

Methodenbericht zur Untersuchung der Reichweite und Wirkung der Silbernetz-Hotline

Oliver Huxhold, Céline Arriagada, Nadiya Kelle, Svenja Spuling & Jenna Wünsche

Deutsches Zentrum für Altersfragen

24. Juni 2026

Einleitung

Der vorliegende Bericht dokumentiert das methodische Vorgehen im Rahmen des Evaluation-Projekts "Reichweite und Wirkung der Silbernetz-Hotline (ReWiSil)". Ziel der Studie ist es, sowohl die Reichweite als auch die Wirkung des Angebots empirisch zu untersuchen.

Unter Reichweite wird dabei verstanden, inwieweit die Nutzerinnen und Nutzer der Silbernetz-Hotline systematisch von der gleichaltrigen Bevölkerung in Deutschland abweichen. Diese Abweichungen werden insbesondere im Hinblick auf soziale Ressourcen Einsamkeit, Altersselbstbilder sowie gesundheitliche und lebenslagen-spezifische Merkmale analysiert. Die Reichweitenanalyse dient somit der Frage, ob

das Angebot gezielt solche Personen erreicht, die im Vergleich zur Bevölkerung in besonderem Maße belastet sind.

Die Wirkungsanalyse zielt darauf ab, Veränderungen in der psychosozialen Situation der Anrufenden nach dem ersten Kontakt mit der Hotline zu untersuchen. Im Mittelpunkt steht hierbei die Frage, ob sich relevante Indikatoren wie Einsamkeit oder wahrgenommene Unterstützung im Zeitverlauf verändern.

Zur Beantwortung dieser Fragestellungen wurde ein Studiendesign gewählt, das sowohl populationsbezogene Vergleiche als auch intraindividuelle Veränderungen ermöglicht.

Studiendesign, Stichprobe und Gewinnung

Um sowohl die Reichweite als auch die Wirkung der Silbernetz-Hotline untersuchen zu können, wurde ein Prä-Post-Design mit zwei Messzeitpunkten gewählt. Dabei werden dieselben Personen zu zwei Zeitpunkten mit einem Abstand von drei Monaten befragt, sodass Veränderungen im Zeitverlauf erfasst werden können.

Ein zentrales Merkmal des Studiendesigns ist die gezielte Auswahl von sogenannten Erstanrufenden, also Personen, die die Silbernetz-Hotline erstmalig nutzen und zuvor keinen Kontakt zu diesem Angebot hatten. Durch dieses Auswahlkriterium wird sichergestellt, dass beobachtete Veränderungen in zeitlichem Zusammenhang zum ersten Kontakt mit dem Angebot stehen. Bei Personen, die das Angebot bereits länger nutzen, könnten mögliche Effekte bereits vor der ersten Befragung eingetreten sein und wären daher nicht mehr eindeutig zuordenbar.

Die Gewinnung der Teilnehmenden erfolgte direkt über die Silbernetz-Hotline.

Alle Telefonistinnen und Telefonisten wurden im Vorfeld geschult und nutzten für die Ansprache potenzieller Teilnehmender einen standardisierten Einladungstext. Dieses Vorgehen gewährleistet eine möglichst einheitliche Gewinnungssituation und reduziert gleichzeitig Selektionsunterschiede, die durch die interviewende Person entstehen können.

Erstanrufende, die an einer Studienteilnahme interessiert waren, hinterließen ihre Kontaktdaten und wurden anschließend von geschulten Interviewerinnen und Interviewern für ein etwa 20-minütiges telefonisches Interview kontaktiert.

Die Erstbefragung (T1) fand im Zeitraum von Juni bis Dezember 2025 statt und schloss unmittelbar an den Erstkontakt an (zwischen Erstkontakt und Interview sollten maximal sieben Tage liegen). Insgesamt nahmen 209 Personen teil. Eine Person wurde aufgrund eines Alters unterhalb der definierten Zielgruppe ausgeschlossen, sodass 208 Personen in die Analysen einbezogen wurden.

Die Zweitbefragung (T2) erfolgte etwa drei Monate nach dem Erstkontakt (September 2025 bis März 2026). An beiden Befragungen nahmen 144 Personen teil.

Zur Auswertung wurden statistische Verfahren eingesetzt, die den Umgang mit fehlenden Werten ermöglichen. Dadurch

konnten alle verfügbaren Informationen der gesamten Stichprobe ($n = 208$) in die Analysen einbezogen werden, auch wenn nicht für alle Personen Daten zum zweiten Zeitpunkt vorlagen. Die Analysen basieren somit auf allen verfügbaren Informationen zu beiden Messzeitpunkten.

Variablenbeschreibungen

Bis auf wenige Ausnahmen, welche an den entsprechenden Stellen näher erläutert werden, wurden alle nachfolgend beschriebenen Items sowohl in der Erst- als auch in der Zweitbefragung erhoben. Der Großteil der verwendeten Items wurde dabei dem Deutschen Alterssurvey (DEAS) aus den Erhebungswellen 2023 und 2024 entnommen (Stuth et al., 2025). Abweichungen hiervon werden an den jeweiligen Stellen explizit ausgewiesen. Der DEAS ist eine bundesweit repräsentative Quer- und Längsschnittbefragung von Personen, die sich in der zweiten Lebenshälfte befinden (Klaus et al., 2017; Vogel et al., 2020) und wird gefördert vom Bundesministerium für Bildung, Familie, Senioren, Frauen und Jugend (BMBFSFJ). Die Darstellung folgt dabei im Wesentlichen der Chronologie des Interviews, mit Ausnahme der soziodemografischen Merkmale, die hier zusammenfassend betrachtet werden.

Soziodemografische Merkmale

Zu Beginn des Telefoninterviews wurden das Alter sowie das Geschlecht der Befragten erfragt, wobei zwischen den Kategorien „männlich“, „weiblich“ und „anderes“ unterschieden wurde. Zudem wurde erfasst, ob die Teilnehmenden in einem Privathaushalt oder einem Altenheim bzw. einer Pflegeeinrichtung leben (Stuth et al., 2025). Die Erfassung des derzeitigen Wohnortes erfolgte mithilfe einer aus dem ALLBUS (GESIS - Leibniz-Institut für Sozialwissenschaften, 2025) übernommenen und hinsichtlich der Antwortkategorien angepassten Frage. Dabei wurde zwischen

Großstadt, Mittel- oder Kleinstadt und Landgemeinde differenziert. Außerdem wurde erhoben, ob die Befragten in West- oder Ostdeutschland leben, wobei Westberlin Westdeutschland und Ostberlin Ostdeutschland zugeordnet wurde (Stuth et al., 2025). Diese Angaben zur Wohnsituation wurden im zweiten Interview nur dann erneut erhoben, wenn die vorgelagerte Filterfrage „Hat sich Ihre Wohnsituation in den letzten drei Monaten verändert?“ bejaht wurde.

