

Verknüpfung von Mikrozensus- Querschnitterhebungen 2016-2019 zu Paneldaten

Backes, Natalie; Pforr, Klaus

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

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

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Backes, N., & Pforr, K. (2023). *Verknüpfung von Mikrozensus-Querschnitterhebungen 2016-2019 zu Paneldaten*. (GESIS Papers, 2023/11). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.90811>

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

2023 | 11

**Verknüpfung von Mikrozensus-
Querschnitterhebungen
2016-2019 zu Paneldaten**

Natalie Backes & Klaus Pforr

GESIS Papers 2023|11

**Verknüpfung von Mikrozensus-
Querschnitterhebungen
2016-2019 zu Paneldaten**

Natalie Backes & Klaus Pforr

GESIS Papers

GESIS – Leibniz-Institut für Sozialwissenschaften
Data and Research on Society
German Microdata Lab
Postfach 12 21 55
68072 Mannheim

E-Mail: klaus.pfarr@gesis.org
<https://www.gesis.org/gml>

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

Zusammenfassung

Das Papier beschreibt die Zusammenführung der Mikrozensus Scientific-Use-Files 2016-2019 zu einem Personenpanel datensatz. Es wird das Vorgehen ausführlich beschrieben, die Anzahl der Auswahlbezirke, Haushalte und Personen und die Rekonstruktion der Rotationsgruppen dargestellt. Die Darstellung des Verfahrens schließt mit Konsistenzprüfungen. Danach wird eine kleine Beispielanalyse gezeigt, die das Analysepotential veranschaulicht.

Schlüsselwörter Mikrozensus, Scientific-Use-File, Rotationspanel

Abstract

The paper describes the combination of the Microcensus Scientific-Use-Files 2016-2019 into a person panel data set. The procedure is described in detail, the number of sampling districts, households and persons and the reconstruction of the rotation groups are presented. The presentation of the procedure concludes with consistency checks. Afterwards, a small sample analysis is shown to illustrate the analysis potential.

Keywords Mikrozensus, Scientific-Use-File, Rotational Panel

Inhaltsverzeichnis

1	Einleitung	6
2	Stichprobendesign.....	8
3	Panelerstellung.....	10
3.1	Vorbereiten der Datensätze	10
3.2	Zusammenführen der Datensätze	12
3.3	Variablen zur Bestimmung der Rotationsgruppe.....	14
4	Beispielanwendung	26
4.1	Deskriptive Verteilungen	26
4.2	Konsistenz: Geschlechts- und Geburtsjahrwechsel	27
4.3	Anwendungsbeispiel: Effekt der Geburt des ersten Kindes auf die Aufteilung der Erwerbsarbeitszeit in Partnerschaften.....	28
5	Schluss	29
6	Literatur.....	31

1 Einleitung

Der Mikrozensus (MZ) ist eine amtliche Befragung, bei der seit 1957 in Westdeutschland und 1991 in den neuen Bundesländern Informationen von ca. 1% der deutschen Wohnbevölkerung erhoben werden. Thematisch werden dabei demographische und sozioökonomische Grundinformationen sowie detaillierte Angaben zu arbeitsmarktbezogenen Eigenschaften abgedeckt. Das Scientific Use File (SUF) des Mikrozensus umfasst eine faktisch-anonymisierte 70%-Substichprobe des Mikrozensus und wird Forschenden zur Verfügung gestellt (GESIS/Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, 2022). Weitere Informationen zum Mikrozensus SUF sind online im Mikrodateninformationssystem (MISSY) (<https://www.gesis.org/missy/>) sowie auf dem Onlineauftritt des Forschungsdatenzentrums (<https://www.forschungsdatenzentrum.de/de/haushalte/mikrozensus>) abrufbar.

Neben der Auswertung von querschnittlichen Fragestellungen bietet der Mikrozensus SUF seit 2012 die Möglichkeit, durch die Verknüpfung der Datensätze Paneldatensätze zu erstellen, die die längsschnittliche Auswertung der Daten erlauben (Herter-Eschweiler/Schimpl-Neimanns, 2018). Dieses Verfahren kann aktuell für die Jahre 2012 bis 2015 sowie 2016 bis 2019 angewandt werden. Im Jahr 2016 wurde die Stichprobe auf der Auswahlgrundlage des Zensus 2011 erneuert, weshalb eine Verknüpfung mit den vorherigen Jahren nicht möglich ist. Durch die konzeptionelle Neugestaltung des Mikrozensus ab 2020 beschränkt sich die Möglichkeit der Verknüpfungen auf die Jahre 2016 bis einschließlich 2019.

Neben der Durchführung von längsschnittlichen Analyseverfahren auf Individual- und Haushaltsebene eignen sich die Panelverknüpfungen zudem dazu, Strukturen, Mobilität, Netto-Merkmalveränderungen und Statusübergänge auf regionaler Ebene zu analysieren. Zudem erlaubt das Stichprobendesign (Klumpen-Flächenstichprobe) Ursachen und Wirkungen von Veränderungen in der unmittelbaren Nachbarschaft (z. B. kleinräumige Mobilität, Segregation) zu untersuchen. Da die Teilnahme an der Befragung im Mikrozensus gesetzlich verpflichtend ist, zeichnet sich der Datensatz nicht nur durch große Bruttofallzahlen geringer Auswahlquote aus. Dies ermöglicht die Analyse von seltenen Veränderungen und spezifischen Subpopulationen sowie die Nutzung des Paneldatensatzes als Goldstandard für zeitliche Verläufe von räumlich immobilien Haushalten, bzw. Personen.¹ Die Daten eignen sich außerdem zur Gewichtung anderer längsschnittlicher Datenquellen, die räumlich immobile Personen und Haushalte abbilden.

Im Folgenden wird der Prozess der Panelerstellung für die Jahre 2016 bis 2019 (Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, 2020, 2021, 2022b, 2022a) für Stata beschrieben, der sich am Vorgehen der Verknüpfung von 2012 und 2013 nach Herter-Eschweiler und Schimpl-Neimanns (2018) orientiert. Dazu wird zunächst das Stichprobendesign erläutert, um die Schritte der Aufbereitung nachvollziehen zu können. Im Anschluss werden das Codeskript im Detail erklärt und Fallzahlen berichtet. Um Anwendungsgebiete zu demonstrieren, wird eine beispielhafte Panelanalyse durchgeführt. Limitationen, die beim Paneldatensatz zu berücksichtigen sind, werden am Schluss

¹ Mit Goldstandard ist hier gemeint, dass viele Umfragen die Verteilungen im Mikrozensus zum nachträglichen Ausgleich selektiver Nichtteilnahme verwenden.

diskutiert. Die Stata Do-Files der Zusammenführungen 2012 bis 2015 und 2016 bis 2019 sind online unter <https://www.gesis.org/missy/materials/MZ/tools/panel2012> verfügbar.

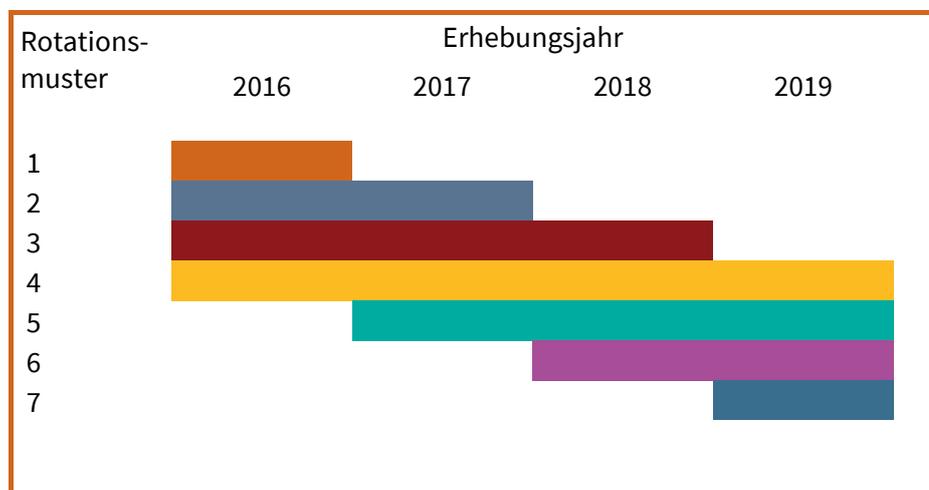
2 Stichprobendesign

Beim Vorgang der Panelverknüpfung müssen verschiedene Besonderheiten der Erhebungsmethodik des Mikrozensus berücksichtigt werden. Diese werden im folgenden Kapitel beschrieben.

