

Hohes Alter in Deutschland (D80+): Kurzbeschreibung des SUF D80+, Version 1.0. Stand Oktober 2023

Stuth, Stefan

Veröffentlichungsversion / Published Version

Kurzbericht / abridged report

Empfohlene Zitierung / Suggested Citation:

Stuth, S. (2023). *Hohes Alter in Deutschland (D80+): Kurzbeschreibung des SUF D80+, Version 1.0. Stand Oktober 2023*. Berlin: Deutsches Zentrum für Altersfragen. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-90936-6>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-SA Lizenz (Namensnennung-Weitergabe unter gleichen Bedingungen) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-sa/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-SA Licence (Attribution-ShareAlike). For more information see: <https://creativecommons.org/licenses/by-sa/4.0>

Oktober 2023

Hohes Alter in Deutschland (D80+):
Kurzbeschreibung des SUF D80+, Version 1.0

Stefan Stuth

Zitationsempfehlung: Stuth, Stefan (2023): Hohes Alter in Deutschland:
Kurzbeschreibung des SUF D80+. Berlin: Deutsches Zentrum für Alters-
fragen.

Deutsches Zentrum für Altersfragen (DZA)
Forschungsdatenzentrum (FDZ-DZA)
Manfred-von-Richthofen-Straße 2
12101 Berlin
Telefon +49 (0)30 – 26 07 40-0
Telefax +49 (0)30 – 26 07 40-33
E-Mail fdz@dza.de

INHALTSVERZEICHNIS

1.	Hohes Alter in Deutschland (D80+) – eine Studie zur Hochaltrigkeit in Deutschland	3
1.1	Stichprobe	4
1.2	Besonderheiten der Stichprobe	4
1.3	Unterbringung und Versorgung	5
1.4	Themenbereiche	6
2	Ausschöpfung und Passungsprüfung	6
3	Stellvertreterinterviews (Proxys)	7
4	Datengewichtung der Hauptbefragung 2020/21	7
5	Datenformate	8
6	Systematik der Variablennamen und Missing-Codes	9
7	Veränderung von Daten im SUF zur Wahrung der faktischen Anonymität	10
8	Das komplexe Surveydesign des D80+	10
9	Dokumentation	11
10	Regionaldaten	11

1. HOHES ALTER IN DEUTSCHLAND (D80+) – EINE STUDIE ZUR HOCHALTRIGKEIT IN DEUTSCHLAND

Die Studie Hohes Alter in Deutschland (D80+) ist eine vom Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) geförderte Studie, die vom cologne center for ethics, rights, economics, and social sciences of health (Ceres) in Kooperation mit dem Deutschen Zentrum für Altersfragen (DZA) durchgeführt wurde. Im Mittelpunkt der Studie stehen hochaltrige Menschen in Deutschland die 80 Jahre oder älter sind. Ziel der Studie ist es, das Leben im hohen und sehr hohen Alter systematisch zu erfassen. Die Studie schließt damit eine Lücke in der empirischen Forschung, bei der Hochaltrige als schwer befragbare Gruppe, einer hohen und nicht zufälligen Selektivität unterliegen. Diese Selektivität resultiert in der Regel aus Samplingplänen, die Altersgrenzen festlegen und Gemeinschaftsunterkünfte (z.B. Altersheime) ausschließen.

Um ein repräsentatives Bild der Hochaltrigen und ihrer Lebensverhältnisse zeichnen zu können, wurde eine repräsentative Stichprobe der Wohnbevölkerung ab 80 Jahren erhoben. Diese Stichprobe erfasste sowohl Personen in Privat- als auch in Pflegeheimen. Die beteiligten Personen wurden umfassend zu ihren objektiven und subjektiven Lebensbedingungen befragt. Hinzu kommen Testverfahren zur Erhebung der kognitiven Leistungsfähigkeit. Um auch Hochaltrige in die Befragung zu integrieren, die aus gesundheitlichen Gründen nicht mehr persönlich teilnehmen konnten, wurden Proxy-Interviews geführt. Die kombinierte Befragung Hochaltriger in privaten Haushalten und in Heimen und die Nutzung der Proxy-Interviews bieten eine einzigartige repräsentative Datenquelle, zur interdisziplinären Erforschung der Lebenssituationen der alten als auch sehr alten Menschen in Deutschland.