Der subjektive Lebensstandard wurde über eine fünfstufige Skala von „sehr gut“ bis „sehr schlecht“ erhoben (Stuth et al., 2025). Zudem wurde die subjektive Einschätzung erfragt, inwieweit der eigene Haushalt mit dem monatlichen Einkommen auskommt, was auf einer vierstufigen Skala von „sehr gut“ bis „sehr schlecht“ beantwortet werden konnte (GESIS - Leibniz-Institut für Sozialwissenschaften, 2022). Ausschließlich im Rahmen der Erstbefragung wurde nach dem höchsten Schul- und Ausbildungsabschluss gefragt, um das Bildungsniveau zu erfassen (Stuth et al., 2025). Diese Angaben bestimmen den Bildungsgrad auf Grundlage einer reduzierten ISCED-Klassifizierung, unterteilt in ein niedriges, mittleres und hohes Bildungsniveau (OECD, 2015). Ebenfalls nur in der Erstbefragung wurde die Staatsangehörigkeit in Anlehnung an den DEAS erhoben (Stuth et al., 2025), wobei die vorgegebenen Antwortkategorien für diese Befragung angepasst wurden; hierzu stan-

den den Befragten die Antwortmöglichkeiten „Deutschland“, „Türkei“, „Polen“ sowie „weitere oder andere Staatsangehörigkeit“ zur Verfügung.

Haushaltszusammensetzung und Familienstand

Im Rahmen der Haushaltszusammensetzung wurde die Haushaltsgröße ermittelt, indem die Anzahl der im Haushalt lebenden Personen inklusive der befragten Person erfragt wurde (Stuth et al., 2025). Ergänzend wurde erfasst, ob es sich bei den weiteren Haushaltsmitgliedern um Ehe-Partner oder Partner, Ehe-Partnerin oder Partnerin, Kinder, (Schwieger-)Elternteile, andere Verwandte, Freundinnen oder Freunde sowie sonstige Personen handelt (Deutsches Zentrum für Altersfragen, 2022).

Ebenso wie die Angaben zum Wohnort wurden diese Fragen zur Haushaltsstruktur im zweiten Interview nur dann erneut abgefragt, wenn sich die Wohnsituation in der Zwischenzeit verändert hatte, das heißt die vorgelagerte Filterfrage „Hat sich Ihre Wohnsituation in den letzten drei Monaten verändert?“ bejaht wurde. Die Teilnehmenden gaben zudem die Anzahl der Kinder an, die bei ihnen aufgewachsen sind oder noch aufwachsen, wobei neben leiblichen Kindern auch Adoptiv- oder Stiefkinder genannt werden konnten; verstorbene Kinder wurden explizit nicht berücksichtigt (Stuth et al., 2025).

Bei der Erhebung des Familienstands wurde zwischen den Kategorien „verheiratet“, „geschieden“, „verwitwet“, „ledig“, „eingetragene Lebenspartnerschaft“, „aufgehobene eingetragene Lebenspartnerschaft“ und „verstorbene eingetragene Lebenspartnerschaft“ unterschieden (Stuth et al., 2025). Der Partnerschaftsstatus wurde über die Frage „Haben Sie einen festen Partner oder eine feste Partnerin?“ mit den Antwortmöglichkeiten „Ja“ oder „Nein“ erhoben (Stuth et al., 2025).

Bewertung der sozialen Beziehungen und Nachbarschaftskontakte

Für eine Einschätzung der sozialen Beziehungen wurden die Befragten anhand von drei einzelnen Items aufgefordert, die Qualität ihrer derzeitigen Partnerschaft, die Beziehung zur Familie sowie die Beziehung zu Freund*innen und Bekannten separat zu bewerten (Stuth et al., 2025). Die Antwortmöglichkeiten reichten hierbei jeweils auf einer fünfstufigen Skala von „sehr gut“ bis „sehr schlecht“. Da der Partnerschaftsstatus im Vorfeld gefiltert wurde, stand bei den Fragen zur Familie sowie zu Freundinnen oder Freunde und Bekannten zusätzlich die Option „habe keine Familienangehörigen mehr“ beziehungsweise „habe keine Freundinnen oder Freunde und Bekannte mehr“ zur Verfügung. Diese Angaben wurden für die Auswertung mit der jeweils schlechtesten Kategorie („sehr schlecht“) gleichgesetzt.

Die Verfügbarkeit von engen Nachbarschaftsbeziehungen wurde über die Frage „Wie eng ist derzeit der Kontakt zu Ihren Nachbarinnen oder Nachbarn?“ erfasst (Stuth et al., 2025). Die fünf vorgegebenen Antwortkategorien lauteten „gar kein Kontakt“, „nur flüchtig“, „weniger eng“, „eng“ und „sehr eng“. Auch hier wurde die zusätzliche Antwortmöglichkeit „habe keinen Kontakt zu Nachbarinnen oder Nachbarn“ in der Analyse mit der Kategorie „gar kein Kontakt“ gleichgesetzt.

Außer Haus verbrachte Zeit

Als objektiver Indikator für soziale Isolation wurde die Dauer der durchschnittlich außer Haus verbrachten Zeit genutzt (Stuth et al., 2025). Die entsprechenden Angaben der Befragten wurden in Stunden und Minuten erfasst.

Altersselbstbilder

Die Einstellungen zum Altern wurden über die Zustimmung zu acht Items erhoben, die Teil einer mehrdimensionalen Skala zu Altersbildern sind (Steверink et al., 2001; Wurm et al., 2007). Die vierstufige Antwortskala reicht von „trifft genau zu“, „trifft eher zu“, „trifft eher nicht zu“ bis hin zu „trifft gar nicht zu“. Vier dieser Items bilden eine Subskala zu sozialem Verlust (beispielsweise „Älterwerden bedeutet für mich, dass ich nicht mehr so recht gebraucht werde“), während die anderen vier Items eine Subskala zur persönlichen Weiterentwicklung abbilden (beispielsweise „Älterwerden bedeutet für mich, dass ich weiterhin viele Ideen realisieren kann“).

Hilfe und Pflege

Der eigene Erhalt von Pflegeleistungen wurde über die Frage ermittelt, ob die Befragten derzeit selbst gesundheitlich so eingeschränkt sind, dass sie deshalb regelmäßig Hilfe, Pflege oder andere Unterstützung benötigen. Eventuelle eigene Unterstützungs- und Pflegeleistungen für andere Personen wurden allgemein über die Frage identifiziert, ob es Personen gibt, die aufgrund ihres schlechten Gesundheitszustandes von den Teilnehmenden privat oder ehrenamtlich betreut beziehungsweise gepflegt werden oder denen regelmäßig Hilfe geleistet wird. Befragte, die diese Frage bejahten, schätzten anschließend auf einer vierstufigen Skala von „gar nicht“ bis „sehr stark“ ein, wie stark sie durch diese geleisteten Hilfen oder Pflegeleistungen belastet sind (Schneekloth & Potthoff, 1993).

