



pairfam Quick Guide


Release 11.0

pairfam Group

Mai 2020

Gefördert als Langfristprojekt der Deutschen Forschungsgemeinschaft (DFG)

Der **Quick Guide** ist eine Kurzanleitung zum Umgang mit den komplexen Daten des Beziehungs- und Familienpanel pairfam. Die Hinweise auf den folgenden Seiten sollen den Einstieg in die Datenanalyse für Nutzer erleichtern, indem auf wesentliche Dokumentationsmaterialien verwiesen wird sowie Erläuterungen zum Aufbau und zur Struktur der Daten gegeben werden. Ein Schwerpunkt bildet das für nahezu alle Analysen unverzichtbare Zusammenfügen einzelner Datensätze.

Mit einem  Symbol gekennzeichnete Wörter verlinken zur jeweiligen Seite unserer [Homepage](#), auf der entsprechende Dokumente bzw. weiterführende Informationen verfügbar sind. Zusätzlich zum Scientific-Use-File können sämtliche Dokumentationsmaterialien auch von der Homepage des Beziehungs- und Familienpanel unter www.pairfam.de heruntergeladen werden.

Über aktuelle Entwicklungen und Hinweise zum Umgang mit den Daten sowie über relevante Veranstaltungen und neuere Publikationen informiert der regelmäßig erscheinende [Newsletter](#). Darüber hinaus können Sie sich mit Fragen und Anregungen jederzeit an den [Nutzerservice](#) von pairfam wenden (support@pairfam.de).

Inhalt

1	Erste Orientierung.....	1
2	Hinweise zu den Daten.....	2
2.1	Datensätze.....	3
2.2	Variablen	5
2.3	Analysen	6
2.4	Kombinieren von Datensätzen.....	7
2.5	pairfam in der Lehre	11
3	Publikationen.....	12
3.1	Literaturdatenbank.....	12
3.2	Zitation	12

1 Erste Orientierung

Das Beziehungs- und Familienpanel pairfam ist eine von der Deutschen Forschungsgemeinschaft (DFG) geförderte, multidisziplinäre Längsschnittstudie zur Untersuchung partnerschaftlicher und familialer Lebensformen. Die Studie startete 2008/09 mit einer Ausgangsstichprobe von 12.402 zufällig ausgewählten Ankerpersonen, die jährlich wiederholt interviewt werden. Im Rahmen des sogenannten Multi-Actor-Designs werden parallel auch die jeweiligen Partner, Eltern (bis Welle 8) und Kinder dieser Ankerpersonen („Alteri“) befragt. Kinder der Ankerpersonen, die an der Kinderbefragung teilnehmen, werden nach ihrem 16. Geburtstag als neue Ankerpersonen („Step-ups“) in das Panel aufgenommen.

Sowohl in Vorbereitung auf die eigenen Analysen als auch während der Nutzung der Daten empfiehlt sich immer wieder ein Blick in die [Dokumentation](#) von pairfam:

- ☞ [Referenzpapier:](#) Einführung in die konzeptionellen und methodischen Grundlagen des Beziehungs- und Familienpanels von Huinink et al. (2011).
- ☞ [Data Manual:](#) Erläuterungen zur Struktur der Datensätze und zur Aufbereitung der Daten, inklusive Beschreibung aller generierten Variablen und Datensätze, außerdem Hinweise zu Gewichtungsfaktoren, zu Inkonsistenzen in den Daten (Flag- bzw. Tag-Variablen) und zu Modifikationen zwischen den Releases.
- ☞ [Methodenberichte:](#) Dokumentation der Feldarbeit durch das Befragungsinstitut mit Informationen zu Stichproben und Rücklaufquoten, Incentivierung und Panelpflege sowie weiteren Maßnahmen der Qualitätssicherung. Eine Zusammenfassung der Methodenberichte ist das ☞ [Technical Paper # 01](#).
- ☞ [Technical Papers:](#) Sammlung von Hilfestellungen für die Datenaufbereitung und -analyse sowie Hinweisen zu Besonderheiten in den Daten.

