `flag_beg` indicates at which degree of institutionalization a partnership starts in a sub-study (see *chapter 6.1.5*). Possible inconsistencies in a couple biography or in the processing are marked with the variable `problem` (see *Table 9* and *chapter 6.1.5*).

4.5 Missing scheme

In the following, the missing scheme of the HaSpaD project is presented (see *Table 10*):

Table 10: Missing scheme

Value	Label
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
-66	Death partner
-77	Separation/divorce, no exact date
-88	Union dissolution/death of partner
-99	Ongoing

The code “-3 Does not apply” is assigned, if a condition does not apply for the respondent (e.g. `cohbeg = -3`, no cohabitation; `marbeg = -3`, no marriage). The variable `relbeg` can never have the value “-3” because the units of observation in the HaSpaD data are partnerships with the status “relationship”.

The missing code “-7 Incomplete data” summarizes several data problems in the original data sources. For instance, this could be missing data (item non-responses such as: “don’t know”, “no response” or “response denied”) or corrections of the data provider (e.g. filter errors, data corrections, technical problems).

The code “-10 Missing by study design” is applied, if an information could not be reported for a subgroup of the total sample (due to filtering) or the whole number of participants due to the study design.

For the variables regarding the end of a partnership additional missing codes exist: in case that the relationship, cohabitation or marriage ends with the death of the anchor’s partner (`dop > 0`), the

value “-66 Death Partner” is assigned to the other variables regarding the end of partnership (`re1end`, `cohend`, `marend`).

In some survey programs the filtering of a questionnaire causes the problem that it is known that a marriage ended with death or a divorce (1) or that cohabitation (at a certain point of time) ended because of a separation or divorce (2). Yet, the exact date of death or divorce in the first case and the exact date of separation or divorce in the second case are unknown. Hence, in the first case the label “-88 Union dissolution/death of partner” is assigned to the date of divorce and death (`marend`, `dop`) for the partner. In the second case the date of the end of cohabitation is marked in `cohend` and the end of relationship (`re1end`) or, respectively, the end of marriage (`marend`) are coded as “-77 Separation/divorce, no exact date”.

The value “-99 Ongoing” indicates that a partnership phase has not finished and continues until the (last) date of interview of the respondent (`right censoring`).

5 Special characteristics of the HaSpaD data due to the linkage of different data sources

To harmonize data on partnership is especially challenging, since partnerships are differently defined and measured across the included studies (e.g. partnerships with a minimum duration, a certain degree of institutionalization, or starting from a particular age). Also, there are varying limits for the maximum number of reported partnerships. Moreover, the period of observation for the biographies may differ across survey programs and sub-studies, which results in different variables of beginnings and ends of partnerships (the end - depending on the sub-study - must be deduced from the divorce, the end of cohabitation or the separation). Furthermore, the original data sources differ regarding the surveyed area and the age structure of the samples, as well as regarding the question, whether only Germans or also non-German parts of the population are interviewed. This chapter introduces and discusses these specific challenges. Furthermore, methodological implications for weights and strategies of month imputation in case of missing or only seasonal information about months is explained and discussed. Finally, the longitudinal processing of panel surveys without the existence of a partner-ID is elaborated for the Life History Studies and the Generation and Gender Survey in the last section.

5.1 Varying definitions of partnerships

The definition of a partnership varies across survey programs with regard to the degree of institutionalization, the duration and the minimum age for respondents to report about their partnerships.⁷ *Table 11* presents the degree of institutionalization and other restrictions by sub-studies. For example, pairfam allows to report LAT-partnerships (living apart together), cohabiting, non-married partnerships and marriages. Whilst there are no restrictions for the report of the current partnership, retrospectively reported partnerships should include all relationships starting from age 14, which fulfill the definitions stated in *Table 11*. In contrast, within the Allbus Cumulation 1980-2016, only marriages are considered without any restrictions on retrospectively reported partnerships.

Table 11: Degree of institutionalization per study

Sub-study	Degree of institutionalization	Further restrictions
Panel Analysis of Intimate Relationships and Family Dynamics (pairfam)	LAT, cohabitation, marriages	<p><i>Current partnership:</i> No restrictions</p> <p><i>Retrospectively reported partnerships:</i> All relationships starting from age 14 are reported, which [...] either lasted longer than 6 months, or where [...] [the anchor person] cohabitated with the partner [...], or from which resulted a child or which were important for [...] [the anchor person] for other reasons" (pairfam Group 2021, p. 11).</p>

⁷ The references to the documentation in the form of questionnaires, codebooks and documentation manuals can be found in the *appendix A. 3*.

ALLBUS– Cumulation 1980 – 2016	Marriages	<i>Current and retrospectively reported partnerships:</i> There are no further restrictions (Baumann & Schulz 2018, pp. 1629).
Family Survey I. Wave West 1988 Family survey I. Wave East 1990 Family survey II. Wave 1994	LAT*, cohabitation*, marriages	<i>Current partnership:</i> There are no further restrictions. The current partnership can last less than a year. (Deutsches Jugendinstitut 1992b, p. 7; 2018d, p. 8; 2018e, p. 11) <i>Retrospectively reported partnerships:</i> Past partnerships are reported, which lasted a minimum of one year and “[...] consisted in more than just passing acquaintances or crushes”. Definitely implies marriages. (Deutsches Jugendinstitut 1992b, p. 7; 2018d, p. 8; 2018e, p. 11)
Family Survey III. Wave 2000	LAT*, cohabitation*, marriages*	<i>Current partnership:</i> There are no further restrictions. The current partnership or marriage can last less than a year. (Deutsches Jugendinstitut 2000, p. 59) <i>Retrospectively reported partnerships:</i> Past partnerships are reported, which lasted a minimum of one year. (Deutsches Jugendinstitut 2000, p. 59)
Mannheim Divorce Study 1996	Marriages*	<i>Current and retrospectively reported partnerships:</i> Only the first marriage of the anchor is reported (Esser, Gostomski & Hartmann 2018c, p. 2).
German Fertility and Family Survey 1992	cohabitation, marriages	<i>Current and retrospectively reported partnerships:</i> There are no further restrictions (Bundesinstitut für Bevölkerungsforschung 2002b, p. 7).
GLHS-West I GLHS-West II A – Personal Interview GLHS-West II T - Telephone Interview GLHS-West III	cohabitation*, marriages	<i>Current partnership:</i> A marriage or a cohabiting, non-married partnership can be reported. The cohabiting, non-married partnership can only be reported, as soon as the respondent currently lives in a non-married cohabitation with the partner (Mayer 2018d, pp. 17, 39; 2018e, p. 53; 1995c, p.53; 2018f, pp. 206). <i>Retrospectively reported partnerships:</i> Marriages, but no non-married cohabitations are reported (Mayer 2018d, p. 17; 2018e: 53; 1995c, p. 53; 1995c, p. 53; 2018f, pp. 206).
GLHS-GDR	cohabitation, marriages	<i>Current and retrospectively reported partnerships:</i> There are no further restrictions. (Mayer 1995d, p. 77).

GLHS-East 71	LAT*, cohabitation*, marriages*	<i>Current partnership and retrospectively reported partnerships:</i> All types of partnerships are reported for the “[...] period since [...] the 16 th birthday until today” (Mayer 2004c, p. 209).
GLHS-East Panel - Birth cohorts 1939-41, 1951-53, 1959-61, GLHS-East Panel - Birth cohort 1929-31	cohabitation, marriages	<i>Current and retrospectively reported partnerships:</i> There are no further restrictions (Mayer 2004d, p. 115).
GLHS-West 64/71	LAT*, cohabitation*, marriages	<i>Current partnership and retrospectively reported partnerships:</i> All partnerships with a minimum duration of one year are reported (Mayer & Kleinhenz 2004b, p. 105).
GLHS-Panel 71	LAT*, cohabitation*, marriages	<i>Current partnership:</i> There are no further restrictions. <i>Retrospectively reported partnerships:</i> All partnerships with a minimum duration of one year are reported (Mayer 2014b, p. 92).
Generations & Gender Survey (2005) Generations & Gender Survey (2008)	LAT*, cohabitation, marriages	<i>Current partnership:</i> Besides non-married cohabitations and marriages, LAT-partnerships can be reported, too (Ruckdeschel et al. 2006, p. 51; Leven 2009, p. 26). <i>Retrospectively reported partnerships:</i> Non-married cohabitations and marriages, but no LAT-partnerships are reported (Ruckdeschel et al. 2006, p. 64; Leven 2009, pp. 22).
SHARE	LAT*, cohabitation, marriages	<i>Current partnership and retrospectively reported partnerships:</i> LAT-partnerships are defined as „long-term relationships” (SHARE-ERIC 2019, p. 24).
Socio-Economic Panel (SOEP)	marriages*	<i>Current partnership:</i> Marriages are only reported starting from age 16 (Hamjediers et al. 2020) <i>Retrospectively reported partnerships:</i> Up to and including 2010, marriages are only reported retrospectively in the biography questionnaire. Marriages are only reported starting from age 16 (Hamjediers et al. 2020).

From the distinct partnership definitions derives that retrospectively reported partnerships can possess a longer duration than current ones due to the study design. Furthermore, it must be considered that LAT-partnerships could be underrepresented in the HaSpaD-data for the life history before the interview and probably historically as well. This stems from the fact that LAT-partnerships are prospectively reported and from the restrictions in retrospectively reported partnerships.

Based on the explained differences of partnership definitions, varying analysis potentials exist regarding different types of partnerships from the harmonized survey programs and sub-studies. These analysis potentials for each type of partnership are explained in the following.

Data sources that contain information about Living apart together (LAT) – partnerships

LAT-partnerships are covered in twelve sub-studies. Couple biographies for respondents who do not live together with their partner can be analyzed for the survey programs pairfam, the Family Surveys, the Life History Studies GLHS-East 71, GLHS-West 64/71 and GLHS-Panel 71, the Generations and Gender Survey and SHARE⁸.

It has to be kept in mind that respondents from the Generations and Gender Survey can only report the current partnership as LAT. Therefore, analyses for past partnerships of such type are not possible.

Data sources that contain cohabiting, non-married partnerships

Cohabiting, non-married partnerships can be analyzed for 19 studies. For the GLHS-West I, both GLHS-West II and the GLHS-West III merely current cohabiting, non-married partnerships are reported, which last until the date of interview. Exceptions are the ALLBUS-Cumulation, the Mannheim Divorce Study and partially the Socio-Economic Panel, which only contain marriage biographies.

Data sources that contain marriages

The minimal requirement for harmonized survey programs in the HaSpaD project is biography data on marriages. Thus, for all survey programs marriages can be classified as the partnership type with the highest degree of institutionalization. Nonetheless, not all harmonized survey programs include complete marriage biographies of the respondent or the respondents have not entered a marriage yet. Gaps in the marriage biography can occur due to restrictions for the report of partnerships per study design and are explained in the following section (see *chapter 5.2*).

5.2 Differences in the maximum number of reported partnerships

The maximum number of reported partnerships is limited for the ALLBUS-Cumulation, the Mannheim Divorce Study and for some partial studies of the Family Surveys, Life History Studies and the Socio-Economic Panel (see *Table 12*). While the Mannheim Divorce Study only surveys biography data for the first marriage, a maximum of four marriages can be reported in the ALLBUS-Cumulation. Likewise, the Family Survey I. Wave 1988, I. Wave 1990 and II. Wave 1994 are restricted to four partnerships: if respondents have more than three partnerships, then the fourth registered partnership is the current one. Consequentially, partnerships between the third and the current one, cannot be reported. Moreover, differences can also be found within the Life History Studies: for the GLHS-West I up to four previous and one current marriage can be stated. Additionally, respondents with an ongoing marriage can state another current partnership, if the marriage is separated, but not yet divorced and the respondent lives together with the new partner. In the GLHS-West III the

⁸ Since the *biomarsy* dataset contains only marriage biographies, no information on LAT-relationships can be given for partnerships from this data source.

maximum number of partnerships is four: two previous as well as one current marriage can be reported. Besides, respondents with an ongoing marriage can state another current partnership, if the marriage is separated, but not yet divorced and the respondent lives together with the new partner. In the Socio-Economic Panel – depending on the survey year – a maximum of three or four partnerships can be registered (Hamjediers et al 2020).

Table 12: Maximum number of reported partnerships per study

Sub-study	Maximum number of reported partnerships	Further restrictions
pairfam	There are no restrictions (Brüderl et al. 2020b, p. 62).	-
ALLBUS Cumulation	A maximum of four marriages can be reported (Baumann & Schulz 2018, pp. 1629).	-
Family Survey I. Wave West 1988; Family Survey I. Wave East 1990; Family Survey II. Wave 1994	A maximum of four partnerships can be reported.	In case of more than three partnerships, the fourth is the most recent one (Deutsches Jugendinstitut 1992b, p. 9, 2018d, p. 10, 2018e, p. 13).
Family Survey III. Wave 2000	There are no restrictions. (Deutsches Jugendinstitut 2000, pp. 59).	-
Mannheim Divorce Study 1996	Only the first marriage is reported (Esser, Gostomski & Hartmann 2018c: 2).	-
German Fertility and Family Survey 1992	There are no restrictions (Bundesinstitut für Bevölkerungsforschung 2002b, p. 8).	-
GLHS-West I	A maximum of six partnerships can be reported.	The respondents can report their four first terminated marriages (Mayer 2018d, p. 41) and a current marriage (Mayer 2018d, p. 40). Insofar as the respondents state “2 Separated from spouse, but living together with another partner“ (Mayer 2018d, p. 39), they can report a current cohabiting, non-married partnership additionally to the current and the four former marriages (Mayer 2018d, p. 46).
GLHS -West II A - Personal Interview; GLHS -West II T - Telephone Interview; GLHS-GDR; GLHS-East 71, GLHS-East Panel, GLHS - West 64/71, GLHS -Panel 71	There are no restrictions (Mayer 2018e, p. 53; Mayer 1995c, p. 53; Mayer 1995d, p. 77; Mayer 2004c, p. 209; Mayer 2004d, p. 115; Mayer & Kleinhenz 2004b, p. 105; Mayer 2014b, p. 92).	-
GLHS -West III	A maximum of four partnerships can be reported.	The respondents can report their two first terminated marriages (Mayer 2018f, p. 99) and a current marriage (Mayer 2018f, p. 100). Insofar as the respondents state “2 Separated from spouse, but living together with another partner“ (Mayer 2018f, p. 98), they can report a current cohabiting, non-

Sub-study	Maximum number of reported partnerships	Further restrictions
		married partnership additionally to the current and the two former marriages (Mayer 2018f, p. 106).
Generations & Gender Survey (2005); Generations & Gender Survey (2008)	There are no restrictions (Ruckdeschel et al. 2006, pp. 50, Leven 2008, pp. 22).	-
SHARE	There are no restrictions (SHARE-ERIC 2019: 12et seq.)	-
Socio-Economic Panel (SOEP)	A maximum number of three or four partnerships – depending on the survey year – can be reported retrospectively.	Until and including 2010 a maximum of three marriages can be reported retrospectively in the biography questionnaire. Since 2011 up to three marriages or civil unions can be reported retrospectively (Hamjediers et al 2020).

In consequence, when comparing partnerships of a higher order or marriage sequences in the HaSpaD-target data set, it must be considered that couple biographies can include gaps by study design. These gaps should not be confused with single episodes. Hence, a low number of cases of partnerships or marriages of a higher order partially derives from the study designs.

5.3 Varying observation periods for partnerships

Since partnerships are defined differently throughout original data sources, observation periods for partnerships differ consequentially. Specifically, observations vary in relation to the beginning of a partnership (start of relationship, start of cohabitation, marriage) and the end (separation, end of cohabitation, divorce), which may influence the analyses.

Beginning of observation period

Across all survey programs and sub-studies, at least marriage biographies are recorded (see *Table 13*). Thus, valid values are available for the start of marriage (*marbeg*) in all studies. However, in some studies, which also include LAT and non-marital cohabitation, a transition to marriage has not (yet) necessarily taken place for some partnerships, so that information on the beginning of marriage is not available. In contrast, the ALLBUS-Cumulations as well as data set *biomarsy* of the Socio-Economic Panel, only provide marriage data. Consequently, neither information about the beginning of the relationship nor about cohabitation is available.

Apart from these exceptions (ALLBUS, SOEP), all other studies provide at least partial information on the beginning of cohabitation. For the GLHS -West I, both GLHS -West II and for the GLHS -West III, the start of cohabitation is only provided for cohabiting, non-married partnerships that were ongoing at the date of interview. In addition to that, the cohabitation begin is reported for the current marriage, if the respondent had been married only once (Mayer 2018f, p. 112) or for the first marriage, if the respondent had been married several times (Mayer 2018f, p. 113).

A particularity for the start of cohabitation exists in case of the GLHS-East Panel - Birth Cohort 1929-31 and the Generations & Gender Survey (2008) due to the panel linkage:

In the GLHS-East Panel - Birth Cohort 1929-31, the beginning of cohabitation cannot be reported (Mayer 2004d, p. 135). For the partnerships, which were already reported in wave 1 and again in

wave 2, the start of cohabitation could be updated based on the value from wave 1. If in the Generations & Gender Survey (2008) the respondent states to have lived together “[...] at the beginning of 2005 [with] a (marriage) partner” (Leven 2009, p. 22), then no start of cohabitation is available for this partnership in wave 2. For partnerships reported in wave 2, which cannot be linked to partnerships from the first wave, the start of cohabitation cannot be transferred from wave 1 and thus, is not available.

The beginning of relationship for all partnerships is only available for a limited set of included studies (pairfam, Family Surveys, Mannheim Divorce Study, GLHS-GDR, GLHS-East 71, GLHS-West 64/71, GLHS-Panel 71, SHARE). For the GLHS-East panel studies the relationship begin is not reported. Nonetheless, the start of relationships can be updated with the value from wave 1 for all partnerships, which are already reported in wave 1 and again in wave 2. Moreover, for the Generation & Gender Survey (2005) and the Generation & Gender Survey (2008) the beginning of relationship is merely accessible for LAT-partnerships, which were ongoing at the date of interview.

