

Ausfallkorrektur und Designgewichtung im Mikrozensuspanel 2012-2015

Schimpl-Neimanns, Bernhard

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:

Schimpl-Neimanns, B. (2021). *Ausfallkorrektur und Designgewichtung im Mikrozensuspanel 2012-2015*. (GESIS Papers, 2021/14). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.75410>

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 | 14

**Ausfallkorrektur und
Designgewichtung im
Mikrozensuspanel 2012 – 2015**

Bernhard Schimpl-Neimanns

GESIS Papers 2021|14

**Ausfallkorrektur und
Designgewichtung im
Mikrozensuspanel 2012 – 2015**

Bernhard Schimpl-Neimanns

GESIS Papers

GESIS – Leibniz-Institut für Sozialwissenschaften
Monitoring Society and Social Change
Postfach 12 21 55
68072 Mannheim

E-Mail: heike.wirth@gesis.org

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

Zusammenfassung

Seit dem Mikrozensus Scientific Use File von 2012 können die jährlichen Daten zu einem Rotationspanel zusammengeführt werden. Gemäß dem Prinzip der Flächenstichprobe werden fortziehende Personen nicht weiter befragt. Das methodische Kernproblem besteht darin, ob der daraus resultierende Ausfall zu verzerrten Ergebnissen führt. Wenn dieser Ausfall mit Hilfe von Designgewichten korrigiert werden soll, fehlen derzeit einige Informationen in den Daten; unter anderem die Unterscheidung der Ausfallgründe Fortzug und Sterbefall. In diesem Bericht werden die fehlenden Informationen zu Sterbefällen ersatzweise modellgestützt imputiert. Anschließend wird beispielhaft gezeigt, wie damit Längsschnittgewichte erstellt werden können. Des Weiteren werden die Auswirkungen fehlender Informationen zu Sterbefällen untersucht. Falls die Längsschnittpopulationen nicht sachgerecht Sterbefälle enthalten, sind die Längsschnittgewichte höher als bei korrektem Vorgehen und führen zu einer Kompensation Verstorbener. Vor diesem Hintergrund wird eine bessere Datenbereitstellung vorgeschlagen.

Schlüsselwörter Mikrozensus, Scientific-Use-File, Rotationspanel, Nonresponse, Designgewichtung

Nonresponse adjustment and design weights in the German Microcensus panel 2012 – 2015

Abstract

Since the Scientific Use File of 2012, the annual data can be merged into a rotation panel. In accordance with the principle of area sampling, residential movers are not followed. The core methodological problem is whether the resulting dropout leads to biased results. If this dropout is to be corrected by means of design weights, some information is currently missing from the data; among other things, the differentiation of the reasons for dropout – spatial mobility and death. In this report, the missing information on deaths is imputed using a model-based procedure. It is then shown by way of examples how longitudinal weights can be created. Furthermore, effects of missing information on deaths are examined. If the longitudinal populations contain deaths incorrectly, the longitudinal weights are higher than with the correct procedure and lead to compensation of deceased persons. Against this background, better data provision is proposed.

Keywords German Microcensus, Scientific Use File, rotation panel, nonresponse, design weight

1 Einleitung

Seit dem Mikrozensus (MZ) Scientific-Use-File (SUF) 2012 können die Querschnittsdaten des MZ zusammengeführt und zu einem Rotationspanel verknüpft werden. Dies ermöglicht die Konstruktion von Längsschnittdatensätzen, die einen Zeitraum von zwei bis vier Jahren umfassen. Aufgrund des Austauschs von Rotationsgruppen kommt es zu systematischen Ausfällen. Diese können als uneingeschränkt zufällig behandelt werden. Nach dem Konzept der Flächenstichprobe werden die aus dem Auswahlbezirk wegziehenden Haushalte und Personen nicht weiter befragt, sondern durch die zugezogenen Haushalte und Personen ersetzt. Damit fehlen für fortgezogene Einheiten die Angaben nach dem Fortzug und für zugezogene Einheiten liegen keine Informationen vor dem Zuzug vor. Falls der Ausfall infolge räumlicher Mobilität als zufällig zustande gekommen gelten kann, ist das für statistische Schlussfolgerungen auf der Basis räumlich immobiler Personen kein Problem. Dies gilt aber nicht, falls der Fortzug mit einem interessierenden Merkmal zusammenhängt. Ziehen beispielsweise Erwerbslose häufiger um als Erwerbstätige und untersucht man Änderungen des Erwerbsstatus, kann die Analyse auf Basis räumlich immobiler Personen zu verzerrten Ergebnissen führen (vgl. Basic und Rendtel 2006 sowie Schimpl-Neimanns 2008).

Üblicherweise werden zur Korrektur von Ausfällen GewichtungsvARIABLEN eingesetzt. Diese liegen im MZ-Panel ab 2012 nicht vor. Herter-Eschweiler und Schimpl-Neimanns (2018) weisen in Bezug auf Ausfallkorrekturen und zur Konstruktion von Längsschnittgewichten auf Ansätze zum MZ-Panel 1996-1999 hin. Forscher stehen dabei vor einigen Schwierigkeiten. Es fehlen Informationen, die in den früheren MZ-Paneln für Ausfallanalysen verwendet werden konnten. Insbesondere sind dies die Kennung der Rotationsgruppe, die Unterscheidung der Ausfälle nach Sterbefällen und Fortzügen und die Unterscheidung der Zugänge nach Geburten und Zuzügen. Somit lässt sich die räumliche Mobilität nicht genau erfassen.

Geht es um die Untersuchung von Veränderungen auf Individualebene mit Paneldaten, werden in der Regel die Personen mit Angaben ab einschließlich dem ersten Befragungszeitpunkt bis zu ihrem Ausscheiden aus der Population ausgewählt. Nach diesem Konzept zählen später zugezogene oder geborene Personen sowie nach dem ersten Befragungszeitpunkt verstorbene oder ins Ausland verzogene Personen nicht zur Längsschnittpopulation (siehe Herter-Eschweiler 2006: 43 ff.). Um Längsschnittpopulationen statistisch korrekt abzugrenzen, sind somit sowohl Informationen über den Einbezug einer Rotationsgruppe in die Stichprobe als auch Informationen über die Gründe eventueller Ausfälle unverzichtbar.

In diesem Bericht werden die Probleme und Lösungsansätze beschrieben, wie Designgewichte zur Ausfallkorrektur mit dem MZ-Panel 2012 – 2015 erstellt werden können. Im zweiten Abschnitt werden das Stichprobendesign und das Konzept zur Schätzung von Ausfallkorrekturen und zur Konstruktion von Designgewichten dargestellt sowie die Arbeiten zur Gewichtung im MZ-Panel 1996 – 1999 zusammengefasst. Im dritten Abschnitt werden am Beispiel der von 2012 bis 2015 befragten Rotationsgruppe Aufbereitungen der Daten beschrieben. Abschnitt 4 befasst sich ausführlich damit, dass zurzeit mit den Daten Sterbefälle nicht von Fortzügen unterschieden werden können. Hier werden die fehlenden Werte mittels eines multinomialen Logitmodells mit vollständigen Daten des MZ-Panels 2001 – 2004 geschätzt. Diese Schätzungen sind Ausgangspunkt für eine einfache modellgestützte Imputation der unbekanntenen Sterbefälle im neuen MZ-Panel. Anschließend wird skizziert, wie die Längsschnittpopulation abgegrenzt und Längsschnittgewichte konstruiert werden können. Der fünfte Abschnitt enthält empirische Ergebnisse und es wird gezeigt, wie sich diese Längsschnittgewichte von jenen unterscheiden, bei denen Sterbefälle bei der Längsschnittpopulation nicht

berücksichtigt werden. Der Bericht schließt mit einer Zusammenfassung und Vorschlägen zu einer verbesserten Datenbereitstellung.¹

¹ Für hilfreiche Fragen und Kommentare zu vorherigen Versionen danke ich Halil Duran, Robert Herter-Eschweiler, Klaus Pforr und Tobias Roth. Ein besonderer Dank geht an Ulrich Rendtel für frühe Hinweise zur notwendigen Differenzierung der Ausfallgründe und weitere konstruktive Vorschläge. Nicht zuletzt danke ich Heike Wirth für Kommentare und ihre Unterstützung, ohne die diesem Bericht zugrunde liegenden Analysen nicht möglich gewesen wären.

2 Konzept

2.1 Stichprobendesign

Für die Erstellung von Designgewichten sind Informationen über das Stichprobendesign notwendig. Der MZ wird jährlich als geschichtete einstufige Klumpenstichprobe (Flächenstichprobe) durchgeführt. Die Primäreinheiten (PSUs) sind Auswahlbezirke, die aus benachbarten Wohnungen in Gebäuden oder Gebäudeteilen bestehen. Die Ziehungswahrscheinlichkeit eines MZ-Querschnitts beträgt 1 %. Die Auswahlbezirke und die darin wohnenden Haushalte und Personen verbleiben vier Jahre lang in der Stichprobe und bilden ein sogenanntes Rotationsviertel. Ein Viertel der Auswahlbezirke wird jährlich ausgetauscht.

Die Ziehung der 70 %-Substichprobe des SUF orientiert sich eng am Stichprobenplan des MZ. Die Auswahlgesamtheiten sind die Auswahlbezirke jedes Rotationsviertels. Die Ziehungswahrscheinlichkeit beträgt 70 % (für Details siehe Statistisches Bundesamt und GESIS 2019). Die Substichprobe kann als Ergebnis eines 2-phasigen Ziehungsverfahrens interpretiert werden. Die erste Phase umfasst die Auswahl des MZ, die zweite Phase die des SUF. Aufgrund des kleinen Auswahlatzes des MZ genügt aber die Annahme einer einstufigen Auswahl (siehe Särndal et al. 1997: 139 ff.).

Bei einer einstufigen Klumpenstichprobe gelten die Ziehungswahrscheinlichkeiten für alle Einheiten (Auswahlbezirke, Haushalte und Personen). Wird nur eine Rotationsgruppe, d. h. ein Rotationsviertel einer 1 %-Stichprobe betrachtet, ist die Ziehungswahrscheinlichkeit mit 0,25 % zu multiplizieren.

Die durch räumliche Mobilität verursachten Ausfälle im Panel können durch eine entsprechende Gewichtung der räumlich immobilen Personen, die im MZ-Panel verbleiben, kompensiert werden. Während die designbedingten Ziehungswahrscheinlichkeiten bekannt sind, müssen die Wahrscheinlichkeiten für Immobilität geschätzt werden.

Konzeptionell kann der Ausfall durch Fortzüge im Kontext eines zweiphasigen Auswahlverfahrens behandelt werden (Lundström und Särndal 2001: 63 f.; Marek 2005; Särndal et al. 1997: 575 ff.). Die erste Phase ist die Ziehung des SUF mit dem Ergebnis der Stichprobe s , aus der in der zweiten Phase die räumlich immobilen und zur Grundgesamtheit zählenden Personen $r \subset s$ ausgewählt werden. Für räumlich immobile Personen k ist das Designgewicht im MZ-Panel $d_k = 1/\pi_k = 1/(0,01 * 0,70)$, bzw. für ein Rotationsviertel $d_k = 1/(0,01 * 0,25 * 0,70)$. Die Wahrscheinlichkeit für räumliche Immobilität $\hat{\theta}_k = P(k \in r|s)$ kann mit einem Logitmodell geschätzt werden. Der Kehrwert des Produkts der designbedingten Ziehungswahrscheinlichkeit und der Wahrscheinlichkeit für Immobilität ergibt das korrigierte Designgewicht $\tilde{d}_k = 1/(\pi_k * \hat{\theta}_k)$.

2.2 Längsschnittpopulation und Längsschnittgewichte

Die Längsschnittpopulation besteht aus den beim Start des Panels ausgewählten Einheiten bis zu ihrem dauerhaften Ausscheiden aus der Stichprobe. In der Regel werden bei Ausfallanalysen und bei der Erstellung von Korrekturgewichten Personen nicht berücksichtigt, die die Grundgesamtheit verlassen haben. Das betrifft die in der Umfrageforschung als „neutrale Ausfälle“ bezeichneten Fortzüge ins Ausland² und Sterbefälle (Basic et al. 2006: 20-21; Porst 2000: 100; ADM und AG MA 1999: 88;

² Fortzüge ins Ausland können allerdings mit dem MZ nicht ermittelt werden.

Spiess 2005: 11). Daher sollten Ausfälle infolge von Sterbefällen bzw. von nicht der Grundgesamtheit zugehörigen Personen nicht durch Korrekturgewichte kompensiert werden. Späteintritte (Zuzüge) nach dem Panelstart und natürliche Bevölkerungsbewegungen (Geburten und Sterbefälle) zählen ebenfalls nicht zur Längsschnittpopulation. Wie in den früheren MZ-Panels werden gemäß dem sogenannten Weiterverfolgungskonzept temporäre Ausfälle ausgeschlossen (siehe Herter-Eschweiler 2006: 43 ff.; Rohloff 2006 :16). In der von 2012 bis 2015 befragten Rotationsgruppe beträgt der Anteil 0,6 % (siehe Tab. 1). Für diese Personen mit fehlenden Angaben in mindestens einem Jahr ist nicht bekannt, ob sie zum Zeitpunkt des Ausfalls zur Population zählen oder im Ausland oder in einem anderen Haushalt im Bundesgebiet wohnen.

Die Ausfallanalysen konzentrieren sich damit auf die wegziehenden Haushalte und Personen. Das Problem, dass zurzeit die Ausfälle infolge von Sterbefällen und Wegzügen von den echten Ausfällen durch räumliche Mobilität nicht unterschieden werden können, wird vorläufig vernachlässigt, aber unten behandelt. Die Schätzungen der Gewichte für die Übergänge ab 2012 können in enger Anlehnung an Kalton (1987) und Lepkowski (1987) formal wie folgt beschrieben werden (siehe auch Chen et al. 2012).³

Die Teilnahme (response) von Personen k in den einzelnen Wellen wird mit $r = (r_1, r_2, r_3, r_4)$ bezeichnet, wobei die Werte 1 für Teilnahme sowie 0 für Ausfall stehen. Für Ausfallanalysen stehen Designinformationen z zur Verfügung.⁴ Für Befragte, die nach der ersten Erhebung in $t + 1$ ausgefallen sind, liegen Angaben x_t zum vorherigen Zeitpunkt vor. Die Antwortwahrscheinlichkeit des Übergangs t_1/t_2 für den Zeitpunkt t_2 kann dann mit einem Logitmodell bzw. mit einer logistischen Regression von r_2 auf z und x_1 für alle zum Zeitpunkt t_1 befragten Personen ($r_1 = 1$) geschätzt werden. x_t steht beispielhaft auch für weitere erklärende Variablen. Um die Notation einfach zu halten, wird der Index k weggelassen.

