

HaSpaD - Datenhandbuch (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 - Datenhandbuch (September 2021)*. (GESIS Papers, 2021/13). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.75134>

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 | 13

HaSpaD - Datenhandbuch

September 2021

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

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG)

GESIS Papers 2021|13

HaSpaD - Datenhandbuch

September 2021

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

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG)

GESIS – Leibniz-Institut für Sozialwissenschaften

GESIS Papers

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

E-Mail: sonja.schulz@gesis.org

Studentische Mitarbeitende

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

Inhaltsverzeichnis

1	Quickstart	7
2	Das Projekt HaSpaD.....	9
2.1	Überblick über alle verwendeten Umfrageprogramme.....	10
2.2	Datenbeschaffung.....	21
2.3	Zitation der Daten	22
3	Von den Ursprungsdaten zum Zieldatensatz - Erläuterungen zum Stata-Code	25
3.1	Systemvoraussetzungen.....	25
3.2	Vor dem Ausführen des Main-Files main.do	26
3.3	Während des Ausführens des Main-Files main.do	30
3.4	Nach dem Ausführen des Main-Files main.do	30
4	Datenstruktur	32
4.1	Identifikation von Ursprungsstudien und -datensätzen im HaSpaD-Datensatz.....	32
4.2	ID-Variablen.....	34
4.3	Biografie-Variablen	35
4.4	Flag-Variablen	36
4.5	Missing-Schema	37
5	Besonderheiten der HaSpaD-Daten durch die Verknüpfung der verschiedenen Datenquellen	39
5.1	Variierende Definitionen des Partnerschaftsbegriffs.....	39
5.2	Unterschiede maximal berichtbarer Partnerschaften	42
5.3	Variierender Beobachtungszeitraum von Partnerschaften.....	44
5.4	Unterschiedliche Referenzpopulationen	52
5.5	Umgang mit Survey-Gewichten.....	52
5.6	Imputation von Datumsangaben im HaSpaD-Datensatz	55
5.7	Längsschnittliche Aufbereitung von Panelbefragungen ohne Partner-ID	56
6	Übersicht über die harmonisierten Variablen	61
6.1	Biografiedaten.....	61
6.1.1	ID-Variablen.....	61
6.1.2	Biografie-Variablen.....	65
6.1.3	Imputed-Variablen	69
6.1.4	Original-Variablen	73
6.1.5	Zusätzliche Flag-Variablen.....	76
6.2	Harmonisierte Zusatzvariablen	78
7	Literaturverzeichnis.....	79

A	Anhang	85
A. 1	Beschaffungsmöglichkeiten der in HaSpaD harmonisierten Umfrageprogramme	85
A. 2	Der HaSpaD-Programm-Code	90
A. 3	Dokumentationen der Ursprungsdatenquellen	95
A. 4	Imputationsgrad der Biografievariablen	98
A. 5	Konstruktion der Original-Variablen der Biografiedaten	101
A. 6	Ausprägungen der Variable problem	109

Abbildungsverzeichnis

Abbildung 1: Systemvoraussetzungen	7
Abbildung 2: Ordnerstruktur vor der Ausführung des Main-Files	27
Abbildung 3: Anleitung zum Entpacken der Ursprungsdatensätze (für Programm 7-Zip)	29
Abbildung 4: Ordnerstruktur nach Entpacken der Source-Datensätze (Ausschnitt)	29
Abbildung 5: Ordnerstruktur nach Ausführung des Main-Files (Ausschnitt)	31
Abbildung 6: Vorgehensweise der Verknüpfung gleicher Partnerschaften	57
Abbildung A 1: Beispielhafter Prozess der Erstellung der kumulierten Partnerschaftsbiografien in Stata	91
Abbildung A 2: Ordnerstruktur nach Aufbereitung des Biografiedatensatzes	92
Abbildung A 3: Ordnerstruktur nach Aufbereitung der Zusatzvariablen	94

Tabellenverzeichnis

<i>Tabelle 1:</i> Überblick über alle Umfrageprogramme und Teilstudien	11
<i>Tabelle 2:</i> Zitationsrichtlinien der harmonisierten Umfrageprogramme	22
<i>Tabelle 3:</i> Im Main-File verwendete Packages	25
<i>Tabelle 4:</i> Bezeichnungen der Umfrageprogramme im HaSpaD-Zieldatensatz	32
<i>Tabelle 5:</i> Übersicht über die ID-Variablen	34
<i>Tabelle 6:</i> Übersicht über die Biografie-Variablen	35
<i>Tabelle 7:</i> Übersicht über Flag-Variablen mit Informationen zum Imputationsgrad der Biografievariablen	36
<i>Tabelle 8:</i> Übersicht über Flag-Variablen mit den Originalwerten	36
<i>Tabelle 9:</i> Übersicht über zusätzliche Flag-Variablen	37
<i>Tabelle 10:</i> Verwendetes Missing-Schema	37
<i>Tabelle 11:</i> Institutionalierungsgrade nach Studie	39
<i>Tabelle 12:</i> Anzahl maximal berichtbarer Partnerschaften nach Studie	43
<i>Tabelle 13:</i> Beginn der Beobachtung nach Studie	45
<i>Tabelle 14:</i> Ende der Beobachtung nach Studie	48
<i>Tabelle 15:</i> Imputierte Werte der Begin- und End-Variablen bei saisonalen Angaben	56
<i>Tabelle 16:</i> Plausibilitätsannahmen zur Identifikation in Welle 1 und Welle 2 berichteter identischer Partnerschaften	58
<i>Tabelle 17:</i> Bestimmung der Ober- und Untergrenzen nach Kombination des Imputationsgrades	59
<i>Tabelle 18:</i> Beispiel der Bestimmung der Unter- und Obergrenze nach Kombination des Imputationsgrades	60
<i>Tabelle 19:</i> Kodierung der Monate bei Vorliegen saisonaler Angaben	60

<i>Tabelle 20:</i> ID-Variable: sid	61
<i>Tabelle 21:</i> ID-Variable: did.....	62
<i>Tabelle 22:</i> ID-Variable: aid.....	63
<i>Tabelle 23:</i> ID-Variable: id.....	63
<i>Tabelle 24:</i> ID-Variable: pid.....	64
<i>Tabelle 25:</i> ID-Variable: pid_o.....	64
<i>Tabelle 26:</i> ID-Variable: partner_pointer	64
<i>Tabelle 27:</i> Biografie-Variable: intd	65
<i>Tabelle 28:</i> Biografie-Variable: relbeg	65
<i>Tabelle 29:</i> Biografie-Variable: relend	66
<i>Tabelle 30:</i> Biografie-Variable: cohbeg	66
<i>Tabelle 31:</i> Biografie-Variable: cohend	67
<i>Tabelle 32:</i> Biografie-Variable: marbeg.....	67
<i>Tabelle 33:</i> Biografie-Variable: marend.....	68
<i>Tabelle 34:</i> Biografie-Variable: dop	69
<i>Tabelle 35:</i> Imputed-Variable: flag_relbeg_imp	69
<i>Tabelle 36:</i> Imputed-Variable: flag_relend_imp	70
<i>Tabelle 37:</i> Imputed-Variable: flag_cohbeg_imp	70
<i>Tabelle 38:</i> Imputed-Variable: flag_cohend_imp	71
<i>Tabelle 39:</i> Imputed-Variable: flag_marbeg_imp.....	71
<i>Tabelle 40:</i> Imputed-Variable: flag_marend_imp.....	72
<i>Tabelle 41:</i> Imputed-Variable: flag_dop_imp	72
<i>Tabelle 42:</i> Original-Variable: flag_relbeg_ori	73
<i>Tabelle 43:</i> Original-Variable: flag_relend_ori	74
<i>Tabelle 44:</i> Original-Variable: flag_cohbeg_ori	74
<i>Tabelle 45:</i> Original-Variable: flag_cohend_ori	74
<i>Tabelle 46:</i> Original-Variable: flag_marbeg_ori.....	75
<i>Tabelle 47:</i> Original-Variable: flag_marend_ori.....	75
<i>Tabelle 48:</i> Original-Variable: flag_dop_ori	76
<i>Tabelle 49:</i> Zusätzliche Flag-Variable: flag_beg.....	76
<i>Tabelle 50:</i> Zusätzliche Flag-Variable: problem.....	77
<i>Tabelle A 1:</i> Beschaffungsmöglichkeiten der in HaSpaD harmonisierten Umfrageprogramme.....	85
<i>Tabelle A 2:</i> Dokumentationen der Ursprungsdatenquellen	95
<i>Tabelle A 3:</i> Imputationsgrad der Biografievariablen.....	98
<i>Tabelle A 4:</i> Konstruktion der Original-Variablen der Biografiedaten	101
<i>Tabelle A 5:</i> Ausprägungen der Variable problem.....	109

1 Quickstart

Das HaSpaD-Projekt ermöglicht Forschenden die Erstellung eines individualisierten, harmonisierten Datensatzes verschiedener paarbiographischer Umfrageprogramme aus Deutschland. Die Harmonisierungen sind syntaxbasiert und über den HaSpaD-Wizard (<https://haspad.gesis.org/wizard>) abrufbar. Dieser Quickstart gibt einen Überblick über die notwendigen Systemvoraussetzungen (*Abbildung 1*) und dient als schrittweise Kurz-Anleitung für die Harmonisierungsprozesse, von der Auswahl der Umfrageprogramme bis zum HaSpaD-Zieldatensatz.

Benötigte Software	<ul style="list-style-type: none">• Stata 14 SE oder aktueller• Datenkompressionsprogramm (zum Beispiel 7-Zip)
Erforderlicher Speicherplatz	<ul style="list-style-type: none">• mindestens 15 GB
Sonstiges	<ul style="list-style-type: none">• Internetzugang (für Download zusätzlicher Stata Packages, siehe <i>Abschnitt 3.1, Verwendete Packages</i>)

Abbildung 1: Systemvoraussetzungen

Folgende Schritte müssen vollzogen werden:

1. Auswahl der Zusatzvariablen und Umfrageprogramme mit Hilfe des HaSpaD- Harmonization Wizards auf <https://haspad.gesis.org/wizard>.
2. Download des ZIP-Ordners `haspad.zip` und Entpacken des Ordners. Die Ordnerstruktur muss wie folgt beibehalten werden:

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

Um ein Überschreiben des Ordners zu vermeiden, wird beim Herunterladen ein Datum in den Dateinamen des Ordners `haspad.zip` eingefügt. Der entpackte Ordner kann in „haspad“ umbenannt werden. Alternativ kann auch der datumsspezifische Pfad in das main-File übertragen werden. Im Folgenden steht „haspad“ für das entpackte .zip-File.

3. Beschaffung der Ursprungsdatenquellen über GESIS und/oder die jeweiligen Forschungsdatenzentren. Falls diese in ZIP-Ordern geliefert werden: bitte noch nicht entpacken.

4. Speichern der Ursprungsdatenquellen bzw. ZIP-Ordner der Ursprungsdatenquellen aus Schritt 3 im Verzeichnis `~haspad/2_source`.
5. Falls die Ursprungsdatenquellen in ZIP-Ordnern geliefert wurden: Entpacken der ZIP-Ordner aus Schritt 4 in `~haspad/2_source`
Bei Problemen beim Entpacken Hinweise in FAQ auf der HaSpaD-Website beachten.
6. Öffnen des Do-Files `main.do` im Ordner `haspad`.
7. In Zeile 40 des `main.do` den eigenen Dateipfad einfügen, in dem der `haspad`-Ordner gespeichert wurde

Beispiel: entpackter Ordner `haspad` wurde nach `c:/mydata` gespeichert

```
> 40 cd "c:/mydata/haspad"
```
8. Ausführen des kompletten Do-Files `main.do`.
9. Der Zieldatensatz `target.dta` mit den von Ihnen gewählten Variablen kann jetzt unter `~haspad/4_target` geöffnet werden.

2 Das Projekt HaSpaD

In dem Projekt „Harmonisierung und Synthese von paarbiografischen Daten (HaSpaD)“, gefördert durch die Deutsche Forschungsgemeinschaft, werden paarbiografische Daten Deutschlands größter Umfrageprogramme harmonisiert und kumuliert^{1,2}. Die Verknüpfung unterschiedlicher Datenquellen dient drei Zielen: (1) Durch die Nutzung aller thematisch relevanten und für Sekundäranalysen verfügbaren deutschen Forschungsdaten wird eine umfassende Datengrundlage für Forschungssynthesen geschaffen. Diese Datengrundlage ermöglicht zuverlässigere Erkenntnisse über verschiedene Erklärungsfaktoren von Partnerschaftsereignissen wie Trennung, Scheidung oder Partnerwahl zu generieren, als dies mit einer einzelnen Datenquelle möglich wäre. (2) Basierend auf dieser Datenkumulation und den daraus resultierenden höheren Fallzahlen können bislang unbeantwortete Fragen zu Ehe- und Partnerschaftsstabilität aus historischer und lebensverlaufstheoretischer Perspektive, insbesondere zum langfristigen Anstieg der Scheidungsraten und zum sozialen Wandel von Risikofaktoren für Trennungen, untersucht werden. (3) Da es in den umfragebasierten Sozialwissenschaften kaum Vorbilder für solche umfangreichen Forschungssynthesen von heterogenen Datenquellen gibt, soll das HaSpaD-Projekt als best-practice Beispiel weitere Aktivitäten in diesem Bereich motivieren. Die verfügbaren Umfragedaten sind dabei in mehrfacher Hinsicht heterogen, etwa in der Datenstruktur, den verwendeten Gewichtungsschemata oder der Messung theoretischer Konstrukte.

Das HaSpaD-Projekt stellt ein webbasiertes Tool zur Verfügung, mit dessen Hilfe ein individuell zugeschnittener Stata-Aufbereitungscode zur Harmonisierung ausgewählter Variablen und Umfrageprogramme erstellt wird. Über ein Web-Interface (HaSpaD Harmonization Wizard) auf <https://haspad.gesis.org/wizard> können Variablen und Umfrageprogrammen ausgewählt werden. Eine ZIP-Datei mit angepasstem Metajob (Main-File als Stata-Do-File) und Aufbereitungs-Programmdateien zur Erstellung eines kumulierten Datensatzes steht dann zum Download bereitherunterladen. Das Main-File muss nur an wenigen Stellen angepasst werden. Für die Beschaffung der Daten der zu harmonisierenden Umfrageprogramme über Nutzungsverträge sind Nutzende selbst verantwortlich (siehe *Kapitel 2.2*). Dieser Schritt ist erforderlich, da das HaSpaD-Projekt selbst keine Primärdaten bereitstellt, sondern ausschließlich Stata-Programmcode zur Aufbereitung eines harmonisierten und kumulierten Gesamtdatensatzes. Die Rechte für die Primärdatenquellen liegen bei den zuständigen Datenerzeugern.

Der verfügbare Datensatz (HaSpaD-Zieldatensatz³) beinhaltet

- 1) Biografiedaten** der Partnerschaften⁴, wie Beziehungs-, Kohabitations-, Ehedaten, sowie Sterbedaten der Partner, Interviewdatum, zusätzliche Identifikationsvariablen und Flag-Variablen

¹ DFG - Projektnummer 316901171 (<https://gepris.dfg.de/gepris/projekt/316901171>)

² Harmonisierung von Messungen bedeutet, dass eine oder mehrere Source-Variablen eines Ursprungsdatensatzes mittels festgelegter Transformationsregeln zu einer Target-Variable (Zielvariable) rekodiert werden. Während sich die Ausprägungen der Variablen aus den Ursprungsdatenquellen unterscheiden können, sind diejenigen der Zielvariablen über jede Ausgangsquelle hinweg identisch. Wolf et al. (2016, S. 514) beschreiben eine Harmonisierung wie folgt: „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 Synonym dafür werden folgende Begriffe verwendet: Zieldatensatz, Targetdatensatz, HaSpaD-Daten.

⁴ Eine Partnerschaft bezeichnet hier eine intime Beziehung zu den Partnern, welche aus den Institutionalisierungsgraden/Institutionalisierungsphasen (synonym: Partnerschaftstyp, Partnerschaftsphase, Partner-

2) Zusatzvariablen der über die Einzelstudien hinweg harmonisierten Variablen. Beispiele hierfür sind Geburtsdaten, Geschlecht, Staatsbürgerschaft oder Bildungsstand.

Analyseeinheiten des Zieldatensatzes sind die Partnerschaften der Befragten. Eine Anleitung zur Erstellung des Zieldatensatzes findet sich im Quickstart, sowie ausführlich im *Kapitel 3*.

Forschende, die die HaSpaD-Daten nutzen, müssen alle verwendeten Ursprungsdatensätze und zusätzlich das HaSpaD-Projekt zitieren (siehe Zitierhinweise in *Kapitel 2.3*). Lediglich die Ursprungsdatensätze zu zitieren ist nicht ausreichend, da das HaSpaD-Projekt zwar keine Primärdaten bereitstellt, jedoch die Harmonisierung und Kumulation der Datenquellen als wissenschaftliche Leistung des Projekts zu nennen sind. Im folgenden Kapitel werden die von HaSpaD harmonisierten Umfrageprogramme vorgestellt.

2.1 Überblick über alle verwendeten Umfrageprogramme

Insgesamt werden Daten von neun verschiedenen Umfrageprogrammen verwendet, die längsschnittliche Informationen zu Partnerschafts- und Ehebiografien enthalten. Die Grundgesamtheit aller Umfrageprogramme setzt sich aus in Deutschland lebenden Personen zusammen, wobei die Datenerhebung durch zufallsbasierte Stichprobenverfahren (oder Nachbefragungen von Personen, die bei Erstinterview zufällig gezogen wurden) realisiert wurde⁵. Dabei ist zu beachten, dass einige Studien im Querschnittsdesign, wie die ALLBUS-Kumulation (GESIS 2018), Mannheimer Scheidungsstudie (Esser, Gostomski & Hartmann 2018a) und der Fertility and Family Survey (Bundesinstitut für Bevölkerungsforschung 2002a) und andere im Paneldesign wie pairfam (Brüderl et al. 2021a), das Sozio-oekonomische Panel (Liebig et al. 2019) und SHARE (SHARE Waves 1, 2 und 3) vorliegen. Eine Zwischenstellung nehmen die Familiensurveys (DJI 1992a, 2018a, 2018b, 2018c), die Lebensverlaufsstudien (Mayer 1995a, 1995b, 2004a, 2004b, 2014a, 2018a, 2018b, 2018c; Mayer & Kleinhenz 2004a) und der Generations and Gender Survey (GGs Wave 1; GGS Wave 2) ein, in welchen eine Teilmenge der Befragten der jeweiligen Ausgangsstudie in einer weiteren Welle erneut befragt werden. Zusätzlich sind das pairfam, SHARE und das Sozio-oekonomische Panel im Multi-Actor-Design angelegt: Sobald die Partner mit den Befragten (Ankerperson) in einem gemeinsamen Haushalt leben, werden sie in die Stichprobe integriert und berichten ebenfalls Partnerschaftsbiografien. Bei der pairfam-Studie werden die Ankerpersonen um Erlaubnis gebeten, ihre Partner ebenfalls zu befragen. Dabei spielt es keine Rolle, ob die Ankerpersonen mit ihren Partnern zusammenleben oder nicht (Brüderl et al. 2021, S. 5). *Tabelle 1* gibt nachfolgend einen Überblick über die Umfrageprogramme, deren Teilstudien, Titel und das Studiendesign. Weiterhin ist dort für jede Teilstudie der dazugehörige Methodenbericht zitiert, in welchem nähere Informationen zur methodischen Umsetzung der Studie gegeben sind.

schaftsgrad) 1. *Beziehungsphase* (Beziehungsbeginn und -ende), 2. *Kohabitationsphase* (Kohabitationsbeginn und -ende) und 3. *Ehephase* (Ehebeginn und -ende) bestehen kann. Der Begriff „Beziehung“ ist somit nur ein Teil einer Partnerschaft und wird im Folgenden nicht synonym zur Partnerschaft als Ganzes verwendet.

⁵ Eine Besonderheit stellt die Mannheimer Scheidungsstudie dar. Hier wurde zunächst ein Screening-Verfahren im Rahmen von anderen, zufallsbasierten Befragungen des Instituts GFM-GETAS durchgeführt. Hieran anschließend erfolgte eine disproportionale geschichtete Stichprobenziehung anhand der Schichtungsmerkmale „Scheidung der ersten Ehe“, „Geschlecht“ und „Wohnort“ (Esser, Gostomski & Hartmann 2018c).

Tabelle 1: Überblick über alle Umfrageprogramme und Teilstudien

Titel der Studie	Erhebungs- jahre	Gesamtanzahl jemals befragter Personen / Gesamtanzahl jemals befragter Personen mit Paarbiografie	Grundgesamtheit	Stichprobe	Erhebungs- design	Dokumentation DOI
Panel Analysis of Intimate Relationships and Family Dynamics (pairfam), Data Release 12.0.0	2008-2020, jährliche Erhebung	18.912 / 16.191	<p>Region</p> <p><i>Hauptstichprobe</i> Individuen in Privathaushalten Gesamtdeutschlands lebend</p> <p><i>Auffrischungstichprobe</i> In Privathaushalten Gesamtdeutschlands lebend</p> <p><i>Zusatzstichprobe DemoDiff</i> In Privathaushalten Ostdeutschlands lebend</p> <p>Staatsangehörigkeit Deutsche und Ausländer (Deutschsprachig)</p> <p>Geschlecht Frauen und Männer</p> <p>Alter / Geburtskohorte</p> <p><i>Hauptstichprobe</i> Geburtskohorten 1971-1973, 1981-1983, 1991-1993</p> <p><i>Auffrischungstichprobe</i> 1981-1983, 1991-1993, 2001-2003</p> <p><i>Zusatzstichprobe DemoDiff</i> 1971-73, 1981-83</p>	<p>Anzahl der Stichproben</p> <p>Drei Stichproben: Hauptstichprobe, Zusatzstichprobe DemoDiff und Auffrischungstichprobe</p> <p>Stichprobendesign</p> <p>zweistufige Stichprobenziehung</p>	Panelbefragung Multi-Actor-Design	Brüderl et al. 2021c DOI: 10.4232/pairfam.5678.12.0.0 .

Titel der Studie	Erhebungsjahre	Gesamtanzahl jemals befragter Personen / Gesamtanzahl jemals befragter Personen mit Paarbiografie	Grundgesamtheit	Stichprobe	Erhebungsdesign	Dokumentation DOI
Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) – Kumulation 1980 – 2016	1980, 1982, 1984, 1986, 1988, 1991, 2000, 2010	24.786 (Nur Erhebungsjahre mit Erhebung der Ehebiografie) / 19.498	Region In Privathaushalten Westdeutschlands lebend (BRD) bis einschließlich 1990, Gesamtdeutschland ab 1991 Staatsangehörigkeit Deutsche und Ausländer (Deutschsprachig) Geschlecht Frauen und Männer Alter / Geburtskohorte ab 18 Jahre	Anzahl der Stichproben Eine Stichprobe pro Erhebungsjahr Stichprobendesign mehrstufig geschichtete Zufallsauswahl (ADM) 1980-1992/1992 zweistufig 1994, 1996, ab 2000	kumulierter Querschnitt Single-Actor-Design	Baumann & Schulz 2018 DOI: 10.4232/1.13029
Wandel und Entwicklung familiärer Lebensformen in Westdeutschland (Familien-survey)	1988	10.043 / 4.562	Region In Privathaushalten Westdeutschlands (BRD) lebend Staatsangehörigkeit Deutsch Geschlecht Frauen und Männer Alter / Geburtskohorte 18-55 Jahre	Anzahl der Stichproben Drei Teilstichproben Stichprobendesign <i>Teilstichprobe 1</i> dreistufige Zufallsauswahl (ADM) <i>Teilstichprobe 2</i> Adressen-Random <i>Teilstichprobe 3</i> Wiederholungsbefragung aus Quotenstichprobe bei Pretest	Querschnitt, teilweise Panel Single-Actor-Design	Eschenlohr et al. 1989 DOI: 10.4232/1.2245

Titel der Studie	Erhebungsjahre	Gesamtanzahl jemals befragter Personen / Gesamtanzahl jemals befragter Personen mit Paarbiografie	Grundgesamtheit	Stichprobe	Erhebungsdesign	Dokumentation DOI
Familie und Partner-beziehungen in Ostdeutschland (Familiensurvey)	1991	1.951 / 1.778	Region Ostdeutschland Staatsangehörigkeit Deutsch Geschlecht Frauen und Männer Alter / Geburtskohorte 18-55 Jahre	Anzahl der Stichproben Eine Stichprobe Stichprobendesign einfache Zufallsauswahl	Querschnitt Single-Actor-Design	Haas 1991 DOI: 10.4232/1.13196
Wandel und Entwicklung familialer Lebensformen - 2. Welle (Familiensurvey)	1994	10.994 / 8.215	Region Gesamtdeutschland Staatsangehörigkeit Deutsch Geschlecht Frauen und Männer Alter / Geburtskohorte 18-61 Jahre	Anzahl der Stichproben Drei Teilstichproben Stichprobendesign <i>Teilstichprobe 1</i> Panel West: Wiederbefragung von Personen aus Familiensurvey 1.1 <i>Teilstichprobe 2</i> Survey Ost - unsystematische Zufallsstichprobe der 18-55-jährigen Wohnbevölkerung (random route) <i>Teilstichprobe 3</i> 18-30jährige West: unsystematische Zufallsstichprobe der 18-30jährigen Wohnbevölkerung (random route)	Querschnitt, teilweise Wiederholungsbefragung Single-Actor-Design	Albers et al. 1994 DOI: 10.4232/1.13197

Titel der Studie	Erhebungsjahre	Gesamtanzahl jemals befragter Personen / Gesamtanzahl jemals befragter Personen mit Paarbiografie	Grundgesamtheit	Stichprobe	Erhebungsdesign	Dokumentation DOI
Wandel und Entwicklung familiärer Lebensformen - 3. Welle (Familiensurvey)	2000	10.318 / 8.862	Region Gesamtdeutschland Staatsangehörigkeit Deutsch Geschlecht Frauen und Männer Alter / Geburtskohorte 18-67 Jahre	Anzahl der Stichproben Zwei Teilstichproben Stichprobendesign <i>Teilstichprobe 1</i> Querschnittserhebung von 18-55jährigen Wohnbevölkerung <i>Teilstichprobe 2</i> Panel West: Wiederbefragung von Panelteilnehmer des Familiensurveys 1.1 und Familiensurveys 2	Querschnitt, teilweise Wiederholungsbefragung Single-Actor-Design	Schneekloth 2000 DOI: 10.4232/1.13198
Mannheimer Scheidungsstudie 1996	1996	5.020 / 5.020	Region Gesamtdeutschland Staatsangehörigkeit Deutsche und Ausländer Geschlecht Frauen und Männer Alter / Geburtskohorte ab 18 Jahre	Anzahl der Stichproben Eine Stichprobe Stichprobendesign Disproportional geschichtete Stichprobenziehung	Querschnitt Single-Actor-Design	Esser, Gostomski & Hartmann 2018c DOI: 10.4232/1.13056
Deutscher Fertility and Family Survey 1992	1992	10.012 / 6.555	Region Gesamtdeutschland Staatsangehörigkeit Deutsch Geschlecht	Anzahl der Stichproben Eine Stichprobe Stichprobendesign mehrstufig geschichtete Zufallsauswahl (ADM und FGW)	Querschnitt Single-Actor-Design	Pohl 1995 DOI: 10.4232/1.3400

Titel der Studie	Erhebungsjahre	Gesamtanzahl jemals befragter Personen / Gesamtanzahl jemals befragter Personen mit Paarbiografie	Grundgesamtheit	Stichprobe	Erhebungsdesign	Dokumentation DOI
Lebensverläufe und gesellschaftlicher Wandel: Lebensverläufe und Wohlfahrtsentwicklung (Lebensverlaufsstudie LV-West I)	1981-1983	2.171 / 2.002	Frauen und Männer Alter / Geburtskohorte 20-39 Jahre Region Westdeutschland (BRD) Staatsangehörigkeit Deutsch Geschlecht Frauen und Männer Alter / Geburtskohorte Geburtskohorten 1929-31, 1939-41, 1949-51	Anzahl der Stichproben Eine Stichprobe Stichprobendesign mehrstufig geschichtete Zufallsauswahl (ADM)	Querschnitt Single-Actor-Design	Mayer et al. 1989 DOI: 10.4232/1.13193
Lebensverläufe und gesellschaftlicher Wandel: Die Zwischenkriegskohorte im Übergang zum Ruhestand (Lebensverlaufsstudie LV-West II A - Persönliche Befragung)	1985-1986	407 / 395	Region In Privathaushalten Westdeutschlands (BRD) lebend Staatsangehörigkeit Deutsch Geschlecht Frauen und Männer Alter / Geburtskohorte Geburtskohorte 1919-21	Anzahl der Stichproben Eine Stichprobe Stichprobendesign mehrstufig geschichtete Zufallsauswahl (ADM)	Querschnitt Single-Actor-Design	Brückner 1993 DOI: 10.4232/1.13194
Lebensverläufe und gesellschaftlicher Wandel:	1987-1988	1.005 / 948	Region Westdeutschland (BRD) Staatsangehörigkeit	Anzahl der Stichproben Eine Stichprobe Stichprobendesign	Querschnitt Single-Actor-Design	Brückner 1993 DOI: 10.4232/1.2647