Beim Mikrozensus handelt es sich um eine einstufige Klumpenstichprobe. Dabei erfolgt die gezogene Grundauswahl durch künstlich abgegrenzte Flächen (Klumpen), den sogenannten Auswahlbezirken (Primäreinheiten). Die Bildung der Auswahlbezirke basiert auf den Gebäudegrößenklassen, die von der Anzahl der Wohnungen in einem Gebäude abhängen. Durchschnittlich umfasst ein Auswahlbezirk 9 Wohnungen (Statistisches Bundesamt, 2021).

Der Stichprobenplan des Mikrozensus folgt der Methode der partiellen Rotation. Die Haushalte in den Auswahlbezirken werden vier Jahre lang jährlich wiederholt befragt. Jedes Jahr wird ein Rotationsviertel der Auswahlbezirke aus der Stichprobe herausgenommen und ein neues Rotationsviertel kommt hinzu. Daraus ergeben sich bei Kumulation der vier Jahre sieben Rotationsmuster, wobei die Wahrscheinlichkeit einer der Gruppen anzugehören für jeden Auswahlbezirk jeweils ein Siebtel beträgt.

Abbildung 1: Rotationsmuster im Mikrozensus Panel 2016 bis 2019.



Innerhalb der vier Erhebungsjahre werden fortziehende Haushalte für die Befragung nicht verfolgt, stattdessen werden die Haushalte befragt, die in die Samplewohnung nachgezogen sind.² Dadurch bleibt die Stichprobe nicht konstant identisch, was bei der Panelverknüpfung zwingend berücksichtigt werden muss. Der Datensatz eignet sich somit zur Beobachtung von Populationsveränderungen auf regionalen Ebenen; für Analysen auf Individualebene folgt aus Umzügen die Einschränkung, dass für regional mobile Haushalte nach dem Fortzug bzw. vor dem Zuzug keine Informationen vorliegen. Daraus folgt, dass mit dem

² Rotationspanel haben den Vorteil, dass die Stichproben die fortlaufende Veränderung der Population durch Geburten, Sterbefälle und Migration mitberücksichtigen.

Mikrozensus Panel Verteilungen von Variablen, die längsschnittliche Prozesse abbilden, verzerrt geschätzt werden können, wenn diese mit räumlicher Mobilität zusammenhängen. Diese Verzerrung kann prinzipiell durch Längsschnittgewichte ausgeglichen werden (Schimpl-Neimanns, 2021), sofern sich die Zusammenhänge zwischen den interessierenden Variablen und der Mobilität mit den im Mikrozensus enthaltenen Variablen abbilden lassen. Dasselbe gilt weitgehend analog für multivariate Zusammenhänge zwischen Variablen.

Aufgrund von Problemen der Erreichbarkeit kann es vorkommen, dass die Erhebungen zeitlich versetzt, d. h. im jeweiligen Folgejahr durchgeführt werden. In diesem Fall sind die Angaben erst im darauffolgenden Jahr verfügbar. Diese werden als „Jahresüberhang aus dem Vorjahr“ bezeichnet und müssen bei der Panelerstellung gesondert berücksichtigt werden.

3 Panelerstellung

Die Schritte der Panelerstellung erfolgen zu Beginn auf Personenebene, danach auf Haushalts- und schließlich auf Auswahlbezirksebene. Dafür müssen die Datensätze der vier Jahre vorbereitet werden, Dubletten identifiziert werden und die Zugehörigkeit zu den Rotationsgruppen ermittelt werden, woraufhin die Zusammenführung der Datensätze erfolgt.

3.1 Vorbereiten der Datensätze

Folgende Schritte werden zunächst für die Datensätze der Jahre 2016 bis 2019 durchgeführt. Die Schritte, die für die einzelnen Jahre identisch sind, werden beispielhaft für das Jahr 2016 dargestellt.

3.1.1 Personenidentifikationsnummer

Im Mikrozensus SUF steht die Variable `idpers` zur Verfügung, die eine längsschnittliche Personenidentifikationsnummer enthält. Zu Zwecken der Nachvollziehbarkeit wird die Konstruktion der Variable anhand der neuen, inhaltlich identischen Variable `persid` nachfolgend dokumentiert. Dazu dienen die folgenden systemfreien, längsschnittlich konsistenten Ordnungsnummern:

- EF1 Bundesland
- EF3 Nr. des Auswahlbezirks (systemfrei)
- EF4 Nr. des Haushalts im Auswahlbezirk (systemfrei)
- EF63 Personennummer

```
gen persid = string(EF1,"%02.0f")+ string(EF3,"%06.0f") /*
            */ + string(EF4,"%03.0f")+ string(EF63,"%02.0f")
```

Die Gebäudeordnungsnummern (EF8 Nr. des Gebäudes im Auswahlbezirk (systemfrei)) und Wohnungsordnungsnummern (EF9 Nr. der Wohnung im Auswahlbezirk (systemfrei)) sind nicht längsschnittlich konsistent und können somit nicht für Panelauswertungen verwendet werden.

3.1.2 Jahresüberhänge

Um Jahresüberhänge zu identifizieren und im weiteren Verlauf „umbuchen“ zu können, wird eine Variable `ueberhang` erstellt, die Jahresüberhänge identifiziert:

- EF5b Jahresüberhang aus dem Vorjahr

```
recode EF5b (-8=-8) (-7=-7) (0=0) (1 2=1), gen(ueberhang)
```

Nur im Fall von 2016 werden Jahresüberhänge aus dem Datensatz entfernt, da die Jahre 2015 und 2016 nicht verknüpfbar sind. Für die weiteren Jahre 2017 bis 2019 entfällt dieser Schritt.

```
drop if inlist(EF5b,1,2)
```

3.1.3 Dubletten

In den Datensätzen muss geprüft werden, ob für eine Person mehrere Fälle vorliegen. Diese Fälle werden als „Dubletten“ bezeichnet. In dem veröffentlichten Daten liegen regelmäßig vereinzelt Dubletten vor, vor allem dadurch, dass die Daten auch ausgefallene Auswahlbezirke enthalten, theoretisch müssen sich diese auf einen validen Fall zurückführen lassen. Um die Dubletten zu identifizieren, wurde die Variable `temp` erzeugt, die die vorhandenen Fälle je längsschnittlicher Personenid mit `ueberhang` zählt.

```
capture noisily isid persid ueberhang
bysort persid ueberhang: gen temp=_N
```

Die dadurch identifizierten Dubletten wurden auf Ähnlichkeit der Variablen geprüft, indem für alle vorhandenen Dubletten jeweils die Varianz aller Ausprägungen der im SUF enthaltenen numerischen Variablen berechnet wird.

```
quietly ds, not(type string)
foreach var in `r(varlist)' {
    quietly bysort persid ueberhang: egen double sd_`var' = /*
    */ sd(`var') if temp>1 & temp!=. & `var'>0 /*
    */ & !missing(`var')
    quietly count if sd_`var'!=.
    if `r(N)'==0 {
        quietly drop sd_`var'
    }
    else {
        quietly sum sd_`var'
        if `r(max)'==0 {
            quietly drop sd_`var'
        }
    }
}
sort persid ueberhang sd_*
capture log close
log using $dataout\log.log, replace
sum sd_*
quietly ds sd_*
foreach var in `r(varlist)' {
    noisily des `=substr("`var'",4,.)'
    noisily list persid ueberhang temp `var' /*
    */ `=substr("`var'",4,.)' if temp>1 & /*
    */ temp!=. & `var'>0 & `var'!=., nol /*
    */ sepby(persid) linesize(255)
}
```

Bei der Prüfung wurden keine problematischen Unterschiede festgestellt. Dubletten kommen ausschließlich bei Auswahlbezirken (AWBs) ohne befragte Haushalte, ausgefallenen Privathaushalten oder Gemeinschaftsunterkünften und gewerblich genutzten oder leerstehenden Wohnungen vor. Diese sind für die praktische Verwendung des MZ SUF nicht von Relevanz und erfordern dementsprechend keine präzisere Diagnose. Daher können alle überschüssigen Fälle in der Datenaufbereitung ohne Einschränkungen gelöscht werden. Für den Ausschluss können Dubletten durch die generierte Variable `dublette` identifiziert werden. Dabei werden diejenigen Fälle als Dublette markiert, die eine höhere Personenid haben.

```
bysort persid ueberhang (ID): gen dublette=_n<_N & temp>1 & temp!=.
```

Wie in Tabelle 1 dargestellt, gab es im Jahr 2017 271 Dubletten, was einem Anteil von 0,05 % der gesamten Fallzahl des Jahres 2016 (N = 546.785) entspricht.

