

Überblick zu den Gewichtungsvariablen für die DEAS-Erhebungsjahre 2002-2017 (Stand: Juni 2019)

Im Jahr 2018 hat Infas im Auftrag des DZA die Datengewichtung für alle vorliegenden DEAS-Wellen, die sich aus unterschiedlichen Stichproben zusammensetzen, d.h. ab dem Erhebungsjahr 2002, grundlegend überarbeitet. Ziel war einerseits die Umstellung der Längsschnittgewichtung auf eine von Welle zu Welle fortgeschriebene Panelgewichtung und andererseits die Bildung einer integrierten Querschnittgewichtung für jedes Erhebungsjahr unter Einbezug jeweils aller Befragten des jeweiligen Jahres aus allen beteiligten Stichproben. Im Ergebnis liegen nun die folgenden neuen Gewichtungsvariablen vor (Stand: Juni 2019)¹:

Nr	Variablen	2002	2008	2011	2014	2017
1	Reziproke Bleibewahrscheinlichkeit, mündliches Interview	pbleib_96_02	pbleib_02_08	pbleib_08_11	pbleib_11_14	pbleib_14_17
2	Reziproke Bleibewahrscheinlichkeit, Drop-off	pbleibdrop_02	pbleibdrop_08	pbleibdrop_11	pbleibdrop_14	pbleibdrop_17
3	Längsschnittgewicht Panel, mündliches Interview	ls96_02	ls02_08	ls08_11	ls11_14	ls14_17
4	Längsschnittgewicht Panel, Drop-off	lsdrop96_02	lsdrop02_08	lsdrop08_11	lsdrop11_14	lsdrop14_17
5	Integriertes Querschnittgewicht, mündliches Interview	qs_02	qs_08	qs_11	qs_14	qs_17
6	Integriertes Querschnittgewicht, Drop-off	qsdrop_02	qsdrop_08	qsdrop_11	qsdrop_14	qsdrop_17
7	Poststratifiziertes integriertes Querschnittgewicht, mdl. Interview	qsps_02	qsps_08	qsps_11	qsps_14	qsps_17
8	Poststratifiziertes integriertes Querschnittgewicht, Drop-off	qspsdrop_02	qspsdrop_08	qspsdrop_11	qspsdrop_14	qspsdrop_17
9	Gewichtungszellen	zelle_02	zelle_08	zelle_11	zelle_14	zelle_17

Eine ausführliche Beschreibung des Gewichtungsverfahrens und der dabei generierten Gewichtungsvariablen findet sich im Methodenbericht von Infas zur Durchführung der 6. Erhebungswelle 2017 (S. 70ff.) und der Ergänzung zum Methodenbericht mit der zusätzlichen Gewichtung für die Erhebungsjahre ab 2002 (S.101ff.)². Die nachfolgenden Ausführungen beruhen – meist wörtlich – auf den Beschreibungen im Methodenbericht.

Längsschnittgewichtung

Zu 1): Reziproke Bleibewahrscheinlichkeit, mündliches Interview

Zur Berechnung der Längsschnittgewichte wurden Ausfallmodelle (logistische Regression) zur Bestimmung der Teilnahmewahrscheinlichkeit von einer Erhebungswelle zur darauf folgenden Erhebungswelle berechnet. Grundgesamtheit eines jeden Modells sind Teilnehmerinnen und Teilnehmer (Panel- und Basisfälle) in der Ausgangswelle abzüglich Personen, die zwischen der Ausgangswelle und der Betrachtungswelle verstorben sind. Berechnet wird

¹ Die Gewichtsdatei DEAS1996-2017_Gewichte_de_SPSS.sav bzw. DEAS1996-2017_Gewichte_de_SPSS.dta erhalten registrierte Datennutzer zum Download. Jeder Daten-SUF ab dem Erhebungsjahr 2002 enthält bereits die beiden Querschnittgewichtungsvariablen qsps_jj und qspsdrop_jj. Für das DEAS-Startjahr 1996 bleibt es bei den beiden querschnittlichen Gewichtungsvariablen gew_b96_int und gew_b96_drop.

² siehe Anhang.

hierüber die individuelle Wahrscheinlichkeit einer Teilnahme in der Folgewelle. Die Variablen pbleib_96_02 bis pbleib_14_17 enthalten den Kehrwert dieser Bleibewahrscheinlichkeit (1/Bleibewahrscheinlichkeit).

Die Längsschnittgewichte (Variablen ls*) beziehen sich auf den Betrachtungszeitraum von einer Welle zur Folgewelle. Längere Zeithorizonte (z.B. über drei Wellen) können erschlossen werden, indem Gewichte auf folgende Weise gebildet werden: Querschnittgewicht der Ausgangswelle (t1) * reziproke Teilnahmewahrscheinlichkeit Folgewelle (t2) * reziproke Teilnahmewahrscheinlichkeit Folgewelle (t3). Auf diese Weise kann jedes denkbare Längsschnittgewicht erstellt werden. Voraussetzung für einen Vergleich über mehrere Wellen ist, dass die befragte Person in jeder der betrachteten Erhebungswellen an der Befragung teilgenommen haben muss.

Zu 2): Reziproke Bleibewahrscheinlichkeit, Drop-off

Zur Berechnung der Längsschnittgewichte der vorliegenden Drop-off- Fragebögen werden Ausfallmodelle (logistische Regressionen) zur Bestimmung der Teilnahmewahrscheinlichkeit am Drop-off gerechnet. Grundgesamtheit eines jeden Modells sind Panelfälle mit Teilnahme am CAPI-Interview in der Vorwelle sowie Teilnahme am CAPI-Interview in der Betrachtungswelle.

Zu 3): Längsschnittgewicht Panel, mündliches Interview

Um die Längsschnittgewichte zu berechnen, wurde das Querschnittgewicht der Ausgangswelle (Startgewicht = integriertes Querschnittgewicht) mit der reziproken Bleibewahrscheinlichkeit (1/Teilnahmewahrscheinlichkeit) multipliziert:

$$ls_{t1_t2} = qs_{t1} * (1/p_{Teilnahme_t2})$$

Dieses Längsschnittgewicht passt die Verteilungen der Panelfälle an die Verteilungen in der Ausgangswelle an. Somit kann die Veränderung von Verteilungen zwischen der Ausgangswelle und der Folgewelle betrachtet werden. Um einen Vergleich zwischen diesen beiden Wellen durchzuführen, muss die Verteilungsausählung für die Ausgangswelle mit dem Querschnittgewicht der Ausgangswelle und die Verteilungsausählung für die Folgewelle mit dem Längsschnittgewicht gewichtet werden.

Man kann auf diese Weise z.B. beschreiben, wie sich die subjektive Gesundheit der im Jahr 2014 Befragten bis zum Jahr 2017 entwickelt hat. Bei einer Beschreibung der individuellen Entwicklung über längere Zeiträume (z.B. 2008 bis 2017) ist für jede Folgewelle das Längsschnittgewicht durch Multiplikation mit der entsprechenden reziproken Bleibewahrscheinlichkeit fortzuschreiben.

Zu 4): Längsschnittgewicht Panel, Drop-off

Die nach 3) berechnete individuelle Teilnahmewahrscheinlichkeit am Drop-off wird als reziproke Teilnahmewahrscheinlichkeit (1/Teilnahmewahrscheinlichkeit) multiplikativ mit dem berechneten Längsschnittgewicht der Betrachtungswelle verknüpft:

$$lsdrop_{t1_t2} = ls_{t1_t2} * (1/p_{TeilnahmeDropoff_t2})$$

Die so entstehenden Längsschnittgewichte für die Drop-off-Fragebögen passen die Verteilungen für die am Drop-off teilnehmenden Panelfälle an die Verteilungen in der Vorwelle (Ausgangswelle) an. Sie beziehen sich somit auf den Betrachtungszeitraum von einer Welle (Ausgangswelle) zum Drop-off der Folgewelle (Betrachtungswelle).

Querschnittgewichtung

Zu 5): Integriertes Querschnittgewicht, mündliches Interview

Für die Bildung von Querschnittgewichten müssen die Teilstichproben in jeder Erhebungswelle in eine gemeinsame Querschnittstichprobe integriert werden. Diese Teilstichproben sind (soweit in der jeweiligen Welle zutreffend):

- Panelfälle mit Teilnahme in der Vorwelle (Wiederteilnehmer),
- Panelfälle ohne Teilnahme in der Vorwelle (temporäre Ausfälle),
- neue Basisfälle.

Das Ausgangsgewicht für die Integration ist bei den wieder teilnehmenden Panelfällen das Längsschnittgewicht in der Betrachtungswelle und für die Basisstichprobe das Querschnittgewicht (vorhandenes Querschnittgewicht der Basisstichprobe) in der Betrachtungswelle. Für die temporären Ausfälle wurde ein Hilfsgewicht als Konvexgewichtung erstellt (Details hierzu siehe Methodenbericht, S.75f).

Das so entstandene Querschnittgewicht berücksichtigt die Auswahl- und Teilnahmewahrscheinlichkeiten über die verschiedenen Stichproben hinweg. Es dient als Startgewicht für eine Anpassung an bekannte Sollverteilungen (poststratifiziertes Querschnittgewicht). Zudem stellt es ein hilfswises Querschnittgewicht dar, um Längsschnittgewichte zu berechnen.