Table 13: Beginning of observation per study

Sub-study	Beginning of observation	Particularities/ Restrictions
pairfam	relbeg cohbeg marbeg	-
ALLBUS-Cumulation	relbeg ¹ cohbeg marbeg	-
Family Survey I. Wave East 1990; Family Survey I. Wave West 1988; Family Survey II. Wave 1994; Family Survey III. Wave 2000	relbeg cohbeg marbeg	
Mannheim Divorce Study 1996	relbeg cohbeg marbeg	-
German Fertility and Family Survey 1992	relbeg cohbeg marbeg	-
GLHS-West I GLHS-West II A - Personal Interview	relbeg cohbeg* marbeg	The cohabitation begin (cohbeg) is only reported for currently ongoing cohabiting, non-married partnerships (Mayer 2018d, pp. 39, 41, 46; 2018e, pp. 53; 2018e, pp. 53).
GLHS-West II T - Telephone Interview GLHS-West III	relbeg cohbeg* marbeg	The cohabitation begin (cohbeg) is only reported for currently ongoing cohabiting, non-married partnerships (Mayer 2018f, pp. 98, 106). Additionally, the cohabitation begin is reported for the current marriage, if the respondent had been married only once (Mayer 2018f, p. 112) or for the first marriage, if the respondent had been married several times (Mayer 2018f, p. 113).
GLHS-GDR	relbeg cohbeg marbeg	-

Sub-study	Beginning of observation	Particularities/ Restrictions
GLHS-East 71)	relbeg cohbeg marbeg	-
GLHS-East Panel - Birth cohort 1939-41, 1951-53, 1959-61	relbeg* cohbeg marbeg	The relationship begin (<i>relbeg</i>) is not reported in the GLHS-East Panel (Mayer 2004d, p. 115). The relationship begin can be updated with the value from wave 1 for all partnerships, which are already reported in wave 1 and again in wave 2.
GLHS-East Panel - Birth cohort 1929-31	relbeg* cohbeg* marbeg	The relationship begin (<i>relbeg</i>) and the cohabitation begin (<i>cohbeg</i>) are not reported in the GLHS-East Panel (Mayer 2004d, p. 135). The relationship begin and the cohabitation begin can be updated with the value from wave 1 for all partnerships, which are already reported in wave 1 and again in wave 2.
GLHS -West 64/71	relbeg cohbeg marbeg	-
GLHS-Panel 71	relbeg cohbeg marbeg	-
Generations & Gender Survey (2005)	relbeg* cohbeg marbeg	The relationship begin (<i>relbeg</i>) is only reported for current partnerships without cohabitation (Ruckdeschel et al. 2006, p. 51).
Generations & Gender Survey (2008)	relbeg* cohbeg* marbeg*	<p>The relationship begin (<i>relbeg</i>) is only reported for current partnerships without cohabitation (Leven 2009, p. 26).</p> <p>If in the Generations & Gender Survey (2008) the respondent states to have lived together “[...] at the beginning of 2005 [with] a (marriage) partner” (Leven 2009, p. 22), then cohabitation begin (<i>cohbeg</i>) is available for this partnership in wave 2. For partnerships reported in wave 2, which cannot be linked to partnerships from the first wave, the cohabitation begin cannot be transferred from wave 1 and thus, is not available.</p> <p>If in the Generations & Gender Survey (2008) the respondent states to have lived together “[...] at the beginning of 2005 [with] a (marriage) partner” (Leven 2009, p. 22), then no marriage begin (<i>marbeg</i>) is available for this partnership in wave 2. For partnerships reported in wave 2, which cannot be linked to partnerships from the first wave, the marriage begin cannot be transferred from wave 1 and thus, is not available. Also, for current partnerships without cohabitation reported only in wave 2, no marriage begin is available.</p>

Sub-study	Beginning of observation	Particularities/ Restrictions
SHARE	relbeg cohbeg marbeg	-
Socio-Economic Panel (SOEP)	relbeg* cohbeg* marbeg	The relationship begin (relbeg) and the cohabitation begin (cohbeg) are not available in the data source biomarsy.

¹relbeg = relationship begin, cohbeg = cohabitation begin, marbeg = marriage begin; if completely gray, then it is unknown.

End of observation period

In general, variables for the end of a partnership depend from the surveyed degrees of institutionalization, or, in other words, if the observation period of a (sub-) study includes this degree of institutionalization.

For example, if respondents have never been married, no data on marriage is available. In case a partnership is ongoing at the time of the interview, the end-variables are assigned a missing value ("-99 Ongoing", see *chapter 4.5*). This means, that for LAT-partnerships the relationship, for cohabiting, non-married partnerships also the cohabitation and for marriages the marriage itself are still ongoing at the time of the interview. Therefore, the construction of ongoing partnerships depends on the type of partnership and the availability of the respective variables in the original survey source ("-10 Missing by study design", if the information is not available). If the partnership ends with the death of the respondent's partner, the value "-66 Death partner" is assigned to all available end-variables and a valid value is given for the date of death. The Family Surveys form an exception, where the date of death of the deceased partner is only given for marriages.

Depending on the type of partnership and the existence of a valid value, dates of separation can be constructed based on the end of relationship for LAT-partnerships, the end of cohabitation for cohabiting, non-married partnerships, and the end of marriage for marriages.⁹

However, for some survey programs more nuanced information are available: pairfam, for instance, collects data about the date of separation and the end of cohabitation in case of a divorce.

Table 14: End of observation per study

Sub-study	End of observation	Particularities/ Restrictions
pairfam	relend cohend marend	-
ALLBUS-Cumulation	relend ¹ cohend marend	-
Family Survey I. Wave West 1988	relend* cohend marend	For cohabiting, non-married and LAT-partnerships the death of partner (dop) and the end of relationship (relend) cannot be distinguished, since it is only asked "[...] in which year [the partnership] ended" (Deutsches Jugendinstitut 1992b, p. 9). The year of the end of partnership is assigned to the end of relationship (relend). The date of death of the partner (dop)
Family Survey II. Wave 1994		
Family Survey III. Wave 2000		

⁹ In the end, it is left to the user to decide which date, if there are several end dates for the respective partnership phase, will be used as the separation date and thus the final end of the partnership.

Sub-study	End of observation	Particularities/ Restrictions
Family Survey I. Wave East 1990	re end* cohend* marend	is labelled “-88 Union dissolution/death of partner”. For cohabiting, non-married and LAT-partnerships the death of partner (dop) and the end of relationship (re end) cannot be distinguished, since it is only asked “[...] in which year [the partnership] ended” (Deutsches Jugendinstitut 2018d, p. 10). The year of the end of partnership is assigned to the end of relationship (re end). The date of death of the partner (dop) is labelled “-88 Union dissolution/death of partner”. The end of cohabitation (cohend) is only known for marriages.
Mannheim Divorce Study 1996	re end* cohend marend	The end of relationship (re end) is only known in the special case of a separation with an ongoing cohabitation (“separation from table and bed”) (Esser, Gostomski & Hartmann 2018b, p. 21).
German Fertility and Family Survey 1992	re end* cohend marend*	If the partnership ended by separation, the end of relationship (re end) is labelled “-77 Separation/divorce, no exact date” because even though the date for the end of cohabitation is known (Bundesinstitut für Bevölkerungsforschung 2002b, p. 8), it does not equal to the end of relationship (re end). If the partnership ended by divorce, the end of marriage (marend) is labelled “-77 Separation/divorce, no exact date” because even though the date for the end of cohabitation is known (Bundesinstitut für Bevölkerungsforschung 2002b, p. 8), it does not equal to the end of marriage (marend).
GLHS-West I	re end cohend* marend	The end of cohabitation (cohend) is only available for current marriages and current cohabiting, non-married partnerships. For current cohabiting, non-married partnerships the end of cohabitation (cohend) is always “-99 Ongoing”, since the cohabitation must be ongoing at the date of interview (Mayer 2018d, p. 39)
GLHS-West II A - Personal Interview	re end cohend* marend	For current cohabiting, non-married partnerships the end of cohabitation (cohend) is always “-99 Ongoing”, since the cohabitation must be ongoing at the date of interview. (Mayer 2018e, p. 54)
GLHS-West II T - Telephone Interview	re end cohend marend	
GLHS -West III	re end* cohend* marend	The end of relationship (re end) is only available for current marriages, which are separated but not yet divorced (Mayer 2018f, p. 101).

Sub-study	End of observation	Particularities/ Restrictions
		The end of cohabitation (<code>cohend</code>) is surveyed only for currently separated or divorced marriages (Mayer 2018f, p. 115). For current cohabiting, non-married partnerships the end of cohabitation (<code>cohend</code>) is always “-99 Ongoing”, since the cohabitation must be ongoing at the date of interview. (Mayer 2018f, p. 106)
GLHS-GDR	<code>reIend</code> <code>cohend</code> <code>marend</code>	-
GLHS-East 71	<code>reIend</code> <code>cohend</code> <code>marend</code>	-
GLHS-East Panel - Birth cohort 1939-41, 1951-53, 1959-61	<code>reIend*</code> <code>cohend*</code> <code>marend</code>	The end of relationship (<code>reIend</code>) is not included in the original data source GLHS-East Panel. For partnerships reported in wave 1 and again in wave 2, the end of relationship from wave 1 can be used if a separation already happened in wave 1. These partnerships are reported again in wave 2 and updated using the information about the end of marriage and cohabitation stemming from wave 2. Due to an error in the filtering, the end of cohabitation (<code>cohend</code>) is not available for a partnership reported in wave 1 and wave 2 (Mayer 2004d, p. 114). Even though this respondent state a partnership in wave 2, the question regarding the end of cohabitation is not asked.
GLHS-East Panel - Birth cohort 1929-31	<code>reIend*</code> <code>cohend*</code> <code>marend</code>	The end of relationship (<code>reIend</code>) cannot be reported in the second wave of the GLHS-East Panel (Mayer 2004d, p. 135). For partnerships reported in wave 1 and again in wave 2, the end of relationship from wave 1 can be used if a separation already happened in wave 1. These partnerships are reported again in wave 2 and updated using the information about the end of marriage and cohabitation stemming from wave 2. The end of cohabitation (<code>cohend</code>) is not obtainable for partnerships, which are reported in wave 2 for the first time. These people are not asked the question about the end of cohabitation due to an error in the filtering (Mayer 2004d, p. 135).
GLHS-West 64/71	<code>reIend</code> <code>cohend</code> <code>marend</code>	-
GLHS-Panel 71	<code>reIend</code> <code>cohend</code> <code>marend</code>	-
Generations & Gender Survey (2005)	<code>reIend</code> <code>cohend</code> <code>marend</code>	-
Generations & Gender Survey (2008)	<code>reIend</code> <code>cohend</code> <code>marend*</code>	For current partnerships, which are reported in wave 2 for the first time, no end of marriage (<code>marend</code>) is available.

Sub-study	End of observation	Particularities/ Restrictions
SHARE	reIend* cohend marend	The end of relationship (reIend) can only be reported for LAT-partnerships (SHARE-ERIC 2019, p. 24).
Socio-Economic Panel (SOEP)	reIend* marend* cohend*	The end of relationship (reIend) is mostly not available in the original data source biomarsy. Only for a few partnerships, the status “married, separated” is stated, which explicitly marks the end of relationship and assigns a date (Hamjediers et al 2020). The end of cohabitation (cohend) is not available in biomarsy (Hamjediers et al 2020) There are married partnerships, for which the end of marriage (marend) is coded “-88 Union dissolution/death of partner” because it is unknown whether the marriage ended with a divorce or with the death of the partner (Hamjediers et al 2020).

*reIend = relationship end, cohend = cohabitation end, marend = marriage end; if completely gray, then it is unknown.

This yields particularities for several studies regarding these three variables and possibly to the operationalization in the questionnaire:

Table 14 displays the end of observation by study. The end of marriage is not reported in the GLHS-Panel 71 and only partially in the Generations and Gender Survey (2008). In the Fertility and Family Survey (Germany) and in the Socio-Economic Panel (data source biomarsy), the end of marriage is coded with a special missing code: question 227 of the Fertility and Family Survey captures “[...] in which month and year [...] your cohabitation [ended] (Bundesinstitut für Bevölkerungsforschung 2002b, p. 7)” and for which reason. If the reason is a divorce, then the end of marriage is coded “-77 Separation/divorce, no exact date”, since merely the cohabitation ended, the date of divorce, however, is unknown. Hereby, it is possible to use the end of cohabitation as partnership end or to impute the end in some other way. Moreover, the Socio-Economic Panel (data source biomarsy) contains partnerships for which it is unclear if the marriage ended by divorce or death of partner (“-88 Union dissolution/death of partner”¹⁰). Thus, no assumptions can be made regarding the partnership end because neither the end of cohabitation nor the end of relationship are available. For all other survey programs, the end of marriage is given in case of a divorce.

When using end of cohabitation as end of partnership, missing data of cohabitation end in the ALLBUS-Cumulation, the Generations and Gender Surveys (2005,2008) and in parts of the Family Surveys (Family Survey I. Wave West 1988, II. Wave 1994, III. Wave 2000), the Life History Studies (GLHS-West II T - Telephone Interview, GLHS-GDR, GLHS-East 71, GLHS-West 64/71, GLHS-Panel 71) and the Socio-Economic Panel (data source biomarsy) must be considered. In the Family Survey I. Wave East 1990 the end of cohabitation is only provided for married couples. Also, in the GLHS-West I the end of cohabitation is only given for current marriages and current cohabiting, and non-married partnerships. In the GLHS-West III it only exists for currently separated or, respectively, divorced marriages. Furthermore, the end of cohabitation is not available for one respective partnership in both birth cohorts of the GLHS-East Panel Study, as described in more detail in the

¹⁰ See Hamjediers et al 2020.

table. For all the studies not mentioned in this section the end of cohabitation – given that one of the partners in fact moved out – is provided.

The end of relationship is not obtainable for any partnership in the ALLBUS-Cumulation and in the GLHS-West I. As a result of the form of questioning for cohabiting, non-married and LAT-partnerships, no distinction between the death of partner and the end of relationship is possible in the Family Surveys. There, the year of the end of partnership is assigned to the end of relationship. In the Mannheim Divorce Study, the definition of separation is based on the date of the “separation from table and bed” (Esser, Gostomski & Hartmann 2018b, p. 21; Questions 175 and 176). This refers to the end of the marital union according to §1353 BGB (BGB 2017, p. 315). According to §1567 BGB living separately does not necessarily imply moving out but can also be exercised “[...] within the marital home (BGB 2017, p. 342)”. Since different questions ask for the move-out from the shared home and the separation from table and bed, it can be assumed that the latter relates to the end of relationship, even if both spouses live in the same home. In the Fertility and Family Survey (Germany) the end of relationship is coded “-77 Separation/ divorce, no exact date” if “separation” is stated as the reason for the end of cohabitation (see end of marriage above). In addition, in the GLHS-West III valid values for the end of relationship only exist for current marriages, which are separated but not divorced. For both birth cohorts in the GLHS-East Panel, the end of relationship is not available. Nonetheless, for partnerships reported in Wave 1 and again in Wave 2 the end of relationship can be extracted from Wave 1, in case that the separation already occurred in Wave 1. These partnerships are reported again in Wave 2 and can be updated with information about the end of marriage and cohabitation from Wave 2. In the SHARE, the end of relationship is only provided for LAT-partnerships. Likewise, it is only partially available in the SOEP: in the data source `biomarsy` only for those, who stated “married, separated”. For all other studies not mentioned here, the end of relationship is provided if a separation happened.

5.4 Different reference population

The (sub-)studies vary regarding the targeted population. The studies differ concerning surveyed areas (East and/or West Germany), included age groups and cohorts as well as targeted populations (e.g. limited samples to only Germans or including foreigners as well) (see *Table 1* in *chapter 2.1*). Similar to Dual-Frame samples, in which people are contacted through landline numbers as well as through mobile phone numbers and where the higher probability of inclusion of the former has to be corrected, a subset of people with certain characteristics has a positive probability of inclusion in several surveys. The literature about the design based (Neyman 1934) handling of multi-frame samples, e.g. with the use of composite factors that (further) weight observations, is vast. A literature review can be found in Xia, Pedlow and Davern (2010). On the other hand, for model-based inference (Fisher 1922), it is possible to include covariates (and corresponding interactions) into the model to control for differing probabilities of inclusion and participation (Pfeffermann & Sverchkov 2009). What control variables are required depends on the model used. Therefore, no general advice for the inclusion of covariates or weights can be given at this point. The topic of different target populations is closely linked to the topic of survey weights of the distinct surveys, which will be discussed in the following *section 5.5*. This section also gives examples of the design-based, hybrid (design-weighted regressions) and model-based approaches.

5.5 Dealing with survey weights

Survey weights exist in the most original data sources. However, these are not included in the HaSpaD- target data set but can be added by the user via the original IDs in the HaSpaD- target data set.

Some considerations and advice on the use of survey weights are presented below. These are not exhaustive due to the complex structure and the wide range of analysis possibilities.

In general, survey weights serve to reduce the sampling bias for estimators and to correct for non-coverage or non-response. Usually, survey weights are further divided into design weights, non-response weights and poststratification weights. While design weights help to correct uneven selection probabilities, non-response weights correct non-response if at least some data about the interviewed and those that refused to be interviewed is available, e.g. remarks about the living environment of the person., made by the interviewer. Post-stratification weights are used to adapt the sample distribution to the population distribution in terms of specific variables. This strategy assumes these variables to be closely correlate with possible target variables (Survey Research Center 2016, p. 657). Also, weights can be combined for the purpose of analysis. As an example, the ALLBUS-Cumulation offers three different person-related weights, one person-related East-West weight that balances the overrepresentation of individuals from East Germany and a transformation weight (a form of a design weight) for person-related statements, as the ALLBUS-Cumulation has been a household sample for several years. Additionally, the product of the transformation weight and the East-West weight forms another combined weight. On the other hand, the Socio-Economic Panel includes a post-stratification weight besides the transformation weight. Its purpose is to adapt the sample of the population (people older than 16 years in Federal Republic of Germany) concerning age groups, civil status, sex and nationality.

Generally, all original data sources included use design weights, post-stratification weights or the combined weights of both. The Family Survey, the Generations and Gender Survey and SHARE have both design weights and post-stratification weights for at least some survey years.

Throughout statistics, the use of survey weights is heavily debated (Gelman 2007, Lavallée & Beaumont 2015; for an introduction see Gabler and Häder 2016; Lavallée and Beaumont 2016, Tillé and Matei 2016, Vannette and Krosnick 2018). On the one hand, survey weights can reduce coverage, sampling and non-response bias. On the other hand, weights can significantly increase the variance of estimators. Hence, the Survey Research Center (2016, p. 659) recommends not to use post-stratification weights if using weights does not dramatically change the point estimators. Furthermore, it is possible to introduce variables used for poststratification or design weights as covariates (including the interaction term with other independent variables, if applicable) into the regression model instead of weighting individual cases in the regression estimation (Winship & Radbill 1994).

An illustrating example: if the correlation between partnership and life satisfaction does not differ systematically between East- and West Germany and it is intended to estimate both variables for Germany as a whole, design weights, such as provided in the ALLBUS-Cumulation in order to correct the overrepresentation of couples from East Germany, are not required. The weights would only impact the estimated variance and increase it. In opposition to that, design weights to correct the overrepresentation of East Germany are necessary if the correlation between the education attainment and the life satisfaction differ between East and West Germany and the interest lies in creating an estimator for Germany as a whole. Otherwise, the point estimator would result biased. However, the East-West weights from the ALLBUS-Cumulation are not required neither if separated analyses are conducted for East and West Germany or, as aforementioned, if an East-West variable

(if necessary, including the interaction term with other independent variables) is added to the statistical models.