Titel der Studie	Erhebungsjahre	Gesamtanzahl jemals befragter Personen / Gesamtanzahl jemals befragter Personen mit Paarbiografie	Grundgesamtheit	Stichprobe	Erhebungsdesign	Dokumentation DOI
Die Zwischenkriegskohorte im Übergang zum Ruhestand (Lebensverlaufsstudie LV-West II T - Telefonische Befragung)			Deutsch Geschlecht Frauen und Männer Alter / Geburtskohorte Geburtskohorte 1919-21	einfache Zufallsauswahl		
Lebensverläufe und gesellschaftlicher Wandel: Berufszugang in der Beschäftigungskrise (Lebensverlaufsstudie LV-West III)	1988-1989	2.008 / 1.521	Region Westdeutschland (BRD) Staatsangehörigkeit Deutsch Geschlecht Frauen und Männer Alter / Geburtskohorte Geburtskohorten 1954-56, 1959-61	Anzahl der Stichproben Eine Stichprobe Stichprobendesign einfache Zufallsauswahl	Querschnitt Single-Actor-Design	Brückner et al. 1995 DOI: 10.4232/1.2648
Lebensverläufe und historischer Wandel in Ostdeutschland (Lebensverlaufsstudie LV-DDR)	1991-1992	2.331 / 2.246	Region Ostdeutschland (ehemals DDR) Staatsangehörigkeit Deutsch Geschlecht Frauen und Männer Alter / Geburtskohorte	Anzahl der Stichproben Eine Stichprobe Stichprobendesign einfache Zufallsauswahl	Querschnitt, Single-Actor-Design	Hess & Smid 1991 DOI: 10.4232/1.2644

Titel der Studie	Erhebungs- jahre	Gesamtanzahl jemals befragter Personen / Gesamtanzahl jemals befragter Personen mit Paarbiografie	Grundgesamtheit	Stichprobe	Erhebungs- design	Dokumentation DOI
Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost 71)	1996-1997	610 / 534	Geburtskohorten 1929-31, 1939-41, 1951-53, 1959-61 Region Ostdeutschland (ehemals DDR) Staatsangehörigkeit Deutsch Geschlecht Frauen und Männer Alter / Geburtskohorte Geburtskohorte 1971	Anzahl der Stichproben Eine Stichprobe Stichprobendesign einfache Zufallsauswahl	Querschnitt, teilweise Panel Single-Actor- Design	Matthes, Lichtwardt & Mayer 2004 DOI: 10.4232/1.3926
Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost Panel)	1996-1997	1.078 / 995	Region Ostdeutschland (ehemals DDR) Staatsangehörigkeit Deutsch Geschlecht Frauen und Männer Alter / Geburtskohorte Geburtskohorten 1939-41, 1951-53, 1959-61	Anzahl der Stichproben Eine Stichprobe Stichprobendesign Wiederbefragung von Personen aus der Lebensverlaufsstudie LV-DDR	Wiederholungs- befragung von Lebensverlaufsstudie LV-DDR) Single-Actor- Design	Goedicke, Lichtwardt & Mayer 2004 DOI: 10.4232/1.3925
Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost Panel)	1996-1997	329 / 321	Region Ostdeutschland (ehemals DDR) Staatsangehörigkeit Deutsch Geschlecht	Anzahl der Stichproben Eine Stichprobe Stichprobendesign Wiederbefragung von Personen aus der Lebensverlaufsstudie LV-DDR	Wiederholungs- befragung von Lebensverlaufsstudie LV-DDR) Single-Actor- Design	Goedicke, Lichtwardt & Mayer 2004 DOI: 10.4232/1.3925

Titel der Studie	Erhebungsjahre	Gesamtanzahl jemals befragter Personen / Gesamtanzahl jemals befragter Personen mit Paarbiografie	Grundgesamtheit	Stichprobe	Erhebungsdesign	Dokumentation DOI
Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland (Lebensverlaufsstudie LV-West 64/71)	1998-1999	2.909 / 2.583	Frauen und Männer Alter / Geburtskohorte Geburtskohorte 1929-31 Region Westdeutschland Staatsangehörigkeit Deutsche und Ausländer Geschlecht Frauen und Männer Alter / Geburtskohorte Geburtskohorte 1964 und 1971	Anzahl der Stichproben Eine Stichprobe Stichprobendesign einfach geschichtete Zufallsauswahl	Querschnitt, teilweise Panel Single-Actor-Design	Hillmert et al. 2004 DOI: 10.4232/1.3927
Frühe Karrieren und Familiengründung: Lebensverläufe der Geburtskohorte 1971 in Ost- und Westdeutschland (Lebensverlaufsstudie LV-Panel 71)	2004-2005	1.069 / 971	Region Ostdeutschland (ehemals DDR) & Westdeutschland Staatsangehörigkeit Deutsch Geschlecht Frauen und Männer Alter / Geburtskohorte Geburtskohorte 1971	Anzahl der Stichproben Zwei Teilstichproben Stichprobendesign <i>Teilstichprobe 1</i> Stichprobe Ost: Wiederbefragung von Personen aus der Lebensverlaufsstudie LV-Ost 71) <i>Teilstichprobe 2</i> Stichprobe West Wiederbefragung von Personen aus der Lebensverlaufsstudie LV-West 64/71, Kohorte 71	Wiederholungsbefragung von Lebensverlaufsstudie LV-Ost 71 und Lebensverlaufsstudie LV-West 64/71, Kohorte 71 Single-Actor-Design	Mayer 2014b DOI: 10.4232/1.5099

Titel der Studie	Erhebungs- jahre	Gesamtanzahl jemals befragter Personen / Gesamtanzahl jemals befragter Personen mit Paarbiografie	Grundgesamtheit	Stichprobe	Erhebungs- design	Dokumentation DOI
Generations and Gender Survey (2005) (Teilstichprobe Deutschland)	Hauptbefragung 2005	10.017 / 8.168	Region In Privathaushalten Gesamtdeutschlands lebend Staatsangehörigkeit Deutsche und Ausländer Geschlecht Frauen und Männer Alter / Geburtskohorte 18-79 Jahre	Anzahl der Stichproben Eine Stichprobe Stichprobendesign dreistufige Zufallsauswahl (ADM)	Querschnitt, teilweise Panel Single-Actor-Design	Fokkema et al. 2016 DOI: 10.14301/llcs.v9i4.500
Generations and Gender Survey (2008) (Teilstichprobe Deutschland)	Hauptbefragung 2008	3.226 / 2.637	Region In Privathaushalten Gesamtdeutschlands lebend Staatsangehörigkeit Deutsche und Ausländer Geschlecht Frauen und Männer Alter / Geburtskohorte 20-83 Jahre	Anzahl der Stichproben Eine Stichprobe Stichprobendesign Wiederbefragung von Personen aus dem Generations and Gender Survey (GGS) 1	Wiederholungsbefragung von GGS (2005) Single-Actor-Design	Fokkema et al. 2016 DOI: 10.14301/llcs.v9i4.500
The Survey of Health, Ageing and Retirement in Europe (SHARE)	2008	1.921/ 1.379	Region Gesamtdeutschland Staatsangehörigkeit Deutsche und Ausländer Geschlecht Frauen und Männer	Anzahl der Stichproben Eine Stichprobe pro Welle Stichprobendesign mehrstufige Stichprobenziehung	Panelbefragung Multi-Actor-Design	Klevmarken, Hesselius & Swensson 2005 DOIs: 10.6103/SHARE.w1.600 , 10.6103/SHARE.w2.600 , 10.6103/SHARE.w3.600

Titel der Studie	Erhebungsjahre	Gesamtanzahl jemals befragter Personen / Gesamtanzahl jemals befragter Personen mit Paarbiografie	Grundgesamtheit	Stichprobe	Erhebungsdesign	Dokumentation DOI
Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2018	1984-2018, jährliche Erhebungen	142.308 / 34.410	<p>Alter / Geburtskohorte Ankerpersonen ab 50 Jahre Ehepartner können jünger als 50 sein</p> <p>Region In Privathaushalten Deutschlands lebend</p> <p>Staatsangehörigkeit Deutsche und Ausländer</p> <p>Geschlecht Frauen und Männer</p> <p>Alter / Geburtskohorte ab 17 Jahre</p>	<p>Anzahl der Stichproben mehrere Stichproben pro Welle möglich (Auffrischungsstichproben, Ergänzungsstichproben (bspw. Zuwanderer, Hochverdiener)</p> <p>Stichprobendesign mehrstufige Stichprobenziehung</p>	Panelbefragung Multi-Actor-Design	Siegers et al. 2020 DOI: 10.5684/soep-core.v35

2.2 Datenbeschaffung

Nach Auswahl der Umfrageprogramme und zusätzlichen Variablen mit Hilfe des Harmonization Wizards muss die Beschaffung der benötigten Daten eigenständig durch die Nutzenden erfolgen. Da die Datenaufbereitung mit der Software Stata erfolgt, müssen die jeweiligen Datenquellen als Stata-Datendateien mit der Endung `*.dta` heruntergeladen werden. Um eine korrekte Aufbereitung gewährleisten zu können, müssen **exakt die Versionen** der Ursprungsdaten ausgewählt werden, die im HaSpaD-Projekt aufbereitet wurden.

Zur Gewährleistung der Nachnutzbarkeit werden in HaSpaD nur Datenquellen einbezogen, die durch ein Forschungsdatenrepositorium als Scientific-Use-File oder Public-Use-File vertrieben werden. *Tabelle A 1* im Anhang bietet einen Überblick über die nach Umfrageprogramm gegliederten Forschungsdateninfrastrukturen und den Datenvertrieb. Weitere Informationen können auf der jeweiligen Homepage des Forschungsdatenzentrums (FDZ) bzw. Datenvertriebs recherchiert werden. Bei der Recherche und Beschaffung der Datenquellen ist Folgendes zu beachten:

- Daten, die durch GESIS vertrieben werden, sind nur in der jeweils aktuellen Version verfügbar. Da es durchaus sein kann, dass ein Update einer im HaSpaD-Projekt verwendeten Datenversion erfolgte, muss die benötigte Version (siehe *Tabelle A 1* im Anhang, Spalte „Benötigte Version“) unter dataservices@gesis.org angefragt werden.
- Die Umfrageprogramme pairfam, Mannheimer Scheidungsstudie und die Lebensverlaufstudien sind Studien der Zugangsklasse C (C-Studien, Stand: 24.11.2019) und damit kostenpflichtig. Unter <https://www.gesis.org/angebot/archivieren-und-registrieren/datenarchivierung/> dem Menüpunkt „Dokumente“ ist eine Preisliste abrufbar.
- GESIS Cryptshare bietet die Möglichkeit, alle angebotenen C-Studien in einer übergeordneten ZIP-Datei (Name `cs-transfer.zip`) herunterzuladen. Das heißt, der Nutzende muss zuerst das Meta-ZIP-File entpacken, um im nächsten Schritt die darin liegenden ZIP-Files nacheinander in den Ordner `~haspad/2_source` zu speichern und nochmals zu entpacken. Bitte beachten, dass die Daten des Sozio-Oekonomisch Panels (SOEP) ebenfalls unter dem Namen `cs-transfer.zip` angeliefert werden. In diesem Fall, jedoch, wird der Ordnername `cs-transfer` beim Entpacken der Daten in `~haspad/2_source` beibehalten.
- Beim Generations and Gender Survey (GGS) steht jeweils nur die aktuellste Daten-Version zum Herunterladen zur Verfügung. Daher sollte unter ggp@nidi.nl die vom HaSpaD-Projekt verwendete Datenquelle beschafft werden, sofern bereits eine neuere Version zum Download bereitsteht.
- Auch bei den SHARE-Daten ist eine Versionskontrolle wichtig. Im Downloadbereich des SHARE-FDZ befindet sich unter der Download-Matrix der aktuellsten Version ein Feld „Archiv, Go to archived data“, unter welchem alle älteren Versionen zum Herunterladen bereitstehen.
- Auch Daten des Sozio-Oekonomischen Panels (SOEP) sind nur in der aktuellsten Version verfügbar. Um frühere Versionen zu beschaffen, muss diese unter soepmail@diw.de angefragt werden.

2.3 Zitation der Daten

Werden Publikationen auf Basis des HaSpaD-Projekts angefertigt, müssen sowohl die verwendeten Ursprungsdatensätze zitiert werden als auch die Arbeiten des HaSpaD-Projekts. Lediglich die Ursprungsdatensätze zu zitieren ist nicht ausreichend, da das HaSpaD-Projekt zwar keine Primärdaten bereitstellt, jedoch die Harmonisierung und Kumulation der Datenquellen als wissenschaftliche Leistung des Projekts zu nennen sind.

Für die korrekte Datenzitation sind die Nutzenden verantwortlich. Die jeweilige aktuelle Zitation findet sich i.d.R. auf der Internetpräsenz des jeweiligen Forschungsdatenzentrums/Datengebers. Zum Zeitpunkt der Veröffentlichung dieses Datenhandbuchs waren die in *Tabelle 2* aufgeführten Zitierweisen gültig.

Das HaSpaD-Projekt wird folgend zitiert:

Schulz, Sonja; Weiß, Bernd; Sterl, Sebastian; Haensch, Anna-Carina; Schmid, Lisa; May, Antonia (2021): *Harmonisierung und Synthese von paarbiografischen Daten: Ein Modellprojekt zur Verknüpfung von Forschungsdaten aus verschiedenen Infrastrukturen (HaSpaD)*. Aufbereitungsfiles zur Erstellung eines Paarbiografiedatensatzes. Version 1.0.0. Köln und Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.7802/2317>

Bei der Zitation des vorliegenden Datenhandbuchs ist folgende Zitation zu beachten:

Schulz, S., Weiß, B., Sterl, S., Haensch, A.-C., Schmid, L. & May, A. (2021). *HaSpaD - Datenhandbuch. Version 1*. (GESIS Papers, 2021/13). Köln: GESIS – Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.75134>

Tabelle 2: Zitationsrichtlinien der harmonisierten Umfrageprogramme

Umfrageprogramm	Datenzitation
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): Beziehungs- und Familienpanel (pairfam). GESIS Datenarchiv, Köln. ZA5678 Datenfile Version 12.0.0, https://doi.org/10.4232/pairfam.5678.12.0.0
ALLBUS	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
Familiensurveys	Deutsches Jugendinstitut (DJI), München (1992). Wandel und Entwicklung familiärer Lebensformen in Westdeutschland (Familiensurvey). GESIS Datenarchiv, Köln. ZA2245 Datenfile Version 1.0.0, doi:10.4232/1.2245 Deutsches Jugendinstitut (DJI), München (2018). Familie und Partnerbeziehungen in Ostdeutschland (Familiensurvey). GESIS Datenarchiv, Köln. ZA2392 Datenfile Version 1.1.0, doi:10.4232/1.13196

Umfrageprogramm	Datenzitation
	Deutsches Jugendinstitut (DJI), München (2018). Wandel und Entwicklung familialer Lebensformen - 2. Welle (Familiensurvey). GESIS Datenarchiv, Köln. ZA2860 Datenfile Version 1.1.0, doi:10.4232/1.13197
Mannheimer Scheidungsstudie 1996	Deutsches Jugendinstitut (DJI), München (2018). Wandel und Entwicklung familialer Lebensformen - 3. Welle (Familiensurvey). GESIS Datenarchiv, Köln. ZA3920 Datenfile Version 1.1.0, doi:10.4232/1.13198 Esser, Hartmut; Gostomski, Christian Babka von; Hartmann, Josef (2018). Mannheimer Scheidungsstudie 1996. GESIS Datenarchiv, Köln. ZA3188 Datenfile Version 2.0.0, doi:10.4232/1.13056
Deutscher Fertility and Family Survey 1992	Bundesinstitut für Bevölkerungsforschung, Wiesbaden (2002). Deutscher Fertility and Family Survey 1992. GESIS Datenarchiv, Köln. ZA3400 Datenfile Version 1.0.0, doi:10.4232/1.3400
Lebensverlaufsstudien	Mayer, Karl U. (2018). 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 Mayer, Karl U. (2018): 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 Mayer, Karl U. (1995). 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, Karl U. (2018). 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 Mayer, Karl U. (1995). Lebensverläufe und historischer Wandel in Ostdeutschland (Lebensverlaufsstudie LV-DDR). GESIS Datenarchiv, Köln. ZA2644 Datenfile Version 1.0.0, doi:10.4232/1.2644 Mayer, Karl U. (2004). Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost 71). GESIS Datenarchiv, Köln. ZA3926 Datenfile Version 1.0.0, doi:10.4232/1.3926 Mayer, Karl U. (2004). Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost Panel). GESIS Datenarchiv, Köln. ZA3925 Datenfile Version 1.0.0, doi:10.4232/1.3925 Mayer, Karl U.; Kleinhenz, Gerhard (2004). 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 Mayer, Karl U. (2014). 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

Umfrageprogramm	Datenzitation
Generations and Gender Survey (GGS)	<p><i>Generations and Gender Survey (2005) (Teilstichprobe Deutschland)</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) (Teilstichprobe Deutschland)</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) / <i>SHARE Wellen 1, 2 und 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-13: 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 (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>(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 style="list-style-type: none"> • 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 • 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 • Börsch-Supan, A. (2017). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 3 – SHARELIFE. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w3.600 • Börsch-Supan, A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaaf, 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
Sozio-oekonomisches Panel (SOEP)	<p>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 Metzger, David Richter, Diana Schacht, Paul Schmelzer, Christian Schmitt, Daniel Schnitzlein, Rainer Siegers, Knut Wenzig, Stefan Zimmermann (2019). Sozio-oekonomisches Panel (SOEP), Daten für die Jahre 1984-2018, Version 35, SOEP, 2019, doi:10.5684/soep.v35.</p>

3 Von den Ursprungsdaten zum Zieldatensatz - Erläuterungen zum Stata-Code

Das HaSpaD-Projekt bietet alle Harmonisierungsprozesse syntaxbasiert an. Nach der Auswahl der Umfrageprogramme und Variablen auf <https://haspad.gesis.org/wizard> steht ein ZIP-File zum download bereit. Dieser beinhaltet alle benötigten Syntaxfiles und ein Main-File zur Ausführung der Harmonisierung, Dokumentationen und die basale HaSpaD-Ordnerstruktur. Nachdem das ZIP-File heruntergeladen und entpacken wurde, müssen die Ursprungsdaten selbstständig beschafft (siehe *Kapitel 2.2*) und in dem dafür vorgesehenen `~haspad/2_source`-Ordner in ihrer ursprünglichen Version abgelegt werden. Dieses Kapitel beginnt mit den Systemvoraussetzungen und führt detailliert über nötige Anpassungen der Dateien schrittweise zum finalen target-Datensatz.

3.1 Systemvoraussetzungen

Verwendete Stata-Version

Zur Erstellung der einzelnen Aufbereitungscode-Dateien wurde Stata SE 15.1 verwendet; der HaSpaD-Code des Main-Files läuft bereits ab Version 14.0. Ältere Stata-Versionen können den Stata-Code nicht verarbeiten, da einige Funktionen (wie beispielsweise die Umwandlung des Datensatzes in das Unicode-Format) erst in neueren Versionen unterstützt werden. Wenn gewünscht, kann die Zeile 32 des `main.do` an die verwendete jüngere Version angepasst werden. Allerdings empfehlen wir, maximal 15.1 zu nutzen, um ein fehlerloses Ausführen des Main-Files zu sichern. Wir empfehlen dringend, mindestens Stata/SE zu nutzen.

Verwendete Packages

Tabelle 3 enthält die Stata-Packages (Stata Ado-files), die zum Aufbereiten der Daten verwendet werden, sowie ihre Funktion und Hinweise zur Dokumentation. Beim Ausführen des Main-Files werden diese Packages, sofern noch nicht installiert, heruntergeladen und installiert. Genutzt werden die Stata-Packages `markdoc`, `weaver`, `dm31` und das selbst geschriebene Ado-file `list_error_cases`. Da diese Packages von Stata beim Ausführen des Main-Files heruntergeladen werden, ist eine Internetverbindung während des Ausführens notwendig.

Tabelle 3: Im Main-File verwendete Packages

Package	Funktion	Dokumentation
markdoc	MarkDoc generiert aus den Do-files der Datenaufbereitung und Harmonisierung eine Dokumentation im HTML-Format.	<p>Dokumentation MarkDoc:</p> <p>http://www.haghish.com/statistics/stata-blog/reproducible-research/markdoc.php</p> <p>(abgerufen am 20.09.2019, 23.20 Uhr)</p> <p>Haghish, E. F. (2014). Dynamic documents in Stata: MarkDoc, Ketchup, and Weaver.</p> <p>(https://www.stata.com/meeting/italy14/abstracts/materials/it14_haghish.pdf)</p> <p>(abgerufen am 20.09.2019, 23.20 Uhr)</p>

Package	Funktion	Dokumentation
weaver	Weaver wird benötigt, damit das Package MarkDoc ausgeführt werden kann.	Dokumentation weaver Haghish, E. F. (2014). Dynamic documents in Stata: MarkDoc, Ketchup, and Weaver. (https://www.stata.com/meeting/italy14/abstracts/materials/it14_haghish.pdf) (abgerufen am 20.09.2019, 23.20 Uhr)
dm31	Der Befehl <code>rmiss2</code> zählt die Anzahl an Fehlern in einer spezifizierten Variablenliste. Im Gegensatz zum Stata-Befehl <code>rmiss</code> akzeptiert <code>rmiss2</code> sowohl string als auch numerisch kodierte Variablen.	Dokumentation rmiss2 (https://www.stata.com/stb/stb26/dm31/rmiss2.hlp) (abgerufen am 20.09.2019, 23.20 Uhr)
list_errorcases	Der Befehl <code>list_errorcases</code> listet alle Partnerschaften innerhalb einer Person, auch wenn nur eine davon bestimmten Kriterien (spezifiziert in einer <code>if</code> -Bedingung) entspricht. Zusätzlich können alle Partnerschaften innerhalb einer Ankerperson in einer Zusatzoption markiert werden.	selbst geschriebenes Ado-file

3.2 Vor dem Ausführen des Main-Files main.do

Um die HaSpaD-Syntax korrekt nutzen zu können, sind einige wenige Schritte nötig, die im Folgenden erläutert werden.

Aufbau des haspad-Ordners und erforderliche Anpassungen

Nach der Auswahl der gewünschten Umfrageprogramme und der zu harmonisierenden Variablen mithilfe des HaSpaD-Harmonization Wizards, wird die Archivdatei `haspad.zip` zum Download angeboten, heruntergeladen und anschließend lokal im gewählten Arbeitsverzeichnis entpackt⁶. Dabei ist darauf zu achten, dass *der Inhalt der* `haspad.zip`-Datei anschließend in einem übergeordneten Verzeichnis mit dem Namen `haspad` liegt. Bei einigen Programmen zum Entpacken von ZIP-Archiven ist es die Voreinstellung, in ein eigenes Verzeichnis mit dem Namen des ZIP-Archivs zu entpacken, bei anderen nicht. Wird korrekt entpackt, liegt das Dokument `README.txt` beispielsweise unter `C:/mydata/haspad/README.txt`. Falsch ist `C:/mydata/README.txt`.

Im Folgenden gehen wir beispielhaft von der Auswahl der Variablen `Geschlecht (sex)` und Erhebungsgebiet (`eastwest`) und der Umfrageprogramme ALLBUS-Kumulation, dem Fertility and Family Survey (Deutschland) und der Mannheimer Scheidungsstudie über den HaSpaD-Harmonization Wizard aus. Der Inhalt des entpackten ZIP-Archivs sieht vor dem Ausführen des Main-Files in diesem Minimalbeispiel wie folgt aus (*Abbildung 2*):

⁶ Eine große Auswahl an Datenkompressionsprogrammen findet sich online und ist teilweise zur freien Nutzung verfügbar.

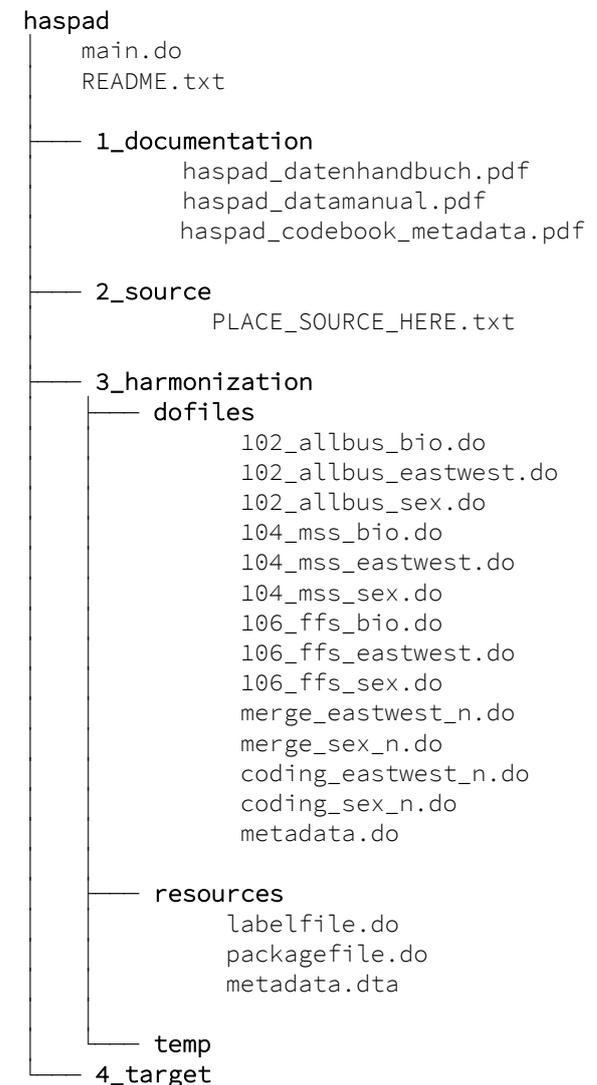


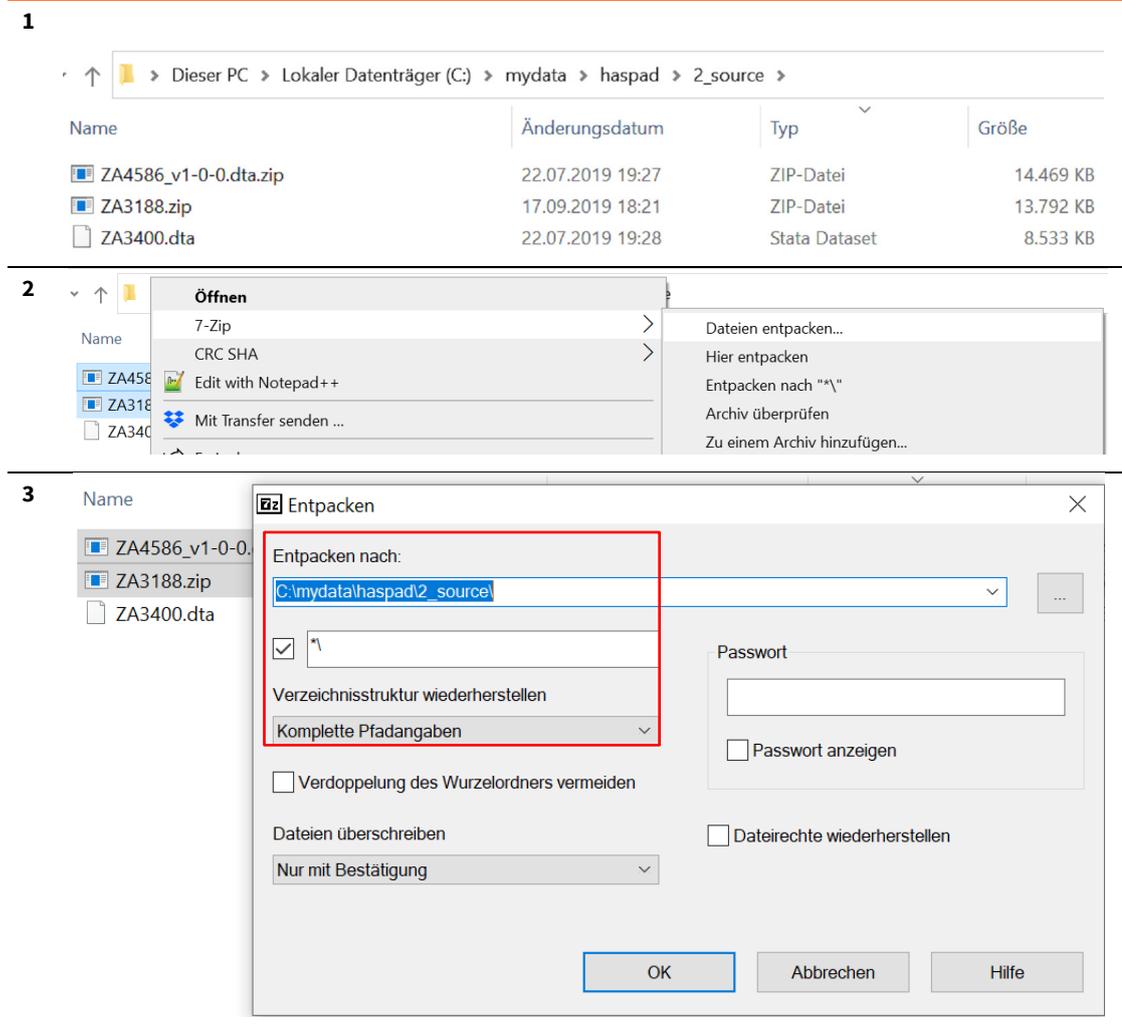
Abbildung 2: Ordnerstruktur vor der Ausführung des Main-Files

Im Ordner `haspad` befinden sich auf der ersten Ebene eine txt-Datei `README.txt` mit einer Erklärung der Ordnerstruktur des `haspad`-Ordners und notwendigen Systemanforderungen (siehe auch *Kapitel 1*), ein Do-file `main.do`, sowie vier Ordner `1_documentation`, `2_source`, `3_harmonization` und `4_target`. Im Ordner `1_documentation` liegt das vorliegende HaSpaD-Datenhandbuch in deutscher und englischer Sprache. Unter `3_harmonization` liegen die benötigten Do-Files zur Datenaufbereitung sowie zusätzlich benötigte Stata-Packages. Der Ordner `4_target` ist noch leer.