Ohne Berücksichtigung des Stichprobendesigns entspricht das Korrekturgewicht $w_{2,1}$ für die zu den Zeitpunkten t_1 und t_2 Befragten, mit dem der Ausfall von Befragten in der zweiten Welle ($r_2 = 0$) kompensiert wird, dem Kehrwert der Antwortwahrscheinlichkeit bzw. der Wahrscheinlichkeit für räumliche Immobilität

$$w_{2,1} = 1 / (\hat{p}(r_2 = 1 | z, x_1, r_1 = 1)).$$

Im MZ-Panel werden die Angaben der Haushalte des Jahresüberhangs, die erst verspätet im Folgejahr $t + 1$ befragt werden konnten, gemäß Stichprobenplan dem Panelzeitpunkt t zugeordnet. Davon ausgehend, dass durch die Zuordnung beim Start des Panels keine weiteren Ausfälle aufgetreten sind, wird w_1 auf 1 gesetzt.⁵ Das Endgewicht für das 2-Wellen-Panel ist $w_2 = w_1 * w_{2,1}$ und das korrigierte Designgewicht

$$\vec{d}_2 = d * w_2.$$

³ Schätzungen zu temporären Ausfällen und Späteintritten (reentry, late entry) beschreibt Lepkowski (1987: 102); siehe auch Rendtel und Harms (2009).

⁴ Im MZ-Panel ab 2012 kommen die Designinformationen Regionalschicht Bundesland (EF1) und die Gebäudegrößenklasse (EF570) infrage.

⁵ Da der MZ 2016 auf einer neuen Stichprobe beruht, können für den Panelzeitpunkt 2015 keine Jahresüberhänge aus dem MZ 2016 zugespielt werden. Im MZ 2012 beträgt der Unit-Nonresponse 1,1 % der Haushalte (ohne Jahresüberhänge), 2013 - 2015 2,4 % (siehe Statistisches Bundesamt und GESIS 2016: 6; 2017, 2018, 2019: 8).

Analog ist das Korrekturgewicht für das 3-Wellen-Panel

$$w_{3.12} = 1 / (\hat{p}(r_3 = 1 | z, x_1, x_2, r_1 = r_2 = 1)),$$

das Endgewicht ist $w_3 = w_1 * w_{2.1} * w_{3.12}$ und das korrigierte Designgewicht

$$\tilde{d}_3 = d * w_3.$$

Für das 4-Wellen-Panel ist das Korrekturgewicht

$$w_{4.123} = 1 / (\hat{p}(r_4 = 1 | z, x_1, x_2, x_3, r_1 = r_2 = r_3 = 1)),$$

das Endgewicht ist $w_4 = w_1 * w_{2.1} * w_{3.12} * w_{4.123}$ und das korrigierte Designgewicht lautet dann

$$\tilde{d}_4 = d * w_4.$$

Abschließend können die korrigierten Designgewichte so normiert werden, dass die Summe der korrigierten Designgewichte der Summe der Designgewichte des Startzeitpunktes entspricht.

Als erklärende Variablen sollten Merkmale verwendet werden, die möglichst eng mit der Teilnahme bzw. dem Ausfall zusammenhängen. Diese Auswahl kann sich an den Analysen zum MZ-Panel 1996 – 1999 orientieren. Letztlich liegt dieser Art der Ausfallkorrektur die nicht überprüfbare Annahme zugrunde, dass die Zusammenhänge der interessierenden Eigenschaften für nicht ausgefallene und ausgefallene Personen gleich sind, der Ausfall also mit dem vorher beobachteten Status, nicht aber mit dem Status beim oder nach dem Ausfall zusammenhängt. Dies entspricht der Annahme eines bedingt zufälligen Ausfalls (missing at random).

2.3 Gewichtung im MZ-Panel 1996 - 1999

Zur Orientierung werden im Folgenden die Arbeiten zum MZ-Panel 1996 – 1999 skizziert. Damals waren Informationen über die Gründe eines Ausfalls (Fortzug, Tod, Zuzug, Geburt) sowohl auf Haushalts- als auch auf Personenebene verfügbar (Herter-Eschweiler 2006: 48-52). Im Längsschnitt 1996 – 1999 (analog 1996 – 1997 und 1996 – 1998) zählten zur Grundgesamtheit alle Personen mit Hauptwohnsitz in einem Privathaushalt, die 1996 oder 1999 in Deutschland wohnten. Die Längsschnitt-Stichprobe wurde zur Vermeidung potenzieller Doppelerfassung eingeschränkter definiert und setzte sich zusammen aus „(...) Personen, die

- in allen vier Jahren in einem Stichprobenauswahlbezirk lebten
- 1996 in einem Stichprobenauswahlbezirk lebten und vor der letzten Befragung im April 1999 verstarben
- 1996 in einem Stichprobenauswahlbezirk lebten und vor der letzten Befragung im April 1999 ins Ausland verzogen
- nach April 1996 geboren wurden und 1999 in einem befragten Haushalt lebten
- nach April 1996 aus dem Ausland in einen Stichprobenauswahlbezirk zuzogen und in diesem 1999 erfasst wurden“ (Rohloff 2006: 8).

Für die Konstruktion von Hochrechnungsgewichten erfolgte eine gleichzeitige Anpassung der Stichprobe (Kalibrierung) an Populationsdaten aus der laufenden Bevölkerungsfortschreibung und Veränderungsdaten (Geburt, Zuzug aus dem Ausland, Sterbefall und Fortzug ins Ausland). Als Eingangsgewichte wurden die mittels Logitmodell geschätzten Wahrscheinlichkeiten („Mobilitätsfaktoren“)

für realisierte Antworten im gesamten Zeitraum 1996 – 1999 berücksichtigt. Bei der Schätzung der Mobilitätsfaktoren wurden Geburten, Zuzüge aus dem Ausland und Sterbefälle ausgeklammert, jedoch bei der Kalibrierung mit extra Hilfsvariablen berücksichtigt (Rohloff 2006: 9, 12).

Im MZ-Panel 1996 – 1999 wurden die Ausfallkorrekturen nur für eine Rotationsgruppe vorgenommen. Im MZ-Panel 2012 – 2015 sind alle im Zeitraum involvierten Rotationsgruppen vertreten. Die Gewichte sind also für alle Rotationsgruppen jeweils getrennt zu ermitteln und anschließend zusammenzuführen. Für die Schätzung des Längsschnittgewichts w_4 wird nur die von 2012 bis 2015 in der Stichprobe enthaltene Rotationsgruppe benötigt (siehe Abschnitt 3.1 und Abb. 1).

Bei der Abgrenzung der Haushalte ist zu berücksichtigen, dass im MZ-Panel 2012 - 2015 auch die Bevölkerung in Gemeinschaftsunterkünften enthalten ist. Diese kann bei der Abgrenzung pragmatisch wie die Bevölkerung in Privathaushalten behandelt werden.

Der MZ ist eine Haushaltsstichprobe. Zwischen den Erhebungszeitpunkten können ganze Haushalte oder einzelne Personen eines Haushaltes ausfallen. Clarke und Chambers (1998) halten für Haushaltsstichproben fest, dass Ausfälle einzelner Personen korreliert sein können und deshalb in Analysen von Individualdaten die üblicherweise getroffene Annahme unabhängig identisch verteilter Schätzfehler nicht zutrifft. Stattdessen wird die Analyse von Haushaltsdaten vorgeschlagen (Clarke und Chambers 1998: 124). Dieser Ansatz wurde im Projekt zum MZ-Panel 1996 – 1999 aufgegriffen: „Ein synthetischer Haushalt ist eine Gruppe von in einem Haushalt zusammenlebenden Individuen, die das gleiche Mobilitätsverhalten aufweisen. Liegt beispielsweise ein Drei-Personenhaushalt vor und eine Person zieht um, so wird dieser Haushalt in zwei synthetische Haushalte untergliedert. Der erste synthetische Haushalt umfasst die räumlich mobile Person und der zweite die beiden räumlich immobilen Personen“ (Marek und Rohloff 2006: 94). Der Vorschlag von Clarke und Chambers lässt sich mit Analysen von Individualdaten umsetzen, wenn die Korrelation in den Klumpen synthetischer Haushalte mit der Option für die Schätzung der Varianz-Kovarianz-Matrix „*vce(cluster clustvar)*“ gewählt wird (siehe StataCorp 2017: 561).

In Bezug auf die Einschränkung der Analysen auf die Bevölkerung am Hauptwohnsitz ist zu beachten, dass es Wechsel des Wohnrechtstatus geben kann. Personen am Nebenwohnsitz werden eventuell ausgeschlossen, obwohl sie zu einem späteren Zeitpunkt zur Bevölkerung am Hauptwohnsitz gehören können. Fallen Personen am Hauptwohnsitz aus, kann dies mit einem nicht beobachteten Wechsel zum Nebenwohnsitz verbunden sein. Zur Abgrenzung der Bevölkerungsgruppen bei der Gewichtung kann eine Alternative überlegt werden.

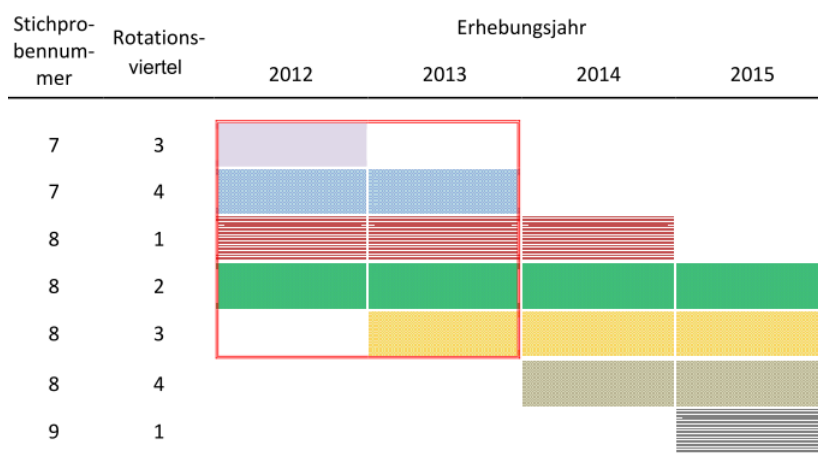
Im neuen MZ-Panel sind die poststratifizierten Anpassungsgewichte – außer im MZ 2012 – für alle Personen eines Haushaltes gleich (siehe zur Hochrechnung im MZ Afentakis und Bihler 2005). Falls ein Haushalt (incl. Gemeinschaftsunterkünfte) nur aus Personen am Nebenwohnsitz besteht, wird diesem das Designgewicht des MZ zugewiesen. Die Schätzung von Gewichten ohne Ausschluss der Bevölkerung am Nebenwohnsitz stellt damit für eine spätere Konstruktion von Anpassungsgewichten kein prinzipielles Problem dar. In die Modelle können geeignete Interaktionen der Design- und Personenangaben mit dem Wohnrechtstatus aufgenommen werden. Dadurch wird die Schätzung zwar etwas aufwändiger, dafür können aber Gewichte sowohl für Haushalte (incl. Gemeinschaftsunterkünfte) als auch für die Bevölkerung am Hauptwohnsitz ermittelt werden.

3 Datenaufbereitung

Werden für die Analysen weitere zeitveränderliche Variablen aufgenommen, sind sie in den SAS- und Stata-Programmen zur Zusammenführung der Querschnittsdaten analog zu den bereits in der Syntax (GESIS 2021) enthaltenen Variablen (z. B. EF46 Geschlecht) zu behandeln.

3.1 Rotationsdesign und Rotationskennung

Mit dem MZ-Panel 2012 – 2015 können verschiedene Übergänge untersucht werden (siehe Abb. 1). Beispielsweise sind beim Übergang 2012/2013 die Rotationsgruppen (ROT) 74, 81 und 82 involviert. Die 2012 ebenfalls enthaltene ROT 73 (Stichprobe 7, Rotationsviertel 3) wurde 2012 das letzte Mal befragt. 2013 kam ROT 83 als Ersatz für die ausgeschiedene ROT 73 hinzu. Das Rotationsdesign erfordert, jeweils die Risikopopulation zu berücksichtigen. D. h., bei den Ausfallanalysen ist die zum Zeitpunkt $t - 1$ letztmals befragte Rotationsgruppe auszuschließen.



Quelle: Herter-Eschweiler und Schimpl-Neimanns 2018: 2.

Abbildung 1: Rotation im Mikrozensus

Für die Definition der Risikopopulation bei Ausfallanalysen ist es notwendig zu wissen, zu welchen Zeitpunkten Auswahlbezirke in die Stichprobe aufgenommen wurden und wie lange sie in der Stichprobe verblieben sind. Leider ist die Rotationskennung im SUF nicht explizit enthalten, sodass Nutzer selbst eine näherungsweise Bestimmung der Rotationsgruppenzugehörigkeit vornehmen müssen. Hierfür kann die Typisierung für Auswahlbezirke „ptyp_psu“ herangezogen werden (siehe z. B. https://www.gesis.org/missy/files/documents/MZ/panel_12131415_v1.do, Zeilen 584-628, 823-848, 1158-1208). Mit den Angaben zur Zahl der Befragten pro Panelzeitpunkt im Auswahlbezirk lässt sich der Verbleib eines Auswahlbezirks im Zeitverlauf ermitteln. In Kombination mit der Angabe -5 „Entfällt (Keine Wiederholungsbefragung)“ in Variable EF45 „Zuzug (i. d. letzten 12 Monaten)“ können erstmals befragte Rotationsgruppen identifiziert werden.

Die Abgrenzung funktioniert jedoch nicht für alle Haushalte. Für Haushalte des Jahresüberhangs, temporär ausgefallene Haushalte und später in einen bereits zuvor in der Stichprobe enthaltenen Auswahlbezirk Zugezogene (Späteintritte) können Informationen zur Zuordnung zu Rotationsgruppen fehlen. Dies betrifft auch sogenannte Ausfallsätze; das sind Auswahlbezirke ohne einen befragten Haushalt, leerstehende Wohnungen, ausgefallene Privathaushalte und Gemeinschaftsunterkünfte.