Mit der Veröffentlichung von **Release 11.0** im Frühjahr 2020 stehen mittlerweile die Daten der ersten elf Befragungswellen von pairfam für wissenschaftliche Analysen zur Verfügung. Um sich einen Überblick über die im Frageprogramm enthaltenen Items und Konstrukte zu verschaffen, haben Sie verschiedene Möglichkeiten:

- ☞ [Variablenübersichten:](#) detaillierter Überblick über sämtliche verfügbaren Items der jeweiligen Wellen, thematisch sortiert, mit Angaben zu Variablenname und -label und Hinweisen zu Modifikationen zwischen den Wellen; separate Darstellung für die einzelnen Befragtengruppen.
- ☞ [Codebücher:](#) vollständige Informationen zum gesamten Befragungsinstrumentarium der verfügbaren Wellen: Darstellung von Fragetexten, Antwortvorgaben, Filterführungen und Interviewerhinweisen in der Reihenfolge der jeweiligen Befragung; separate Darstellung für die einzelnen Wellen und Befragtengruppen.
- ☞ [Scales Manual:](#) Information zu den in pairfam verwendeten Skalen und ihrer Herkunft. Weiterhin wird die Bildung von Indizes bzw. Skalen auf Grundlage der erhobenen Variablen, inklusive der erreichten internen Konsistenz erläutert.

2 Hinweise zu den Daten

Die aufbereiteten Daten der Befragten werden ein Jahr nach dem Ende der jeweiligen Feldphase als Scientific-Use-File veröffentlicht. Bei einem Intervall ca. zwölf Monaten zwischen den Erhebungswellen bedeutet dies ein jährliches Update. Aktuell erhältlich ist Release 11.0 mit den Daten der ersten elf Wellen des Beziehungs- und Familienpanels pairfam. Ebenfalls im Scientific-Use-File enthalten sind die Daten der drei Wellen der Zusatzstudie ☞ [DemoDiff](#), einer parallelen Befragung ostdeutscher Ankerpersonen der Kohorten 1971-73 und 1981-83 sowie deren jeweiliger Partner. Die erste DemoDiff-Erhebungswelle fand zeitverzögert um ein Jahr nach dem Start von pairfam statt, die letzte DemoDiff-Welle zeitgleich mit der vierten Welle von pairfam. Mit Welle 5 wurden die Befragten von DemoDiff in die Stichprobe von pairfam übernommen. In Welle 11 kommt nun eine weitere Stichprobe mit etwa 5.000 Befragten hinzu, mit der die beiden Kohorten 1981-83 und 1991-93 aufgefrischt und eine neue Kohorte 2001-03 hinzugefügt wurde. Auf der pairfam-Website finden Sie eine Übersicht, wie sich die Fallzahlen der einzelnen ☞ [Stichproben](#) über die Wellen entwickelt haben.

2.1 Datensätze

Der Scientific-Use-File enthält eine Vielzahl von Datensätzen, da für jede Befragten-Gruppe jeweils ein separater Querschnittsdatensatz pro Erhebungswelle ausgeliefert wird. Die Verknüpfung dieser Datensätze, z. B. für längsschnittliche oder dyadische Analysen, wird durch eine einheitliche Identifikatorvariable gewährleistet (siehe 2.2 und 2.4). Die Daten von DemoDiff sind teilweise als eigene Datensätze im SUF enthalten, teilweise in den pairfam-Datensätzen integriert. Die Daten der Aufstockungsstichprobe sind in den Datensätzen jeweils enthalten. Identifiziert werden können die verschiedenen Stichproben über die Variable *sample*.

Anker	Partner	Kind	Eltern	Erziehung	Erziehung Jugendlicher	Erziehung U6 (Partner)
Welle 1 (2008/2009)						
anchor1	partner1	-	-	-	-	-
anchor1_DD	partner1_DD	-	-	-	-	-
Welle 2 (2009/2010)						
anchor2	partner2	child2	parent2	parenting2	-	-
anchor2_DD	-	-	-	-	-	-
Welle 3 (2010/2011)						
anchor3	partner3	child3	parent3	parenting3	-	-
Welle 4 (2011/2012)						
anchor4	partner4	child4	parent4	parenting4	-	-
Welle 5 (2012/2013)						
anchor5	partner5	child5	parent5	parenting5	-	-
Welle 6 (2013/2014)						
anchor6	partner6	child6	parent6	parenting6	-	-
Welle 7 (2014/2015)						
anchor7	partner7	child7	parent7	parenting7	-	-
Welle 8 (2015/2016)						
anchor8	partner8	child8	parent8	parenting8	-	-
Welle 9 (2016/2017)						
anchor9	partner9	child9	-	parenting9	paya9	-
Welle 10 (2017/2018)						
anchor1	partner10	child10	-	parenting10	paya10	-
Welle 11 (2018/2019)						
anchor11	partner11	child11	-	parenting11	paya11	parenting-U6partner11

Darüber hinaus gibt es mehrere generierte Datensätze mit biografischen Informationen, die wellenübergreifend verfügbar sind ([☞ Data Manual](#), Abschnitt 4.6).