Als nächstes müssen Nutzende die beschafften Datenquellen in Ordner `2_source` platzieren und ggf. entpacken (wenn diese durch den Datengeber in einem ZIP-Archiv bereitgestellt wurden). Der Stata-Code zur Datenaufbereitung arbeitet mit relativen Pfaden innerhalb des `haspad`-Ordners. Werden Änderungen an der Verzeichnisstruktur vorgenommen, können die Programmdateien nicht mehr auf die Source- bzw. Zieldateien zugreifen. Daher ist es wichtig, dass beim Entpacken der zu harmonisierenden Daten *keine Datei- oder Ordnernamen verändert* oder einzelne Dateien gelöscht

bzw. in andere Ordner verschoben werden und somit *alle Verzeichnisstrukturen der bereitgestellten Daten erhalten bleiben*.

Nutzende müssen darauf achten, dass das gewählte Datenkompressionsprogramm das ZIP-Archiv in einen neuen Ordner entpackt, der exakt so heißt und eine identische Ordnerstruktur aufweist wie die ZIP-Datei selbst. Beim Programm 7-Zip ist dies die Programmvoreinstellung (vgl. *Abbildung 3*). In *Abbildung 3* wird beispielhaft die ZIP-Datei für die ALLBUS-Kumulation 1980-2016 (ZA4586_v1-0-0.dta.zip) und der Mannheimer Scheidungsstudie (ZA3188.zip) in den Ordner 2_source gespeichert und entpackt. Der Fertilty and Family Survey (ZA3400.dta) wird direkt im dta-Format ausgeliefert und muss lediglich nach haspad/2_source gespeichert werden. Wichtig ist in Screenshot 3 der *Abbildung 3*, dass das Häkchen innerhalb des rot markierten Feldes aktiviert bleibt und die Voreinstellung „Komplette Pfadangaben“ beibehalten wird, damit die Verzeichnisstruktur aller ZIP-Ordner erhalten bleibt; also für jeden entpackten ZIP-Ordner ein eigener Unterordner angelegt wird:



4 > Dieser PC > Lokaler Datenträger (C:) > mydata > haspad > 2_source >

Name	Änderungsdatum	Typ	Größe
ZA4586_v1-0-0.dta	23.04.2020 18:56	Dateiordner	
ZA3188	23.04.2020 18:56	Dateiordner	
ZA4586_v1-0-0.dta.zip	22.07.2019 19:27	ZIP-Datei	14.469 KB
ZA3188.zip	17.09.2019 18:21	ZIP-Datei	13.792 KB
ZA3400.dta	22.07.2019 19:28	Stata Dataset	8.533 KB

Abbildung 3: Anleitung zum Entpacken der Ursprungsdatensätze (für Programm 7-Zip)

Für die ALLBUS-Kumulation lautet der korrekte Speicherort der Stata-Datei `ZA4586_v1-0-0.dta` nach dem Entpacken entsprechend `C:/mydata/haspad/2_source/ZA4586_v1-0-0.dta/ZA4586_v1-0-0.dta`. Zu Fehlermeldungen würde folgender Speicherort führen: `C:/mydata/haspad/ZA4586_v1-0-0.dta`.

Abbildung 4 informiert über die haspad-Ordnerstruktur nach Entpacken der Ursprungsdatensätze.

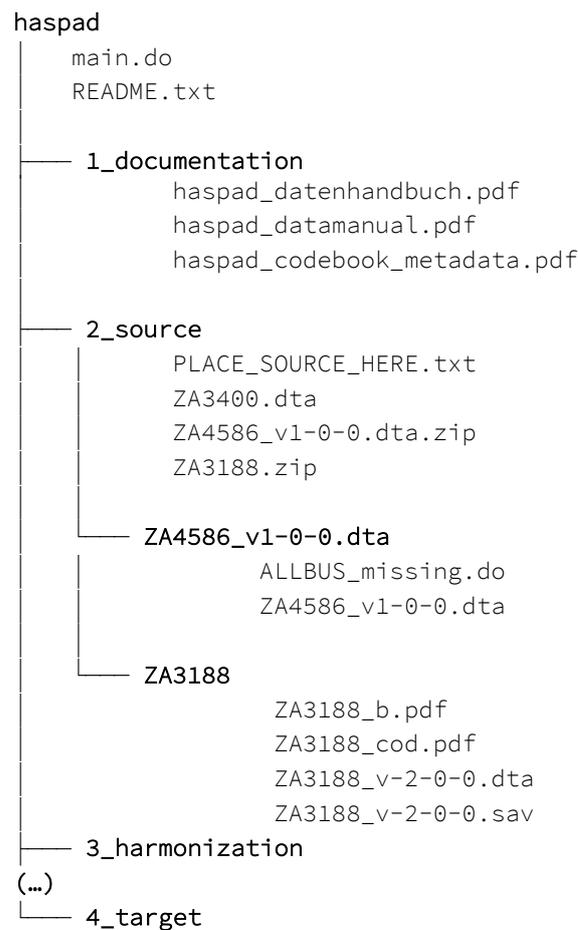


Abbildung 4: Ordnerstruktur nach Entpacken der Source-Datensätze (Ausschnitt)

Notwendige Anpassungen des Main-Files main.do

Nachdem die Auswahl der Variablen mit Hilfe des Harmonization Wizards getroffen wurde, der HaSpaD-Download erfolgte, die Ursprungsdaten beschafft, korrekt abgespeichert und entpackt wurden, wird das Main-File `main.do` geöffnet und angepasst. Zeile 40, die den Pfad des Ordners angibt, in welchem `haspad` entpackt wurde, muss an den Speicherort des entzippten HaSpaD-Ordners angepasst werden. Liegt der entpackte Ordner `haspad` beispielsweise in `C:/mydata`, dann wird in Zeile 40 angegeben:

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

3.3 Während des Ausführens des Main-Files main.do

Das Ausführen des Main-Files `main.do` dauert je nach Rechnerleistung, Anzahl der gewählten Variablen und Umfrageprogrammen zwischen ca. 30 Sekunden (z.B. wenn der Nutzende nur ein Umfrageprogramm ohne Zusatzvariablen auswählt) und ca. 15 Minuten (wenn der Nutzende alle Umfrageprogramme und alle Zusatzvariablen auswählt). Es werden alle nötigen Packages geladen und installiert, die Aufbereitungs-Do-files aufgerufen und ausgeführt, die Datensätze kumuliert und gelabelt und die zugehörige Dokumentation und der Zieldatensatz gespeichert. Für interessierte Nutzende sind in *Anhang A. 2* die einzelnen Schritte der Datenaufbereitung ausführlicher beschrieben.

3.4 Nach dem Ausführen des Main-Files main.do

Im HaSpaD-Ordner befinden sich nun mehrere neue Dateien und Ordner (siehe *Abbildung 5*):

- die Dokumentationen der Datenaufbereitung im HTML-Format unter `haspad/1_documentation/harmonization_targets/` bzw. `haspad/1_documentation/preparation_couplebio/`.
- der analysefähige HaSpaD-Datensatz `target.dta` unter `haspad/4_target/`.

Dabei dienen die HTML-Files dem Nachvollziehen der Aufbereitung der Paarbiografien (`preparation_couplebio`) und der harmonisierten Zusatzvariablen (`harmonization_targets`). Der Zieldatensatz `target.dta` besteht im ausgewählten Minimalbeispiel insgesamt aus 33 Variablen: 31 Variablen des Paarbiografiedatensatzes und zwei Zusatzvariablen Geschlecht (`sex`) und Erhebungsgebiet (`eastwest`). Zusätzlich können – nach Anpassen der Stata Versionsnummer und des Dateipfades – in Do-File `metadata.do` unter `haspad/3_harmonization/dofiles` verschiedene Metadaten auf Studieneben (Stichprobe, Erhebungsdesign, Erhebungsjahre etc.) an den vorher erstellten HaSpaD-Zieldatensatz herangespielt werden.

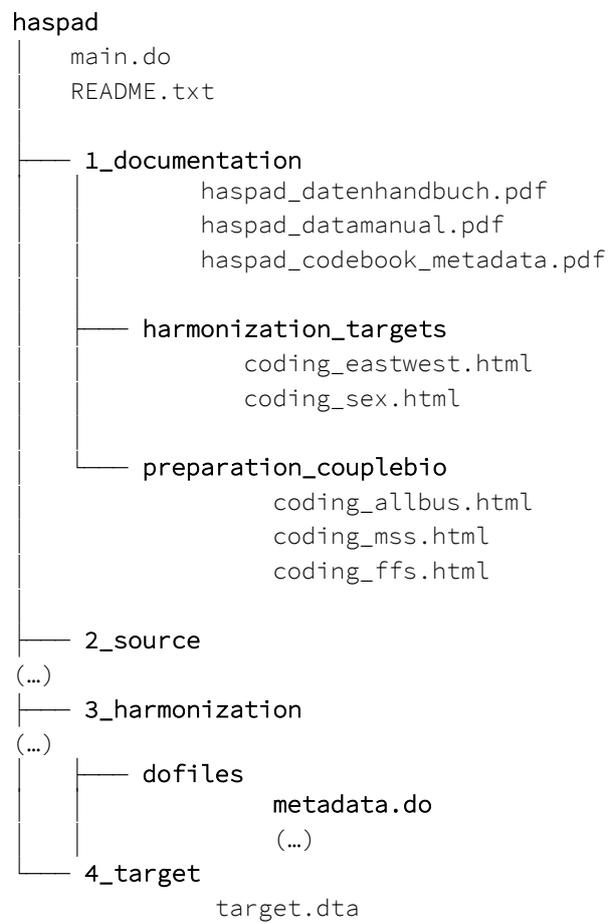


Abbildung 5: Ordnerstruktur nach Ausführung des Main-Files (Ausschnitt)

4 Datenstruktur

Der Zieldatensatz des HaSpaD-Projektes besteht aus harmonisierten Biografiedaten und zusätzlichen harmonisierten Variablen. Der Datensatz liegt im long-Format vor. Das heißt, dass jede Zeile die Information einer Partnerschaft enthält. Individuen sind über die Anchor-ID *aid* identifizierbar. Befragte aus den Ursprungsdaten, die keine Partnerschaftsbiografien berichten, sind nicht Bestandteil der HaSpaD-Daten. Der Biografiedatensatz enthält insgesamt 31 Variablen, unterteilt in drei Bereiche: ID-Variablen, Biografie-Variablen und Flag-Variablen (siehe Tabelle 5).

4.1 Identifikation von Ursprungsstudien und -datensätzen im HaSpaD-Datensatz (Datensatz-ID)

Die im HaSpaD-Projekt verwendeten Umfrageprogramme setzen sich häufig aus verschiedenen Teilstudien zusammen. In HaSpaD wurden Studien- und Teilstudien-IDs vergeben, die Aufschluss darüber geben, aus welcher Ursprungsstudie und Teilstudie eine Partnerschaftsbiografie stammt. Die Vergabe der Teilstudien-ID orientiert sich daran, aus welcher Ursprungsstudie eine Partnerschaft stammt. Die Teilstudien-ID wird daher im Folgenden auch als *Datensatz-ID* bezeichnet. Werden Partnerschaften in der gleichen Ursprungsstudie in verschiedenen Dateien abgelegt (zum Beispiel, weil ein anderer Fragebogen zugrunde lag), bekommen die Ursprungsdaten verschiedene Teilstudien-IDs bzw. Datensatz-IDs (auch dann, wenn sie beispielsweise mit der gleichen DOI zitiert werden; dies ist beispielsweise bei der Lebensverlaufsstudie der Fall). Der Datensatz-ID der Lebensverlaufsstudie wurde teilweise zusätzlich eine „0“ angehängt (z.B. 40 für die LV-DDR oder 50 für die LV-Ost 71), da sich bei der Erstellung der Personenidentifikationsnummer (Anker-ID) ansonsten Überschneidungen im HaSpaD-Zieldatensatz ergeben hätten. Dies wird im Abschnitt zur Erstellung der Anker-ID (*aid*) näher erläutert. *Tabelle 4* dient zur Veranschaulichung der Systematik der Studien- und Datensatz-IDs.

Bereits an dieser Stelle sollte darauf hingewiesen werden, dass bei Partnerschaften, die im Längsschnitt aktualisiert wurden (und den Partnerschaften somit mehrere Ursprungsdatensätze zugrunde liegen), die Datensatz-ID durch Verkettung der Datensatz-ID der Ursprungsdatenquellen vergeben wurde. Wurde beispielsweise eine Partnerschaft, die erstmalig in der Lebensverlaufsstudie LV-Ost 71 (*did* = 50) berichtet wurde, mit Angaben aus der Lebensverlaufsstudie LV-Panel 71 (*did* = 80) aktualisiert, wurde der Partnerschaft die Datensatz-ID „5080“ vergeben. D.h., die Datensatz-IDs der Lebensverlaufsstudie LV-Ost 71 („50“) und diejenige des LV-Panel 71 („80“) wurden zu „5080“ verkettet.

Tabelle 4: Bezeichnungen der Umfrageprogramme im HaSpaD-Zieldatensatz

Studiennr. / Datensatz-ID / Akronym	Titel der Studie	Name des Datensatzes zur Partnerbiografie, Mit Partnerbiografiedaten- satz verknüpfte Datensätze	Anzahl an Part- nerschaften
101 / 1 / pairfam	pairfam Data Release 12.0	biopart	36.994
102 / 1 / all- bus	Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) – Ku- mulation 1980 – 2016	ZA4586_v1-0-0	21.536

Studiennr. / Datensatz-ID / Akronym	Titel der Studie	Name des Datensatzes zur Partnerbiografie, Mit Partnerbiografiedatensatz verknüpfte Datensätze	Anzahl an Partnerschaften
103 / 11 / fs	Wandel und Entwicklung familialer Lebensformen in Westdeutschland (Familiensurvey)	ZA2245_a	6.781
103 / 12 / fs	Familie und Partnerbeziehungen in Ostdeutschland (Familiensurvey)	ZA2392_Partner	2.584
103 / 2 / fs	Wandel und Entwicklung familialer Lebensformen - 2. Welle (Familiensurvey)	ZA2860_PartnerBiographie_NoMissing	11.838
103 / 3 / fs	Wandel und Entwicklung familialer Lebensformen - 3. Welle (Familiensurvey)	ZA3920_Partnerbiographie	12.879
104 / 1 / mss	Mannheimer Scheidungsstudie 1996	ZA3188_v2-0-0	5.020
106 / 1 / ffs	Deutscher Fertility and Family Survey 1992	ZA3400	7.905
107 / 1 / glhs	Lebensverläufe und gesellschaftlicher Wandel: Lebensverläufe und Wohlfahrtsentwicklung (Lebensverlaufsstudie LV-West I)	REC1	2.129
107 / 21 / glhs	Lebensverläufe und gesellschaftlicher Wandel: Die Zwischenkriegskohorte im Übergang zum Ruhestand (Lebensverlaufsstudie LV-West II A - Persönliche Befragung)	FP, PA	463
107 / 22 / glhs	Lebensverläufe und gesellschaftlicher Wandel: Die Zwischenkriegskohorte im Übergang zum Ruhestand (Lebensverlaufsstudie LV-West II T - Telefonische Befragung)	FP, PA	1.128
107 / 3 / glhs	Lebensverläufe und gesellschaftlicher Wandel: Berufszugang in der Beschäftigungskrise (Lebensverlaufsstudie LV-West III)	FP, PA, VB, GT	1.633
107 / 40 / glhs	Lebensverläufe und historischer Wandel in Ostdeutschland (Lebensverlaufsstudie LV-DDR)	PARTNER, PARTQUER, KONTAKT, pa3456p	1.432
107 / 50 / glhs	Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost 71)	fp7p, pa7, public_caseid	796
107 / 61 / glhs	Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost Panel, Geburtskohorte 1939-41, 1951-53, 1959-61); Geburtskohorten 1939-41, 1951-53, 1959-61	fpp456p, pa3456p, fpq456p	1.001 (hiervon 942 im Längsschnitt aktualisierte Partnerschaften)
107 / 62 / glhs	Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost Panel, Geburtskohorte 1929-31); Geburtskohorte 1929-31	fa3p, pa3456p	275 (hiervon 267 im Längsschnitt aktualisierte Partnerschaften)

Studiennr. / Datensatz-ID / Akronym	Titel der Studie	Name des Datensatzes zur Partnerbiografie, Mit Partnerbiografiedatensatz verknüpfte Datensätze	Anzahl an Partnerschaften
107 / 70 / glhs	Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland (Lebensverlaufsstudie LV-West 64/71)	rec16sp , rec1, public_caseid	3.635
107 / 80 / glhs	Frühe Karrieren und Familiengründung: Lebensverläufe der Geburtskohorte 1971 in Ost- und Westdeutschland (Lebensverlaufsstudie LV-Panel 71)	fpl , public_caseid, pa	1.189 (hiervon 395 im Längsschnitt aktualisierte Partnerschaften)
108 / 1 / ggs	Generations and Gender Survey (2005) (Teilstichprobe Deutschland)	GGG_Wave1_Germany_V.4.3 , Germany_Wave2_V.1.3	7.463
108 / 2 / ggs	Generations and Gender Survey (2008) (Teilstichprobe Deutschland)	Germany_Wave2_V.1.3	2.681 (hiervon 2.223 im Längsschnitt aktualisierte Partnerschaften)
109 / 1 / share	The Survey of Health, Ageing and Retirement in Europe (SHARE); Welle 1, 2, 3 (Deutsche Stichprobe)	sharew3_rel6-0-0_rp , sharew3_rel6-0-0_cv_r, sharew3_rel6-0-0_st	2.483
110 / 1 / soep	Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2018	biomarsy , pgen	38.446
		Gesamtanzahl an Partnerschaften	170.291

4.2 ID-Variablen

Tabelle 5: Übersicht über die ID-Variablen

Variablenname	Variablenlabel
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

Insgesamt enthält der Zieldatensatz sieben verschiedene Identifikationsvariablen, die für die korrekte Identifikation, aus welchem Umfrageprogramm (Studien-ID, `sid`) bzw. welcher Teilstudie (Datensatz-ID, `did`) eine Partnerschaft berichtet wird, wichtig sind. Die Anker-ID (`aid`) und die Partner-ID (`pid`) dienen der Verknüpfung von Biografiedaten mit den von den Nutzenden ausgewählten Zusatzvariablen.

Die `aid` ergibt sich aus der Verkettung von Informationen zur Ursprungsstudie (`sid` + `did`) und den Identifikationsmerkmalen innerhalb der Ursprungsstudie (`id`).

In ähnlicher Weise wurde die `pid` konstruiert: Hier wurden Informationen zur Ursprungsstudie (`sid` + `did`), und zur jeweiligen Partnerschaft verkettet. Die Kennzeichnung der jeweiligen Partnerschaft erfolgt bei Ursprungsstudien, die eine Partner-ID bereitstellen (`pairfam`, `SHARE`), mit der originalen Partner-ID (`pid_o`). Die `pid` ergibt sich somit aus der Verkettung von `sid` + `did` + `pid_o`. Bei Studien, in der es kein eigenständiges Identifikationsmerkmal für eine Partnerschaft gibt (z.B. `ALLBUS`, `Lebensverlaufsstudien`), erfolgt die Kennzeichnung einer Partnerschaft über die `aid` (`sid` + `did` + `id`) verkettet mit der Variable `partner_pointer`. Diese Variable ergibt sich aus der Position der Ursprungsinformationen zu einer Partnerschaft im Ursprungsdatensatz und folgt damit in jeder Ursprungsstudie notwendigerweise einer eigenen Logik (z.B. erster, zweiter, dritter Spell; erstes, zweites, drittes Set an Variablen zu Partnerschaften). Nutzende, die weitere Informationen an den Zieldatensatz anspielen möchten als die bereitgestellten Zusatzvariablen, empfehlen wir, sich mit der Systematik der Variable `partner_pointer` anhand der HTML-Dokumentation der Datenaufbereitung und Variablenkonstruktion vertraut machen.

4.3 Biografie-Variablen

Tabelle 6: Übersicht über die Biografie-Variablen

Variablenname	Variablenlabel
<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.

Die Biografie-Variablen enthalten Angaben des Beginns und Endes von Beziehungs-, Kohabitations- und Ehephasen für alle von der Ankerperson berichteten Partnerschaften und werden in Jahrhundertmonaten seit 1900 angegeben (der Wert 1 entspricht Januar 1900, siehe *Tabelle 6*). Das Interviewdatum der Letztbefragung der Ankerperson kann zur Identifikation von Rechtszensierungen von Partnerschaftsdauern genutzt werden. Falls die Person, die mit der Ankerperson in einer Partnerschaft lebte, verstorben ist, enthält die Variable `dop` das Todesdatum jener verstorbenen Person.

4.4 Flag-Variablen

In den Flag-Variablen sind verschiedene Metadaten über einzelne Biografievariablen abgelegt.

Imputed-Variablen

Tabelle 7: Übersicht über Flag-Variablen mit Informationen zum Imputationsgrad der Biografievariablen

Variablenname	Variablenlabel
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

Die Imputationsvariablen geben an, ob die Biografievariablen auf monatlicher, saisonaler oder jährlicher Basis vorliegen (siehe *Tabelle 7*; siehe *Kapitel 5.6* zur näheren Erläuterung). Dabei besteht kein Unterschied, ob in der jeweiligen Teilstudie beispielsweise nur jahresgenaue Angaben berichtet werden konnten oder die Befragten aufgrund von Erinnerungsschwierigkeiten nur ungenaue Angaben machen konnten.

Original-Variablen

Tabelle 8: Übersicht über Flag-Variablen mit den Originalwerten

Variablenname	Variablenlabel
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

Die Originalvariablen enthalten die aus der Ursprungs-Datenquelle stammenden, nicht harmonisierten Variablenwerte der jeweiligen Biografievariable und machen die Kodierung der Sourcevari-

ablen transparent. Da diese zumeist auf Angaben zum Kalenderjahr (ggf. auch Monat) beruhen, können diese Variablen einen schnellen Überblick über die Plausibilität bestimmter Angaben der Befragten geben (siehe *Tabelle 8*, siehe *Kapitel 6.1.4* zur näheren Erläuterung).

Weitere Flag-Variablen (`flag_beg`, `problem`)

Tabelle 9: Übersicht über zusätzliche Flag-Variablen

Variablenname	Variablenlabel
<code>flag_beg</code>	Beginning partnership biography
<code>problem</code>	Flag used to mark problem cases

In einigen Teilstudien liegen nur Informationen über kohabitierende Partnerschaften vor, in anderen nur Angaben über Ehephasen. Nur in manchen Teilstudien kann eine Partnerschaft ab Beziehungsbeginn und bis zu einer eventuellen Scheidung analysiert werden. In der Variable `flag_beg` wird angegeben, mit welchem Institutionalisierungsgrad eine Partnerschaft in einer Teilstudie beginnt (siehe *Kapitel 6.1.5*). Mögliche Inkonsistenzen in den Partnerschaftsbiografien oder Aufbereitungsschritten werden in der Variablen `problem` markiert (siehe *Tabelle 9* und *Kapitel 6.1.5*).

4.5 Missing-Schema

Im Folgenden wird das Missing-Schema des HaSpaD-Projekts vorgestellt (siehe *Tabelle 10*):

Tabelle 10: Verwendetes Missing-Schema

Wert	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

Der Code „-3 Does not apply“ wird vergeben, falls ein Zustand für eine befragte Person nicht zutrifft (z.B. `cohbeg` = -3; keine Kohabitation; `marbeg` = -3, keine Ehe). Die Variable `relbeg` kann niemals den Wert „-3“ annehmen, da die Untersuchungseinheiten in den HaSpaD-Daten Partnerschaften mit dem Status „Beziehung“ sind.

Mit dem Missing-Code „-7 Incomplete Data“ werden verschiedene Datenprobleme in den Ursprungsdaten zusammengefasst. Dies können beispielsweise fehlende Angaben sein (Item Non-Response wie: „weiß nicht“, „keine Angabe“ oder „Antwort verweigert“), oder auch Korrekturen der Datengeber (z.B. Filterfehler, Datenkorrekturen, technische Probleme).

Der Code „-10 Missing by study design“ wird angewendet, falls eine Information aufgrund des Studiendesigns für eine bestimmte Teilmenge der Gesamtstichprobe (aufgrund der Filterführung) nicht berichtet werden konnte bzw. für alle Teilnehmende nicht abgefragt wurde.

Für die Variablen zum Zeitpunkt des Endes einer Beziehung gibt es zusätzliche Missing-Codes: Falls die Beziehung, die Kohabitation oder die Ehe mit dem Tod der Person, mit der die Ankerperson in einer Partnerschaft lebte, endete ($dop > 0$), wird den anderen Variablen zum Partnerschaftsende ($re\text{lend}$, $cohend$, $marend$) der Wert „-66 Death partner“ zugewiesen.

In einigen Umfrageprogrammen ergibt sich aufgrund von Fragebogenfilterung das Problem, dass zwar bekannt ist, dass entweder eine Ehe aufgrund eines Todesfalls oder einer Scheidung endete (1), oder in anderen Studien ein Zusammenleben (zu einem bestimmten Zeitpunkt) *aufgrund* einer Trennung oder Scheidung endete (2). Im ersten Fall ist das genaue Datum vom Tod des Partnes oder der Scheidung, im zweiten Fall das genaue Datum des Trennungs- beziehungsweise Scheidungsdatums aber unbekannt. Im ersten Fall wird die Ausprägung „-88 Union dissolution/death of partner“ beim Scheidungs- und Todesdatum ($marend$, dop) der Person vergeben, mit der die Ankerperson in einer Partnerschaft lebte. Im zweiten Fall wird das Datum des Auseinanderzuges im Kohabitationsende ($cohend$) vergeben und das Beziehungsende ($re\text{lend}$) bzw. das Eheende ($marend$) mit „-77 Separation/divorce, no exact date“ kodiert.

Der Wert „-99 Ongoing“ zeigt an, dass eine Partnerschaftsphase nicht beendet ist und bis zum (letzten) Interviewzeitpunkt der Befragten andauert (Rechtszensur).

5 Besonderheiten der HaSpaD-Daten durch die Verknüpfung der verschiedenen Datenquellen

Besonderheiten bei der Harmonisierung paarbiographischer Daten ergeben sich zunächst darin, dass in den einzelnen Teilstudien Partnerschaften unterschiedlich definiert und gemessen werden (z.B. Partnerschaften einer bestimmten Mindestdauer, eines bestimmten Institutionalierungsgrades, oder ab einem bestimmten Lebensalter). Des Weiteren gab es abweichende Beschränkungen der maximal berichtbaren Anzahl von Partnerschaften. Darüber hinaus unterscheiden sich Beobachtungszeitraum der Biografien zwischen den einzelnen Umfrageprogrammen und ihren Teilstudien, so dass Beginn und Ende einer Partnerschaft über verschiedene Variablen abgebildet werden (das Ende einer Beziehung muss – je nach Teilstudie – aus der Scheidung, dem Auseinanderziehen oder der Trennung abgeleitet werden). Weiterhin unterscheiden sich die Ursprungsdaten hinsichtlich der Erhebungsgebiete und der Altersstruktur der Stichprobe sowie der Frage, ob nur Deutsche oder auch die nichtdeutsche Bevölkerung befragt wurden. Dieses Kapitel erläutert und diskutiert diese Herausforderungen der Harmonisierung mit möglichen Konsequenzen für die Nutzung. Methodische Konsequenzen der Gewichtung der Daten sowie die Imputation von Monatsangaben, falls für den Monat keine oder nur saisonale Informationen vorliegen, werden ebenso erläutert. Abschließend wird die längsschnittliche Aufbereitung von Panelbefragungen ohne Vorliegen einer Partner-ID der Lebensverlaufsstudien und des Generations and Gender Surveys diskutiert.

5.1 Variierende Definitionen des Partnerschaftsbegriffs

Die Definition einer Partnerschaft variiert zwischen den einzelnen Umfrageprogrammen hinsichtlich des Institutionalierungsgrades, sowie der Dauer und dem Lebensalter, ab der die Befragten von Partnerschaften berichten sollen.⁷ In *Tabelle 11* wird der vorliegende Institutionalierungsgrad und weitere Einschränkungen nach Teilstudie dargestellt. Beispielsweise können im pairfam Paarbeziehungen ohne gemeinsamen Haushalt („living apart together-Partnerschaften“ (LAT)), Nichteheliche Lebensgemeinschaften (NEL) und Ehen berichtet werden. Während bei aktuell berichteten Partnerschaften keine Einschränkungen bestehen, sollen bei retrospektiv erhobenen Partnerschaften alle Beziehungen ab dem 14. Lebensjahr aufgenommen werden, die der in *Tabelle 11* erläuterten Definition entsprechen. Im Kontrast dazu steht beispielsweise der ALLBUS, in dem lediglich paarbiografischen Daten zu Ehen der Befragten erhoben wurden.

Tabelle 11: Institutionalierungsgrade nach Studie

Teilstudie	Institutionalisierungsgrad	Weitere Einschränkungen
pairfam	LAT, NEL, Ehen	<p><i>Aktuelle Partnerschaft:</i> Es gibt keine weiteren Einschränkungen.</p> <p><i>Retrospektiv erhobene Partnerschaften:</i> Es werden alle Beziehungen ab dem 14. Lebensjahr erfasst, die entweder länger als 6 Monate dauerten, oder bei denen die Ankerperson mit einem Partner zusammenlebte, oder aus denen ein Kind hervorging</p>

⁷ Die Verweise auf die Dokumentationen in Form von Fragebögen, Codebücher und Dokumentationshandbücher befinden sich in *Anhang A. 3*.