Die Mikrodaten der Studie Hohes Alter in Deutschland (D80+) stehen der wissenschaftlichen Forschung für nicht-gewerbliche Zwecke kostenfrei zur Verfügung. Das Forschungsdatenzentrum gibt Wissenschaftlerinnen und Wissenschaftlern Zugang zu den im Rahmen des Hochaltrigen surveys erhobenen Informationen und berät sie bei deren Verwendung. Der anonymisierte D80+-Datensatz und die Dokumentationsmaterialien sind kostenlos über das FDZ-DZA (<https://www.dza.de/forschung/fdz>) erhältlich.

Die Studie D80+ baut methodisch und inhaltlich auf die Studie NRW80+ auf. Die Studie Hohes Alter in Deutschland war ursprünglich als persönliches Interview geplant (CAPI). Aufgrund der Corona-Pandemie und der besonderen Gefährdung der Ältesten durch COVID-19, musste das geplante Erhebungsdesign vollständig geändert werden. Das Befragungsprogramm wurde in zwei Blöcke unterteilt: Ein Block erfasst essentielle Informationen, die einfach zu erheben sind (Modul 1). Ein weiterer Block erfasst weiterführende Informationen, die komplexer in ihrer Abfrage sind (Modul2). Modul 1 wurde schriftlich erfasst, Modul 2 etwa ein halbes Jahr später telefonisch. Die Erhebung beider Module fand von Ende 2020 bis 2021 statt. Die Feldarbeit, Datenerfassung und Datengewichtung verantwortete das Institut für angewandte Sozialwissenschaft GmbH (Infas) mit Sitz in Bonn.

1.1 Stichprobe

Die Grundgesamtheit der D80+ Stichprobe basiert auf allen Personen in Deutschland die vor dem 1.3.1940 geboren sind. Die Stichprobe wurde in zwei Stufen gezogen: 1) Ziehung einer bevölkerungsproportionalen Gemeindestichprobe. 2) Ziehung einer disproportional geschichteten Einwohnerstichprobe aus der Gemeindestichprobe (geschichtet nach Alter und Geschlecht mit 6 Schichtungszellen). Für die erste Befragungsphase, die schriftliche Befragung (Modul 1), wurden 40.209 Personen angeschrieben. 10.360 auswertbare Fragebögen kamen zurück. Die zweite Befragungsphase richtete sich an alle Personen der Einsatzstichprobe der schriftlichen Befragung, für die eine Telefonnummer vorlag und die eine Teilnahme nicht explizit verweigert haben. Für Personen die bereits an der schriftlichen Befragung teilgenommen haben, wurde Modul 2 telefonisch abgefragt. In wenigen Fällen wurde bei Unklarheiten im vorliegenden schriftlichen Fragebogen zusätzliche Nachfragen gestellt, um diese aufzuklären. Personen die an der schriftlichen Befragung nicht teilgenommen haben, wurden in der zweiten Phase die Fragen aus dem Modul 1 und dem Modul 2 per CATI gestellt. Insgesamt haben an der Phase zwei, 3.260 Personen teilgenommen. Von diesen 3.260 Personen wurde für 218 Personen auch das Modul 1 im CATI-Modus erhoben (Personen, die nicht an der schriftlichen Erhebung der Phase 1 teilnahmen). Von diesen 218 Personen haben 27 nicht am Modul 2 teilgenommen, so dass insgesamt für 3.233 Personen Angaben zum Modul 2 vorliegen.

Der Datensatz SUF D80+ enthält Informationen zu den Untersuchungsteilnehmerinnen und -teilnehmern aus den folgenden Phasen und Modulen:

Stichprobe D80+	Modul 1	Modul 2
Phase 1 (PAPI)	10.360	-
Phase 2 (CATI)	218	3.233
Gesamt	10.578	3.233

1.2 Besonderheiten der Stichprobe

Die Grundgesamtheit in D80+ kommt ohne drei Beschränkungen aus, die in den meisten anderen Befragungen üblich sind.