Potenzielle Unterstützung

Die potenzielle Unterstützung durch andere Personen wurde durch die Zustimmung zu sechs Items auf einer vierstufigen Skala von „trifft genau zu“ bis „trifft gar nicht zu“ erhoben (Stuth et al., 2025). Die Aussagen lauteten beispielhaft: „Wenn ich unsicher bin, wie ich mich entscheiden soll, dann habe ich jemanden, der mich

beraten kann“ oder „Wenn ich Sorgen habe, dann gibt es jemanden, der mir richtig zuhören kann“.

Einsamkeit und soziale Exklusion

Die Einsamkeit der Befragten wurde mit der Kurzversion einer etablierten Einsamkeitsskala gemessen (De Jong-Gierveld & van Tilburg, 2006). Die Skala umfasst drei positive und drei negative Aussagen, denen auf einer vierstufigen Antwortskala von „trifft genau zu“ bis „trifft gar nicht zu“ zugestimmt oder widersprochen werden konnte. Die Aussagen lauten beispielhaft: „Ich vermisse Leute, bei denen ich mich wohlfühle“ oder „Ich kenne viele Menschen, auf die ich mich wirklich verlassen kann“. Das Instrument ermöglicht eine Differenzierung zwischen sozialer und emotionaler Einsamkeit, wobei drei der Items die emotionale und die anderen drei Items die soziale Einsamkeit erfassen.

Um die Scham im Zusammenhang mit Einsamkeit zu erfassen, wurden für das Projekt ReWiSil zusätzliche, nicht im DEAS enthaltene Aussagen integriert. Die Teilnehmenden reagierten hierbei auf einer ebenfalls vierstufigen Skala von „trifft genau zu“ bis „trifft gar nicht zu“ auf die folgenden Aussagen: „Wenn ich mich einsam fühle, schäme ich mich dafür“, „Wenn ich mich einsam fühle, ist es mir peinlich, es anderen gegenüber zuzugeben“ und „Wenn ich mich einsam fühle, rede ich nicht darüber“ (Barreto et al., 2022; nach eigener Übersetzung).

Die wahrgenommene soziale Exklusion zur Einschätzung der gesellschaftlichen Zugehörigkeit (Bude & Lantermann, 2006) wurde über drei Aussagen erhoben. Die Befragten wählten hierfür auf einer vierstufigen Skala von „trifft genau zu“ bis „trifft gar nicht zu“ aus, wie sehr die Formulierungen auf sie zutreffen oder nicht, wie beispielsweise: „Ich habe das Gefühl, gar nicht richtig zur Gesellschaft zu gehören“.

Selbstwertgefühl

Zur Messung des Selbstwertgefühls wurden fünf Items der Rosenberg-Skala verwendet (Rosenberg, 1965). Die Formulierungen wurden auf einer vierstufigen Skala von „trifft gar nicht zu“ bis „trifft voll und ganz zu“ beantwortet. Beispiele für die Aussagen sind „Hin und wieder denke ich, dass ich gar nichts tauge“ oder „Ich fürchte, es gibt nicht viel, worauf ich stolz sein kann“.

Gesundheit

Für die Erfassung der subjektiven Gesundheit wurden die Befragten aufgefordert, ihren derzeitigen Gesundheitszustand auf einer fünfstufigen Skala von „sehr gut“ über „gut“, „mittel“ und „schlecht“ bis „sehr schlecht“ zu bewerten (Fayers & Sprangers, 2002). Das Vorhandensein lang andauernder chronischer Erkrankungen konnte entweder mit „Ja, eine“, „Ja, mehrere“ bejaht oder mit „Nein“ verneint werden (Stuth et al., 2025). Beide Items sind im Minimum European Health Module (MEHM) enthalten.

Um gesundheitliche Einschränkungen im Alltag festzustellen, wurde für drei spezifische Tätigkeiten erfragt, ob die Teilnehmenden durch ihren derzeitigen Gesundheitszustand „stark eingeschränkt“, „etwas eingeschränkt“ oder „überhaupt nicht eingeschränkt“ waren. Die untersuchten Aktivitäten umfassten das Steigen eines Treppenabsatzes (Bullinger & Kirchberger, 1998), das Herumgehen in der Wohnung sowie die Fortbewegung außerhalb der Wohnung, wie beispielsweise das Autofahren oder das Nutzen öffentlicher Verkehrsmittel (Spector & Fleishman, 1998).

Lebenszufriedenheit

Die allgemeine Lebenszufriedenheit wurde über die Frage „Ganz allgemein gefragt, wie zufrieden sind Sie mit Ihrem Leben insgesamt?“ erfasst (SOEP Group, 2025). Die Befragten gaben ihre Einschätzung auf einer Skala von 0 bis 10 ab, wobei der

Wert 0 für „überhaupt nicht zufrieden“ und der Wert 10 für „vollkommen zufrieden“ stand.

Nutzung und subjektive Bewertung der Silbernetz-Hotline

Ausschließlich im Rahmen der Zweitbefragung wurde über eigens für ReWiSil entwickelte Fragen erhoben, ob die Silbernetz-Hotline zwischen dem Erst- und Zweitinterview genutzt wurde. Hierzu wurden die Befragten gebeten anzugeben, ob sie nach ihrem ersten Anruf erneut das Silbertelefon genutzt hatten. Die Antwortmöglichkeiten lauteten „Ja“ oder „Nein“; diese Angabe sollte sich ausschließlich auf das Silbertelefon beziehen und nicht auf andere Angebote wie Silbernetz-Freundschaft oder Silberinfo. Zudem wurde erfasst, ob die Absicht besteht, das Silbertelefon weiterzuempfehlen sowie die Absicht der eigenen Weiternutzung (jeweils mit den Antwortmöglichkeiten „Ja“ oder „Nein“).

Während die drei Fragen zur Nutzung und Weiterempfehlung des Silbertelefons allen Teilnehmenden der Zweitbefragung gestellt wurden, gingen die folgenden Fragen ausschließlich an Personen, die die Frage nach der erneuten Nutzung des Silbertelefons bejaht haben: Um die subjektive Wirkung der Gespräche zu analysieren, wurde bewertet, ob und inwieweit sich eine Veränderung des Affekts und des Einsamkeitsempfindens nach dem ersten Gespräch eingestellt hatte. Die Teilnehmenden konnten auf einer vierstufigen Skala von „trifft genau zu“ bis „trifft gar nicht zu“ auf die folgenden Aussagen antworten: „Ich fühle mich durch die Gespräche unbeschwerter“, „Ich fühle mich durch die Gespräche akzeptierter“, „Ich fühle mich durch die Gespräche weniger allein“, „Ich habe durch die Gespräche das Gefühl, von jemandem verstanden zu werden“ und „Ich habe durch die Gespräche das Gefühl, dass es jemanden gibt, der mich unterstützen kann“.