For the analysis of a harmonized data set, these general considerations regarding possible bias and variance can only be the first step. In comparison to a single data set, the HaSpaD-data set contains data from various decades, in other words, populations at different points of time are compared. Furthermore, in comparative social research two questions emerge: (1) How should we deal with weights included in the data sets when comparing several, country-specific populations? And (2) should all observations within a survey also receive an additional survey-specific weighting factor? Concerning the problem of weighting in comparative cases an overview paper from Joye, Sapin and Wolf (2019) was published in the Survey Data Harmonization Newsletter. We will transfer some recommendations to HaSpaD and other harmonization projects, which harmonize over time but not across countries.

For the first point, the use of survey weights stemming from the original data sources, Joye, Sapin and Wolf (2019) recommend a thorough examination whether the use of weights facilitates the comparison between populations or not. An example: the surveys in the HaSpaD project partially have different populations, e.g. they comprise only German citizens (like the Family Surveys) or the general population over 17 years in private households (Socio-Economic Panel). If foreign citizens in the Socio-Economic Panel hold a lower probability of response and thus, are assigned a post-stratification weight, the comparison between both surveys becomes complicated.

The HaSpaD dataset also considers a sub-population (persons in partnerships). Weighted variance estimates for a subpopulation are not trivial and require information on the number of excluded units and their survey weights (see West, Berglund & Heeringa 2008, and Graubard & Korn 1996).

Regarding an additional weight for individual surveys, Joye, Sapin and Wolf (2019) refer to Kish (1999, reprint 2003) and suggest six different approaches. We discuss these using an example that is conceivable with HaSpaD data. We assume that couples in Germany are to be compared in different years or decades using survey data from different survey programs. The following options available:

1. No comparison between surveys from different populations, as differences between the surveys in terms of population are considered too serious
2. Harmonization of the data, but no analysis with the combined data set. Estimators for each survey are then calculated separately and compared. The disadvantage of this procedure is that the variance estimators can become quite large for small data sets. In the case of the HaSpaD project, for example, such a data set is GLHS-East 71.
3. Each data set in the analysis based on the combined data set is equally weighted. As a result, data sets with a high number of respondents will have a greater weight than data sets with a lower number of respondents. For example, if the proportion of same-sex couples in all couples since 2000 is to be estimated and two surveys from 2000 and 2010 with different case numbers have information on same-sex couples, then the estimates are biased if the proportion has changed between 2000 and 2010. However, if the proportion has remained the same, weighting the samples equally gives a smaller variance estimator than weighting the surveys with the size of the sample (see next point), so weighting is not necessary or even not recommended.
4. Another possibility is to weight the surveys with the sample size, or with the effective sample size (after weighting with the included design and post-stratification weights). Surveys with above-average case numbers are given a smaller weight than surveys with below-average case numbers.
5. The use of population scaled weights (Kaminska & Lynn 2017) and careful consideration of what the target reference population is (voters, residents, nationals, etc.) is a fifth option. This scaling is particularly recommended for comparative research when combining samples with very unequal

population sizes. For example, if 1000 respondents come from European countries as different in size as Germany, Estonia or Italy, the Estonian data will have too much influence on a pan-European estimator, depending on the research question, when weights are not scaled. When combining data on couples in Germany over time, this suggestion will be less important.

6. Finally, Joye, Sapin and Wolf (2019) suggest the estimation of a multilevel model without population weights, but with survey weights. This allows differences between surveys to be modelled directly by including random or fixed effects. The two levels of the model are then persons/couples on the one hand and surveys on the other. However, the multilevel models with weighting, as recommended by Joye, Sapin and Wolf (2019), may require a correction of the survey weights (see Carle 2009 for details).

The previous considerations make clear that the usage of survey weights is not always desirable. Nevertheless, if they are necessary, their selection will always depend on the research question. For this reason, no “HaSpaD” weights or harmonized weights exist and it is up to the user to add weights from the original data sources by their ID (as it is the case for additional variables).

5.6 Imputation of dates in the HaSpaD data set

Dates in the harmonized HaSpaD data set are provided in century months (the value 1 is equal to January 1900). Insofar, as dates are reported for month and year, no imputation was required for the transformation into century months (“0 No imputation”). In case that no information about the month was presented in the original data source (due to memory difficulties of the interviewed or because it was only asked for the information about the year), the month was imputed according to the following rules:

- (1) If only the year is reported or asked in the questionnaire, “January” is imputed for the beginning-variables and “December” for the end-variables as month. The respective imputation-variable is assigned the value “1 Only year information”.
- (2) If respondents can state a season instead of the month given memory difficulties, the month is imputed as indicated in *Table 15*. This applies for the Mannheim Divorce Study and the Life History Study GLHS-West 64/71 and GLHS-Panel 71. The corresponding imputation-variable has the same missing code, as soon as the related biography variable contains a missing value.¹¹

¹¹ If one partner dies in the marriage, the end of the marriage and the corresponding flag-imp variable are assigned the following: `marend = "-66 Death Partner"; flag_marend_imp = "-66 Death partner"`.

Table 15: Imputed values of the beginning and end variables for seasonal dates

Study	Value and label of the source variables	Value and label of the beginning-variable	Value and label of the end-variable
Mannheim Divorce Study	17 Begin of the year	1 January	4 April
	18 Middle of the year	5 May	8 August
	19 End of the year	9 September	12 December
GLHS-West 64/71	21 Beginning of the year	1 January	2 February
	24 Spring/Eastern	3 March	5 May
	27 Summer	6 June	8 August
	30 Fall	9 September	10 October
	31 End of the year	11 November	12 December
GLHS-Panel 71	32 Winter	11 November	12 December
	21 Beginning of the year /Winter	1 January	2 February
	24 Spring/Eastern	3 March	5 May
	27 Middle of the year/Summer	6 June	8 August
	30 Fall	9 September	10 October
	32 End of the year/Winter	11 November	12 December

In the Mannheim Divorce Study – in case that respondents can only remember the season – the date of the event is retrieved in the categories “Beginning”, “Middle” and “End” of the year. In this special case, for the beginning-variables the values “1 January” (beginning), “5 May” (middle) and “9 September” (end of the year) were adopted. For the end-variables the coding is “4 April” for the beginning, “8 August” for the middle and “12 December” for the end of the year. The value “2 Only seasonal information” is given to the respective imputation-variable. Following a similar logic, the imputation is conducted for the Life History Study GLHS-West 64/71 and GLHS-Panel 71 (see table 15). Likewise, pairfam contains biography data on a monthly, yearly and seasonal basis. Nonetheless, the original data sources are already imputed in century months. The scheme for the marking of the degree of imputation in special flag-variables was adopted from pairfam. A list of the degrees of imputation of all biography variables per sub-study is presented in *appendix A. 4*.

5.7 Longitudinal processing of panel surveys without partner ID (Life History Studies, Generations and Gender Survey)

Initial situation

Some sub-studies of the Life History Study (GLHS-East Panel, GLHS-Panel 71) as well as the second wave of the Generations and Gender Survey 2008 are panel surveys of a subsample of the respondents from the respective first wave (GLHS-GDR, GLHS-East 71/GLHS-West 64/71; Generations and Gender Survey 2005). In the follow-up survey it is possible that respondents state the same partnership as already reported in the first wave. However, a partner ID is missing to identify the partner across two waves. A step-by-step approach according to plausibility criteria was chosen. This procedure was applied to update couple biographies at the time of wave 1 with values from wave 2 and to reduce the problem of partnership duplicates (double reported partnerships).

Procedure of processing

Initially, both partnerships reported as already completed in wave 1 and partnerships reported in wave 2 that started after the interview date of the first wave were excluded from further longitudinal

data processing (see *Figure 6*). These can be added to the HaSpaD data set without longitudinal linkage of the information from the different panel waves. After the exclusion of these partnerships, only those remain that are stated in wave 1 and are ongoing until the date of interview of wave 1 and those that are reported in wave 2 and started before the date of interview of wave 1.

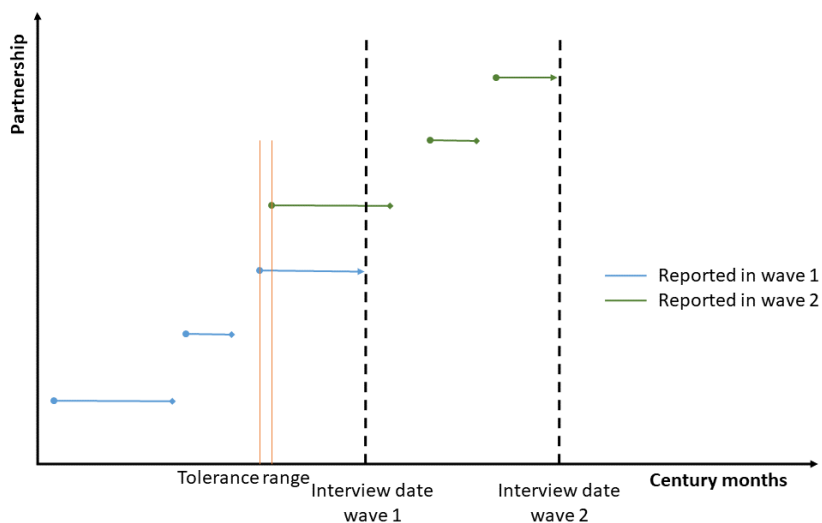


Figure 6: Procedure for linking equal partnerships

Based on filter questions, for some respondents of the LV-East Panel as well as of the second wave of the Generations and Gender Surveys (2008), identical partnerships were identified. For the former, marital status and number of partnerships since December 1989 served as filter questions (LV East Panel). For the latter, the question of whether the respondents still live with their partner or spouse indicated in wave 1 served to identify identical partners. (GGs, Leven 2009, S. 22, Frage 305).

For the remaining partnerships, the dates in both survey waves were used to check whether an identical partnership was plausible (for example: start of marriage of both partnerships identical, see *Table 16*) and whether this could be updated using the information in wave 2. If an identical partnership did not seem to exist due to the plausibility criteria, the partnership existing in wave 1 was not updated longitudinally with the information from wave 2. Partnerships that were reported in wave 2 and started before the survey time in wave 1, but for which (against the background of the formulated plausibility criteria) no match could be found with the information from wave 1, are marked in the variable Problem with "2 Problems due to panel linkage (GLHS, GGS)" or "4 Data inconsistency + problems due to panel linkage (GLHS, GGS)". Often these partnerships are likely to be identical to those reported in wave 1 but affected by respondents' memory (e.g. incorrect reporting of the year of marriage in one of the survey waves). The variable problem serves to identify these partnerships and, if necessary, to process them further or to remove them from the target data set (see chapter 6.1.5).

Table 16 gives an overview of plausibility assumptions the identification of identical partnerships across longitudinal data has been based on:

Table 16: Assumptions about plausibility for the identification of identical partnerships reported in wave 1 and 2

Case	Prerequisite	Conditions that must be fulfilled
1	Valid information in century months about the marriage begin in wave 1 (<i>marbeg_w1</i>) and in wave 2 (<i>marbeg_w2</i>) exist.	The marriage begins in wave 1 and 2 coincide within the assumed limits (see <i>Table 17</i>): if this criterion is fulfilled, the reported partnership in wave 1 and respectively in wave 2 is the same and can be updated.
2	Valid information in century months about the relationship begin and cohabitation begin reported in wave 1 (<i>relbeg_w1</i> and <i>cohbeg_w1</i>) and 2 (<i>relbeg_w1</i> and <i>cohbeg_w2</i>) exist. A valid information in century months about the marriage begin is not provided in both waves.	The relationship begins reported in wave 1 and 2 AND the cohabitation begins reported in wave 1 and 2 coincide within the assumed limits (see <i>Table 17</i>): if this criterion is fulfilled, the reported partnership in wave 1 and respectively in wave 2 is the same and can be updated.
3	Valid information in century months about the relationship begin reported in wave 1 (<i>relbeg_w1</i>) and 2 (<i>relbeg_w2</i>) exist. A valid information in century months about the marriage begin and cohabitation begin is not provided in both waves.	The relationship begins in wave 1 and 2 coincide perfectly : if the relationship begin perfectly coincides in both waves, the reported partnership in wave 1 and respectively in wave 2 is the same and can be updated.
4	Valid information in century months about the cohabitation begin reported in wave 1 (<i>cohbeg_w1</i>) and 2 (<i>cohbeg_w2</i>) exist. A valid information in century months about the marriage begin and relationship begin is not provided in both waves.	The cohabitation begins in wave 1 and 2 coincide perfectly: if the cohabitation begin perfectly coincides in both waves, the reported partnership in wave 1 and respectively in wave 2 is the same and can be updated.

marbeg_w1 = marriage begin reported in wave 1, *marbeg_w2* = marriage begin reported in wave 2, *cohbeg_w1* = cohabitation begin reported in wave 1, *cohbeg_w2* = cohabitation begin reported in wave 2, *relbeg_w1* = relationship begin reported in wave 1, *relbeg_w2* = relationship begin reported in wave 2

In case 1 and 2 of the table above, certain limits are assumed that restrain the values of the respective beginning-variable in order to be identifiable as identical partnership. In the following *Table 17*, upper and lower limits are presented:

Table 17: Determination of upper and lower limits according to the combination of the level of imputation

Case	Beginning-variable reported in wave 1	Value of flag_begin_imp	Beginning-variable reported in wave 2	Value of flag_begin_imp	L_limit (lower limit in century months)	u_limit (upper limit in century months)
1	No imputation	0	No imputation	0	-3	3
2	No imputation	0	Only season information	2	-3	3
3	No imputation	0	Only year information	1	0	11
4	Only season information	2	No imputation	0	-3	3
5	Only season information	2	Only season information	2	-3	3
6	Only season information	2	Only year information	1	0	10
7	Only year information	1	No imputation	0	-11	0
8	Only year information	1	Only season information	2	-10	0
9	Only year information	1	Only year information	1	0	0

Beginning-variable in wave 1: *relbeg_w1*, *cohbeg_w1*, *marbeg_w1*

Beginning-variable in wave 2: *relbeg_w2*, *cohbeg_w2*, *marbeg_w2*

Only beginning-variables with the same degree of institutionalization are compared (e.g.: *relbeg_w1* with *relbeg_w2*).

There are nine distinguishable variations of the degree of imputation for both beginning-variables. The values of both waves 1 and 2 can be reported on a monthly basis (lowest degree of imputation, case 1) up to a yearly basis (highest degree of imputation, case 9). Mixed combinations of the three degrees of imputation can also occur. The lower and the upper limit indicate the minimal and maximal difference (in century months) between both beginning-variables, which allows to declare a partnership reported in wave 1 and 2 identical.

The following examples illustrate how the difference between *marbeg_w1* and *marbeg_w2* is determined for the nine variations of the degree of imputation:

In Table 18, examples for *marbeg_w1* and *marbeg_w2* are presented, whose difference $\text{marbeg}_w1 - \text{marbeg}_w2$ lies within the range of tolerance of the upper and lower limit (both limits included) and hence, can be identified as identical.

If there is only information about the year for a beginning-variable, then it is coded as January. Thus, in case 3, the difference between the monthly based date and the yearly based date can be a maximum of 11, given that the marriage begin was reported in the same year. An identical partnership is identified if the same year is denoted in both waves.

Table 18: Example of the determination of the upper and lower limits after the combination of the degree of imputation

Case	marbeg_w1	marbeg_w2	l_limit (lower limit)	marbeg_w1 – marbeg_w2	u_limit (upper limit)
1	January 1990 1081	December 1989 1080	-3	1 (= 1081-1080)	3
2	January 1990 1081	Winter 1989 1079	-3	2	3
3	December 1990 1092	Year 1990 1081	0	11	11
4	Winter 1989 1079	January 1990 1081	-3	-2	3
5	Winter 1989 1079	Start of the year 1990 1081	-3	-2	3
6	Winter 1990 1091	Year 1990 1081	0	10	10
7	Year 1990 1081	December 1990 1092	-11	-11	0
8	Year 1990 1081	Winter 1990 1091	-10	-10	0
9	Year 1990 1081	Year 1990 1081	0	0	0

First row in column marbeg_w1/marbeg_w2: displayed as month/season/year

Second row in column marbeg_w1/marbeg_w2: displayed in century months

Table 19 describes the coding for seasonal information within a year. Accordingly, winter 1989 is coded as November 1989 (century months = 1079). In comparison to Table 18, there is a maximum and minimum tolerance of 3 and -3 for the degree of imputation “2 Only seasonal information” of the beginning-data in both waves. Therefore, an identical partnership is identified, when the distance between the data is at most one season (example case 5 in Table 18: Winter 1989 reported in wave 1, start of the year reported in wave 2).

Table 19: Coding of the months given seasonal information

Season	Coding of beginning-variables
Start of year	1: January
Spring/Eastern	3: March
Summer	6: June
Fall	9: September
End of year	11: November
Winter	11: November

6 Overview of the harmonized variables

6.1 Biography data

The biography data set consists in ID variables to identify the survey program and sub-study, the anchor person (respondent) and the partner. Furthermore, it contains information about the (latest) interview date, the beginning and the end of relationship, cohabitation and marriage episodes of a partnership as well as the date of the partner's death in case of a decease during the partnership. These dates are available in century months. Moreover, for all of these dates the information from the original data source and the degree of imputation are integrated as single variables. The missing and valid values in the form of values and the maximum and minimum values as stated in the tables are related to the complete HaSpaD-target data set, which includes all survey programs.

6.1.1 ID variables

The ID variables serve to identify single respondents, data sources and partnerships and are discussed in the following.

Table 20: ID variable: `sid`

Variable label:	Study id
Missing values	none
Valid values:	
101	Panel Analysis of Intimate Relationships and Family Dynamics (pairfam)
102	German General Social Survey (ALLBUS-Cumulation)
103	Family Survey (FS)
104	Mannheim Divorce Study (MSS)
106	Fertility and Family Survey (FFS)
107	German Life History Study (GLHS)
108	Generations and Gender Survey (GGS)
109	The Survey of Health, Ageing and Retirement in Europe (SHARE)
110	Socio-Economic Panel (SOEP)
Data type:	numeric
Explanation:	The variable <code>sid</code> returns the identification number of all used survey programs.