1) D80+ verwendet eine Stichprobe ohne eine obere Altersgrenze. Es gibt allerdings eine untere Altersgrenze: Alle Befragten müssen zum Zeitpunkt der Ziehung mindestens 80 Jahre alt gewesen sein bzw. vor dem 1. März 1940 geboren worden sein. Das führt zu der folgenden Verteilung der Befragten über drei Altersgruppen: 80-84-Jährige machen ca. 48 Prozent des D80+-Samples aus, 85-89-Jährige 31 Prozent und Befragte die 90 Jahre oder älter waren, machen ca. 21 Prozent des D80+-Samples aus.

2) D80+ verwendet eine Stichprobe die nicht auf spezifische Wohnformen beschränkt ist. Das heißt, institutionelle Anschriften von Altenpflegeheimen wurden in den Samplingprozess einbezogen. Von allen Befragten im D80+, geben 83 Prozent an, in einer privaten Wohnung bzw. einem privaten Haus zu leben. 5,6 Prozent des Samples leben in einer institutionellen Wohnform. Der Rest der Befragten wohnt in Mischformen, die weder eindeutig privaten noch eindeutig institutionellen Wohnformen zuzuordnen sind.

3) Befragungen können in der Regel nur Personen einbeziehen, die gesundheitlich in der Lage sind teilzunehmen. Insbesondere im hohen Alter, erhöht sich der Anteil von Personen, die wegen Krankheit oder Behinderung keine eigenen Auskünfte mehr geben können. Sie auszuschließen würde heißen, eine Verzerrung mindestens bei gesundheitlichen Merkmalen und Merkmalen des psychischen Wohlbefindens, ins positive in Kauf zu nehmen. Um diese potentiell positive Verzerrung zu vermeiden, sind Stellvertreterinterviews vorgesehen gewesen. (Für mehr Informationen, siehe Kapitel 3.)

1.3 Unterbringung und Versorgung

Das Besondere der D80+ Studie ist, dass erstmalig auch stationär untergebrachte Personen explizit in das Sample integriert wurden. Die Studie erfasst die Art der Unterbringung und differenziert die institutionelle Unterbringung von der privaten Unterbringung. Privates Wohnen ist dabei nicht gleichzusetzen mit einer privaten Mietwohnung. Als private Unterbringungen gelten private Wohnungen, Mehrgenerationenhäuser und ambulant betreutes Wohnen. Altenwohnheime bzw. Wohnanlagen mit betreuten Wohnungen und Seniorenresidenzen gelten als privat, solange die BewohnerInnen nicht vollstationär versorgt werden. Altenheime, Pflegeeinrichtungen gelten als institutionelle Unterbringung. Altenwohnheime bzw. Wohnanlagen mit betreuten Wohnungen und Seniorenresidenzen werden als institutionelles Wohnen eingestuft, wenn die BewohnerInnen vollstationär versorgt werden. Insgesamt wurden 587 Personen als institutionell untergebracht eingestuft.

Die vollstationäre Versorgung erfasst den Sachverhalt, dass umfangreiche Pflegeleistungen kontinuierlich beansprucht werden. Die exakte Definition, die den Befragten vorgegeben wurde, lautet: Als vollstationär versorgt gilt, wer in einem Pflegeheim/Altenheim lebt **und** pflegebedürftig ist **und** 24h am Tag Zugang zu Versorgung, Betreuung und Pflege durch eine Pflegefachkraft hat. Private Pflege, ambulante Pflege oder die Pflege durch einen Pflegedienst gelten **nicht** als vollstationäre Pflege.

Neben den Original-Abfragen in den Variablen steht auch eine generierte Variable zur Verfügung (wohnf1_g). Diese versucht Inkonsistenzen im Antwortverhalten aufzulösen. Gemäß dieser generierten Variable, waren 1.018 Befragte vollstationär versorgt. Dass die Auflösung der Inkonsistenzen nur bedingt erfolgreich war, zeigt die Kreuzung mit der Wohnform: 456 Personen wurden als vollstationär versorgt **und** als in einer privaten Wohnform lebend eingeordnet. Die Quelle der Inkonsistenzen kann nicht abschließend geklärt werden, da der Erhebungsmodus (Selbstaussfüller-Fragebogen) keine Nachfragen gestattet. Denkbar ist, dass vollstationäre

Pflege teilweise auch im privaten Umfeld erbracht wird, dass der Begriff vollstationäre Versorgung durch die Befragten anders verstanden wurde als vom Forschungsteam oder dass der Fragebogen nicht von jedem mit der notwendigen Sorgfalt ausgefüllt wurde.