Reichweitenanalysen

Bei der Reichweitenanalyse wurde die ReWiSil-Stichprobe mit einer repräsentativen Referenzstichprobe verglichen. Zur Untersuchung der Reichweite wurde ein fünfstufiges analytisches Vorgehen eingesetzt. Zunächst wurde durch ein exaktes Matching auf Alter und Geschlecht die strukturelle Vergleichbarkeit zwischen der ReWiSil- und der jeweiligen Referenzstichprobe hergestellt. Anschließend wurden für die Referenzstichprobe Gewichtungen auf Basis des Mikrozensus berechnet und in einem weiteren Schritt rekaliert, um sowohl populationsbezogene Repräsentativität als auch die durch das Matching vorgegebene Altersstruktur zu berücksichtigen. Darauf aufbauend wurden die Variablen auf eine populationsbezogene T-Skala standardisiert, um eine einheitliche Verteilung mit Mittelwert 50 und Standardabweichung 10 zu erhalten, und die Unterschiede zwischen den Stichproben mithilfe von Multi-Gruppen-Strukturgleichungsmodellen geschätzt. Die Ergebnisse wurden schließlich in Perzentile überführt und grafisch dargestellt.

Dieses Vorgehen ermöglicht eine normreferenzierte Einordnung der ReWiSil-Stichprobe, bei der Unterschiede nicht nur als absolute Abweichungen, sondern als Position innerhalb der Verteilung der gleichaltrigen Bevölkerung interpretiert werden können.

Als Referenzstichprobe diente der Deutsche Alterssurvey (DEAS), eine bundesweit repräsentative Quer- und Längsschnittbefragung von Personen, die sich in der zweiten Lebenshälfte befinden (Klaus et al., 2017; Vogel et al., 2020). Für die Analysen wurde eine Teilstichprobe des DEAS im Alter von 50 bis 93 Jahren ausgewählt, um eine direkte Vergleichbarkeit mit der ReWiSil-Stichprobe zu gewährleisten.

Erster Schritt: Matching

Zur Herstellung einer möglichst hohen Vergleichbarkeit zwischen der ReWiSil-Stichprobe und der jeweiligen Referenzstichprobe aus dem DEAS wurde zunächst ein exaktes Matching durchgeführt (Stuart, 2010). Ziel dieses Schrittes ist es, strukturelle Unterschiede zwischen den Stichproben in zentralen demografischen Merkmalen zu kontrollieren, die andernfalls die Interpretation von Gruppenunterschieden verzerren könnten.

Das Matching erfolgte auf Basis der Variablen Alter und Geschlecht. Diese Merkmale wurden ausgewählt, da sie in engem Zusammenhang mit zentralen Untersuchungsgrößen wie sozialen Ressourcen, Einsamkeit und Gesundheit stehen und somit potenzielle Confounder darstellen.

Es wurde ein one-to-many Matching ohne Zurücklegen mit einem festen Verhältnis von 1:5 realisiert (sogenanntes *fixed ratio matching*; Stuart, 2010). Das bedeutet, dass jeder Person der ReWiSil-Stichprobe fünf Personen aus der jeweiligen DEAS-Stichprobe zugeordnet wurden, die in den Matching-Variablen identisch sind. Keine DEAS-Person wurde dabei mehr als einmal gematcht. Dieses Vorgehen ermöglicht eine effiziente Nutzung der verfügbaren Informationen aus der bevölkerungsbasierten Referenzstichprobe.

Die Wahl eines 1:5-Matching-Verhältnisses verfolgt zwei zentrale Ziele: Zum einen erhöht sie die statistische Präzision der Schätzungen, da mehrere Vergleichspersonen pro Fall berücksichtigt werden. Zum anderen ermöglicht sie eine stabilere Schätzung der Referenzverteilung, da zufällige Schwankungen innerhalb der jeweiligen DEAS-Stichprobe reduziert werden.

Durch das exakte Matching wird sichergestellt, dass sich die ReWiSil- und die jeweilige DEAS-Stichprobe hinsichtlich Alter

und Geschlecht nicht unterscheiden. Unterschiede zwischen den Gruppen können somit nicht auf diese Merkmale zurückgeführt werden, sondern sind auf andere Faktoren zurückzuführen.

Das Matching führt zur Bildung von sogenannten Matching-Clustern, bestehend aus jeweils einer Person der ReWiSil-Stichprobe und den zugeordneten fünf Personen aus der jeweiligen DEAS-Stichprobe. Diese Clusterstruktur wird in den weiteren Analysen berücksichtigt, da Beobachtungen innerhalb eines Clusters nicht unabhängig voneinander sind.

Zweiter Schritt: Gewichtung

Der Deutsche Alterssurvey (DEAS) stellt aufgrund seines Stichprobendesigns sowie selektiver Teilnahmeprozesse keine unverzerrte Abbildung der Bevölkerungsverteilung dar. Insbesondere kommt es zu systematischen Abweichungen in zentralen Merkmalen wie Alter, Geschlecht, Region (Ost- vs. Westdeutschland) und Bildung. Ohne Gewichtung würde die DEAS-Stichprobe daher keine geeignete Referenz für populationsbezogene Vergleiche darstellen.

Aus diesem Grund wurden für die vorliegenden Analysen Gewichtungen der DEAS-Stichproben neu berechnet. Die Neuberechnung der Gewichte erfolgte auf Basis von Verteilungen des Mikrozensus 2022, dem zum Zeitpunkt der Analyse aktuellsten öffentlich verfügbaren Datensatz. Der Mikrozensus basiert auf einer verpflichtenden Teilnahme und gilt daher als besonders zuverlässige und populationsrepräsentative Datenquelle zur Beschreibung der Bevölkerungsstruktur in Deutschland. Die Gewichtung wurde für die Altersgruppe von 50 bis 93 Jahren durchgeführt und berücksichtigte zentrale demografische Merkmale: Alter, Geschlecht, Region (Ost- vs. Westdeutschland) und Bildung. Dadurch wird sichergestellt, dass die jeweilige DEAS-Stichprobe hinsichtlich die-

ser Merkmale an die Bevölkerungsverteilung angepasst ist und somit als populationsnahe Referenz interpretiert werden kann.

Die ReWiSil-Stichprobe wurde hingegen nicht gewichtet, denn das Ziel der Reichweitenanalyse besteht in der Untersuchung der systematischen Abweichungen dieser spezifischen, erreichten Zielgruppe von der gleichaltrigen Bevölkerung. Die Gewichtung wird somit ausschließlich auf die jeweilige Referenzstichprobe angewendet und dient der Sicherstellung ihrer interpretativen Funktion als populationsnahe Vergleichsbasis.