Um eine bessere Zuordnung zu Rotationsgruppen zu erreichen, werden in einem ersten Schritt die Wertebereiche der längsschnittkonsistenten Auswahlbezirksnummern der gemäß Stichprobenplan erfolgreich befragten Haushalte ermittelt. Um alle Haushalte zuzuordnen, werden in einem nächsten Schritt diese Wertebereiche so vergrößert, dass sie mindestens ein Viertel der Anzahl der Auswahlbezirke eines MZ-Querschnitts (siehe Statistisches Bundesamt 2016: 5) umfassen und gleichzeitig keine Überschneidungen mit Wertebereichen anderer vermuteter Rotationsgruppen auftreten. Auf diese Weise können für die von 2012 bis 2015 befragte Rotationsgruppe rund 2,5 % Personen (ohne Ausfallsätze) mehr der Rotationsgruppe 82 zugeordnet werden. Für diese Rotationsgruppe können in Tab. 1 folgende Teilnahme- bzw. Ausfallmuster unterschieden werden.

Tabelle 1: Teilnahme-/Ausfallmuster der von 2012 bis 2015 befragten Rotationsgruppe (ROT 82; Proxy)

Teilnahme / Ausfall	n	%
[0] Kein Ausfall, Teilnahme 2012 - 2015	88.087	56,0
[1] Dauerhafter Ausfall nach 2012 (attrition)	32.858	20,9
[2] Temporärer Ausfall nach 2012 mit Wiedereintritt	993	0,6
[3] Späteintritt nach 2012 (Zuzug, Geburt)	35.241	22,4
Zusammen	157.179	100,0
[-1] Ausfallsätze	33.206	
Insgesamt	190.385	

Quelle: FDZ der Statistischen Ämter des Bundes und der Länder, Mikrozensus 2012-2015, Scientific Use Files (SUF), eigene Berechnungen. Fallzahlen ohne Hochrechnung; Auswahl: wohnberechtigte Bevölkerung.

3.2 Längsschnittkonsistente Angaben zum Geburtsjahr und Geschlecht

In Panels sind häufig Inkonsistenzen von im Prinzip zeitkonstanten Merkmalen festzustellen. Unter den im MZ-Panel 2012 bis 2015 Befragten sind rund 1 % der Angaben zum Geschlecht und rund 5 % der Angaben zum Geburtsjahr inkonsistent (siehe GESIS 2021). Bei den Ausfallanalysen sollten aber solche Inkonsistenzen aufgrund von Mess- oder Klassifikationsfehlern möglichst vermieden werden. In Anlehnung an die früheren MZ-Panels 1996 – 1999 und 2001 – 2004 werden deshalb die Angaben bereinigt. Dabei wird der am häufigsten vorkommende Wert übernommen. Falls dies nicht möglich ist, wie z. B. im Fall der Antwortfolge männlich – weiblich – männlich – weiblich, wird die erste Antwort übernommen.

⁶ Siehe dazu Herter-Eschweiler 2006: 38: „Bei diesen temporären Ausfällen ist allerdings zu beachten, dass es sich grundsätzlich um Ausfälle aller Haushaltsmitglieder handelt, weil der zu befragende Haushalt im betreffenden Erhebungsjahr nicht erreicht wurde.“

4 Unterscheidung der Ausfälle nach Ausfallgründen

Für die Unterscheidung der Ausfallgründe Sterbefall vs. Fortzug stehen in den früheren MZ-Panels 1996 – 1999 und 2001 – 2004 sowie im Campus-File MZ-Panel 1996 – 1999 entsprechende Informationen sowohl für Haushalte als auch für Personen bereit. Die den Typisierungen zugrunde liegenden Fragen werden auch im MZ ab 2012 gestellt und lauten:

- Frage 2: „Falls Ihr Haushalt bereits im Mikrozensus befragt wurde: Sind in den letzten 12 Monaten Mitglieder Ihres Haushalts verstorben?“ (Variable EF22 der Originaldaten enthält bei der Antwort „Ja“ die Anzahl der Verstorbenen.)
- Frage 3: Falls Ihr Haushalt bereits im Mikrozensus befragt wurde: Sind in den letzten 12 Monaten Mitglieder Ihres Haushalts fortgezogen?“ (Variable EF21 der Originaldaten enthält bei der Antwort „Ja“ die Anzahl der Fortgezogenen.)

Die Retrospektivangaben beziehen sich auf den gesamten Zeitraum seit der letzten Erhebung und können damit auch Ereignisse für zwischenzeitlich zugezogene oder nicht mehr im Haushalt wohnende Personen enthalten. Außerdem liegen die Angaben nur auf der Haushaltsebene vor.⁷ Eindeutige Bezüge zu Personen sind somit nur eingeschränkt möglich (Herter-Eschweiler 2006: 48). Sind beispielsweise zwei Personen ausgefallen und im Haushalt ist eine Person fortgezogen und eine Person verstorben, kann keiner Person eines der Ereignisse zugeordnet werden. Dies erklärt, dass häufig nicht zwischen Sterbefall und Fortzug unterschieden werden kann. Außerdem ist davon auszugehen, dass bei einem vollständigen Ausfall des Haushalts keine Informationen vorliegen. Ausgenommen sind Konstellationen, in denen zwar alle zum Zeitpunkt t befragten Haushaltsmitglieder bei der nachfolgenden Befragung $t + 1$ ausgefallen sind, aber zwischenzeitlich neu zugezogene Haushaltsmitglieder Auskunft über fortgezogene oder verstorbene Personen geben können. Ausgenommen sind auch temporäre Ausfälle von Personen. Außer beim letzten beobachteten Ausfall können diese den Fortzügen zugeordnet werden.

Ohne die Angaben zu den Ausfallgründen (EF21 und EF22) können mit dem Rotationspanel des MZ 2012-2015 somit nur folgende Sachverhalte ermittelt werden:

- Ausfälle einzelner Haushaltsmitglieder, ohne Angabe zu den Ausfallgründen: Sterbefall (1), Fortzug (2) und Fortzug oder Sterbefall (3).
- Vollständiger Ausfall des Haushalts ohne zwischenzeitliche Zuzüge (4).

Eine Erstellung von Gewichten zur Ausfallkorrektur wäre deshalb nur durch den Einbezug der unbekanntesten Sterbefälle möglich. Ab dem Alter von 65 Jahren liegen aber überdurchschnittlich viele Sterbefälle vor. Ein fehlender Ausschluss der Sterbefälle führt zu einem höheren Gewicht als bei einem fachlich korrekten Ausschluss der Sterbefälle und damit zu einer Kompensation von Verstorbene.

⁷ Im Folgenden umfasst der technische Begriff Haushalt sowohl Privathaushalte als auch Gemeinschaftsunterkünfte.

4.1 Daten zur Imputation der Sterbefälle

Falls nicht klar zwischen Sterbefall oder Fortzug unterschieden werden kann, lässt sich eventuell ein Imputationsmodell aus dem MZ-Panel 2001 – 2004 benutzen. Zuerst sind aber die Daten zu beschreiben. Diese enthalten eine Unterstichprobe der von 2001 bis 2004 befragten Rotationsgruppe 53 (Stichprobe 5, Rotationsviertel 3) im Umfang von 70 % des ursprünglichen MZ. Für die Zusammenführung der Querschnittsangaben wurden die Ordnungsnummern des Auswahlbezirks und des Haushalts, das Geburtsjahr und das Geschlecht einer Person, ergänzt um das Eheschließungsjahr und Informationen über haushalts- oder personenbezogene Veränderungen verwendet. Damit konnten 97 % der Datensätze von Personen aus Privathaushalten und 89 % der Personen aus Gemeinschaftsunterkünften längsschnittplausibel zusammengeführt werden.

Tabelle 2: Art der Ausfälle nach Teilnahmemuster im MZ-Panel 2001-2004

Art der Ausfälle	Teilnahmemuster (20..; X=Ausfall)			Insgesamt
	01XXXXXX	0102XXXX	010203XX	
Insgesamt	120.758	103.421	93.334	120.758
Kein Ausfall	103.421	93.334	84.443	84.443
Ausfälle insgesamt	17.337	10.087	8.891	36.315
Ausfälle insgesamt (in %)	14,4	9,8	9,5	30,1
Darunter (in %):				
[1] Sterbefall	4,6	6,5	5,2	5,3
[2] Fortzug	11,3	13,2	12,9	12,2
[3] Fortzug oder Sterbefall	20,4	8,9	21,5	17,5
[4] Vollständiger Ausfall des Haushalts	63,7	71,4	60,4	65,0
Anteil Sterbefall an Ausfällen der Typen 1 - 3 (in %)	12,5	22,8	13,2	15,1

Quelle: FDZ der Statistischen Ämter des Bundes und der Länder, Mikrozensus-Panel 2001-2004, Scientific Use File (SUF), DOI: 10.21242/12211.2004.00.01.3.1.0, eigene Berechnungen; Auswahl: Weiterverfolgungskonzept; plausibel zusammengeführte Personensätze 2001 – 2004 ($perkl \geq 1$ & $perkl \leq 4$); wohnberechtigte Bevölkerung. Fallzahltabellen ohne Hochrechnung.

Nach dem Ausschluss von Haushalten mit temporären Ausfällen stehen insgesamt für rund 120.000 der ab 2001 befragten Personen Längsschnittangaben zur Verfügung (Statistisches Bundesamt 2009: 33-50). Die Angaben zum Ausfall in der Typisierung verlustp „Kennzeichnung der Art des Verlustes auf Personenebene“ werden nachträglich analog zur oben skizzierten Abgrenzung recodiert. Tab. 2 zeigt, dass sich sowohl die Ausfallquoten als auch die Anteile der Ausfallgründe zwischen den einzelnen Übergängen unterscheiden.

Ab dem MZ 2012 sind alle jeweils in einem MZ-Querschnitt vertretenen Rotationsgruppen in der Auswahlgesamtheit des SUF enthalten (Herter-Eschweiler und Schimpl-Neimanns 2018). Zum Vergleich mit Tab. 2 enthält Tab. 3 für das neue MZ-Panel nur Angaben der von 2012 bis 2015 befragten Rotationsgruppe nach dem Weiterverfolgungskonzept (siehe Abschnitt 2.2), also wie im älteren Panel ohne temporäre Ausfälle, Geburten und Zuzüge. In Bezug auf die Ausfallquoten zeigen sich geringere Unterschiede zwischen den einzelnen Übergängen als im älteren Panel. Jedoch variieren die Verhältnisse der unbekannteren Ausfallgründe (.) vs. vollständig ausgefallener Haushalte (4). In Bezug auf diese vierte Kategorie ist zu anmerken, dass sich in beiden Panels die Art des Ausfalls nicht exakt bestimmen lässt.

Tabelle 3: Art der Ausfälle der 2012 bis 2015 befragten Rotationsgruppe nach Teilnahmemuster im MZ-Panel 2012-2015

Art der Ausfälle	Teilnahmemuster (20.; X=Ausfall)			Insgesamt
	12XXXXXX	1213XXXX	121314XX	
Insgesamt	120.945	107.948	97.964	120.945
Kein Ausfall	107.948	97.964	88.087	88.087
Ausfälle insgesamt	12.997	9.984	9.877	32.858
Ausfälle insgesamt (in %)	10,8	9,2	10,1	27,2
Darunter (in %)				
[.] Art unbekannt	25,3	26,5	22,7	29,6
[4] Vollständiger Ausfall des Haushalts	74,7	73,5	77,3	70,4

Quelle: Siehe Tab. 1.

4.2 Imputationsmodell

Da bei einem vollständigen Ausfall des Haushalts keine Informationen zum Ausfallgrund vorliegen (s. o.) konzentriert sich die Modellierung auf die Kategorien 1 „Sterbefall“, 2 „Fortzug“ und 3 „Fortzug oder Sterbefall“ des MZ-Panels 2001 – 2004 (siehe Tab. 2). Hierfür kann ein multinomiales Logitmodell verwendet werden. Im MZ-Panel 2012 – 2015 entspricht die Auswahl der Kategorie „Art unbekannt“ mit fehlenden Angaben des Ausfallgrundes (siehe Tab. 3).

Falls angenommen werden kann, dass sich die Zusammenhänge zwischen dem Ausfall und den Ausfallgründen nicht gravierend verändert haben, können die Schätzungen mit dem MZ-Panel 2001 – 2004 als Analytestichprobe für Vorhersagen des Ausfallgeschehens im MZ-Panel 2012 – 2015 verwendet werden. Technisch lässt sich dies einfach mit einer sogenannten „out-of-sample estimation“ durchführen (StataCorp 2017: 2015) umsetzen. Dabei werden die Daten des neuen MZ-Panels, in dem die Angaben zum Ausfallgrund fehlen und mit „.“ codiert sind, an die Daten des älteren MZ-Panels mit gültigen Angaben zum Ausfallgrund angehängt. Durch das Modell werden Prädiktoren sowie dazugehörige Standardfehler etc. auch für die Daten mit fehlenden Werten geschätzt. Damit liegen Schätzungen nicht nur für jede Kombination der in beiden Daten vorkommenden Kovariaten, sondern auch für eventuell nicht enthaltene Variablenkonstellationen vor.

Einerseits sollten mit dem Modell die systematischen Zusammenhänge zwischen den typisierten Ausfallgründen und den erklärenden Variablen möglichst gut geschätzt werden können. Andererseits sollte das Modell sparsam sein, damit möglichst wenige Besonderheiten der alten Daten zum Tragen kommen. Es sind keine Varianzschätzungen geplant, sodass einfache Imputationen infrage kommen.

Eigentlich sind die Übergänge separat zu behandeln. Weil sich aber im MZ-Panel 2001 – 2004 die Ausfallquoten und Anteile der Ausfallgründe zwischen den einzelnen Übergängen unterscheiden (s. o.), erscheint es ratsam, im MZ-Panel 2001 – 2004 die Ausfälle pro Übergang zusammenzufassen und den Durchschnitt der drei Übergänge zu betrachten. Entsprechend wird den erklärenden Variablen jeweils der Status vor dem Ausfall zugewiesen. Dieses Vorgehen ist unproblematisch, da die Analytestichprobe nach dem Weiterverfolgungskonzept ausgewählt wurde und die Ereignisse deshalb nur ein Mal gezählt werden. Den einzelnen Übergängen des MZ-Panels 2012 – 2015 werden also bei der Schätzung jeweils die Durchschnitte des älteren MZ-Panels 2001 – 2004 gegenübergestellt.

In erster Linie sollen alters- und geschlechtsspezifische Zusammenhänge mit der Sterblichkeit untersucht werden. Als erklärende Variablen werden verwendet: Die Altersgruppe (im Wesentlichen in

Schritten von 5 Jahren: unter 1 bis 4 Jahre, ..., 85 bis 89 Jahre, 90 Jahre und älter), die Interaktion von Geschlecht mit dem Alter (in Jahren) und dem quadrierten Alter. Hierbei werden die längsschnittkonsistenten Angaben verwendet. Zur näherungsweise Erfassung der Schwierigkeiten der Typisierung der Ausfallgründe mit Hilfe der Retrospektivfragen und den Antworten auf Haushaltsebene (s. o.) wird ergänzend die Differenz der Zahl der Haushaltsmitglieder zwischen dem Zeitpunkt t und dem vorherigen Zeitpunkt $t - 1$ ermittelt und wie folgt kategorisiert: ≤ 0 Zuzug, einschließlich keine Veränderung; 1 Person ausgefallen; 2 Personen ausgefallen; 3 und mehr Personen ausgefallen.