Tabelle 1: Dubletten nach Jahren. Total und in Prozent.

Jahr	2016	2017	2018	2019	Total
Dubletten (N)	/	271	270	1.009	1.500
Prozent	/	0,05	0,05	0,17	0,09

3.1.4 Erhebungsjahr

Für eine sinnvolle Auswertung im zeitlichen Verlauf wurde eine Variable `survey_year` mit dem Wert des jeweiligen Jahres erstellt. Diese enthält Informationen über das Erhebungsjahr und unterscheidet sich inhaltlich von der Variable `year`, die das Berichtsjahr abbildet und deren Generierung in Kapitel 3.2 beschrieben wird. Die Unterscheidung des Erhebungsjahres und des Berichtsjahrs ist notwendig, da im Falle von Jahresüberhängen das Erhebungsjahr nicht dem Berichtsjahr entspricht und Informationen zum Erhebungsjahr benötigt werden, um das Berichtsjahr generieren zu können.

```
gen survey_year=2016
```

Abschließend werden die Datensätze jeweils vor dem Zusammenführen sortiert und temporär abgespeichert.

```
sort persid ueberhang
compress
save `temp16', replace
```

3.2 Zusammenführen der Datensätze

Beginnend mit dem temporären Datensatz 2016, werden die Datensätze nacheinander zusammengeführt.

```
use `temp16'
append using `temp17'
append using `temp18'
append using `temp19'
```

Wie oben erwähnt, bildet die erstellte Variable `survey_year` lediglich das Erhebungsjahr ab; um das inhaltliche Jahr – das “Berichtsjahr” – festzustellen, wurde die Jahresvariable `year` erstellt. Dabei werden die Jahresüberhänge dem Vorjahr zugeordnet.

```
gen year=2016      if (survey_year==2016 & ueberhang==0) | /*
*/ (survey_year==2017 & ueberhang==1)
replace year=2017 if (survey_year==2017 & ueberhang==0) | /*
*/ (survey_year==2018 & ueberhang==1)
replace year=2018 if (survey_year==2018 & ueberhang==0) | /*
*/ (survey_year==2019 & ueberhang==1)
replace year=2019 if (survey_year==2019 & ueberhang==0) | /*
*/ (survey_year==2020 & ueberhang==1)
replace year=-7 if ueberhang==--7
replace year=-8 if ueberhang==--8
bysort dublette year persid ueberhang: gen temp=_N
table year (dublette temp), totals(year#dublette /*
*/ dublette#temp) missing zero
```

Die letzte Zeile produziert die Tabelle 2.

Tabelle 2: Verteilung der Duplikate in den Berichtsjahren.

year	dublette				Total	1		Total
	0	1	2	3		1	2	
temp	1	2	3	4				
-8	70.187	41.692	19.386	7.452	138.717	1.532	2	1.534
-7	58.067	770	3	0	58.840	16	0	16
2016	530.049	0	0	0	530.049	0	0	0
2017	533.222	0	0	0	533.222	0	0	0
2018	530.308	0	0	0	530.308	0	0	0
2019	507.997	0	0	0	507.997	0	0	0
Total	2.229.830	42.462	19.389	7.452	0	1.548	2	0

Nach dieser Zuordnung muss erneut auf Dubletten geprüft werden. Die Dubletten sind mit der Variable `dublette` identifizierbar, wobei erneut diejenigen Fälle als Dubletten markiert wurden, die die höhere Personen-Id aufweisen. Nach der Durchführung dieses Schrittes beläuft sich die Fallzahl der als Dubletten markierten Fälle auf 41.296. Die hohe Anzahl ist durch den Codierungsschritt bedingt, der ungültige Fälle, d.h. leerstehende Wohnungen, ausgefallene Privathaushalte und Gemeinschaftsunterkünfte und Auswahlbezirke ohne befragten Haushalt, ebenso als Dubletten markiert werden. Da es sich bei genannten ungültigen Fällen ausschließlich um ausgefallene HH- oder AWB-Fälle handelt, welche für

Analysen nicht relevant sind, können diese als Dubletten markierten Fälle in der Datenaufbereitung ausgeschlossen werden.

```
bysort temp dublette year persid ueberhang (ID): replace /*
      */ dublette=(dublette >0)|(_n<_N & temp>1 & temp!=.)
```

3.3 Variablen zur Bestimmung der Rotationsgruppe

Durch die gesetzliche Auskunftspflicht im Mikrozensus sind übliche Probleme der Panelattrition nicht gegeben (Mikrozensusgesetz - MZG, 2016). Jedoch bestehen Einschränkungen durch fortziehende Haushalte und die Erhebungsmethodik mit Rotationsvierteln. Fortziehende Personen und Haushalte werden nicht verfolgt, d. h. dies muss gegebenenfalls durch Längsschnittgewichte, die die Selektivität ausgleichen berücksichtigt werden. Die Ersetzung der Rotationsviertel führt dazu, dass im Zeitraum 2016-2019 die eingangs dargestellten Rotationsgruppen vorliegen. Hier ist zu beachten, dass im Scientific Use File keine Variable enthalten ist, die die Zuordnung der Haushalte zu den Rotationsgruppen anzeigt. Um die Auswahlbezirke, Haushalte und entsprechend Personen den Rotationsgruppen zuzuordnen zu können, muss die Rotationsgruppe heuristisch gebildet werden. Für diese Zuordnung werden entsprechende Teilnahmemustervariablen gebildet, was im Folgenden erläutert wird.

Zunächst wird ein Auswahlbezirk als teilnehmend betrachtet, wenn für diesen Auswahlbezirk in dem jeweiligen Jahr mindestens eine gültige Beobachtung vorliegt. Daraus sollten sich die sieben Rotationsgruppen ergeben, die in Abbildung 1 dargestellt sind. Da dieses Vorgehen aber nicht nur eine Schätzung darstellt, liegen auch Fälle vor, die nicht zu diesen sieben Rotationsgruppen passen.

3.3.1 AWB-Teilnahmemuster

Zur Feststellung der Teilnahmemuster einzelner Auswahlbezirke wurde die Variable `awb_panel_pattern` erstellt. Die Variable wurde anhand der folgenden Variablen gebildet:

- EF1 Bundesland
- EF3 Nr. des Auswahlbezirks (systemfrei)
- year Berichtsjahr

Die Variable bildet das Teilnahmemuster ab, wobei die Ziffern die jeweiligen teilgenommenen Berichtsjahre anzeigen.

```
forvalues i=16/19 {
    bysort EF1 EF3: egen y`i'=total(year==20`i')
    replace y`i'=y`i'>0 & y`i'!=.
}
gen awb_panel_pattern=""
replace awb_panel_pattern=cond(y16==1,"16","__")+ /*
      */ + cond(y17==1,"17","__")+cond(y18==1,"18","__") /*
      */ + cond(y19==1,"19","__")
```

Die Verteilung der im Code gebildete Variable `awb_panel_pattern` ist in Tabelle 3 abgetragen.

Tabelle 3: Verteilung der Auswahlbezirksteilmuster und der Rotationsgruppen.