Integration der Ausländerstichprobe 2002:

Da die Fälle in der Ausländerstichprobe ausschließlich im Jahr 2002 befragt wurden, waren sie lediglich für das Querschnittgewicht, nicht aber für die (bereits erstellten) Längsschnittgewichte zu berücksichtigen. Für die Integration in den Querschnitt erfolgte zunächst eine separate Randanpassung der Ausländerstichprobe. Hierbei wurden die Merkmale Geschlecht, Alter und Landesteil an bekannte Verteilungen in der Grundgesamtheit (Mikrozensus 2002) angepasst. Dies geschah nicht, wie bei den anderen Basisstichproben, durch eine Kombination der Merkmale (Zellen), da die Zellenbesetzung teilweise zu gering ist, sondern anhand der separaten Merkmale. Die Ausländerstichprobe mit ihrem Gewicht wurde an den bereits existierenden integrierten Querschnitt für das Jahr 2002 angehängt.³

Zu 6): Integriertes Querschnittgewicht, Drop-off

Zur Bestimmung der Querschnittgewichte des Drop-offs wurden auf Basis aller Teilnehmer am Interview Ausfallmodelle (logistische Regression) zur Bestimmung der Wahrscheinlichkeit, den Drop-off-Fragebogen in derselben Welle zu beantworten, gerechnet. Das Querschnittgewicht für den Drop-off wurde gebildet, indem das Querschnittgewicht der Betrachtungswelle mit der reziproken Teilnahmewahrscheinlichkeit am Drop-off ($1/\text{Teilnahmewahrscheinlichkeit}$) multipliziert wurde.

Dieses hilfswise erstellte Querschnittgewicht dient als Startgewicht für eine Anpassung von Merkmalen an bekannte Sollverteilungen (poststratifiziertes Gewicht).

Zu 7): Poststratifiziertes integriertes Querschnittgewicht, mündliches Interview

Für die Erhebungsjahre 2002 bis 2017 erfolgte eine Anpassung der Verteilungen Alter*Geschlecht*Landesteil in der Stichprobe an bekannte Verteilungen in der Grundgesamtheit (Mikrozensus) mittels „iterative proportional fitting“ (IPF). Im Erhebungsjahr 2002 wurde zusätzlich als weiteres Merkmal die Staatsbürgerschaft (deutsch vs. nicht deutsch) angepasst. In den Erhebungsjahren 2002 und 2008 erhalten Teilnehmer mit einem Alter von 86 Jahren und

³ Dies ist möglich, da die Auswahlwahrscheinlichkeiten nicht überlappend waren, d.h., dass die Befragten entweder in der Ausländerstichprobe oder in der Stichprobe deutscher Staatsbürger, nicht aber in beiden gleichzeitig, ausgewählt werden konnten.

älter und in den Jahren 2011 bis 2017 mit einem Alter von 91 Jahren und älter wegen zu gering besetzter Zellen kein poststratifiziertes/kalibriertes Querschnittsgewicht. Als Startgewicht diente jeweils das Gewicht welches aus der Integration in den jeweiligen Querschnitt entstanden ist.

Zu 8): Poststratifiziertes integriertes Querschnittsgewicht, Drop-off

Ausgehend vom nach 6) gebildeten integrierten Querschnittsgewicht für den Drop-off erfolgte in gleicher Weise wie für das mündliche Interview mittels IPF eine Anpassung der Verteilungen Alter*Geschlecht*Landesteil. Im Erhebungsjahr 2002 wurde zusätzlich anhand des Merkmals Staatsbürgerschaft kalibriert.

Zu 9): Gewichtungszellen

Die Gewichtungszellen enthalten die Fallzahlen der einzelnen Merkmalskombinationen aus Altersgruppe, Geschlecht und Landesteil der integrierten Stichproben im jeweiligen Erhebungsjahr. Sie dienen als Grundlage der Poststratifizierung mittels IPF. In den Jahren 2002 und 2008 enthält die Alterseinteilung die Kategorien 40-54, 55-69 und 70-85 Jahre. Ab dem Jahr 2011 gibt es die zusätzliche Altersgruppe 86-90 Jahre. In den Jahren 2011 und 2017 ohne Basisstichprobe umfasst die jüngste Altersgruppe den Altersbereich 43-54 Jahre.

Alle Gewichtungsvariablen (ls^* , qs^*) sind fallzahlnormiert auf die gültige Fallzahl der Betrachtungswelle, d.h. ihr Mittelwert beträgt jeweils 1.

Methodenbericht

Deutscher Alterssurvey (DEAS): Durchführung der 6. Erhebungswelle 2017

Autoren: Stefan Schiel, Jonathan Ruiz Marcos, Christian Dickmann, Folkert Aust, Lena Middendorf

Auszug

infas Institut für angewandte
Sozialwissenschaft GmbH

Friedrich-Wilhelm-Straße 18
D-53113 Bonn
Tel. +49 (0)228/38 22-0
Fax +49 (0)228/31 00 71
info@infas.de
www.infas.de

7 Gewichtung der Daten

Rückwirkend für alle Erhebungsjahre (2002, 2008, 2011, 2014 und 2017) wurde eine Gewichtung der Daten durchgeführt. Sie umfasst zum einen eine Längsschnittgewichtung für die CAPI-Interviews und die auswertbaren Drop-off-Fragebögen über alle Erhebungsjahre, zum anderen eine Querschnittgewichtung: die Stichproben wurden für jedes Erhebungsjahr in einen gemeinsamen Querschnitt integriert. Für die Erhebungsjahre 2011 und 2017 wurden jeweils separate poststratifizierte bzw. kalibrierte Querschnittgewichte für die CAPI-Interviews und die Drop-off-Fragebögen erstellt.

7.1 Längsschnittgewichtung

7.1.1 CAPI-Interview

Zur Berechnung der Längsschnittgewichte wurden Ausfallmodelle (logistische Regression) zur Bestimmung der Teilnahmewahrscheinlichkeit von einer Erhebungswelle zur darauf folgenden Erhebungswelle berechnet.²⁰ Grundgesamtheit eines jeden Modells sind Teilnehmerinnen und Teilnehmer (Panel- und Basisfälle) in der Ausgangswelle²¹ abzüglich Personen, die zwischen der Ausgangswelle und der Betrachtungswelle verstorben sind. Berechnet wird hierüber die individuelle Wahrscheinlichkeit einer Teilnahme in der Folgewelle. Prädikatoren²² eines jeden Modells sind:

- Landesteil: West, Ost
- Kreistypen nach BBSR gruppiert: kreisfreie Großstädte, städtische Kreise, ländliche Kreise/dünn besiedelte ländliche Kreise
- Alter gruppiert²³: bis 49 Jahre, 50 bis 59 Jahre, 60 bis 69 Jahre, 70 Jahre und älter
- Geschlecht: männlich, weiblich
- Bildungsabschluss ISCED gruppiert: niedrig, mittel, hoch
- Netzwerkgröße gruppiert: bis 2 Personen, 3 bis 5 Personen, 6 und mehr Personen
- Äquivalenzeinkommen gruppiert: niedriges Einkommen/keine Angabe, mittleres bis hohes Einkommen

²⁰ Die Ergebnisse der Ausfallmodelle sind im Anhang dargestellt.

²¹ Ausgangswelle bezeichnet die Startwelle, d.h. die vorangegangene Erhebung der betrachteten Erhebungswelle, nicht die Startwelle einer Basisstichprobe.

²² Die Auswahl der Prädikatoren erfolgte in Abstimmung mit dem DZA.

²³ Die Berechnung des Alters erfolgte über das Geburtsjahr der Befragungsperson: $\text{Alter} = \text{Erhebungsjahr} - \text{Geburtsjahr}$.

– Subjektiver Gesundheitsstatus gruppiert: sehr gut/gut, mittel, schlecht/sehr schlecht

Berücksichtigt wurden die Ausprägungen einer Variablen in der Ausgangswelle. Das Geschlecht ist konstant, Landesteil und Bildung bezieht sich auf die Erstbefragung der Basisstichprobe eines Falls.

Um die Längsschnittgewichte zu berechnen, wurde das Querschnittgewicht der Ausgangswelle (Startgewicht = integriertes Querschnittgewicht) mit der reziproken Bleibewahrscheinlichkeit (1/Teilnahmewahrscheinlichkeit) multipliziert:

$$ls_{t1,t2} = qs_{t1} * (1/p_{\text{Teilnahme}_{t2}})$$

Dieses Längsschnittgewicht passt die Verteilungen der Panelfälle an die Verteilungen in der Ausgangswelle an. Somit kann die Veränderung von Verteilungen zwischen der Ausgangswelle und der Folgewelle betrachtet werden.

Hinweise zur Erstellung und Benutzung der Gewichte: Die an das DZA übergebenen Längsschnittgewichte beziehen sich auf den Betrachtungszeitraum von einer Welle zur Folgewelle. Um einen Vergleich zwischen diesen beiden Wellen durchzuführen, muss die Verteilungsauszahlung für die Ausgangswelle mit dem Querschnittgewicht der Ausgangswelle und die Verteilungsauszahlung für die Folgewelle mit dem Längsschnittgewicht gewichtet werden. Längere Zeithorizonte (Beispielsweise 3 Wellen) können erschlossen werden, indem Gewichte auf folgende Weise gebildet werden: Querschnittgewicht der Ausgangswelle (t1) * reziproke Teilnahmewahrscheinlichkeit Folgewelle (t2) * reziproke Teilnahmewahrscheinlichkeit Folgewelle (t3). Auf diese Weise kann jedes denkbare Längsschnittgewicht erstellt werden. Voraussetzung für einen Vergleich über mehrere Wellen ist, dass die befragte Person in jeder der betrachteten Erhebungswelle an der Befragung teilgenommen haben muss.

7.1.2 Drop-off

Zur Berechnung der Längsschnittgewichte der vorliegenden Drop-off-Fragebögen werden Ausfallmodelle (logistische Regressionen) zur Bestimmung der Teilnahmewahrscheinlichkeit am Drop-off gerechnet.²⁴ Grundgesamtheit eines jeden Modells sind Panelfälle mit Teilnahme am CAPI-Interview in der Vorwelle sowie Teilnahme am CAPI-Interview in der Betrachtungswelle.