Teilstudie	Institutionalisierungsgrad	Weitere Einschränkungen
ALLBUS-Kumulation 1980-2016	Ehen	oder die aus anderen Gründen für die Ankerperson wichtig war (pairfam Group 2021, S. 11). <i>Aktuelle und retrospektiv erhobene Partnerschaften:</i> Es gibt keine weiteren Einschränkungen (Baumann & Schulz 2018, S. 1629ff.).
Familiensurvey I. Welle West 1988 Familiensurvey I. Welle Ost 1990 Familiensurvey II. Welle 1994	LAT*, NEL*, Ehen	<i>Aktuelle Partnerschaft:</i> Es gibt keine weiteren Einschränkungen. Die aktuelle feste Partnerschaft oder Ehe kann weniger als ein Jahr andauern. (Deutsches Jugendinstitut 1992b, S. 7; 2018d, S. 8; 2018e, S.11) <i>Retrospektiv erhobene Partnerschaften:</i> Es werden vergangene Partnerschaften mit einer Mindestdauer von einem Jahr erfasst und „[...] mehr waren als vorübergehende Bekanntschaften oder Schwärmereien“. Ehen sind auf jeden Fall gemeint. (Deutsches Jugendinstitut 1992b, S. 7, 2018d, S. 8, 2018e, S.11)
Familiensurvey III. Welle 2000	LAT*, NEL*, Ehen*	<i>Aktuelle Partnerschaft:</i> Es gibt keine weiteren Einschränkungen. Die aktuelle feste Partnerschaft oder Ehe kann weniger als ein Jahr andauern. (Deutsches Jugendinstitut 2000, S. 59) <i>Retrospektiv erhobene Partnerschaften:</i> Es werden vergangene Partnerschaften mit einer Mindestdauer von einem Jahr erfasst. (Deutsches Jugendinstitut 2000, S. 59)
Mannheimer Scheidungsstudie	Ehen*	<i>Aktuelle und retrospektiv erhobene Partnerschaften:</i> Es wird nur die erste Ehe des Ankers erfasst (Esser, Gostomski & Hartmann 2018c, S. 2).
Deutscher Fertility and Family Survey	NEL, Ehen	<i>Aktuelle und retrospektiv erhobene Partnerschaften:</i> Es gibt keine weiteren Einschränkungen (Bundesinstitut für Bevölkerungsforschung 2002b, S. 7).
LV-West I LV-West II A - Persönliche Befragung LV-West II T - Telefonische Befragung LV-West III	NEL*, Ehen	<i>Aktuelle Partnerschaft:</i> Es kann eine Ehe oder eine NEL berichtet werden. Eine NEL kann nur berichtet werden, insofern Befragte zum Befragungszeitpunkt in einer Partnerschaft unverheiratet zusammenwohnen. (Mayer 2018d, S. 17, 39, 2018e, S. 53; 1995c, S. 53, 2018f, S. 98, 206f.) <i>Retrospektiv erhobene Partnerschaften:</i> Es werden Ehen, aber keine NELs erfasst (Mayer 2018d, S. 17; 2018e, S.53, 1995c, S.53 2018f, S. 206f.).
LV-DDR	NEL, Ehen	Es gibt keine weiteren Einschränkungen (Mayer 1995d, S. 77).
LV-Ost 71	LAT*, NEL*, Ehen*	<i>Aktuelle Partnerschaft und retrospektiv erhobene Partnerschaften:</i> Alle Partnerschaftstypen werden ab dem „[...] Zeitraum seit [...] [dem] 16. Geburtstag bis heute“ erfasst (Mayer 2004c, S. 209).
LV-Ost Panel, Geburtskohorten 1939-41, 1951-53, 1959-61,	NEL, Ehen	<i>Aktuelle und retrospektiv erhobene Partnerschaften:</i> Es gibt keine weiteren Einschränkungen (Mayer 2004d, S. 115).

Teilstudie	Institutionalisierungsgrad	Weitere Einschränkungen
LV-Ost Panel; Geburtskohorte 1929-31		
LV-West 64/71	LAT*, NEL*, Ehen	<i>Aktuelle Partnerschaft und retrospektiv erhobene Partnerschaften:</i> Alle Partnerschaften werden ab einer Mindestdauer von einem Jahr erhoben (Mayer & Kleinhenz 2004b, S. 105).
LV-Panel 71	LAT*, NEL*, Ehen	<i>Aktuelle Partnerschaft:</i> Es gibt keine weiteren Einschränkungen. <i>Retrospektiv erhobene Partnerschaften:</i> Alle Partnerschaften werden ab einer Mindestdauer von einem Jahr erfasst (Mayer 2014b, S. 92).
Generations and Gender Survey (2005)	LAT*, NEL, Ehen	<i>Aktuelle Partnerschaft:</i> Neben NELs und Ehen können auch LAT-Partnerschaften berichtet werden (Ruckdeschel et al. 2006, S. 51; Leven 2009, S. 26).
Generations and Gender Survey (2008)		<i>Retrospektiv erhobene Partnerschaften:</i> Es können NELs und Ehen, aber keine LAT-Partnerschaften berichtet werden (Ruckdeschel et al. 2006, S. 64; Leven 2009, S. 22ff.).
SHARE	LAT*, NEL, Ehen	<i>Aktuelle Partnerschaft und retrospektiv erhobene Partnerschaften:</i> LAT-Partnerschaften werden definiert als „langfristige Beziehung“ (SHARE-ERIC 2019, S. 24).
Sozio-oekonomisches Panel (SOEP)	Ehen*	<i>Aktuelle Partnerschaft:</i> Ehen werden erst ab dem 16. Lebensjahr dokumentiert (Hamjediers et al 2020) <i>Retrospektiv erhobene Partnerschaften:</i> Bis einschließlich 2010 wurden im Biografiefragebogen des SOEPS nur Ehen retrospektiv erfasst Ehen werden erst ab dem 16. Lebensjahr dokumentiert (Hamjediers et al 2020).

Aus den unterschiedlichen Partnerschaftsdefinitionen folgt, dass Partnerschaften, die retrospektiv erhoben wurden, per Studiendesign eine höhere Dauer aufweisen können, als aktuell berichtete. Zudem muss berücksichtigt werden, dass durch das prospektive Erheben von LAT-Beziehungen, aber Einschränkungen der Surveys beim retrospektiven Berichten der Beziehungen, das Auftreten dieser Beziehungsform in der Lebensphase vor dem Interview und vermutlich auch historisch in den harmonisierten HaSpaD-Daten unterrepräsentiert ist.

Durch die erläuterten Unterschiede der Definition einer Partnerschaft ergeben sich unterschiedliche Analysepotentiale hinsichtlich der Partnerschaftstypen aus den harmonisierten Umfrageprogrammen bzw. Teilstudien. Diese Analysepotentiale auf Partnerschaftstypenebene werden im Folgenden skizziert.

Datenquellen, die Informationen zu Living apart together (LAT) - Partnerschaften enthalten

LAT-Partnerschaften werden in zwölf Teilstudien erhoben. Dabei können Partnerschaftsbiografien, in denen die Befragten nicht mit ihren Partnerinnen oder Partnern zusammenwohnen, in den Umfrageprogrammen pairfam, den Familiensurveys, in den Lebensverlaufsstudien LV-Ost 71, LV-West 64/71 und LV-Panel 71, dem Generations and Gender Survey und dem SHARE⁸.

Dabei ist zu beachten, dass die Befragten aus dem Generations and Gender Survey lediglich die aktuellen Personen, mit denen sie in einer Partnerschaft leben, als LAT-Partnerschaft angeben können und somit keine Analysen über vergangene Partnerschaften dieses Typs möglich sind.

Datenquellen, die nichteheliche Lebensgemeinschaften (NEL) beinhalten

In 19 Studien können nichteheliche Lebensgemeinschaften untersucht werden. Anzumerken ist, dass in der LV-West I, den beiden LV-West II und der LV-West III lediglich aktuelle NEL-Partnerschaften berichtet werden konnten, die bis zum Interviewzeitpunkt andauern. Ausnahmen bilden die ALLBUS-Kumulation, die Mannheimer Scheidungsstudie und teilweise das Sozio-oekonomische Panel, welche lediglich Ehebiografien beinhalten.

Datenquellen, die Ehen beinhalten

Die Minimalanforderung an die harmonisierten Umfrageprogramme im HaSpaD-Projekt sind Biografie-Angaben zur Ehe. Damit können in allen Umfrageprogrammen Ehen als Partnerschaftstyp mit höchstem Institutionalierungsgrad identifiziert werden. Jedoch beinhalten nicht alle harmonisierten Umfrageprogramme vollständige Ehebiografien der Befragungspersonen oder die Befragungspersonen sind noch nicht in eine Ehe übergegangen. Lückenhafte Ehebiografien können durch Beschränkungen der berichtbaren Partnerschaften per Studiendesign entstehen und werden im folgenden Abschnitt erklärt (siehe *Kapitel 5.2*).

5.2 Unterschiede maximal berichtbarer Partnerschaften

Bei der ALLBUS-Kumulation, der Mannheimer Scheidungsstudie, Teilstudien der Familiensurveys und der Lebensverlaufsstudien und dem Sozio-oekonomischen Panel ist die Anzahl der zu berichteten Partnerschaften begrenzt (vgl. *Tabelle 12*). Während die Mannheimer Scheidungsstudie nur Biografiedaten zur ersten Ehe erhebt, können in der ALLBUS-Kumulation maximal vier Ehen angegeben werden. Auch die Familiensurveys I. Welle West 1988, I. Welle Ost 1990 und II. Welle 1994 beschränken sich auf vier Partnerschaften: Sind Befragte mehr als drei Partnerschaften im bisherigen Lebensverlauf eingegangen, so ist die vierte berichtbare Partnerschaft die aktuelle; d.h. Partnerschaften, die zwischen der dritten und der aktuellen liegen, können nicht berichtet werden. Unterschiede gibt es ebenso innerhalb der Lebensverlaufsstudien: In der LV-West I können bis zu vier frühere Ehen und eine aktuelle Ehe berichtet werden. Zusätzlich können Befragte einer noch andauernden Ehe eine aktuelle Partnerschaft angeben, wenn die aktuelle Ehe getrennt, aber nicht geschieden ist und Befragte in einer neuen Partnerschaft zusammenwohnen. In der LV-West III beschränkt sich die Maximalangabe der Partnerschaften auf vier: Es können zwei frühere, abgeschlossene Ehen und eine aktuelle Ehe angegeben werden. Darüber hinaus ist es möglich eine Partnerschaft anzugeben, inso-

⁸ Da der Datensatz *biomarsy* lediglich Ehebiografien beinhaltet, können für Partnerschaften aus dieser Datenquelle keine Angaben zu LAT-Beziehungen gemacht werden.

fern die befragte Person noch verheiratet ist, aber mit einer anderen Person in einer neuen Partnerschaft lebt und mit ihr zusammenwohnt. Im Sozio-oekonomischen Panel können, je nach Erhebungsjahr, maximal drei bzw. vier Partnerschaften berichtet werden (Hamjediers et al 2020).

Tabelle 12: Anzahl maximal berichtbarer Partnerschaften nach Studie

Teilstudie	Anzahl maximal berichtbarer Partnerschaften	Weitere Einschränkungen
pairfam	Es gibt keine Begrenzung (Brüderl et al. 2020b, S. 62).	-
ALLBUS-Kumulation	Es können maximal vier Ehen berichtet werden (Baumann & Schulz 2018, S. 1629ff.).	-
Familiensurvey I. Welle West 1988; Familiensurvey I. Welle Ost 1990; Familiensurvey II. Welle 1994	Es können maximal vier Partnerschaften berichtet werden.	Wenn die Befragten mehr als drei Partnerschaften haben, ist die vierte berichtete Partnerschaft die letzte (Deutsches Jugendinstitut 1992b, S. 9; 2018d, S. 10; 2018e, S. 13).
Familiensurvey III. Welle 2000	Es gibt keine Begrenzung. (Deutsches Jugendinstitut 2000, S. 59ff.).	-
Mannheimer Scheidungsstudie	Es wird nur die erste Ehe der Befragten berichtet (Esser, Gostomski & Hartmann 2018c, S. 2).	-
Deutscher Fertility and Family Survey	Es gibt keine Begrenzung (Bundesinstitut für Bevölkerungsforschung 2002b, S. 8).	-
LV-West I	Es können maximal sechs Partnerschaften berichtet werden.	Die Befragten können vier frühere, abgeschlossene Ehen (Mayer 2018d, S. 41) und eine aktuell geführte Ehe (Mayer 2018d, S. 40) berichten. Sofern die Befragten den Familienstand „2 verheiratet, lebe von meinem Ehemann (meiner Ehefrau) getrennt und wohne mit meinem neuen Partner (meiner neuen Partnerin) zusammen“ (Mayer 2018d, S. 39) angeben, können sie zusätzlich zur aktuellen Ehe und den vier früheren Ehen noch eine aktuell geführte NEL-Partnerschaft (Mayer 2018d, S. 46) angeben.
LV-West II A - Persönliche Befragung; LV-West II T - Telefonische Befragung, LV-DDR; LV-Ost 71, LV-Ost	Es gibt keine Begrenzung (Mayer 2018e, S. 53; Mayer 1995c, S. 53; Mayer 1995d, S. 77; Mayer 2004c,	-

Teilstudie	Anzahl maximal berichtbarer Partnerschaften	Weitere Einschränkungen
Panel, LV-West 64/71, LV-Panel 71	S. 209; Mayer 2004d, S. 115; Mayer & Kleinhenz 2004b, S. 105; Mayer 2014b, S. 92).	
LV-West III	Es können maximal vier Partnerschaften berichtet werden.	Die Befragten können zwei frühere, nämlich die ersten beiden abgeschlossenen Ehen (Mayer 2018f, S. 99) und eine aktuell geführte Ehe (Mayer 2018f, S. 100) berichten. Sofern die Befragten „2 von Ihrem Ehepartner getrennt, aber mit anderem Partner zusammen“ (Mayer 2018f, S. 98) angeben, können sie zusätzlich zur aktuellen Ehe und den zwei früheren ersten Ehen noch eine aktuell geführte NEL-Partnerschaft (Mayer 2018f, S. 106) angeben.
Generations and Gender Survey (2005); Generations and Gender Survey (2008)	Es gibt keine Begrenzung (Ruckdeschel et al. 2006, S. 50ff., Leven 2009, S. 22ff.).	-
SHARE	Es gibt keine Begrenzung (SHARE-ERIC 2019, S. 12ff.)	-
Sozio-oekonomisches Panel (SOEP)	Retrospektiv können, je nach Erhebungsjahr, maximal drei bzw. vier Partnerschaften berichtet werden.	Bis einschließlich 2010 können im Biografiefragebogen retrospektiv maximal drei Ehen berichtet werden. Ab 2011 können im Biografiefragebogen retrospektiv bis zu drei Ehen oder eingetragene Lebenspartnerschaften (Hamjedi et al 2020).

Folglich muss beim Vergleich von Partnerschaften höherer Ordnung oder Ehesequenzen im HaSpaD-Zieldatensatz berücksichtigt werden, dass die Partnerschaftsbiografien teilweise per Studiendesign Lücken aufweisen. Diese Lücken sind nicht zu verwechseln mit Single-Episoden. Geringe Fallzahlen von Partnerschaften oder Ehen höherer Ordnung sind demnach teilweise auch den Studiendesigns geschuldet.

5.3 Variierender Beobachtungszeitraum von Partnerschaften

Weiter oben wurde beschrieben, dass sich die Ursprungsstudien dahingehend unterscheiden, was als relevante beziehungsweise berichtenswerte Partnerschaft eingestuft wird (Definition der Partnerschaften). Hieraus ergeben sich unmittelbar Konsequenzen für den Beobachtungszeitraum der Partnerschaften, das heißt, ab wann Nutzende den Beginn einer Partnerschaft analysieren können (Beziehungsbeginn, Zusammenzug, Heirat) und wann der Beobachtungszeitraum endet (mit Trennung, Auseinanderzug, Scheidung).

Beginn des Beobachtungszeitraums

In allen Umfrageprogrammen bzw. Teilstudien werden mindestens Ehebiografien erhoben (vgl. *Tabelle 13*). Das heißt, dass für die Variable Ehebeginn (*marbeg*) gültige Werte bei allen Studien vorhanden sind. Zu erwähnen ist, dass bei Studien, die LAT- und NEL-Partnerschaften einbeziehen, ein Übergang zu einer Ehe (noch) nicht stattgefunden haben muss und somit für diese Partnerschaften kein Ehebeginn vorliegt. Sowohl in der ALLBUS-Kumulation als auch in der Datenquelle *biomarsy* des Sozio-oekonomisches Panel liegen ausschließlich Ehedaten vor, d.h. es fehlen Angaben zum Beziehungs- und Kohabitationsbeginn.

In allen weiteren Studien sind – zumindest teilweise – Informationen über den Kohabitationsbeginn verfügbar. Besonders ist, dass in der LV-West I, den beiden LV-West II und der LV-West III der Kohabitationsbeginn nur in zum Befragungszeitpunkt bestehenden NEL-Partnerschaften berichtet werden konnte. Zusätzlich dazu kann der Kohabitationsbeginn bei der aktuellen Ehe berichtet werden, wenn die Befragte nur einmal verheiratet war (Mayer 2018f, S. 112) oder vom ersten Ehepartner, falls die Person mehrmals verheiratet war (Mayer 2018f, S. 113).

Besonderheiten des Kohabitationsbeginns aufgrund der Panelverknüpfung ergeben sich in der LV-Ost Panel, Geburtskohorte 1929-31 und im Generations and Gender Survey Welle 2:

In der LV-Ost Panel, Geburtskohorte 1929-31 konnte der Kohabitationsbeginn nicht berichtet werden (Mayer 2004d, S. 135). Bei den Partnerschaften, die bereits in Welle 1 und erneut in Welle 2 berichtet wurden, konnte der Kohabitationsbeginn mit dem Wert aus der Welle 1 aktualisiert werden. Wenn im Generations and Gender Survey (2008) die Befragte angibt, „[...] Anfang 2005 [mit] einem/r (Ehe-)Partner/in“ zusammengelebt zu haben (Leven 2009, S. 22), kann für diese Partnerschaft in Welle 2 kein Kohabitationsbeginn angegeben werden. Für in Welle 2 berichtete Partnerschaften, welche nicht mit Partnerschaften aus der ersten Welle verknüpft werden konnten, konnte der Kohabitationsbeginn nicht aus Welle 1 übernommen werden und ist demnach nicht verfügbar.

Der Beziehungsbeginn ist nur bei einer Teilmenge der Studien für alle Partnerschaften verfügbar (*pairfam*, *Familiensurveys*, *Mannheimer Scheidungsstudie*, *LV-DDR*, *LV-Ost 71*, *LV-West 64/71*, *LV-Panel 71*, *SHARE*). Bei den beiden LV-Ost Panelstudien konnte der Beziehungsbeginn nicht berichtet werden. Jedoch konnte bei den Partnerschaften, die bereits in Welle 1 und erneut in Welle 2 berichtet wurden, der Beziehungsbeginn mit dem Wert aus der Welle 1 aktualisiert werden. Hervorzuheben ist weiterhin, dass beim *Generations and Gender Survey (2005)* und *Generations and Gender Survey (2008)* ein Beziehungsbeginn lediglich bei zum Befragungszeitpunkt bestehenden LAT-Beziehungen vorlag.

Tabelle 13: Beginn der Beobachtung nach Studie

Teilstudie	Beginn der Beobachtung	Besonderheiten/ Einschränkungen
<i>pairfam</i>	<i>relbeg cohbeg marbeg</i>	-
ALLBUS-Kumulation	<i>relbeg¹ cohbeg marbeg</i>	-
Familiensurvey I. Welle Ost 1990; Familiensurvey Welle West 1988; Familiensurvey II. Welle 1994; Familiensurvey III. Welle 2000	<i>relbeg cohbeg marbeg</i>	-

Teilstudie	Beginn der Beobachtung	Besonderheiten/ Einschränkungen
Mannheimer Scheidungsstudie	relbeg cohbeg marbeg	-
Deutscher Fertility and Family Survey	relbeg cohbeg marbeg	-
LV-West I LV-West II A - Persönliche Befragung LV-West II T - Telefonische Befragung	relbeg cohbeg* marbeg	Der Kohabitationsbeginn (cohbeg) kann nur in aktuell geführten NEL-Partnerschaften berichtet werden (Mayer 2018d, S. 39, 41, 46; 2018e, S. 53f.; 2018e, S. 53f.).
LV-West III	relbeg cohbeg* marbeg	Der Kohabitationsbeginn (cohbeg) kann nur in aktuell geführten NEL-Partnerschaften berichtet werden (Mayer 2018f, S. 98, 106). Zusätzlich dazu kann der Kohabitationsbeginn bei der aktuellen Ehe berichtet werden, wenn die Befragte nur einmal verheiratet war (Mayer 2018f, S. 112) oder vom ersten Ehepartner, falls die Person mehrmals verheiratet war (Mayer 2018f, S. 113).
LV-DDR	relbeg cohbeg marbeg	-
LV-Ost 71	relbeg cohbeg marbeg	-
LV-Ost Panel, Geburtskohorten 1939-41, 1951-53, 1959-61	relbeg* cohbeg marbeg	Der Beziehungsbeginn (relbeg) konnte im LV-Ost Panel nicht berichtet werden (Mayer 2004d, S. 115). Bei den Partnerschaften, die bereits in Welle 1 und erneut in Welle 2 berichtet wurden, konnte der Beziehungsbeginn mit dem Wert aus der Welle 1 aktualisiert werden.
LV-Ost Panel, Geburtskohorte 1929-31	relbeg* cohbeg* marbeg	Der Beziehungsbeginn (relbeg) und der Kohabitationsbeginn (cohbeg) konnte im LV-Ost Panel nicht berichtet werden (Mayer 2004d, S. 135). Bei den Partnerschaften, die bereits in Welle 1 und erneut in Welle 2 berichtet wurden, konnte der Beziehungsbeginn und der Kohabitationsbeginn mit dem Wert aus der Welle 1 aktualisiert werden.
LV-West 64/71	relbeg cohbeg marbeg	-
LV-Panel 71	relbeg cohbeg marbeg	-
Generations and Gender Survey (2005)	relbeg* cohbeg marbeg	Der Beziehungsbeginn (relbeg) kann nur in aktuell geführten Partnerschaften berichtet werden, in welchem die Partner nicht zusammenwohnen (Ruckdeschel et al. 2006, S. 51).
Generations and Gender Survey (2008)	relbeg* cohbeg* marbeg*	Der Beziehungsbeginn (relbeg) kann nur in aktuell geführten Partnerschaften berichtet

Teilstudie	Beginn der Beobachtung	Besonderheiten/ Einschränkungen
		werden, in welchem die Partner nicht zusammenwohnen (Leven 2009, S. 26).
		Wenn die Befragte angibt, „[...] Anfang 2005 [mit] einem/r (Ehe-)Partner/in“ zusammengelebt zu haben (Leven 2009, S. 22), kann für diese Partnerschaft in Welle 2 kein Kohabitationsbeginn (<i>cohbeg</i>) angegeben werden. Für Partnerschaften berichtet in Welle 2, welche nicht mit Partnerschaften aus der ersten Welle verknüpft werden konnten, konnte der Kohabitationsbeginn nicht aus Welle 1 übernommen werden.
		Wenn die Befragte angibt, „[...] Anfang 2005 [mit] einem/r (Ehe-)Partner/in“ zusammengelebt zu haben (Leven 2009, S. 22), kann für diese Partnerschaft in Welle 2 kein Ehebeginn (<i>marbeg</i>) angegeben werden. Für Partnerschaften berichtet in Welle 2, welche nicht mit Partnerschaften aus der ersten Welle verknüpft werden konnten, konnte der Ehebeginn nicht aus Welle 1 übernommen werden. Weiterhin ist für aktuelle Partnerschaften neu berichtet in Welle 2, in welchen die Partner nicht zusammenwohnen, kein Ehebeginn verfügbar.
SHARE	<i>relbeg cohbeg marbeg</i>	-
Sozio-oekonomisches Panel (SOEP)	<i>relbeg cohbeg marbeg</i>	Der Beziehungsbeginn (<i>relbeg</i>) und Kohabitationsbeginn (<i>cohbeg</i>) ist in der Datenquelle <i>biomarsy</i> nicht verfügbar.

¹*relbeg* = Beziehungsbeginn, *cohbeg* = Kohabitationsbeginn, *marbeg* = Ehebeginn; wenn vollständig ausgegaut, dann nicht bekannt.

Ende des Beobachtungszeitraums

Grundsätzlich können die Variablen zum Partnerschaftsende nur gültige Werte annehmen, wenn eine Partnerschaft des jeweiligen Institutionalisierungsgrads auch begonnen wurde, beziehungsweise der Beobachtungszeitraum einer (Teil-)Studie diesen Institutionalisierungsgrad umfasst.

Wenn z.B. keine Ehe eingegangen wurde, liegen entsprechend keine Angaben zu *marend* vor. Ist die Partnerschaft bis zum Interviewzeitpunkt andauernd, enthalten die End-Variablen des jeweiligen Partnerschaftstypen einen Fehlwert („-99 Ongoing“, siehe *Kapitel 4.5*). Das heißt, für Living apart together-Partnerschaften ist die Beziehung für NEL-Partnerschaften zusätzlich die Kohabitation und für Ehen das Eheende bis zum Interviewzeitpunkt andauernd. Die Konstruktion andauernder Partnerschaften ist somit abhängig vom vorliegenden Partnerschaftstyp und der Verfügbarkeit der jeweiligen Variablen in der Ausgangsstudie („-10 Missing by study design“ falls Information nicht verfügbar ist). Sofern die Partnerschaft mit dem Tod der Person endet, mit der die befragte Person in eben jener Partnerschaft lebte, wird „-66 Death partner“ bei allen verfügbaren End-Variablen und ein gültiger Wert im Todesdatum vergeben. Eine Ausnahme bilden die Familiensurveys, in welchen nur in Ehen ein Todesdatum verstorbener Partnerinnen bzw. Partner verfügbar ist.

Je nach berichtetem Partnerschaftstyp und Angabe eines gültigen Wertes können Nutzende das Trennungsdatum nun über das Beziehungsende bei LAT-Partnerschaften, das Kohabitationsende bei NEL-Partnerschaften und das Eheende bei Ehen konstruieren.⁹

Für einige Umfrageprogramme liegen aber auch differenzierte Informationen vor: Zum Beispiel sind im pairfam bei einer Scheidung auch Informationen zum Trennungsdatum und zum Auseinanderzug erhoben worden.

Tabelle 14: Ende der Beobachtung nach Studie

Teilstudie	Ende der Beobachtung	Besonderheiten/ Einschränkungen
pairfam	re end cohend marend	-
ALLBUS-Kumulation	re end ¹ cohend marend	-
Familiensurvey I. Welle West 1988	re end* cohend marend	Bei NEL- und LAT-Partnerschaften kann nicht zwischen dem Tod des Partners (dop) und dem Beziehungsende (re end) unterschieden werden, da lediglich gefragt wird, „[...] in welchem Jahr [die Partnerschaft] endete“ (Deutsches Jugendinstitut 1992b, S. 9; 2018e, S. 13; 2000, S. 71ff.). Das Jahr des Partnerschaftsendes wird dem Beziehungsende (re end) zugeordnet. Das Kohabitationsende (cohend) kann nur in Ehen berichtet werden. Dem Todesdatum des Partners (dop) wird “-88 Union dissolution/death of partner” zugewiesen.
Familiensurvey II. Welle 1994		
Familiensurvey III. Welle 2000		
Familiensurvey I. Welle Ost 1990	re end* cohend* marend	Bei NEL- und LAT-Partnerschaften kann nicht zwischen dem Tod des Partners und dem Beziehungsende (re end) unterschieden werden, da lediglich gefragt wird, „[...] in welchem Jahr [die Partnerschaft] endete“ (Deutsches Jugendinstitut 2018d, S. 10). Das Jahr des Partnerschaftsendes wird dem Beziehungsende (re end) zugeordnet. Dem Todesdatum des Partners (dop) wird “-88 Union dissolution/death of partner” zugewiesen. Das Kohabitationsende (cohend) kann nur in Ehen berichtet werden.
Mannheimer Scheidungsstudie	re end* cohend marend	Das Beziehungsende (re end) ist nur bei der Besonderheit der Trennung mit weiterem Zusammenwohnen („Trennung von Tisch und Bett“) verfügbar (Esser, Gostomski & Hartmann 2018b, S. 21).
Deutscher Fertility and Family Survey	re end* cohend marend*	Falls die Partnerschaft getrennt wurde, wird das Beziehungsende (re end) auf „-77 Separation/divorce, no exact date“ gesetzt, da zwar das Datum des Endes des Zusammenwohnens bekannt ist (Bundesinstitut für Bevölkerungsforschung 2002b, S. 8), dies aber nicht mit dem Beziehungsende (re end) gleichzusetzen ist.

⁹ Am Ende bleibt es den Nutzenden überlassen, welches Datum bei Vorliegen mehrere Enddaten für die jeweilige Partnerschaftsphase als Trennungsdatum und damit endgültiges Partnerschaftsende verwendet wird.