Auswahlbezirk Panelmuster	N	%	Rotationsgruppe	%
16_____	140.206	6,09	1	6,09
1617_____	286.240	12,44	2	12,44
161718__	431.264	18,75	3	18,75
16171819	567.728	24,68	4	24,68
__171819	438.625	19,06	5	19,06
____1819	291.042	12,65	6	12,65
_____19	130.458	5,67	7	5,67
1617__19	608	0,03		
16__1819	531	0,02		
16__18__	141	0,01		
16____19	52	0,00		
__1718__	5.091	0,22		
__17__19	416	0,02		
__17_____	2.411	0,10		
____18__	1.548	0,07		
_____	4.322	0,19		
Total	2.300.683	100,00		99,34

Anhand der Variable `panelok` werden die Auswahlbezirke festgestellt, die den sieben Rotationsgruppen der Abbildung 1 entsprechen.

```
gen panelok=inlist(awb_panel_pattern,/*
    */ "16_____", /*
    */ "1617_____", /*
    */ "161718__", /*
    */ "16171819", /*
    */ "__171819", /*
    */ "____1819", /*
    */ "_____19")
```

Die Nummern der gültigen Rotationsgruppen entsprechend der Abbildung 1 wurden in der Variable `rotationsgruppe` zugewiesen. Die Verteilung kann der Tabelle 3 entnommen werden. Der Anteil ungültiger Rotationsgruppen hat nur geringfügige Auswirkungen, da lediglich 0,66% der Fälle keiner gültigen Rotationsgruppe zuordenbar sind

```
egen rotationsgruppe = group(awb_panel_pattern)
recode rotationsgruppe (8 = 1) (4 = 2) (2 = 3) (1 = 4) (9 = 5) /*
    */ (13 = 6) (15 = 7) (else=.)
label var rotationsgruppe "Rotationsgruppe"
```

3.3.2 Personen-Teilnahmemuster

Im nächsten Schritt wurde die generierte Variable `panelok` verwendet, um Teilnahmemuster auf Personenebene feststellen zu können. Die generierte Variable `pers_panel_pattern` wurde analog zur AWB-Panel Variable gebildet. Tabelle 4 zeigt die Auszählung der gebildeten Variable

```
forvalues i=16/19 {
    bysort persid: egen yp`i'=total(year==20`i')
    replace yp`i'=yp`i'>0 & yp`i'!=.
}

gen pers_panel_pattern=""
replace pers_panel_pattern=cond(yp16==1,"16","__") /*
    */ + cond(yp17==1,"17","__") + cond(yp18==1,"18","__") /*
    */ + cond(yp19==1,"19","__")
```

Tabelle 4: Verteilung der Personen-Teilnahmemuster.

Personen-Teilnahmemuster	N	%
16171819	382.698	16,63
161718__	344.853	14,99
1617__19	1.893	0,08
1617____	279.108	12,13
16__1819	2.181	0,09
16__18__	2.556	0,11
16____19	296	0,01
16_____	177.087	7,70
__171819	345.009	15,00
__1718__	56.028	2,44
__17__19	2.394	0,10
__17_____	38.197	1,66
____1819	273.703	11,90
____18__	37.805	1,64
_____19	157.768	6,86
_____	199.107	8,65
Total	2.300.683	100,00

Es ist zu beachten, dass die 199.107 Fälle mit dem ungültigen Teilnahmemuster „_____“, die in keinem Jahr teilgenommen haben, angesichts der verpflichtenden Teilnahme eine hohe Zahl an Fällen darstellt. Von den 199.107 Fällen sind 149.251 Fälle leerstehende Wohnungen und ausgefallene Privathaushalte und Gemeinschaftsunterkünfte, und 58.856 Fälle Auswahlbezirke ohne befragten Haushalt. Ausfallzahlen wie bei den 56.028 Fällen mit dem Muster „__1718__“, den 38.197 Fällen mit dem Muster „__17_____“ und den 37.805 Fällen mit dem Muster „____18__“ sind dagegen trotz verpflichtender Teilnahme nicht mit so hohen Fallzahlen zu erwarten.

Anschließend erfolgte die Identifikation von Dubletten durch den Abgleich der gültigen vorhandenen Personenjahre und dem jeweiligen AWB-Teilnahmemuster, welcher in Tabelle 5 abgelesen werden kann. 33 Dubletten wurden dabei identifiziert. Insgesamt erhöht sich die Fallzahl der Dubletten auf 49.587, da in diesem Schritt zudem Fälle markiert wurden, die keinem gültigen Teilnahmemuster entsprechen. Da diese in der Analyse nicht von Nutzen sind, können diese bedenkenlos mit den Dubletten ausgeschlossen werden.

```
bysort persid: gen N=_N if dublette==0 & panelok==1
replace dublette=1 if (pers_panel_pattern=="16171819" & N!=4) | /*
*/ (N!=3&inlist(pers_panel_pattern,"161718__", /*
*/ "1617__19","16__1819","__171819")) | /*
*/ (N!=2&inlist(pers_panel_pattern,"1617____", "16__18__" /*
*/ ,"16____19","__17__19","__1718__","____1819")) | /*
*/ (N!=1 &inlist(pers_panel_pattern,"16_____", /*
*/ "__17_____", "____18__", "______19"))
table pers_panel_pattern (dublette N), totals(pers_panel_pattern /*
dublette#N) missing zerocounts
drop N
```

Tabelle 5: Verteilung der Personen-Teilnahmemuster und gezählte gültige Personenjahre.

Personen- Teilnahmemuster	dublette										Total
	1	2	3	4	Missing	2	3	5	Missing		
16171819		0	0	0	382.688	0	0	0	10	0	382.698
161718__		0	0	344.853	0	0	0	0	0	0	344.853
1617__19		0	0	1.677	0	0	0	0	0	216	1.893
1617____		0	279.014	0	0	0	0	6	0	88	279.108
16__1819		0	0	1.965	0	0	0	0	0	216	2.181
16__18__		0	2.498	0	0	0	0	0	0	58	2.556
16____19		0	290	0	0	0	0	0	0	6	296
16_____	176.996		0	0	0	0	0	0	0	91	177.087
__171819		0	0	345.009	0	0	0	0	0	0	345.009
__1718__		0	52.664	0	0	0	0	0	0	3.364	56.028
__17__19		0	2.228	0	0	0	0	0	0	166	2.394
__17____	35.809		0	0	0	0	2	0	0	2.386	38.197
____1819		0	273.636	0	0	0	0	15	0	52	273.703
____18__	36.311		0	0	0	0	0	0	0	1.494	37.805
______19	157.647		0	0	0	0	0	0	0	121	157.768
_____	123.366	21.058	6.397	1.893	5.097	0	0	0	0	41.296	199.107
Total	530.129	631.388	699.901	384.581	5.097	2	21	10	49.554	0	

3.3.3 Personen-AWB-Teilnahmemuster

Die Variable `pers_awb_panel_pattern` kombiniert die Teilnahmemuster auf der Ebene von Personen und Auswahlbezirken. Dabei geben die ersten acht Stellen Information über das Teilnahmemuster von Personen und die darauffolgenden acht Stellen Information über das Teilnahmemuster von Auswahlbezirken.

```
gen pers_awb_panel_pattern=pers_panel_pattern+"|"+awb_panel_pattern
```

In Tabelle 6 kann die Verteilung der Variable eingesehen werden. Die Personen-Auswahlbezirk-Teilnahmemuster können Aufschluss über Wegzüge oder Sterbefälle geben: Beispielsweise zeigt die Ausprägung mit dem Muster „161718_|16171819“ an, dass die Person in den Jahren 2016 bis 2018, jedoch nicht im Jahr 2019 erfolgreich befragt wurde; im Auswahlbezirk jedoch für alle Jahre gültige Werte vorliegen. Daraus kann geschlossen werden, dass die Person 2019 weggezogen oder verstorben ist. Umgekehrt können auch Zuzüge und Geburten anhand der Muster identifiziert werden: wenn beispielsweise für einen Fall das Muster „_171819|16171819“ vorliegt, kann angenommen werden, dass die Person im Jahr 2017 zugezogen oder geboren ist.