Generierte Datensätze	Datei	Inhalt
Aktivitätenbiografie	bioact bioact_rtr	prospektive & retrospektive Informationen zu Tätigkeiten der Ankerperson: Bildung, Ausbildung und Erwerbstätigkeit
Fertilitätsbiografie	biochild	prospektive & retrospektive Informationen zu allen Kindern der Ankerperson: Geburt und Zusammenleben
Mobilitätsbiografie	biomob_	prospektive & retrospektive Informationen zur Mobilität der Ankerperson: Wohnsitze, Umzüge, Auszug aus dem Elternhaus
Elternbiografie	bioparent	prospektive & retrospektive Informationen zu allen Elternteilen des Ankers: Art der Elternschaft, Kohabitation, Familienstand
Partnerschaftsbiografie	biopart	prospektive & retrospektive Informationen zu Partnerschaften des Ankers ab 14. Lebensjahr: Beziehung, Kohabitation, Ehe
Haushaltsraster	household	prospektive Informationen zum Haushalt der Ankerperson (Wellen 1-3): Wohnsituation und Haushaltsmitglieder
Multi-Aktor-Überblick	overview_multi_actor	prospektive Informationen zur Teilnahme der Alteri an pairfam

Die Daten der Step-ups sind in einem eigenen Ordner abgelegt. Die Struktur der Step-up-Datensätze ist in der folgenden Tabelle dargestellt.

Step-up-Datensätze	Anker	Partner
Welle 4	stepup_anchor4	stepup_partner4
2011/12	stepup_transition_anchor4	-
Welle 5	stepup_anchor5	stepup_partner5
2012/13	stepup_transition_anchor5	-
Welle 6	stepup_anchor6	stepup_partner6
2013/14	stepup_transition_anchor6	-
Welle 7	stepup_anchor7	stepup_partner7
2014/15	stepup_transition_anchor7	-
Welle 8	stepup_anchor8	stepup_partner8
2015/16	stepup-transition_anchor8	-
Welle 9	stepup_anchor9+transition.dta	stepup_partner9
2016/17	-	-
Welle 10	stepup_anchor10+transition.dta	stepup_partner10
2017/18	-	-
Welle 11	stepup_anchor11+transition.dta	stepup_partner11
2018/19	-	-
generierte Datensätze	stepup_biochild	prospektive & retrospektive Informationen zu allen Kindern der Ankerperson
	stepup_biopart	prospektive & retrospektive Informationen zu Partnerschaften des Ankers ab 14. Lebensjahr

2.2 Variablen

Wir versuchen den Nutzern eine anwendungsfreundliche und verständliche Datenressource bereitzustellen. Dazu zählt unter anderem eine Systematik bei der **Bezeichnung der Variablen** ([☞ Data Manual](#), Abschnitt 2.3). Dieser Systematik zufolge tragen inhaltlich identische Variablen, die in mehreren Wellen abgefragt wurden, auch den gleichen Variablennamen. Die Variablen in den Alteri-Datensätzen sind jeweils durch ein entsprechendes Präfix gekennzeichnet:

- *Beispiel: Variablenname Geschlecht: Anker = sex | Partner = **psex** | Eltern = **parsex** | Kinder = **csex***
- *Variablen aus dem PAYA-Fragebogen haben das Präfix paya, Variablen aus dem U6-Erziehungsfragebogen für Partner haben das Präfix p.*

Von wesentlicher Bedeutung für die Arbeit mit den Daten sind die **Identifikationsvariablen** ([☞ Data Manual](#), Abschnitt 2.2), die eine eindeutige Zuordnung der jeweiligen Angaben zu den entsprechenden Befragungspersonen ermöglichen. Insbesondere für das Zusammenführen der Einzeldatensätze sind die Identifikatoren unverzichtbar. Die Benennung folgt dem üblichen System der Variablenbezeichnung:

- *ID-Variablen: Anker = id | Partner = pid | Eltern = parid | Kinder = cid(x) | Geschwister = sibid(x)
Mutter = mid | Vater = fid | Stiefmutter = smid | Stiefvater = sfid*

Neben den eigentlichen Befragungsdaten enthalten die Datensätze eine Reihe zusätzlicher Variablen, die im Zuge der Datenaufbereitung erstellt wurden. Diese **generierten Variablen** ([☞ Data Manual](#), Abschnitt 4.3) geben Auskunft über häufig in Analysen verwendete Merkmale. Mit der Bereitstellung der generierten Variablen entfällt für die Nutzer der oftmals erhebliche Aufwand, die entsprechenden Informationen selbst aus den einzelnen Angaben der Befragten zusammenzustellen:

- *Beispiele für generierte Variablen Anker: Kohortenzugehörigkeit = cohort
Migrationshintergrund = ethni, migstatus
Familienstand, Beziehungsdauer = marstat, reldur
Anzahl aller Kinder = nkids
Bildungsklassifikation = isced, casmin, yeduc
Berufsklassifikation = egp, isei, kldb2010, isco08, mps
Einkommen = hhincgee, incnet, hhincnet*

Zur Korrektur der durch das Stichprobendesign und durch Verweigerungen verursachten Diskrepanzen zwischen Stichprobe und Gesamtbevölkerung innerhalb der drei betrachteten Geburtskohorten stehen vier **Gewichtungsvariablen** ([☞ Data Manual](#), Abschnitt 4.6) in den Ankerdatensätzen zur Verfügung. Über die Anwendung der Gewichtungsvariablen ist je nach Art und Ziel der Analyse zu entscheiden. In Welle 5 wurden neue Gewichte zur Verfügung gestellt, die eine gemeinsame Gewichtung von pairfam und DemoDiff ermöglichen.

Gewichtungsvariable	Sample	Beschreibung	Anwendung
dweight	<i>pairfam</i>	Designgewicht: Korrektur für disproportionale Größe der Bruttostichproben für die drei Kohorten	
ddweight	<i>DemoDiff</i>		
d1weight	<i>pairfam & DemoDiff</i>		
psweight	<i>pairfam</i>	Post-Stratification-Gewicht: Korrektur für systematisches Non-Response in Welle 1 durch Anpassung der Randhäufigkeiten an Mikrozensus 2008	Stata: svyset [pweight=weightvar] svy: command command [pweight=weightvar], options
caweight	<i>DemoDiff</i>		
ca1weight	<i>pairfam & DemoDiff</i>		
dxpsweight	<i>pairfam</i>	Kombiniertes Gewicht: simultane Korrektur für Non-Response- und Designeffekte in Welle 1	
ddcaweight	<i>DemoDiff</i>		
d1ca1weight	<i>pairfam & DemoDiff</i>		
lweight	<i>pairfam, DemoDiff, pairfam & DemoDiff</i>	Längsschnittgewicht: Korrektur für Panelausfall über kohortenspezifische Wiederbefragungsbereitschaft	SPSS: WEIGHT BY weight- var . COMMAND WEIGHT OFF .
rpsweight1 rpsweight2	<i>Aufstockung</i>	Post-Stratification-Gewicht: Korrektur für systematische Nonresponse in Welle 11 (nur Aufstockung) durch Anpassung der Randhäufigkeiten an Mikrozensus	

2.3 Analysen

Wichtig ist, dass man über einige „Fallstricke“ der pairfam Daten Bescheid weiß. Deshalb gib es im Kapitel 3 des [Data Manual](#) einen Überblick über „**Tipps, Tricks und Probleme der pairfam Daten**“. Jeder Nutzer sollte sich dieses Kapitel vor (!) den ersten Analysen ansehen.