²⁴ Die Ergebnisse der Ausfallmodelle sind im Anhang dargestellt.

Prädikatoren²⁵ der Modelle sind:

- Landesteil: West, Ost
- Geschlecht: männlich, weiblich
- Alter gruppiert: bis 49 Jahre, 50 bis 59 Jahre, 60 bis 69 Jahre, 70 Jahre und älter
- Bildungsabschluss ISCED gruppiert: niedrig, mittel, hoch

Der Landesteil bezieht sich auf den Zeitpunkt der Erstbefragung der Basisstichprobe, das Geschlecht ist konstant. Das Alter ist das Alter zum Zeitpunkt der Betrachtungswelle.

Die so berechnete individuelle Teilnahmewahrscheinlichkeit wird als reziproke Teilnahmewahrscheinlichkeit (1/Teilnahmewahrscheinlichkeit) multiplikativ mit dem berechneten Längsschnittgewicht der Betrachtungswelle verknüpft:

$$ls_{drop_t1_t2} = ls_{t1_t2} * (1/p_{TeilnahmeDropoff_t2})$$

Die so entstehenden Längsschnittgewichte für die Drop-off-Fragebögen passen die Verteilungen für die am Drop-off teilnehmenden Panelfälle an die Verteilungen in der Vorwelle (Ausgangswelle) an. Sie beziehen sich somit auf den Betrachtungszeitraum von einer Welle (Ausgangswelle) zum Drop-off der Folgewelle (Betrachtungswelle).

7.1.3 Kennziffern der Gewichtungsfaktoren

Zur Beurteilung der Längsschnittgewichte wird im Folgenden neben dem Mittelwert und der Standardabweichung auch das Effektivitätsmaß angegeben (E)²⁶. Das Effektivitätsmaß E basiert auf der Varianz des Gewichtungsfaktors. Das Effektivitätsmaß gibt in Prozent der realisierten Fallzahl an, wie groß bei Verwendung des Gewichts die effektive Fallzahl (n') bei einem passiven Merkmal ist, das mit den aktiven Merkmalen nicht korreliert (Worst-case-Szenario).²⁷ Die effektive Fallzahl entspricht der Anzahl Befragter, die bei einer uneingeschränkten Zufallsauswahl, gegeben die Varianz des Merkmals in der Stichprobe, den gleichen Stichprobenfehler produziert hätte. Das Effektivitätsmaß drückt das Verhältnis von n zu n' als Prozentwert aus und wird berechnet über:

²⁵ Die Auswahl der Prädikatoren erfolgte in Abstimmung mit dem DZA.

²⁶ Vergleiche zum Effektivitätsmaß und Designeffekt (E = 1/Designeffekt) auch:

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Kish, L. (1987). Weighting in Deft2. *The Survey Statistician*. June 1987.

Little, R. J.A., S. Lewitzky, S. Heeringa, J. Lepkowski & R.C. Kessler (1997): Assessment of weighting methodology for the National Comorbidity Survey. *American Journal of Epidemiology*, 146, 439–449.

Rösch, G. (1994): Kriterien der Gewichtung einer nationalen Bevölkerungsstichprobe: Gabler, S., J. H. P. Hoffmeyer-Zlotnik & D. Krebs (Hg.): *Gewichtung in der Umfragepraxis*, Westdeutscher Verlag, 7-26.

²⁷ Aktive Merkmale sind Merkmale, die bei der Kalibrierung eingesetzt wurden. Alle übrigen erhobenen Informationen sind passive Merkmale.

$$E = (n'/n) * 100, \text{ wobei: } n' = ((\sum_i g_i)^2 / \sum_i g_i^2)$$

Die folgende Tabelle stellt die Kennziffern für die Längsschnittgewichte dar:

Tabelle 36 Kennziffern für die Längsschnittgewichtungsfaktoren

Kennwert	Erhebung 2002		Erhebung 2008		Erhebung 2011		Erhebung 2014		Erhebung 2017	
	CAPI-Interview	Drop-off	CAPI-Interview	Drop-off	CAPI-Interview	Drop-off	CAPI-Interview	Drop-off	CAPI-Interview	Drop-off
Mittelwert	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Standardabweichung	0,53	0,53	0,83	0,75	1,15	1,30	0,94	0,89	1,06	1,13
Minimum	0,21	0,21	0,18	0,17	0,11	0,99	0,10	0,10	0,05	0,05
Maximum	4,64	4,80	13,19	9,05	41,02	40,91	21,95	12,43	18,81	20,60
Effektivitätsmaß	78,10%	77,91%	59,28%	63,86%	43,25%	37,26%	52,70%	55,72%	47,24%	44,10%
Effektive Fallzahl	1.190	1.119	1.031	923	1.809	1.291	1.885	1.727	2.960	2.360
Anzahl gewichtete Fälle	1.524	1.437	1.739	1.445	4.183	3.468	3.577	3.103	6.265	5.351

Quelle: infas-eigene Berechnung

Bei der Bewertung der dargestellten Kennziffern ist das Ziel der Gewichtung zu berücksichtigen. Das Hauptziel der Gewichtung ist der Ausgleich von Verzerrungen, die durch das Stichprobendesign (mit ungleichen Auswahlwahrscheinlichkeiten) und durch selektive Teilnahme bzw. Nichtteilnahme (ungleiche Realisierungswahrscheinlichkeiten) entstehen. Unter Verwendung der Gewichte lassen sich Populationswerte aus der Stichprobe erwartungstreu²⁸ schätzen. Weisen die Gewichte allerdings eine starke Streuung auf, so kann dies zu einer großen Varianz der Schätzfunktionen führen. Hier tritt der in der Statistik immer wieder anzutreffende Trade-off zwischen Bias und Varianz auf. Die Gewichtung reduziert den Bias, eine allzu große Steigerung der Varianz durch die Gewichtung sollte allerdings auch vermieden werden.

²⁸ Ein Schätzer heißt erwartungstreu bzw. unverzerrt, wenn sein Erwartungswert gleich dem wahren Wert des zu schätzenden Parameters ist. Ist eine Schätzfunktion nicht erwartungstreu, spricht man davon, dass der Schätzer verzerrt ist. Das Ausmaß der Abweichung seines Erwartungswerts vom wahren Wert heißt Verzerrung oder Bias. Die Verzerrung drückt den systematischen Fehler des Schätzers aus. Erwartungstreue ist eine wichtige Eigenschaft eines Schätzers, da die Varianz der meisten Schätzer mit steigendem Stichprobenumfang gegen Null konvergiert. Bei erwartungstreuen Schätzern lässt sich erwarten, dass die Differenz zwischen dem aus der Stichprobe berechneten Schätzwert und dem wahren Wert umso kleiner ist, je größer der Stichprobenumfang ist. Die Varianz eines erwartungstreuen Schätzers konvergiert also gegen 0, wenn der Stichprobenumfang n gegen unendlich geht.

Hinsichtlich der Kennziffern bedeutet dies, dass eine hohe Standardabweichung der Gewichte und in der Folge ein geringes Effektivitätsmaß und eine geringe effektive Fallzahl möglichst vermieden werden sollten, da in der Folge die Konfidenzintervalle der Populationsschätzungen breiter werden, d.h. der Stichprobenfehler entsprechend steigt. Aus dieser Perspektive wäre eine Effektivität von 100 Prozent (effektive Fallzahl entspricht der tatsächlichen Fallzahl) ideal.

Andererseits bedeutet ein sehr hohes Effektivitätsmaß (eine geringe Varianz der Gewichte), dass die Gewichtung keinen Effekt auf die Populationsschätzungen hat, eine vermutete Verzerrung der Schätzungen also nicht ausgeglichen wird. Idealerweise gibt es keine Verzerrung und die Gewichtung ist deshalb unnötig. Möglich ist aber auch, dass die Gewichtung eine vorhandene Verzerrung nicht ausgleicht. Aus dieser Perspektive sind hohe Standardabweichungen der Gewichtungsfaktoren und geringe Effektivitätsmaße ein Ausdruck dafür, dass die Gewichtung zumindest einen Teil der Verzerrungen ausgleicht und damit die Populationsschätzungen verbessert, allerdings um den Preis eines höheren Stichprobenfehlers, d.h. breiterer Konfidenzintervalle um den geschätzten Wert.

Bei Längsschnitterhebungen tritt häufig ein kontinuierliches Absinken der Effektivität auf, da sich die Selektivitätseffekte über die Wellen kumulieren. Selten lässt sich beobachten, dass die Selektivitäten (beispielsweise durch die verstärkte Teilnahme von temporären Ausfällen) wieder zurückgehen und die Effektivität steigt.

Die Kennziffern beim DEAS mit einem Längsschnittansatz und regelmäßigen Auffrischungstichproben bewegen sich insgesamt in einem üblichen Rahmen und sinken nicht dramatisch ab. Wesentlich ist auch, dass die effektive Fallzahl in der Regel noch so hoch bleibt, dass die Konfidenzintervalle in einer akzeptablen Größenordnung bleiben. Hierbei ist auch noch einmal darauf hinzuweisen, dass es sich bei den effektiven Fallzahlen um ein Worst-Case-Szenario handelt (gilt nur für Merkmale, die überhaupt nicht mit den aktiven Gewichtungsmerkmalen korrelieren), d.h. bei der überwiegenden Zahl der Merkmale wird die effektive Fallzahl (deutlich) höher sein.

7.2 Integration der Stichproben in einen gemeinsamen Querschnitt

7.2.1 CAPI-Interview

Für die Bildung von Querschnittsgewichten müssen die Teilstichproben in jeder Erhebungswelle in eine gemeinsame Querschnittstichprobe integriert werden. Diese Teilstichproben sind (soweit in der jeweiligen Welle zutreffend):

- Panelfälle mit Teilnahme in der Vorwelle (Wiederteilnehmer),
- Panelfälle ohne Teilnahme in der Vorwelle (temporäre Ausfälle),
- neue Basisfälle.