Teilstudie	Ende der Beobachtung	Besonderheiten/ Einschränkungen
		Falls die Partnerschaft geschieden wurde, wird das Eheende (marend) auf „-77 Separation/divorce, no exact date“ gesetzt, da zwar das Datum des Endes des Zusammenwohnens bekannt ist (Bundesinstitut für Bevölkerungsforschung 2002b, S. 8), dies aber nicht mit dem Eheende (marend) gleichzusetzen ist.
LV-West I	re end cohend* marend	Das Kohabitationsende (cohend) ist nur für aktuelle Ehen und aktuelle NEL-Partnerschaften verfügbar. Für aktuelle NEL-Partnerschaften ist das Kohabitationsende (cohend) konstant „-99 Ongoing“, da diese mit der Partnerin oder dem Partner zum Interviewzeitpunkt zusammenwohnen (Mayer 2018d, S. 39).
LV-West II A - Persönliche Befragung	re end cohend* marend	Für aktuelle NEL-Partnerschaften ist das Kohabitationsende (cohend) konstant „-99 Ongoing“, da diese mit der Partnerin oder dem Partner momentan zusammenwohnen (Mayer 2018e, S. 54).
LV-West II T - Telefonische Befragung	re end cohend marend	-
LV-West III	re end* cohend* marend	Das Beziehungsende (re end) ist nur bei aktuellen Ehen verfügbar, die getrennt, aber noch nicht geschieden sind (Mayer 2018f, S. 101). Das Kohabitationsende (cohend) wird nur für aktuell getrennte bzw. geschiedene Ehen erhoben (Mayer 2018f: 115). Für aktuelle NEL-Partnerschaften ist das Kohabitationsende (cohend) konstant „-99 Ongoing“, da diese mit der Partnerin oder dem Partner momentan zusammenwohnen (Mayer 2018f, S. 106).
LV-DDR	re end cohend marend	-
LV-Ost 71	re end cohend marend	-
LV-Ost Panel, Geburtskohorten 1939-41, 1951-53, 1959-61	re end* cohend* marend	Das Beziehungsende (re end) ist in dem Ursprungsdatensatz der LV-Ost Panel nicht enthalten. Für Partnerschaften, die in Welle 1 und erneut in Welle 2 berichtet wurden, kann das Beziehungsende, berichtet in Welle 1 übernommen werden, sofern eine Trennung in Welle 1 bereits stattgefunden hat. Diese Partnerschaften wurden in Welle 2 nochmals genannt und werden mit den aus Welle 2 stammenden Daten des Ehe- und Kohabitationsendes aktualisiert. Das Kohabitationsende (cohend) ist für eine Partnerschaft, die in Welle 1 und Welle 2 berichtet wurde, aufgrund eines Filterfehlers nicht verfügbar (Mayer 2004d, S. 114). Obwohl diese Person eine Partnerschaft in Welle 2 angab, wurde die Frage zum Kohabitationsende nicht gestellt.

Teilstudie	Ende der Beobachtung	Besonderheiten/ Einschränkungen
LV-Ost Panel, Geburtskohorte 1929-31	re ¹ lend* cohend* mare ¹ nd	<p>Das Beziehungsende (re¹lend) kann in der zweiten Welle der LV-Ost Panel nicht berichtet werden (Mayer 2004d, S. 135). Für Partnerschaften, die in Welle 1 und erneut in Welle 2 berichtet wurden, kann das Beziehungsende übernommen werden, sofern eine Trennung in Welle 1 bereits stattgefunden hat. Diese Partnerschaften wurden in Welle 2 nochmals genannt und werden mit den aus Welle 2 stammenden Daten des Ehe- und Kohabitationsendes aktualisiert.</p> <p>Das Kohabitationsende (cohend) ist für eine neu in Welle 2 berichtete Partnerschaft nicht verfügbar. Dieser Person wurde aufgrund der Filterführung im aktuellen Ehestand (Mayer 2004d, S. 135) die Frage zum Kohabitationsende nicht gestellt.</p>
LV-West 64/71	re ¹ lend cohend mare ¹ nd	-
LV-Panel 71	re ¹ lend cohend mare ¹ nd	-
Generations and Gender Survey (2005)	re ¹ lend cohend mare ¹ nd	-
Generations and Gender Survey (2008)	re ¹ lend cohend mare ¹ nd*	Für aktuelle Partnerschaften, die neu in Welle 2 berichtet wurden und in welchen die Partner nicht zusammenwohnen, ist kein Eheende (mare ¹ nd) verfügbar.
SHARE	re ¹ lend* cohend mare ¹ nd	Das Beziehungsende (re ¹ lend) kann nur in LAT-Partnerschaften berichtet werden (SHARE-ERIC 2019, S. 24).
Sozio-oekonomisches Panel (SOEP)	re ¹ lend* cohend mare ¹ nd*	<p>Das Beziehungsende (re¹lend) ist in der Datenquelle biomarsy überwiegend nicht verfügbar. Nur bei wenigen Partnerschaften wird in einem Partnerschaftsspell der Zustand „verheiratet, getrennt“ angegeben der explizit das Beziehungsende markiert und datiert (Hamjediers et al 2020).</p> <p>Das Kohabitationsende (cohend) ist in der Datenquelle biomarsy nicht verfügbar (Hamjediers et al 2020).</p> <p>Es gibt verheiratete Partnerschaften, in welchen das Eheende (mare¹nd) mit „88 Union dissolution/death of partner“ codiert ist, da unbekannt ist, ob die Ehe mit der Scheidung oder dem Tod der Partnerin / des Partners geendet ist (Hamjediers et al 2020).</p>

¹re¹lend = Beziehungsende, cohend = Kohabitationsende, mare¹nd = Eheende; wenn vollständig ausgegraut, dann nicht bekannt.

Daraus ergeben sich für einige Studien Besonderheiten hinsichtlich dieser drei End-Variablen und ggf. der Operationalisierung im Fragebogen:

In obiger *Tabelle 14* ist das Ende der Beobachtung nach Studie dargestellt. Das Eheende wird in der LV-Panel 71 nicht und im Generations and Gender Survey (2008) nur teilweise abgefragt. Beim Fertility and Family Survey (Deutschland) und im Sozio-oekonomischen Panel (Datenquelle *biomarsy*) wird das Eheende mit einem besonderen Missing codiert: Im Fertility and Family Survey wird in Frage 227 erfasst, „[...] in welchem Monat und Jahr [...] damals Ihr Zusammenwohnen [endete] (Bundesinstitut für Bevölkerungsforschung 2002b, S. 7)“ und welcher Grund dafür vorlag. Wird als Grund „Scheidung“ angegeben, so liegt beim Eheende „-77 Separation/divorce, no exact date“ vor, da lediglich das Zusammenwohnen endete; das Scheidungsdatum jedoch unklar ist. Möglich ist hier, das Kohabitationsende als Trennungsdatum zu nutzen oder das Ende auf andere Weise zu imputieren. Weiterhin gibt es im Sozio-oekonomischen Panel (Datenquelle *biomarsy*) Partnerschaften, bei denen unklar ist, ob die Ehe mit der Scheidung oder dem Tod der Person endete, mit der die befragte Person verheiratet war („-88 Union dissolution/death of partner“¹⁰) und somit keinerlei Aussagen über das Trennungsdatum getroffen werden können, da sowohl ein Kohabitations- als auch ein Beziehungsende nicht vorliegen. Bei allen anderen Umfrageprogrammen bzw. Studien liegt das Eheende bei Scheidung vor.

Falls das Kohabitationsende als Trennungsdatum genutzt werden sollte, ist darauf zu achten, dass dieses bei den Studien der ALLBUS-Kumulation, des Generations and Gender Surveys (2005, 2008) und in Teilen der Familiensurveys (Familiensurvey I. Welle West 1988, II. Welle 1994, III. Welle 2000), Lebensverlaufsstudien (LV-West II T - Telefonische Befragung, LV-DDR, LV-Ost 71, LV-West 64/71, LV-Panel 71) und des Sozio-oekonomischen Panels (Datenquelle *biomarsy*) nicht verfügbar ist. Im Familiensurvey I. Welle Ost 1990 kann nur in verheirateten Partnerschaften das Kohabitationsende berichtet werden. In der LV-West I ist das Ende des Zusammenwohnens nur für aktuelle Ehen und aktuelle NEL-Partnerschaften verfügbar; in der LV-West III ist das Ende des Zusammenwohnens nur für aktuell getrennte bzw. geschiedene Ehen erhoben. Das Kohabitationsende ist in den beiden Geburtskohorten der LV-Ost Panelstudien für jeweils eine Partnerschaft nicht verfügbar und in der Tabelle genauer beschrieben. In allen hier nicht genannten Studien ist ein Kohabitationsende – falls ein Auseinanderzug stattgefunden hat – verfügbar.

Das Beziehungsende ist sowohl in der ALLBUS-Kumulation und der LV-West I für alle Partnerschaften nicht verfügbar. Aufgrund der Fragestellung für NEL- und LAT-Partnerschaften kann in allen Familiensurveys nicht zwischen Tod des Partners und dem Beziehungsende unterschieden werden. Das Jahr des Partnerschaftsendes wird dort dem Beziehungsende zugeordnet. In der Mannheimer Scheidungsstudie ist bei der Definition der Trennung darauf zu achten, dass das Beziehungsende durch das Datum der „Trennung von Tisch und Bett“ (Esser, Gostomski & Hartmann 2018b, S. 21; Frage 175 und 176) operationalisiert wird. Gemeint ist die Trennung von Tisch und Bett als die Aufgabe der ehelichen Lebensgemeinschaft gemäß §1353 BGB (BGB 2017, S. 315). Nach §1567 BGB muss ein Getrenntleben nicht unbedingt ein Auseinanderzug bedeuten und kann auch „[...] innerhalb der ehelichen Wohnung (BGB 2017, S. 342)“ stattfinden. Da der Auszug aus der gemeinsamen Wohnung und die Trennung von Tisch und Bett getrennt abgefragt werden, kann davon ausgegangen werden, dass mit der Trennung von Tisch und Bett das Beziehungsende gemeint ist, auch wenn beide Ehepartner in der gleichen Wohnung leben. Im Fertility and Family Survey (Deutschland) wird das Beziehungsende als „-77 Separation/divorce, no exact date“ codiert, falls als Grund des Endes des Zusammenwohnens „Trennung“ angegeben wurde (siehe Eheende oben). Weiterhin ist in der LV-West III nur für aktuelle Ehen, die getrennt aber noch nicht geschieden sind, ein gültiger Wert im Beziehungsende. Für die beiden Geburtskohorten der LV-Ost Panel ist das Beziehungsende nicht

¹⁰ Siehe dazu Hamjediers et al 2020.

verfügbar. Jedoch wurde für Partnerschaften, die in Welle 1 und erneut in Welle 2 berichtet wurden, das Beziehungsende berichtet in Welle 1 übernommen, sofern eine Trennung in Welle 1 bereits stattgefunden hat. Diese Partnerschaften wurden in Welle 2 nochmals genannt und werden mit den aus Welle 2 stammenden Daten des Ehe- und Kohabitationsendes aktualisiert. Weiterhin kann beim SHARE das Beziehungsende nur in LAT-Partnerschaften berichtet werden. Ebenso ist beim SOEP das Beziehungsende nur teilweise verfügbar: In der Datenquelle `biomarsy` nur für diejenigen, die „verheiratet, getrennt“ angegeben haben. Bei allen anderen Studien, die hier nicht genannt wurden, liegt bei Trennung ein Beziehungsende vor.

5.4 Unterschiedliche Referenzpopulationen

Die einbezogenen (Teil-)Studien unterscheiden sich hinsichtlich der Grundgesamtheit. Unterschiede bestehen unter anderem im Erhebungsgebiet (Ost- und/oder Westdeutschland), in den einbezogenen Altersgruppen und Kohorten sowie der Frage, ob nur Deutsche oder auch Ausländer in die Befragung einbezogen wurden (vgl. *Tabelle 1* in *Kapitel 2.1*). Ähnlich, wie z.B. bei einer Dual-Frame-Stichprobe bei denen Personen über Mobilfunknummern als auch Festnetznummern erreicht werden können und für diese erhöhte Inklusionswahrscheinlichkeit korrigiert werden muss, hat ein Subset an Personen mit bestimmten Eigenschaften eine positive Inklusionswahrscheinlichkeit für mehrere Surveys. Die Literatur zum design-basierten (Neyman 1934) Umgang mit Multi-Frame-Stichproben, z.B. durch sogenannte Composite Faktoren mit denen Beobachtungen (zusätzlich) gewichtet werden, ist umfangreich, ein Literaturüberblick findet sich zum Beispiel in Xia, Pedlow und Davern (2010). Bei modell-basierter Inferenz (Fisher 1922) ist es stattdessen möglich, Kovariaten (und zugehörige Interaktionen) in das Modell aufzunehmen, um für unterschiedliche Inklusions- und Teilnahmewahrscheinlichkeiten zu kontrollieren (Pfeffermann & Sverchkov 2009). Welche Kontrollvariablen benötigt werden, hängt vom Modell ab, daher können an dieser Stelle keine allgemeingültigen Empfehlungen zur Aufnahme von Kovariaten oder der Inklusion von Gewichten gegeben werden. Das Thema der unterschiedlichen Zielpopulationen ist eng mit dem Thema der Survey-Gewichte der verschiedenen Surveys verwandt, der im nächsten *Abschnitt 5.5* folgt und auch einige Beispiele zum design-basierten bzw. hybriden (design-gewichtete Regressionen) und dem modell-basierten Ansatz anführt.

5.5 Umgang mit Survey-Gewichten

In den meisten Originaldatensätzen sind Survey-Gewichte vorhanden. Diese sind allerdings nicht im HaSpaD-Zieldatensatz enthalten, können aber durch die Datennutzenden über die im HaSpaD-Zieldatensatz enthaltenen Original-IDs dem Datensatz selbstständig zugespült werden.

Im Folgenden sollen einige Überlegungen und Hinweise bzgl. der möglichen Verwendungen von Survey-Gewichten gegeben werden. Diese können aufgrund der komplexen Struktur und der vielfältigen Analysemöglichkeiten nicht erschöpfend sein.

Survey-Gewichte dienen im Allgemeinen dazu, Sampling Bias für Schätzer zu reduzieren und für Non-Coverage oder Non-Response zu korrigieren. In der Regel werden Survey-Gewichte noch einmal in Design-Gewichte, Non-Response-Gewichte und Poststratifikation-Gewichte unterschieden. Während Design-Gewichte zur Korrektur von ungleichen Ziehungswahrscheinlichkeiten erstellt werden, dienen Non-Response-Gewichte zur Korrektur von Non-Response, wenn zumindest einige Daten zu den befragten und den verweigernden Personen verfügbar sind, wie zum Beispiel Beobachtungen des Interviewers zur Wohnumgebung der Personen. Poststratifikation-Gewichte werden dazu benutzt, um die Verteilung des Samples dem der Population in Bezug auf einzelne Variab-

len anzupassen. Dies geschieht unter der Annahme, dass diese Variablen eng mit möglichen Zielvariablen zusammenhängen (Survey Research Center 2016, S. 657). Verschiedene Gewichte können bei Analysen auch kombiniert werden. Die ALLBUS-Kumulation verfügt beispielsweise über drei verschiedene personenbezogene Gewichte, nämlich ein personenbezogenes Ost-West-Gewicht, das die Überrepräsentierung von Personen aus den neuen Bundesländern aufhebt, ein Transformationsgewicht (eine Form des Design-Gewichts) für personenbezogene Aussagen, da der ALLBUS in früheren Erhebungen eine Haushaltsstichprobe war und ein kombiniertes Gewicht, das Produkt aus Transformations- und Ost-West-Gewicht. Das Sozio-oekonomische Panel hingegen hat neben einem Transformationsgewicht außerdem ein Poststratifikation-Gewicht, um die Stichprobe der Population (Personen > 16 Jahre in der Bundesrepublik) in Hinblick auf Altersgruppen, Familienstand, Geschlecht und Nationalität anzupassen.

Generell sind in den von HaSpaD verwendenden Originaldatensätzen Design-Gewichte oder Poststratifikation-Gewichte enthalten, bzw. daraus kombinierte Gewichte. Die Familiensurveys und der Generations and Gender Survey sowie SHARE verfügen sowohl über Design-Gewichte als auch Poststratifikation-Gewichte zumindest für einige Erhebungsjahre.

Die Verwendung von Survey-Gewichten wird in der Statistik intensiv diskutiert (Gelman 2007, Lavallée & Beaumont 2015; für Einführungen siehe Gabler & Häder 2016; Lavallée & Beaumont 2016, Tillé & Matei 2016, Vannette & Krosnick 2018). Survey-Gewichte können auf der einen Seite Coverage, Sampling und Non-Response Bias reduzieren. Auf der anderen Seite können Survey-Gewichte aber auch die Varianz der Schätzer erheblich erhöhen. Deshalb empfiehlt auch das Survey Research Center (2016, S. 659) zumindest Poststratifikations-Gewichte nicht zu verwenden, wenn die Verwendung der Gewichte die Punktschätzer nicht dramatisch verändern. Auch besteht die Möglichkeit, Variablen, die zur Poststratifikation oder zur Design-Gewichtung genutzt wurden, als Kovariaten (gegebenenfalls inklusive Interaktionstermen mit anderen unabhängigen Variablen) in ein Regressionsmodell aufzunehmen, statt die einzelnen Fälle in der Regressionsschätzung zu gewichten (Winship & Radbill 1994).

Ein illustrierendes Beispiel: Wenn sich der Zusammenhang zwischen Partnerschafts- und Lebenszufriedenheit zwischen Ost- und Westdeutschland nicht systematisch unterscheidet und man diese für Gesamtdeutschland schätzen möchte, werden Design-Gewichte wie sie bspw. für die ALLBUS-Kumulation vorliegen, um für eine Überrepräsentierung der ostdeutschen Paare zu korrigieren, nicht benötigt. Die Verwendung der Gewichte würde sich nur auf die geschätzte Varianz auswirken und diese erhöhen. Wenn sich hingegen der Zusammenhang zwischen Bildungsabschluss und Lebenszufriedenheit zwischen Ost- und Westdeutschland unterscheidet und man an einem Schätzer für Gesamtdeutschland interessiert ist, müssen die Design-Gewichte zur Korrektur der Überrepräsentation Ostdeutschlands verwendet werden, da sonst der Punktschätzer verzerrt ist. Die Ost-West-Gewichte des ALLBUS müssen ebenfalls nicht verwendet werden, wenn eine Analyse getrennt nach Ost- und Westdeutschland vorgenommen wird oder, wie oben bereits erwähnt, eine Ost-West-Variable (und gegebenenfalls Interaktionsterme dieser Variable mit anderen unabhängigen Variablen) in die statistischen Modelle aufgenommen wird.

Bei der Analyse eines harmonisierten Datensatzes kann diese generelle Abwägung zwischen möglichem Bias und Varianz aber nur ein erster Schritt sein. Im Vergleich zur Analyse eines einzelnen Datensatzes kommt beim HaSpaD-Datensatz hinzu, dass dieser Daten aus mehreren Jahrzehnten enthält, also damit Populationen zu unterschiedlichen Zeitpunkten miteinander verglichen werden können. Auch in der komparativen Sozialforschung stellen sich zwei Fragen: (1) Wie soll beim Vergleich mehrerer, länderspezifischer Populationen mit in den Datensätzen erhaltenen Gewichten umgegangen werden? Und (2) Sollen alle Observations innerhalb eines Surveys auch noch einen zusätzlichen survey-spezifischen Gewichtungsfaktor erhalten? Zur Problematik der Gewichtung im komparativen Fall ist im Survey Data Harmonization Newsletter ein Überblicksartikel von Joye,

Sapin und Wolf (2019) erschienen. Wir werden einige Empfehlungen auf HaSpaD und andere Harmonisierungsprojekte, die über die Zeit, aber nicht über verschiedene Länder harmonisieren, übertragen.

Zum ersten Punkt, der Verwendung von Survey-Gewichten aus den Originaldatensätzen, empfehlen Joye, Sapin und Wolf (2019) sorgfältig abzuwägen, ob die Verwendung von Gewichten den Vergleich zwischen Populationen erleichtert oder erschwert. Als Beispiel: Die in HaSpaD beinhalteten Surveys unterscheiden sich teilweise in ihrer Grundgesamtheit, etwa, ob sie nur deutsche Staatsangehörige umfassen (wie die Familiensurveys) oder die gesamte Bevölkerung in Privathaushalten über 17 Jahren als Grundgesamtheit betrachten (Sozio-oekonomisches Panel). Wenn nun ausländische Staatsangehörige im Sozio-oekonomischen Panel eine geringere Antwortwahrscheinlichkeit haben und dafür per Poststratifikation-Gewicht höher gewichtet werden können, würde diese Gewichtung den Vergleich zwischen den beiden Surveys erschweren.

Im HaSpaD-Datensatz wird zudem eine Subpopulation (Personen in Partnerschaften) betrachtet. Gewichtete Varianzschätzungen für eine Subpopulation sind nicht trivial und erfordern Informationen über die Anzahl der ausgeschlossenen Einheiten und deren Survey Gewichte (siehe West, Berglund & Heeringa 2008, und Graubard & Korn 1996).

Zum zweiten Punkt bzgl. eines zusätzlichen spezifischen Gewichtes für die einzelnen Surveys verweisen Joye, Sapin und Wolf (2019) auf Kish (1999, Reprint 2003) und schlagen sechs verschiedene Vorgehensweisen vor. Wir diskutieren diese anhand eines Beispiels, das mit HaSpaD-Daten denkbar ist. Wir nehmen an, dass Paare in Deutschland in verschiedenen Jahren oder Jahrzehnten mit Hilfe von Surveydaten aus verschiedenen Surveyprogrammen verglichen werden sollen. Es gibt nun folgende Möglichkeiten:

1. Kein Vergleich zwischen Surveys aus verschiedenen Populationen, da Unterschiede zwischen den Surveys in Bezug auf die Grundgesamtheit als zu gravierend erachtet werden.
2. Harmonisierung der Daten, aber keine Analyse auf dem kombinierten Datensatz. Schätzer für jeden Survey werden dann getrennt berechnet und anschließend verglichen. Dieses Vorgehen hat den Nachteil, dass die Varianzschätzer für kleine Datensätze recht groß werden können. Ein solcher Datensatz ist im Fall des HaSpaD-Projektes beispielsweise die LV-Ost 71.
3. Jeder Datensatz in der Analyse auf Grundlage des kombinierten Datensatzes wird gleich gewichtet. Dies führt dazu, dass Datensätze mit einer hohen Anzahl an Befragten ein größeres Gewicht erhalten als Datensätze mit einer niedrigeren Anzahl an Befragten. Soll beispielsweise der Anteil an gleichgeschlechtlichen Paaren an allen Paaren seit 2000 geschätzt werden und zwei Surveys aus den Jahren 2000 und 2010 mit unterschiedlichen Fallzahlen haben Informationen über gleichgeschlechtliche Paare, dann sind die Schätzungen verzerrt, wenn sich der Anteil zwischen 2000 und 2010 verändert hat. Falls der Anteil aber gleichgeblieben ist, erhält man bei der Gleichgewichtung der Samples einen kleineren Varianzschätzer als bei einer Gewichtung der Surveys mit der Größe des Samples (siehe nächsten Punkt), eine Gewichtung ist also nicht nötig, bzw. sogar nicht zu empfehlen.
4. Eine weitere Möglichkeit ist die Gewichtung der Surveys mit der Größe des Samples, bzw. der effektiven Größe des Samples (nach der Gewichtung mit den enthaltenen Design- und Poststratifikation-Gewichten). Surveys mit *überdurchschnittlich* großen Fallzahlen erhalten dabei ein kleineres Gewicht als Surveys mit *unterdurchschnittlich* großen Fallzahlen.
5. Die Verwendung von populationsskalierten Gewichten (Kaminska & Lynn 2017) und die sorgfältige Beachtung dessen, was die angestrebte Referenzpopulation ist (Wähler, Einwohner, Staatsangehörige, etc.), ist eine fünfte Option. Diese Skalierung ist vor allem für die komparative Forschung bei der Kombination von Samples mit sehr ungleichen Populationsgrößen empfehlenswert. Wenn

beispielsweise je 1000 Befragte aus so unterschiedlich großen europäischen Staaten wie Deutschland, Estland oder Italien stammen, haben die estnischen Daten je nach Fragestellung bei nicht skalierten Gewichten einen zu großen Einfluss auf einen gesamteuropäischen Schätzer. Bei der Kombination von Daten zu Paaren in Deutschland über die Zeit hinweg, wird dieser Vorschlag von geringerer Bedeutung sein.

6. Schließlich schlagen Joye, Sapin und Wolf (2019) noch die Schätzung eines Multilevel-Modells ohne Populationsgewichten, aber mit Survey-Gewichten vor. Dieses ermöglicht Unterschiede zwischen Surveys direkt durch den Einbezug von Random oder Fixed Effects zu modellieren. Die beiden Ebenen des Modells sind dann einerseits Personen/Paare und andererseits Surveys. Allerdings erfordern die von Joye, Sapin und Wolf (2019) angeführten Multilevel-Modelle mit Gewichtung u.U. eine Korrektur der Survey-Gewichte (siehe ausführlich Carle 2009).

Die vorangegangenen Überlegungen verdeutlichen, dass eine Verwendung von Survey-Gewichten durchaus nicht immer notwendig ist. Falls aber erforderlich, wird die Auswahl an Survey-Gewichten und ihre Verwendung aber immer abhängig von der Forschungsfrage sein. Es gibt daher auch keine „HaSpaD“-Gewichte, bzw. harmonisierten Gewichte und es bleibt den Datennutzenden überlassen, im gegebenen Fall aus den Originaldatensätzen per ID die Gewichte (wie natürlich auch weitere Variablen) anzufügen.

5.6 Imputation von Datumsangaben im HaSpaD-Datensatz

Datumsangaben im harmonisierten HaSpaD-Datensatz liegen in Jahrhundertmonaten (der Wert 1 entspricht Januar 1900) vor. Sofern Datumsangaben sowohl monats- als auch jahresgenau berichtet wurden, war bei der Umrechnung in Jahrhundertmonaten keine Imputation nötig („0 No imputation“). Lagen hingegen die Informationen im Ursprungsdatensatz nicht monatsgenau vor (aufgrund von Erinnerungsschwierigkeiten der Befragten oder weil nur jahresgenaue Informationen erfragt wurden), wurden monatsgenaue Angaben nach folgenden Regeln imputiert:

(1) Wird nur das Jahr berichtet oder im Fragebogen nur jahresgenau abgefragt, so wird für eine Begin-Variable „Januar“, für eine End-Variable „Dezember“ als Monat imputiert. Der jeweiligen Imputationsvariable wird der Wert „1 Only year information“ zugewiesen.

(2) Haben Befragte die Möglichkeit bei Erinnerungsschwierigkeiten eine Jahreszeit anstelle einer Monatsangabe anzugeben, wird der Monat so imputiert, wie in *Tabelle 15* aufgezeigt. Dies betrifft die Mannheimer Scheidungsstudie und die Lebensverlaufsstudien LV-West 64/71 und LV-Panel 71. Die entsprechende Imputationsvariable trägt den gleichen Missing-Code, sobald die dazugehörige Biografievariable einen Fehlwert aufweist.¹¹

¹¹ Angenommen, eine Partnerin oder ein Partner verstirbt in der Ehe, werden dem Eheende und der dazugehörigen flag-imp-Variable folgendes zugewiesen: marend = „-66 Death Partner“; flag_marend_imp = „-66 Death partner“.

Tabelle 15: Imputierte Werte der Begin- und End-Variablen bei saisonalen Angaben

Studie	Wert und Label der Source-Variablen	Wert und Label der Begin-Variablen	Wert und Label der End-Variablen
Mannheimer Scheidungsstudie	17 Anfang	1 Januar	4 April
	18 Mitte	5 Mai	8 August
	19 Ende	9 September	12 Dezember
LV-West 64/71	21 Jahresanfang	1 Januar	2 Februar
	24 Frühjahr/Ostern	3 März	5 Mai
	27 Sommer	6 Juni	8 August
	30 Herbst	9 September	10 Oktober
	31 Jahresende	11 November	12 Dezember
	32 Winter	11 November	12 Dezember
LV-Panel 71	21 Jahresanfang/Winter	1 Januar	2 Februar
	24 Frühjahr/Ostern	3 März	5 Mai
	27 Jahresmitte/Sommer	6 Juni	8 August
	30 Herbst	9 September	10 Oktober
	32 Jahresende/Winter	11 November	12 Dezember

In der Mannheimer Scheidungsstudie wird – falls sich die Befragten nur saisonal an das Ereignis erinnern können – das Datum des Ereignisses nach „Anfang“, „Mitte“ und „Ende“ des Jahres abgefragt. In diesem speziellen Fall wurde für die Begin-Variablen „1 Januar“ für den Anfang, „5 Mai“ für die Mitte und „9 September“ für das Ende des Jahres angenommen. Für die End-Variablen wird „4 April“ für den Anfang, „8 August“ für die Mitte und „12 Dezember“ für das Ende des Jahres codiert. Der entsprechenden Imputationsvariable wird „2 Only saison information“ zugewiesen. Nach ähnlicher Logik wurde bei den Lebensverlaufsstudien LV-West 64/71 und LV-Panel 71 verfahren (siehe Tabelle 15). Pairfam enthält die Biografiedaten ebenfalls auf monatlicher, saisonaler oder jährlicher Basis. Jedoch liegen die Ursprungsdaten bereits imputiert in Jahreshundermonaten vor. Das Schema der Markierung des Imputationsgrades in speziellen Flag-Variablen wurde aus dem pairfam übernommen. Eine Auflistung der Imputationsgrade aller Biografievariablen je nach Teilstudie ist im Anhang A. 4 zu finden.

5.7 Längsschnittliche Aufbereitung von Panelbefragungen ohne Partner-ID (Lebensverlaufsstudie, Generations and Gender Survey)

Ausgangslage

Sowohl ein Teil der Lebensverlaufsstudien (LV-Ost Panel, LV-Panel 71) als auch die zweite Welle des Generations and Gender Surveys (2008) sind Wiederholungsbefragungen einer Teilmenge der Befragten aus der jeweils ersten Welle (LV-DDR, LV-Ost 71/LV-West 64/71 bzw. Generations and Gender Survey 2005). In der Wiederholungsbefragung ist es möglich, dass die Befragten dieselbe Partnerschaft, die bereits in der ersten Welle berichtet wurde, nochmals angeben. Es fehlt jedoch eine Partner-ID zur eindeutigen Identifikation der Partner über zwei Befragungswellen hinweg. Es wurde ein

schrittweises Vorgehen nach Plausibilitätskriterien gewählt. Diese Vorgehensweise wurde angewandt, um die Biografien von Partnerschaften zum Zeitpunkt der Welle 1 mit Werten der Welle 2 aktualisieren zu können und das Problem doppelt berichteter Partnerschaften zu reduzieren.