Table 21: ID variable: did

Variable label:	Dataset id
Missing values:	none
Valid values:	
1	<ul style="list-style-type: none"> - Couple biography derives from only one data source (pairfam, ALLBUS-Cumulation, Mannheim Divorce Study, Fertility and Family Survey, SHARE). - Couple biographies derive from the Life History Study GLHS-West I. - Couple biographies derive only from Generations & Gender Survey (2005). - Couple biographies derive from the data set <code>biomarsy</code> of the Socio-Economic Panel.
2	<ul style="list-style-type: none"> - Couple biographies derive from the Family Survey II. Wave 1994. - Couple biographies derive from the follow-up survey of the Generations and Gender Survey (2008)
3	<ul style="list-style-type: none"> - Couple biographies derive from the Family Survey III. Wave 2000. - Couple biographies derive from the Life History Study GLHS-West III.
11	<ul style="list-style-type: none"> - Couple biographies derive from the first survey of the Family Survey in West Germany.
12	<ul style="list-style-type: none"> - Couple biographies derive from the first survey of the Family Survey in East Germany. - Partnerships, which were reported in the first wave of the Generations & Gender Survey and also were reported and updated in the second wave of the Generations & Gender Survey
21	<ul style="list-style-type: none"> - Couple biographies derive from the Life History Study GLHS-West II A – Personal Interview.
22	<ul style="list-style-type: none"> - Couple biographies derive from the Life History Study GLHS-West II T – Telephone Interview.
40	<ul style="list-style-type: none"> - Couple biographies derive from the Life History Study GLHS-GDR.
50	<ul style="list-style-type: none"> - Couple biographies derive from the Life History Study GLHS-East 71.
61	<ul style="list-style-type: none"> - Couple biographies derive from the Life History Study GLHS-East Panel (Birth cohorts 1939-41, 1951-53, 1959-61).
62	<ul style="list-style-type: none"> - Couple biographies derive from the Life History Study GLHS-East Panel (Birth cohort 1929-31).
70	<ul style="list-style-type: none"> - Couple biographies derive from the Life History Study GLHS-West 64/71.
80	<ul style="list-style-type: none"> - Couple biographies derive from the Life History Study GLHS-Panel 71.
4061	<ul style="list-style-type: none"> - Partnerships reported in the Life History Study GLHS-GDR and reported and updated in the Life History Study GLHS-East Panel (Birth cohorts 1939-41, 1951-53, 1959-61).
4062	<ul style="list-style-type: none"> - Partnerships reported in the Life History Study GLHS-GDR and reported and updated in the Life History Study GLHS-East Panel (Birth cohort 1929-31).
5080	<ul style="list-style-type: none"> - Partnerships reported in the Life History Study GLHS-East 71 and reported and updated in the Life History Study GLHS-Panel 71.
7080	<ul style="list-style-type: none"> - Partnerships reported in the Life History Study GLHS-West 64/71 (only birth cohort 1971) and reported and updated in the Life History Study GLHS-Panel 71.

Variable label:	Dataset id
Data type:	numeric
Explanation:	The variable <code>did</code> indicates the data set ID within each survey program. If only one data source for the creation of couple biographies exists, this variable is assigned the value 1.
Construction specifics:	<p>If various studies within a survey program were used to create the cumulated data set (Family Survey, Life History Study, Generations & Gender Survey), the data set IDs are ordered chronologically according to the survey year. A concatenation of two data set IDs (see the identification numbers with four digits in this table in “valid values”) is conducted, when respondents participated in two waves of a survey program and reported partnerships at both waves.</p> <p>An exception regarding the construction of the <code>did</code> is the Socio-Economic Panel: the information for the construction of the cumulated data set stems from three data sources in the Socio-Economic Panel, which contain data about the couple biographies and the civil status of the respondent and are not structured in a chronological order based on the date of survey.</p>

Table 22: ID variable: `aid`

Variable label:	Anchor id
Missing values:	none
Minimum:	10322
Maximum:	10212085082000
Data type:	numeric
Explanation:	The variable <code>aid</code> indicates the anchor-ID. This value is a result of the concatenation of the <code>sid</code> , <code>did</code> and <code>id</code> of a respondent (<code>sid + did + id</code>). It enables to identify the respondent for all individual studies.
Construction specifics:	<p>For the panel linkage of specific Life History Studies (GLHS-GDR with both GLHS-East Panel, GLHS-East 71 with GLHS-Panel 71, GLHS-West 64/71 with GLHS-Panel 71) and the Generations & Gender Survey (2005) with the Generations & Gender Survey (2008), the <code>did</code> of the first wave is always used to create the <code>aid</code>.</p> <p>As the original anchor ID contains duplicates for the single cross-sections of the ALLBUS-Cumulation, the <code>aid</code> is composed of the linkage of the original anchor ID and the respective survey year.</p>

Table 23: ID variable: `id`

Variable label:	Original anchor id
Missing values:	none
Data type:	string
Explanation:	The variable <code>id</code> indicates the original identification number as contained in the original data sources.

Table 24: ID variable: pid

Variable label:	Partner id
Missing values:	none
Data type:	string
Explanation:	The variable pid serves as link variable of a partner variable with the corresponding partnership.
Construction specifics:	In the case of origin studies without an identification feature of a partnership, the pid is constructed from the combination of aid and partner_pointer (aid + partner_pointer). For source studies that allow for an individual identification of a partner by means of an ID (pairfam, SHARE and Socio-Economic Panel), information on the source study is concatenated with the original partner ID (sid + did + pid_o). The partner ID is assigned aid + partner_pointer for multi-actor studies (pairfam, SHARE and Socio-Economic Panel) if no information on the original partner ID is available.

Table 25: ID variable: pid_o

Variable label:	Original partner id
Missing values:	
-10	Missing by study design
Data type:	string
Explanation:	The variable pid_o contains the original identification number of the respondent's partner as assigned in the original data set.

Table 26: ID variable: partner_pointer

Variablenlabel:	Partner pointer
Missing values:	None
Minimum	0
Maximum	17
Data type:	numeric
Explanation:	The variable partner_pointer indicates the position of the partner information in the original data (e.g. first, second, third spell; first, second, third set of variables regarding partnerships).

6.1.2 Biography variables

Table 27: Biography variable: `intd`

Variable label:	Date of interview (in century months)
Missing values:	
-10	Missing by study design
Minimum:	961
Maximum:	1395
Data type:	numeric
Explanation:	The variable <code>intd</code> indicates the date of interview in century months.
Construction specifics:	Insofar as the date of interview is unknown, the last month of the field phase is chosen.

Table 28: Biography variable: `relbeg`

Variable label:	Beginning relationship (in century months)
Missing values:	
-7	Incomplete data
-10	Missing by study design
Minimum:	325
Maximum:	1394
Data type:	numeric
Explanation:	The variable <code>relbeg</code> indicates the beginning of a relationship in a partnership in century months.

Table 29: Biography variable: reLend

Variable label:	End relationship (in century months)
Missing values:	
-7	Incomplete data
-10	Missing by study design
-66	Death partner
-77	Separation/divorce, no exact date
-99	Ongoing
Minimum:	481
Maximum:	1393
Data type:	numeric
Explanation:	The variable reLend indicates the end of a relationship in a partnership in century months.
Construction specifics:	If only the date of moving out and a separation as its reason are known, no exact separation date can be assigned because it does not necessarily coincide temporally with the end of cohabitation. Therefore, reLend is coded “-77 Separation/divorce, no exact date”.

Table 30: Biography variable: cohbeg

Variable label:	Beginning cohabitation (in century months)
Missing values:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
Minimum:	369
Maximum:	1394
Data type:	numeric
Explanation:	The variable cohbeg indicates the beginning of cohabitation in a common household with the respective partner in century months.

Table 31: Biography variable: cohend

Variable label:	End cohabitation (in century months)
Missing values:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
-66	Death partner
-99	Ongoing
Minimum:	484
Maximum:	1394
Data type:	numeric
Explanation:	The variable <code>cohend</code> indicates the end of cohabitation in a common household with the respective partner in century months.

Table 32: Biography variable: marbeg

Variable label:	Beginning marriage (in century months)
Missing values:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
Minimum:	145
Maximum:	1392
Data type:	numeric
Explanation:	The variable <code>marbeg</code> indicates the beginning of marriage in a partnership in century months.

Table 33: Biography variable: `marend`

Variable label:	End marriage (in century months)
Missing values:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
-66	Death partner
-77	Separation/divorce, no exact date
-88	Divorce/death of partner
-99	Ongoing
Minimum:	264
Maximum:	1391
Data type:	numeric
Explanation:	The variable <code>marend</code> indicates the end of marriage in a partnership in century months.
Construction specifics:	If only the date of moving out and a divorce as its reason are known, no exact divorce date can be assigned because it does not necessarily coincide temporally with the end of cohabitation. Therefore, <code>marend</code> is coded “-77 Separation/divorce, no exact date”. The value “-88 Union dissolution/death of partner” is assigned if it is not clear whether a marriage ended with the death of partner or with divorce. ¹²

¹² „SPELLTYP has one additional category ‘divorced or widowed’ which indicates that a marriage definitely ended, though we do not know whether via divorce or death of the spouse. This may be due to missing information from the biographical questionnaires or due to a respondent’s frequent shifts between both categories without ever reporting the death of the partner or divorce as an event.” (Hamjediers et al 2020)

Table 34: Biography variable: dop

Variable label:	Date of partner's death (in century months)
Missing values:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
-88	Divorce/death of partner
Minimum:	180
Maximum:	1390
Data type:	numeric
Explanation:	The variable dop indicates the date of death of the respective partner in century months.

6.1.3 Imputed variables

The imputed variables (*imp*-variables) show the degree of imputation of the biography variables.

Table 35: Imputed variable: flag_relbeg_imp

Variable label:	Imputed date of beginning relationship
Missing values:	
-7	Incomplete data
-10	Missing by study design
Valid values:	
0	No imputation
1	Only year information
2	Only season information
Data type:	numeric
Explanation:	The variable flag_relbeg_imp indicates the degree of imputation of the variable relbeg.

Table 36: Imputed variable: flag_relend_imp

Variable label:	Imputed date of end relationship
Missing values:	
-7	Incomplete data
-10	Missing by study design
-66	Death partner
-77	Separation/divorce, no exact date
-99	Ongoing
Valid values:	
0	No imputation
1	Only year information
2	Only season information
Data type:	numeric
Explanation:	The variable flag_relend_imp indicates the degree of imputation of the variable relend.

Table 37: Imputed variable: flag_cohbeg_imp

Variable label:	Imputed date of beginning cohabitation
Missing values:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
Valid values:	
0	No imputation
1	Only year information
2	Only season information
Data type:	numeric
Explanation:	The variable flag_cohbeg_imp indicates the degree of imputation of the variable cohbeg.

Table 38: Imputed variable: `flag_cohend_imp`

Variable label:	Imputed date of end cohabitation
Missing values:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
-66	Death partner
-99	Ongoing
Valid values:	
0	No imputation
1	Only year information
2	Only season information
Data type:	numeric
Explanation:	The variable <code>flag_cohend_imp</code> indicates the degree of imputation of the variable <code>cohend</code> .

Table 39: Imputed variable: `flag_marbeg_imp`

Variable label:	Imputed date of beginning marriage
Missing values:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
Valid values:	
0	No imputation
1	Only year information
2	Only season information
Data type:	numeric
Explanation:	The variable <code>flag_marbeg_imp</code> indicates the degree of imputation of the variable <code>marbeg</code> .

Table 40: Imputed Variable: flag_marend_imp

Variable label:	Imputed date of end marriage
Missing values:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
-66	Death partner
-77	Separation/divorce, no exact date
-88	Union dissolution/death of partner
-99	Ongoing
Valid values:	
0	No imputation
1	Only year information
2	Only season information
Data type:	numeric
Explanation:	The variable flag_marend_imp indicates the degree of imputation of the variable marend.

Table 41: Imputed Variable: flag_dop_imp

Variable label:	Imputed date of partner's death
Missing values:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
-88	Union dissolution/death of partner
Valid values:	
0	No imputation
1	Only year information
2	Only season information
Data type:	numeric
Explanation:	The variable flag_dop_imp indicates the degree of imputation of the variable dop.

6.1.4 Original variables

The original variables (*ori*-variables) include the non-harmonized variable values from the original data sources of the respective biography-variable and ensure the transparency of the coding of the source variables. Since most variables are based on calendar years (or sometimes months), these variables can offer a quick overview of the plausibility of certain statements from the respondents.

The rules for the construction of the *ori*-variables are explained in the following. If information about month and year are provided in the original data source, both values are concatenated as string variables (“month” “+” “year”) and result in the value of the respective *ori*-variable. If only the year is known (this means that only this information is asked for), the value is adopted without any concatenation. For instance, if the respondent states August 1989 as marriage begin (*marbeg_month* = 08; *marbeg_year* = 1989), both variables are concatenated in *flag_marbeg_ori* as “08+1989”. In the *appendix A. 5*, a more detailed list of all original variables is presented.

Apart from that one special case exist: There is (partially) no information for the *ori*-variables, for instance, if the required information is not asked for in the original data set. In this case, the *ori*-variable has the value “-10 Missing by study design”.

Table 42: Original variable: *flag_relbeg_ori*

Variable label:	Original date of beginning relationship
Missing values:	
-10	Missing by study design
Data type:	string
Explanation:	The variable <i>flag_relbeg_ori</i> contains the information about the beginning of relationship in the original data set.
Construction specifics:	In the Life History Study 4 the variable <i>flag_relbeg_ori</i> is calculated as the difference between the beginning of cohabitation and the duration of time that both respondent and partner had known each other previously. All required variables are concatenated.

Table 43: Original variable: *flag_relend_ori*

Variable label:	Original date of end relationship
Missing values:	
-10	Missing by study design
Data type:	string
Explanation:	The variable <i>flag_relend_ori</i> contains the information about the end of relationship in the original data set.
Construction specifics:	The data sets of the Socio-Economic Panel include time periods (spells), which mark the end of the respective phase of the partnership and thus mark events. Therefore, besides year-information, the <i>_ori</i> -variables additionally include censoring information since they report left- and right-censors. For more information, please also consult the html-documentation “coding_soep”.

Table 44: Original variable: `flag_cohbeg_ori`

Variable label:	Original date of beginning cohabitation
Missing values:	
-10	Missing by study design
Data type:	string
Explanation:	The variable <code>flag_cohbeg_ori</code> contains the information about the beginning of cohabitation in the original data set.

Table 45: Original variable: `flag_cohend_ori`

Variable label:	Original date of end cohabitation
Missing values:	
-10	Missing by study design
Data type:	string
Explanation:	The variable <code>flag_cohend_ori</code> contains the information about the end of cohabitation in the original data set.
Construction specifics:	<p>In the GLHS-West I <code>cohend</code> is constructed based on three variables. Hence, <code>flag_cohend_ori</code> in this case is a concatenation of the source variables “duration of separation before divorce”, “month of divorce” and “year of divorce”. In the GLHS-West II A – Personal Interview the duration of separation is adopted for the variable <code>flag_cohend_ori</code>. In case that a date for the divorce is reported by the respondent, “duration of separation before divorce” and “year of divorce” are concatenated.</p> <p>The data sets of the Socio-Economic Panel include time periods (spells), which mark the end of the respective phase of the partnership and thus mark events. Therefore, besides year-information, the <code>_ori</code>-variables additionally include censoring information since they report left- and right-censors. For more information, please also consult the html-documentation “coding_soep”.</p>

Table 46: Original variable: `flag_marbeg_ori`

Variable label:	Original date of beginning marriage
Missing values:	
-10	Missing by study design
Data type:	string
Explanation:	The variable <code>flag_marbeg_ori</code> contains the information about the beginning of marriage in the original data set.
Construction specifics:	<p>The data sets of the Socio-Economic Panel include time periods (spells), which mark the end of the respective phase of the partnership and thus mark events. Therefore, besides year-information, the <code>_ori</code>-variables additionally include censoring information since they report left- and right-censors. For more information, please also consult the html-documentation “coding_soep”.</p>

Table 47: Original variable: `flag_marend_ori`

Variable label:	Original date of end marriage
Missing values:	
-10	Missing by study design
Data type:	string
Explanation:	The variable <code>flag_marend_ori</code> contains the information about the end of marriage in the original data set.
Construction specifics:	The data sets of the Socio-Economic Panel include time periods (spells), which mark the end of the respective phase of the partnership and thus mark events. Therefor, besides year-information, the <code>_ori</code> -variables additionally include censoring information since they report left- and right-censors. For more information, please also consult the html-documentation "coding_soep".

Table 48: Original variable: `flag_dop_ori`

Variable label:	Original date of partner's death
Missing values:	
-10	Missing by study design
Data type:	string
Explanation:	The variable <code>flag_dop_ori</code> contains the information about the date of the death of partner in the original data set.
Construction specifics:	<p>In the Family Surveys, the variable <code>dop</code> only exists for married couples. Unmarried couples have the value "-10 Missing by study design".</p> <p>For the Fertility and Family Survey (Germany) has to be considered that <code>flag_dop_ori</code> can be concatenated as month and year of the end of cohabitation if the respondent states the death of partner as the reason for the separation. In case that the reason for separation is not the death of partner, <code>flag_dop_ori</code> is coded "-10 Missing by study design".</p> <p>The data sets of the Socio-Economic Panel include time periods (spells), which mark the end of the respective phase of the partnership and thus mark events. Therefor, besides year-information, the <code>_ori</code>-variables additionally include censoring information since they report left- and right-censors. For more information, please also consult the html-documentation "coding_soep".</p>

6.1.5 Additional flag variables

In this section, two additional flag variables are explained.

Table 49: Additional flag variable: `flag_beg`

Variable label:	Beginning partnership biography
Missing values:	
-10	Missing by study design
Valid values:	
0	With relationship
1	With cohabitation
2	With marriage
Data type:	numeric
Explanation:	The variable <code>flag_beg</code> indicates with which degree of institutionalization (degree of partnership) the couple biography starts in a data source.

Table 50: Additional flag variable: `problem`

Variable label:	Flag used to mark problem cases
Missing values:	none
Valid values:	
0	No data inconsistency
1	Data inconsistency
2	Problem due to panel linkage (GLHS, GGS)
3	Higher order partnerships to same partner (pairfam, FFS, SOEP)
4	Data inconsistency + problem due to panel linkage (GLHS, GGS)
5	Data inconsistency + higher order partnerships to same partner (pairfam, FFS)
6	Duplicate Multi-Actor-Design (SHARE)
Data type:	numeric
Construction specifics:	<p>The variable <code>problem</code> helps the user to identify inconsistencies, which already existed in the original data sources or were discovered during the data processing. In <i>appendix A. 6</i>, all values and data checks are discussed in more detail.</p> <p>The value „0 No data inconsistency“ indicates that no inconsistencies were found.</p> <p>The value „1 Data inconsistency“ indicates a problem with the dates stated in the couple biographies. For example, if the end of relationship occurred before its beginning, the biography data was surveyed inconsistently in the original data source.</p> <p>The value „2 Problems due to panel linkage (GLHS, GGS)“ is assigned to partnerships, for which problems occur during the linkage of the Life History</p>

Variable label:	Flag used to mark problem cases
-----------------	---------------------------------

Studies and the Generations & Gender Surveys. In *section 5.7* the process of panel linkages and related problems are discussed.

In pairfam and Fertility and Family Survey (FFS), the special case of several independent partnerships with the same partner can be found due to the assignment of the partnership ID (`pid`). In the event that the respondent states multiple partnerships with the same partner, all partnerships starting from the second are marked “3 Higher order partnerships to same partner (pairfam, FFS, SOEP)”.

The values “4 Data inconsistency + problems due to panel linkage (GLHS, GGS)” and “5 Data inconsistency + higher order partnerships to same partner (pairfam, FFS,)” are combinations of the value “1 Data inconsistency” with the values “2 Problems due to panel linkage (GLHS, GGS)” and, respectively, “3 Higher order partnerships to same partner (pairfam, FFS, SOEP)”. More details about the values and their meaning is presented in *Table A 5* in the appendix.