Das Ausgangsgewicht für die Integration ist bei den wieder teilnehmenden Panelfällen das Längsschnittgewicht in der Betrachtungswelle und für die Basisstichprobe das Querschnittgewicht (vorhandenes Querschnittgewicht der Basisstichprobe) in der Betrachtungswelle. Für die temporären Ausfälle wurde ein Hilfsgewicht erstellt. Basis für dieses Hilfsgewicht war das Querschnittgewicht der Welle, in dem der Fall das letzte Mal teilgenommen hat. Dieses Gewicht wurde anhand der Ergebnisse zweier Ausfallmodelle adjustiert: (1) Die individuelle Wahrscheinlichkeit einer Nichtteilnahme in der Folgewelle nach der letzten Teilnahme (1-Teilnahmewahrscheinlichkeit) wurde in den Ausfallmodellen der Längsschnittgewichtung berechnet. (2) Zur Bestimmung der individuellen Wahrscheinlichkeit, in der Betrachtungswelle teilzunehmen, unter der Voraussetzung in der Vorwelle (bzw. in den Vorwellen) nicht teilgenommen zu haben (Rückkehrwahrscheinlichkeit), wurden separate logistische Regressionen gerechnet.²⁹ Grundgesamtheit der jeweiligen Modelle sind alle Panelfälle ohne Teilnahme in der Ausgangswelle (Vorwelle der Betrachtungswelle). Prädikatoren eines jeden Modells sind:

- Landesteil: West, Ost.
- Geschlecht: männlich, weiblich.
- Alter gruppiert: bis 49 Jahre, 50 bis 59 Jahre, 60 bis 69 Jahre, 70 Jahre und älter.

Der Landesteil bezieht sich auf die erste Erhebungswelle eines Falls, das Geschlecht ist konstant. Das Alter ist das Alter zum Zeitpunkt der Ausgangswelle des Modells.

Zur Bildung des Hilfsgewichts werden folgende Elemente multiplikativ verknüpft: Querschnittgewicht der Welle, in welcher der Fall das letzte Mal teilgenommen hat, reziproke Wahrscheinlichkeit der Nichtteilnahme in der Folgewelle nach der letzten Teilnahme (1-Teilnahmewahrscheinlichkeit aus den Ausfallmodellen zur Berechnung der Längsschnittgewichte), reziproke Rückkehrwahrscheinlichkeit in der Betrachtungswelle:

$$hw_{t3} = qs_{t1} * (1/(1-p_{\text{Teilnahme}_{t2}})) * (1/p_{\text{Teilnahme}_{t3}})$$

Die Zusammenführung der Teilstichproben in eine gemeinsame Stichprobe muss berücksichtigen, dass die Auswahlgesamtheiten jeder der beiden Teilstichproben überlappend waren. Befragte konnten also sowohl in der einen als auch in der anderen Stichprobe ausgewählt werden. Die Auswahlwahrscheinlichkeiten, auf deren Basis das Designgewicht berechnet wird, müssen dementsprechend angepasst werden.

Dies kann auf im Wesentlichen zwei unterschiedliche Arten erfolgen, zum einen über die fallweise Bestimmung der Wahrscheinlichkeiten, jeweils in den beiden Stichproben enthalten zu sein. Dies setzt allerdings weitreichende Informationen z.B. über die Mobilität (Zuzüge und Fortzüge) voraus und ist, insbesondere auch für die temporären Ausfälle, schwierig bzw. gar nicht möglich. Alternativ

²⁹ Die Ergebnisse der Modelle sind im Anhang dargestellt.

kann die Integration auch über eine Konvexkombination (composite weighting) erfolgen, was für den Alterssurvey die zweckmäßige Vorgehensweise darstellt.³⁰

Ist die Voraussetzung erfüllt, dass zwei Stichproben vorliegen, die erwartungstreue Schätzungen für den wahren Parameter erlauben, dann ist jede Konvexkombination, d.h. die Berechnung von Konvexgewichten, die sich zur Gesamtheit summieren, der erwartungstreuen Schätzer ebenfalls erwartungstreu. Bei der Konvexgewichtung werden die ursprünglichen Ausgangsgewichte jeder Stichprobe mit einem beliebigen Faktor multipliziert, wobei die Gewichte der Stichprobe 1 mit einem beliebigen Faktor α , die Gewichte der Stichprobe 2 mit $1-\alpha$ multipliziert wird. Als Faktor kann dabei auch die Fallzahl verwendet werden:

$$\alpha = n_1/(n_1 + n_2) \text{ und } 1 - \alpha = n_2/(n_1 + n_2).$$

Dies entspricht der einfachen Regel, die ursprünglichen Gewichte mit dem Anteil der jeweiligen Teilstichprobe am Gesamtstichprobenumfang zu multiplizieren.

Eine Konvexkombination der Gewichte führt also zu erwartungstreuen Schätzern, allerdings kann unter Umständen die Varianz der Gewichte höher sein als bei der Bestimmung der Auswahlwahrscheinlichkeiten.

Die Konvexkombination erfolgte in zwei Stufen: zunächst wurden die Panelteilnehmer mit Wiederteilnahme mit solchen mit temporärem Ausfall zusammengefügt. In einem weiteren Schritt erfolgte die Zusammenführung der Panelfälle mit der Basisstichprobe. Dabei wurden die ursprünglichen Gewichte jeweils mit dem Anteil der jeweiligen Teilstichprobe am Gesamtstichprobenumfang multipliziert.

Das so entstandene Querschnittsgewicht berücksichtigt die Auswahl- und Teilnahmewahrscheinlichkeiten über die verschiedenen Stichproben hinweg. Es dient als Startgewicht für eine Anpassung an bekannte Sollverteilungen (poststratifiziertes Querschnittsgewicht). Zudem stellt es ein hilfswises Querschnittsgewicht dar, um Längsschnittgewichte zu berechnen.

³⁰ Lohr, Sharon L. (2011). Alternative survey sample designs: Sampling with multiple overlapping frames. *Statistics Canada - Survey Methodology*, 37:197–213.

Brick, J. M.; Dipko, S.; Presser S. Tucker C.; Yuan Y. (2006). Nonresponse Bias in a Dual Frame Sample of Cell and Landline Numbers. *Public Opinion Quarterly*, 70:780 – 793.

Xia, K.; Pedlow, S.; Davern M. (2010): Dual-Frame Weights (Landline and Cell) for the 2009 Minnesota Health Access Survey. *American Statistical Association - Proceedings of the Survey Research Methods Section*, S. 3912 – 3922.

Spieß, M. & Rendtel, U. (2000): Combining an ongoing panel with a new cross-sectional sample. *DIW-Discussion Papers* 198. Berlin.

Rendtel, U. (1999): The Application of the Convex Weighting Estimator to Household Panel Surveys. Mimeo, Frankfurt.

7.2.2 Drop-off

Zur Bestimmung der Querschnittgewichte des Drop-offs werden auf Basis aller Teilnehmer am Interview Ausfallmodelle (logistische Regression) zur Bestimmung der Wahrscheinlichkeit, den Drop-off-Fragebogen in derselben Welle zu beantworten, gerechnet³¹. Prädiktoren des Modells sind:

- Landesteil: West, Ost
- Geschlecht: männlich, weiblich
- Alter gruppiert: bis 49 Jahre, 50 bis 59 Jahre, 60 bis 69 Jahre, 70 Jahre und älter
- Bildungsabschluss ISCED gruppiert: niedrig, mittel, hoch

Der Landesteil bezieht sich auf den Zeitpunkt der Ersterhebung, das Geschlecht ist konstant. Das Alter ist das Alter zum Zeitpunkt der Betrachtungswelle.

Das Querschnittgewicht für den Drop-off für die Erhebungsjahre 2011 und 2017 wurde gebildet, indem das Querschnittgewicht der Betrachtungswelle mit der reziproken Teilnahmewahrscheinlichkeit am Drop-off (1/Teilnahmewahrscheinlichkeit) multipliziert wurde. Dieses hilfsweise erstellte Querschnittgewicht dient als Startgewicht für eine Anpassung von Merkmalen an bekannte Sollverteilungen (poststratifiziertes Gewicht).

7.3 Poststratifizierte Querschnittgewichtung

Für die Erhebungsjahre 2011 und 2017 erfolgte eine Anpassung der Verteilungen Alter*Geschlecht*Landesteil in der Stichprobe an bekannte Verteilungen in der Grundgesamtheit (Mikrozensus 2011 und 2017) mittels „iterative proportional fitting“ (IPF³²). Dies geschah sowohl für das CAPI-Interview als auch für die Drop-off-Interviews.

Als Startgewicht diente das Gewicht welches aus der Integration in den jeweiligen Querschnitt bzw. aus der zusätzlichen Berücksichtigung der Teilnahmewahrscheinlichkeit am Drop-off-Fragebogen, entstanden ist. Teilnehmer mit einem Alter von 91 Jahren und älter erhalten nach Absprache mit dem DZA kein poststratifiziertes/kalibriertes Querschnittgewicht.

³¹ Die Ergebnisse der Ausfallmodelle sind im Anhang dargestellt.

³² Ein Verteilungsvergleich der ungewichteten und gewichteten Stichprobe ist im Anhang dargestellt.