Dritter Schritt: Rekalisierung der Gewichte

Ein zentrales methodisches Problem ergibt sich aus der Kombination von Matching und Gewichtung. Während das Matching sicherstellt, dass die ReWiSil- und die jeweilige DEAS-Stichprobe hinsichtlich Alter und Geschlecht vergleichbar sind, enthalten die auf Basis des Mikrozensus 2022 neu berechneten Gewichte ebenfalls Informationen über die Bevölkerungsverteilung dieser Merkmale, insbesondere über die Altersstruktur.

Würde die gematchte DEAS-Stichprobe direkt mit diesen Gewichten gewichtet, könnte die durch das Matching hergestellte Übereinstimmung in der Altersstruktur wieder beeinträchtigt werden. Die DEAS-Stichprobe würde in diesem Fall zwar stärker an die Bevölkerungsverteilung angenähert, wäre jedoch nicht mehr unmittelbar mit der ReWiSil-Stichprobe vergleichbar.

Um diesen Zielkonflikt zu lösen, wurden die Gewichte der DEAS-Stichprobe rekali-briert. Ziel dieser Rekalisierung war es, die durch das Matching vorgegebene Altersstruktur beizubehalten, da diese einen zentralen Bezugspunkt für die Interpretation der Ergebnisse darstellt und zugleich in den zuvor berechneten Gewichten enthalten ist.

Andere Merkmale, insbesondere Region (Ost- vs. Westdeutschland) und Bildung, wurden weiterhin populationsbezogen gewichtet, sodass die Referenzfunktion der DEAS-Stichprobe erhalten bleibt. Durch dieses Vorgehen wird sichergestellt, dass die gewichtete DEAS-Stichprobe einerseits eine angemessene Annäherung an die Bevölkerungsverteilung darstellt und andererseits die für den Vergleich entscheidende Altersstruktur mit der ReWiSil-Stichprobe übereinstimmt.

Vierter Schritt: T-Standardisierung

Zur Vereinheitlichung der Skalen und zur Verbesserung der Interpretierbarkeit wurden alle untersuchten Variablen sowohl in der ReWiSil-Stichprobe als auch in den jeweiligen Vergleichsstichproben des DEAS auf eine T-Skala transformiert. Die Standardisierung erfolgte auf Basis der gewichteten und gematchten DEAS-Stichproben, die als populationsnahe Referenzen dienten.

Dabei wurde der gewichtete Mittelwert der Referenzstichprobe auf 50 und die Standardabweichung auf 10 gesetzt. Diese Transformation ermöglicht eine direkte Einordnung der Werte relativ zur Verteilung der gleichaltrigen Bevölkerung. Ein T-Wert von 50 entspricht dem Bevölkerungs-

durchschnitt, während höhere oder niedrigere Werte die Position innerhalb dieser Verteilung anzeigen.

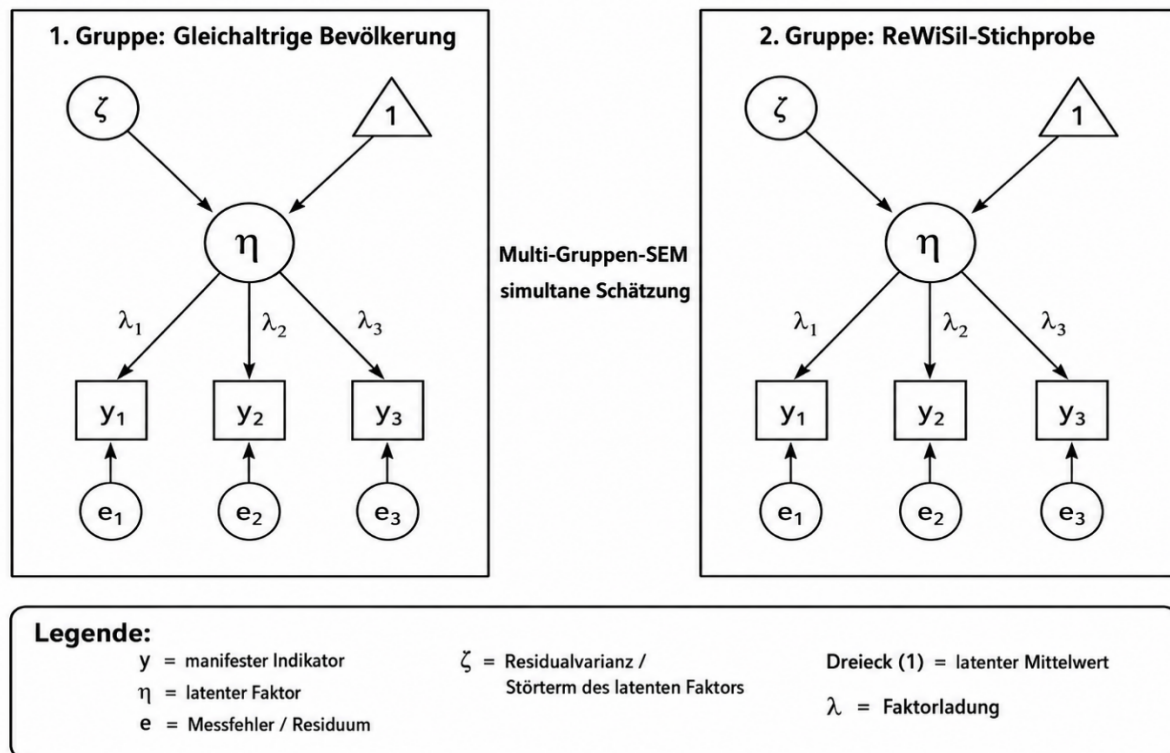
Ein wesentlicher Vorteil dieser Vorgehensweise besteht darin, dass unterschiedliche Variablen auf eine gemeinsame Metrik gebracht werden. Dadurch können Unterschiede zwischen der ReWiSil- und der Referenzstichproben über verschiedene Indikatoren hinweg konsistent interpretiert werden.

Darüber hinaus erleichtert die T-Standardisierung die Ableitung standardisierter Effektgrößen, da Mittelwertsunterschiede direkt in Relation zur populationsbezogenen Standardabweichung interpretiert werden können. Dies ermöglicht eine intuitive und vergleichbare Quantifizierung der Abweichungen von der Bevölkerungsreferenz.

Fünfter Schritt: Multi-Gruppen-Strukturgleichungsmodelle

Die Unterschiede zwischen der ReWiSil-Stichprobe und der Referenzstichprobe wurden mit Hilfe von Multi-Gruppen-Strukturgleichungsmodellen (Vandenberg & Lance, 2000) geschätzt (siehe Abbildung 1). Dieses Verfahren ermöglicht es, Mittelwerte simultan für beide Gruppen zu modellieren und Gruppenunterschiede direkt zu testen.