1.4 Themenbereiche

Beide Module erhoben Auskünfte zu den folgenden Themenbereichen:

- Soziodemographie
- Wohnen
- Familie und Haushalt
- Finanzielle Situation
- Umgang mit Alter
- Gesundheit
- Alltagsgestaltung und Lebensstil
- Techniknutzung
- Wertesystem
- Biografie
- Soziale Eingebundenheit (nur Modul 1)
- Wohlbefinden und Lebenszufriedenheit (nur Modul 1)
- Persönlichkeit (nur Modul 2)
- Kritische Lebensereignisse (nur Modul 2)
- Beeinträchtigung durch Covid19-Pandemie und -Maßnahmen (nur Modul 2)
- Demenz (nur Modul 2)

Manche Themenbereiche werden detaillierter in Modul 1 erfasst (z.B. Soziodemographie und Familie und Haushalt). Andere Themenbereiche werden in Modul 2 detaillierter abgefragt (z.B. Techniknutzung oder finanzielle Situation). Einige Themenbereiche werden nur in einem der beiden Module erfasst. Dazu gehört zum Beispiel die soziale Eingebundenheit oder das Wohlbefinden und die Lebenszufriedenheit, die nur in Modul 1 erfasst wurden. Fragen zur Persönlichkeit und kritischen Lebensereignissen wurden dagegen nur in Modul 2 gestellt. In Modul 2 wurde außerdem im Gesundheitsbereich ein komplexes Befragungsinstrument zur Messung von Demenz eingesetzt.

2 AUSSCHÖPFUNG UND PASSUNGSPRÜFUNG

In Phase 1, der schriftlichen Befragung, wurde aus einer Bruttoeinsatzstichprobe von 40.209 Adressen 10.677 Interviews realisiert. Das entspricht einer Ausschöpfung von fast 27 Prozent. Die Phase 2 (telefonische Befragung), setzte das Vorhandensein einer Telefonnummer voraus. Weiterhin wurden Personen die in der Zwischenzeit verstorben waren oder (weitere) Befragungen ablehnten aus der Bruttostichprobe der Phase 2 ausgeschlossen. Diese Bedingungen erfüllten 6.265 der 10.677 Personen die den Fragebogen in Phase 1 ausgefüllt hatten. Für 37,9

Prozent der 6.265 Personen konnten auswertbare Interviews realisiert werden (N = 2.377). Um die Ausschöpfung unter den Respondenten zu erhöhen, wurden diejenigen unter ihnen angeschrieben, für die bisher keine Telefonnummer vorlag, und mit einem beiliegenden Bar-Incentive von 5 Euro und einem in Aussicht gestelltem weiteren von 20 Euro um die Teilnahme am Telefoninterview gebeten. Auf diese Weise konnten von weiteren 4.289 Teilnehmenden der schriftlichen Befragung mit 703 Telefoninterviews realisiert werden. Von 29.532 Personen die sich nicht an Phase 1 beteiligt haben, erfüllten 8.568 Personen die Bedingungen für die Teilnahme (Telefonnummer, nicht verstorben, keine Verweigerung). Allerdings konnte nur für 218 dieser Fälle Interviews realisiert werden.

In zwei weiteren Prüfschritten musste die Zahl der auswertbaren Fälle reduziert werden. Fälle, die mehr als 50% der Inhalte des Modul 1 nicht beantwortet haben wurden ebenso ausgeschlossen, wie Fälle, bei denen der Verdacht bestand, dass die Interviews unlegitimiert von einer anderen Person, als der Zielperson ausgefüllt wurde. Somit stehen im SUF 10.578 Fälle zur Verfügung die Modul 1 realisiert haben und 3.233 Fälle die Modul 1 und Modul 2 realisiert haben.