7.3.1 Kennziffern der Gewichtungsfaktoren

Analog zu den Längsschnittgewichten, werden in folgender Tabelle die Kennziffern der poststratifizierten Querschnittgewichte dargestellt:

Tabelle 37 Kennziffern für die Querschnittgewichtungsfaktoren

Kennwert	Erhebung 2011		Erhebung 2017	
	CAPI-Interview	Drop-off	CAPI-Interview	Drop-off
Mittelwert	1,00	1,00	1,00	1,00
Standardabweichung	0,92	1,00	1,12	1,20
Minimum	0,07	0,06	0,05	0,04
Maximum	13,31	14,63	20,35	19,84
Effektivitätsmaß	54,43%	50,01%	44,45%	40,95%
Effektive Fallzahl	2.631	1.997	2.921	2.284
Anzahl gewichtete Fälle	4.834	3.993	6.572	5.576

Quelle: infas-eigene Berechnung

Anhang

**Tabelle 38 Ausfallmodellierung Panelstichprobe: CAPI-Interview in 2002,
Basis: Fälle mit Interview in 1996 (logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,27	0,00
Kreistyp (BBSR)		
Kreisfreie Großstadt	0,90	0,23
Städtische Kreise	Referenz	
(Dünn besiedelte) Ländliche Kreise	1,04	0,60
Alter		
Bis 49 Jahre	0,96	0,64
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	0,86	0,10
70 Jahre und älter	0,55	0,00
Geschlecht		
Männer	Referenz	
Frauen	0,93	0,30
Bildung (ISCED)		
Niedrig	0,73	0,00
Mittel	Referenz	
Hoch	1,42	0,00
Netzwerkgröße		
Bis 2 Personen	0,72	0,00
3 bis 5 Personen	Referenz	
6 und mehr Personen	1,11	0,18
Einkommen		
Niedriges Einkommen/keine Angabe	Referenz	
Mittleres bis hohes Einkommen	1,53	0,00
Subjektiver Gesundheitsstatus		
Sehr gut/gut	1,14	0,08
Mittel	Referenz	
Schlecht/sehr schlecht	0,72	0,01
Konstante	0,45	0,00
n		4.553
Pseudo-R²		0,05

Hinweis zur Fallzahl: CAPI-Interview Erhebung 1996, ohne Verstorbene
Quelle: infas-eigene Berechnung

**Tabelle 39 Ausfallmodellierung Panelstichprobe: CAPI-Interview in 2008,
Basis: Fälle mit Interview in 2002 (logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	0,93	0,34
Kreistyp (BBSR)		
Kreisfreie Großstadt	1,06	0,45
Städtische Kreise	Referenz	
(Dünn besiedelte) Ländliche Kreise	1,10	0,24
Alter		
Bis 49 Jahre	0,78	0,01
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,07	0,47
70 Jahre und älter	0,67	0,00
Geschlecht		
Männer	Referenz	
Frauen	1,12	0,08
Bildung (ISCED)		
Niedrig	0,60	0,00
Mittel	Referenz	
Hoch	1,34	0,00
Netzwerkgröße		
Bis 2 Personen	0,74	0,00
3 bis 5 Personen	Referenz	
6 und mehr Personen	1,25	0,00
Einkommen		
Niedriges Einkommen/keine Angabe	Referenz	
Mittleres bis hohes Einkommen	1,31	0,00
Subjektiver Gesundheitsstatus		
Sehr gut/gut	1,12	0,11
Mittel	Referenz	
Schlecht/sehr schlecht	0,53	0,00
Konstante	0,58	0,00
n		4.299
Pseudo-R²		0,05

Hinweis zur Fallzahl: CAPI-Interview Erhebung 2002, ohne Verstorbene
Quelle: infas-eigene Berechnung

**Tabelle 40 Ausfallmodellierung Panelstichprobe: CAPI-Interview in 2011,
Basis: Fälle mit Interview in 2008 (logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,18	0,00
Kreistyp (BBSR)		
Kreisfreie Großstadt	0,86	0,02
Städtische Kreise	Referenz	
(Dünn besiedelte) Ländliche Kreise	0,78	0,00
Alter		
Bis 49 Jahre	0,80	0,00
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,26	0,00
70 Jahre und älter	0,98	0,78
Geschlecht		
Männer	Referenz	
Frauen	1,12	0,02
Bildung (ISCED)		
Niedrig	0,73	0,00
Mittel	Referenz	
Hoch	1,46	0,00
Netzwerkgröße		
Bis 2 Personen	0,65	0,00
3 bis 5 Personen	Referenz	
6 und mehr Personen	1,08	0,17
Einkommen		
Niedriges Einkommen/keine Angabe	Referenz	
Mittleres bis hohes Einkommen	1,68	0,00
Subjektiver Gesundheitsstatus		
Sehr gut/gut	0,98	0,72
Mittel	Referenz	
Schlecht/sehr schlecht	0,75	0,00
Konstante		
	0,92	0,30
n		7.914
Pseudo-R²		0,05

Hinweis zur Fallzahl: CAPI-Interview Erhebung 2008, ohne Verstorbene
Quelle: infas-eigene Berechnung

**Tabelle 41 Ausfallmodellierung Panelstichprobe: CAPI-Interview in 2014,
Basis: Fälle mit Interview in 2011 (logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,27	0,01
Kreistyp (BBSR)		
Kreisfreie Großstadt	1,25	0,02
Städtische Kreise	Referenz	
(Dünn besiedelte) Ländliche Kreise	1,04	0,64
Alter		
Bis 49 Jahre	0,78	0,08
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	0,97	0,76
70 Jahre und älter	0,81	0,03
Geschlecht		
Männer	Referenz	
Frauen	1,26	0,00
Bildung (ISCED)		
Niedrig	0,70	0,01
Mittel	Referenz	
Hoch	1,22	0,01
Netzwerkgröße		
Bis 2 Personen	0,78	0,01
3 bis 5 Personen	Referenz	
6 und mehr Personen	1,19	0,03
Einkommen		
Niedriges Einkommen/keine Angabe	Referenz	
Mittleres bis hohes Einkommen	1,15	0,06
Subjektiver Gesundheitsstatus		
Sehr gut/gut	1,15	0,08
Mittel	Referenz	
Schlecht/sehr schlecht	0,66	0,00
Konstante		
	2,44	0,00
n		4.661
Pseudo-R²		0,02

Hinweis zur Fallzahl: CAPI-Interview Erhebung 2011, ohne Verstorbene
Quelle: infas-eigene Berechnung

**Tabelle 42 Ausfallmodellierung Panelstichprobe: CAPI-Interview in 2017,
Basis: Fälle mit Interview in 2014 (logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,15	0,01
Kreistyp (BBSR)		
Kreisfreie Großstadt	0,93	0,18
Städtische Kreise	Referenz	
(Dünn besiedelte) Ländliche Kreise	0,97	0,60
Alter		
Bis 49 Jahre	0,72	0,00
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,17	0,01
70 Jahre und älter	0,95	0,39
Geschlecht		
Männer	Referenz	
Frauen	0,98	0,58
Bildung (ISCED)		
Niedrig	0,60	0,00
Mittel	Referenz	
Hoch	1,24	0,00
Netzwerkgröße		
Bis 2 Personen	0,77	0,00
3 bis 5 Personen	Referenz	
6 und mehr Personen	1,22	0,00
Einkommen		
Niedriges Einkommen/keine Angabe	Referenz	
Mittleres bis hohes Einkommen	2,78	0,00
Subjektiver Gesundheitsstatus		
Sehr gut/gut	1,25	0,00
Mittel	Referenz	
Schlecht/sehr schlecht	0,89	0,11
Konstante		
	1,22	0,01
n		9.963
Pseudo-R²		0,05

Hinweis zur Fallzahl: CAPI-Interview Erhebung 2014, ohne Verstorbene
Quelle: infas-eigene Berechnung

**Tabelle 43 Ausfallmodellierung Panelstichprobe temporäre Ausfälle:
CAPI-Interview in 2008, Basis: Fälle mit Nichtteilnahme in 2002
(logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Alter		
Bis 49 Jahre	1,01	0,96
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	0,65	0,01
70 Jahre und älter	0,31	0,00
Geschlecht		
Männer	Referenz	
Frauen	0,78	0,06
Landesteil		
West	Referenz	
Ost	1,53	0,00
Konstante	0,15	0,00
n		2.839
Pseudo-R²		0,04

Quelle: infas-eigene Berechnung

**Tabelle 44 Ausfallmodellierung Panelstichprobe temporäre Ausfälle:
CAPI-Interview in 2011, Basis: Fälle mit Nichtteilnahme in 2008
(logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Alter		
Bis 49 Jahre	1,07	0,72
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,01	0,93
70 Jahre und älter	0,34	0,00
Geschlecht		
Männer	Referenz	
Frauen	0,84	0,04
Landesteil		
West	Referenz	
Ost	1,36	0,00
Konstante	0,24	0,00
n		4.880
Pseudo-R²		0,05

Quelle: infas-eigene Berechnung

**Tabelle 45 Ausfallmodellierung Panelstichprobe temporäre Ausfälle:
CAPI-Interview in 2014, Basis: Fälle mit Nichtteilnahme in 2011
(logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Alter		
Bis 49 Jahre	0,87	0,35
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	0,72	0,00
70 Jahre und älter	0,31	0,00
Geschlecht		
Männer	Referenz	
Frauen	0,94	0,39
Landesteil		
West	Referenz	
Ost	0,76	0,00
Konstante	0,21	0,00
n		7.693
Pseudo-R²		0,04

Quelle: infas-eigene Berechnung

**Tabelle 46 Ausfallmodellierung Panelstichprobe temporäre Ausfälle:
CAPI-Interview in 2017, Basis: Fälle mit Nichtteilnahme in 2014
(logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Alter		
Bis 49 Jahre	1,27	0,32
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	0,74	0,04
70 Jahre und älter	0,34	0,00
Geschlecht		
Männer	Referenz	
Frauen	1,00	0,97
Landesteil		
West	Referenz	
Ost	0,81	0,09
Konstante	0,09	0,00
n		7.808
Pseudo-R²		0,03