Due to the multi-actor design of SHARE, there are partnerships reported from both partners. To allow users to remove one of the two partnerships, the value “6 Duplicate Multi-Actor-Design (SHARE)” was assigned to the male respondents of such partnerships for heterosexual partnerships and to those respondents with a higher `aid` for homosexual partnerships.

6.2 Harmonized additional variables

Additional variables at respondent and partner level, such as dates of birth, citizenship, education, religious denomination, but also information on divorce and separation of the respondent's parents, are part of the harmonised variables offered in addition to the couple biographies. An overview of the processed variables can be found on the HaSpaD Wizard pages. In case further harmonized variables have been selected, documentation of codings and further literature references can be found in the respective HTML documentation under

`~haspad/1_documentation/harmonization_targets`. These documents are generated during the execution of the do file `main.do` and contain both the coding scheme and the individual processing documentation of the respective additional variables of all survey programs selected in the Harmonization Wizard. For example, if the user selects the anchor's sex (`sex`), the coding scheme and the harmonization process of all selected survey programs are documented under `~haspad/1_documentation/harmonization_targets/coding_sex.html`. Further information about the folder structure can be found in *chapter 3* (and additionally in *appendix A. 2*).

7 References

- Albers, I., Burgmeier, B., Preiss, K., Thum, M & Schneekloth, U. (1994). Familie und Partnerbeziehungen in der Bundesrepublik Deutschland (Familien- Survey 1994). Methodenbericht zur Haupterhebung 1994. München: Infratest Sozialforschung.
- Baumann, H. & Schulz, S. (2018). ALLBUS - Kumulation 1980-2016. Variable Report. GESIS Datenarchiv, Köln. Studie ZA4586. [doi:10.4232/1.13029](https://doi.org/10.4232/1.13029)
- Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S. & Zuber, S. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology* 42(4), 992–1001. doi: [10.1093/ije/dyt088](https://doi.org/10.1093/ije/dyt088)
- Brückner, E. (1993). Lebensverläufe und gesellschaftlicher Wandel. Konzeption, Design und Methodik der Erhebung von Lebensläufen der Geburtsjahrgänge 1919-1921. Max-Planck-Berlin: Institut für Bildungsforschung.
- Brückner, H., Mayer, K. U., Geis, A., Kortmann, K., Nuthmann, R. & Wehner, S. (1995). Lebensverläufe und gesellschaftlicher Wandel. Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1954-1956 und 1959-1961. Berlin: Max-Planck-Institut für Bildungsforschung.
- Brüderl, J., Drobníč, S., Hank, K., Neyer, F.J., Walper, S., Alt, P., Borschel, E., Bozoyan, C., Garrett, M., Geissler, S., Gonzalez Avilés, T., Gröpler, N., Hajek, K., Herzig, M., Huyer-May, B., Lenke, R., Lorenz, R., Lutz, K., Minkus, L., Peter, T., Phan, T., Preetz, R., Reim, J., Sawatzki, B., Schmiedeberg, C., Schütze, P., Schumann, N., Thönnissen, C., Timmermann, K., Wetzel, M. (2021a). The German Family Panel (pairfam). GESIS Data Archive, Cologne. ZA5678 Data file Version 12.0.0, [doi:10.4232/pairfam.5678.12.0.0](https://doi.org/10.4232/pairfam.5678.12.0.0).
- Brüderl, J., Garrett, M., Hajek, K., Herzig, M., Lenke, R., Lorenz, R., Lutz, Katharina, Phan, T., Schütz, P., Schumann, N. (2021b) pairfam Data Manual, Release 12.0, LMU Munich: Technical Report. GESIS Data Archive, Cologne. ZA5678 Data File Version 12.0.0, <https://doi.org/10.4232/pairfam.5678.12.0.0>.
- Brüderl, J., Schmiedeberg, C., Castiglione, L., Arránz Becker, O., Buhr, P., Fuß, D., Ludwig, V., Schröder, J. & Schumann, N. (2021c). The German Family Panel: Study Design and Cumulated Field Report (Waves 1 to 12). Release 12.0, LMU Munich, Technical report. Online document. https://www.pairfam.de/fileadmin/user_upload/redakteur/publis/Dokumentation/TechnicalPapers/TP01%20Cumulated%20Field%20Report%2C%20pairfam%202021.pdf [23/09/2021, 11:47]
- Bürgerliches Gesetzbuch - BGB (2017). Bürgerliches Gesetzbuch. Mit den Nebengesetzen zum Verbraucherschutz, Mietrecht und Familienrecht (15. Auflage). Regensburg: Walhalla Fachverlag.
- Bundesinstitut für Bevölkerungsforschung, Wiesbaden (2002a). Deutscher Fertility and Family Survey 1992. GESIS Datenarchiv, Köln. ZA3400 Datenfile Version 1.0.0. [doi:10.4232/1.3400](https://doi.org/10.4232/1.3400)
- Bundesinstitut für Bevölkerungsforschung, Wiesbaden (2002b). Deutscher Fertility and Family Survey 1992. GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA3400. [doi:10.4232/1.3400](https://doi.org/10.4232/1.3400)
- Carle, A. (2009). Fitting multilevel models in complex survey data with design weights: Recommendations. *BMC Medical Research Methodology* 9(49). [doi:10.1186/1471-2288-9-49](https://doi.org/10.1186/1471-2288-9-49)

- Deutsches Jugendinstitut (DJI), München (1992a). Wandel und Entwicklung familialer Lebensformen in Westdeutschland (Familiensurvey). GESIS Datenarchiv, Köln. ZA2245 Datenfile Version 1.0.0. [doi:10.4232/1.2245](https://doi.org/10.4232/1.2245)
- Deutsches Jugendinstitut (DJI), München (1992b). Wandel und Entwicklung familialer Lebensformen in Westdeutschland (Familiensurvey). GESIS Datenarchiv, Köln. Originaler Fragebogen zur Studie. ZA2245. [doi:10.4232/1.2245](https://doi.org/10.4232/1.2245)
- Deutsches Jugendinstitut (DJI), München (2000). Wandel und Entwicklung familialer Lebensformen - 3. Welle (Familiensurvey), GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA3920. [doi:10.4232/1.13198](https://doi.org/10.4232/1.13198)
- Deutsches Jugendinstitut (DJI), München (2018a). Familie und Partnerbeziehungen in Ostdeutschland (Familiensurvey). GESIS Datenarchiv, Köln. ZA2392 Datenfile Version 1.1.0. [doi:10.4232/1.13196](https://doi.org/10.4232/1.13196)
- Deutsches Jugendinstitut (DJI), München (2018b). Wandel und Entwicklung familialer Lebensformen - 2. Welle (Familiensurvey). GESIS Datenarchiv, Köln. ZA2860 Datenfile Version 1.1.0. [doi:10.4232/1.13197](https://doi.org/10.4232/1.13197)
- Deutsches Jugendinstitut (DJI), München (2018c). Wandel und Entwicklung familialer Lebensformen - 3. Welle (Familiensurvey). GESIS Datenarchiv, Köln. ZA3920 Datenfile Version 1.1.0. [doi:10.4232/1.13198](https://doi.org/10.4232/1.13198)
- Deutsches Jugendinstitut (DJI), München (2018d). Familie und Partnerbeziehungen in Ostdeutschland (Familiensurvey). GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2392. [doi:10.4232/1.13196](https://doi.org/10.4232/1.13196)
- Deutsches Jugendinstitut (DJI), München (2018e). Wandel und Entwicklung familialer Lebensformen - 2. Welle (Familiensurvey). GESIS Datenarchiv, Köln. Originaler Fragebogen zur Studie. ZA2860. [doi:10.4232/1.13197](https://doi.org/10.4232/1.13197)
- Eschenlohr, J., Haberkorn, M., von der Heyde, C., Preiss, K., von Rosenblatt, B., & Löhr, H. (1989). Familie und Partnerbeziehungen in der Bundesrepublik (DJI-Familiensurvey). Methodenbericht zur Haupterhebung 1988. München: Infratest Sozialforschung.
- Esser, H., Gostomski, C. Babka von & Hartmann, J. (2018a). Mannheimer Scheidungsstudie 1996. GESIS Datenarchiv, Köln. ZA3188 Datenfile Version 2.0.0. [doi:10.4232/1.13056](https://doi.org/10.4232/1.13056)
- Esser, H., Gostomski, C. Babka von & Hartmann, J. (2018b). Mannheimer Scheidungsstudie 1996. GESIS Datenarchiv, Köln. Codebuch zur Studie. ZA3188. [doi:10.4232/1.13056](https://doi.org/10.4232/1.13056)
- Esser, H., Gostomski, C. Babka von & Hartmann, J. (2018c). Mannheimer Scheidungsstudie 1996. GESIS Datenarchiv, Köln. Abschlussbericht zur Studie. ZA3188. [doi:10.4232/1.13056](https://doi.org/10.4232/1.13056)
- Fisher, R. (1922). On the Mathematical Foundations of Theoretical Statistics. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 222(594-604), 309–368.
- Fokkema, T., Kveder, A., Hiekel, N., Emery, T. & Liefbroer, A. C. (2016). Generations and Gender Programme Wave 1 data collection: An overview and assessment of sampling and fieldwork methods, weighting procedures, and cross-sectional representativeness. Demographic Research 34(18), 499-524. [doi:10.4054/DemRes.2016.34.18](https://doi.org/10.4054/DemRes.2016.34.18)
- Gabler, S. & Häder, S. (2016). Special challenges of sampling for comparative surveys. In C. Wolf, D. Joye, T. W. Smith & Yang-chih Fu (Hrsg.), The SAGE Handbook of Survey Methodology (S. 346-355) Los Angeles: SAGE. [doi:10.4135/9781473957893](https://doi.org/10.4135/9781473957893).

- Gauthier, A. H., Cabaço, S. L. F. & Emery, T. (2018). Generations and Gender Survey study profile. *Longitudinal and Life Course Studies*. *International Journal* 9 (4), 456-465. [doi:10.14301/lcs.v9i4.500](https://doi.org/10.14301/lcs.v9i4.500)
- Gelman, A. (2007). Struggles with Survey Weighting and Regression Modeling. *Statistical Science* 22 (2), 153-164. Institute of Mathematical Statistics. [doi:10.1214/088342306000000691](https://doi.org/10.1214/088342306000000691)
- GESIS - Leibniz-Institut für Sozialwissenschaften (2018). Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS - Kumulation 1980-2016. GESIS Datenarchiv, Köln. ZA4586 Datenfile Version 1.0.0. [doi:10.4232/1.13029](https://doi.org/10.4232/1.13029)
- GGG Wave 1 ([DOI: 10.17026/dans-z5z-xn8g](https://doi.org/10.17026/dans-z5z-xn8g)), see [Gauthier, A. H. et al. \(2018\)](#) or visit the GGP website (<https://www.ggp-i.org/>) for methodological details. Working Version: GGS Wave 1 Version 4.3. – Consolidated identification: GGSW1.Cons.
- GGG Wave 2 ([DOI: 10.17026/dans-xm6-a262](https://doi.org/10.17026/dans-xm6-a262)), see [Gauthier, A. H. et al. \(2018\)](#) or visit the GGP website (<https://www.ggp-i.org/>) for methodological details. Working Version: GGS Wave 2 Version 1.3. – Consolidated identification: GGSW2.Cons.
- Goedicke, A., Lichtwardt, B. & Mayer, K. U. (2004). Dokumentationshandbuch Ostdeutsche Lebensverläufe im Transformationsprozess. LV-Ost Panel. Teil 2: Dokumentation verwendeter Materialien. Berlin: Max-Planck-Institut für Bildungsforschung.
- Graubard, B. & Korn, E. (1996). Survey inference for subpopulations. *American Journal of Epidemiology* 144 (1), 102-106. [doi:10.1093/oxfordjournals.aje.a008847](https://doi.org/10.1093/oxfordjournals.aje.a008847)
- Haas, W. (1991). Familie und Partnerbeziehungen in der Bundesrepublik (Ost). DJI-Familiensurvey. Methodenbericht zur Datenerhebung 1990/91. Leipzig: Leipziger Institut für Empirische Forschung.
- Haghighi, E.-F. (2014). Dynamic documents in Stata: MarkDoc, Ketchup, and Weaver. Verfügbar unter http://haghighi.com/talk/reproducible_report.php
- Hamjediers, M., Schmelzer, P., Geschke, A.-C. & SOEP Group (2020). SOEP-Core v35 - The couple history files BICOUPLM and BICOUPLY, and marital history files BIOMARSM and BIOMARSY SOEP Survey Papers 871: Series D. Berlin: DIW/SOEP.
- Hess, D. & Smid, M. (1991). Lebensverläufe und historischer Wandel in der ehemaligen DDR. Methodenbericht zur Hauptstudie. Bonn: Infas Institut für angewandte Sozialwissenschaft.
- Hillmert, S., Künster, R., Spengemann, P., Mayer, K. U. (2004). Projekt „Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland“. Dokumentation. Teil IX. Infas-Metodenberichte. Berlin: Max-Planck-Institut für Bildungsforschung.
- Infratest Sozialforschung (2011). SOEP 1984 – Methodenbericht zum Befragungsjahr 1984 (Welle 1) des Sozio-oekonomischen Panels. SOEP Survey Papers 1: Series B. Berlin: DIW/SOEP.
- Joye, D., Sapin, M. & Wolf, C. (2019). Weights in Comparative Surveys? A Call for Opening the Black Box. *Newsletter on Harmonization in the Social Sciences* 5 (2), 2-16.
- Kaminska, O. & Lynn, P. (2017). Survey-based cross-country comparisons where countries vary in sample design: issues and solutions. *Journal of Official Statistics* 33, 123-136. [doi:10.1515/jos-2017-0007](https://doi.org/10.1515/jos-2017-0007)
- Kish, L. (1999). Cumulating/Combining Population Surveys. *Survey Methodology* 25 (2), 129-138. In Kalton G. & Heeringa S. (Hrsg.) (2003). Wiley, Hoboken.

- Kish, L., Heeringa, S. & Kalton, G. (2003). *Leslie Kish: Selected Papers*. Wiley series in survey methodology. Hoboken: Wiley. [ISBN: 978-0-471-26661-7](#)
- Klevmarcken, A., Hesselius, P. & Swensson, B. (2005). The SHARE Sampling Procedures and Calibrated Designs Weights. In A. Börsch-Supan & H. Jürgens (Hrsg.), *The Survey of Health, Aging, and Retirement in Europe – Methodology* (S. 28-69). Mannheim: Mannheim Research Institute for the Economics of Ageing (MEA).
- Lavallée, P. & Beaumont, J.-F. (2015). Why We Should Put Some Weight on Weights. *Survey Insights: Methods from the Field. Weighting: Practical Issues and 'How to' Approach*. Invited article. [doi:10.13094/SMIF-2015-00001](#)
- Lavallée, P. & Beaumont, J.-F. (2016). Weighting: Principles and Practicalities. In C. Wolf, D. Joye; T. W. Smith & Yang-chih Fu (Hrsg.), *The SAGE Handbook of Survey Methodology* (S. 346-355). Los Angeles: SAGE. [doi:10.4135/9781473957893](#)
- Leven, I. (2009). *Generations and Gender Survey 2. Welle Methodenbericht der ersten Wiederholungsbefragung*. München: TNS Infratest Sozialforschung.
- Liebig, S., Goebel, J., Kroh, M., Schröder, C., Grabka, M., Schupp, J., Bartels, C., Fedorets, A., Franken, A., Jacobsen, J., Kara, S., Krause, P., Kröger, H., Metzinger, M., Richter, D., Schacht, D., Schmelzer, P., Schmitt, C., Schnitzlein, D., Siegers, R., Wenzig, K. & Zimmermann, S. (2019) *Socio-Economic Panel (SOEP), data for the years 1984-2018, version 35, SOEP, 2019*, [doi: 10.5684/soep-core.v35](#).
- Matthes, B., Lichtwardt, B. & Mayer, K. U. (2004). *Dokumentationshandbuch Ostdeutsche Lebensverläufe im Transformationsprozess. LV-Ost 71. Teil 2: Dokumentation verwendeter Materialien*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Mayer, K. U. (1995a). *Lebensverläufe und gesellschaftlicher Wandel: Die Zwischenkriegskohorte im Übergang zum Ruhestand (Lebensverlaufsstudie LV-West II T - Telefonische Befragung)*. GESIS Datenarchiv, Köln. ZA2647 Datenfile Version 1.0.0. [doi:10.4232/1.2647](#)
- Mayer, K. U. (1995b). *Lebensverläufe und historischer Wandel in Ostdeutschland (Lebensverlaufsstudie LV-DDR)*. GESIS Datenarchiv, Köln. ZA2644 Datenfile Version 1.0.0. [doi:10.4232/1.2644](#)
- Mayer, K. U. (1995c). *Lebensverläufe und gesellschaftlicher Wandel: Die Zwischenkriegskohorte im Übergang zum Ruhestand (Lebensverlaufsstudie LV-West II T - Telefonische Befragung)*. GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2647. [doi:10.4232/1.2647](#)
- Mayer, K. U. (1995d). *Lebensverläufe und historischer Wandel in Ostdeutschland (Lebensverlaufsstudie LV-DDR)*. GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2644. [doi:10.4232/1.2644](#)
- Mayer, K. U. (2004a). *Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost 71)*. GESIS Datenarchiv, Köln. ZA3926 Datenfile Version 1.0.0. [doi:10.4232/1.3926](#)
- Mayer, K. U. (2004b). *Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost Panel)*. GESIS Datenarchiv, Köln. ZA3925 Datenfile Version 1.0.0. [doi:10.4232/1.3925](#)
- Mayer, K. U. (2004c). *Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost 71)*. GESIS Datenarchiv, Köln. Codebuch zur Studie. ZA3926. [doi:10.4232/1.3926](#)
- Mayer, K. U. (2004d). *Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost Panel)*. GESIS Datenarchiv, Köln. Codebuch zur Studie. ZA3925. [doi:10.4232/1.3925](#)