Abbildung 1: Multigruppen SEM



Quelle: Mit ChatGPT erstellte eigene Darstellung auf Grundlage der RAM-Notation (Reticular Action Model notation; McArdle & McDonald, 1984).

Für Variablen, die als mehrere Items umfassende Skalen vorlagen, wurden latente Faktoren spezifiziert. Der Einsatz latenter Variablen ermöglicht eine Trennung von wahrer Varianz und Messfehler und führt damit zu präziseren Parameterschätzungen. Dadurch können auch kleinere Gruppenunterschiede zuverlässiger erfasst werden.

Für diese Konstrukte wurde zusätzlich die Messäquivalenz (Messinvarianz) zwischen der ReWiSil- und der DEAS-Stichprobe überprüft. Ziel dieser Prüfung ist es sicherzustellen, dass beobachtete Mittelwertsunterschiede tatsächlich Unterschiede im zugrunde liegenden Konstrukt widerspiegeln und nicht auf Unterschiede in der Messstruktur zurückzuführen sind (Vandenberg & Lance, 2000).

Für Variablen, die nur als Einzelitems vorlagen – Gesundheitsindikatoren sowie die Qualität von Familien- und Freundschafts-

beziehungen und Kontakte in der Nachbarschaft – wurden manifeste Variablen direkt in die Modelle integriert. Der Einsatz von Strukturgleichungsmodellen ist auch bei der Kombination von latenten und manifesten Variablen vorteilhaft, da er eine konsistente Schätzung unter Berücksichtigung fehlender Daten, komplexer Stichprobenstrukturen sowie gewichteter Analysen ermöglicht.

Die Schätzung der Modelle erfolgte unter Verwendung von Full Information Maximum Likelihood (FIML). Dieses Verfahren ermöglicht es, auch unvollständige Daten zu berücksichtigen, indem alle verfügbaren Informationen in die Parameterschätzung einbezogen werden. Dadurch konnte die gesamte Stichprobe in die Analysen einbezogen werden, ohne dass Fälle mit fehlenden Werten ausgeschlossen werden mussten.

Zusätzlich wurden Geschlecht, Region, Partnerschaftsstatus und subjektive Gesundheit als Hilfsvariablen in die Modelle aufgenommen. Hilfsvariablen können die Schätzung unter FIML unterstützen, wenn sie mit fehlenden Werten oder mit den untersuchten Variablen zusammenhängen. Dadurch werden Verzerrungen durch selektiven Ausfall reduziert und die verfügbaren Informationen effizienter genutzt.

Zur Berücksichtigung der durch das Matching entstandenen Abhängigkeiten zwischen den Beobachtungen wurden robuste Standardfehler sowie eine Clustering nach Matching-Identifikatoren verwendet. Dies verhindert eine Unterschätzung der Unsicherheit der Parameterschätzungen.

Die Gewichtung wurde entsprechend dem oben beschriebenen Vorgehen berücksichtigt. Während die Fälle der ReWiSil-Stichprobe mit einem Gewicht von 1 in die Analysen eingingen, wurden für die DEAS-Stichprobe die rekalierten Gewichte verwendet.

Der Fokus der Analysen lag auf der Schätzung von Mittelwertsunterschieden zwischen der ReWiSil- und der Referenzstichprobe. Die Streuung der Variablen in der Referenzstichprobe ist durch die T-Standardisierung vorgegeben, während die Streuung der ReWiSil-Stichprobe modellbasiert geschätzt und in der anschließenden Visualisierung berücksichtigt wird. Die Testung von Parameterunterschieden zwischen den beiden Gruppen (ReWiSil und gleichaltrige Bevölkerung) erfolgte auf Basis von Chi-Quadrat-Modelltests bei einem Signifikanzniveau von $\alpha = 0,05$.

Sechster Schritt: Überführung in Perzentile und Visualisierung

Zur anschaulichen Darstellung der Reichweitenanalysen wurden die geschätzten Mittelwerte der ReWiSil-Stichprobe in Perzentile der gleichaltrigen Bevölkerung überführt. Grundlage hierfür war die zuvor vorgenommene T-Standardisierung der

Untersuchungsvariablen. In dieser Standardisierung entspricht der Mittelwert der Referenzstichprobe einem T-Wert von 50 und die Standardabweichung einem Wert von 10. Die Mittelwerte der ReWiSil-Stichprobe wurden ebenfalls in T-Werten ausgedrückt.

Für jeden geschätzten Mittelwert der ReWiSil-Stichprobe wurde anschließend bestimmt, welchem Perzentil dieser Wert innerhalb der populationsbezogenen Referenzverteilung entspricht. Ein Wert am 50. Perzentil zeigt an, dass der Mittelwert der ReWiSil-Stichprobe dem Durchschnitt der gleichaltrigen Bevölkerung entspricht. Ein Wert oberhalb des 50. Perzentsils bedeutet, dass die ReWiSil-Stichprobe im Vergleich zur Referenzbevölkerung höhere Ausprägungen aufweist; ein Wert unterhalb des 50. Perzentsils entsprechend niedrigere Ausprägungen der ReWiSil-Stichprobe im Vergleich zur Referenzbevölkerung. Beispielsweise entspricht ein T-Wert von 60 ungefähr dem 84. Perzentil, ein T-Wert von 70 ungefähr dem 98. Perzentil.

Die Perzentildarstellung erleichtert damit die inhaltliche Einordnung der Ergebnisse. Gruppenunterschiede werden durch dieses Vorgehen nicht nur als Abweichungen auf der T-Skala beschrieben, sondern als relative Position der ReWiSil-Stichprobe innerhalb der Verteilung der gleichaltrigen Bevölkerung. Dadurch wird unmittelbar sichtbar, ob die Silbernetz-Hotline Personen erreicht, deren Werte im Bevölkerungsvergleich durchschnittlich, über- bzw. unterdurchschnittlich oder besonders stark bzw. schwach ausgeprägt sind.

In den grafischen Darstellungen zur Reichweite (siehe Abbildungen 4-7 in der Broschüre) wurden demnach zwei Informationen kombiniert. Erstens wurde der mittlere Perzentilwert der ReWiSil-Stichprobe dargestellt (lila Linie). Zweitens wurde die modellbasiert geschätzte Streuung innerhalb der ReWiSil-Stichprobe berücksichtigt (Schraffur). Hierfür wurde der Bereich von

einer Standardabweichung oberhalb bzw. unterhalb des mittleren Perzentilwerts abgebildet. Diese Darstellung zeigt nicht nur, wo die ReWiSil-Stichprobe im Mittel innerhalb der Bevölkerungsverteilung verortet ist, sondern auch, wie stark die erreichten

Personen hinsichtlich der jeweiligen Merkmale streuen. Damit erfüllt die Visualisierung eine doppelte Funktion: Sie macht die Ergebnisse leichter verständlich und unterstützt zugleich die Interpretation der Reichweite.