Grundsätzlich gilt bei der Arbeit mit den pairfam-Daten die Empfehlung, sich vor jeder Analyse eine Häufigkeitsverteilung für alle einbezogenen Variablen anzeigen zu lassen. Auf diese Weise lässt sich unter anderem sicherstellen, dass fehlende Werte in den Angaben der Befragten auch als solche in die Berechnungen eingehen. Da es jeweils Sache der Nutzer ist zu entscheiden, welche Angaben als **Missings** ([Data Manual](#)) zu behandeln sind, wurden fehlende bzw. unvollständige Informationen nach einem einheitlichen Schema mit negativen Werten kodiert. Um entsprechende Fälle von der Analyse auszuschließen, ist es erforderlich, die negativen Werte vorab als Missings zu definieren.

Kodierung	Bedeutung	Definition als Missing
-1	Weiß nicht	
-2	Keine Angabe / Das möchte ich nicht beantworten	
-3	Trifft nicht zu (durch Filterführung ausgeschlossen)	Stata: mvdecode _all, mv(-9/-1)
-4	Fehleingabe / Filterfehler	mvdecode _all, mv(-1=.a \...\ -9=.i)
-5	Inkonsistenter Wert	mvdecode varlist, mv(-1=.a \...\ -9=.i)
-6	Offene Angabe nicht lesbar	
-7	Unvollständige Angabe	
-9	Ungültige Mehrfachantwort	
-10	Nicht in DemoDiff: Variable nur in pairfam enthalten	SPSS: MISSING VALUES ALL (-9 THRU -1).
-11	Nicht in pairfam: Variable nur in DemoDiff enthalten	MISSING VALUES varlist (-9 THRU -1).
-12	Nicht in Aufstockung (W11)	RECODE varlist (-9 THRU -1 = SYSMIS).
-13	Nicht für die Step-ups	

Für den schnellen Einstieg in die Datenanalyse stehen zudem zahlreiche **Syntax-Dateien** zur freien Verfügung. Die standardmäßig für Stata erstellten Befehlszeilen ermöglichen zum einen das Nachvollziehen ausgewählter Analyseverfahren. Zum anderen dienen sie der Transparenz, indem sämtliche Schritte der Generierung von Variablen und Datensätzen durch entsprechende Syntax-Dateien dokumentiert werden. Alle Befehle sind hinreichend kommentiert, so dass sie auch als Vorlage für eigene Analysevorhaben genutzt werden können.

Skalenbildung: Syntax-Dateien zur Bildung der im Skalenhandbuch beschriebenen Skalen bzw. Indexvariablen; für Stata und für SPSS verfügbar; separat für jede Befragten-Gruppe und jede Welle; nur im Scientific-Use-File verfügbar.

Generierte Variablen: Syntax-Dateien zur Erstellung der mit dem Scientific-Use-File ausgelieferten generierten Datensätze (siehe 2.1) sowie der in den Datensätzen enthaltenen generierten Variablen (siehe 2.2); nur für Stata verfügbar; nur im Scientific-Use-File verfügbar.

Quick Starts: Syntax-Dateien zur Erleichterung des Einstiegs in die Datenanalyse, Anleitungen und Beispiele für das Zusammenführen von Datensätzen/Matching (siehe 2.4); für den Nachvollzug der im [Technical Paper 02](#) beschriebenen Verfahren sowie für die Durchführung von Panelregressionen und Ereignisdatenanalysen; aktuell nur für Stata verfügbar; nur über die Website verfügbar.

2.4 Kombinieren von Datensätzen

Für die meisten Analysen ist es erforderlich, Daten aus den verschiedenen Datensätzen zusammenzuspielen. Dazu empfiehlt es sich, zunächst die relevanten Variablen aus den einzelnen Datensätzen zu bestimmen. Aufgrund der Komplexität und des Umfangs der pairfam-Daten erleichtert es die Analysen erheblich, wenn die zusammengespielten Datensätze nur diejenigen Variablen enthalten, die benötigt werden. Beim Zusammenführen von Daten („Matching“) sind je nach Art der beabsichtigten Analyse mehrere Entscheidungen zu treffen.