- Mayer, K. U. (2014a). Frühe Karrieren und Familiengründung: Lebensverläufe der Geburtskohorte 1971 in Ost- und Westdeutschland (Lebensverlaufsstudie LV-Panel 71). GESIS Datenarchiv, Köln. ZA5099 Datenfile Version 1.0.0. [doi:10.4232/1.5099](https://doi.org/10.4232/1.5099)
- Mayer, K. U. (2014b). Frühe Karrieren und Familiengründung: Lebensverläufe der Geburtskohorte 1971 in Ost- und Westdeutschland (Lebensverlaufsstudie LV-Panel 71). GESIS Datenarchiv, Köln. Codebuch zur Studie. ZA5099. [doi:10.4232/1.5099](https://doi.org/10.4232/1.5099)
- Mayer, K. U. (2018a). Lebensverläufe und gesellschaftlicher Wandel: Lebensverläufe und Wohlfahrtsentwicklung (Lebensverlaufsstudie LV-West I). GESIS Datenarchiv, Köln. ZA2645 Datenfile Version 1.1.0. [doi:10.4232/1.13193](https://doi.org/10.4232/1.13193)
- Mayer, K. U. (2018b). Lebensverläufe und gesellschaftlicher Wandel: Die Zwischenkriegskohorte im Übergang zum Ruhestand (Lebensverlaufsstudie LV-West II A - Persönliche Befragung). GESIS Datenarchiv, Köln. ZA2646 Datenfile Version 1.1.0. [doi:10.4232/1.13194](https://doi.org/10.4232/1.13194)
- Mayer, K. U. (2018c). Lebensverläufe und gesellschaftlicher Wandel: Berufszugang in der Beschäftigungskrise (Lebensverlaufsstudie LV-West III). GESIS Datenarchiv, Köln. ZA2648 Datenfile Version 1.1.0. [doi:10.4232/1.13195](https://doi.org/10.4232/1.13195)
- Mayer, K. U. (2018d). Lebensverläufe und gesellschaftlicher Wandel: Lebensverläufe und Wohlfahrtsentwicklung (Lebensverlaufsstudie LV-West I). GESIS Datenarchiv, Köln. Deutsche Version des Fragebogens zur Studie. ZA2645. [doi:10.4232/1.13193](https://doi.org/10.4232/1.13193)
- Mayer, K. U. (2018e). Lebensverläufe und gesellschaftlicher Wandel: Die Zwischenkriegskohorte im Übergang zum Ruhestand (Lebensverlaufsstudie LV-West II A - Persönliche Befragung). GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2646. [doi:10.4232/1.13194](https://doi.org/10.4232/1.13194)
- Mayer, K. U. (2018f). Lebensverläufe und gesellschaftlicher Wandel: Berufszugang in der Beschäftigungskrise (Lebensverlaufsstudie LV-West III). GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2648. [doi:10.4232/1.13195](https://doi.org/10.4232/1.13195)
- Mayer, K. U. & Kleinhenz, G. (2004a). Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland (Lebensverlaufsstudie LV-West 64/71). GESIS Datenarchiv, Köln. ZA3927 Datenfile Version 1.0.0. [doi:10.4232/1.3927](https://doi.org/10.4232/1.3927)
- Mayer, K. U. & Kleinhenz, G. (2004b). Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland (Lebensverlaufsstudie LV-West 64/71). GESIS Datenarchiv, Köln. Codebuch zur Studie. ZA3927. [doi:10.4232/1.3927](https://doi.org/10.4232/1.3927)
- Mayer, K. U., Brückner, E., Blossfeld, H.-P., Früh, W., Geis, A., Kirschner, H.-P., Papastefanou, G., Tölke, A., Wagner, M., Wiedenbeck, M. (1989). Lebensverläufe und Wohlfahrtsentwicklung. Konzeption, Design und Methodik der Erhebung von Lebensläufen der Geburtsjahrgänge 1929-1931, 1939-1941, 1949-1951. Berlin: Max-Planck-Institut für Bildungsforschung.
- Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* (109), 558–606.
- Pairfam Group (2021). Anchor Codebook, Wave 1 (2008/2009), Release 12.0, GESIS Data Archive, Cologne. ZA5678 Data file Version 12.0.0, <https://doi.org/10.4232/pairfam.5678.12.0.0>.
- Pfeffermann, D. & Sverchkov, M. (2009). Chapter 39 - Inference under Informative Sampling. In C. Rao (Hrsg.), *Handbook of Statistics, Volume 29 of Handbook of Statistics* (S.455 – 487).
- Pohl, K. (1995). Design und Struktur des deutschen FFS. Heft 82a. Wiesbaden: Bundesinstitut für Bevölkerungsforschung (BiB).

- Ruckdeschel, K., Ette, A., Hullen, G. & Leven, I. (2006). Generations and Gender Survey. Dokumentation der ersten Welle der Hauptbefragung in Deutschland. Wiesbaden: Bundesinstitut für Bevölkerungsforschung beim Statistitischen Bundesamt.
- Schneekloth, U. (2000). Familie und Partnerbeziehungen in der Bundesrepublik Deutschland (Familiensurvey 2000). Methodenbericht. München: Infratest Burke.
- SHARE Waves 1, 2 and 3 (DOIs: [10.6103/SHARE.w1.600](https://doi.org/10.6103/SHARE.w1.600), [10.6103/SHARE.w2.600](https://doi.org/10.6103/SHARE.w2.600), [10.6103/SHARE.w3.600](https://doi.org/10.6103/SHARE.w3.600)), see Börsch-Supan et al. (2013) for methodological details.(1) The SHARE data collection has been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982) and [Horizon 2020](https://doi.org/10.6103/SHARE.w3.600) (SHARE-DEV3: GA N°676536, SERISS: GA N°654221) and by DG Employment, Social Affairs & Inclusion. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged(see www.share-project.org).
- SHARE-ERIC (2019). Sharelife Questionnaire Germany. Sharelife: 50+ in Europe main interview version 3.9.4b. Online document [accessed 26/09/2019, 13.30]
- Siegers, R., Belcheva, V. & Silbermann, T. (2020) SOEP-Core v35 – Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (SOEP) (1984 until 2018). SOEP Survey Papers 826: Series C. Berlin: DIW/SOEP
- Survey Research Center. (2016). Guidelines for Best Practice in Cross-Cultural Surveys. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. access via <http://www.ccsr.isr.umich.edu/> [accessed 06.02.2020, 11.30]
- Tillé, Y. & Matej, A. (2016). Basics of Sampling for Survey Research. In C. Wolf, D. Joye; T. W. Smith & Yang-chih Fu (Hrg.), *The SAGE Handbook of Survey Methodology* (S. 346-355). Los Angeles: SAGE. [doi:10.4135/9781473957893](https://doi.org/10.4135/9781473957893)
- UNESCO (2006). International Standard Classification of Education ISCED 1997 (May 2006 re-edition), UNESCO-Institute for Statistics. access via http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-1997-en_0.pdf [accessed 22.06.2020, 10.30]
- Vannette, D & Krosnick, J. (2018). *The Palgrave Handbook of Survey Research*. Palgrave Handbooks. Palgrave Macmillan, Cham. [doi:10.1007/978-3-319-54395-6](https://doi.org/10.1007/978-3-319-54395-6)
- West, B. T., Berglund, P. & Heeringa, S. G. (2008). A Closer Examination of Subpopulation Analysis of Complex-Sample Survey Data. *The Stata Journal* 8 (4), 520-531. [doi:10.1177/1536867X0800800404](https://doi.org/10.1177/1536867X0800800404)
- Winship, C. & Radbill, L. (1994). Sampling Weights and Regression Analysis. *Sociological Methods & Research*, 23(2), 230–257. <https://doi.org/10.1177/0049124194023002004>
- Wolf, C., Schneider, S. L., Behr, D., Joye, D. (2016). Chapter 33: Harmonizing Survey Questions Between Cultures and Over Time. In C. Wolf, D. Joye, T. W. Smith & Yang-chih Fu (Hrsg.), *The SAGE Handbook of Survey Methodology* (S. 502-524). L.A. u.a.: SAGE reference.
- Xia, K., Pedlow, S. & Davern, M. (2010). Dual-Frame Weights (Landline and Cell) for the 2009 Minnesota Health Access Survey. In *American Statistical Association - Proceedings of the Survey Research Methods Section* (S. 3912–3922).

Appendix

A. 1 Acquisition possibilities for the survey programs harmonized in HaSpaD

Table A 1 gives an overview of the acquisition possibilities for the studies used in the HaSpaD project. The link to the data distribution enables a direct access to the download of the data sources. It is important to obtain the required version (last column in the table) because otherwise a successful creation of the HaSpaD-target data set cannot be ensured.

Table A 1: Acquisition possibilities for the survey programs harmonized in HaSpaD

Study title	Research data center (RDC) / data distributor / data producer	Link to data provider / research data center	Link to data distributor / DOI	Required version
<i>pairfam Data Release 12.0.0</i>	RDC pairfam	Link to pairfam Homepage https://www.pairfam.de/en/	Link to GESIS Search https://doi.org/10.4232/pairfam.5678.12.0.0	Version 12.0.0
<i>ALLBUS – Cumulation 1980 – 2016</i>	RDC ALLBUS	Link to RDC ALLBUS https://www.gesis.org/en/allbus/allbus-home	Link to GESIS Search http://dx.doi.org/10.4232/1.13029	Version 1.0.0
<i>Change and Development of Forms of Family Life in West Germany (Survey of Families)</i>	RDC DJI (Deutsches Jugendinstitut, German Youth Institute)	Link to research database DJI https://surveys.dji.de/index.php?m=msg,0&fid=2	Link to GESIS Search http://dx.doi.org/10.4232/1.2245	Version 1.0.0
<i>Family and Partner Relations in Eastern Germany (Survey of Families)</i>	RDC DJI (Deutsches Jugendinstitut, German Youth Institute)	Link to research database DJI https://surveys.dji.de/index.php?m=msg,0&fid=2	Link to GESIS Search http://dx.doi.org/10.4232/1.13196	Version 1.1.0
<i>Change and Development of Ways of Family Life - 2nd Wave (Survey of Families)</i>	RDC DJI (Deutsches Jugendinstitut, German Youth Institute)	Link to research database DJI https://surveys.dji.de/index.php?m=msg,0&fid=2	Link to GESIS Search http://dx.doi.org/10.4232/1.13197	Version 1.1.0

Study title	Research data center (RDC) / data distributor / data producer	Link to data provider / research data center	Link to data distributor / DOI	Required version
<i>Change and Development of Families` Way of Life - 3rd Wave (Family Survey)</i>	RDC DJI (Deutsches Jugendinstitut, German Youth Institute)	Link to research database DJI https://surveys.dji.de/index.php?m=msg,0&fID=2	Link to GESIS Search http://dx.doi.org/10.4232/1.13198	Version 1.1.0
<i>Mannheim Divorce Study 1996</i>	Mannheimer Zentrum für Europäische Sozialforschung (MZES)	Link to Mannheim Centre for European Social Research (MZES) http://www.mzes.uni-mannheim.de/d7/de/projects/determinanten-der-ehescheidung	Link to GESIS Search http://dx.doi.org/10.4232/1.13056	Version 2.0.0
<i>German Fertility and Family Survey 1992</i>	Federal Institute for Population Research (Bundesinstitut für Bevölkerungsforschung (BiB))	Link to info-site of the Federal Institute for Population Research (Bundesinstituts für Bevölkerungsforschung) https://www.bib.bund.de/EN/Research/Surveys/FFS/family-and-fertility-survey.html	Link to GESIS Search http://dx.doi.org/10.4232/1.3400	Version 1.0.0
<i>Courses of Life and Social Change: Courses of Life and Welfare Development (Life History Study LV-West I)</i>	Max Planck Institute for Human Development	Link zur info-site of MPI for Human Development https://www.mpib-berlin.mpg.de/research/concluded-areas/center-for-sociology-and-the-study-of-the-life-course	Link to GESIS Search http://dx.doi.org/10.4232/1.13193	Version 1.1.0
<i>Courses of Life and Social Change: The Between-the-War Cohort in Transition to Retirement</i>	Max Planck Institute for Human Development	Link zur info-site of MPI for Human Development https://www.mpib-berlin.mpg.de/research/concluded-areas/center-for-sociology-and-the-study-of-the-life-course	Link to GESIS Search http://dx.doi.org/10.4232/1.13194	Version 1.1.0

Study title	Research data center (RDC) / data distributor / data producer	Link to data provider / research data center	Link to data distributor / DOI	Required version
<i>(Life History Study LV-West II A - Personal Interview)</i>		ch/concluded-areas/center-for-sociology-and-the-study-of-the-life-course		
<i>Courses of Life and Social Change: The Between-the-War Cohort in Transition to Retirement (Life History Study LV-West II T - Telephone Interview)</i>	Max Planck Institute for Human Development	Link zur info-site of MPI for Human Development	Link to GESIS Search http://dx.doi.org/10.4232/1.2647	Version 1.0.0
<i>Courses of Life and Social Change: Access to Occupation in Employment Crisis (Life History Study LV-West III)</i>	Max Planck Institute for Human Development	Link zur info-site of MPI for Human Development	Link to GESIS Search http://dx.doi.org/10.4232/1.13195	Version 1.1.0
<i>Courses of Life and Historical Change in East Germany (Life History Study LV DDR)</i>	Max Planck Institute for Human Development	Link zur info-site of MPI for Human Development	Link to GESIS Search http://dx.doi.org/10.4232/1.2644	Version 1.0.0
<i>East German Life Courses After Unification (Life History Study LV Ost 71)</i>	Max Planck Institute for Human Development	Link zur info-site of MPI for Human Development	Link to GESIS Search http://dx.doi.org/10.4232/1.3926	Version 1.0.0
		https://www.mpib-berlin.mpg.de/research/concluded-areas/center-for-sociology-and-the-study-of-the-life-course		
		https://www.mpib-berlin.mpg.de/research/concluded-areas/center-for-sociology-and-the-study-of-the-life-course		

Study title	Research data center (RDC) / data distributor / data producer	Link to data provider / research data center	Link to data distributor / DOI	Required version
		study-of-the-life-course		
<i>East German Life Courses After Unification (Life History Study LV-Ost Panel)</i>	Max Planck Institute for Human Development	Link zur info-site of MPI for Human Development	Link to GESIS Search http://dx.doi.org/10.4232/1.3925	Version 1.0.0
<i>Education, Training, and Occupation: Life Courses of the 1964 and 1971 Birth Cohorts in West Germany (Life History Study LV-West 64/71)</i>	Max Planck Institute for Human Development	Link zur info-site of MPI for Human Development	Link to GESIS Search http://dx.doi.org/10.4232/1.3927	Version 1.0.0
<i>Early Careers and Starting a Family: Life Courses of the 1971 Birth Cohorts in East and West Germany (Life History Study LV-Panel 71)</i>	Max Planck Institute for Human Development	Link zur info-site of MPI for Human Development	Link to GESIS Search http://dx.doi.org/10.4232/1.5099	Version 1.0.0
<i>Generations and Gender Survey (2005) (Subsample Germany)</i>	Federal Institute for Population Research (Bundesinstitut für Bevölkerungsforschung (BiB)) Generations and Gender Programme (GGP)	Link to contact of Generations and Gender Programme	Link to GGP-Data Access Site https://www.ggp-i.org/form/accounts/login/?next=/form/	GGG Wave 1 Version 4.3. - Consolidated

Study title	Research data center (RDC) / data distributor / data producer	Link to data provider / research data center	Link to data distributor / DOI	Required version
<i>Generations and Gender Survey (2008) (Subsample Germany)</i>	Federal Institute for Population Research (Bundesinstitut für Bevölkerungsforschung (BiB)) Generations and Gender Programme (GGP)	Link to contact of Generations and Gender Programme https://www.ggp-i.org/	Link to GGP-Data Access Site https://www.ggp-i.org/form/accounts/login/?next=/form/	GGG Wave 2 Version 1.3. - Consolidated
<i>SHARE Waves 1, 2 and 3</i>	SHARE Research Data Center	Link to SHARE Project http://www.share-project.org/home0.html	Link to SHARE RDC http://www.share-project.org/data-access/user-registration.html	Version 3-0-0
<i>Socio-Economic Panel (SOEP), data for years 1984-2018, version 35</i>	German Institute for Economic Research (DIW Berlin) RDC SOEP	Link to RDC SOEP https://www.diw.de/en/diw_02.c.222518.en/research_data_center_of_the_soep.html	Link to Data Access of SOEP https://www.diw.de/en/diw_02.c.222829.en/access_and_ordering.html	Version 35

A.2 The HaSpaD program code

For the sake of clearness, the survey programs ALLBUS-Cumulation and the German Fertility and Family Survey (FFS) are used in the following example. In the folder `3_harmonization\dofiles` all Stata-program files (do-files) are stored, which are required to create the target data set consisting of biography data and additional variables. Thus, no changes should be done by the user in this folder, since the Main-do-file would not be able to access to the required program files.

The names of the do-files match the subsequent system:

sid_Studyacronym_Suffix.do

The study ID (*sid*) corresponds to the identification number assigned to each survey program in the HaSpaD project. In this example, 102 belongs to the ALLBUS-Cumulation – 1980-2016 (study acronym: *allbus*), 106 to the German Fertility and Family Survey (study acronym: *ffs*). Besides, the suffix describes the function of the do-file. Do-files with the ending *_bio* is used to process the couple biographies; the ending **_sex* labels the harmonized variable *sex* from the original data sources, **_eastwest* the survey area.

In the folder *temp* all temporary Stata-files are saved, which are necessary to create the final target data set.

Moreover, help-do-files needed for the processing are stored in the folder `3_harmonization/resources`. The do-file *labelfile.do* assigns variable labels to all biography-variables after running the single processing files of **_bio.do*. Packages for the processing are loaded through the do-file *packages.do*.

The final target data set *target.dta* with the biography-data and additional variables from the survey programs and variables as requested from the user is saved in the initially empty folder *4_target* after the successful execution of the main-file.

The main-file designs both paths to the execution of the processing and harmonization do-files, the creation of the temporary folders and HTML-files as well as the save location of the source data sets (data sources downloaded and unpacked by the user) via the macro commands *global drive* and *global source*.

During the execution of the main file, the two folders *harmonization_targets* and *preparation_couplebio* are stored in *1_documentation*, which contain the respective documentation of the processing of the couple biographies (*preparation_couplebio*) and the harmonization of the additional variables (*harmonization_targets*).

Creation of the cumulated couple biographies

After the automatic creation of both folders *harmonization_target* and *preparation_target* the main file harmonizes and cumulates the couple biographies of the selected survey programs in `# RUN PREFILES FROM COUPLEBIO` (see *main.do*). Hereby, the following flow chart helps to illustrate (see *Figure A 1*, from top to bottom):

First, the main file calls the processing do-files *102_allbus_bio.do* and *106_ffs_bio.do*. There, the couple biography data sets *102_allbus_bio.dta* and *106_ffs_bio.dta* are created based on the source data sets (see *Figure A 1*, the very top) and saved in the folder *temp* (see *Figure A 2*). Furthermore, the HTML-documentations *coding_allbus.html* and *coding_ffs.html* are created in *preparation_couplebio*.

In a next step, both target data sets from the folder `temp` are appended (append), stored in the folder `4_target` as `target.dta` and eventually deleted from the folder `temp`.