Tabelle 6: Verteilung der Personen-Auswahlbezirk-Teilnahmemuster.

Personen-Auswahlbezirk-Teilnahmemuster	N	%
16171819 16171819	382.698	16,63
161718_ 16171819	28.296	1,23
161718_ 161718__	316.557	13,76
1617__19 16171819	1.677	0,07
1617__19 1617__19	216	0,01
1617____ 16171819	20.518	0,89
1617____ 161718__	22.554	0,98
1617____ 1617__19	88	0
1617____ 1617____	235.948	10,26
16__1819 16171819	1.965	0,09
16__1819 16__1819	216	0,01
16__18_ 16171819	486	0,02
16__18_ 161718__	2.012	0,09
16__18_ 16__1819	6	0
16__18_ 16__18__	52	0
16____19 16171819	290	0,01
16____19 1617__19	6	0
16_____ 16171819	14.122	0,61
16_____ 161718__	14.872	0,65
16_____ 1617__19	26	0
16_____ 1617____	15.267	0,66
16_____ 16__1819	40	0
16_____ 16__18__	13	0
16_____ 16____19	12	0
16_____ 16_____	132.735	5,77
__171819 16171819	26.655	1,16
__171819 __171819	318.354	13,84

Personen-Auswahlbezirk-Teilnahmemuster	N	%
__1718__ 16171819	4.454	0,19
__1718__ 161718__	22.592	0,98
__1718__ __171819	25.618	1,11
__1718__ __1718__	3.364	0,15
__17__19 16171819	356	0,02
__17__19 1617__19	12	0
__17__19 __171819	1.872	0,08
__17__19 __17__19	154	0,01
__17____ 16171819	3.314	0,14
__17____ 161718__	3.590	0,16
__17____ 1617__19	16	0
__17____ 1617____	14.137	0,61
__17____ __171819	14.770	0,64
__17____ __1718__	406	0,02
__17____ __17__19	66	0
__17____ __17____	1.898	0,08
____1819 16171819	17.986	0,78
____1819 16__1819	52	0
____1819 __171819	22.259	0,97
____1819 ____1819	233.406	10,15
____18__ 16171819	3.086	0,13
____18__ 161718__	11.903	0,52
____18__ 16__1819	8	0
____18__ 16__18__	18	0
____18__ __171819	3.830	0,17
____18__ __1718__	385	0,02
____18__ ____1819	17.492	0,76
____18__ ____18__	1.083	0,05
______19 16171819	10.750	0,47
______19 1617__19	53	0
______19 16__1819	11	0
______19 16____19	11	0
______19 __171819	12.162	0,53
______19 __17__19	46	0
______19 ____1819	14.136	0,61
______19 ______19	120.599	5,24
_____ 16171819	51.075	2,22
_____ 161718__	37.184	1,62
_____ 1617__19	191	0,01
_____ 1617____	20.888	0,91
_____ 16__1819	198	0,01
_____ 16__18__	58	0
_____ 16____19	29	0
_____ 16____	7.471	0,32
_____ __171819	39.760	1,73
_____ __1718__	936	0,04
_____ __17__19	150	0,01
_____ __17____	513	0,02
_____ ____1819	26.008	1,13

Personen-Auswahlbezirk-Teilnahmemuster	N	%
_____18__	465	0,02
_____19	9.859	0,43
_____	4.322	0,19
Total	2.300.683	100,00

Tabelle 6 zeigt für alle sieben Rotationsgruppen niedrige Nichtteilnahmequoten auf Personenebene, zumindest, soweit man dies ohne weitere Informationen beurteilen kann. Die Interpretation der Teilnahmemuster ist zur Veranschaulichung beispielhaft anhand der Rotationsgruppe 3 mit dem Auswahlbezirkmuster "161718__" in Tabelle 6b dargestellt (nicht im Code enthalten). Die Muster auf Personenebene können Hinweise darauf geben, ob es sich um Wegzüge und Sterbefälle, auf Zuzüge und Geburten, und auf Nichtteilnahme aus anderen Gründen handelt.

Tabelle 6b: Verteilung der Personen-Auswahlbezirk-Teilnahmemuster mit AWB-Muster „161718__“

Personen-Auswahlbezirk-Teilnahmemuster	N	%	Interpretation
161718__ 161718__	316.557	73,40	Vollständige Teilnahme
1617____ 161718__	22.554	5,23	Auszug oder Tod zwischen 2017 und 2018
16__18__ 161718__	2.012	0,47	Nicht-Teilnahme im Jahr 2017
16_____ 161718__	14.872	3,45	Auszug oder Tod zwischen 2016 und 2017
__1718__ 161718__	22.592	5,24	Zuzug oder Geburt zwischen 2016 und 2017
__17____ 161718__	3.590	0,83	Einzug oder Geburt zwischen 2016 und 2017 und Auszug oder Tod zwischen 2017 und 2018
____18__ 161718__	11.903	2,76	Einzug oder Geburt zwischen 2017 und 2018
_____ 161718__	37.184	8,62	Vollständiger Ausfall des gesamten AWBs
Total	431.261	100,00	

In der zweiten Zeile stehen Personen aus AWBs, in denen mindestens eine Person in 2016, 2017 und 2018 aber keine Person in 2019 teilgenommen hat (AWB_Muster „161718__“), die aber selbst nicht in 2018 teilgenommen haben. Ausfälle dieser Art können dadurch zustande kommen, dass die Person wegzieht, stirbt oder nicht teilnimmt. Angesichts der verpflichtenden Teilnahme ist der Wegzug oder Tod das wahrscheinlichere Ereignis. Dagegen stehen in der fünften Zeile Personen aus AWBs, in denen mindestens eine Person in 2016, 2017 und 2018 aber keine Person in 2019 teilgenommen hat (AWB_Muster „161718__“), die aber selbst nicht in 2016 teilgenommen haben. Für dieser Personen ist aus derselben Überlegung zu erwarten, dass sie erst in 2017 zugezogen sind oder neugeboren wurden. Die entsprechenden Interpretationen gelten für alle anderen Muster abgesehen für die Personemuster „16__18__“ und „____“ . Der erste Fall lässt sich nur durch Nichtteilnahme erklären, was dem niedrigen Anteil von weniger als einem Prozent entspricht. Der zweite Fall entspricht Auswahlbezirken, die komplett ausgefallen sind, vermutlich dadurch, dass Auswahlbezirke mit langfristig unbewohnten Wohneinheiten gezogen wurden.

In Bezug auf die Bewertung der Datenqualität ist festzuhalten, dass es zu erwartende Nichtteilnahmen durch Zu- und Fortzüge einerseits und Geburten und Sterbefälle andererseits gibt. Echte Nichtteilnahmen sind selten. Dagegen ist der Anteil der komplett ausgefallenen Auswahlbezirke hoch.

3.3.4 Haushalts-Teilnahmemuster

Das Teilnahmemuster der Haushalte (HH) kann der Variable `hh_panel_pattern` entnommen werden. Diese wurde nach dem Schema der vorherigen Teilnahmemustervariablen erstellt und in Tabelle 7 ausgezählt. Man beachte, dass hier lediglich ein Indikator zur Darstellung erstellt wird, der einen Fall pro Haushalt und Jahr markiert.

```
* Zur Darstellung ein Fall von Haushalt und Jahr
bysort EF1 EF3 EF4 year: gen hh_index=_n==1

forvalues i=16/19 {
    bysort EF1 EF3 EF4: egen yh`i'=total(year==20`i')
    replace yh`i'=yh`i'>0 & yh`i'!=.
}
gen hh_panel_pattern=""
replace hh_panel_pattern=cond(yh16==1,"16","__") /*
    */ + cond(yh17==1,"17","__")+cond(yh18==1,"18","__") /*
    */ + cond(yh19==1,"19","__")
tab hh_panel_pattern if hh_index==1
```

Tabelle 7: Verteilung der Haushalt-Teilnahmemuster.