- Ziel Datensatz:
 - Akteurspanel = Matching der Daten einer Befragtengruppe aus mehreren Wellen
 - Multi-Actor-Querschnitt = Matching der Daten verschiedener Befragtengruppe aus einer Welle
 - Multi-Actor-Panel = Matching der Daten verschiedener Befragtengruppen aus mehreren Wellen
- Datenformat:
 - WIDE = die zusätzlichen Daten werden in der jeweiligen Datenzeile zugespielt
 - LONG = die zusätzlichen Daten werden als neue Datenzeilen zugespielt
 - LONG-WIDE = Kombination aus beiden Datenformaten
- Umgang mit Nonresponse:
 - balanciert = Ziel Datensatz enthält nur Fälle, die überall gültige Angaben enthalten
 - unbalanciert = Ziel Datensatz enthält alle Fälle der gematchten Datensätze (z. B. Panalaussteiger)

Aus den genannten Optionen ergeben sich zahlreiche Kombinationsmöglichkeiten. Aus diesem Grund ist es erforderlich, sich vor dem Zusammenführen von Einzeldatensätzen sowohl über die Analyseverfahren und -ziele als auch über die dafür erforderliche Datenstruktur Klarheit zu verschaffen. Die auf den folgenden Seiten dargestellten Beispiele illustrieren einige gebräuchliche Matching-Prozeduren. Ausführlich kommentierte Stata-Syntaxen finden sich als [Quick Start](#) auf der Webseite von pairfam.

Auf eine Besonderheit der Daten des Beziehungs- und Familienpanels sei in diesem Zusammenhang nochmals hingewiesen: Inhaltlich gleiche Variablen innerhalb einer Befragtengruppe sind über die Wellen mit identischen Variablennamen versehen. Für das Zusammenführen entsprechender Daten im WIDE-Format bedeutet dies, dass die zu matchenden Variablen vorab umbenannt werden müssen, da das Matchen mehrere Variablen mit dem gleichen Namen in einem Datensatz in aller Regel zu Problemen führt. Die Änderung von Variablennamen ist auch erforderlich, wenn inhaltlich gleiche Informationen aus verschiedenen Befragtengruppen (z. B. Anker und Partner) im LONG-Format zusammengespielt werden sollen. In diesem Fall müssen die entsprechenden Variablennamen harmonisiert, also angeglichen werden. Die folgenden Beispiele veranschaulichen beide Strategien.

Die hier dargestellten Beispiele sowie die QuickStart-Datei „Matching“ beziehen sich auf das Statistikprogramm Stata. Die für SPSS erforderlichen Befehle sind im Wesentlichen ADD FILES, MATCH FILES und RENAME VARIABLES.

Beispiel 1: Anker-Akteurspanel im WIDE-Format

Anker (Welle 1)						Anker (Welle 2)					
	id	wave	var1	var2	var3		id	wave	var1	var2	var3
1	100000	1	1	2	3	1	100000	2	2	3	4
2	101000	1	5	4	3	2	101000	2	4	3	2
3	102000	1	-3	3	-3	3	102000	2	2	-3	2

Ziel Datensatz: Anker (Wellen 1+2)

	id	wave_01	var1_01	var2_01	var3_01	wave_02	var1_02	var2_02	var3_02
1	100000	1	1	2	3	2	2	3	4
2	101000	1	5	4	3	2	4	3	2
3	102000	1	-3	3	-3	2	2	-3	2

Hinweis:

- Umbenennung der gleichlautenden Variablennamen erforderlich
- keine Umbenennung der Schlüsselvariable id → muss in den zu matchenden Datensätzen jeweils identisch benannt sein

DO-File für Stata:

```

use id wave var1 var2 var3 using anchor1, clear // Auswahl der Variablen aus Anker W1
foreach x in wave var1 var2 var3 {           // Variablen für Umbenennung (NICHT id)
    rename `x' `x'_01                         // Umbenennung durch Postfix „_01“
}                                              // Abschluss der Schleife
save anchor1_01, replace                       // Speicherung des Zwischendatensatzes

use id wave var1 var2 var3 using anchor2, clear // Vorgehen analog für Anker W2
foreach x in wave var1 var2 var3 {
    rename `x' `x'_02
}
save anchor2_02, replace

use anchor1_01, clear                          // Aufruf Zwischendatensatz Anker W1
merge 1:1 id using anchor2_02                // Anfügen der Daten aus Zwischendatensatz Anker W2
keep if _merge==3                              // Erzeugung eines balancierten Datensatzes
drop _merge                                     // Löschen der Hilfsvariable _merge
erase anchor1_01.dta                           // Löschen der Zwischendatensätze
erase anchor2_02.dta