4.3 Klassifikation der Modellergebnisse

Mit dem multinomialen Logitmodell werden Anteile für alle Kategorien der abhängigen Variablen geschätzt. Um die Prognosegüte zu beurteilen, ist eine Klassifikation der geschätzten Anteile hilfreich. Dabei werden in Anlehnung an einen Vorschlag in StataCorp (2017: 1678-1679) Aspekte der Sensibilität und Spezifität berücksichtigt. Die „Trefferquote“ sollte möglichst gut sein. Das heißt, der Anteil der korrekt klassifizierten Sterbefälle sollte einerseits hoch sein. Dieses „richtig-positive“ Ergebnis wird als Sensibilität (sensitivity) bezeichnet. Falls Fortzüge sowie Fortzüge oder Sterbefälle vorliegen, sollten außerdem die geschätzten Anteile für Sterbefälle möglichst gering sein. Dieser Aspekt des „richtig-negativen“ Ergebnisses ist die Spezifität (specificity). Je größer also die Summe der Sensibilität und der Spezifität ist, desto besser ist die Schätzung zu beurteilen. Konkret wird wie folgt vorgegangen:

- Klassifikation als Sterbefall, falls der geschätzte Anteil für einen Sterbefall (1) größer als die Summe der geschätzten Anteile für Fortzug (2) und Fortzug oder Sterbefall (3) ist (d. h. größer als 50 %) und der z -Wert der Differenz der Prädiktoren für Sterbefall (1) und Fortzug (2) größer als 1,96 ist. Es wird eine Irrtumswahrscheinlichkeit von 5 % angenommen.
- Klassifikation als Fortzug, falls der geschätzte Anteil für Fortzug (2) größer als der geschätzte Anteil für Sterbefall (1) und größer als der geschätzte Anteil für Fortzug oder Sterbefall (3) ist und der z -Wert der Differenz der Prädiktoren für Sterbefall (1) und Fortzug (2) kleiner als -1,96 ist.
- Die übrigen Fälle werden der nicht eindeutigen Restkategorie Fortzug oder Sterbefall (3) zugewiesen.

Um zu testen, ob sich diese Ideen für den Einsatz mit dem MZ-Panel 2012 - 2015 eignen, wird der Datensatz des MZ-Panels 2001 – 2004 zufällig in zwei Teile geteilt. Der erste Teil enthält gültige Angaben zum Ausfallgrund, während diese im zweiten Teil als fehlend behandelt werden.

4.4 Validierung mit dem MZ-Panel 2001 – 2004

Die zufällige Aufteilung der Daten des MZ-Panels 2001 – 2004 mit den Ausfallgründen 1 bis 3 in zwei Teile bietet für die im zweiten Teil als fehlend behandelten Angaben zum Ausfallgrund den Vergleich der Schätzungen mit den „beobachteten“ bzw. typisierten Angaben. Mit dem MZ-Panel 2012 – 2015 ist diese Prüfung nicht möglich. Die Daten wurden vor der zufälligen Aufteilung nach den Ausfallgründen pro Übergang sowie nach Bundesland (West vs. Ost) und Geschlecht geschichtet. Zwischen den beiden Teilen kann es also Unterschiede der bei der Schichtung unberücksichtigten Merkmale geben. Informationen dazu finden sich im Anhang in Tab. A1. Außerdem ist zu beachten, dass diese Aufteilung nur ein Ergebnis von vielen möglichen Aufteilungen ist.

Wie oben erläutert, werden die Angaben zum Ausfallgrund der zweiten Teilstichprobe als fehlend behandelt. Die Daten der ersten Teilstichprobe mit den gültigen Angaben bilden die Analytestichprobe für das multinomiale Logitmodell, mit dem unter den für die erste Teilstichprobe geltenden Zusammenhängen auch für die in der zweiten Teilstichprobe vorliegenden Kovariaten Schätzungen der Anteile und Prädiktionen usw. erfolgen.

Das nur für die erste, als Validierungsstichprobe dienende Teilstichprobe geltende Pseudo- R^2 beträgt 0,4741. Die Ergebnisse werden wie beschrieben klassifiziert und für beide Teilstichproben in Tab. 4 dargestellt. Im Wesentlichen werden für beide Teile gleiche Ergebnisse erzielt. Die geringen Differenzen hängen mit zufälligen Merkmalsunterschieden der Teilstichproben zusammen.

Wie Tab. 4 zeigt, ist die Prognosegüte des Modells eher mäßig. Zwar ist die Diagonale der Schätzwerte für jede Teilstichprobe am stärksten besetzt und damit der Anteil der jeweils korrekt klassifizierten Kategorien hoch. Es gibt aber auch beträchtliche Fehlklassifikationen. Beispielsweise ist in Bezug auf Sterbefälle in der Teilstichprobe 2 mit fehlenden Angaben zum Ausfallgrund festzuhalten, dass diese durch das Modell um rund 5 % ($= 1 - (913/961)$) unterschätzt werden. Die Sensibilität (jeweils binär betrachtet) beträgt rund 72 %, d. h. 28 % sind anders klassifiziert, nämlich 7 % als Fortzug und 21 % als Fortzug oder Sterbefall. Dagegen ist die Spezifität mit rund 96 % sehr hoch.

Nimmt man die geschätzten und klassifizierten Sterbefälle für Imputationen bei der Konstruktion von Ausfallkorrekturen bzw. Gewichten, ist zu beachten, dass lediglich 76 % auch „beobachtete“ bzw. typisierte Sterbefälle sind. Weiter sind unter den klassifizierten Sterbefällen 9 % als Fortzug und 15 % als Fortzug oder Sterbefall „beobachtet“ bzw. typisiert. Somit ist Ausfallkorrektur mit dem Risiko verbunden, dass die durch Fortzug entstandenen Ausfälle nicht korrekt kompensiert werden.

Die in Tab. 4 berichteten Statistiken Sensibilität und Spezifität beziehen sich jeweils auf die betreffende Kategorie vs. eine andere Kategorie; z. B. Sterbefall vs. Sonst. Betrachtet man nur die Kategorien Sterbefall und Fortzug, stehen den $n = 85$ fehlklassifizierten Fortzügen unter den klassifizierten Sterbefällen $n = 71$ als Fortzüge klassifizierte, aber typisierte Sterbefälle gegenüber.

Konzentriert man sich auf die Kategorien Sterbefall und die Restkategorie, stehen $n = 135$ Fehlklassifikationen der Gruppe 3 „Fortzug oder Sterbefall“ in der Klasse der Sterbefälle $n = 197$ Sterbefälle gegenüber, die als Fortzug oder Sterbefall klassifiziert sind. Der Anteil von Gruppe 3 „Fortzug oder Sterbefall“ unter den klassifizierten Sterbefällen lässt sich damit verantworten, dass immerhin eine Schätzung für Sterbefälle in der nicht eindeutigen Restkategorie vorliegt.

Tabelle 4: Typisierte Art der Ausfälle und Klassifikation der Ergebnisse des Logitmodells für beide Teilstichproben (Schätzwerte, Zeilen- und Spaltenprozentwerte (kursiv))

Klassifikation der Modellergebnisse								
Typisierte Art der Ausfälle	Teilstichprobe 1 (gültige Angaben zum Ausfallgrund)				Teilstichprobe 2 (Ausfallgrund als fehlend recodiert)			
	[1]	[2]	[3]	Insg.	[1]	[2]	[3]	Insg.
[1] Sterbefall	711	71	170	952	693	71	197	961
	74,7	7,5	17,9	100,0	72,1	7,4	20,5	100,0
	77,7	2,7	6,2	15,1	75,9	2,7	7,0	15,0
[2] Fortzug	89	1.860	264	2.213	85	1.864	262	2.211
	4,0	84,1	11,9	100,0	3,8	84,3	11,9	100,0
	9,7	71,0	9,6	35,2	9,3	69,6	9,3	34,6
[3] Fortzug oder Sterbefall	115	689	2.324	3.128	135	743	2.349	3.227
	3,7	22,0	74,3	100,0	4,2	23,0	72,8	100,0
	12,6	26,3	84,3	49,7	14,8	27,7	83,7	50,4
Insgesamt	915	2.620	2.758	6.293	913	2.678	2.808	6.399
	14,5	41,6	43,8	100,0	14,3	41,9	43,9	100,0
	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Klassifikation								
Über-/Unterschätzung (%)	-3,9	18,4	-11,8		-5,0	21,1	-13,0	
Klassifikation korrekt (%)	77,7	71,0	84,3		75,9	69,6	83,7	
Sensibilität (%)	74,7	84,1	74,3		72,1	84,3	72,9	
Spezifität (%)	96,2	81,4	86,3		95,9	80,6	85,5	

Quelle: Siehe Tab. 2.

4.5 Ergebnisse des Logitmodells mit den MZ-Panels 2001 – 2004 und 2012 – 2015

Nach dieser Validierung werden zunächst die Verteilungen der beiden Datensätze dargestellt. Die Daten des MZ-Panels 2001 – 2004 enthalten die Ausfallgründe 1 – 3 und den vollen Datensatz ohne Aufteilung in Teilstichproben (siehe Tab. 2). Der Einfachheit halber wird für das neuere Panel nur der Übergang 2012/2013 für die Rotationsgruppe 82 und die Auswahl der unbekanntenen Ausfallgründe (siehe Tab. 3) vorgestellt.

Tab. 5 zeigt, dass sich die Verteilungen teilweise beträchtlich unterscheiden. Dies betrifft insbesondere die Differenz der Haushaltsgrößen zum Zeitpunkt $t - 1$ und t . In der obigen Validierung mit den Teilstichproben des MZ-Panels 2001 – 2004 ist die Unterscheidung danach sehr wichtig (vgl. Tab. A2 und A3). Während im neueren Panel der Anteil der Kategorie „Zuzug oder keine Veränderung“ nur rund 8 % beträgt, sind es im älteren MZ-Panel durchschnittlich 37 %. Ähnliche Unterschiede treten bei der Haushaltsgröße zum Zeitpunkt $t-1$ auf. Im MZ-Panel 2001 – 2004 haben 1-Personenhaushalte einen Anteil von durchschnittlich rund 8 %. Im neuen Panel liegt der Anteil 2012 dagegen bei 0,4 %. Da für 1-Personenhaushalte mehrheitlich keine Zuzüge feststellbar sind, zählen sie zu der in den Analysen ausgeschlossenen 4. Kategorie „Vollständiger Ausfall des Haushalts ohne Zuzug und ohne Angabe zu den Ausfallgründen“. Worauf diese Änderungen der Zusammenhänge zurückzuführen sind, kann letztlich nicht geklärt werden. Die gravierenden Unterschiede sprechen dafür, das Merkmal Differenz der Haushaltsgrößen nicht weiter zu verwenden. Um die Konsequenzen der Verteilungsunterschiede für die Schätzungen zu prüfen, wird es zunächst dennoch berücksichtigt.

Tabelle 5: Statische Kennwerte zu den Stichproben der MZ-Panels 2001 – 2004 und 2012 – 2015 (Übergang 2012/2013)

	2001 - 2004	2012/2013
Stichprobenumfang (n)	12.692	3.285
Alter in Jahren		
1. Perzentil	13	18
1. Quartil	21	22
Median	32	28
3. Quartil	61	59
9. Perzentil	79	82
Mittelwert	40,1	40,2
Standardabweichung	24,8	24,9
Variationskoeffizient	0,6197	0,6200
Geschlecht (%)		
[1] Männlich	50,1	52,6
[2] Weiblich	49,9	47,4
Familienstand (%)		
[1] ledig	56,4	63,7
[2] verheiratet, eingetr. Lebenspartnersch. (eLP)	29,8	25,3
[3] verwitwet, eLP verstorben	9,4	7,5
[4] geschieden, eLP aufgehoben	4,4	3,5
Haushaltstyp (%)		
[0] Privathaushalt	95,4	91,5
[1] Gemeinschaftsunterkunft	4,6	8,5
Haupt-/Nebenwohnsitz (%)		
[0] Hauptwohnsitz	97,4	96,6
[1] Nebenwohnsitz	2,6	3,4
Differenz der Haushaltsgrößen zum Zeitpunkt t-1 - t (%)		
[0] Zuzüge oder keine Veränderung (Diff. ≤ 0)	36,9	7,8
[1] 1 Person ausgefallen	47,5	69,6
[2] 2 Personen ausgefallen	10,0	15,5
[3] 3 und mehr Personen ausgefallen	5,6	7,1

Quelle: Siehe Tab. 1 und Tab. 2.

Tab. 6 enthält Kriterien für die Modellwahl. Sie konzentriert sich auf die klassifizierten Sterbefälle im Vergleich zu den „beobachteten“ bzw. typisierten Sterbefällen. Das Modell M3 enthält das kritische Merkmal Differenz der Haushaltsgrößen. Mit ihm werden im Vergleich zur Typisierung weniger Sterbefälle geschätzt (-2,4 %). Für das MZ-Panel 2012 – 2015 und den Übergang 2012/2013 wird aber mit rund 20 % ein höherer Anteil als Sterbefall klassifiziert als im älteren Panel (zur Klassifikation siehe Abschnitt 4.3). Der demographische Wandel führte dazu, dass der Anteil älterer Personen, die eine höhere statistische Sterblichkeit haben, im Vergleich zu 2001 gestiegen ist. Einerseits ist deshalb für die neuen Daten mit einem höheren geschätzten Anteil von Sterbefällen zu rechnen. Andererseits hat sich im Zeitverlauf die Sterbewahrscheinlichkeiten nach Alter und Geschlecht 2012 – 2014 um durchschnittlich rund 0,3 Prozentpunkte gegenüber 2001 – 2003 verringert (siehe Statistisches Bundesamt 2021). Der Anteil klassifizierter Sterbefälle von 20 % im neuen MZ-Panel erscheint deshalb unplausibel hoch und dürfte mit den Verteilungsunterschieden des Merkmals Differenz der Haushaltsgrößen zusammenhängen.