Haushalt Teilnahmemuster	N	%
16171819	196.564	16,43
161718__	176.752	14,77
1617__19	1.310	0,11
1617____	146.340	12,23
16__1819	1.432	0,12
16__18__	1.926	0,16
16____19	200	0,02
16_____	104.527	8,74
__171819	173.929	14,53
__1718__	32.530	2,72
__17__19	1.753	0,15
__17_____	28.418	2,37
____1819	135.287	11,31
____18__	27.186	2,27
______19	77.650	6,49
_____	90.835	7,59
Total	1.196.639	100,00

Wie in Tabelle 7 abzulesen ist, ist die Zahl der Fälle, die in keinem Jahr teilgenommen haben, vergleichsweise hoch.

3.3.5 HH-AWB-Teilnahmemuster

Um die Informationen der Haushalts- und Auswahlbezirksteilnahmemuster zu kombinieren, wurde die Variable `hh_awb_panel_pattern` erstellt. Die ersten acht Stellen zeigen die Haushaltsmuster und die letzten acht Stellen die AWB-Muster dar.

```
gen hh_awb_panel_pattern=hh_panel_pattern+"|"+awb_panel_pattern
tab hh_awb_panel_pattern if hh_index==1
```

In Tabelle 8 kann die Verteilung der Variable eingesehen werden. Anhand der Haushalts-Auswahlbezirk-Teilnahmemuster können zu- oder weggezogene Haushalte identifiziert werden. So kann exemplarisch angenommen werden, dass bei dem Haushalts-Auswahlbezirks-Teilnahmemuster „`__171819|16171819`“ ein Zuzug im Jahr 2017 vorliegt, während das Muster „`1617____|16171819`“ auf einen Fortzug im Jahr 2018 hinweist. Bei wegfallenden Haushalten ist jedoch zu beachten, dass Ausfälle alternativ implizieren können, dass sämtliche Haushaltsmitglieder verstorben sind.

Tabelle 8: Verteilung der Haushalts-Auswahlbezirks-Teilnahmemuster.

Haushalts-Auswahlbezirks Teilnahmemuster	N	%
16171819 16171819	196.564	16,43
161718_ 16171819	16.049	1,34
161718_ 161718__	160.703	13,43
1617__19 16171819	1.149	0,10
1617__19 1617__19	161	0,01
1617____ 16171819	13.559	1,13
1617____ 161718__	14.851	1,24
1617____ 1617__19	67	0,01
1617____ 1617____	117.863	9,85
16__1819 16171819	1.298	0,11
16__1819 16__1819	134	0,01
16__18_ 16171819	406	0,03
16__18_ 161718__	1.478	0,12
16__18_ 16__1819	3	0,00
16__18_ 16__18__	39	0,00
16____19 16171819	191	0,02
16____19 1617__19	6	0,00
16____19 16__1819	3	0,00
16_____ 16171819	12.359	1,03
16_____ 161718__	13.217	1,10
16_____ 1617__19	34	0,00
16_____ 1617____	13.214	1,10
16_____ 16__1819	45	0,00
16_____ 16__18__	15	0,00
16_____ 16____19	16	0,00
16_____ 16____	65.627	5,48
__171819 16171819	12.386	1,04
__171819 __171819	161.543	13,50
__1718_ 16171819	3.115	0,26
__1718_ 161718__	10.972	0,92
__1718_ __171819	16.748	1,40
__1718_ __1718__	1.695	0,14
__17__19 16171819	269	0,02
__17__19 1617__19	7	0,00
__17__19 __171819	1.387	0,12
__17__19 __17__19	90	0,01
__17____ 16171819	3.141	0,26
__17____ 161718__	3.316	0,28
__17____ 1617__19	17	0,00
__17____ 1617____	7.509	0,63
__17____ __171819	13.077	1,09
__17____ __1718__	369	0,03
__17____ __17__19	70	0,01
__17____ __17____	919	0,08
____1819 16171819	8.218	0,69
____1819 16__1819	47	0,00
____1819 __171819	10.897	0,91

Haushalts-Auswahlbezirks Teilnahmemuster	N	%
___1819 ___1819	116.125	9,70
___18_ 16171819	2.658	0,22
___18_ 161718__	5.801	0,48
___18_ 16__1819	8	0,00
___18_ 16__18__	7	0,00
___18_ __171819	3.483	0,29
___18_ __1718__	166	0,01
___18_ ___1819	14.503	1,21
___18_ ___18__	560	0,05
____19 16171819	5.094	0,43
____19 1617__19	34	0,00
____19 16__1819	5	0,00
____19 16____19	7	0,00
____19 __171819	5.799	0,48
____19 __17__19	29	0,00
____19 ___1819	7.374	0,62
____19 _____19	59.308	4,96
_____ 16171819	17.153	1,43
_____ 161718__	14.478	1,21
_____ 1617__19	58	0,00
_____ 1617____	9.801	0,82
_____ 16__1819	70	0,01
_____ 16__18__	25	0,00
_____ 16____19	8	0,00
_____ 16_____	7.198	0,60
_____ __171819	15.361	1,28
_____ __1718__	438	0,04
_____ __17__19	51	0,00
_____ __17____	314	0,03
_____ __1819	12.705	1,06
_____ __18__	269	0,02
_____ ____19	9.799	0,82
_____ _____	3.107	0,26
Total	1.196.639	100,00

Ähnlich wie auf der Personenebene lässt sich hier erkennen, wie hoch die Ausfallraten auf Haushaltsebene sind, die sich vermutlich durch Weg- und Zuzüge und Nichtteilnahmen ganzer Haushalte ergeben. Ausfälle durch Todesfälle sind auf Haushaltsebene unwahrscheinlicher, da entsprechend alle Haushaltsmitglieder von einem Jahr auf das nächste Jahr versterben müssen. Zur Veranschaulichung der Muster betrachten wir hier erneut die in Tabelle 8b dargestellten Teilnahmemuster für Auswahlbezirke, in denen in 2016, 2017 und 2018 mindestens eine Person teilgenommen hat und in 2019 keine Person teilgenommen hat.

Tabelle 8b: Verteilung der Haushalts-Auswahlbezirks-Teilnahmemuster für AWB-Muster „161718__“.

Haushalts-Auswahlbezirks Teilnahmemuster	N	%	Interpretation
161718__ 161718__	160.703	71,48	Vollständige Teilnahme
1617____ 161718__	14.851	6,61	Auszug zwischen 2017 und 2018
16__18__ 161718__	1.478	0,66	Nicht-Teilnahme des gesamten Haushalts im Jahr 2017
16_____ 161718__	13.217	5,88	Auszug zwischen 2016 und 2017
__1718__ 161718__	10.972	4,88	Zuzug zwischen 2016 und 2017
__17____ 161718__	3.316	1,47	Einzug zwischen 2016 und 2017 und Auszug zwischen 2017 und 2018
____18__ 161718__	5.801	2,58	Einzug zwischen 2017 und 2018
_____ 161718__	14.478	6,44	Vollständiger Ausfall des gesamten AWBs
Total	224.816	100,00	

Hier ist noch deutlicher als auf Personenebene erkennbar, dass Nichtteilnahmen aller Wahrscheinlichkeit nach durch Zu- und Wegzüge dominiert werden, während zwischenzeitliche Ausfälle ganzer Haushalte sehr selten auftreten. Vollständige Ausfälle treten dagegen über alle vier Jahre unerwartet häufig auf.

3.3.6 Compress

Im nächsten Schritt wird der entstandene Datensatz von Person-Jahr-Ebene in eine Personen-Ebene überführt, so dass für jede Person, die in den vier Jahren 2016 bis 2019 mindestens einmal enthalten war, genau ein Fall vorliegt, in dem die Information über die Teilnahmen in den vier Jahren enthalten ist. Dabei wird ausschließlich die Variablen behalten, die der Identifikation der Personen, der Dubletten und der Rotationsgruppe dienen. Dieser Schritt erfolgt, um einen Datensatz zu erhalten, an den die Variablen und Jahre zugespielt werden können, die für den Samplezuschnitt und die geplante Analyse relevant sind.

```
keep EF1 EF3 EF4 EF63 panelok hh_panel_pattern /*
      */ hh_awb_panel_pattern pers_panel_pattern /*
      */ pers_awb_panel_pattern idpers idhh dublette
bysort EF1 EF3 EF4 EF63: keep if _n==1
sort EF1 EF3 EF4 EF63
compress
```

4 Beispielanwendung

4.1 Deskriptive Verteilungen

Für die folgenden Ausführungen werden ungültige Fälle wie in Tabelle 9 dargestellt ausgeschlossen, d.h. es werden Auswahlbezirke ohne befragten Haushalt, leerstehende Wohnungen, ausgefallene Privathaushalte und Gemeinschaftsunterkünfte sowie Jahre, in denen Personen aus anderen Gründen nicht teilgenommen haben, Fälle außerhalb der gültigen Rotationsgruppen und Fälle, die nicht eindeutig einer Rotationsgruppe zuordenbar waren, ausgeschlossen.