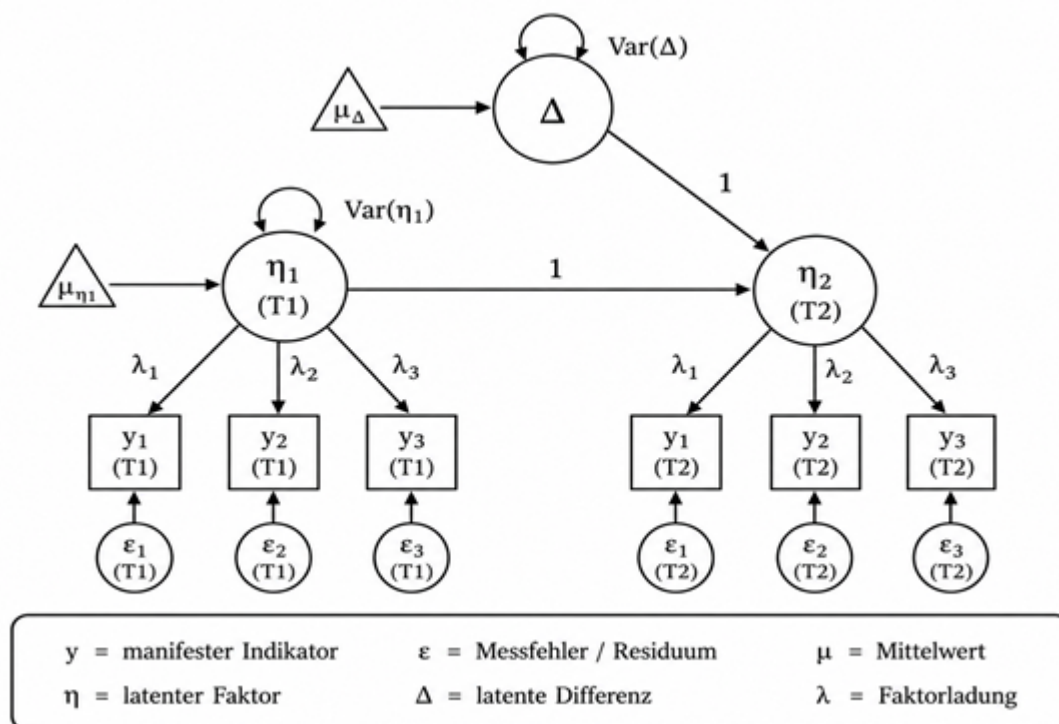
Wirkungsanalysen

Zur Untersuchung der Veränderungen zwischen Erst- und Zweitbefragung wurden Strukturgleichungsmodelle mit zwei Messzeitpunkten spezifiziert (siehe Abbildung 2). Für diese Analysen wurde nur die ReWiSil-Stichprobe untersucht.

Analog zu den Reichweitenanalysen wurden die untersuchten Variablen zunächst

auf eine populationsbezogene T-Skala transformiert. Die Standardisierung erfolgte auf Basis der gewichteten DEAS-Stichprobe, die als Referenz für die gleichaltrige Bevölkerung dient. Der Mittelwert der Referenzverteilung wurde auf 50 und die Standardabweichung auf 10 gesetzt.

Abbildung 2: Latente Veränderungs-messung



Quelle: Mit ChatGPT erstellte eigene Darstellung auf Grundlage der RAM-Notation (Reticular Action Model notation; McArdle & McDonald, 1984).

Durch diese Standardisierung können die Werte der ReWiSil-Stichprobe zu beiden Messzeitpunkten relativ zur Verteilung der gleichaltrigen Bevölkerung interpretiert werden. Veränderungen zwischen Erst- und Zweitbefragung beschreiben damit nicht nur Veränderungen innerhalb der ReWiSil-Stichprobe, sondern zugleich Veränderungen in ihrer Position innerhalb der populationsbezogenen Referenzverteilung. Eine Abnahme belastender Merkmale zeigt demnach eine Annäherung an den Bevölkerungsdurchschnitt an. Eine Zunahme ressourcenbezogener Merkmale, verweist auf eine günstigere Position innerhalb der gleichaltrigen Bevölkerung.

Die Veränderung zwischen den beiden Messzeitpunkten wurde mithilfe von Latent-Difference-Score-Modellen geschätzt (McArdle & Hagami, 2001). In diesen Modellen wird der Wert zum zweiten Messzeitpunkt als Funktion des Ausgangswerts zum ersten Messzeitpunkt und einer Veränderungsvariable modelliert. Diese Veränderungsvariable beschreibt die durchschnittliche Veränderung zwischen Erst- und Zweitbefragung und kann direkt auf statistische Bedeutsamkeit geprüft werden.

Für Konstrukte, die über mehrere Items erfasst wurden, wurden latente Faktoren zu beiden Messzeitpunkten spezifiziert. Die Veränderung wurde dann auf Ebene des latenten Konstrukts modelliert. Dieses Vorgehen hat den Vorteil, dass die Veränderung nicht auf Basis einfacher beobachteter Differenzwerte geschätzt wird. Beobachtete Differenzwerte enthalten Messfehler aus beiden Messzeitpunkten und weisen deshalb häufig eine geringere Reliabilität auf als die zugrunde liegenden Einzelmessungen. Durch die Modellierung latenter Faktoren können Messfehleranteile von den stabilen und veränderlichen Anteilen des Konstrukts getrennt werden. Die geschätzten Veränderungen sind dadurch

um die Messfehler bereinigt und können präziser interpretiert werden.

Für Variablen, die nur durch Einzelitems erfasst wurden, wurden manifeste Veränderungsmodelle geschätzt. Auch hier wurde die Veränderung zwischen Erst- und Zweitbefragung innerhalb eines Strukturgleichungsmodells analysiert. Dadurch konnten latente und manifeste Variablen nach einer einheitlichen Logik ausgewertet werden. Bei manifesten Einzelindikatoren ist allerdings zu berücksichtigen, dass Messfehler nicht in gleicher Weise explizit modelliert werden können wie bei latenten Konstrukten.

Fehlende Werte wurden mithilfe von Full Information Maximum Likelihood (FIML) berücksichtigt (McArdle & Hagami, 2001). Auf diese Weise konnten alle verfügbaren Informationen der ReWiSil-Stichprobe in die Schätzung einbezogen werden, auch wenn nicht für alle Personen Angaben zum zweiten Messzeitpunkt vorlagen. Zusätzlich wurden zentrale Hintergrundmerkmale als Hilfsvariablen in die Modelle aufgenommen, um den Umgang mit fehlenden Werten weiter zu verbessern. Hierzu gehörten Geschlecht, Region, Partnerschaftsstatus und subjektive Gesundheit. Diese Variablen wurden nicht als inhaltliche Prädiktoren der Veränderung interpretiert, sondern dienten ausschließlich dazu, die Schätzung unter FIML zu stabilisieren und mögliche Verzerrungen durch selektiven Ausfall zu reduzieren.