Mit dem Modell M2 wird auf die Differenz der Haushaltsgrößen verzichtet. Stattdessen werden der Familienstand, der Haushaltstyp und der Wohnrechtstatus aufgenommen, obwohl es ebenfalls Verteilungsunterschiede gibt. Im MZ-Panel 2001 – 2004 werden Sterbefälle stärker unterschätzt (-9,7 %) als mit Modell M3. Der Anteil klassifizierter Sterbefälle im neuen MZ-Panel (Übergang 2012/2013) beträgt mit Modell M2 rund 18 %.

Modell M1 enthält nur die erklärenden Variablen Altersgruppe und die Interaktionen von Alter und Alter² mit Geschlecht (siehe Tab. A3). Die klassifizierten Sterbefälle unterschätzen die typisierten Sterbefälle im MZ-Panel 2001 – 2004 (-5,5 %). Für den Übergang 2012/13 im neuen Panel werden 17 % Sterbefälle geschätzt.

Tabelle 6: Klassifikation der Ergebnisse der Logitmodelle mit dem MZ-Panel 2001-2004 als Analysestichprobe

Modelle	M1	M2	M3
Log Pseudolikelihood Ratio Chi ²	6.292,24	7.032,45	11.789,98
d.f.	44	54	50
BIC	19.499,38	18.853,65	14.058,32
Pseudo R ² (McFadden)	0,2481	0,2773	0,4650
Klassifikation des Ausfalltyps (%)			
\hat{P} [1] Sterbefall	14,2	13,6	14,7
\hat{P} [2] Fortzug	31,8	32,6	42,3
\hat{P} [3] Fortzug/Sterbefall	54,0	53,8	43,0
Über-/Unterschätzung \hat{P} [1]	-5,5	-9,7	-2,4
Klassifikation \hat{P} [1] korrekt	61,8	72,6	76,7

Quelle: Siehe Tab. 1 und Tab. 2. Modellterme: M1: Altersgruppe, (Alter, Alter²)*Geschlecht. M2: M1 + Familienstand + Haushaltstyp + Haupt-/Nebenwohnsitz. M3: M1 + Differenz der Haushaltsgrößen.

Auf das Merkmal Differenz der Haushaltsgrößen wird aufgrund der gravierenden Verteilungsunterschiede verzichtet. Damit entfällt Modell M3. Bei der Entscheidung für ein Modell sprechen die Höhe des Pseudo-R², der geringere BIC-Wert und der höhere Anteil korrekt klassifizierter Sterbefälle für M2. Für M1 spricht die geringere Unterschätzung der Sterbefälle. Da dies auch für das neue MZ-Panel anzunehmen ist, wird M1 präferiert.

In Tab. 7 sind die nach dem Modell M1 klassifizierten Modellergebnisse für das MZ-Panel 2012 – 2015 zusammenfassend dargestellt (siehe zum Vergleich Tab. 2 die Verteilungen des MZ-Panels 2001 – 2004).

Tabelle 7: Art der Ausfälle nach Teilnahmemuster und Klassifikation der Modellergebnisse für das MZ-Panel 2012 – 2015, Rotationsgruppe 82 (Proxy) (in %)

Art der Ausfälle	Teilnahmemuster (20..; X=Ausfall)			Insgesamt
	12XXXXXX	1213XXXX	121314XX	
Ausfälle insgesamt	10,7	9,2	10,1	27,2
Klassifikation geschätzter Ausfalltypen (1 bis 3)				
[1] Sterbefall	4,3	4,9	4,7	4,6
[2] Fortzug	11,2	11,9	10,5	11,2
[3] Fortzug oder Sterbefall	9,8	9,7	7,5	9,1
[4] Vollständiger Ausfall des Haushalts	74,7	73,5	77,3	75,1
Anteil Sterbefall an Ausfällen der Typen 1 - 3	17,1	18,5	20,8	18,6

Quelle: Siehe Tab. 1.

Die Analysen haben gezeigt, dass es zwischen 2001 – 2004 und 2012 – 2015 entgegen der anfangs getroffenen Annahme einige Änderungen der Zusammenhänge zwischen dem Ausfall und den Ausfallgründen gegeben hat. Es muss deshalb mit dem Modell M1 ein weniger gutes Modell verwendet werden. Wie die Validierung mit den zufällig aufgeteilten Teilstichproben des MZ-Panels 2001 – 2004 belegt, führt das gewählte Verfahren zu Fehlklassifikationen (siehe Abschnitt 4.4 und Tab. 4). Dennoch ist auf der Basis des Modells plausibel anzunehmen, dass die geschätzten und klassifizierten Sterbefälle im MZ-Panel 2012-2015 insgesamt nahe bei den unbekanntem Sterbefällen liegen. Unter den klassifizierten Sterbefällen dürfte der Anteil typisierter Sterbefälle über 60 % betragen. Im Einzelfall kann das Ergebnis anders ausfallen.

Werden für das MZ-Panel 2001 – 2004 zum Beispiel die geschätzten und klassifizierten Sterbefälle mit den in den Daten vorliegenden typisierten Sterbefällen nach dem Alter verglichen, liegt das Mindestalter bei den klassifizierten vs. „beobachteten“ Sterbefällen bei 65 vs. 5 Jahren und der Median bei 79 vs. 75 Jahren. Offensichtlich werden mit dem Verfahren Sterbefälle bei den Älteren überschätzt; insbesondere bei den über 80-Jährigen. Dies hängt mit der Abgrenzung von starken Effekten und dem Einsatz von z-Werten bei der Klassifikation der Modellergebnisse zusammen (siehe Abschnitt 4.3). Das schränkt die Verwendung der Schätzungen bei den nachfolgenden Modellen zum Ausfall ein.

Des Weiteren ist zu beachten, dass aufgrund der auf Haushaltsebene vorliegenden Angaben zum Ausfallgrund teilweise nicht zwischen Sterbefall und Fortzug unterschieden werden kann. Auch sind bei einem vollständigen Ausfall des Haushalts, der durchaus mit einem Sterbefall verbunden sein kann, keine Informationen zum Ausfallgrund vorhanden (siehe Abschnitt 4). Zudem können mit dem MZ Fortzüge ins Ausland nicht ermittelt werden (siehe Abschn. 2.2). Trotz dieser Einschränkungen ist es sinnvoll, die geschätzten Angaben bei den Ausfallanalysen zu berücksichtigen.

Nimmt man die Schätzungen der Sterbefälle als Ersatz für die in den Daten fehlenden Werte, entspricht dies einer einfachen Imputation mit den bekannten Nachteilen. Sie sind zwar in weiteren Analysen wesentlich einfacher zu verarbeiten als mehrfach imputierte Daten.⁸ Die Unsicherheiten oder Variabilitäten der Schätzungen bleiben aber bei einfacher Imputation unberücksichtigt, sodass die Standardfehler zu klein und keine validen Aussagen über statistische Inferenz möglich sind.

⁸ Nach einer Faustregel wären für Schätzungen mit multiplen Imputationen für jeden Übergang 60 und mehr Datensätze zu erstellen gewesen (siehe StataCorp 2017).

5 Konstruktion der Designgewichte

Aufgrund der genannten Einschränkungen beschränken sich die folgenden Analysen auf die Beschreibung, wie die im 2. Abschnitt skizzierten Ansätze im Prinzip umzusetzen und Designgewichte zur Korrektur des Ausfalls zu bilden sind. Wie zuvor konzentrieren sich die Analysen auf die Auswahl des neuen MZ-Panels mit der von 2012 bis 2015 befragten Rotationsgruppe.

In Anlehnung an die Arbeiten zum MZ-Panel 1996 – 1999 werden folgende erklärende Variablen verwendet.

- Bundesland als Designinformation (recodiert in: [1] West; [2] Ost incl. West-Berlin)
- Altersgruppe (recodiertes längsschnittkonsistentes Alter: [0] 0-4 Jahre; ...; [85] 85-89; [90] 90 Jahre und älter)
- Familienstand: [1] Ledig, verheiratet oder eingetragene Lebenspartnerschaft (eLP) getrennt lebend; [2] Verheiratet oder eLP zusammenlebend; [3] Geschieden, eLP aufgehoben; [4] Verwitwet, eLP verstorben)
- Wohnrechtstatus ([0] Hauptwohnsitz; [1] Nebenwohnsitz)
- Haushaltstyp ([0] Privathaushalt; [1] Gemeinschaftsunterkunft)
- Staatsangehörigkeit ([1] Deutsch; [2] Ausländer)
- Geschlecht (längsschnittkonsistent recodiert: [1] männlich; [2] weiblich)
- CASMIN-Bildungsklassifikation (GESIS 2016; Version „casa“ recodiert und ergänzt um Kategorien zum Schul- und Hochschulbesuch: [0] Kein Abschluss, keine Angabe; [1] Haupt-/Realschule ohne Berufsausbildung; [2] Haupt-/Realschule mit Berufsausbildung; [3] Abitur, Fachhochschule, Hochschule; [4] Schulbesuch (Klassen 1-9/10); [5] Schul-/Hochschulbesuch (Sonst))
- Persönliches Nettoeinkommen (näherungsweise in Quartilen zusammengefasst: [1] 1-699 Euro; [2] 700-1.299; [3] 1.300-1.999; [4] 2.000 Euro und mehr; [5] Ohne Angabe)

Die Wahrscheinlichkeiten räumlicher Immobilität werden für jeden Übergang geschätzt. Dabei wird die Korrelation in den Klumpen synthetischer Haushalte (siehe Abschnitt 2.3) berücksichtigt. In den Längsschnittpopulationen sind die imputierten Sterbefälle jeweils ausgeschlossen. Die korrigierten Designgewichte werden wie in Abschnitt 2.2 dargestellt ermittelt. Anschließend wird eine Normierung der Gewichte für die Bevölkerung am Hauptwohnsitz 2012 durchgeführt. Die Ergebnisse sind in den Tab. A4 – A6 dokumentiert. Um darzustellen, welche Konsequenzen auftreten, falls Sterbefälle in den Längsschnittpopulationen mit aufgenommen werden, werden die Analysen mit diesem nicht sachgemäßen Vorgehen wiederholt. Tab. 8 und Abb. 2 enthalten die Mittelwerte der normierten und korrigierten Längsschnittgewichte mit und ohne Ausschluss imputierter Sterbefälle.

Den größten Einfluss bei der Erklärung räumlicher Immobilität hat das Alter, gefolgt vom Familienstand. Dies hängt insbesondere damit zusammen, dass jüngere Personen im Alter von etwa 20 bis 30 Jahren aus ausbildungs- oder berufsbedingten Gründen das Elternhaus verlassen oder umziehen. In dieser Lebensphase sind auch Familiengründungen und Partnerwahlen ein Grund für Wohnungswechsel. Dagegen steigt für ältere Personen die Wahrscheinlichkeit räumlich immobil zu bleiben. Die Umzugsneigung nimmt bei Personen im höheren Alter wieder zu; z. B. nach dem Tod von Lebenspartnern oder wegen Wechsels in Seniorenheime o. Ä. Eventuell ist dieser Anstieg auch auf nicht erkannte Sterbefälle der Kategorien 3 „Fortzug oder Sterbefall“ oder 4 „Vollständiger Ausfall des Haushalts“ zurückzuführen. Für die anderen Variablen sind schwächere Effekte festzustellen, wobei die Differenzierung des Bundeslandes in West und Ost sowie das Geschlecht praktisch einen vernachlässigbaren Zusammenhang aufweisen.

In Tab. 8 und Abb. 2 sind bei den unter 65-Jährigen keine wesentlichen Unterschiede der Verteilungen festzustellen. Beschränken sich die Analysen auf diese Altersgruppen, treten daher vermutlich nur geringe Verzerrungen auf. Für ältere Personen steigen die Differenzen stetig an, da ohne Ausschluss der imputierten Sterbefälle Ausfälle Verstorbener durch die Längsschnittgewichte kompensiert werden. Zudem kumulieren sich die Differenzen der ohne Ausschluss der Sterbefälle designgewichteten Längsschnittpopulationen. Sie steigen für den Längsschnitt 2012 – 2015 auf 860.000. Die relativen Differenzen sind gering (2012 – 2015: 1,3 %) und hängen mit dem Anteil von knapp 5 % imputierter Sterbefälle unter den Ausfällen zusammen (siehe Tab. 7).

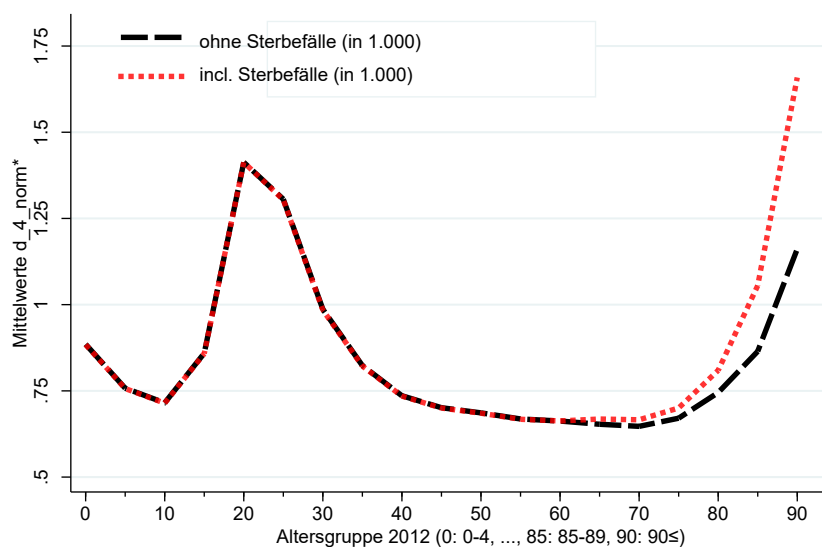


Abbildung 2: Mittelwerte korrigierter und normierter Längsschnittgewichte im MZ-Panel 2012 – 2015 (siehe Tab. 8)

Tabelle 8: Mittelwerte korrigierter Längsschnittgewichte, normiert für die Bevölkerung am Hauptwohnsitz 2012 (in 1.000) mit und ohne Ausschluss imputierter Sterbefälle nach Altersgruppen (2012) im MZ-Panel 2012 – 2015