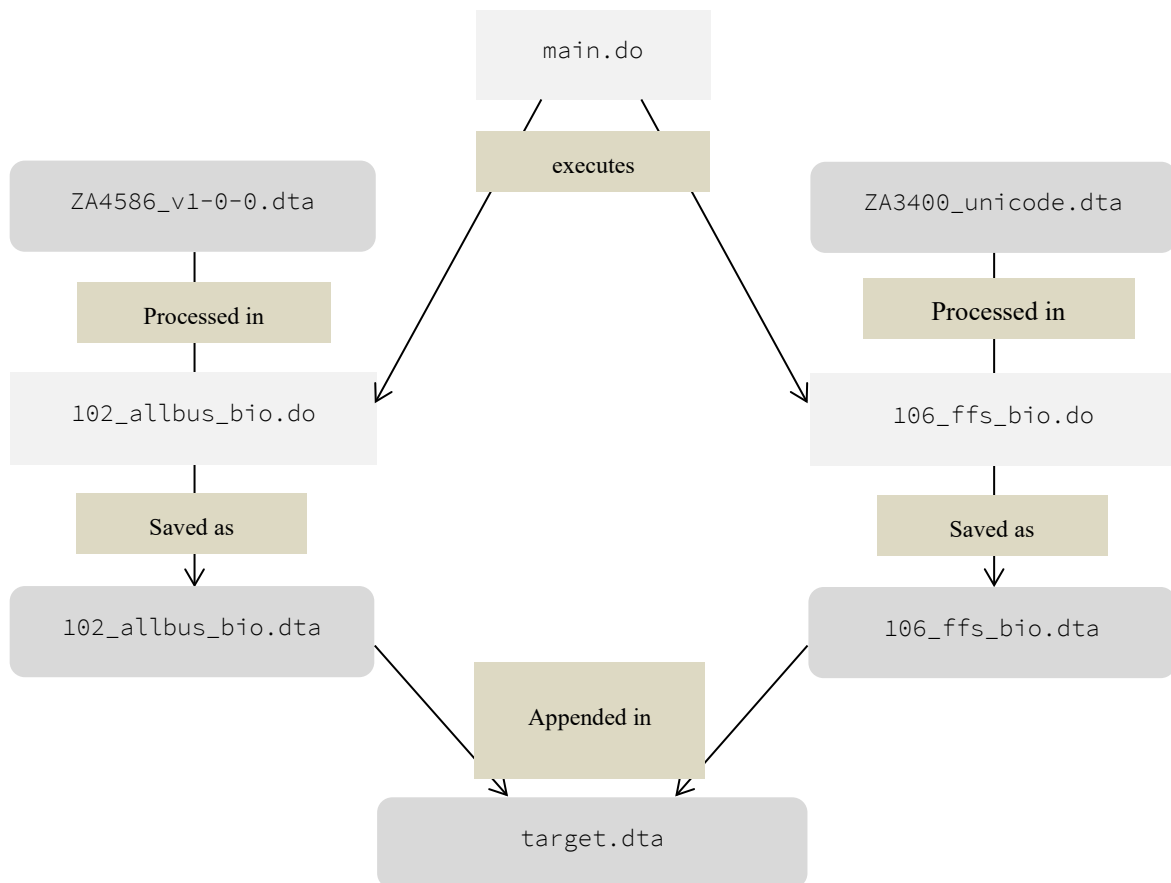


Figure A1: Exemplary process of the creation of cumulative partnership biographies in Stata

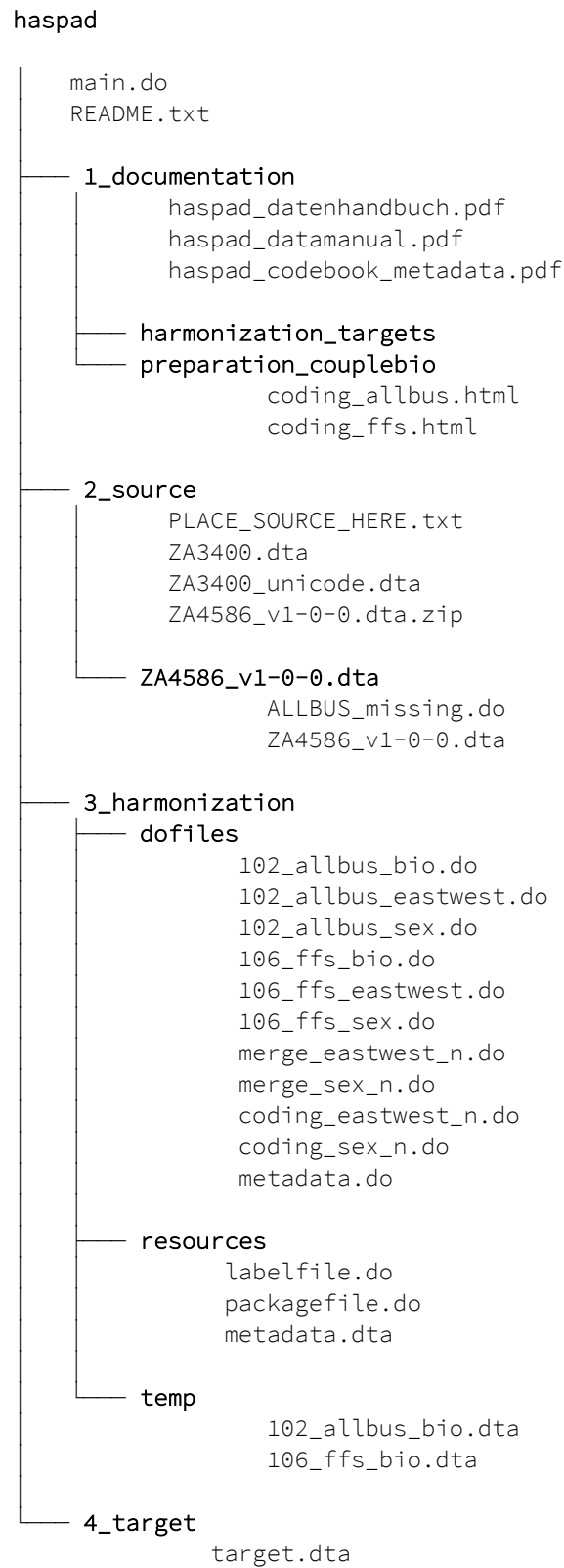
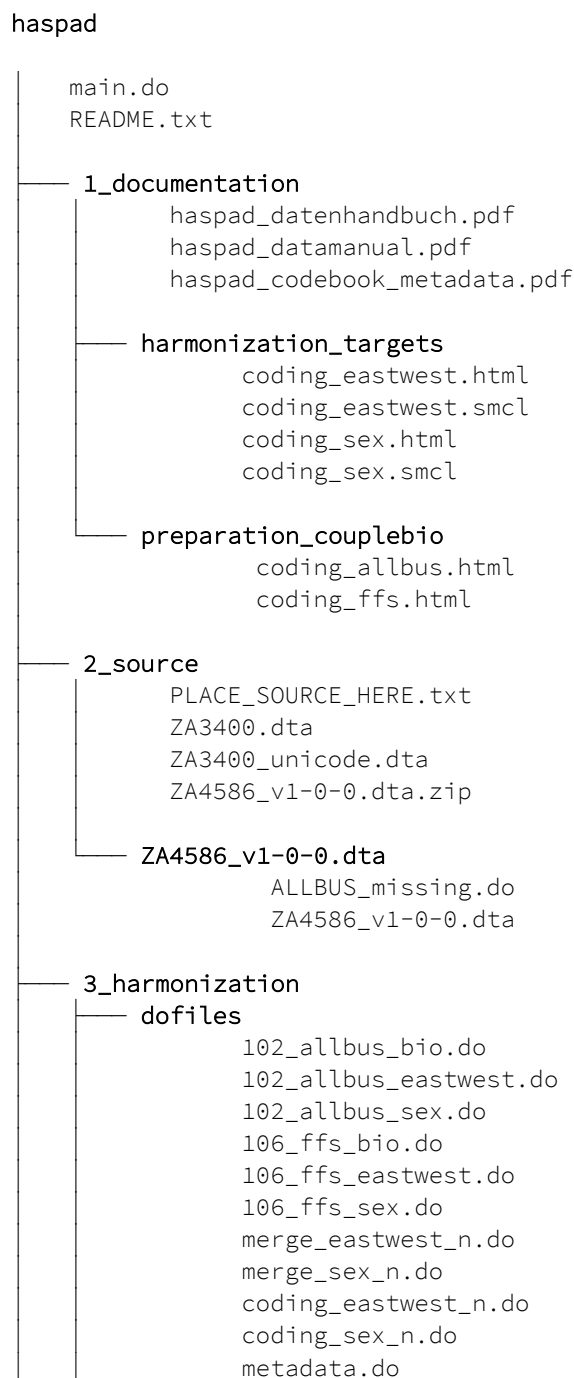


Figure A2: Folder structure after processing the biography data set

Creation of the harmonized additional variables and linkage with target.dta

In # RUN AND MERGE PREPFILES FROM HARMONIZEDVAR in the main-file the additional variables selected by the user are created by running the do-files 102_allbus_sex.do, 102_allbus_eastwest.do, 106_ffs_sex.do and 106_ffs_eastwest.do. After the execution of the harmonization files all identically named dta.-files are saved in temp and at the same time the harmonization documentation is saved in the folder 1_documentation under the name harmonization_targets (see the change of the folder structure in *Figure A 3*).



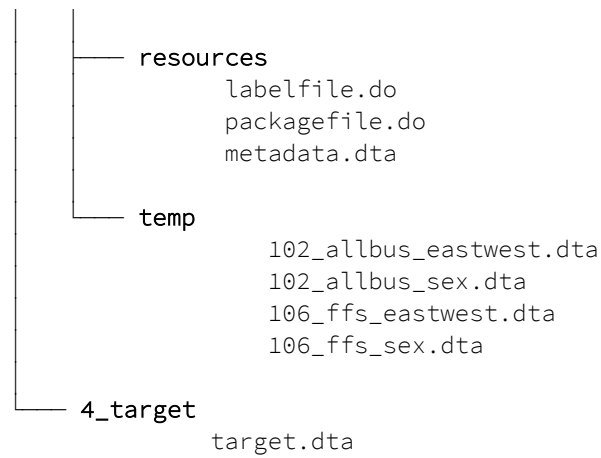


Figure A3: Folder structure after processing the additional variables

Once the data is created in the folder `temp`, they are merged with the couple biographies in `target.dta` (merge) and afterwards deleted from the folder `temp`. The final data set `target.dta` is stored in `4_target` and is now available for the end user.

A.3 Documentations of the source data

The following *Table A 2* gives an overview of the documentation of the original data sources. The page reference in column 4 refers to the section of the questionnaire in which the anchor person could report partnerships.

Table A2: Documentation of the source data

Short study title	Citation of the documentation of source variables	Name of the pdf-document / Link for automatic download
pairfam	Brüderl, Josef, Madison Garrett, Kristin Hajek, Michel Herzig, Rüdiger Lenke, Renate Lorenz, Katharina Lutz, Trang Phan, Philipp Schütze, and Nina Schumann (2021): pairfam Data Manual, Release 12.0. LMU Munich: Technical Report. GESIS Data Archive, Cologne. ZA5678 Data File Version 12.0.0, https://doi.org/10.4232/pairfam.5678.12.0.0	Data_Manual_pairfam_12.0.pdf https://dbk.gesis.org/dbksearch/download.asp?id=53708
pairfam	pairfam Group (2021). Anchor Codebook, Wave 1 (2008/2009), Release 12.0, GESIS Data Archive, Cologne. ZA5678 Data file Version 12.0.0, https://doi.org/10.4232/pairfam.5678.12.0.0	Codebuch Anker de, pairfam Welle 1 2008-09.pdf https://dbk.gesis.org/dbksearch/sdesc2.asp?no=5678&db=e&doi=10.4232/pairfam.5678.12.0.0
ALLBUS - Cumulation	Baumann, Horst; Schulz, Sonja (2018): ALLBUS - Kumulation 1980-2016. Variable Report. GESIS Datenarchiv, Köln. Studie ZA4586, doi:10.4232/1.13029	ZA4586_cdb.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=64351
Family Survey I. Wave West 1988	Deutsches Jugendinstitut (DJI), München (1992): Wandel und Entwicklung familialer Lebensformen in Westdeutschland (Familiensurvey). GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2245, doi:10.4232/1.2245	ZA2245_fb.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=2073
Family Survey I. Wave East 1990	Deutsches Jugendinstitut (DJI), München (2018): Familie und Partnerbeziehungen in Ostdeutschland (Familiensurvey). GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2392, doi:10.4232/1.13196	ZA2392_fb.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=2225
Family Survey II. Wave 1994	Deutsches Jugendinstitut (DJI), München (2018): Wandel und Entwicklung familialer Lebensformen - 2. Welle (Familiensurvey). GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2860, doi:10.4232/1.13197	ZA2860_fb.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=3015
Family Survey III. Wave 2000	Deutsches Jugendinstitut (DJI), München (2000): Wandel und Entwicklung familialer Lebensformen - 3. Welle (Familiensurvey), GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA3920, doi:10.4232/1.13198	ZA3920_fb.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=17019

Short study title	Citation of the documentation of source variables	Name of the pdf-document / Link for automatic download
Mannheim Divorce Study	Esser, Hartmut; Gostomski, Christian Babka von; Hartmann, Josef (2018): Mannheimer Scheidungsstudie 1996. GESIS Datenarchiv, Köln. Codebuch zur Studie. ZA3188, doi:10.4232/1.13056	ZA3188_cod.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=38434
German Fertility and Family Survey	Bundesinstitut für Bevölkerungsforschung, Wiesbaden (2002): Deutscher Fertility and Family Survey 1992. GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA3400, doi:10.4232/1.3400	ZA3400_fb.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=3808
GLHS-West I	Mayer, Karl U. (2018): Lebensverläufe und gesellschaftlicher Wandel: Lebensverläufe und Wohlfahrtsentwicklung (Lebensverlaufsstudie LV-West I). GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2645, doi:10.4232/1.13193	ZA2645_fb.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=47640
GLHS -West II A - Personal Interview	Mayer, Karl U. (2018): Lebensverläufe und gesellschaftlicher Wandel: Die Zwischenkriegskohorte im Übergang zum Ruhestand (Lebensverlaufsstudie LV-West II A - Persönliche Befragung). GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2646, doi:10.4232/1.13194	ZA2646-47_fb.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=47634
GLHS -West II T - Telephone Interview	Mayer, Karl U. (1995): Lebensverläufe und gesellschaftlicher Wandel: Die Zwischenkriegskohorte im Übergang zum Ruhestand (Lebensverlaufsstudie LV-West II T - Telefonische Befragung). GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2647, doi:10.4232/1.2647	ZA2646-47_fb.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=65007
GLHS-West III	Mayer, Karl U. (2018): Lebensverläufe und gesellschaftlicher Wandel: Berufszugang in der Beschäftigungskrise (Lebensverlaufsstudie LV-West III). GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2648, doi:10.4232/1.13195	ZA2648_fb.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=47641
GLHS-GDR	Mayer, Karl U. (1995): Lebensverläufe und historischer Wandel in Ostdeutschland (Lebensverlaufsstudie LV-DDR). GESIS Datenarchiv, Köln. Fragebogen zur Studie. ZA2644, doi:10.4232/1.2644	ZA2644_fb.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=5634
GLHS -East 71	Mayer, Karl U. (2004): Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost 71). GESIS Datenarchiv, Köln. Codebuch zur Studie. ZA3926, doi:10.4232/1.3926	ZA3926_cod.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=46256
GLHS -East Panel, birth cohorts 1939-	Mayer, Karl U. (2004): Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost Panel). GESIS	ZA3925_cod.pdf

Short study title	Citation of the documentation of source variables	Name of the pdf-document / Link for automatic download
41, 1951-52, 1959-61	Datenarchiv, Köln. Codebuch zur Studie. ZA3925, doi:10.4232/1.3925	https://dbk.gesis.org/dbksearch/download.asp?db=D&id=46258
GLHS-East Panel, birth cohort 1929-31	Mayer, Karl U. (2004): Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost Panel). GESIS Datenarchiv, Köln. Codebuch zur Studie. ZA3925, doi:10.4232/1.3925	https://dbk.gesis.org/dbksearch/download.asp?db=D&id=46258 ZA3925_cod.pdf
GLHS -West 64/71	Mayer, Karl U.; Kleinhenz, Gerhard (2004): Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland (Lebensverlaufsstudie LV-West 64/71). GESIS Datenarchiv, Köln. Codebuch zur Studie. ZA3927, doi:10.4232/1.3927	https://dbk.gesis.org/dbksearch/download.asp?db=D&id=46255 ZA3927_cod.pdf
GLHS -Panel 71	Mayer, Karl U. (2014): Frühe Karrieren und Familiengründung: Lebensverläufe der Geburtskohorte 1971 in Ost- und Westdeutschland (Lebensverlaufsstudie LV-Panel 71). GESIS Datenarchiv, Köln. Codebuch zur Studie. ZA5099, doi:10.4232/1.5099	https://dbk.gesis.org/dbksearch/download.asp?db=D&id=53821 ZA5099_cod.pdf
Generations and Gender Survey (2005)	Ruckdeschel, Kerstin; Ette, Andreas; Hullen, Gert; Leven, Ingo (2006): Generations and Gender Survey. Dokumentation der ersten Welle der Hauptbefragung in Deutschland. Bundeinstitut für Bevölkerungsforschung beim Statistitischen Bundesamt, Wiesbaden.	GermanyOreWave1.pdf https://www.ggp-i.org/sites/default/files/questionnaires/GermanyOreWave1.pdf
Generations and Gender Survey (2008)	Leven, Ingo (2009): Generations and Gender Survey 2. Welle. Methodenbericht der erstenWiederholungsbefragung. TNS Infratest Sozialforschung, München.	Methodenbericht_12_11_09.pdf https://www.ggp-i.org/sites/default/files/questionnaires/Methodenbericht_12_11_09.pdf
SHARE	SHARE-ERIC: Sharelife Questionnaire Germany. Sharelife: 50+ in Europe main interview version 3.9.4b. Online-Dokument [26.09.2019, 13.30 Uhr]	DE_Sharelife_Questionnaire - http://www.share-project.org/t3/share/fileadmin/pdf_sharelife/DE_Sharelife_Questionnaire.pdf
Socio-Economic Panel (biomarsy)	Hamjediers, M., Schmelzer, P., Geschke, A.-C. & SOEP Group (2020). SOEP-Core v35 - The couple history files BIOCUPLM and BIOCUPLY, and marital history files BIOMARSM and BIOMARSY SOEP Survey Papers 871: Series D. Berlin: DIW/SOEP.	diw_ssp0871.pdf https://www.diw.de/documents/publikationen/73/diw_01.c.793991.de/diw_ssp0871.pdf

A. 4 Degree of imputation of the biography variables

The following *Table A 3* provides an overview of the respective imputation levels after the study.