Vorgehensweise der Aufbereitung

Zunächst wurden sowohl die in Welle 1 als bereits abgeschlossen berichteten Partnerschaften als auch die in Welle 2 berichteten Partnerschaften, die *nach* dem Interviewdatum der ersten Welle begonnen haben, von der weiteren Längsschnittdatenaufbereitung ausgeschlossen (siehe *Abbildung 6*). Diese können ohne längsschnittliche Verknüpfung der Angaben aus den verschiedenen Panelwellen dem HaSpaD Datensatz hinzugefügt werden. Nach Ausschluss dieser Partnerschaften bleiben nur diejenigen übrig, die in Welle 1 berichtet wurden und bis zum Interviewdatum Welle 1 andauern und diejenigen, die in Welle 2 berichtet wurden und *vor* dem Interviewzeitpunkt der Welle 1 begannen.

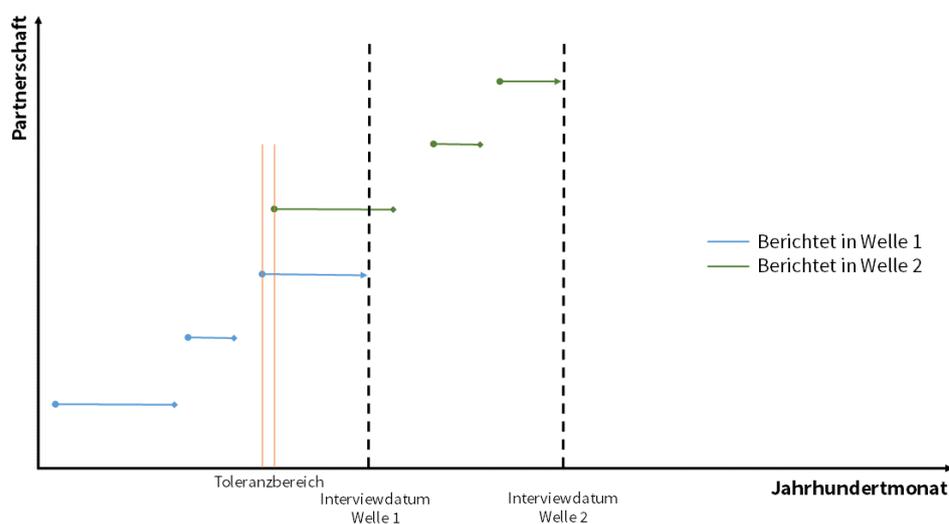


Abbildung 6: Vorgehensweise der Verknüpfung gleicher Partnerschaften

Für einige Befragte konnten in LV-Ost Panel sowie in der zweiten Welle des Generations and Gender Surveys (2008) identische Partnerschaften anhand von Filterfragen identifiziert werden, z.B. zum Familienstand und der Anzahl an Partnerschaften seit Dezember 1989 (LV-Ost Panel) beziehungsweise anhand der Frage, ob die Befragten mit derselben Partnerin/demselben Partner bzw. Ehegattin/Ehegatte der Welle 1 zusammenwohnen (GGs, Leven 2009, S. 22, Frage 305).

Bei den verbliebenen Partnerschaften wurde anhand der Datumsangaben in beiden Erhebungswellen geprüft, ob eine identische Partnerschaft plausibel ist (zum Beispiel: Ehebeginn der beiden Partnerschaften identisch, siehe *Tabelle 16*) und diese anhand der Angaben in Welle 2 aktualisiert werden kann. Erschien eine identische Partnerschaft aufgrund der Plausibilitätskriterien als nicht gegeben, wurde die zu Welle 1 bestehende Partnerschaft nicht mit den Angaben aus Welle 2 längsschnittlich aktualisiert. Partnerschaften, die in Welle 2 berichtet wurden und vor dem Interviewzeitpunkt in Welle 1 begannen, zu denen aber (vor dem Hintergrund der formulierten Plausibilitätskriterien) keine Übereinstimmung mit Angaben aus Welle 1 gefunden werden konnte, werden in der Variable *problem* mit „2 Problems due to panel linkage (GLHS, GGS)“ bzw. „4 Data inconsistency + problems due to panel linkage (GLHS, GGS)“ markiert. Häufig dürfte es sich um dieselben Partnerschaften wie

in Welle 1 berichtet handeln und ein Erinnerungsproblem vorliegen (z.B. falsches Berichten des Eheschlussjahres in einer der Erhebungswellen). Die Variable *problem* dient dazu, diese Partnerschaften zu identifizieren und ggf. weiter zu bearbeiten oder aus dem Zieldatensatz zu entfernen (siehe Kapitel 6.1.5).

Bei der längsschnittlichen Verknüpfung der Angaben wurden folgende Plausibilitätsannahmen getroffen (siehe *Tabelle 16*):

Tabelle 16: Plausibilitätsannahmen zur Identifikation in Welle 1 und Welle 2 berichteter identischer Partnerschaften

Fall	Voraussetzung	Bedingung, die erfüllt sein muss
1	Es liegen gültige Angaben in Jahrhundertmonaten für den Ehebeginn berichtet in Welle 1 (<i>marbeg_w1</i>) und Welle 2 (<i>marbeg_w2</i>) vor.	Der Ehebeginn in Welle 1 und Welle 2 berichtet stimmt in den angenommenen Grenzen (siehe <i>Tabelle 17</i>) überein: Wenn dieses Kriterium erfüllt wird, so ist die jeweils in Welle 1 und 2 berichtete Partnerschaft dieselbe und kann aktualisiert werden.
2	Es liegen gültige Angaben in Jahrhundertmonaten für den Beziehungsbeginn und den Kohabitationsbeginn berichtet in Welle 1 (<i>relbeg_w1</i> bzw. <i>cohbeg_w1</i>) und Welle 2 (<i>relbeg_w2</i> bzw. <i>cohbeg_w2</i>) vor. Eine gültige Angabe in Jahrhundertmonaten für den Ehebeginn ist nicht in beiden Wellen vorhanden	Der Beziehungsbeginn in Welle 1 und Welle 2 berichtet UND der Kohabitationsbeginn in Welle 1 und 2 berichtet stimmen in den angenommenen Grenzen (siehe <i>Tabelle 17</i>) überein: Wenn diese beiden Kriterien erfüllt sind, so ist die jeweils in Welle 1 und 2 berichtete Partnerschaft dieselbe und kann aktualisiert werden.
3	Es liegen gültige Angaben in Jahrhundertmonaten für den Beziehungsbeginn berichtet in Welle 1 (<i>relbeg_w1</i>) und Welle 2 (<i>relbeg_w2</i>) vor. Eine gültige Angabe in Jahrhundertmonaten für den Ehebeginn und den Kohabitationsbeginn ist nicht in beiden Wellen vorhanden	Der Beziehungsbeginn in Welle 1 und Welle 2 berichtet stimmt perfekt überein: Wenn der Beziehungsbeginn in beiden Wellen perfekt übereinstimmt, so ist die jeweils in Welle 1 und 2 berichtete Partnerschaft dieselbe und kann aktualisiert werden.
4	Es liegen gültige Angaben in Jahrhundertmonaten für den Kohabitationsbeginn berichtet in Welle 1 (<i>cohbeg_w1</i>) und Welle 2 (<i>cohbeg_w2</i>) vor. Eine gültige Angabe in Jahrhundertmonaten für den Ehebeginn und den Beziehungsbeginn ist nicht in beiden Wellen vorhanden	Der Kohabitationsbeginn in Welle 1 und Welle 2 berichtet stimmt perfekt überein: Wenn der Kohabitationsbeginn in beiden Wellen perfekt übereinstimmt, so ist die jeweils in Welle 1 und 2 berichtete Partnerschaft dieselbe und kann aktualisiert werden.

marbeg_w1 = Ehebeginn berichtet in Welle 1, *marbeg_w2* = Ehebeginn berichtet in Welle 2, *cohbeg_w1* = Kohabitationsbeginn berichtet in Welle 1, *cohbeg_w2* = Kohabitationsbeginn berichtet in Welle 2, *relbeg_w1* = Beziehungsbeginn berichtet in Welle 1, *relbeg_w2* = Beziehungsbeginn berichtet in Welle 2

In Fall 1 und Fall 2 der Tabelle oben werden bestimmte Grenzen angenommen, in welchen die beiden Werte der jeweiligen Begin-Variable liegen müssen, um als identische Partnerschaft identifizierbar zu sein. In der nachfolgenden *Tabelle 17* werden die Unter- und Obergrenzen angegeben:

Tabelle 17: Bestimmung der Ober- und Untergrenzen nach Kombination des Imputationsgrades

Möglichkeit	Begin-Vari- able be- richtet in Welle 1	Ausprägung von flag_ begin_imp	Begin-Vari- able be- richtet in Welle2	Ausprägung von flag_ begin_imp	l_limit (Unter- grenze in century months)	u_limit (Ober- grenze in century months)
1	No imputa- tion	0	No imputa- tion	0	-3	3
2	No imputa- tion	0	Only season information	2	-3	3
3	No imputa- tion	0	Only year information	1	0	11
4	Only season information	2	No imputa- tion	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 imputa- tion	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

Begin-Variable in Welle 1: *relbeg_w1*, *cohbeg_w1*, *marbeg_w1*

Begin-Variable in Welle 2: *relbeg_w2*, *cohbeg_w2*, *marbeg_w2*

Es werden nur Begin-Variablen mit gleichem Institutionalierungsgrad verglichen (bspw: *relbeg_w1* mit *relbeg_w2*).

Es sind neun verschiedene Möglichkeiten des Imputationsgrades beider Begin-Variablen zu unterscheiden. Die Werte beider Wellen 1 und 2 können monatsgenau (niedrigster Imputationsgrad, Möglichkeit 1) bis lediglich jahresgenau vorliegen (höchster Imputationsgrad, Möglichkeit 9). Gemischte Kombinationen der drei Imputationsgrade können ebenso auftreten. Die Unter- bzw. Obergrenze gibt die minimale bzw. maximale Differenz (in Jahrhundertmonaten) zwischen beiden Begin-Variablen an, um eine in Welle 1 und Welle 2 berichtete Partnerschaft als identisch zu kennzeichnen.

Folgende Beispiele dienen zur Illustration der Bestimmung der Differenz zwischen *marbeg_w1* und *marbeg_w2* für die neun verschiedenen Möglichkeiten der Imputationsgrade:

In *Tabelle 18* werden Beispiele für *marbeg_w1* und *marbeg_w2* gegeben, für welche die Differenz $\text{marbeg}_w1 - \text{marbeg}_w2$ im Toleranzbereich der Ober- und Untergrenze (beide Grenzen eingeschlossen) liegt und demnach als *gleiche Partnerschaft* identifiziert werden kann.

Liegt nur eine jährliche Information einer Begin-Variable vor, wird der Januar des angegebenen Jahres codiert. Im Möglichkeit 3 also kann die Differenz zwischen der monatsgenauen Angabe und Jahresangabe maximal 11 betragen, sofern der Ehebeginn im gleichen Jahr berichtet wurde. Eine identische Partnerschaft wird identifiziert, sobald das gleiche Jahr in beiden Wellen angegeben wird.

Table 18: Beispiel der Bestimmung der Unter- und Obergrenze nach Kombination des Imputationsgrades

Möglichkeit	marbeg_w1	marbeg_w2	L_limit (Untergrenze)	marbeg_w1 – marbeg_w2	u_limit (Obergrenze)
1	Januar 1990 1081	Dezember 1989 1080	-3	1 (= 1081-1080)	3
2	Januar 1990 1081	Winter 1989 1079	-3	2	3
3	Dezember 1990 1092	Jahr 1990 1081	0	11	11
4	Winter 1989 1079	Januar 1990 1081	-3	-2	3
5	Winter 1989 1079	Jahresanfang 1990 1081	-3	-2	3
6	Winter 1990 1091	Jahr 1990 1081	0	10	10
7	Jahr 1990 1081	Dezember 1990 1092	-11	-11	0
8	Jahr 1990 1081	Winter 1990 1091	-10	-10	0
9	Jahr 1990 1081	Jahr 1990 1081	0	0	0

erste Zeile in Spalte marbeg_w1/marbeg_w2: Angabe in Monaten/Saison/Jahr
zweite Zeile in Spalte marbeg_w1/marbeg_w2: Angabe in Century Months

Table 19 beschreibt die Kodierung bei Vorliegen einer saisonalen Angabe innerhalb eines Jahres. Somit wird „Winter 1989“ als „November 1989“ codiert (Jahrhundertmonate = 1079). Verglichen mit Table 18 wird bei dem Imputationsgrad „2 Only season information“ in den Begin-Angaben beider Wellen eine maximale bzw. minimale Toleranz von 3 bzw. -3 angegeben. Somit wird eine gleiche Partnerschaft identifiziert, wenn sich der Abstand beider Angaben um maximal eine Saison unterscheidet (Beispiel siehe Möglichkeit 5 in Table 18: Winter 1989 berichtet in Welle 1, Jahresanfang 1990 berichtet in Welle 2).

Table 19: Kodierung der Monate bei Vorliegen saisonaler Angaben

Saison	Kodierung der Begin-Variablen
Jahresanfang	1: Januar
Frühjahr/Ostern	3: März
Sommer	6: Juni
Herbst	9: September
Jahresende	11: November
Winter	11: November

6 Übersicht über die harmonisierten Variablen

6.1 Biografiedaten

Der Biografiedatensatz besteht aus ID-Variablen zur Identifizierung des Umfrageprogrammes und Teilstudie, der Ankerperson (Befragten) und deren Partnerin oder Partner. Weiterhin beinhalten die Biografiedaten Angaben zum (letzten) Interviewdatum, Beginn und Ende von Beziehungs-, Kohabitations- und Ehephasen einer Partnerschaft sowie das Todesdatum der Partnerin oder des Partners, wenn er oder sie in der Partnerschaft verstorben ist. Diese Daten liegen in Jahrhundertmonaten vor. Des Weiteren ist für jede Datumsvariable die Angabe in der Originaldatenquelle sowie der vorliegende Imputationsgrad als einzelne Variable integriert. Die in den Tabellen aufgeführten fehlenden und gültigen Werte in Form von Ausprägungen oder Maximal- bzw. Minimalwerten beziehen sich auf den vollständigen HaSpaD-Ziel Datensatz, welcher alle Umfrageprogramme enthält.

6.1.1 ID-Variablen

Die ID-Variablen dienen der Identifikation einzelner befragter Personen, Datenquellen und Partnerschaften und werden im Folgenden erläutert.

Tabelle 20: ID-Variable: `sid`

Variablenlabel:	Study id
Fehlende Werte:	Keine
Gültige Werte:	
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)
Datentyp:	numerisch
Erläuterung:	Die Variable <code>sid</code> gibt die Identifikationsnummer aller verwendeten Umfrageprogramme an.

Tabelle 21: ID-Variable: did

Variablenlabel:	Dataset id
Fehlende Werte:	Keine
Gültige Werte:	
1	<ul style="list-style-type: none"> - Partnerschaftsbiografie stammt aus lediglich einer Datenquelle (pairfam, ALLBUS-Kumulation, Mannheimer Scheidungsstudie, Fertility and Family Survey, SHARE). - Partnerschaftsbiografien stammen aus der Lebensverlaufsstudie LV-West I. - Partnerschaftsbiografien stammen ausschließlich aus dem Generations and Gender Survey (2005). - Partnerschaftsbiografien stammen aus dem Datensatz biomarsy des Sozio-ökonomischen Panels.
2	<ul style="list-style-type: none"> - Partnerschaftsbiografien stammen aus dem Familiensurvey II. Welle 1994. - Partnerschaftsbiografien stammen aus der Wiederholungsbefragung des Generations and Gender Survey (2008)
3	<ul style="list-style-type: none"> - Partnerschaftsbiografien stammen aus dem Familiensurvey III. Welle 2000.
11	<ul style="list-style-type: none"> - Partnerschaftsbiografien stammen aus der ersten Befragung des Familiensurveys in West-Deutschland.
12	<ul style="list-style-type: none"> - Partnerschaftsbiografien stammen aus der ersten Befragung des Familiensurveys in Ost-Deutschland. - Partnerschaften, die sowohl in der ersten Welle als auch in der zweiten Welle des Generations and Gender Surveys berichtet und aktualisiert wurden.
21	<ul style="list-style-type: none"> - Partnerschaftsbiografien stammen aus der Lebensverlaufsstudie LV-West II A - Persönliche Befragung.
22	<ul style="list-style-type: none"> - Partnerschaftsbiografien stammen aus der Lebensverlaufsstudie LV-West II T - Telefonische Befragung
40	<ul style="list-style-type: none"> - Partnerschaftsbiografien stammen aus der Lebensverlaufsstudie LV-DDR.
50	<ul style="list-style-type: none"> - Partnerschaftsbiografien stammen aus der Lebensverlaufsstudie LV-Ost 71.
61	<ul style="list-style-type: none"> - Partnerschaftsbiografien stammen aus der Lebensverlaufsstudie LV-Ost Panel (Geburtskohorten 1939-41, 1951-53, 1959-61).
62	<ul style="list-style-type: none"> - Partnerschaftsbiografien stammen aus der Lebensverlaufsstudie LV-Ost Panel (Geburtskohorte 1929-31).
70	<ul style="list-style-type: none"> - Partnerschaftsbiografien stammen aus der Lebensverlaufsstudie LV-West 64/71.
80	<ul style="list-style-type: none"> - Partnerschaftsbiografien stammen aus der Lebensverlaufsstudie LV-Panel 71.
4061	<ul style="list-style-type: none"> - Partnerschaften, die sowohl in der Lebensverlaufsstudie LV-DDR als auch in der Lebensverlaufsstudie LV-Ost Panel (Geburtskohorten 1939-41, 1951-53, 1959-61) berichtet und aktualisiert wurden.
4062	<ul style="list-style-type: none"> - Partnerschaften, die sowohl in der Lebensverlaufsstudie LV-DDR als auch in der Lebensverlaufsstudie LV-Ost Panel (Geburtskohorte 1929-31) berichtet und aktualisiert wurden.
5080	<ul style="list-style-type: none"> - Partnerschaften, die sowohl in der Lebensverlaufsstudie LV-Ost 71 als auch in der Lebensverlaufsstudie LV-Panel 71 berichtet und aktualisiert wurden.
7080	<ul style="list-style-type: none"> - Partnerschaften, die sowohl in der Lebensverlaufsstudie LV-West 64/71 (nur Geburtskohorte 1971) als auch in der Lebensverlaufsstudie LV-Panel 71 berichtet und aktualisiert wurden.
Datentyp:	Numerisch
Erläuterung:	Die Variable <code>did</code> gibt die Datensatz-ID innerhalb des jeweiligen Umfrageprogrammes an. Sobald es nur eine verwendete Datenquelle zur Erstellung der Partnerschaftsbiografie gibt, nimmt diese Variable immer den Wert 1 an.

Variablenlabel:	Dataset id
Besonderheit bei der Konstruktion	<p>Falls mehrere Studien innerhalb eines Umfrageprogrammes zur Erstellung des kumulierten Datensatzes verwendet werden (Familiensurvey, Lebensverlaufsstudie, Generations and Gender Survey) werden die Datensatz-IDs chronologisch nach Erhebungsjahr vergeben. Eine Verkettung zweier Datensatz-IDs (siehe die vierstelligen Identifikationsnummern in dieser Tabelle bei „gültige Werte“) wird vorgenommen, falls Befragte an zwei Erhebungswellen innerhalb eines Umfrageprogrammes teilgenommen haben und zu beiden Zeitpunkten Partnerschaften berichten.</p> <p>Eine Ausnahme bei der Konstruktion der <code>did</code> stellt das Sozio-oekonomische Panel dar: Die Informationen zur Erstellung des kumulierten Datensatzes des Sozio-oekonomischen Panels stammen aus drei Datenquellen, die Informationen zur Partnerschaftsbiografien und Ehestand der Befragten enthalten und nach keiner chronologischen Abfolge von Erhebungszeitpunkten sortiert sind.</p>

Tabelle 22: ID-Variable: `aid`

Variablenlabel:	Anchor id
Fehlende Werte:	Keine
Minimum:	10322
Maximum:	10212085082000
Datentyp:	Numerisch
Erläuterung:	Die Variable <code>aid</code> gibt die Anchor-ID an. Dieser Wert ergibt sich aus Verkettung der <code>sid</code> , <code>did</code> und <code>id</code> einer befragten Person ($sid + did + id$). Dieser Wert ermöglicht es einzelne Befragte über alle Einzelstudien hinweg identifizieren zu können.
Besonderheit bei der Konstruktion:	<p>Bei der Panelverknüpfung bestimmter Lebensverlaufsstudien (LV-DDR mit den beiden LV-Ost Panels, LV-Ost 71 mit LV-Panel 71, LV-West 64/71 mit LV-Panel 71) und dem Generations and Gender Survey (2005) mit dem Generations and Gender Survey (2008) wird zur Erstellung der <code>aid</code> immer die <code>did</code> der ersten Welle genutzt.</p> <p>Da zwischen den einzelnen Querschnitten der ALLBUS-Kumulation die originale Anker-ID Duplikate aufweist, wurde die <code>aid</code> aus der Verkettung der originalen Anker-ID und des jeweiligen Erhebungsjahres gebildet.</p>

Tabelle 23: ID-Variable: `id`

Variablenlabel:	Original anchor id
Fehlende Werte:	Keine
Datentyp:	String
Erläuterung:	Die Variable <code>id</code> gibt die originale Identifikationsnummer an, wie sie in den Ursprungsdatensätzen enthalten ist.

Tabelle 24: ID-Variable: pid

Variablenlabel:	Partner id
Fehlende Werte:	Keine
Datentyp:	String
Erläuterung:	Die Variable pid dient als Verknüpfungsvariable einer Partnervariablen mit der entsprechenden Partnerschaft.
Besonderheit bei der Konstruktion:	Bei Ursprungsstudien ohne eigenes Identifikationsmerkmal einer Partnerschaft, wird die pid aus der Kombination von aid und partner_pointer konstruiert (aid + partner_pointer). Bei Ursprungsstudien, die eine individuelle Identifikation eines Partners anhand einer ID erlauben (pairfam, SHARE und Sozio-oekonomisches Panel), werden Angaben zur Ursprungsstudie mit der originalen Partner-ID verkettet (sid + did + pid_o). Die Partner-ID erhält aid + partner_pointer bei Multi-Actor-Studien (pairfam, SHARE und Sozio-oekonomisches Panel), wenn keinerlei Information über die originale ID des Partners vorliegt.

Tabelle 25: ID-Variable: pid_o

Variablenlabel:	Original partner id
Fehlende Werte:	
-10	Missing by study design
Datentyp:	String
Erläuterung:	Die Variable pid_o gibt die originale Identifikationsnummer der Partner der Befragten so an, wie sie im Ursprungsdatensatz vergeben wurde.

Tabelle 26: ID-Variable: partner_pointer

Variablenlabel:	Partner pointer
Fehlende Werte:	Keine
Minimum	0
Maximum	17
Datentyp:	Numerisch
Erläuterung:	Die Variable partner_pointer gibt die Position der Partnerinformationen in den Ursprungsdaten an (z.B. erster, zweiter, dritter Spell; erstes, zweites, drittes Set an Variablen zu Partnerschaften).

6.1.2 Biografie-Variablen

Tabelle 27: Biografie-Variablen: `intd`

Variablenlabel:	Date of interview (in century months)
Fehlende Werte:	
-10	Missing by study design
Minimum:	961
Maximum:	1395
Datentyp:	Numerisch
Erläuterung:	Die Variable <code>intd</code> gibt das Interviewdatum in Jahrhundertmonaten an.
Besonderheit bei der Konstruktion:	Sofern das Interviewdatum nicht bekannt ist, wird der Endmonat der Feldphase gewählt.

Tabelle 28: Biografie-Variablen: `relbeg`

Variablenlabel:	Beginning relationship (in century months)
Fehlende Werte:	
-7	Incomplete data
-10	Missing by study design
Minimum:	325
Maximum:	1394
Datentyp:	Numerisch
Erläuterung:	Die Variable <code>relbeg</code> gibt den Beginn der Beziehung in einer Partnerschaft in Jahrhundertmonaten an.

Tabelle 29: Biografie-Variable: re_lend

Variablenlabel:	End relationship (in century months)
Fehlende Werte:	
-7	Incomplete data
-10	Missing by study design
-66	Death partner
-77	Separation/divorce, no exact date
-99	Ongoing
Minimum:	481
Maximum:	1393
Datentyp:	Numerisch
Erläuterung:	Die Variable re_lend gibt das Ende der Beziehung in einer Partnerschaft in Jahrhundertmonaten an.
Besonderheit bei der Konstruktion:	Sofern lediglich das Datum des Auseinanderzuges und eine Trennung als Grund für diesen bekannt ist, kann kein exaktes Trennungsdatum vergeben werden, da dieses Ereignis nicht unbedingt zeitgleich mit dem Kohabitationsende aufgetreten sein muss. In diesem Fall wird re_lend = „-77 Separation/divorce, no exact date“ zugewiesen.

Tabelle 30: Biografie-Variable: cohbeg

Variablenlabel:	Beginning cohabitation (in century months)
Fehlende Werte:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
Minimum:	369
Maximum:	1394
Datentyp:	Numerisch
Erläuterung:	Die Variable cohbeg gibt den Beginn des Zusammenlebens in einem gemeinsamen Haushalt mit den jeweiligen Partnern in Jahrhundertmonaten an.

Tabelle 31: Biografie-Variable: cohend

Variablenlabel:	End cohabitation (in century months)
Fehlende Werte:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
-66	Death partner
-99	Ongoing
Minimum:	484
Maximum:	1394
Datentyp:	Numerisch
Erläuterung:	Die Variable <code>cohend</code> gibt das Ende des Zusammenlebens in einem gemeinsamen Haushalt mit den jeweiligen Partnern in Jahrhundertmonaten an.

Tabelle 32: Biografie-Variable: marbeg

Variablenlabel:	Beginning marriage (in century months)
Fehlende Werte:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
Minimum:	145
Maximum:	1392
Datentyp:	Numerisch
Erläuterung:	Die Variable <code>marbeg</code> gibt den Beginn der Ehe in einer Partnerschaft in Jahrhundertmonaten an.

Tabelle 33: Biografie-Variable: marend

Variablenlabel:	End marriage (in century months)
Fehlende Werte:	
-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
Minimum:	264
Maximum:	1391
Datentyp:	Numerisch
Erläuterung:	Die Variable marend gibt das Ende der Ehe in einer Partnerschaft in Jahrhundertmonaten an.
Besonderheit bei der Konstruktuon:	Sofern lediglich das Datum des Auseinanderzuges und eine Scheidung als Grund für diesen bekannt ist, kann kein exaktes Scheidungsdatum vergeben werden, da dieses Ereignis nicht unbedingt zeitgleich mit dem Kohabitationsende aufgetreten sein muss. In diesem Fall wird marend = „-77 Separation/divorce, no exact date“ zugewiesen. Die Ausprägung „-88 Union dissolution/death of partner“ wird vergeben, sofern unklar ist, ob eine Ehe mit dem Tod der Partnerin oder des Partners oder der Scheidung endet. 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, S. 15)

Tabelle 34: Biografie-Variable: dop

Variablenlabel:	Date of partner's death (in century months)
Fehlende Werte:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
-88	Union dissolution/death of partner
Minimum:	180
Maximum:	1390
Datentyp:	Numerisch
Erläuterung:	Die Variable <code>dop</code> gibt den Zeitpunkt des Todes der jeweiligen Partner in Jahrhundertmonaten an.

6.1.3 Imputed-Variablen

Die Imputed-Variablen (`imp`-Variablen) geben den Imputationsgrad der Biografievariablen an.

Tabelle 35: Imputed-Variable: `flag_relbeg_imp`

Variablenlabel:	Imputed date of beginning relationship
Fehlende Werte:	
-7	Incomplete data
-10	Missing by study design
Gültige Werte:	
0	No imputation
1	Only year information
2	Only season information
Datentyp:	Numerisch
Erläuterung:	Die Variable <code>flag_relbeg_imp</code> gibt den Imputationsgrad der Variable <code>relbeg</code> an.

Tabelle 36: Imputed-Variable: flag_relend_imp

Variablenlabel:	Imputed date of end relationship
Fehlende Werte:	
-7	Incomplete data
-10	Missing by study design
-66	Death partner
-77	Separation/divorce, no exact date
-99	Ongoing
Gültige Werte:	
0	No imputation
1	Only year information
2	Only season information
Datentyp:	Numerisch
Erläuterung:	Die Variable flag_relend_imp gibt den Imputationsgrad der Variable relend an.

Tabelle 37: Imputed-Variable: flag_cohbeg_imp

Variablenlabel:	Imputed date of beginning cohabitation
Fehlende Werte:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
Gültige Werte:	
0	No imputation
1	Only year information
2	Only season information
Datentyp:	Numerisch
Erläuterung:	Die Variable flag_cohbeg_imp gibt den Imputationsgrad der Variable cohbeg an.

Tabelle 38: Imputed-Variable: flag_cohend_imp

Variablenlabel:	Imputed date of end cohabitation
Fehlende Werte:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
-66	Death partner
-99	Ongoing
Gültige Werte:	
0	No imputation
1	Only year information
2	Only season information
Datentyp:	Numerisch
Erläuterung:	Die Variable flag_cohend_imp gibt den Imputationsgrad der Variable cohend an.