```

Beispiel 2: Anker-Akteurspanel im LONG-Format

Anker (Welle 1)

	id	wave	var1	var2	var3
1	100000	1	1	2	3
2	101000	1	5	4	3
3	102000	1	-3	3	-3

Anker (Welle 2)

	id	wave	var1	var2	var3
1	100000	2	2	3	4
2	101000	2	4	3	2
3	102000	2	2	-3	2

Zieldatensatz: Anker (Wellen 1+2)

	id	wave	var1	var2	var3
1	100000	1	1	2	3
2	100000	2	2	3	4
3	101000	1	5	4	3
4	101000	2	4	3	2
5	102000	1	-3	3	-3
6	102000	2	2	-3	2

Hinweis:

- keine Umbenennung der gleichlautenden Variablennamen erforderlich
- auch Vorgehensweise für Verknüpfung mit Daten aus dem DemoDiff-Projekt

DO-File für Stata:

```

use id wave var1 var2 var3 using anchor1, clear // Auswahl der Variablen aus Anker W1
append using anchor2, keep (id wave var1 var2 var3) // Anfügen der ausgewählten Variablen Anker W2
sort id wave // Sortierung nach ID und Welle

```

Beispiel 3: Multi-Actor-Querschnitt mit Anker und Partner im WIDE-Format

Anker (Welle 1)

	id	wave	var1	var2	var3
1	100000	1	1	2	3
2	101000	1	5	4	3
3	102000	1	-3	3	-3

Partner (Welle 1)

	id	pid	wave	pvar1	pvar2	pvar3
1	100000	100101	1	2	3	4
2	101000	101101	1	4	3	2
3	102000	102101	1	2	-3	2

Zieldatensatz: Anker + Partner (Welle 1)

	id	wave	var1	var2	var3	pid	pvar1	pvar2	pvar3
1	100000	1	1	2	3	100101	2	3	4
2	101000	1	5	4	3	101101	4	3	2
3	102000	1	-3	3	-3	102101	2	-3	2

Hinweis:

- keine Umbenennung der anderslautenden Variablennamen erforderlich

DO-File für Stata:

```
use id wave var1 var2 var3 using anchor1, clear // Auswahl der Variablen aus Anker W1
merge 1:1 id using partner1, keepusing (pid pvar1 pvar2 pvar3) // Anfügen der Variablen aus Partner W1
keep if _merge==3 // Erzeugung eines balancierten Datensatzes
drop _merge // Löschen der Hilfsvariable _merge
```

Beispiel 4: Multi-Actor-Querschnitt mit Anker und Partner im LONG-Format

Anker (Welle 1)

	id	pid	wave	var1	var2
1	100000	100101	1	1	2
2	101000	101101	1	5	4
3	102000	102101	1	-3	3

Partner (Welle 1)

	id	pid	wave	pvar1	pvar2
1	100000	100101	1	2	3
2	101000	101101	1	4	3
3	102000	102101	1	2	-3

Zieldatensatz: Anker + Partner (Welle 1)

	id	pid	anker	wave	var1	var2
1	100000	100101	1	1	1	2
2	100000	100101	0	1	2	3
3	101000	101101	1	1	5	4
4	101000	101101	0	1	4	3
5	102000	102101	1	1	-3	3
6	102000	102101	0	1	2	-3

Hinweis:

- Anpassung der anderslautenden Variablennamen von Partnerdaten an Ankerdaten erforderlich
- keine Umbenennung der Schlüsselvariable id → muss in den zu matchenden Datensätzen jeweils identisch benannt sein
- Erzeugung einer Herkunftsvariable zur Identifikation des Ursprungs der Angaben im Anker- oder Partnerdatensatz sinnvoll

DO-File für Stata:

```
use id pid wave pvar1 pvar2 using partner1, clear // Auswahl der Variablen aus Partner W1
foreach x in var1 var2 { // Variablen für Umbenennung (NICHT id)
    rename p`x' `x' // Umbenennung = Angleichung an Anker W1
} // Abschluss der Schleife
```

```
append using anchor1, keep (id pid wave var1 var2) generate (anker) // Matching + Herkunftvariable anker
order id pid anker wave var1 var2 // Reihenfolge der Variablen
gsort id -anker // Sortierung nach ID & Herkunftsvar.
```