Altersgruppe (2012)	Längsschnitte mit 2, 3 und 4 Wellen					
	2012 - 2013		2012 - 2014		2012 - 2015	
	Mit Ausschluss	Ohne A.	Mit Ausschluss	Ohne A.	Mit Ausschluss	Ohne A.
[0] 0-4 Jahre	0,6729	0,6728	0,7668	0,7663	0,8839	0,8828
[5] 5-9	0,6311	0,6310	0,6878	0,6875	0,7575	0,7564
[10] 10-14	0,6180	0,6180	0,6627	0,6623	0,7149	0,7139
[15] 15-19	0,6419	0,6418	0,7277	0,7272	0,8572	0,8560
[20] 20-24	0,7645	0,7649	1,0050	1,0058	1,3633	1,3654
[25] 25-29	0,7588	0,7592	0,9924	0,9924	1,2893	1,2877
[30] 30-34	0,6967	0,6966	0,8224	0,8214	0,9770	0,9746
[35] 35-39	0,6519	0,6519	0,7264	0,7260	0,8188	0,8178
[40] 40-44	0,6245	0,6245	0,6751	0,6746	0,7341	0,7330
[45] 45-49	0,6136	0,6136	0,6533	0,6529	0,6998	0,6988
[50] 50-54	0,6083	0,6082	0,6399	0,6396	0,6849	0,6839
[55] 55-59	0,6013	0,6012	0,6279	0,6275	0,6675	0,6666
[60] 60-64	0,5990	0,5990	0,6258	0,6260	0,6619	0,6622
[65] 65-69	0,5958	0,6002	0,6200	0,6292	0,6524	0,6673
[70] 70-74	0,5917	0,5979	0,6159	0,6275	0,6467	0,6655
[75] 75-79	0,5980	0,6066	0,6269	0,6456	0,6704	0,6995
[80] 80-84	0,6223	0,6366	0,6708	0,7054	0,7445	0,8085
[85] 85-89	0,6419	0,6911	0,7299	0,8353	0,8638	1,0521
[90] 90 ≤	0,6996	0,7793	0,8575	1,0761	1,1553	1,6510
Summe*	67.742	68.061	67.466	68.061	67.201	68.061

Quelle: Siehe Tab. 1. Auswahl: Bevölkerung am Hauptwohnsitz 2012. * Summe normierter Gewichte (in 1.000)

6 Schluss

Aufgrund des Stichprobendesigns des MZ als Flächenstichprobe besteht das methodische Problem der Panelselektivität des MZ-Panels in der Frage, ob es durch die aus einem Auswahlbezirk wegziehenden Haushalte und Personen, die nach dem Prinzip der Flächenstichprobe nicht weiter befragt werden, zu verzerrten Ergebnissen kommt. Sollen die Ausfälle mittels Designgewichten korrigiert werden, fehlen gegenwärtig wichtige Informationen im SUF. Dies sind die Kennung der Rotationsgruppe und die Angaben, ob die Ausfälle mit räumlicher Mobilität (Fort- und Zuzug) einerseits oder demographischen Gründen (Geburt und Sterbefall) andererseits zusammenhängen.

In diesem Bericht wurde versucht, die Zugehörigkeit zu den Rotationsgruppen mithilfe längsschnittkonsistenter Auswahlbezirksnummern näherungsweise zu ermitteln. Die Angaben der nicht zur Längsschnittpopulation zählenden Sterbefälle wurden modellgestützt auf Basis von Verteilungen des MZ-Panels 2001 – 2004 geschätzt und einfach imputiert. Das Verfahren ist zwar mit Fehlklassifikationen und Unsicherheiten verbunden. Mit diesen Einschränkungen konnte dennoch gezeigt werden, wie Designgewichte zur Korrektur der Ausfälle im Prinzip erstellt werden können. Die Befunde entsprechen den theoretisch zu erwartenden Ergebnissen. Verzichtet man mangels entsprechender Informationen auf den Ausschluss der Sterbefälle aus der Längsschnittpopulation, führt das zu höheren Gewichten als bei korrekter Vorgehensweise und damit zur Kompensation verstorbener Personen. Insgesamt betrachtet erscheint es also besser, vorläufig die hier vorgeschlagenen Längsschnittgewichte unter Berücksichtigung ihrer Einschränkungen zu verwenden.⁹

Angesichts der durch die gegenwärtig nutzbaren Daten gegebenen Schwierigkeiten wird eine bessere Datenbereitstellung vorgeschlagen.

Bei der Analyse von Rotationspanels wird die Rotationskennung für eine sichere Abgrenzung der Rotationsgruppen benötigt. Es wird deshalb vorgeschlagen, die zurzeit im SUF nicht enthaltenen Merkmale EF572 Stichprobennummer und EF573 Rotationsviertel bereitzustellen.

Für Ausfallanalysen und -korrekturen sind entsprechende Typisierungen zu den Ausfallgründen nötig, die in den älteren MZ-Panels 1996-1999 und 2001-2004 sowie im Campus-File MZ-Panel 1996 – 1999 enthalten sind. Es handelt sich dabei um die im MZ regulär erfragten (EF21 und EF22) bzw. organisatorisch erstellten Angaben nach § 4, Abs. 1, Nr. 1 MZG 2005.

Falls der Aufwand für die Erstellung von Typisierungen zu hoch erscheint, kommt alternativ auch die Bereitstellung der gegenwärtig gesperrten Ausgangsinformationen (siehe z. B. FDZ-L 2019) durch die Statistischen Ämter im SUF infrage:

- EF21 Zahl der fortgezogenen Haushaltsmitglieder
- EF22 Zahl der gestorbenen Haushaltsmitglieder
- EF23 Veränderung des Haushalts
- EF615 Veränderung der Person

Die Notwendigkeit dieser Merkmale wurde im Gesetzentwurf der Bundesregierung wie folgt begründet (BT-Drucksache 15/2543 2004: 11): „Angaben über die Veränderung der Haushaltsgröße und -zusammensetzung seit der letzten Befragung sind erforderlich, um Veränderungen zwischen den einzelnen Erhebungszeitpunkten festzustellen; sie sind ein wichtiger Baustein für Verlaufsunter-

⁹ Die unter der Adresse <https://www.gesis.org/missy/materials/MZ/tools/panel2012> bereitgestellten Stata-Programme können den für die Konstruktion von Längsschnittgewichten nötigen Aufwand verringern.

suchungen.” Aus Sicht der Wissenschaft kann dieser Bedarf für sachgerechte Analysen des MZ-Panels nur unterstrichen werden.

Literatur

- ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute; AG.MA Arbeitsgemeinschaft Media-Analysen (1999): Stichproben-Verfahren in der Umfrageforschung. Opladen: Leske + Budrich.
- Afentakis, Anja; Bihler, Wolf (2005): Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005. *Wirtschaft und Statistik* (10): 1039-1048.
https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEAusgabe_derivate_00000035/1010200051104.pdf [20.09.2021].
- Basic, Edin; Marek, Ivo; Rendtel, Ulrich (2006): The German Microcensus as a tool for longitudinal data analysis: An evaluation using SOEP data. Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe", Arbeitspapier Nr.: 3. Berlin: Freie Universität.
https://web.archive.org/web/20161117000116/https://www.destatis.de/DE/Methoden/Methodenpapiere/Mikrozensus/Arbeitspapiere/Arbeitspapier3.pdf?__blob=publicationFile [20.09.2021].
- BT-Drucksache 15/2543 (2004): Entwurf eines Gesetzes zur Durchführung einer Repräsentativstatistik über die Bevölkerung und den Arbeitsmarkt sowie die Wohnsituation der Haushalte (Mikrozensusgesetz 2005) [Gesetzentwurf der Bundesregierung mit Begründung]
https://dserver.bundestag.de/btd/15/025/150254_3.pdf [20.09.2021].
- Chen, Qixuan; Gelman, Andrew; Tracy, Melissa; Norris, Fran H.; Galea, Sandro (2012): Weighting Adjustments for Panel Nonresponse. New York: Columbia University.
<http://www.stat.columbia.edu/~gelman/research/unpublished/weighting%20adjustments%20for%20panel%20surveys.pdf> [20.09.2021].
- Clarke, Paul S.; Chambers, Ray L. (1998): Estimating Labour Force Gross Flows From Surveys Subject to Household-level Nonignorable Nonresponse. *Survey Methodology* 24(2): 123-129.
<https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1998002/article/4349-eng.pdf?st=CUOfMEC> [20.09.2021].
- Forschungsdatenzentrum der Statistischen Ämter der Länder [FDZ-L] (2019): Schlüsselverzeichnis Mikrozensus 2015. https://www.forschungsdatenzentrum.de/sites/default/files/mz_2015_on-site_svz_v1.pdf [18.08.2021].
- GESIS (2021): Scientific Use Files des Mikrozensus ab 2012 als Rotationspanel. Version 1 (Längsschnittorientierung): Zusammenführung 2012, 2013, 2014 und 2015.
<https://www.gesis.org/missy/materials/MZ/tools/panel2012> [20.09.2021].
- GESIS (2016): Die Bildungsklassifikation CASMIN.
<https://www.gesis.org/missy/materials/MZ/tools/casmin> [20.09.2021].
- Herter-Eschweiler, Robert (2006): Der Mikrozensus als Datenquelle für Panelauswertungen, S. 33-59 in: Methodenverbund Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe (2006): Handbuch Mikrozensus-Panel 1996-1999. Version 0.2 - Juli 2006. Hrsg.: Statistisches Bundesamt, Gruppe VIII C (Mikrozensus): Wiesbaden, Zweigstelle Bonn.
https://www.gesis.org/missy/files/documents/MZ/MZP9699_Handbuch_1_.pdf [20.09.2021].
- Herter-Eschweiler, Robert; Schimpl-Neimanns, Bernhard (2018): Möglichkeiten der Verknüpfung von Mikrozensus-Querschnitterhebungen ab 2012 zu Panels. Bonn/Mannheim:
https://www.gesis.org/missy/files/documents/MZ/panelbildung_suf2012.pdf [20.09.2021].
- Kalton, Graham(1987): Handling Wave Nonresponse in Panel Surveys. Chapter 2, S. 24-42 in: The Treatment of Person-Wave Nonresponse in Longitudinal Surveys, SIPP Working Paper #26. U.S.

- Department of Commerce U.S. CENSUS BUREAU. https://www2.census.gov/library/working-papers/1987/demo/SEHSD-WP_1987-04.pdf [20.09.2021].
- Lepkowski, James M. (1987): The Treatment of Wave Nonresponse in Panel Surveys. S. 90-127 in: The Treatment of Person-Wave Nonresponse in Longitudinal Surveys, SIPP Working Paper #26. U.S. Department of Commerce U.S. CENSUS BUREAU. <https://www2.census.gov/library/working-papers/1987/demo/SEHSD-WP1987-04.pdf> [20.09.2021].
- Lundström, Sixten; Särndal, Carl-Erik (2001): Estimation in the presence of Nonresponse and Frame Imperfections. Örebro: Statistics Sweden.
- Marek, Ivo (2005): Weighting adjustments in the presence of non-coverage due to residential mobility in the German Microcensus-Panel. Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe", Arbeitspapier Nr.: 10. Berlin: Freie Universität. https://web.archive.org/web/20161117000116/https://www.destatis.de/DE/Methoden/Methodenpapiere/Mikrozensus/Arbeitspapiere/Arbeitspapier10.pdf?__blob=publicationFile [20.09.2021].
- Marek, Ivo; Rohloff, Sandra (2006): Das Hochrechnungsverfahren für Längsschnittauswertungen im Mikrozensus-Panel 1996-1999, S. 91-104 in: Methodenverbund Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe (2006): Handbuch Mikrozensus-Panel 1996-1999. Version 0.2 - Juli 2006. Hrsg.: Statistisches Bundesamt, Gruppe VIII C (Mikrozensus): Wiesbaden, Zweigstelle Bonn. https://www.gesis.org/missy/files/documents/MZ/MZP9699_Handbuch_1_.pdf [20.09.2021].
- Methodenverbund Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe (2006): Handbuch Mikrozensus-Panel 1996-1999. Version 0.2 – Juli 2006. Hrsg.: Statistisches Bundesamt, Gruppe VIII C (Mikrozensus): Wiesbaden, Zweigstelle Bonn. https://www.gesis.org/missy/files/documents/MZ/MZP9699_Handbuch_1_.pdf [20.09.2021].
- MZG 2005: Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und den Arbeitsmarkt sowie die Wohnsituation der Haushalte (Mikrozensusgesetz 2005) vom 24.6.2004 (BGBl. I S. 1350). http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl104s1350.pdf [20.09.2021].
- Porst, Rolf (2000): Praxis der Umfrageforschung. Wiesbaden: Teubner.
- Rendtel, Ulrich; Harms, Torsten (2009): Weighting and Calibration for Household Panels, S. 265-286 in: Lynn, Peter (Hrsg.): Methodology of Longitudinal Surveys. Chichester, West Sussex: Wiley.
- Rohloff, Sandra (2006): Das Hochrechnungsverfahren für Längsschnittauswertungen aus dem Mikrozensus. Methodenverbund "Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe", Arbeitspapier Nr.: 6. https://web.archive.org/web/20161117000116/https://www.destatis.de/DE/Methoden/Methodenpapiere/Mikrozensus/Arbeitspapiere/Arbeitspapier6.pdf?__blob=publicationFile [20.09.2021].
- Särndal, Carl-Erik; Swensson, Bengt; Wretman, Jan (1997) Model Assisted Survey Sampling. New York: Springer.
- Schimpl-Neimanns, Bernhard (2008): Bildungsverläufe und Stichprobenselektivität. Analysen zur Stichprobenselektivität des Mikrozensuspanels 1996-1999 am Beispiel bildungsstatistischer Fragestellungen. GESIS-Forschungsberichte, Reihe Sozialwissenschaftliche Datenanalyse, Band 1. Bonn: GESIS. <https://www.ssoar.info/ssoar/handle/document/26223> [20.09.2021].

- Spiess, Martin (2005): Analyse von Längsschnittdaten mit fehlenden Werten: Grundlagen, Verfahren und Anwendungen. Universität Bremen.
<http://nbn-resolving.de/urn:nbn:de:gbv:46-diss000012631> [20.09.2021].
- StataCorp (2017): Stata: Release 15. Statistical Software. College Station, TX: StataCorp LLC.
- Statistisches Bundesamt (2021): Sterbetafel (Periodensterbetafel): Deutschland, Jahre, Geschlecht, Vollendetes Alter (12621-0001). <https://www-genesis.destatis.de/genesis/online> [20.09.2021]
- Statistisches Bundesamt (2016): Qualitätsbericht Mikrozensus 2015. Wiesbaden.
https://www.destatis.de/DE/Methoden/Qualitaet/Qualitaetsberichte/Bevoelkerung/mikrozensus-2015.pdf?__blob=publicationFile [20.09.2021].
- Statistisches Bundesamt (2009): Handbuch zum Mikrozensus-Panel 2001–2004. Hrsg.: Statistisches Bundesamt, Gruppe VIII C (Mikrozensus): Wiesbaden, Zweigstelle Bonn.
https://www.gesis.org/missy/files/documents/MZ/MZP0104_Handbuch.pdf [20.09.2021s].
- Statistisches Bundesamt; GESIS (2016): Wichtige Informationen zur Nutzung des Mikrozensus Scientific Use Files 2012. Bonn und Mannheim.
https://www.gesis.org/missy/files/documents/MZ/readme/readme_suf2012.pdf [20.09.2021].
- Statistisches Bundesamt; GESIS (2017): Wichtige Informationen zur Nutzung des Mikrozensus Scientific Use Files 2013. Bonn und Mannheim.
https://www.gesis.org/missy/files/documents/MZ/readme/readme_suf2013.pdf [20.09.2021].
- Statistisches Bundesamt; GESIS (2018): Wichtige Informationen zur Nutzung des Mikrozensus Scientific Use Files 2014. Bonn und Mannheim.
https://www.gesis.org/missy/files/documents/MZ/readme/readme_suf2014.pdf [20.09.2021].
- Statistisches Bundesamt; GESIS (2019): Wichtige Informationen zur Nutzung des Mikrozensus Scientific Use Files 2015. Bonn und Mannheim.
https://www.gesis.org/missy/files/documents/MZ/readme/readme_suf2015.pdf [20.09.2021].