Tabelle 9: Konstruktion der gültigen Fälle.

	Ausgeschlossene Fälle (N)	Resultierend (N)	%
Gesamtsample		4.869.880	100,00
Dubletten entfernt (keep if dublette==0)	151.488	4.718.392	96,89
Gültige Rotationsgruppe	17.504	4.700.888	96,53
Gültige Privathaushalte und Gemein- schaftsunterkünfte (keep if inlist(EF31,1,2))	2.637.000	2.063.888	42,38
Total		2.063.888	42,38

Nachdem alle ungültigen Fälle ausgeschlossen worden sind, umfasst der resultierende Datensatz 2.063.888 gültige Personen-Jahre und 1.020.508 Haushalts-Jahre (siehe Tabelle 10), die für Analysezwecke geeignet sind.

Tabelle 10: Gültige Personen- und Haushalts-Jahre.

Jahr	Personen-Jahre	%	Haushalts-Jahre	%
2016	507.996	24,61	250.717	24,57
2017	521.160	25,25	257.886	25,27
2018	518.377	25,12	256.259	25,11
2019	516.355	25,02	255.646	25,05
Total	2.063.888	100,00	1.020.508	100,00

Nach Zusammenführung ergeben sich die Fallzahlen der einzelnen Jahre aus Umbuchungen der Jahresüberhänge. Die Jahresüberhänge sind in Tabelle 12 aufgeführt.

Tabelle 11: Verteilung der Jahresüberhänge.

	2016	2017	2018	2019
Kein Jahresüberhang	503.985	503.270	504.419	501.733
Jahresüberhang	4.011	17.890	13.958	14.622
Total	507.996	521.160	518.377	516.355

In der Tabelle 12 kann die Verteilung der Rotationsgruppen abgelesen werden. Demnach liegen 507.525 Personenjahre vor, in denen alle vier Wellen (Rotationsgruppe 4) verfügbar sind. 780.470 Beobachtungen gehören der Gruppe an, für die in drei Jahren gültige Daten vorliegen (Rotationsgruppe 3 und 5) und 522 559 der Gruppe, für die zwei gültige Beobachtungsjahre vorliegen (Rotationsgruppe 2 und 6). Für 253.334 Personen liegt nur eine Beobachtung in einem Jahr vor (Rotationsgruppe 1 und 7).

Tabelle 13: Verteilung der Rotationsgruppen.

Rotationsgruppe	N	%
1 (2016)	132.735	6,43
2 (2016, 2017)	262.067	12,70
3 (2016, 2017, 2018)	388.505	18,82
4 (2016, 2017, 2018, 2019)	507.525	24,59
5 (2017, 2018, 2019)	391.965	18,99
6 (2018, 2019)	260.492	12,62
7 (2019)	120.599	5,84
Total	2.063.888	100,00

4.2 Konsistenz: Geschlechts- und Geburtsjahrwechsel

Um die Güte des zusammengeführten Paneldatensatzes festzustellen, wurde die Konsistenz der Merkmalsausprägungen der im Prinzip unveränderlichen Variablen Geschlecht und Geburtsjahr geprüft. Die Ergebnisse der Konsistenzkontrolle sind in Tabelle 14 protokolliert. Beim Geschlecht sind bei 99,4 % konstante Angaben festzustellen, beim Geburtsjahr liegt der prozentuale Anteil konsistenter Angaben bei 96,5 %. Änderungen des Geschlechts sind zwar möglich, aber es ist zu erwarten, dass diese so selten sind, dass Wechsel des Geschlechts eher auf fehlerhaft verknüpfte Personen mit anderem Geschlecht oder geänderte Angaben hinweisen als auf dieselbe Person mit geändertem Geschlecht. Dies kann teilweise durch variierende Zusammenfassungen der Geburtsjahre in den verschiedenen Mikrozensen bedingt sein: beispielsweise werden im MZ 2017 die Geburtsjahre 1914 oder früher; 1915/1916 und 1917/1918 zusammengefasst, während im Jahr 2019 diese drei Kategorien als „1918 oder früher“ zusammengefasst werden. Die Inkonsistenz bei der Kombination des Geschlechts und des Geburtsjahres liegt lediglich bei 3,8 %. Diese Werte deuten auf eine hohe Qualität der generierten Panelverknüpfung hin.

Tabelle 14: Konsistenz der Angaben in Bezug auf Geschlecht und Geburtsjahr.

	Konsistente Angaben	
	N	%
Zusammenführungsergebnis		
darunter konsistente Angaben bezüglich...	2.063.888	100,0
... des Geschlechts	2.051.394	99,4
... des Geburtsjahres	1.990.932	96,5
... des Geburtsjahres und des Geschlechts	1.984.539	96,2

4.3 Anwendungsbeispiel: Effekt der Geburt des ersten Kindes auf die Aufteilung der Erwerbsarbeitszeit in Partnerschaften

4.3.1 Deskriptives

Für eine beispielhafte Anwendung wird im Folgenden der Effekt der Geburt des ersten Kindes auf die Aufteilung der Arbeitsstunden im Zeitraum 2016–2019 bei zusammenlebenden Paare untersucht (siehe auch Dribe & Stanfors, 2009; Kühhirt, 2012). Erwartet wird ein „traditionelleres“ Verhältnis der Arbeitszeiten nach der Geburt des ersten Kindes.

Als Analysesample werden gegengeschlechtliche Paare, die in Privathaushalten leben betrachtet, die bei Erstbeobachtung noch kein Kind geboren hatten. Weiterhin werden nur Paare betrachtet, bei denen beide Partner zwischen 25 und 60 Jahre alt sind.

Hierfür betrachten wir 75.499 Beobachtungen von 42.205 gegengeschlechtlichen Paaren untersucht, die bei der ersten Beobachtung kinderlos sind. Im Beobachtungszeitraum 2016 bis 2019 gibt es in dieser Gruppe 7.336 Erstgeburten. Die durchschnittliche Arbeitsteilung im Gesamtsample deutet eine traditionelle innerpartnerschaftliche Verteilung der Arbeitszeit an, wobei Männer in diesem Analysesample im Durchschnitt 7,1 mehr Arbeitsstunden übernehmen als Frauen.

4.3.2 Modelle

Zur Untersuchung wird die abhängige Variable „Arbeitsteilung“ gebildet. Dabei betrachten wir die Differenz aus den wöchentlichen Arbeitsstunden des männlichen Partners und der wöchentlichen Arbeitsstunden der Partnerin. Die unabhängige Variable ist die Anzahl der Kinder im Haushalt. Damit die Koeffizienten kausal interpretiert werden können, muss für Variablen kontrolliert werden, die die unabhängige und abhängige Variable beeinflussen. In diesem Beispiel verwenden wir als zeitlich variierende Kontrollvariablen den Familienstand, Fortbildungen, Stellung im Beruf der beiden Partner und Schichtarbeit der Frau. Die Paneldaten erlauben es, mit einem Fixed-Effects-Schätzer alle anderen zeitkonstanten konfundierenden Variablen implizit zu kontrollieren. Zum Vergleich schätzen wir auch ein Random-Effects-Modell, bei dem wir auch das Alter, die Bildung und den Migrationshintergrund der Partner kontrollieren. Zusätzlich kontrollieren wir für Periodeneffekte.