Table A 3: Degree of imputation of the biography variables

Short study title	Beginning relationship (relbeg)	End relationship (relend)	Beginning cohabitation (cohbeg)	End cohabitation (cohend)	Beginning marriage (marbeg)	End marriage (marend)	Date of partner's death (dop)
pairfam	0 No imputation	0 No imputation	0 No imputation	0 No imputation	0 No imputation	0 No imputation	-10 Missing by study design
	1 Only year information	1 Only year information	1 Only year information	1 Only year information	1 Only year information	1 Only year information	
	2 Only season information	2 Only season information	2 Only season information	2 Only season information	2 Only season information	2 Only season information	
ALLBUS - Cumulation	-10 Missing by study design	-10 Missing by study design	-10 Missing by study design	-10 Missing by study design	0 No imputation	0 No imputation	0 No imputation
					1 Only year information	1 Only year information	1 Only year information
Family Survey I. Wave West 1988	1 Only year information	1 Only year information	1 Only year information	-10 Missing by study design	0 No imputation 1 Only year information	1 Only year information	1 Only year information
Family Survey I. Wave East 1990	1 Only year information	1 Only year information	1 Only year information	1 Only year information	0 No imputation 1 Only year information	1 Only year information	1 Only year information
Family Survey II. Wave 1994	1 Only year information	1 Only year information	1 Only year information	-10 Missing by study design	0 No imputation 1 Only year information	1 Only year information	1 Only year information

Short study title	Beginning relationship (relbeg)	End relationship (relend)	Beginning cohabitation (cohbeg)	End cohabitation (cohend)	Beginning marriage (marbeg)	End marriage (marend)	Date of partner's death (dop)
Family Survey III. Wave 2000	1 Only year information	1 Only year information	1 Only year information	-10 Missing by study design	0 No imputation	1 Only year information	1 Only year information
Mannheim Divorce Study	0 No imputation	0 No imputation	0 No imputation	0 No imputation	0 No imputation	0 No imputation	0 No imputation
	1 Only year information	1 Only year information	1 Only year information	1 Only year information	1 Only year information	1 Only year information	1 Only year information
	2 Only season information	2 Only season information	2 Only season information	2 Only season information	2 Only season information	2 Only season information	2 Only season information
German Fertility and Family Survey	-10 Missing by study design		0 No imputation	0 No imputation	0 No imputation		0 No imputation
			1 Only year information	1 Only year information	1 Only year information		1 Only year information
GLHS-West I	-10 Missing by study design	-10 Missing by study design	0 No imputation	0 No imputation	0 No imputation	0 No imputation	0 No imputation
				1 Only year information	1 Only year information	1 Only year information	1 Only year information
GLHS-West II A – Personal Interview	-10 Missing by study design	1 Only year information	0 No imputation	1 Only year information	0 No imputation	1 Only year information	1 Only year information
					1 Only year information		
GLHS-West II T - Telephone Interview	-10 Missing by study design	1 Only year information	0 No imputation	-10 Missing by study design	0 No imputation	1 Only year information	1 Only year information
			1 Only year information		1 Only year information		
GLHS-West III	-10 Missing by study design	1 Only year information	0 No imputation	0 No imputation	0 No imputation	1 Only year information	1 Only year information
			1 Only year information	1 Only year information	1 Only year information		
GLHS-GDR	0 No imputation	1 Only year information	0 No imputation	-10 Missing by study design	0 No imputation	1 Only year information	1 Only year information

Short study title	Beginning relationship (relbeg)	End relationship (relend)	Beginning cohabitation (cohbeg)	End cohabitation (cohend)	Beginning marriage (marbeg)	End marriage (marend)	Date of partner's death (dop)
	1 Only year information		1 Only year information		1 Only year information		
GLHS-East 71	0 No imputation	0 No imputation	0 No imputation	-10 Missing by study design	0 No imputation	0 No imputation	
	1 Only year information	1 Only year information	1 Only year information				
GLHS-East Panel, birth cohorts 1939-41, 1951-52, 1959-61	-10 Missing by study design	-10 Missing by study design	0 No imputation	0 No imputation	0 No imputation	0 No imputation	
			1 Only year information	1 Only year information			
GLHS-Ost Panel, birth cohort 1929-31	-10 Missing by study design	-10 Missing by study design	-10 Missing by study design	-10 Missing by study design	0 No imputation		
GLHS-West 64/71	0 No imputation	0 No imputation	0 No imputation	-10 Missing by study design	0 No imputation	0 No imputation	0 No imputation
	1 Only year information	1 Only year information	1 Only year information		1 Only year information	1 Only year information	
	2 Only season information	2 Only season information	2 Only season information		2 Only season information	2 Only season information	
GLHS-Panel 71	0 No imputation	0 No imputation	0 No imputation	-10 Missing by study design	0 No imputation	-10 Missing by study design	0 No imputation
	1 Only year information	1 Only year information	1 Only year information		1 Only year information		
	2 Only season information	2 Only season information	2 Only season information				
Generations and Gender Survey (2005)	0 No imputation	0 No imputation	0 No imputation	-10 Missing by study design	0 No imputation	0 No imputation	
	1 Only year information	1 Only year information	1 Only year information		1 Only year information	1 Only year information	

Short study title	Beginning relationship (relbeg)	End relationship (relend)	Beginning cohabitation (cohbeg)	End cohabitation (cohend)	Beginning marriage (marbeg)	End marriage (marend)	Date of partner's death (dop)
Generations and Gender Survey (2008)	0 No imputation 1 Only year information	0 No imputation 1 Only year information	0 No imputation 1 Only year information	-10 Missing by study design	0 No imputation	0 No imputation	0 No imputation
SHARE	1 Only year information	1 Only year information	1 Only year information	1 Only year information	1 Only year information	1 Only year information	1 Only year information
Socio-Economic Panel (biomarsy)	-10 Missing by study design	1 Only year information	-10 Missing by study design	-10 Missing by study design	1 Only year information	1 Only year information	1 Only year information

A.5 Constructing the original variables of the biography data

In the following *Table A 4* the construction of the original variables of the biographical data set is described in more detail.

Table A 4: Constructing the original variables of the biography data

Short study title	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
pairfam	“Century months“	“Century months“	“Century months“	“Century months“	“Century months“	“Century months“	“Century months“
ALLBUS - Cumulation	“-10 Missing by study design“	“-10 Missing by study design“	“-10 Missing by study design“	“-10 Missing by study design“	“Month“ “+“ “Year“	“Month“ “+“ “Year“	““Month“ “+“ “Year“
Family Survey I. Wave West 1988	“Year“	Surveyed in two separate variables "Year" for unmarried partnerships "Year" for married partnerships	“Year“	"-10 Missing by study design"	“Month“ “+“ “Year“	“Year“	“Year“ for married partnerships "-10 Missing by study design" for unmarried partnerships
Family Survey I. Wave East 1990	“Jahr“	Surveyed in two separate variables “Year“ for unmarried partnerships “Year“ for married partnerships	“Year“	“Year“ for married partnerships "-10 Missing by study design" for unmarried partnerships	“Month“ “+“ “Year“	“Year“	“Year“ for married partnerships "-10 Missing by study design" for unmarried partnerships
Family Survey II. Wave 1994	“Jahr“	Surveyed in two separate variables	“Year“	“-10 Missing by study design"	“Month“ “+“ “Year“	“Year“	“Year“ for married partnerships "-10 Missing by study design" for

Short study title	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
		“Year“ for unmarried partnerships					unmarried partnerships
		“Year“ for married partnerships					
Family Survey III. Wave 2000	“Jahr“	Surveyed in two separate variables	“Year“	“-10 Missing by study design“	“Month“ “+“ “Year“	“Year“	“Year“ for married partnerships
		“Year“ for unmarried partnerships					“-10 Missing by study design“ for unmarried partnerships
		“Year“ for married partnerships					
Mannheim Divorce Study	“Month/Year“	“Month/Year“	“Month/Year“	“Month/Year“	“Month/Year“	“Month/Year“	“Month/Year“
German Fertility and Family Survey	“-10 Missing by study design“	“-10 Missing by study design“	“Month“ “+“ “Year“	“Month“ “+“ “Year“	“Month“ “+“ “Year“	“-10 Missing by study design“	“Month“ “+“ “Year“
							“-10 Missing by study design“ for partnership with reason for separation that is not the death of the partner
GLHS-West I	“-10 Missing by study design“	“-10 Missing by study design“	“Month“ “+“ “Year“ for current unmarried partnership	“Duration of separation before divorce“ + “+“ + “divorce month“	“Month“ “+“ “Year“ for past marriages and current marriage	“Month“ “+“ “Year“ for past marriages and current marriage, if divorce	“Month“ “+“ “Year“ for past marriages and current

Short study title	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
			"-10 Missing by study design" for past marriages and current marriage	"+" "divorce year" (only for divorcees in current marriage) "Month" "+" "Year" (only for married but separated partnerships in current marriage) "-10 Missing by study design" für past marriages "-10 Missing by study design" for married, widowed, single persons in current marriage "-10 Missing by study design" for current unmarried partnership	"-10 Missing by study design" for current unmarried partnership	is the reason for separation "-10 Missing by study design" for current unmarried partnership or if partner died	marriage, if partner died "-10 Missing by study design" for current unmarried partnership
GLHS-West II A - Personal Interview	"-10 Missing by study design"	"Year"	"Month" "+" "Year" for currently unmarried partnerships "-10 Missing by study design" for	"Duration of separation before divorce" "+" "Divorce year" for partnerships with divorce	"Month" "+" "Year"	"Year"	"Year"

Short study title	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
GLHS-West II T - Telephone Interview	"-10 Missing by study design"	"Year"	current married partnerships "Month" "+" "Year" for currently unmarried partnerships"-10 Missing by study design" for marriages	"Duration of separation before divorce" for rest "-10 Missing by study design"	"Month" "+" "Year"	"Year"	"Year"
GLHS-West III	"-10 Missing by study design"	"Year" "-10 Missing by study design" for currently divorced marriages and retrospective marriages	"Month" "+" "Year" for current unmarried partnerships and first marriage of the respondent "-10 Missing by study design" for rest	"Month" "+" "Year" for currently separated marriages "-10 Missing by study design" for currently unmarried partnerships and retrospective partnerships	"Month" "+" "Year"	"Year"	"Year"
GLHS-GDR	"Weeks known before cohabitation" "+" "Months known before cohabitation" "+" + "Years known before cohabitation" "+" "Month of	"Year"	"Month" "+" "Year"	"-10 Missing by study design"	"Month" "+" "Year"	"Year"	"Year"

Short study title	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
GLHS-East 71	cohabitation" "+" "Year of cohabitation" "Month" "+" "Year"	"Month" "+" "Year" "-10 Missing by study design" for ongoing partnerships	"Month" "+" "Year"	"-10 Missing by study design"	"Month" "+" "Year"	"Month" "+" "Year" "-10 Missing by study design" for unmarried partnerships	"Month" "+" "Year" "-10 Missing by study design" for partnerships with reason for separation that is not the death of the partner
GLHS-East Panel, birth cohorts 1939-41, 1951-52, 1959-61	"-10 Missing by study design"	"-10 Missing by study design"	"Month" "+" "Year" "-10 Missing by study design" for married couples without a partner who have no other partner in wave 2 AND for couples who have the same partner in wave 2 as in wave 1	"Month" "+" "Year" "-10 Missing by study design" for married persons, without a partner, who do not have another partner in 2nd wave	"Month" "+" "Year" "-10 Missing by study design" for persons who are married and live with the same partner (same as from wave 1) or are divorced/widowed and have no other partner	"Month" "+" "Year" "-10 Missing by study design" for married persons, without a partner, who do not have another partner in 2nd wave	"Month" "+" "Year" "-10 Missing by study design" for partnerships with reason for separation that is not the death of the partner
GLHS-East Panel, birth cohort 1929-31	"-10 Missing by study design"	"-10 Missing by study design"	"-10 Missing by study design"	"Month" "+" "Year" "-10 Missing by study design" for	"Month" "+" "Year"	"Month" "+" "Year" "-10 Missing by study design" for	"Month" "+" "Year" "-10 Missing by study design" for

Short study title	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
				divorced partnerships and partnerships with reason for separation "death of partner"		partnerships with separation reason "death of partner"	partnerships with reason for separation that is not the death of the partner
GLHS-West 64/71	"Month" "+" "Year"	"Month" "+" "Year"	"Month" "+" "Year"	"-10 Missing by study design"	"Month" "+" "Year"	"Month" "+" "Year"	"Monat" "+" "Year"
		"-10 Missing by study design" for divorced partnerships and partnerships with reason for separation "death of partner"					"-10 Missing by study design" for partnerships with reason for separation that is not the death of the partner
GLHS-Panel 71	"Month" "+" "Year"	"Month" "+" "Year"	"Month" "+" "Year"	"-10 Missing by study design"	"Month" "+" "Year"	"-10 Missing by study design"	"Month" "+" "Year"
		"-10 Missing by study design" for partnerships with separation reason "death of partner"					"-10 Missing by study design" for partnership with reason for separation that is not the death of the partner
Generations and Gender	"Month" "+" "Year"	"Month" "+" "Year"	"Month" "+" "Year"	"-10 Missing by study design"	"Month" "+" "Year"	"Month" "+" "Year"	"Month" "+" "Year"

Short study title	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
Survey (2005)	"-10 Missing by study design" for retrospective partnerships and current cohabiting partnerships	"-10 Missing by study design" when partner died					"-10 Missing by study design" if partnership ended or in case of currently cohabiting partnerships
Generations and Gender Survey (2008)	"Month" "+" "Year" "-10 Missing by study design" for retrospective partnerships and current cohabiting partnerships	"Month" "+" "Year" "-10 Missing by study design" when partner died	"Month" "+" "Year" "-10 Missing by study design" for same partnerships as wave 1	"-10 Missing by study design"	"Month" "+" "Year" "-10 Missing by study design" for same partnerships as wave 1	"Month" "+" "Year" "-10 Missing by study design" in current married partnerships	"Month" "+" "Year" "-10 Missing by study design" if partnership has ended or in case of currently cohabiting partnerships
SHARE	"Year"	"Year" "-10 Missing by study design" for cohabiting partnerships and marriages	"Year" "-10 Missing by study design" for non-cohabiting partnerships	"Year" "-10 Missing by study design" for non-cohabiting partnerships	"Year" "-10 Missing by study design" for non-cohabiting unmarried partnerships and cohabiting unmarried partnerships	"Year" "-10 Missing by study design" for non-cohabiting unmarried partnerships and cohabiting unmarried partnerships	"Year" "-10 Missing by study design" for non-cohabiting partnerships

Short study title	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
Socio-Economic Panel (biomarsy)	"-10 Missing by study design"	"Year "+" censor "original censor code"	"-10 Missing by study design"	"-10 Missing by study design"	"Year "+" censor "original censor code"	"Year "+" censor "original censor code"	"Year "+" censor "original censor code"
		"-10 Missing by study design"			"-10 Missing by study design"	"-10 Missing by study design"	"-10 Missing by study design"

In the cell: possible values of the ori-variable

A.6 Values of the variable problem

In the following *Table A 5*, the values of the variable `problem` are discussed in more detail.

Table A 5: Values of the variable `problem`

Value	Value label	Survey programs concerned	Explanation
0	No data inconsistency	all	No inconsistencies or noticeable problems were found in the original data.
1	Data inconsistency	all	<p>Following 20 checks were conducted:</p> <ol style="list-style-type: none"> 1. <code>relbeg > relend</code> 2. <code>cohbeg > cohend</code> 3. <code>marbeg > marend</code> 4. <code>relbeg > marbeg</code> 5. <code>relend = -99 & dop > 0</code> 6. <code>cohend = -99 & dop > 0</code> 7. <code>marend = -99 & dop > 0</code> 8. <code>relend > dop & dop > 0</code> 9. <code>cohend > dop & dop > 0</code> 10. <code>marend > dop & dop > 0</code> 11. <code>relbeg > cohend</code> 12. <code>relbeg > marend</code> 13. <code>cohbeg > marend</code> 14. <code>relend = -99 & marend > 0</code> 15. <code>cohbeg = -3 & cohend != -3</code> 16. <code>marbeg = -3 & marend != -3</code> 17. <code>relbeg = -3</code> 18. <code>marbeg > relend</code> 19. <code>cohbeg > relend</code> 20. <code>marbeg > cohend</code>

Value	Value label	Survey programs concerned	Explanation
2	Problems due to panel linkage (GLHS, GGS)	GLHS, GGS	<p>As soon as one or more of the 20 conditions in the original data apply, the value <code>problem = "1 Data inconsistency"</code> is assigned.</p> <p>The value <code>problem = "2 Problems due to panel linkage (GLHS, GGS)"</code> is assigned for two different reasons.</p> <p>In the follow-up studies of the Life History Studies (Mayer 2004b, 2014a) and Generations and Gender Surveys (Wave 2, 2008), respondents indicated partnerships that they had already reported in the first wave and that lasted until the interview date of wave 1. Section 5.7 explains the plausibility checks that were used to identify the same partnership reported in wave 1 and again in wave 2. After identifying the same partnerships reported in both Wave 1 and 2, partnerships that began before the interview date in wave 1 and continued until the interview date in wave 2 remain. These are marked with <code>problem = "2 Problem due to panel linkage (GLHS, GGS)"</code>, as these are very likely the same partnerships that were reported in the first wave but cannot be linked due to the plausibility check. These can be identified by <code>problem</code> and removed from the data set.</p> <p>Furthermore, there are partnerships, reported in the follow-up study, for which it is not clear whether they were already reported in wave 1 due to missing information on the start of the partnership or the filtering in the questionnaire.</p>
3	Higher order partnerships to same partner (pairfam, FFS)	pairfam, FFS, SOEP	<p>In the Socio-Economic Panel, Fertility and Family Survey (FFS) and pairfam, there is the special case of independent partnerships with the same partner. In the original data, these partnerships were identified by duplicates of the partner ID within the partnership biography of a respondent. Furthermore, in the Socio-Economic Panel, the original data set is constructed in such a way that a single-spell is inserted between two partnerships on final completion and a partnership with the same partner can then be restarted.</p>
4	Data inconsistency + problems due to panel linkage (GLHS, GGS)	GLHS, GGS	<p>Since both <code>problem = "1 Data inconsistency"</code> and <code>problem = "2 Problems due to panel linkage (GLHS, GGS)"</code> can occur together, <code>problem = "4 Data inconsistency + problems due to panel linkage (GLHS, GGS)"</code> is assigned for this combination.</p>

Value	Value label	Survey programs concerned	Explanation
5	Data inconsistency + higher order partnerships to same partner (pairfam, FFS)	pairfam, FFS	Since both <code>problem = "1 Data inconsistency"</code> and <code>problem = "3 Higher order partnerships to same partner (pairfam, FFS)"</code> can occur together, <code>problem = "5 Data inconsistency + higher order partnerships to same partner (pairfam, FFS)"</code> is assigned for this combination.
6	Duplicate Multi-Actor-Design (SHARE)	SHARE	Due to the multi-actor design of SHARE, there are partnerships in the data set that were reported from the perspective of both respondents. To enable users to remove one of the partnerships for their analyses, the partnerships for one partner each are marked with the value <code>problem = "6 Duplicate Multi-Actor-Design (SHARE)"</code> . For heterosexual partnerships the value 6 was assigned to men and for homosexual partnerships to respondents with a higher <code>aid</code> .