3 STELLVERTRETERINTERVIEWS (PROXYS)

Während für den schriftlichen Teil der Befragung (Phase 1) keine Stellvertreter-Interviews zugelassen waren, ist es möglich gewesen, für die telefonische Befragung (Phase 2) eine stellvertretende Person zu benennen. Wiesen bereits Rückmeldungen aus der Phase 1 darauf hin, dass gesundheitliche Gründe einer Teilnahme an der Befragung im Wege stehen, wurde nachgefragt, ob Angehörige der Zielperson kontaktiert werden dürfen. Diese Person sollte in der Lage sein, stellvertretend für die Zielperson Auskunft geben zu können. Dafür kamen Angehörige, Freunde, gesetzliche Vertreter und Pflegekräfte der Zielperson in Frage. Erklärte sich eine stellvertretende Person bereit das Interview zu führen, musste zusätzlich noch die Zustimmung der Zielperson erfolgen. Die Zustimmung konnte am Telefon gegeben werden oder schriftlich mit Hilfe einer extra zugesandten Zustimmungstformulars.

Insgesamt liegen für 363 Personen Angaben von Proxys vor: Für 2 Personen nur in Modul 1 (keine Teilnahme an Modul 2), für 322 nur in Modul 2 (Teilnahme an Modul 1 per Fragebogen – hier waren keine Proxyinterviews zugelassen) und für 39 Personen in Modul 1 und 2 (Teilnahme an Modul 1 per Telefoninterview).

4 DATENGEWICHTUNG DER HAUPTBEFRAGUNG 2020/21

Die Gewichtungsfaktoren wurden für Modul eins und Modul zwei getrennt berechnet. Dafür wurde jeweils ein Designgewicht erstellt. Dieses erfasst die Inverse des Produkts der Gesamtauswahlwahrscheinlichkeit einer Person mit der Wahrscheinlichkeit, dass eine Person in die Bruttostichprobe für den Feldeinsatz gelangt. Das Designgewicht wurde in einem zweiten

Schritt für non-response angepasst. Die dafür verwendeten Ausfallmodelle beinhalteten die folgenden Variablen:

<i>Ausfallmodell Modul 1</i>	<i>Ausfallmodell Modul 2</i>
Bundesland	CATII
Gemeindegrößenklasse	Gemeindegrößenklasse
Heimunterbringung	Heimunterbringung
Telefonnummer liegt vor	Haushaltsgröße
Alter x Geschlecht	Pflegegrad
	Bildung x Geschlecht
	Allgemeine Gesundheitseinschätzung
	Altersgruppe der Designgruppe

Mit Hilfe der Ausfallmodelle wurden selektive Ausfall- bzw. Teilnahmewahrscheinlichkeiten berechnet und als inverse mit dem Designgewicht multiplikativ verknüpft. Die auf diese Weise erstellten Gewichte, wurden abschließend in einem dritten Schritt kalibriert bzw. poststratifiziert. Mit Hilfe von IPF (Iterative Proportional Fitting) wurde die Verteilung der Stichprobe an die bekannte Verteilung der Grundgesamtheit (basierend auf Informationen aus dem Mikrozensus 2019) angepasst. Folgende Informationen aus dem Mikrozensus wurden berücksichtigt: Bundesland, Gemeindegrößenklasse, Altersgruppe, Geschlecht, Haushaltsgröße, Familienstand und Heimunterbringung.

Es liegen insgesamt 4 Gewichtungsfaktoren vor: Die Variablen „gewicht“ und „desgew“ beziehen sich auf Informationen, die im Modul 1 erfasst wurden. Die Variablen cati2desgnorm und cati2kal beziehen sich auf Informationen die in Modul zwei erhoben worden sind. Standardmäßig werden im SUF nur die vollständigen Gewichte (gewicht und cati2kal) ausgeliefert. Wer Interesse an den Designgewichten hat, kann Sie beim FDZ gesondert anfragen.

5 DATENFORMATE

Der Scientific Use File (SUF) der D80+-Studie ist sowohl im SPSS-Format als auch im Stata-Datenformat verfügbar. Sollte es versionsbedingte Probleme mit dem Einlesen der Daten geben, dann wenden Sie sich bitte direkt an die Beratung des FDZ-DZA (fdz@dza.de). Alle Daten und die dazugehörigen Dokumentationen sind in Deutsch und Englisch verfügbar.