Anhang

Tabelle A1: Statische Kennwerte zu den zufälligen Teilstichproben des MZ-Panels 2001 - 2004

	Teilstichprobe 1	Teilstichprobe 2
Stichprobenumfang (n)	6.293	6.399
Art des Ausfalls (%)		
[1] Sterbefall	15,1	15,0
[2] Fortzug	35,2	34,6
[3] Fortzug oder Sterbefall	49,7	50,4
Alter in Jahren		
1. Quartil	21	21
Median	32	32
3. Quartil	61	61
Mittelwert	40,2	40,0
Standardabweichung	25,0	24,7
Variationskoeffizient	0,6221	0,6173
Geschlecht (%)		
[1] Männlich	51,1	49,1
[2] Weiblich	48,9	50,9
Differenz der Haushaltsgrößen zum Zeitpunkt $t-1 - t$ (%)		
[0] Zuzüge oder keine Veränderung (Diff. ≤ 0)	37,0	36,8
[1] 1 Person ausgefallen	47,4	47,6
[2] 2 Personen ausgefallen	10,3	9,7
[3] 3 und mehr Personen ausgefallen	5,3	5,9

Quelle: Siehe Tab. 2. Teilstichprobe 1 enthält Sätze mit gültigen Angaben zum Ausfallgrund, in Teilstichprobe 2 sind die Angaben zum Ausfallgrund als fehlend recodiert.

Tabelle A2: Ergebnisse des multinomialen Logitmodells - MZ-Panel 2001 – 2004 für beide Teilstichproben (Teilstichprobe 1: gültige Angaben zum Ausfallgrund (sel=0); Teilstichprobe 2: fehlende Angaben zum Ausfallgrund (sel=1))

Modell: *mlogit yverlust o0.altr ib90.altr (c.alt c.altq)##ib1.sex ib1.hhgr_diff if (sel==0 & yverlust>=1 & yverlust<=3) | (sel==1 & yverlust==), baseoutcome(2) vce(cluster idpsu)*

Art der Ausfälle [yverlust]	Coef.	Std. Err.	z	P> z
_1__Sterbefall				
0.altr [Altersgruppe: unter 1 bis 4 Jahre]	0,000000	.	.	.
5.altr	8,450583	4,04	2,090	0,037
10.altr	8,193736	3,52	2,327	0,020
15.altr	6,438992	3,10	2,079	0,038
20.altr	4,040554	2,87	1,407	0,160
25.altr	4,298957	2,62	1,642	0,101
30.altr	5,566979	2,32	2,403	0,016
35.altr	5,226635	2,11	2,479	0,013
40.altr	4,674168	1,93	2,416	0,016
45.altr	4,569687	1,79	2,555	0,011
50.altr	4,922584	1,67	2,951	0,003
55.altr	4,249944	1,54	2,757	0,006
60.altr	4,390766	1,41	3,120	0,002
65.altr	4,763647	1,26	3,780	0,000
70.altr	3,786460	1,12	3,380	0,001
75.altr	2,990032	0,93	3,210	0,001
80.altr	1,821603	0,76	2,411	0,016
85.altr	1,659282	0,61	2,741	0,006
alt [Alter in Jahren]	0,113533	0,13	0,883	0,377
altq [quadrirtes Alter in Jahren]	0,000580	0,00	0,604	0,546
2.sex [Geschlecht = weiblich]	1,180523	0,87	1,360	0,174
2.sex#c.alt	-0,043900	0,03	-1,354	0,176
2.sex#c.altq	0,000193	0,00	0,651	0,515
[hhgr_diff: Diff. n Haushaltsmitgl. t / t-1]				
0.hhgr_diff [Zuzüge/keine Veränderung]	-0,246111	0,23	-1,088	0,277
2.hhgr_diff [2 Person ausgefallen]	-1,432428	0,27	-5,296	0,000
3.hhgr_diff [≥3 Personen ausgefallen]	-0,597350	0,36	-1,678	0,093
_cons	-12,305820	4,75	-2,591	0,010
_2__Fortzug	(base outcome)			

Fortsetzung Tabelle A2

Art der Ausfälle	Coef.	Std. Err.	z	P> z
_3__Fortzug_Sterbefall				
0.altr	0,000000			
5.altr	1,032004	0,34	3,048	0,002
10.altr	1,338685	0,49	2,725	0,006
15.altr	0,928996	0,67	1,394	0,163
20.altr	0,878658	0,77	1,144	0,253
25.altr	1,317293	0,90	1,470	0,141
30.altr	2,109369	1,00	2,100	0,036
35.altr	2,255010	1,08	2,092	0,036
40.altr	2,812702	1,13	2,496	0,013
45.altr	2,577105	1,15	2,232	0,026
50.altr	3,130750	1,16	2,710	0,007
55.altr	3,012668	1,12	2,682	0,007
60.altr	2,893153	1,07	2,700	0,007
65.altr	3,340335	0,99	3,358	0,001
70.altr	2,654388	0,90	2,937	0,003
75.altr	1,843871	0,78	2,379	0,017
80.altr	1,268828	0,65	1,953	0,051
85.altr	1,379216	0,58	2,374	0,018
alt	-0,175563	0,05	-3,309	0,001
altq	0,001928	0,00	3,444	0,001
2.sex	-0,578811	0,26	-2,202	0,028
2.sex#c.alt	0,040345	0,01	2,760	0,006
2.sex#c.altq	-0,000500	0,00	-2,924	0,003
0.hhgr_diff	3,718839	0,15	25,452	0,000
2.hhgr_diff	0,117336	0,12	0,961	0,337
3.hhgr_diff	-0,046529	0,22	-0,208	0,835
_cons	0,601184	0,29	2,073	0,038
Modellstatistiken (für Teilstichprobe 1)				
N	6293			
Pseudo R ² (McFadden)	0,4741			
Log Pseudolikelihood (Null Modell)	-6297,37			
Log Pseudolikelihood (Modell)	-3311,80			
Pseudolikelihood Ratio Chi ²	5971,13			
d.f.	50			
Prob > Chi ²	0,0000			
BIC	7078,45			

Quelle: Siehe Tab. 2.

Nicht in der Tabelle aufgeführt sind die auf Null gesetzten Koeffizienten der Referenzkategorien: Altersgruppe: [90] 90 Jahre und älter; Geschlecht: [1] männlich; Differenz der Zahl der Haushaltsmitglieder zwischen den Zeitpunkten t-1 und t: [1] 1 Person ausgefallen.

Tabelle A3: Ergebnisse des multinomialen Logitmodells M1 mit dem MZ-Panel 2001 – 2004 als Analysestichprobe (zu Tab. 6)

Modell M1:

mlogit verlust o0.altr ib90.altr (c.alt c.altq)##ib1.sex if verlust>=1 & verlust<=3, baseoutcome(2) vce(cluster idpsu)

Art der Ausfälle [verlust]	Coef.	Std. Err.	z	P> z
_1__Sterbefall				
0.altr [Altersgruppe: unter 1 bis 4 Jahre]	0,000000			
5.altr	8,487803	2,627330	3,23	0,001
10.altr	7,081320	2,386576	2,97	0,003
15.altr	5,267072	2,034277	2,59	0,010
20.altr	3,564167	1,840511	1,94	0,053
25.altr	3,103964	1,683713	1,84	0,065
30.altr	3,680999	1,468808	2,51	0,012
35.altr	3,249104	1,325000	2,45	0,014
40.altr	2,692183	1,199359	2,24	0,025
45.altr	2,553966	1,104318	2,31	0,021
50.altr	2,532614	1,016921	2,49	0,013
55.altr	2,383437	0,934732	2,55	0,011
60.altr	2,511765	0,855667	2,94	0,003
65.altr	2,868363	0,768876	3,73	0,000
70.altr	2,249678	0,694543	3,24	0,001
75.altr	1,936854	0,580641	3,34	0,001
80.altr	1,052733	0,471002	2,24	0,025
85.altr	1,297471	0,417993	3,10	0,002
alt [Alter in Jahren]	0,246917	0,087943	2,81	0,005
altq [quadrirtes Alter in Jahren]	-0,000737	0,000637	-1,16	0,247
2.sex [Geschlecht = weiblich]	1,492463	0,688519	2,17	0,030
2.sex#c.alt	-0,051236	0,023969	-2,14	0,033
2.sex#c.altq	0,000248	0,000207	1,20	0,231
_cons	-13,777800	3,254129	-4,23	0,000
_2__Fortzug	(base outcome)			

Fortsetzung Tabelle A3

Art der Ausfälle	Coef.	Std. Err.	z	P> z
<u>_3__Fortzug_Sterbefall</u>				
0.altr	0,000000			
5.altr	0,919412	0,213487	4,31	0,000
10.altr	1,101116	0,289509	3,80	0,000
15.altr	0,527018	0,376712	1,40	0,162
20.altr	0,200094	0,441486	0,45	0,650
25.altr	0,688959	0,511131	1,35	0,178
30.altr	1,717940	0,570012	3,01	0,003
35.altr	2,132328	0,613258	3,48	0,001
40.altr	2,584411	0,638197	4,05	0,000
45.altr	2,633153	0,651924	4,04	0,000
50.altr	3,095866	0,649248	4,77	0,000
55.altr	2,899934	0,635545	4,56	0,000
60.altr	3,089016	0,603393	5,12	0,000
65.altr	3,307888	0,563626	5,87	0,000
70.altr	2,856920	0,521820	5,47	0,000
75.altr	2,393092	0,445844	5,37	0,000
80.altr	1,568996	0,373597	4,20	0,000
85.altr	1,590256	0,375455	4,24	0,000
Alt	-0,172727	0,030124	-5,73	0,000
Altq	0,001818	0,000320	5,69	0,000
2.sex	-0,529009	0,155954	-3,39	0,001
2.sex#c.alt	0,032724	0,008620	3,80	0,000
2.sex#c.altq	-0,000324	0,000102	-3,17	0,002
<u>_cons</u>	<u>2,020110</u>	<u>0,188144</u>	<u>10,74</u>	<u>0,000</u>
Modellstatistiken (für MZP 2001-2004)				
N	12692			
Pseudo R ² (McFadden)	0,2481			
Log Pseudolikelihood (Null Modell)	-12678,49			
Log Pseudolikelihood (Modell)	-9532,37			
Pseudolikelihood Ratio Chi ²	6292,24			
d.f.	44			
Prob > Chi ²	0,0000			
BIC	19499,38			

Quelle: Siehe Tab. 2.

Nicht in der Tabelle aufgeführt sind die auf Null gesetzten Koeffizienten der Referenzkategorien: Altersgruppe: [90] 90 Jahre und älter; Geschlecht: [1] männlich.

Tabelle A4: Ergebnisse des Logitmodells zum Ausfall beim Übergang 2012/2013 – MZ-Panel 2012 – 2015

Modell: *logit r2 ib1.west_ost ib50.altr_1 ib2.famst_1 ib1.hws_1 ib0.hhtyp_1 ib1.nat_1 ib1.sex_1 ib2.casr_1 ib2.eink_1 if rot==82 & att>=0 & att<=1 & r1==1 & r2>=0 & r2<=1 & verltypr_2!=1, vce(cluster idhhs_2)*

r2	Coef.	Std. Err.	z	P> z
2.west_ost [Bundesland]	0,005374	0,035896	0,15	0,881
0.altr_1 [Altersgruppe]	-0,585125	0,101229	-5,78	0,000
5.altr_1	-0,272304	0,098068	-2,78	0,005
10.altr_1	-0,203980	0,102711	-1,99	0,047
15.altr_1	-0,460109	0,082115	-5,60	0,000
20.altr_1	-1,313708	0,063981	-20,53	0,000
25.altr_1	-1,356537	0,059592	-22,76	0,000
30.altr_1	-1,055867	0,060447	-17,47	0,000
35.altr_1	-0,694359	0,062313	-11,14	0,000
40.altr_1	-0,328935	0,062863	-5,23	0,000
45.altr_1	-0,112249	0,060137	-1,87	0,062
55.altr_1	0,187097	0,067755	2,76	0,006
60.altr_1	0,256326	0,073755	3,48	0,001
65.altr_1	0,396482	0,084550	4,69	0,000
70.altr_1	0,583526	0,083616	6,98	0,000
75.altr_1	0,392111	0,084303	4,65	0,000
80.altr_1	-0,156382	0,087478	-1,79	0,074
85.altr_1	-0,376553	0,100524	-3,75	0,000
90.altr_1	-0,876627	0,119407	-7,34	0,000
1.famst_1 [Familienstand]	-0,668814	0,038326	-17,45	0,000
3.famst_1	-0,842502	0,046776	-18,01	0,000
4.famst_1	-0,690853	0,059045	-11,70	0,000
2.hws_1	-0,925681	0,068150	-13,58	0,000
1.hhtyp_1 [Haushaltstyp]	-0,073449	0,241625	-0,30	0,761
2.nat_1 [Staatsangehörigkeit]	-0,427534	0,047795	-8,95	0,000
2.sex_1 [Geschlecht]	0,016554	0,018818	0,88	0,379
0.casr_1 [Casmin-Bildungsklassifikation]	-0,240738	0,067227	-3,58	0,000
1.casr_1	-0,240071	0,038453	-6,24	0,000
3.casr_1	-0,157726	0,031396	-5,02	0,000
4.casr_1	0,168935	0,075798	2,23	0,026
5.casr_1	0,015230	0,045254	0,34	0,736
1.eink_1 [Persönliches Nettoeinkommen]	-0,052529	0,033725	-1,56	0,119
3.eink_1	0,145964	0,032212	4,53	0,000
4.eink_1	0,158542	0,038461	4,12	0,000
5.eink_1	0,183518	0,045840	4,00	0,000
_cons	3,075780	0,058343	52,72	0,000
N	120382			
Pseudo R2 (McFadden)	0,0872			
Pseudolikelihood Ratio Chi ²	6973,50			
d.f.	35			
Prob > Chi ²	0,0000			

Quelle: Siehe Tab. 1. Auswahl: Rotationsgruppe 82 (Proxy), Weiterverfolgungskonzept, ohne imputierte Sterbefälle.