Tabelle 39: Imputed-Variable: flag_marbeg_imp

Variablenlabel:	Imputed date of beginning marriage
Fehlende Werte:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
Gültige Werte:	
0	No imputation
1	Only year information
2	Only season information
Datentyp:	Numerisch
Erläuterung:	Die Variable flag_marbeg_imp gibt den Imputationsgrad der Variable marbeg an.

Tabelle 40: Imputed-Variable: flag_marend_imp

Variablenlabel:	Imputed date of end marriage
Fehlende Werte:	
-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
Gültige Werte:	
0	No imputation
1	Only year information
2	Only season information
Datentyp:	Numerisch
Erläuterung:	Die Variable flag_marend_imp gibt den Imputationsgrad der Variable marend an.

Tabelle 41: Imputed-Variable: flag_dop_imp

Variablenlabel:	Imputed date of partner's death
Fehlende Werte:	
-3	Does not apply
-7	Incomplete data
-10	Missing by study design
-88	Union dissolution/death of partner
Gültige Werte:	
0	No imputation
1	Only year information
2	Only season information
Datentyp:	Numerisch
Erläuterung:	Die Variable flag_dop_imp gibt den Imputationsgrad der Variable dop an.

6.1.4 Original-Variablen

Die Original-Variablen (*ori*-Variablen) enthalten die aus der Ursprungs-Datenquelle stammenden, nicht harmonisierten Variablenwerte der jeweiligen Biografievariable und machen die Kodierung der Sourcevariablen transparent. Da diese zumeist auf Angaben zum Kalenderjahr (ggf. auch Monat) beruhen, können diese Variablen einen schnellen Überblick über die Plausibilität bestimmter Befragtenangaben geben.

Die Regeln zur Konstruktion der *ori*-Variablen werden im Folgenden erklärt. Sind Informationen zu Monat und Jahr im Ursprungsdatensatz gegeben, werden beide Werte als Stringvariablen miteinander verkettet („Monat“ „+“ „Jahr“) und ergeben die Ausprägung der entsprechenden *ori*-Variable. Ist nur die Jahreszahl bekannt (d.h. nur diese Informationen wird abgefragt), wird der Wert ohne Verkettung übernommen. Beispielsweise gibt die befragte Person einen Ehebeginn im August 1989 an (*marbeg_monat* = 08; *marbeg_jahr* = 1989), so werden die Informationen beider Variablen in *flag_marbeg_ori* als „08+1989“ verkettet. In *Anhang A. 5* befindet sich eine genauere Auflistung aller Original-Variablen.

Auf einen Sonderfall soll hier besonders hingewiesen werden:

Es liegen (teilweise) keine Informationen für die *ori*-Variablen vor, wenn die benötigten Informationen nicht im Ursprungsdatensatz abgefragt worden sind. In diesem Fall trifft „-10 Missing by study design“ auf die entsprechende *ori*-Variable zu.

Tabelle 42: Original-Variable: *flag_relbeg_ori*

Variablenlabel:	Original date of beginning relationship
Fehlende Werte:	
-10	Missing by study design
Datentyp:	String
Erläuterung:	Die Variable <i>flag_relbeg_ori</i> gibt die im Ursprungsdatensatz angegebenen Informationen zum Beginn einer Beziehung an.
Besonderheit bei der Konstruktion:	In der Lebensverlaufsstudie LVS-DDR wird die Variable <i>flag_relbeg_ori</i> aus der Differenz zwischen dem Beginn der Kohabitation und der Dauer, wie lang sich die befragte Person und ihre Partnerin oder Partner vorher kannten, ermittelt. Alle dafür benötigten Variablen werden verkettet.

Tabelle 43: Original-Variable: `flag_relend_ori`

Variablenlabel:	Original date of end relationship
Fehlende Werte:	
-10	Missing by study design
Datentyp:	String
Erläuterung:	Die Variable <code>flag_relend_ori</code> gibt die im Ursprungsdatensatz angegebenen Informationen zum Ende einer Beziehung an.
Besonderheit bei der Konstruktion:	Die Datensätze des Sozio-oekonomischen Panels enthalten Zeitabschnitte (Spells), die jeweiligen Abschnitte einer Partnerschaft und somit Ereignisse markiert. Für das Soziökonomische Panel wurden neben Jahreszahlen auch die censor-Informationen in den <code>_ori</code> -Variablen abgelegt, da diese Auskunft über eine mögliche recht- oder links-zensierung geben. Für mehr Informationen, siehe html-documentation „coding_soep“.

Tabelle 44: Original-Variable: `flag_cohbeg_ori`

Variablenlabel:	Original date of beginning cohabitation
Fehlende Werte:	
-10	Missing by study design
Datentyp:	String
Erläuterung:	Die Variable <code>flag_cohbeg_ori</code> gibt die im Ursprungsdatensatz angegebenen Informationen zum Beginn einer Kohabitation an.

Tabelle 45: Original-Variable: `flag_cohend_ori`

Variablenlabel:	Original date of end cohabitation
Fehlende Werte:	
-10	Missing by study design
Datentyp:	String
Erläuterung:	Die Variable <code>flag_cohend_ori</code> gibt die im Ursprungsdatensatz angegebenen Informationen zum Ende einer Kohabitation an.
Besonderheit bei der Konstruktion:	Bei der LV-West I, <code>cohend</code> setzt sich aus drei Variablen zusammen. Deswegen ist <code>flag_cohend_ori</code> in diesem Fall eine Verkettung aus den Ursprungsvariablen „Dauer des Getrenntlebens vor Scheidung“, „Scheidungsmonat“ und „Scheidungs-jahr“. Für die LV-West II A - Persönliche Befragung wird bei der Variable <code>flag_cohend_ori</code> die Dauer des Getrenntlebens übernommen. Falls die befragte Person ein Scheidungsdatum genannt hat, werden dort die „Dauer des Getrenntlebens vor Scheidung“ und das „Scheidungs-jahr“ miteinander verkettet.

Tabelle 46: Original-Variable: `flag_marbeg_ori`

Variablenlabel:	Original date of beginning marriage
Fehlende Werte:	
-10	Missing by study design
Datentyp:	String
Erläuterung:	Die Variable <code>flag_marbeg_ori</code> gibt die im Ursprungsdatensatz angegebenen Informationen zum Beginn einer Ehe an.
Besonderheit bei der Konstruktion:	Die Datensätze des Sozio-oekonomischen Panels enthalten Zeitabschnitte (Spells), die jeweiligen Abschnitte einer Partnerschaft und somit Ereignisse markiert. Für das Sozioökonomische Panel wurden neben Jahreszahlen auch die censor-Informationen in den <code>_ori</code> -Variablen abgelegt, da diese Auskunft über eine mögliche recht- oder links-zensurierung geben. Für mehr Informationen, siehe html-documentation „coding_soep“ .

Tabelle 47: Original-Variable: `flag_marend_ori`

Variablenlabel:	Original date of end marriage
Fehlende Werte:	
-10	Missing by study design
Datentyp:	String
Erläuterung:	Die Variable <code>flag_marend_ori</code> gibt die im Ursprungsdatensatz angegebenen Informationen zum Ende einer Ehe an.
Besonderheit bei der Konstruktion:	Die Datensätze des Sozio-oekonomischen Panels enthalten Zeitabschnitte (Spells), die jeweiligen Abschnitte einer Partnerschaft und somit Ereignisse markiert. Für das Sozioökonomische Panel wurden neben Jahreszahlen auch die censor-Informationen in den <code>_ori</code> -Variablen abgelegt, da diese Auskunft über eine mögliche recht- oder links-zensurierung geben. Für mehr Informationen, siehe html-documentation „coding_soep“ .

Tabelle 48: Original-Variable: `flag_dop_ori`

Variablenlabel:	Original date of partner's death
Fehlende Werte:	
-10	Missing by study design
Datentyp:	String
Erläuterung:	Die Variable <code>flag_dop_ori</code> gibt die im Ursprungsdatensatz angegebenen Informationen zum Zeitpunkt des Todes des Partners an.
Besonderheit bei der Konstruktion:	Bei den Familiensurveys gilt die Variable <code>dop</code> nur für Verheiratete. Den Unverheirateten wird der Wert „-10 Missing by study design“ zugewiesen. Beim Fertility and Family Survey (Deutschland) ist darauf zu achten, dass <code>flag_dop_ori</code> als Monat und Jahr des Endes des Zusammenwohnens verkettet werden kann, sofern die befragte Person den Tod der Partnerin oder des Partners als Trennungsgrund angibt. Sofern der Trennungsgrund nicht der Tod der Partnerin oder des Partners ist, ist <code>flag_dop_ori</code> = „-10 Missing by study design“. Die Datensätze des Sozio-oekonomischen Panels enthalten Zeitabschnitte (Spells), die jeweiligen Abschnitte einer Partnerschaft und somit Ereignisse markiert. Für das Sozioökonomische Panel wurden neben Jahreszahlen auch die censor-Informationen in den <code>_ori</code> -Variablen abgelegt, da diese Auskunft über eine mögliche recht- oder links-zensierung geben. Für mehr Informationen, siehe <code>html-documentation „coding_soep“</code> .

6.1.5 Zusätzliche Flag-Variablen

In diesem Abschnitt werden zwei weitere Flag-Variablen erklärt.

Tabelle 49: Zusätzliche Flag-Variable: `flag_beg`

Variablenlabel:	Beginning partnership biography
Fehlende Werte:	
-10	Missing by study design
Gültige Werte:	
0	With relationship
1	With cohabitation
2	With marriage
Datentyp:	numerisch
Erläuterung:	Die Variable <code>flag_beg</code> gibt an, mit welchem Institutionalierungsgrad (Partnerschaftsgrad) die Paarbiografie in einer Datenquelle beginnt.

Tabelle 50: Zusätzliche Flag-Variable: `problem`

Variablenlabel:	Flag used to mark problem cases
Fehlende Werte:	Keine
Gültige Werte:	
0	No data inconsistency
1	Data inconsistency
2	Problems due to panel linkage (GLHS, GGS)
3	Higher order partnerships to same partner (pairfam, FFS, SOEP)
4	Data inconsistency + problems due to panel linkage (GLHS, GGS)
5	Data inconsistency + higher order partnerships to same partner (pairfam, FFS)
6	Duplicate Multi-Actor-Design (SHARE)
Datentyp:	Numerisch
Besonderheit bei der Konstruktion:	<p>Die Variable <code>problem</code> dient den Nutzenden zur Identifikation von Inkonsistenzen, die bereits in den Ursprungsdaten vorlagen oder während der Datenaufbereitung festgestellt wurden. In <i>Anhang A. 6</i> sind die Ausprägungen und die einzeln durchgeführten Datenprüfungen näher erläutert.</p> <p>Der Wert „0 No data inconsistency“ gibt an, dass keine Inkonsistenzen festgestellt wurden.</p> <p>Der Wert „1 Data inconsistency“ weist auf ein Problem bei den Datumsangaben der Partnerschaftsbiografien hin. Wenn beispielweise das Ende einer Beziehung vor deren Beginn datiert wird, wurden die Biografiedaten in der Ursprungsdatenquelle inkonsistent aufgenommen.</p> <p>Der Wert „2 Problems due to panel linkage (GLHS, GGS)“ wird bei Partnerschaften vergeben, die bei der Panelverknüpfung in den Lebensverlaufsstudien und Generations and Gender Surveys Probleme aufweisen. In <i>Abschnitt 5.7</i> wird das Vorgehen der Panelverknüpfung und die damit verbundenen Probleme näher erläutert.</p> <p>Sowohl im Sozio-oekonomischen Panel als auch im pairfam und Fertility and Family Survey (FFS) lassen sich aufgrund der Vergabe der Partner-ID (<code>pid</code>) der Sonderfall von mehreren unabhängig geführten Partnerschaften zur gleichen Partnerin oder zum gleichen Partner nachweisen. Falls die Befragten mehrere Partnerschaften zur gleichen Partnerin oder zum gleichen Partner angeben, sind alle ab der zweiten Partnerschaft berichteten mit dem Wert „3 Higher order partnerships to same partner (pairfam, FFS, SOEP)“ markiert.</p> <p>Die Werte „4 Data inconsistency + problems due to panel linkage (GLHS, GGS)“ und „5 Data inconsistency + higher order partnerships to same partner (pairfam, FFS)“ sind Kombinationen aus Wert „1 Data inconsistency“ mit den Ausprägungen „2 Problems due to panel linkage (GLHS, GGS)“ bzw. „3 Higher order partnerships to same partner (pairfam, FFS, SOEP)“. Näheres zu den Werten und ihrer Bedeutung ist in <i>Tabelle A 5</i> im Anhang zu finden.</p> <p>Aufgrund des Multi-Actor-Designs des SHARE befinden sich im Datensatz Partnerschaften, die aus der Perspektive beider Befragter berichtet wurden. Damit Nutzende für ihre Analysen eine der beiden Partnerschaften entfernen können, sind für heterosexuelle Partnerschaften die Partnerschaften des Mannes und für homosexuelle Partnerschaften der Person mit einer größeren <code>aid</code> mit dem Wert „6 Duplicate Multi-Actor-Design (SHARE)“ in der <code>problem</code>-Variable gekennzeichnet.</p>

6.2 Harmonisierte Zusatzvariablen

Zusatzvariablen auf Befragten- und Partnerebene, wie Geburtsdaten, Staatsbürgerschaften, Bildung, Konfession aber auch Informationen über Scheidung- und Trennung der Eltern des Befragten gehören neben den Paarbiografien zu dem Angebot der harmonisierten Variablen. Eine Übersicht über die aufbereiteten Variablen findet sich auf den Seiten des HaSpaD-Wizards. Detaillierte Dokumentationen der Kodierung und weitere Literaturangaben befinden sich unter `~haspad/1_documentation/harmonization_targets` in den jeweiligen HTML-Dokumentationen, sofern diese für den Harmonisierungsprozess ausgewählt wurden. Diese Dokumente werden während des Ausführens des Do-Files `main.do` generiert und beinhalten sowohl Kodierschema als auch die einzelnen Aufbereitungsdokumentationen der jeweiligen Zusatzvariablen aller im Harmonization Wizard ausgewählten Umfrageprogramme. Wurde beispielsweise das Geschlecht des Befragten (`sex`) ausgewählt, ist unter `~haspad/1_documentation/harmonization_targets/coding_sex.html` das Kodierschema und die Aufbereitung aller ausgewählten Umfrageprogramme dokumentiert. Weitere Information über die Ordnerstruktur finde sich im *Kapitel 3* (und zusätzlich in *Anhang A. 2*).

7 Literaturverzeichnis

- 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., Drobnič, 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). Beziehungs- und Familienpanel (pairfam). GESIS Datenarchiv, Köln. ZA5678 Datenfile 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. access via https://www.pairfam.de/fileadmin/user_upload/redakteur/publish/Dokumentation/TechnicalPapers/TP01%20Cumulated%20Field%20Report%2C%20pairfam%202021.pdf [22/09/2021, 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 BIOCUPLE and BIOCUPLY, 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., Metzing, M., Richter, D., Schacht, D., Schmelzer, P., Schmitt, C., Schnitzlein, D., Siegers, R., Wenzig, K. & Zimmermann, S. (2019) *Sozio-oekonomisches Panel (SOEP), Version 35, Daten der Jahre 1984-2018 (SOEP-Core v35)*. 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 Statistisches 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-Dokument [abgerufen am 26.09.2019, 13.30 Uhr]
- 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. Verfügbar unter <http://www.ccsr.isr.umich.edu/> [abgerufen am 06.02.2020, 11.30 Uhr]
- 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
- 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. Verfügbar unter http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-1997-en_0.pdf [abgerufen am 22.06.2020, 10.30 Uhr]
- 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. 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).

A Anhang

A.1 Beschaffungsmöglichkeiten der in HaSpaD harmonisierten Umfrageprogramme

Tabelle A 1 gibt einen Überblick über die Beschaffungsmöglichkeiten der im HaSpaD-Projekt verwendeten Studien. Über den Link zum Datenvetrieb gelangt man direkt zum Download der Datenquellen. Wichtig ist, dass die benötigte Version (letzte Spalte in der Tabelle) beschafft werden muss, da ansonsten die erfolgreiche Erstellung des HaSpaD-Zieldatensatzes nicht gewährleistet werden kann.

Tabelle A 1: Beschaffungsmöglichkeiten der in HaSpaD harmonisierten Umfrageprogramme

Titel der Studie	Forschungsdaten-zentrum (FDZ) / Datengeber / Datenerzeuger	Link zum FDZ / Datengeber	Link zum Datenvertrieb / DOI	Benötigte Version
<i>pairfam Data Release 12.0.0</i>	FDZ pairfam	Link zur pairfam Homepage https://www.pairfam.de	Link zur GESIS Suche https://doi.org/10.4232/pairfam.5678.12.0.0	Version 12.0.0
<i>ALLBUS – Kumulation 1980 – 2016</i>	FDZ ALLBUS	Link zum FDZ ALLBUS https://www.gesis.org/allbus/allbus	Link zur GESIS Suche http://dx.doi.org/10.4232/1.13029	Version 1.0.0
<i>Wandel und Entwicklung familiärer Lebensformen in Westdeutschland (Familiensurvey)</i>	FDZ DJI (Deutsches Jugendinstitut)	Link zur Forschungsdatenbank der DJI https://surveys.dji.de/index.php?m=msg,0&fID=2	Link zur GESIS Suche http://dx.doi.org/10.4232/1.2245	Version 1.0.0
<i>Familie und Partnerbeziehungen in Ostdeutschland (Familiensurvey)</i>	FDZ DJI (Deutsches Jugendinstitut)	Link zur Forschungsdatenbank der DJI https://surveys.dji.de/index.php?m=msg,0&fID=2	Link zur GESIS Suche http://dx.doi.org/10.4232/1.13196	Version 1.1.0

Titel der Studie	Forschungsdaten- zentrum (FDZ) / Da- tengeber / Datener- zeuger	Link zum FDZ / Da- tengeber	Link zum Datenver- trieb / DOI	Benötigte Version
<i>Wandel und Ent- wicklung familia- ler Lebensformen - 2. Welle (Famili- ensurvey)</i>	FDZ DJI (Deutsches Ju- gendinstitut)	Link zur For- schungsdatenbank der DJI	Link zur GESIS Suche http://dx.doi.org/10. 4232/1.13197	Version 1.1.0
<i>Wandel und Ent- wicklung familia- ler Lebensformen - 3. Welle (Famili- ensurvey)</i>	FDZ DJI (Deutsches Ju- gendinstitut)	Link zur For- schungsdatenbank der DJI	Link zur GESIS Suche http://dx.doi.org/10. 4232/1.13198	Version 1.1.0
<i>Mannhei- mer Scheidungs- studie 1996</i>	Mannheimer Zentrum für Europäi- sche Sozialforschung (MZES)	Link zum Mannhei- mer Zentrum für europäische Sozial- forschung	Link zur GESIS Suche http://dx.doi.org/10. 4232/1.13056	Version 2.0.0
<i>Deutscher Fertil- ity and Family Survey 1992</i>	Bundesinstitut für Be- völkerungs-forschung (BiB)	Link zur Info-Seite des Bundesinsti- tuts für Bevölke- rungsforschung	Link zur GESIS Suche http://dx.doi.org/10. 4232/1.3400	Version 1.0.0
<i>Lebensverläufe und gesellschaftli- cher Wandel: Le- bensverläufe und Wohlfahrtsent- wicklung (Lebens- verlaufsstudie LV- West I)</i>	Max-Planck-Institut für Bildungsforschung	Link zur Info-Seite des MPI für Bil- dungsforschung	Link zur GESIS Suche http://dx.doi.org/10. 4232/1.13193	Version 1.1.0

Titel der Studie	Forschungsdaten- zentrum (FDZ) / Da- tengeber / Datener- zeuger	Link zum FDZ / Da- tengeber	Link zum Datenver- trieb / DOI	Benötigte Version
<i>Lebensverläufe und gesellschaftli- cher Wandel: Die Zwischenkriegs- kohorte im Über- gang zum Ruhe- stand (Lebensver- laufsstudie LV- West II A - Persön- liche Befragung)</i>	Max-Planck-Institut für Bildungsforschung	Link zur Info-Seite des MPI für Bil- dungsforschung	Link zur GESIS Suche http://dx.doi.org/10. 4232/1.13194	Version 1.1.0
<i>Lebensverläufe und gesellschaftli- cher Wandel: Die Zwischenkriegs- kohorte im Über- gang zum Ruhe- stand (Lebensver- laufsstudie LV- West II T - Telefo- nische Befragung)</i>	Max-Planck-Institut für Bildungsforschung	Link zur Info-Seite des MPI für Bil- dungsforschung	Link zur GESIS Suche http://dx.doi.org/10. 4232/1.2647	Version 1.0.0
<i>Lebensverläufe und gesellschaftli- cher Wandel: Be- rufszugang in der Beschäftigungs- krise (Lebensver- laufsstudie LV- West III)</i>	Max-Planck-Institut für Bildungsforschung	Link zur Info-Seite des MPI für Bil- dungsforschung	Link zur GESIS Suche http://dx.doi.org/10. 4232/1.13195	Version 1.1.0
<i>Lebensverläufe und historischer Wandel in Ost- deutschland (Le- bensverlaufsstu- die LV-DDR)</i>	Max-Planck-Institut für Bildungsforschung	Link zur Info-Seite des MPI für Bil- dungsforschung	Link zur GESIS Suche http://dx.doi.org/10. 4232/1.2644	Version 1.0.0

Titel der Studie	Forschungsdaten-zentrum (FDZ) / Da-tengeber / Datener-zeuger	Link zum FDZ / Da-tengeber	Link zum Datenver-trieb / DOI	Benötigte Version
<i>Ostdeutsche Lebensverläufe im Transformations-prozeß (Lebens-verlaufsstudie LV-Ost 71)</i>	Max-Planck-Institut für Bildungsforschung	Link zur Info-Seite des MPI für Bildungsforschung	Link zur GESIS Suche http://dx.doi.org/10.4232/1.3926	Version 1.0.0
<i>Ostdeutsche Lebensverläufe im Transformations-prozeß (Lebens-verlaufsstudie LV-Ost Panel)</i>	Max-Planck-Institut für Bildungsforschung	Link zur Info-Seite des MPI für Bildungsforschung	Link zur GESIS Suche http://dx.doi.org/10.4232/1.3925	Version 1.0.0
<i>Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutsch-land (Lebensver-laufsstudie LV-West 64/71)</i>	Max-Planck-Institut für Bildungsforschung	Link zur Info-Seite des MPI für Bildungsforschung	Link zur GESIS Suche http://dx.doi.org/10.4232/1.3927	Version 1.0.0
<i>Frühe Karrieren und Familien-gründung: Lebensverläufe der</i>	Max-Planck-Institut für Bildungsforschung	Link zur Info-Seite des MPI für Bildungsforschung	Link zur GESIS Suche	Version 1.0.0

Titel der Studie	Forschungsdaten- zentrum (FDZ) / Da- tengeber / Datener- zeuger	Link zum FDZ / Da- tengeber	Link zum Datenver- trieb / DOI	Benötigte Version
<i>Geburtskohorte 1971 in Ost- und Westdeutschland (Lebensverlaufs- studie LV-Panel 71)</i>		https://www.mpib-berlin.mpg.de/forschung/beendete-bereiche/forschungsbereich-bildung-arbeit-und-gesellschaftliche-entwicklung	http://dx.doi.org/10.4232/1.5099	
<i>Generations and Gender Survey (2005) (Teilstich- probe Deutsch- land)</i>	Bundesinstitut für Bevölkerungsforschung (BiB) Generations and Gender Programme (GGP)	Link zum Kontakt des Generations and Gender Programme	Link zur GGP-Data Access Seite https://www.ggp-i.org/form/accounts/login/?next=/form/	GGS Wave 1 Version 4.3. - Consolidated
<i>Generations and Gender Survey (2008) (Teilstich- probe Deutsch- land)</i>	Bundesinstitut für Bevölkerungsforschung (BiB) Generations and Gender Programme (GGP)	Link zum Kontakt des Generations and Gender Programme	Link zur GGP-Data Access Seite https://www.ggp-i.org/form/accounts/login/?next=/form/	GGS Wave 2 Version 1.3. - Consolidated
<i>SHARE Wellen 1, 2 und 3</i>	SHARE Research Data Center FDZ SHARE	Link zum SHARE Project	Link zum SHARE RDC http://www.share-project.org/data-access/user-registration.html	Version 6-0-0
<i>Sozio-oekonomi- sches Panel (SOEP), Daten der Jahre 1984-2018</i>	Deutsches Institut für Wirtschaftsforschung (DIW) FDZ SOEP	Link zum FDZ SOEP	Link zum Datenzugang des SOEP https://www.diw.de/de/diw_02.c.221180.de/fdz_soep.html https://www.diw.de/de/diw_02.c.222836.de/datenzugang.html	Version 35

A. 2 Der HaSpaD-Programm-Code

Zur Demonstration werden im folgenden Beispiel die Datensätze der ALLBUS-Kumulation und des Deutschen Fertility and Family Survey (FFS) genutzt.

Im Ordner `3_harmonization/dofiles` liegen alle Stata-Programmdateien (Do-Files), die zur Erstellung des Zieldatensatzes, bestehend aus Biografiedaten und Zusatzvariablen, benötigt werden. In diesem Ordner sollten Nutzende keine Änderungen vornehmen, da der Main-Do-File sonst nicht mehr auf benötigte Programmdateien zugreifen kann.

Die Namen der Do-files entsprechen der folgenden Systematik:

sid_Studienakronym_Suffix.do

Dabei entspricht die Studien-ID (*sid*) der im HaSpaD-Projekt zugeordneten Identifikationsnummer eines bestimmten Umfrageprogramms. Hier entspricht 102 der ALLBUS-Kumulation - 1980-2016 (Studienakronym: *allbus*), 106 dem Deutschen Fertility and Family Survey (Studienakronym: *ffs*). Das Suffix bezeichnet die Funktion eines Do-files. Do-files mit Endung *_bio* werden zur Aufbereitung der Partnerschaftsbiografien verwendet; die Endung *_sex* bezeichnet die aus den Ursprungsdaten harmonisierte Variable Geschlecht, *_eastwest* das Erhebungsgebiet.

Im Ordner `temp` werden temporäre Stata-Dateien zwischengespeichert, die zur Erstellung des finalen Zieldatensatzes verwendet werden.

Für die Aufbereitung benötigte Hilfs-Do-files sind im Ordner `3_harmonization/resources` gespeichert. Das Do-file `labelfile.do` vergibt nach dem Ausführen der einzelnen Aufbereitungsfiles von **_bio.do* Variablenlabels für alle Biografievariablen. Packages, die für die Aufbereitung benötigt werden, werden über das Do-File `packages.do` geladen.

In dem zunächst leeren Ordner `4_target` wird nach erfolgreichem Ausführen des Main-Files ein finaler Zieldatensatz `target.dta` mit Biografiedaten und Zusatzvariablen für die vom Nutzenden gewünschten Umfrageprogramme und Variablen gespeichert.

Das Main-File legt unter den Makrobefehlen `global drive` bzw. `global source` die beiden Pfade zum Ausführen der Aufbereitungs- und Harmonisierungs-Do-Files, Erstellung der Zwischendatensätze und HTML-Files bzw. den Ort der Source-Datensätze (vom Nutzenden heruntergeladene und entpackte Datenquellen) an.

Während des Ausführens des Main-Files werden unter `1_documentation` zwei Ordner `harmonization_targets` und `preparation_couplebio` angelegt, in welchen die jeweiligen Dokumentationen der Aufbereitung der Paarbiografien (`preparation_couplebio`) und der Harmonisierung der Zusatzvariablen (`harmonization_targets`) liegen.