Der Stata-Datensatz wird mit zwei verschiedenen Labelsets (deutsch und englisch) ausgeliefert. Um zwischen den beiden Sprachen für Variablen- und Wertelabels hin- und herzuschalten,

verwenden Sie den Stata-Befehl `mlanguage set en` bzw. `mlanguage set de`. Für SPSS sind die deutschen und englischen Labels in zwei distinkten Datensätzen abgelegt.

6 SYSTEMATIK DER VARIABLENNAMEN UND MISSING-CODES

Variablen die aus dem Modul 2 stammen, haben den Suffix `_c2`. Offene Nennungen wurden durch das D80+-Team rekodiert. Variablen mit diesen rekodierten offenen Nennungen, haben ein x im Variablennamen. Die Benennung der Variablen folgte deren Inhalten. Sie haben demnach sprechende Namen, die in der Regel einen groben Einblick in ihren Inhalt gewähren.

Die Vergabe der Missing-Codes lehnt sich an den DEAS an.

Die folgende Tabelle beschreibt die im Datensatz vorkommenden Codes für fehlende Werte.

Wert SPSS	Wert Stata	Label
-1	.a	Verweigert
-2	.b	Weiß nicht
-3	.c	Trifft nicht zu (überfiltert)
-4	.d	Nicht erhoben (Stichprobe) <i>[Fragen in Modul 2, wenn die Befragten nur Modul 1 beantwortet haben]</i>
-6	.f	Angabe fehlt
-7	.g	Gelöscht in Datenaufbereitung
-8	.h	Mehrfachnennung
-9	.i	Nicht möglich (Sehschwierigkeiten)
-10	.j	Nicht möglich (Probleme mit dem Schreiben)
-11	.k	Fehlende Originalwerte [Qualitätssicherung – es lagen nicht ausreichend valide Angaben in den Originalvariablen vor, um die Skala bzw. das Aggregat zu bilden]
-12	.l	Nicht möglich (Handgreifkraft)

Die Missing-Codes wurden im Datensatz als user-defined-system missings formatiert.

7 VERÄNDERUNG VON DATEN IM SUF ZUR WAHRUNG DER FAKTISCHEN ANONYMITÄT

Um die faktische Anonymität der Untersuchungsteilnehmerinnen und -teilnehmer zu wahren, wurden einzelne Variablen nicht in den Scientific Use File (SUF) D80+ aufgenommen und bei einer Reihe von Variablen Vergrößerungen der Ausprägungen vorgenommen.

- Gelöscht wurden unter anderem alle Variablen zum räumlichen Kontext, die unter Umständen eine Bestimmung der Wohngemeinde der Befragten ermöglichen würden. Entfernt wurden auch die Angaben zum Geburtstag und -monat der Befragten.
- Variablen mit offenen Angaben wurden gelöscht. Die offenen Angaben wurden durch das Team der D80+-Studie rekodiert. Die rekodierten Angaben stehen den Datennutzenden zur Verfügung.
- Seltene familiendemografische Merkmale wurden vergrößert (etwa Herkunftsland eines Elternteils).
- Bei einer Reihe von Variablen wurden Werte mit geringer Fallzahl zu Kategorien zusammengefasst (sogenanntes „topcoding“; etwa bei Haushaltsgröße, Anzahl Kinder/Enkel, Jahreszahlen). Diese sind mit einem Label versehen, das angibt welche Wertespanne zusammengelegt wurde.

8 DAS KOMPLEXE SURVEYDESIGN DES D80+

Der D80+ war ursprünglich als persönliche face-to-face Befragung geplant. Aus diesem Grund ist die Stichprobe nicht nur geschichtet, sondern auch geklumpt. Es ist logistisch und kostentechnisch nahezu unmöglich, die gesamte Bundesrepublik abzudecken. Die Reise- und Organisationskosten für die Durchführung der persönlichen Interviews wären unverhältnismäßig hoch. Es wurde ein größen-proportionales Sample an Gemeinden aus allen Bundesländern gezogen. Größen-proportional heißt, dass größere Gemeinden eine höhere Ziehungswahrscheinlichkeit hatten als kleinere Gemeinden. Im Ergebnis wurden 461 Gemeinden gezogen bzw. eingesetzt.