Beispiel 5: Multi-Actor-Querschnitt mit Anker und Eltern im LONG-WIDE-Format

Anker (Welle 2)

	id	wave	var1	var2
1	100000	2	1	2
2	101000	2	5	4
3	102000	2	-3	3
4	103000	2	5	5

Eltern (Welle 2 - LONG)

	id	parid	wave	parvar3
1	100000	100301	2	1
2	100000	100302	2	2
3	100000	100304	2	3
4	101000	101301	2	4
5	101000	101302	2	5
6	102000	102303	2	-3

Zieldatensatz: Anker + Eltern (Welle 2)

	id	wave	var1	var2	parid	parvar3
1	100000	2	1	2	100301	1
2	100000	2	1	2	100302	2
3	100000	2	1	2	100304	3
4	101000	2	5	4	101301	4
5	101000	2	5	4	101302	5
6	102000	2	-3	3	102303	-3
7	103000	2	5	5	.	.

Hinweis:

- keine Umbenennung der anderslautenden Variablenamen erforderlich
- Erzeugung eines unbalancierten Zieldatensatzes, bei dem auch die Angaben der Ankerpersonen enthalten sind, zu denen keine Elternangaben vorliegen (z. B. Fall id=1030000)

DO-File für Stata:

```
use id wave var1 var2 using anchor2, clear // Auswahl der Variablen aus Anker W2
merge 1:m id using parent2, keepusing (parid parvar3) // Anfügen der Variablen aus Eltern W2
drop _merge // Löschen der Hilfsvariable _merge
sort id wave // Sortierung nach ID und Eltern-ID
```

2.5 pairfam in der Lehre

Um die pairfam-Daten in der Lehre einsetzen zu können – etwa für Beispielrechnungen der Studierenden – sollte der Datensatz aus Datenschutzgründen reduziert werden. Auch bei der Weitergabe dieser reduzierten Version muss von allen Teilnehmer*innen bzw. Studierenden ein [Weitergabeformular](#) unterschrieben und an den pairfam Nutzerservice gesendet werden.

Im Scientific-Use-File ist die nötige Syntax enthalten, um aus den ursprünglichen Daten eine solche Lehrversion zu generieren. Aus allen vorhandenen Datensätzen werden dabei neue Datensätze geschaffen, die nur die Hälfte der Fälle enthalten.

3 Publikationen

Das Beziehungs- und Familienpanel pairfam ist ein Infrastrukturprojekt, das umfangreich aufbereitete Daten für wissenschaftliche Analysen zur Verfügung stellt. Der Wert der Studie bemisst sich nicht zuletzt an der Anzahl und der Qualität von Veröffentlichungen auf Basis der pairfam-Daten.

3.1 Literaturdatenbank

Eine Zusammenstellung der bislang erschienenen Arbeiten im Kontext der pairfam-Studie findet sich auf der Webseite. Die Übersichten zu allen [pairfam-Publikationen](#) werden permanent aktualisiert. Voraussetzung für eine vollständige Literaturdatenbank ist die Mitwirkung der Autorinnen und Autoren: **Bitte informieren Sie den Nutzerservice per Email** (support@pairfam.de) **über Ihre Veröffentlichungen auf Basis der Daten des Beziehungs- und Familienpanels!** Die registrierten Veröffentlichungen bilden zugleich die Grundlage für die Verleihung des [Best Publication Award](#).

[Bibliografie](#): Liste aller bekannten pairfam-Publikationen in wissenschaftlichen Zeitschriften, Büchern und Sammelbänden (inkl. Abstracts)

[Studentische Arbeiten](#): Liste aller gemeldeten Abschlussarbeiten von Bachelor- und Masterstudierenden, die auf Analysen mit den pairfam-Daten basieren

3.2 Zitation

Denken Sie bei eigenen Veröffentlichungen bitte stets daran, die Studie in Ihrer Arbeit zu erwähnen. Hierzu sollten Sie sowohl das Referenzpapier [\(Huinink et al., 2011\)](#) als auch den Datensatz [\(Brüderl et al., 2020\)](#) zitieren. Beachten Sie, dass die Zitation der Daten jeweils das verwendete Release berücksichtigen sollte, da sich Publikationsjahr, Autorschaft und doi bei jedem Release ändern. Die korrekte bibliografische Kennzeichnung der Nutzung von Daten finden Sie auf der [pairfam-Website](#).