4.3.3 Ergebnisse

Die Ergebnisse beider Modelle sind in Tabelle 15 abzulesen. Der Effekt der Geburt des ersten Kindes auf die Arbeitsverteilung von gegengeschlechtlichen Partnern beträgt unter Konstanthaltung der genannten Kontrollvariablen im Random-Effects-Modell 4,56 und statistisch signifikant ($p < 0,01$). Demnach ist in Paaren im ersten Jahr nach der Geburt die Arbeitszeit des männlichen Partners um 4,56 Stunden höher als die Arbeitszeit der Frau höher als bei Paaren, die keine Kinder bekommen haben. Dies lässt auf eine Traditionalisierung in der Verteilung der Arbeitszeit nach der Erstgeburt schließen.

Dieser Effekt bleibt auch unter der Berücksichtigung von konstanten unbeobachteten Merkmalen auf Paarebene im Fixed-Effects-Modell bestehen: Ceteris paribus hat der männliche Partner im Jahr nach der Geburt im Vergleich zum Jahr vor der Geburt eine um 4,38 Stunden höhere Arbeitszeit im Vergleich zur Frau als bei Paaren, die kein Kind

bekommen haben. Der geringe Unterschied zwischen den Schätzern des Fixed-Effects- und des Random-Effects-Modells deutet auf einen schwachen einen Selbstselektionseffekt von Paaren mit höherer Neigung zu traditioneller Arbeitsteilung, in die Entscheidung ein Kind zu haben, hin.

Diese Analyse zeigen, dass durch die Nutzung des Paneldatensatzes, wie hier gezeigt, unbeobachtete Selektionseffekte berücksichtigt werden können, was die kausale Interpretation von Effektschätzung plausibler macht. Trotz der strengen Fallauswahl bietet das generierte Mikrozensuspanel ein ausreichend großes Analysesample, um eine zufriedenstellende Schätzgenauigkeit zu erhalten. Die große Anzahl der im Mikrozensus enthaltenen Variablen bietet die Möglichkeit, für zahlreiche konfundierende Variablen zu kontrollieren, um kausale Effekte bestmöglich zu identifizieren.

Tabelle 15: Regressionsergebnisse mit Arbeitszeitverhältnis als Outcome.

	Random-Effects Modell		Fixed-Effects Modell	
	b	se	be	se
Ref: Kein Kind				
Geburt des ersten Kindes	4,56***	0,26	4,38***	0,27
des zweiten Kindes	4,63***	0,28	4,18***	0,29
weiterer Kinder	3,89***	0,55	3,02***	0,57
Weiterbildung Frau	-2,18***	0,16	-0,46*	0,20
Weiterbildung Mann	1,72***	0,16	0,39*	0,20
Schichtdienst Frau	-4,60***	0,20	-2,76***	0,32
Verheiratet vs. Kohabitierend	2,83***	0,16		
Alter Frau	0,09***	0,02		
Alter Mann	0,01	0,02		
Geboren in D vs. geboren im Ausland Frau	-4,51***	0,29		
Geboren in D vs. geboren im Ausland Mann	0,49	0,30		
Ref: Bildung der Frau Kein Schulabschluss, Hauptschule				
Realschule, POS	-2,75***	0,23		
Fachabi, Abi	-4,00***	0,26		
Ref: Bildung der Mannes Kein Schulabschluss, Hauptschule				
Realschule, POS	0,31	0,21		
Fachabi, Abi	0,88***	0,23		
Ref: 2016				
2017	-0,03	0,13	0,07	0,14
2018	-1,05***	0,15	-0,67***	0,17
2019	0,02	0,15	0,50**	0,19
Paar-Jahre	75.499		75.499	
Paare	42.291		42.291	

Quelle: Mikrozensus SUF 2016-2019; *** p<0,01, ** p<0,05, * p<0,10.

5 Schluss

Im vorliegenden Bericht wurde die Erstellung des Mikrozensus Panels der Jahre 2016 bis 2019 skizziert, der daraus resultierende Datensatz vorgestellt und eine beispielhafte Analyse durchgeführt. Insgesamt betrachten wir die Zusammenführung als erfolgreich. Es resultiert ein Datensatz mit 2.300.683 gültigen Personenjahren und 1.035.014 gültigen Haushaltsjahren. Die geringe Anzahl der identifizierten Dubletten sowie das seltene Vorkommen von Ausfällen führt zu keinen wesentlichen Einschränkungen. Des Weiteren sind die Ausfälle auf Personen- und Haushaltsebene durch die Informationen zu den Teilnahmemustern inhaltlich nachvollziehbar auf Zu-/Wegzüge, Geburten oder Sterbefälle zurückzuführen und sind zahlenmäßig unbedenklich, wobei auf einen hohen Anteil kompletter Ausfälle über alle vier Jahre hingewiesen werden muss. Die Konsistenzprüfungen von Geschlechts- und Geburtsjahrveränderungen zeigen keine ungewöhnlichen Muster an. Das vorgestellte do-File kann online abgerufen werden: <https://www.gesis.org/missy/materials/MZ/tools/panel2012>.

Bei der Verwendung des Datensatzes müssen trotz der eingangs genannter Vorteile auch diverse Limitationen berücksichtigt werden. Für das Mikrozensuspanel liegen derzeit keine Längsschnittgewichte vor. Diese können jedoch analog zum Papier von Schimpl-Neimanns (Schimpl-Neimanns, 2021) erstellt werden. Die vierjährlichen Zusatzerhebungen können nicht längsschnittlich ausgewertet werden, da aufgrund des Rotationsdesigns jeder Haushalt an maximal einer Befragungsrunde teilgenommen hat.

6 Literatur

- Dribe, M., & Stanfors, M. (2009). Does Parenthood Strengthen a Traditional Household Division of Labor? Evidence From Sweden. *Journal of Marriage and Family*, 71(1), 33–45. <https://doi.org/10.1111/j.1741-3737.2008.00578.x>
- Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. (2020). *Mikrozensus 2016, SUF, Version 0* (Version 0) [dataset]. RDC of the Federal Statistical Office and the statistical offices of the Länder (RDC). <https://doi.org/10.21242/12211.2016.00.00.3.1.0>
- Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. (2021). *Mikrozensus 2017, SUF, Version 0* (Version 0) [dataset]. RDC of the Federal Statistical Office and the statistical offices of the Länder (RDC). <https://doi.org/10.21242/12211.2017.00.00.3.1.0>
- Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. (2022a). *Mikrozensus 2018, SUF, Version 1* (Version 1) [dataset]. RDC of the Federal Statistical Office and the statistical offices of the Länder (RDC). <https://doi.org/10.21242/12211.2018.00.00.3.1.1>
- Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. (2022b). *Mikrozensus 2019, SUF, Version 0* (Version 0) [dataset]. RDC of the Federal Statistical Office and the statistical offices of the Länder (RDC). <https://doi.org/10.21242/12211.2019.00.00.3.1.0>
- Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und die Arbeitsmarkt-beteiligung sowie die Wohnsituation der Haushalte (Mikrozensusgesetz—MZG), (2016).
- GESIS/Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. (2022). *Wichtige Informationen zur Nutzung des Mikrozensus Scientific Use Files 2019*.
- Herter-Eschweiler, R., & Schimpl-Neimanns, B. (2018). *Möglichkeiten der Verknüpfung von Mikrozensus-Querschnitterhebungen ab 2012 zu Panels*.
- Kühhirt, M. (2012). Childbirth and the Long-Term Division of Labour within Couples: How do Substitution, Bargaining Power, and Norms affect Parents' Time Allocation in West Germany? *European Sociological Review*, 28(5), 565–582. <https://doi.org/10.1093/esr/jcr026>
- Schimpl-Neimanns, B. (2021). Ausfallkorrektur und Designgewichtung im Mikrozensuspanel 2012-2015. *GESIS Papers*. <https://doi.org/10.21241/SSOAR.75410>
- Statistisches Bundesamt. (2021). *Qualitätsbericht—Mikrozensus 2018*. Destatis. <https://www.destatis.de/DE/Methoden/Qualitaet/Qualitaetsberichte/Bevoelkerung/mikrozensus-2018.html>