Erstellung der kumulierten Partnerschaftsbiografien

Nach dem automatisierten Anlegen der beiden Ordner `harmonization_target` und `preparation_target` harmonisiert und kumuliert das Main-File die Paarbiografien der vom Nutzenden ausgewählten Umfrageprogramme unter `# RUN PREFILES FROM COUPLEBIO` (siehe `main.do`). Dabei dient folgendes Flussdiagramm zur Anschauung (siehe *Abbildung A 1*, von oben nach unten):

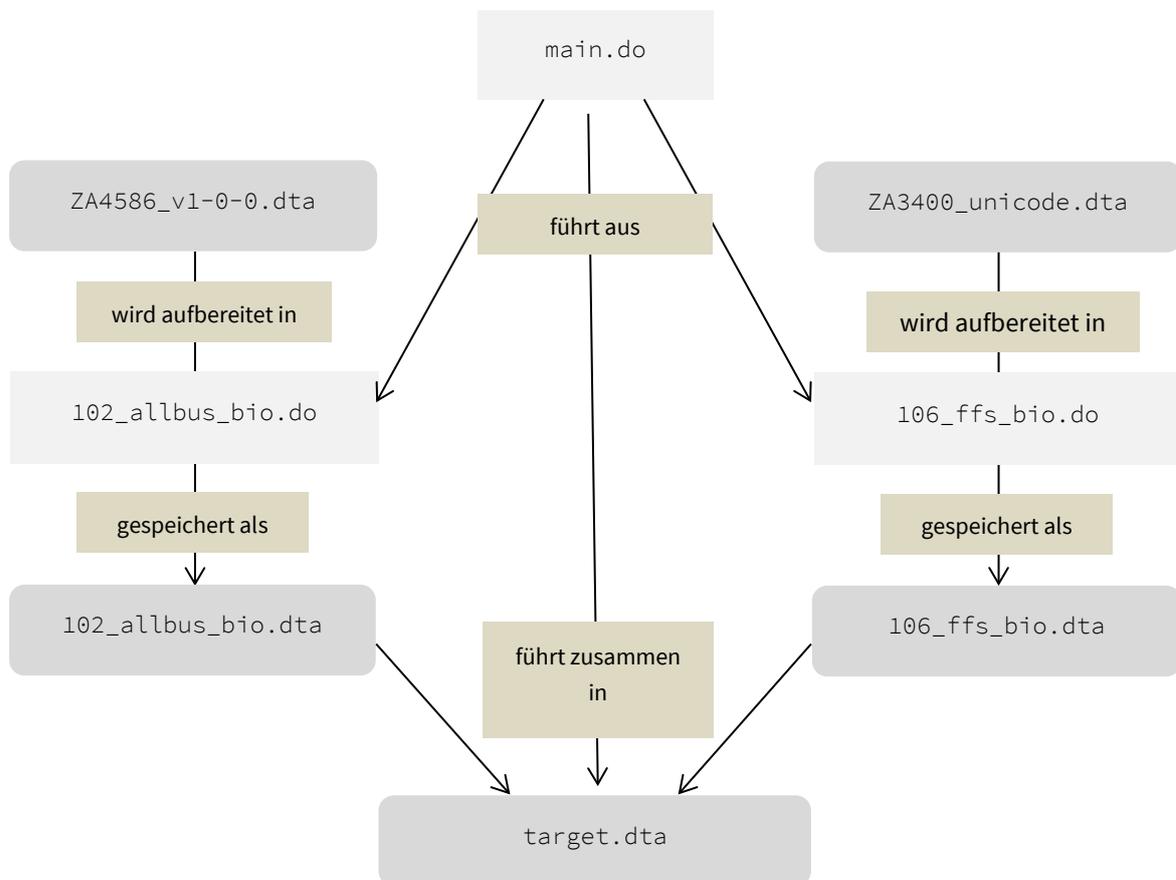


Abbildung A 1: Beispielhafter Prozess der Erstellung der kumulierten Partnerschaftsbiografien in Stata

Hierbei ruft das Main-File zunächst die Aufbereitungs-Do-Files `102_allbus_bio.do` und `106_ffs_bio.do` auf. Dort werden aus den Ausgangsdatensätzen (siehe Abbildung A 1, ganz oben) die Paarbiografiedatensätze `102_allbus_bio.dta` und `106_ffs_bio.dta` erstellt und im Ordner `temp` abgespeichert (siehe Abbildung A 2), sowie die HTML-Dokumentationen `coding_allbus.html` und `coding_ffs.html` unter `preparation_couplebio` erstellt.

In einem weiteren Schritt werden die beiden Zieldatensätze aus dem `temp`-Ordner miteinander kumuliert (append), als `target.dta` im Ordner `4_target` abgespeichert und anschließend aus dem `temp`-Ordner entfernt.

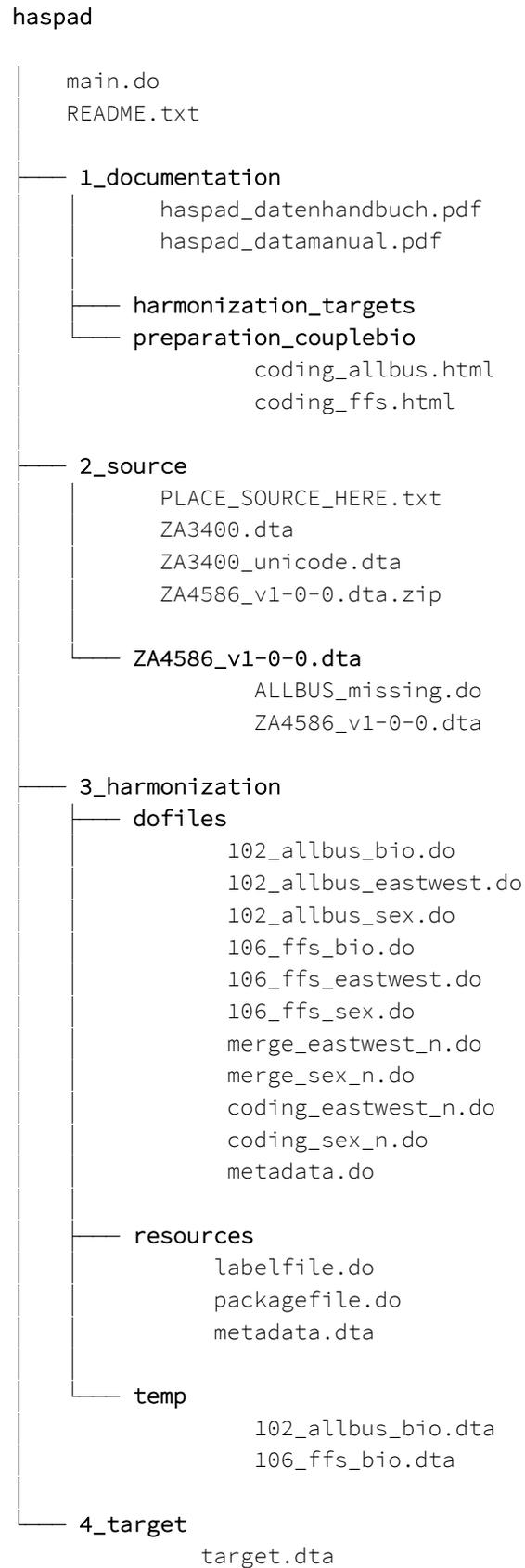
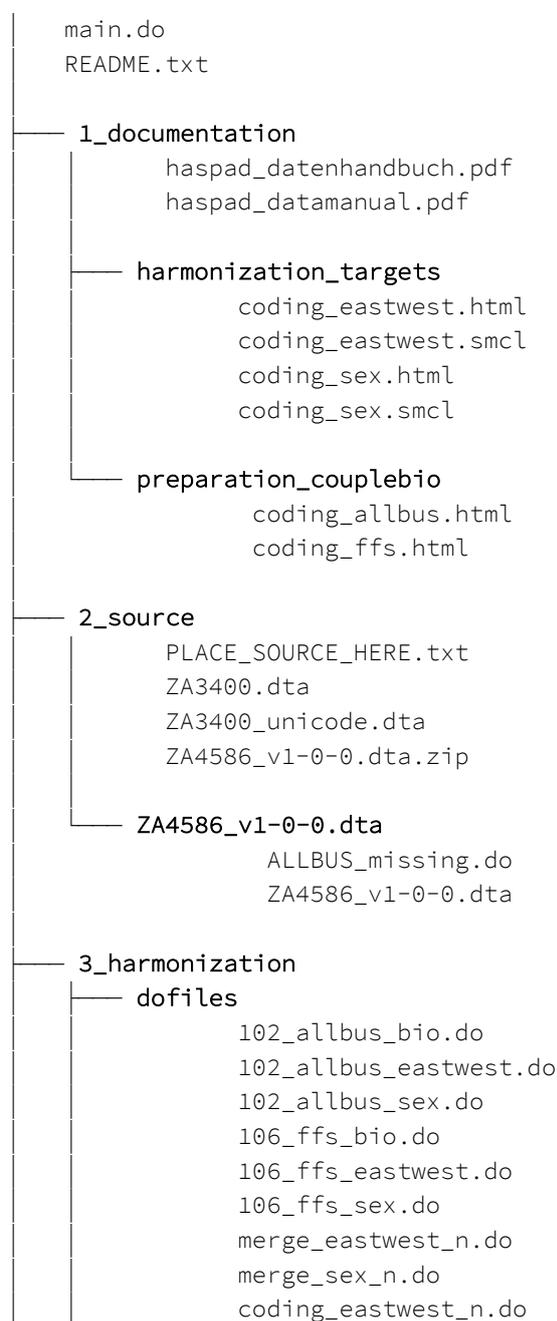


Abbildung A 2: Ordnerstruktur nach Aufbereitung des Biografiedatensatzes

Erstellung der harmonisierten Zusatzvariablen und Verknüpfen mit target.dta

Unter # RUN AND MERGE PREPFILES FROM HARMONIZEDVAR im Main-File werden die vom Nutzenden ausgewählten Zusatzvariablen erstellt, indem die Do-Files 102_allbus_sex.do, 102_allbus_eastwest.do, 106_ffs_sex.do und 106_ffs_eastwest.do angesprochen werden. Nach dem Ausführen dieser Harmonisierungsfiles werden die identisch benannten dta.-Files nach temp gespeichert und parallel die Harmonisierungsdokumentationen unter harmonization_targets im Ordner 1_documentation gespeichert (siehe Veränderung der Ordnerstruktur in *Abbildung A 3*).

haspad



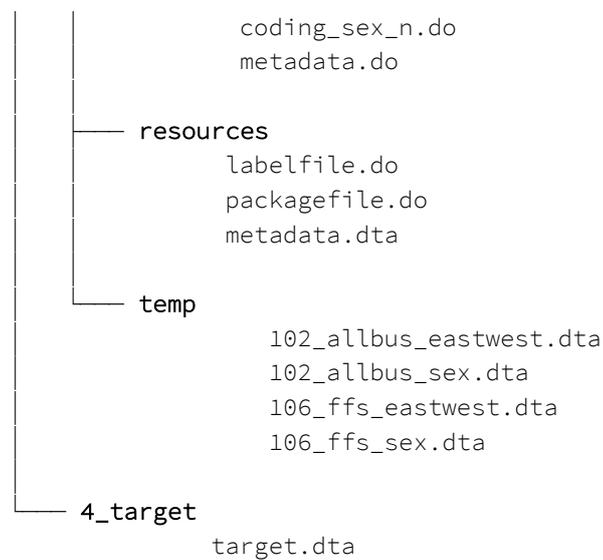


Abbildung A3: Ordnerstruktur nach Aufbereitung der Zusatzvariablen

Sind die Dateien im temp-Ordner erstellt, werden sie im nächsten Schritt mit den Paarbiografien in target.dta verknüpft (merge) und danach aus dem temp-Ordner entfernt. Der finale Datensatz target.dta wird nach 4_target abgelegt und steht nun dem Endnutzenden zu Verfügung.

A. 3 Dokumentationen der Ursprungsdatenquellen

In folgender *Tabelle A 2* wird ein Überblick über die Dokumentationen der Ursprungsdatenquellen gegeben. Die Seitenangabe in *Spalte 4* verweist auf den Abschnitt des Fragebogens, in welchem die Ankerperson Partnerschaften berichten konnte.

Tabelle A 2: Dokumentationen der Ursprungsdatenquellen

Kurztitel der Studie	Zitation der Dokumentation der Source-Variablen	Name des PDF-Dokuments / Link für automatischen 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://search.gesis.org/research_data/ZA5678
ALLBUS - Kumulation	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
Familiensurvey I. Welle 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
Familiensurvey I. Welle Ost 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
Familiensurvey II. Welle 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
Familiensurvey III. Welle 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
Mannheimer Scheidungsstudie	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

Kurztitel der Studie	Zitation der Dokumentation der Source-Variablen	Name des PDF-Dokuments / Link für automatischen Download
Deutscher 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
LV-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
LV-West II A - Persönliche Befragung	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
LV-West II T - Telefonische Befragung	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
LV-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
LV-DDR	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
LV-Ost 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
LV-Ost Panel, Geburtskohorten 1939-41, 1951-52, 1959-61	Mayer, Karl U. (2004): Ostdeutsche Lebensverläufe im Transformationsprozeß (Lebensverlaufsstudie LV-Ost Panel). GESIS Datenarchiv, Köln. Codebuch zur Studie. ZA3925, doi:10.4232/1.3925	ZA3925_cod.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=46258
LV-Ost Panel, Geburtskohorte 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	ZA3925_cod.pdf https://dbk.gesis.org/dbksearch/download.asp?db=D&id=46258
LV-West 64/71	Mayer, Karl U.; Kleinhenz, Gerhard (2004): Ausbildungs- und Berufsverläufe der Geburtskohorten	ZA3927_cod.pdf

Kurztitel der Studie	Zitation der Dokumentation der Source-Variablen	Name des PDF-Dokuments / Link für automatischen Download
	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/dbkse-arch/download.asp?db=D&id=46255
LV-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	ZA5099_cod.pdf https://dbk.gesis.org/dbkse-arch/download.asp?db=D&id=53821
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.	GermanyQreWave1.pdf https://www.ggp-i.org/sites/default/files/questionnaires/GermanyQre-Wave1.pdf
Generations and Gender Survey (2008)	Leven, Ingo (2009): Generations and Gender Survey 2. Welle. Methodenbericht der ersten Wiederholungsbefragung. 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
Sozio-oekonomisches Panel (biomarsy)	Hamjediers, M., Schmelzer, P., Geschke, A.-C. & SOEP Group (2020). SOEP-Core v35 - The couple history files BIOCOUPLM and BIOCOUPLY, and marital history files BIOMARSM and BIOMARSY SOEP Survey Papers 871: Series D. Berlin: DIW/SOEP.	diw_ssp0871.pdf http://panel.gsoep.de/soep-docs/surveypapers/diw_ssp0418.pdf#page=98

Kurztitel der Studie	Beziehungsbeginn (relbeg)	Beziehungsende (relend)	Kohabitationsbeginn (cohbeg)	Kohabitationsende (cohend)	Ehebeginn (marbeg)	Eheende (marend)	Tod des Partners (dop)
Deutscher Fertility and Family Survey	2 Only season information -10 Missing by study design	2 Only season information	2 Only season information 0 No imputation 1 Only year information	2 Only season information 0 No imputation 1 Only year information	2 Only season information 0 No imputation 1 Only year information	2 Only season information	2 Only season information 0 No imputation 1 Only year information
LV-West I	-10 Missing by study design	-10 Missing by study design	0 No imputation	0 No imputation 1 Only year information	0 No imputation 1 Only year information	0 No imputation 1 Only year information	0 No imputation 1 Only year information
LV-West II A - Persönliche Befragung	-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
LV-West II T - Telefonische Befragung	-10 Missing by study design	1 Only year information	0 No imputation 1 Only year information	-10 Missing by study design	0 No imputation 1 Only year information	1 Only year information	1 Only year information
LV-West III	-10 Missing by study design	1 Only year information	0 No imputation 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
LV-DDR	0 No imputation 1 Only year information	1 Only year information	0 No imputation 1 Only year information	-10 Missing by study design	0 No imputation 1 Only year information	1 Only year information	1 Only year information
LV-Ost 71	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	
LV-Ost Panel, Geburtskohorten 1939-41, 1951-52, 1959-61	-10 Missing by study design	-10 Missing by study design	0 No imputation 1 Only year information	0 No imputation 1 Only year information	0 No imputation	0 No imputation	

Kurztitel der Studie	Beziehungsbeginn (relbeg)	Beziehungsende (relend)	Kohabitationsbeginn (cohbeg)	Kohabitationsende (cohend)	Ehebeginn (marbeg)	Eheende (marend)	Tod des Partners (dop)
LV-Ost Panel, Geburtskohorte 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		
LV-West 64/71	0 No imputation 1 Only year information 2 Only season information	0 No imputation 1 Only year information 2 Only season information	0 No imputation 1 Only year information 2 Only season information	-10 Missing by study design	0 No imputation 1 Only year information 2 Only season information	0 No imputation 1 Only year information 2 Only season information	0 No imputation
LV-Panel 71	0 No imputation 1 Only year information 2 Only season information	0 No imputation 1 Only year information 2 Only season information	0 No imputation 1 Only year information 2 Only season information	-10 Missing by study design	0 No imputation 1 Only year information	-10 Missing by study design	0 No imputation
Generations and Gender Survey (2005)	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 1 Only year information	0 No imputation 1 Only year information	
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
Sozio-oekonomisches 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 Konstruktion der Original-Variablen der Biografiedaten

In folgender *Tabelle A 4* wird näher auf die Konstruktion der Originalvariablen des Biografiedatensatzes eingegangen.

Tabelle A 4: Konstruktion der Original-Variablen der Biografiedaten

Kurztitel der Studie	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
pairfam	„Jahrhundert-mo- nate“	„Jahrhundert-mo- nate“	„Jahrhundert-mo- nate“	„Jahrhundert-mo- nate“	„Jahrhundert-mo- nate“	„Jahrhundert-mo- nate“	„Jahrhundert-mo- nate“
ALLBUS - Kumulation	„-10 Missing by study design“	„-10 Missing by study design“	„-10 Missing by study design“	„-10 Missing by study design“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“
Familien-survey I. Welle West 1988	„Jahr“	Abgefragt in zwei getrennten Variab- len „Jahr“ in unverhei- rateten Partner- schaften „Jahr“ in verheirate- ten Partnerschaften	„Jahr“	"-10 Missing by study design"	„Monat“ „+“ „Jahr“	„Jahr“	„Jahr“ für verheira- tete Partnerschaf- ten "-10 Missing by study design" für unverheirateten Partnerschaften
Familien-survey I. Welle Ost 1990	„Jahr“	Abgefragt in zwei getrennten Variab- len „Jahr“ in unverhei- rateten Partner- schaften „Jahr“ in verheirate- ten Partnerschaften	„Jahr“	„Jahr“ in verheirate- ten Partnerschaften "-10 Missing by study design" in un- verheirateten Part- nerschaften	„Monat“ „+“ „Jahr“	„Jahr“	„Jahr“ in verheirate- ten Partnerschaften "-10 Missing by study design" in un- verheirateten Part- nerschaften

Kurztitel der Studie	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
Familien-survey II. Welle 1994	„Jahr“	Abgefragt in zwei getrennten Variab- len „Jahr“ in unverhei- rateten Partner- schaften „Jahr“ in verheirate- ten Partnerschaften	„Jahr“	„-10 Missing by study design“	„Monat“ „+“ „Jahr“	„Jahr“	„Jahr“ in verheirate- ten Partnerschaften "-10 Missing by study design" in un- verheirateten Part- nerschaften
Familien-survey III. Welle 2000	„Jahr“	Abgefragt in zwei getrennten Variab- len „Jahr“ in unverhei- rateten Partner- schaften „Jahr“ in verheirate- ten Partnerschaften	„Jahr“	„-10 Missing by study design“	„Monat“ „+“ „Jahr“	„Jahr“	„Jahr“ in verheirate- ten Partnerschaften „-10 Missing by study design" in un- verheirateten Part- nerschaften
Mannheimer Schei- dungs-studie	„Monat/Jahr“	„Monat/Jahr“	„Monat/Jahr“	„Monat/Jahr“	„Monat/Jahr“	„Monat/Jahr“	„Monat/Jahr“
Deutscher Fertility and Family Survey	„-10 Missing by study design“	„-10 Missing by study design“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“	„-10 Missing by study design“	„Monat“ „+“ „Jahr“ "-10 Missing by study design" für Partnerschaft mit Trennungsgrund, der nicht der Tod des Partners ist

Kurztitel der Studie	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
LV-West I	„-10 Missing by study design“	„-10 Missing by study design“	„Monat“ „+“ „Jahr“ für aktuelle unverheiratete Partnerschaft „-10 Missing by study design“ für vergangene Ehen und aktuelle Ehe	„Dauer des Getrenntlebens vor Scheidung“ + „+“ + „Scheidungsmonat“ „+“ „Scheidungs-jahr“ (nur für geschiedene in aktueller Ehe) „Monat“ „+“ „Jahr“ (nur für verheiratete, aber getrenntlebende Partnerschaften in aktueller Ehe) „-10 Missing by study design“ für vergangene Ehen „-10 Missing by study design“ für Verheiratete, Verwitwete, Ledige) in aktueller Ehe "-10 Missing by study design" für	„Monat“ „+“ „Jahr“ für vergangene Ehen und aktuelle Ehe „-10 Missing by study design“ für aktuelle unverheiratete Partnerschaft	„Monat“ „+“ „Jahr“ für vergangene Ehen und aktuelle Ehe, wenn Scheidung als Trennungsgrund -10 Missing by study design" für aktuelle unverheiratete Partnerschaft oder wenn Partner gestorben ist	„Monat“ „+“ „Jahr“ für vergangene Ehen und aktuelle Ehe, wenn Partner verstorben ist „-10 Missing by study design“ für aktuelle unverheiratete Partnerschaft

Kurztitel der Studie	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
LV-West II A - Persönliche Befragung	„-10 Missing by study design“	„Jahr“	„Monat“ „+“ „Jahr“ für aktuell unverheiratete Partnerschaften “-10 Missing by study design“ für aktuelle verheiratete Partnerschaften	aktuelle unverheiratete Partnerschaft „Dauer des Getrenntlebens vor Scheidung“ „+“ „Scheidungs-jahr“ für Partnerschaften mit Scheidung “Dauer des Getrenntlebens vor Scheidung“ für Rest	„Monat“ „+“ „Jahr“	„Jahr“	„Jahr“
LV-West II T - Telefonische Befragung	„-10 Missing by study design“	„Jahr“	„Monat“ „+“ „Jahr“ für aktuell unverheiratete Partnerschaften „-10 Missing by study design“ für Ehen	„-10 Missing by study design“	„Monat“ „+“ „Jahr“	„Jahr“	„Jahr“
LV-West III	„-10 Missing by study design“	„Jahr“ „-10 Missing by study design“ für aktuell geschiedene Ehen und retrospektive Ehen	„Monat“ „+“ „Jahr“ für unverheiratete aktuelle Partnerschaften und erste Ehe des Befragten	„Monat“ „+“ „Jahr“ für aktuell getrenntlebende Ehen „-10 Missing by study design“ für aktuell unverheiratete Partnerschaften	„Monat“ „+“ „Jahr“	„Jahr“	„Jahr“

Kurztitel der Studie	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“ für Rest	tete Partnerschaften und retrospektive Partnerschaften			
LV-DDR	“Wochen gekannt bevor Kohabitation” „+“ “Monate gekannt bevor Kohabitation” + „+“ + „Jahre gekannt bevor Kohabitation” „+“ „Monat des Zusammenzuges“ „+“ „Jahr des Zusammenzuges“	„Jahr“	„Monat“ „+“ „Jahr“	„-10 Missing by study design“	„Monat“ „+“ „Jahr“	„Jahr“	„Jahr“
LV-Ost 71	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“	„-10 Missing by study design“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“
		„-10 Missing by study design“ für andauernde Partnerschaften				„-10 Missing by study design“ für unverheiratete Partnerschaften	„-10 Missing by study design“ für Partnerschaft mit Trennungsgrund, der nicht der Tod des Partners ist
LV-Ost Panel, Geburtskohorten 1939-41, 1951-52, 1959-61	„-10 Missing by study design“	„-10 Missing by study design“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“
			„-10 Missing by study design“ für verheiratete, ohne	„-10 Missing by study design“ für verheiratete, ohne	„-10 Missing by study design“ für	„-10 Missing by study design“ für verheiratete, ohne	„-10 Missing by study design“ für Partnerschaft mit

Kurztitel der Studie	flag_relbeg_ori	flag_relend_ori	flag_cohbeg_ori	flag_cohend_ori	flag_marbeg_ori	flag_marend_ori	flag_dop_ori
			Partner, die keinen weiteren Partner in 2. Welle haben UND für Partnerschaften, die in Welle 2 gleichen Partner haben wie in Welle 1	Partner, die keinen weiteren Partner in 2. Welle haben	Personen, die verheiratet sind und mit gleichem Partner zusammenleben (gleichen, wie aus Welle 1) oder geschieden/verwitwet sind und keinen weiteren Partner haben	Partner, die keinen weiteren Partner in 2. Welle haben	Trennungsgrund, der nicht der Tod des Partners ist
LV-Ost Panel, Geburtskohorte 1929-31	„-10 Missing by study design“	„-10 Missing by study design“	„-10 Missing by study design“	„Monat“ „+“ „Jahr“ „-10 Missing by study design“ für geschiedene Partnerschaften und Partnerschaften mit Trennungsgrund „Tod des Partners“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“ „-10 Missing by study design“ für Partnerschaften mit Trennungsgrund „Tod des Partners“	„Monat“ „+“ „Jahr“ „-10 Missing by study design“ für Partnerschaft mit Trennungsgrund, der nicht der Tod des Partners ist
LV-West 64/71	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“ „-10 Missing by study design“ für geschiedene Partnerschaften und Partnerschaften mit Trennungsgrund „Tod des Partners“	„Monat“ „+“ „Jahr“	„-10 Missing by study design“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“	„Monat“ „+“ „Jahr“ „-10 Missing by study design“ für Partnerschaft mit Trennungsgrund, der nicht der Tod des Partners ist

Kurztitel der Studie	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“ für kohabitierende Partnerschaften und Ehen	„-10 Missing by study design“ für nicht-kohabitierende Partnerschaften	„-10 Missing by study design“ für nicht-kohabitierende Partnerschaften	„-10 Missing by study design“ für nicht-kohabitierende, unverheiratete Partnerschaften und kohabitierende, unverheiratete Partnerschaften	„-10 Missing by study design“ für nicht-kohabitierende, unverheiratete Partnerschaften und kohabitierende, unverheiratete Partnerschaften	„-10 Missing by study design“ für nicht-kohabitierende Partnerschaften
Sozio-oekonomisches Panel (biomarsy)	„-10 Missing by study design“	„Jahr“ + censor “original censor code”	„-10 Missing by study design“	„-10 Missing by study design“	„Jahr“ + censor “original censor code”	„Jahr“ + censor “original censor code”	„Jahr“ + 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 der Zelle: mögliche Ausprägungen der Ori-Variable

A. 6 Ausprägungen der Variable problem

In folgender *Tabelle A 5* wird näher auf die Ausprägungen der Variablen `problem` eingegangen.

Tabelle A 5: Ausprägungen der Variable `problem`

Wert	Wertelabel	Betreffende Umfrageprogramme	Erklärung
0	No data inconsistency	alle	Es konnten keine Inkonsistenzen oder Auffälligkeiten in den Originaldaten festgestellt werden.
1	Data inconsistency	alle	<p>Folgende 20 Checks wurden durchgeführt:</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>

Wert	Wertelabel	Betreffende Umfrageprogramme	Erklärung
2	Problems due to panel linkage (GLHS, GGS)	LVS, GGS	<p>Sobald eine oder mehrere der 20 Bedingungen in den Originaldaten zutreffen, wird der Wert <code>problem = „1 Data inconsistency“</code> vergeben.</p> <p>Die Ausprägung <code>problem = „2 Problems due to panel linkage (GLHS, GGS)“</code> wird aus zwei unterschiedlichen Gründen vergeben.</p> <p>In den Wiederholungsbefragungen der Lebensverlaufsstudien (Mayer 2004b, 2014a) und Generations and Gender Surveys (Wave 2, 2008) geben die Befragten Partnerschaften an, die sie bereits in der ersten Welle berichteten und bis zum Interviewdatum der Welle 1 andauerten. In <i>Abschnitt 5.7</i> werden die Plausibilitätschecks erläutert, die zur Identifikation derselben Partnerschaft berichtet in Welle 1 und nochmals in Welle 2 angewandt wurden. Nach der Feststellung gleicher Partnerschaften, die sowohl in Welle 1 als auch 2 berichtet wurden, bleiben Partnerschaften übrig, die <i>vor</i> dem Interviewzeitpunkt in Welle 1 begannen und die Befragten in der zweiten Welle angaben. Diese werden mit <code>problem = „2 Problem due to panel linkage (GLHS, GGS)“</code> markiert, da diese mit hoher Wahrscheinlichkeit dieselben Partnerschaften sind, die in der Erstbefragung berichtet wurden, aber aufgrund der Plausibilitätsprüfung nicht verknüpft werden können. Diese können durch <code>problem</code> identifiziert und aus dem Datensatz entfernt werden.</p> <p>Weiter gibt es Partnerschaften, berichtet in der Wiederholungsbefragung, bei welchen aufgrund fehlender Informationen des Partnerschaftsbeginns oder der Filterführung im Fragebogen nicht eindeutig ist, ob sie bereits in Welle 1 berichtet wurden.</p>
3	Higher order partnerships to same partner (pairfam, FFS, SOEP)	Pairfam, FFS, SOEP	<p>Im Sozio-oekonomischen Panel, Fertility and Family Survey (FFS) und pairfam gibt es den Sonderfall unabhängiger Partnerschaften zum gleichen Partner. In den Originaldaten wurden diese Partnerschaften über Duplikate der Partner-ID innerhalb der Partnerschaftsbiografie eines Befragten identifiziert. Weiterhin ist im Sozio-oekonomischen Panel der Ursprungsdatensatz so konstruiert, dass zwischen zwei Partnerschaften bei endgültigem Abschluss ein Single-Spell eingefügt ist und danach eine Partnerschaft mit dem gleichen Partner erneut beginnen kann.</p>
4	Data inconsistency + problems due to panel linkage (GLHS, GGS)	LVS, GGS	<p>Da sowohl <code>problem = „1 Data inconsistency“</code> als auch <code>problem = „2 Problems due to panel linkage (GLHS, GGS)“</code> gemeinsam auftreten können, wird <code>problem = „4 Data inconsistency + problems due to panel linkage (GLHS, GGS)“</code> für diese Kombination vergeben.</p>

Wert	Wertelabel	Betreffende Umfrageprogramme	Erklärung
5	Data inconsistency + higher order partnerships to same partner (pairfam, FFS)	pairfam, FFS	Da sowohl problem = „1 Data inconsistency“ als auch problem = „3 Higher order partnerships to same partner (pairfam, FFS)“ gemeinsam auftreten können, wird problem = „5 Data inconsistency + higher order partnerships to same partner (pairfam, FFS)“ für diese Kombination vergeben.
6	Duplicate Multi-Actor-Design (SHARE)	SHARE	Aufgrund des Multi-Actor- Designs des SHARE befinden sich im Datensatz Partnerschaften, die aus der Perspektive beider Befragter berichtet wurden. Damit Nutzende für ihre Analysen eine der beiden Partnerschaften entfernen können, sind die Partnerschaften für je einen Partner oder Partnerin mit dem Wert problem = „6 Duplicate Multi-Actor-Design (SHARE)“ gekennzeichnet. Für heterosexuelle Partnerschaften wurde der Wert 6 Männern und für homosexuelle Partnerschaften Befragten mit einer höheren aid zugewiesen.