Quelle: infas-eigene Berechnung

Tabelle 47 Ausfallmodellierung Panelstichprobe: Drop-off-Fragebogen in 2002, Basis: Fälle mit CAPI-Interview in 1996 und CAPI-Interview in 2002 (logistische Regression)

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,03	0,88
Geschlecht		
Männer	Referenz	
Frauen	1,12	0,63
Alter		
Bis 49 Jahre	1,09	0,82
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,78	0,06
70 Jahre und älter	1,21	0,50
Bildung (ISCED)		
Niedrig	0,56	0,09
Mittel	Referenz	
Hoch	0,89	0,66
Konstante	13,94	0,00
n		1.524
Pseudo-R²		0,01

Quelle: infas-eigene Berechnung

Tabelle 48 Ausfallmodellierung Panelstichprobe: Drop-off-Fragebogen in 2008, Basis: Fälle mit CAPI-Interview in 2002 und CAPI-Interview in 2008 (logistische Regression)

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,25	0,11
Geschlecht		
Männer	Referenz	
Frauen	0,94	0,65
Alter		
Bis 49 Jahre	0,45	0,00
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,11	0,56
70 Jahre und älter	1,03	0,85
Bildung (ISCED)		
Niedrig	0,92	0,72
Mittel	Referenz	
Hoch	1,11	0,45
Konstante	4,68	0,00
n		1.739
Pseudo-R²		0,01

Quelle: infas-eigene Berechnung

**Tabelle 49 Ausfallmodellierung Panelstichprobe: Drop-off-Fragebogen in 2011,
Basis: CAPI-Interview in 2008 und CAPI-Interview in 2011
(logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,89	0,00
Geschlecht		
Männer	Referenz	
Frauen	1,34	0,00
Alter		
Bis 49 Jahre	0,75	0,06
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,21	0,11
70 Jahre und älter	1,22	0,07
Bildung (ISCED)		
Niedrig	0,56	0,00
Mittel	Referenz	
Hoch	1,18	0,06
Konstante	3,16	0,00
n		4.183
Pseudo-R²		0,02

Quelle: infas-eigene Berechnung

Tabelle 50 Ausfallmodellierung Panelstichprobe: Drop-off-Fragebogen in 2014, Basis: CAPI-Interview in 2011 und CAPI-Interview in 2014 (logistische Regression)

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,19	0,10
Geschlecht		
Männer	Referenz	
Frauen	1,22	0,06
Alter		
Bis 49 Jahre	0,65	0,05
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,87	0,00
70 Jahre und älter	1,29	0,04
Bildung (ISCED)		
Niedrig	0,45	0,00
Mittel	Referenz	
Hoch	1,06	0,60
Konstante	4,56	0,00
n		3.577
Pseudo-R²		0,02

Quelle: infas-eigene Berechnung

Tabelle 51 Ausfallmodellierung Panelstichprobe: Drop-off-Fragebogen in 2017, Basis: Fälle mit CAPI-Interview in 2014 und CAPI-Interview in 2017 (logistische Regression)

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,02	0,84
Geschlecht		
Männer	Referenz	
Frauen	1,27	0,00
Alter		
Bis 49 Jahre	0,80	0,13
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,67	0,00
70 Jahre und älter	1,63	0,00
Bildung (ISCED)		
Niedrig	0,44	0,00
Mittel	Referenz	
Hoch	1,32	0,00
Konstante	3,52	0,00
n		6.265
Pseudo-R²		0,02

Quelle: infas-eigene Berechnung

**Tabelle 52 Ausfallmodellierung: Drop-off-Fragebogen in 2011,
Basis: Fälle mit CAPI-Interview in 2011 (logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,69	0,00
Geschlecht		
Männer	Referenz	
Frauen	1,26	0,00
Alter		
Bis 49 Jahre	0,77	0,07
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,22	0,06
70 Jahre und älter	1,15	0,16
Bildung (ISCED)		
Niedrig	0,57	0,00
Mittel	Referenz	
Hoch	1,18	0,05
Konstante	3,30	0,00
n		4.854
Pseudo-R²		0,02

Quelle: infas-eigene Berechnung

**Tabelle 53 Ausfallmodellierung: Drop-off-Fragebogen in 2017,
Basis: Fälle mit CAPI-Interview in 2017 (logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,04	0,64
Geschlecht		
Männer	Referenz	
Frauen	1,28	0,00
Alter		
Bis 49 Jahre	0,85	0,26
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,70	0,00
70 Jahre und älter	1,63	0,00
Bildung (ISCED)		
Niedrig	0,43	0,00
Mittel	Referenz	
Hoch	1,33	0,00
Konstante	3,25	0,00
n		6.626
Pseudo-R²		0,02

Quelle: infas-eigene Berechnung

**Tabelle 54 Verteilungsvergleich ungewichtete und kalibrierte Stichprobe:
Querschnittgewicht Interview 2011**

Spalten%	Ungewichtet		Gewichtet	Bekannte Soll-Verteilung*
	abs.	%	%	%
Gesamt	4.834	100,0	100	100
Alter*Geschlecht*Gebiet				
43-54, Mann West	276	5,7	15	15
43-54, Frau West	352	7,3	14	14
43-54, Mann Ost	133	2,8	3	3
43-54, Frau Ost	184	3,8	3	3
55-69, Mann West	630	13,0	13	13
55-69, Frau West	689	14,3	14	14
55-69, Mann Ost	339	7,0	3	3
55-69, Frau Ost	347	7,2	3	3
70-85, Mann West	675	14,0	9	9
70-85, Frau West	500	10,3	12	12
70-85, Mann Ost	312	6,5	2	2
70-85, Frau Ost	283	5,9	3	3
86-90, Mann West	48	1,0	1	1
86-90, Frau West	34	0,7	1	1
86-90, Mann Ost	18	0,4	0	>0
86-90, Frau Ost	14	0,3	0	>0

*Mikrozensus 2011, Zahlen sind ohne Nachkommastellen gerundet, daher Abweichungen zur Summe 100 möglich
Quelle: infas-eigene Berechnung

**Tabelle 55 Verteilungsvergleich ungewichtete und kalibrierte Stichprobe:
Querschnittgewicht Drop-off 2011**

Spalten%	Ungewichtet		Gewichtet	Bekannte Soll-Verteilung*
	abs.	%	%	%
Gesamt	3.993	100,0	100	100
Alter*Geschlecht*Gebiet				
43-54, Mann West	204	5,1	15	15
43-54, Frau West	275	6,9	14	14
43-54, Mann Ost	106	2,7	3	3
43-54, Frau Ost	161	4,0	3	3
55-69, Mann West	504	12,6	13	13
55-69, Frau West	569	14,3	14	14
55-69, Mann Ost	300	7,5	3	3
55-69, Frau Ost	315	7,9	3	3
70-85, Mann West	546	13,7	9	9
70-85, Frau West	404	10,1	12	12
70-85, Mann Ost	269	6,7	2	2
70-85, Frau Ost	255	6,4	3	3
86-90, Mann West	42	1,1	1	1
86-90, Frau West	19	0,5	1	1
86-90, Mann Ost	15	0,4	0	>0
86-90, Frau Ost	9	0,2	0	>0

*Mikrozensus 2011, Zahlen sind ohne Nachkommastellen gerundet, daher Abweichungen zur Summe 100 möglich
Quelle: infas-eigene Berechnung

**Tabelle 56 Verteilungsvergleich ungewichtete und kalibrierte Stichprobe:
Querschnittgewicht Interview 2017**

Spalten%	Ungewichtet		Gewichtet	Bekannte Soll-Verteilung*
	abs.	%	%	%
Gesamt	6.572	100,0	100	100
Alter*Geschlecht*Gebiet				
43-54, Mann West	313	4,8	14	14
43-54, Frau West	363	5,5	14	14
43-54, Mann Ost	136	2,1	3	3
43-54, Frau Ost	152	2,3	3	3
55-69, Mann West	958	14,6	16	16
55-69, Frau West	974	14,8	16	16
55-69, Mann Ost	434	6,6	3	3
55-69, Frau Ost	519	7,9	4	4
70-85, Mann West	893	13,6	9	9
70-85, Frau West	778	11,8	11	11
70-85, Mann Ost	455	6,9	2	2
70-85, Frau Ost	392	6,0	3	3
86-90, Mann West	89	1,4	1	1
86-90, Frau West	57	0,9	1	1
86-90, Mann Ost	28	0,4	0	>0
86-90, Frau Ost	31	0,5	0	>0

*Mikrozensus 2017, Zahlen sind ohne Nachkommastellen gerundet, daher Abweichungen zur Summe 100 möglich
Quelle: infas-eigene Berechnung

**Tabelle 57 Verteilungsvergleich ungewichtete und kalibrierte Stichprobe:
Querschnittgewicht Drop-off 2017**

Spalten%	Ungewichtet		Gewichtet	Bekannte Soll-Verteilung*
	abs.	%	%	%
Gesamt	5.576	100,0	100	100
Alter*Geschlecht*Gebiet				
43-54, Mann West	233	4,2	14	14
43-54, Frau West	291	5,2	14	14
43-54, Mann Ost	110	2,0	3	3
43-54, Frau Ost	132	2,4	3	3
55-69, Mann West	802	14,4	16	16
55-69, Frau West	839	15,1	16	16
55-69, Mann Ost	364	6,5	3	3
55-69, Frau Ost	452	8,1	4	4
70-85, Mann West	784	14,1	9	9
70-85, Frau West	663	11,9	11	11
70-85, Mann Ost	395	7,1	2	2
70-85, Frau Ost	351	6,3	3	3
86-90, Mann West	71	1,3	1	1
86-90, Frau West	43	0,8	1	1
86-90, Mann Ost	23	0,4	0	>0
86-90, Frau Ost	23	0,4	0	>0