Tabelle A5: Ergebnisse des Logitmodells zum Ausfall beim Übergang 2013/2014 – MZ-Panel 2012 – 2015

Modell: *logit r3 ib1.west_ost ib50.altr_1 ib2.famst_1 ib1.hws_1 ib0.hhtyp_1 ib1.nat_1 ib1.sex_1 ib2.casr_1 ib2.eink_1 ib50.altr_2 ib2.famst_2 ib1.hws_2 ib0.hhtyp_2 ib1.nat_2 ib1.sex_2 ib2.casr_2 ib2.eink_2 if rot==82 & att>=0 & att<=1 & verlytpr_2!=1 & verlytpr_3!=1 & r1==1 & r2==1 & r3>=0 & r3<=1, vce(cluster idhhs_3)*

r3	Coef.	Std. Err.	z	P> z
2.west_ost	0,079654	0,040956	1,94	0,052
0.altr_1	-0,255593	0,355488	-0,72	0,472
5.altr_1	-0,283064	0,334653	-0,85	0,398
10.altr_1	-0,182840	0,302247	-0,60	0,545
15.altr_1	-0,537497	0,262206	-2,05	0,040
20.altr_1	-0,796702	0,246034	-3,24	0,001
25.altr_1	-0,835795	0,232791	-3,59	0,000
30.altr_1	-0,537068	0,216862	-2,48	0,013
35.altr_1	-0,358251	0,190990	-1,88	0,061
40.altr_1	-0,314690	0,152914	-2,06	0,040
45.altr_1	-0,215080	0,110950	-1,94	0,053
55.altr_1	0,092000	0,126165	0,73	0,466
60.altr_1	-0,005392	0,187506	-0,03	0,977
65.altr_1	-0,045252	0,252450	-0,18	0,858
70.altr_1	0,148314	0,293083	0,51	0,613
75.altr_1	0,355432	0,321603	1,11	0,269
80.altr_1	0,093712	0,364354	0,26	0,797
85.altr_1	-0,477462	0,410497	-1,16	0,245
90.altr_1	-1,075769	0,469099	-2,29	0,022
1.famst_1	-0,426916	0,117549	-3,63	0,000
3.famst_1	-0,572642	0,139537	-4,10	0,000
4.famst_1	-0,129351	0,152790	-0,85	0,397
2.hws_1	-0,210036	0,109360	-1,92	0,055
1.hhtyp_1	-0,075155	0,989364	-0,08	0,939
2.nat_1	-0,360181	0,146354	-2,46	0,014
2.sex_1	-0,006349	0,021193	-0,30	0,765
0.casr_1	-0,019568	0,103236	-0,19	0,850
1.casr_1	-0,029710	0,058886	-0,50	0,614
3.casr_1	-0,011390	0,063827	-0,18	0,858
4.casr_1	0,225273	0,106061	2,12	0,034
5.casr_1	-0,229314	0,069265	-3,31	0,001
1.eink_1	-0,066015	0,044755	-1,48	0,140
3.eink_1	-0,005676	0,043829	-0,13	0,897
4.eink_1	0,049300	0,061813	0,80	0,425
5.eink_1	0,091380	0,061209	1,49	0,135

Fortsetzung Tabelle A5

r3	Coef.	Std. Err.	z	P> z
0.altr_2	-0,599977	0,352227	-1,70	0,088
5.altr_2	-0,331690	0,332728	-1,00	0,319
10.altr_2	-0,402595	0,299355	-1,34	0,179
15.altr_2	-0,197730	0,262245	-0,75	0,451
20.altr_2	-0,638683	0,243286	-2,63	0,009
25.altr_2	-0,713136	0,229707	-3,10	0,002
30.altr_2	-0,599346	0,214058	-2,80	0,005
35.altr_2	-0,417883	0,190271	-2,20	0,028
40.altr_2	-0,127245	0,151308	-0,84	0,400
45.altr_2	0,022907	0,108129	0,21	0,832
55.altr_2	0,042713	0,126151	0,34	0,735
60.altr_2	0,106263	0,187340	0,57	0,571
65.altr_2	0,334221	0,247578	1,35	0,177
70.altr_2	0,183905	0,289284	0,64	0,525
75.altr_2	-0,193060	0,319423	-0,60	0,546
80.altr_2	-0,349607	0,360481	-0,97	0,332
85.altr_2	-0,262354	0,405503	-0,65	0,518
90.altr_2	-0,110210	0,461968	-0,24	0,811
1.famst_2	-0,103951	0,116126	-0,90	0,371
3.famst_2	-0,218428	0,139059	-1,57	0,116
4.famst_2	-0,462875	0,151191	-3,06	0,002
2.hws_2	-0,774262	0,121291	-6,38	0,000
1.hhtyp_2	0,184884	1,029590	0,18	0,857
2.nat_2	-0,054592	0,150809	-0,36	0,717
2.sex_2	0,000000			
0.casr_2	-0,208618	0,101322	-2,06	0,039
1.casr_2	-0,161554	0,060185	-2,68	0,007
3.casr_2	-0,148040	0,063085	-2,35	0,019
4.casr_2	-0,023610	0,114388	-0,21	0,836
5.casr_2	0,094487	0,069937	1,35	0,177
1.eink_2	0,018430	0,045180	0,41	0,683
3.eink_2	0,056178	0,042630	1,32	0,188
4.eink_2	0,099330	0,059739	1,66	0,096
5.eink_2	0,248329	0,062134	4,00	0,000
_cons	3,232944	0,071394	45,28	0,000
N	107458			
Pseudo R2 (McFadden)	0,0792			
Pseudolikelihood Ratio Chi ²	5083,20			
d.f.	68			
Prob > Chi ²	0,0000			

Quelle: Siehe Tab. 1. Auswahl: Rotationsgruppe 82 (Proxy), Weiterverfolgungskonzept, ohne imputierte Sterbefälle.

Tabelle A6: Ergebnisse des Logitmodells zum Ausfall beim Übergang 2014/2015 – MZ-Panel 2012 – 2015

Modell: *logit r4 ib1.west_ost ib50.altr_1 ib2.famst_1 ib1.hws_1 ib0.hhtyp_1 ib1.nat_1 ib1.sex_1 ib2.casr_1 ib2.eink_1 ib50.altr_2 ib2.famst_2 ib1.hws_2 ib0.hhtyp_2 ib1.nat_2 ib1.sex_2 ib2.casr_2 ib2.eink_2 ib50.altr_3 ib2.famst_3 ib1.hws_3 ib0.hhtyp_3 ib1.nat_3 ib1.sex_3 ib2.casr_3 ib2.eink_3 if rot==82 & att>=0 & att<=1 & verlypr_2!=1 & verlypr_3!=1 & verlypr_4!=1 & r1==1 & r2==1 & r3==1 & r4>=0 & r4<=1, vce(cluster idhhs_4)*

r4	Coef.	Std. Err.	z	P> z
2.west_ost	0,073794	0,041758	1,77	0,077
0.altr_1	0,107080	0,369901	0,29	0,772
5.altr_1	0,067461	0,346590	0,19	0,846
10.altr_1	0,113407	0,321741	0,35	0,724
15.altr_1	-0,655848	0,277025	-2,37	0,018
20.altr_1	-0,897942	0,257193	-3,49	0,000
25.altr_1	-0,729736	0,239653	-3,04	0,002
30.altr_1	-0,543027	0,220160	-2,47	0,014
35.altr_1	-0,313117	0,193100	-1,62	0,105
40.altr_1	-0,294212	0,149898	-1,96	0,050
45.altr_1	-0,096591	0,106830	-0,90	0,366
55.altr_1	0,118372	0,116197	1,02	0,308
60.altr_1	0,288839	0,179325	1,61	0,107
65.altr_1	0,493021	0,230458	2,14	0,032
70.altr_1	0,717033	0,270836	2,65	0,008
75.altr_1	0,667510	0,309712	2,16	0,031
80.altr_1	0,444853	0,349545	1,27	0,203
85.altr_1	0,049631	0,394407	0,13	0,900
90.altr_1	-0,245811	0,443471	-0,55	0,579
1.famst_1	-0,285402	0,126169	-2,26	0,024
3.famst_1	-0,326647	0,141375	-2,31	0,021
4.famst_1	-0,116067	0,147774	-0,79	0,432
2.hws_1	-0,213671	0,127606	-1,67	0,094
1.hhtyp_1	-1,811498	0,737495	-2,46	0,014
2.nat_1	0,046525	0,168703	0,28	0,783
2.sex_1	0,003498	0,021567	0,16	0,871
0.casr_1	-0,028314	0,103864	-0,27	0,785
1.casr_1	0,030957	0,064676	0,48	0,632
3.casr_1	-0,001449	0,067528	-0,02	0,983
4.casr_1	-0,144564	0,114557	-1,26	0,207
5.casr_1	-0,245932	0,076184	-3,23	0,001
1.eink_1	-0,004306	0,049333	-0,09	0,930
3.eink_1	-0,000722	0,047093	-0,02	0,988
4.eink_1	-0,080486	0,067050	-1,20	0,230
5.eink_1	0,089667	0,068693	1,31	0,192

Fortsetzung Tabelle A6

r4	Coef.	Std. Err.	z	P> z
0.altr_2	-0,388714	0,447292	-0,87	0,385
5.altr_2	-0,038677	0,420242	-0,09	0,927
10.altr_2	-0,069185	0,385196	-0,18	0,857
15.altr_2	0,190168	0,338682	0,56	0,574
20.altr_2	0,018135	0,313272	0,06	0,954
25.altr_2	-0,050762	0,291713	-0,17	0,862
30.altr_2	-0,022377	0,267366	-0,08	0,933
35.altr_2	-0,174230	0,236126	-0,74	0,461
40.altr_2	0,367438	0,189961	1,93	0,053
45.altr_2	0,257031	0,136357	1,88	0,059
55.altr_2	0,125424	0,139629	0,90	0,369
60.altr_2	0,254846	0,207572	1,23	0,220
65.altr_2	0,331771	0,259991	1,28	0,202
70.altr_2	-0,126959	0,321379	-0,40	0,693
75.altr_2	-0,178139	0,363392	-0,49	0,624
80.altr_2	-0,035425	0,405619	-0,09	0,930
85.altr_2	-0,054237	0,456828	-0,12	0,905
90.altr_2	-0,129995	0,520942	-0,25	0,803
1.famst_2	0,147506	0,155103	0,95	0,342
3.famst_2	0,154879	0,176720	0,88	0,381
4.famst_2	0,272119	0,181364	1,50	0,134
2.hws_2	-0,216583	0,171466	-1,26	0,207
1.hhtyp_2	-11,220490	1,279334	-8,77	0,000
2.nat_2	-0,032178	0,208228	-0,15	0,877
2.sex_2	0,000000			
0.casr_2	0,025306	0,117257	0,22	0,829
1.casr_2	-0,058058	0,072071	-0,81	0,420
3.casr_2	-0,064289	0,072940	-0,88	0,378
4.casr_2	0,291572	0,125342	2,33	0,020
5.casr_2	-0,106489	0,083284	-1,28	0,201
1.eink_2	-0,021167	0,052705	-0,40	0,688
3.eink_2	0,073816	0,049955	1,48	0,140
4.eink_2	0,145405	0,071355	2,04	0,042
5.eink_2	0,127951	0,073903	1,73	0,083

Fortsetzung Tabelle A6

r4	Coef.	Std. Err.	z	P> z
0.altr_3	-0,316692	0,369643	-0,86	0,392
5.altr_3	-0,336980	0,348660	-0,97	0,334
10.altr_3	-0,037536	0,315995	-0,12	0,905
15.altr_3	-0,094883	0,287564	-0,33	0,741
20.altr_3	-0,317861	0,261164	-1,22	0,224
25.altr_3	-0,519032	0,242014	-2,14	0,032
30.altr_3	-0,415855	0,221615	-1,88	0,061
35.altr_3	-0,113358	0,196774	-0,58	0,565
40.altr_3	-0,323556	0,165054	-1,96	0,050
45.altr_3	-0,248020	0,118205	-2,10	0,036
55.altr_3	-0,170417	0,109876	-1,55	0,121
60.altr_3	-0,311184	0,162626	-1,91	0,056
65.altr_3	-0,592490	0,205944	-2,88	0,004
70.altr_3	-0,259352	0,268893	-0,96	0,335
75.altr_3	-0,268010	0,301864	-0,89	0,375
80.altr_3	-0,745803	0,329836	-2,26	0,024
85.altr_3	-0,586168	0,366704	-1,60	0,110
90.altr_3	-0,924294	0,417510	-2,21	0,027
1.famst_3	-0,372238	0,120400	-3,09	0,002
3.famst_3	-0,537335	0,145389	-3,70	0,000
4.famst_3	-0,826531	0,135021	-6,12	0,000
2.hws_3	-0,285860	0,162317	-1,76	0,078
1.hhtyp_3	13,161160	1,017180	12,94	0,000
2.nat_3	-0,327659	0,192940	-1,70	0,089
2.sex_3	0,000000			
0.casr_3	-0,241417	0,110685	-2,18	0,029
1.casr_3	-0,143902	0,069640	-2,07	0,039
3.casr_3	-0,180697	0,067065	-2,69	0,007
4.casr_3	-0,300217	0,120172	-2,50	0,012
5.casr_3	0,137829	0,074014	1,86	0,063
1.eink_3	-0,073612	0,048621	-1,51	0,130
3.eink_3	-0,020607	0,045741	-0,45	0,652
4.eink_3	-0,016761	0,063616	-0,26	0,792
5.eink_3	-0,038994	0,070620	-0,55	0,581
<u>_cons</u>	<u>3,011450</u>	<u>0,071531</u>	<u>42,10</u>	<u>0,000</u>
N	97497			
Pseudo R2 (McFadden)	0,0646			
Pseudolikelihood Ratio Chi ²	3995,44			
d.f.	101			
Prob > Chi ²	0,0000			

Quelle: Siehe Tab. 1. Auswahl: Rotationsgruppe 82 (Proxy), Weiterverfolgungskonzept, ohne imputierte Sterbefälle.