Der D80+ ist auf der zweiten Auswahlstufe (nach der Ziehung der Personen aus den Melderegistern der gezogenen Gemeinden) disproportional geschichtet. Anhand der Kriterien Alter und Geschlecht wurden 6 Stichprobenzellen gebildet.

Wird dieses komplexe Stichprobendesign nicht berücksichtigt (insbesondere die Klumpung), werden die Befragten als voneinander unabhängige Beobachtungseinheiten gewertet. Befragte die jedoch aus den gleichen Gemeinden stammen, können einander aber ähnlich sein (z.B. Befragte aus armen/reichen Gemeinden sind sich in ihrem Einkommen wahrscheinlich sehr ähnlich). Wird die Klumpung nicht berücksichtigt, wird die Varianz unterschätzt und im Ergebnis werden signifikante Effekte/Zusammenhänge gefunden, die bei der korrekten Berücksichtigung des Designs, nicht gefunden werden würden.

Aus diesem Grund werden im D80+ zusätzliche Variablen bereitgestellt, um den statistischen Analyseprogrammen die Komplexität des Surveydesigns mitteilen zu können. In Stata kann dies mit der svy-Suite geschehen. Im Datensatz werden dafür die folgenden Merkmale bereitgestellt: PSU_rand (Primary Sampling Unit), fallnr (Secondary Sampling Unit), und strata (Stichprobenzellen die für die Stratifizierung der SSUs verwendet wurden).

Der svyset Befehl in Stata, mit dem die Surveydesignmerkmale des D80+ definiert werden sollten, sieht folgendermaßen aus:

```
svyset PSU_rand, vce(linearized) singleunit(scaled) || fallnr [pw=gewicht], strata(strata)
```

Das passende Gewicht (gewicht) muss gemäß dem Analysedesign selbständig gewählt oder erstellt werden. Beinhalten die Analysen nur Informationen aus Modul1, sollte das Gewicht „gewicht“ gewählt werden. Sobald Informationen aus dem Modul 2 genutzt werden, sollte das Gewicht „catii2kal“ verwendet und bei svy-Set angegeben werden. Alle folgenden deskriptiven Analysen, sollten dann den Suffix -svy- beinhalten, da dann das komplexe Surveydesign des D80+ berücksichtigt wird, das mit Hilfe des svyset-Befehls definiert wurde. Bei komplexeren Analysen wie Regressionen muss das svy-Kommando nicht zwangsläufig Anwendung finden. Stattdessen sollte die Variable PSU_rand genutzt werden, um geklusterte Standardfehler zu berechnen. Weiterhin sollten die Ausprägungen der Variable strata als Dummies in die Analyse eingehen und der passende Gewichtungsfaktor berücksichtigt werden. In SPSS kann die Einbeziehung des komplexen Stichprobendesigns mittels einer Plandatei erfolgen, die als Clustervariable PSU_rand hat und als Stichprobengewichtungsvariable entweder gewicht (nur Informationen aus Modul 1) oder catii2kal ((auch) Informationen aus Modul 2).

9 DOKUMENTATION

Auf der [Website](#) des FDZ-DZA befinden sich im Bereich der Dokumentation verschiedene Dokumentationsmaterialien zum Download. Die Materialien umfassen das vollständige Erhebungsinstrument, den infas-Methodenbericht, das Codebuch und die ausführliche Skalendokumentation. Diese Dokumente sind zusätzlich in Englisch verfügbar. Die englische Dokumentation wurde aber nur maschinell (DeepL) und ohne zusätzliche Prüfung oder Überarbeitung erstellt.

Ebenfalls auf der [FDZ-DZA-Seite](#) sind alle uns bekannten Publikationen auf Basis des D80+ gelistet, diese sind auch als komplettes PDF-Dokument erhältlich. Wenn Sie eigene Publikationen auf Basis des D80+ für diese Dokumentation melden möchten, dann bitte an fdz@dza.de.

10 REGIONALDATEN

Der SUF des D80+ kann nicht auf regionaler Ebene ausgewertet werden.