*Mikrozensus 2017, Zahlen sind ohne Nachkommastellen gerundet, daher Abweichungen zur Summe 100 möglich
Quelle: infas-eigene Berechnung

Bericht

Ergänzung zum Methodenbericht der DEAS-Erhebungswelle 2017: zusätzliche Gewichtung für die Erhebungsjahre 2002, 2008 und 2014

Autorin: Lena Middendorf

infas Institut für angewandte
Sozialwissenschaft GmbH

Friedrich-Wilhelm-Straße 18
D-53113 Bonn
Tel. +49 (0)228/38 22-0
Fax +49 (0)228/31 00 71
info@infas.de
www.infas.de

Bericht an

Deutsches Zentrum für Altersfragen (DZA)
Manfred-von-Richthofen-Straße 2
12101 Berlin

Vorgelegt von

infas Institut für angewandte
Sozialwissenschaft GmbH
Friedrich-Wilhelm-Straße 18
53113 Bonn

Kontakt

Stefan Schiel
Senior-Projektleiter Sozialforschung

Tel.: 0228/3822-424
Fax: 0228/310071
E-Mail: s.schiel@infas.de

Autorin

Lena Middendorf

© infas Institut für angewandte Sozialwissenschaft GmbH
Der Inhalt dieses Dokuments darf ganz oder teilweise
nur mit unserer schriftlichen Genehmigung veröffentlicht,
vervielfältigt, gedruckt oder in Informations- und
Dokumentationssystemen (information storage and
retrieval systems) gespeichert, verarbeitet oder ausgegeben
werden.

Projekt

5413
Bonn, November 2018
Sl, Rj, MI

infas ist zertifiziert

nach ISO 20252 für die Markt-,
Meinungs- und Sozialforschung

ISO 20252



infas ist Mitglied im
Arbeitskreis Deutscher Markt- und
Sozialforschungsinstitute e.V.
(ADM) und ESOMAR



ESOMAR
member

Vorbemerkung

Die Vorgehensweise bei der Berechnung der Längsschnittgewichte, der Integration der Teilstichproben in die Querschnitte sowie die Berechnung von poststratifizierten Querschnittgewichten für die Erhebungsjahre 2011 und 2017 sind im Methodenbericht zur 6. Welle des DEAS beschrieben.

Nachfolgend wird die Berechnung von poststratifizierten Querschnittgewichten für die Erhebungsjahre 2002, 2008 und 2014 beschrieben. Ebenso die Integration der Ausländerstichprobe 2002 in den Querschnitt und das Querschnittgewicht.

Inhaltsverzeichnis

1	Integration der Ausländerstichprobe	6
2	Poststratifizierte Querschnittsgewichte CAPI-Interview	7
3	Poststratifizierte Querschnittsgewichte Drop-off-Fragebogen	8
4	Kennziffern der Gewichtungsfaktoren	9
	Anhang	10

Tabellenverzeichnis

Tabelle 1	Kennziffern der Gewichtungsfaktoren	9
Tabelle 2	Ausfallmodellierung: Drop-off-Fragebogen in 2002, Basis: Fälle mit CAPI-Interview in 2002 (logistische Regression)	10
Tabelle 3	Ausfallmodellierung: Drop-off-Fragebogen in 2008, Basis: Fälle mit CAPI-Interview in 2008 (logistische Regression)	11
Tabelle 4	Ausfallmodellierung: Drop-off-Fragebogen in 2014, Basis: Fälle mit CAPI-Interview in 2014 (logistische Regression)	12
Tabelle 5	Verteilungsvergleich ungewichtete und kalibrierte Stichprobe: Querschnittgewicht Interview 2002	13
Tabelle 6	Verteilungsvergleich ungewichtete und kalibrierte Stichprobe: Querschnittgewicht Drop-off 2002	14
Tabelle 7	Verteilungsvergleich ungewichtete und kalibrierte Stichprobe: Querschnittgewicht Interview 2008	14
Tabelle 8	Verteilungsvergleich ungewichtete und kalibrierte Stichprobe: Querschnittgewicht Drop-off 2008	15
Tabelle 9	Verteilungsvergleich ungewichtete und kalibrierte Stichprobe: Querschnittgewicht Interview 2014	15
Tabelle 10	Verteilungsvergleich ungewichtete und kalibrierte Stichprobe: Querschnittgewicht Drop-off 2014	16

1 Integration der Ausländerstichprobe

Die im Erhebungsjahr 2002 gezogene separate Ausländerstichprobe wurde bisher nicht in die Gesamtstichprobe integriert. Im Zuge der Querschnittgewichtung wurde dies nun vorgenommen. Da die Fälle in der Ausländerstichprobe ausschließlich im Jahr 2002 befragt wurden, sind sie lediglich für das Querschnittgewicht, nicht aber für die (bereits erstellten) Längsschnittgewichte zu berücksichtigen.

Für die Integration in den Querschnitt erfolgte zunächst eine separate Randanpassung der Ausländerstichprobe. Hierbei wurden die Merkmale Geschlecht, Alter und Landesteil an bekannte Verteilungen in der Grundgesamtheit (Mikrozensus 2002) angepasst. Dies geschah nicht, wie bei den anderen Basisstichproben, durch eine Kombination der Merkmale (Zellen), da die Zellenbesetzung teilweise zu gering ist, sondern anhand der separaten Merkmale.

Die Ausländerstichprobe mit ihrem Gewicht wurde an den bereits existierenden integrierten Querschnitt für das Jahr 2002 angehängt. Dies ist möglich, da die Auswahlwahrscheinlichkeiten nicht überlappend waren, d.h., dass die Befragten entweder in der Ausländerstichprobe oder in der Stichprobe deutscher Staatsbürger, nicht aber in beiden gleichzeitig, ausgewählt werden konnten.

2 Poststratifizierte Querschnittgewichte CAPI-Interview

Analog zu den Jahren 2011 und 2017, für die bereits ein poststratifiziertes Querschnittsgewicht erstellt wurde, geschieht dies für die Jahre 2002, 2008 und 2014. Dabei wird die Verteilung Alter*Geschlecht*Landesteil in der Stichprobe an bekannte Verteilungen in der Grundgesamtheit (Mikrozensus 2002, 2008 und 2014) mittels „iterative proportional fitting“ (IPF) angepasst.¹ Im Erhebungsjahr 2002 wird zusätzlich als weiteres Merkmal die Staatsbürgerschaft (deutsch vs. nicht deutsch) angepasst. In den Erhebungsjahren 2002 und 2008 erhalten Teilnehmer mit einem Alter von 86 Jahren und älter und im Jahr 2014 mit einem Alter von 91 Jahren und älter wegen zu gering besetzter Zellen nach Absprache mit dem DZA kein poststratifiziertes/kalibriertes Querschnittsgewicht.

¹ Ein Vergleich der ungewichteten und gewichteten Stichprobe ist im Anhang dargestellt.

3 Poststratifizierte Querschnittgewichte Drop-off-Fragebogen

Zur Bestimmung der Querschnittgewichte des Drop-offs werden auf Basis aller Teilnehmer am CAPI-Interview Ausfallmodelle (logistische Regression) zur Bestimmung der Wahrscheinlichkeit, den Drop-off-Fragebogen in derselben Welle zu beantworten, gerechnet². Prädiktoren des Modells sind:

- Landesteil: West, Ost.
- Geschlecht: männlich, weiblich.
- Alter gruppiert: bis 49 Jahre, 50 bis 59 Jahre, 60 bis 69 Jahre, 70 Jahre und älter.
- Bildungsabschluss ISCED gruppiert: niedrig, mittel, hoch.

Der Landesteil bezieht sich auf den Zeitpunkt der Ersterhebung, das Geschlecht ist konstant. Das Alter ist das Alter zum Zeitpunkt der Betrachtungswelle.

Das Querschnittgewicht für den Drop-off wurde gebildet, indem das Querschnittgewicht der Betrachtungswelle mit der reziproken Teilnahmewahrscheinlichkeit am Drop-off ($1/\text{Teilnahmewahrscheinlichkeit}$) multipliziert wurde. Daraufhin erfolgte ebenso wie beim poststratifizierten Querschnittgewicht der CAPI-Interviews eine Anpassung der Verteilungen Alter*Geschlecht*Landesteil in der Stichprobe an bekannte Verteilungen in der Grundgesamtheit (Mikrozensus 2011 und 2017) mittels „iterative proportional fitting“ (IPF).³ Im Erhebungsjahr 2002 wurde zusätzlich anhand des Merkmals Staatsbürgerschaft kalibriert.

² Die Ergebnisse der Ausfallmodelle sind im Anhang dargestellt.

³ Ein Verteilungsvergleich der ungewichteten und gewichteten Stichprobe ist im Anhang dargestellt.

4 Kennziffern der Gewichtungsfaktoren

Nachfolgend sind die Kennziffern für die berechneten poststratifizierten Querschnittsgewichte dargestellt.

Tabelle 1 Kennziffern der Gewichtungsfaktoren

Kennwert	Erhebung 2002		Erhebung 2008		Erhebung 2014	
	CAPI-Interview	Drop-off	CAPI-Interview	Drop-off	CAPI-Interview	Drop-off
Mittelwert	1,00	1,00	1,00	1,00	1,00	1,00
Standardabweichung	0,59	0,59	0,59	0,66	0,88	0,94
Minimum	0,14	0,14	0,13	0,11	0,05	0,05
Maximum	5,88	6,27	9,53	7,55	18,04	17,13
Effektivitätsmaß	74,18%	74,03%	74,43%	69,74%	56,35%	53,02%
Effektive Fallzahl	3.829	3.462	6.051	4.214	5.802	4.254
Anzahl gewichtete Fälle	5.162	4.676	8.129	6.043	10.295	8.023

Quelle: infas-eigene Berechnung

Anhang

**Tabelle 2 Ausfallmodellierung: Drop-off-Fragebogen in 2002,
Basis: Fälle mit CAPI-Interview in 2002 (logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,32	0,01
Geschlecht		
Männer	Referenz	
Frauen	1,09	0,37
Alter		
Bis 49 Jahre	0,84	0,20
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,45	0,01
70 Jahre und älter	0,97	0,83
Bildung (ISCED)		
Niedrig	0,53	0,00
Mittel	Referenz	
Hoch	1,22	0,09
Konstante	8,96	0,00
n		5.194
Pseudo-R²		0,02

Quelle: infas-eigene Berechnung

**Tabelle 3 Ausfallmodellierung: Drop-off-Fragebogen in 2008,
Basis: Fälle mit CAPI-Interview in 2008 (logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	1,43	0,00
Geschlecht		
Männer	Referenz	
Frauen	0,96	0,44
Alter		
Bis 49 Jahre	0,74	0,00
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,38	0,00
70 Jahre und älter	1,15	0,04
Bildung (ISCED)		
Niedrig	0,71	0,00
Mittel	Referenz	
Hoch	1,20	0,00
Konstante	2,42	0,00
n		8.196
Pseudo-R²		0,02

Quelle: infas-eigene Berechnung

**Tabelle 4 Ausfallmodellierung: Drop-off-Fragebogen in 2014,
Basis: Fälle mit CAPI-Interview in 2014 (logistische Regression)**

Abhängige Variable: Y=1: Teilnahme, Y=0: Keine Teilnahme	Odds ratio	p-Wert
Landesteil		
West	Referenz	
Ost	0,99	0,79
Geschlecht		
Männer	Referenz	
Frauen	1,40	0,00
Alter		
Bis 49 Jahre	0,73	0,00
50 bis 59 Jahre	Referenz	
60 bis 69 Jahre	1,42	0,00
70 Jahre und älter	1,27	0,00
Bildung (ISCED)		
Niedrig	0,50	0,00
Mittel	Referenz	
Hoch	1,25	0,00
Konstante	2,57	0,00
n		10.324
Pseudo-R²		0,02

Quelle: infas-eigene Berechnung

**Tabelle 5 Verteilungsvergleich ungewichtete und kalibrierte Stichprobe:
Querschnittgewicht Interview 2002**

Spalten%	Ungewichtet		Gewichtet	Bekannte Sollverteilung*
	abs.	%	%	%
Gesamt	5.162	100,0	100	100
Alter*Geschlecht*Gebiet				
40-54, Mann West	634	12,3	17	17
40-54, Frau West	684	13,3	17	17
40-54, Mann Ost	267	5,2	4	4
40-54, Frau Ost	246	4,8	4	4
55-69, Mann West	681	13,2	15	15
55-69, Frau West	592	11,5	15	15
55-69, Mann Ost	313	6,1	3	3
55-69, Frau Ost	295	5,7	4	4
70-85, Mann West	521	10,1	7	7
70-85, Frau West	462	9,0	10	10
70-85, Mann Ost	234	4,5	1	1
70-85, Frau Ost	233	4,5	3	3
Staatsangehörigkeit				
Deutsch	4.576	88,7	94	94
Nicht deutsch	586	11,4	6	6

*Mikrozensus 2002, Zahlen sind ohne Nachkommastellen gerundet, daher Abweichungen zur Summe 100 möglich
Quelle: infas-eigene Berechnung

**Tabelle 6 Verteilungsvergleich ungewichtete und kalibrierte Stichprobe:
Querschnittgewicht Drop-off 2002**

Spalten%	Ungewichtet		Gewichtet	Bekannte Sollverteilung*
	abs.	%		
Gesamt	4.676	100,0	100	100
Alter*Geschlecht*Gebiet				
40-54, Mann West	560	12,0	17	17
40-54, Frau West	606	13,0	17	17
40-54, Mann Ost	242	5,2	4	4
40-54, Frau Ost	228	4,9	4	4
55-69, Mann West	629	13,5	15	15
55-69, Frau West	541	11,6	15	15
55-69, Mann Ost	296	6,3	3	3
55-69, Frau Ost	279	6,0	4	4
70-85, Mann West	467	10,0	7	7
70-85, Frau West	402	8,6	10	10
70-85, Mann Ost	213	4,6	1	1
70-85, Frau Ost	213	4,6	3	3
Staatsangehörigkeit				
Deutsch	4.192	89,7	94	94
Nicht deutsch	484	10,4	6	6

*Mikrozensus 2002, Zahlen sind ohne Nachkommastellen gerundet, daher Abweichungen zur Summe 100 möglich
Quelle: infas-eigene Berechnung

**Tabelle 7 Verteilungsvergleich ungewichtete und kalibrierte Stichprobe:
Querschnittgewicht Interview 2008**

Spalten%	Ungewichtet		Gewichtet	Bekannte Sollverteilung*
	abs.	%		
Gesamt	8.129	100,0	100	100
Alter*Geschlecht*Gebiet				
40-54, Mann West	742	9,1	18	18
40-54, Frau West	869	10,7	18	18
40-54, Mann Ost	376	4,6	4	4
40-54, Frau Ost	441	5,4	4	4
55-69, Mann West	1.014	12,5	13	13
55-69, Frau West	1.004	12,4	14	14
55-69, Mann Ost	517	6,4	3	3
55-69, Frau Ost	517	6,4	3	3
70-85, Mann West	991	12,2	8	8
70-85, Frau West	776	9,6	10	10
70-85, Mann Ost	469	5,8	2	2
70-85, Frau Ost	413	5,1	3	3

*Mikrozensus 2008, Zahlen sind ohne Nachkommastellen gerundet, daher Abweichungen zur Summe 100 möglich
Quelle: infas-eigene Berechnung

**Tabelle 8 Verteilungsvergleich ungewichtete und kalibrierte Stichprobe:
Querschnittgewicht Drop-off 2008**

Spalten%	Ungewichtet		Gewichtet	Bekannte Sollverteilung*
	abs.	%	%	%
Gesamt	6.043	100,0	100,0	100,00
Alter*Geschlecht*Gebiet				
40-54, Mann West	507	8,4	18	18
40-54, Frau West	585	9,7	18	18
40-54, Mann Ost	278	4,6	4	4
40-54, Frau Ost	336	5,6	4	4
55-69, Mann West	759	12,6	13	13
55-69, Frau West	756	12,5	14	14
55-69, Mann Ost	409	6,8	3	3
55-69, Frau Ost	426	7,1	3	3
70-85, Mann West	762	12,6	8	8
70-85, Frau West	514	8,5	10	10
70-85, Mann Ost	387	6,4	2	2
70-85, Frau Ost	324	5,4	3	3

*Mikrozensus 2008, Zahlen sind ohne Nachkommastellen gerundet, daher Abweichungen zur Summe 100 möglich
Quelle: infas-eigene Berechnung

**Tabelle 9 Verteilungsvergleich ungewichtete und kalibrierte Stichprobe:
Querschnittgewicht Interview 2014**

Spalten%	Ungewichtet		Gewichtet	Bekannte Sollverteilung*
	abs.	%	%	%
Gesamt	10.295	100,0	100	100
Alter*Geschlecht*Gebiet				
40-54, Mann West	737	7,2	17	17
40-54, Frau West	892	8,7	17	17
40-54, Mann Ost	353	3,4	3	3
40-54, Frau Ost	384	3,7	3	3
55-69, Mann West	1.364	13,3	13	13
55-69, Frau West	1.416	13,8	14	14
55-69, Mann Ost	603	5,9	3	3
55-69, Frau Ost	719	7,0	3	3
70-85, Mann West	1.340	13,0	9	9
70-85, Frau West	1.076	10,5	11	11
70-85, Mann Ost	674	6,6	2	2
70-85, Frau Ost	574	5,6	3	3
86-90, Mann West	63	0,6	1	1
86-90, Frau West	51	0,5	1	1
86-90, Mann Ost	25	0,2	0	>0
86-90, Frau Ost	24	0,2	0	>0

*Mikrozensus 2014, Zahlen sind ohne Nachkommastellen gerundet, daher Abweichungen zur Summe 100 möglich
Quelle: infas-eigene Berechnung

**Tabelle 10 Verteilungsvergleich ungewichtete und kalibrierte Stichprobe:
Querschnittgewicht Drop-off 2014**

Spalten%	Ungewichtet		Gewichtet	Bekannte Sollverteilung*
	abs.	%	%	%
Gesamt	8.023	100,0	100	100
Alter*Geschlecht*Gebiet				
40-54, Mann West	500	6,2	17	17
40-54, Frau West	671	8,4	17	17
40-54, Mann Ost	244	3,0	3	3
40-54, Frau Ost	300	3,7	3	3
55-69, Mann West	1.063	13,3	13	13
55-69, Frau West	1.169	14,6	14	14
55-69, Mann Ost	469	5,9	3	3
55-69, Frau Ost	591	7,4	3	3
70-85, Mann West	1.044	13,0	9	9
70-85, Frau West	859	10,7	11	11
70-85, Mann Ost	540	6,7	2	2
70-85, Frau Ost	458	5,7	3	3
86-90, Mann West	44	0,6	1	1
86-90, Frau West	35	0,4	1	1
86-90, Mann Ost	21	0,3	0	>0
86-90, Frau Ost	15	0,2	0	>0

*Mikrozensus 2014, Zahlen sind ohne Nachkommastellen gerundet, daher Abweichungen zur Summe 100 möglich
Quelle: infas-eigene Berechnung