Visualisierung der Veränderungen

Zur Darstellung der Wirkungsanalysen wurden die geschätzten Veränderungen zunächst als standardisierte Veränderungswerte ausgewiesen. Da die Variablen auf Basis der Referenzstichprobe T-standardisiert wurden, können Differenzen zwischen Erst- und Zweitbefragung direkt als Effektgrößen interpretiert werden. Eine Veränderung von 10 T-Wert-Punkten entspricht dabei einer Standardabweichung in der gleichaltrigen Bevölkerung. Entsprechend kann die Veränderung als Cohen's d ausgedrückt werden, indem die geschätzte Veränderung durch 10 geteilt wird. Cohen's d ist ein Effektstärkemaß, das einordnet, wie praktisch bedeutsam ein vorgefundener Effekt ist. Wir haben uns bei der Einordnung der Effekte an der gängigen Klassifikation orientiert: $d < 0,2$: sehr kleiner Effekt, $0,2 \leq d < 0,5$: kleiner Effekt, $0,5 \leq d < 0,8$: mittlerer Effekt, $d \geq 0,8$: starker Effekt.

Zusätzlich wurden die Werte zu beiden Messzeitpunkten in Perzentile der gleichaltrigen Bevölkerung überführt. Dadurch wird sichtbar, an welcher Position innerhalb der populationsbezogenen Referenzverteilung sich die ReWiSil-Stichprobe vor und nach dem Beobachtungszeitraum befindet. Die Perzentildarstellung ergänzt die standardisierten Veränderungswerte, da sie die Veränderung nicht nur als metrische Differenz, sondern als Verschiebung innerhalb der Bevölkerungsverteilung veranschaulicht.

Die Visualisierung in Abbildung 9 in der Broschüre kombiniert damit wieder zwei Perspektiven: Die d-Werte quantifizieren die Größe der Veränderung auf einer standardisierten Skala des Effektstärkemaßes Cohen's d (Pfeile), während die Perzentile zeigen, wie sich die relative Position der ReWiSil-Stichprobe im Vergleich zur gleichaltrigen Bevölkerung verändert (Prozentwerte). Dadurch können die Ergebnisse sowohl statistisch als auch inhaltlich anschaulich interpretiert werden.

Literatur

- Barreto, M., Van Breen, J., Victor, C., Hammond, C., Eccles, A., Richins, M. T., & Qualter, P. (2022). Exploring the nature and variation of the stigma associated with loneliness. *Journal of Social and Personal Relationships*, 39(9), 2658–2679. <https://doi.org/10.1177/02654075221087190>
- Bude, H., & Lantermann, E.-D. (2006). Soziale Exklusion und Exklusionsempfinden. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 58(2), 233–252. <https://doi.org/10.1007/s11575-006-0054-1>
- Bullinger, M., & Kirchberger, I. (1998). *Der SF-36 Fragebogen zum Gesundheitszustand. Handbuch für die deutschsprachige Fragebogenversion*. Göttingen: Hogrefe Verlag.
- De Jong-Gierveld, J., & van Tilburg, T. G. (2006). A 6-item scale for overall, emotional, and social loneliness: Confirmatory tests on survey data. *Research on Aging*, 28(5), 582–598. <https://doi.org/10.1177/0164027506289723>
- Dittmann-Kohli, F., Kohli, M., Künemund, H., Motel, A., Westerhof, G., & infas Sozialforschung. (1997). *Lebenszusammenhänge, Selbst- und Lebenskonzeptionen. Erhebungsdesign und Instrumente des Alters-Survey* [Forschungsbericht 61]. Berlin: Freie Universität Berlin. http://www.fall-berlin.de/lit/FALL_Forschungsbericht_61.pdf
- Fayers, P. M., & Sprangers, M. A. (2002). Understanding self-rated health. *The Lancet*, 359(9302), 187–188. [https://doi.org/10.1016/S0140-6736\(02\)07466-4](https://doi.org/10.1016/S0140-6736(02)07466-4)
- GESIS - Leibniz-Institut für Sozialwissenschaften, Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, Statistisches Bundesamt. (2022). *Datenhandbuch zum Mikrozensus Scientific Use File 2022* [Version 2]. https://www.gesis.org/missy/files/documents/MZ/DHB_2022_v2.pdf
- Klaus, D., Engstler, H., Mahne, K., Wolff, J. K., Simonson, J., Wurm, S., & Tesch-Römer, C. (2017). Cohort profile: The German Ageing Survey (DEAS). *International Journal of Epidemiology*, 46(4), 1105–1105g. <https://doi.org/10.1093/ije/dyw326>
- McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In: *New methods for the analysis of change* (S. 139–175). Washington, DC, US: American Psychological Association.
- OECD, Eurostat, UNESCO Institute for Statistics. (2015). *ISCED 2011 Operational manual: Guidelines for classifying national education programmes and related qualifications*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264228368-en>
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Schneekloth, U., & Potthoff, P. (1993). *Hilfe- und Pflegebedürftige in privaten Haushalten: Bericht zur Repräsentativerhebung im Forschungsprojekt „Möglichkeiten und Grenzen selbstständiger Lebensführung“*. [Bd. 20.2 - Schriftenreihe des BMFuS]. Stuttgart: Kohlhammer.
- Spector, W. D., & Fleishman, J. A. (1998). Combining activities of daily living with instrumental activities of daily living to measure functional disability. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 53(1), S46–S57. <https://doi.org/10.1093/geronb/53B.1.S46>
- Stuth, S., Hameister, N., Schwichtenberg-Hilmert, B., Erdmann-Linge, A., Engstler, H., Lozano Alcántara, A., Luitjens, M., Klaus, D., & Kortmann, L. (2025). German Ageing Survey- Deutscher Alterssurvey (DEAS): Documentation of instruments and variables- Instrumenten-und Variablen-dokumentation 1996–2023.
- GESIS - Leibniz-Institut für Sozialwissenschaften, Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, Statistisches Bundesamt. (2025). *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUScompact 2023* [Version 1.3.0]. <https://doi.org/10.4232/1.14545>
- SOEP Group. (2025). *SOEP-Core– 2023: Person (PAPI, mit Verweis auf Variablen). SOEP Survey Papers 1527: Series A– Survey Instruments (Erhebungsinstrumente)*. Berlin: DIW Berlin/SOEP. Online: https://www.diw.de/documents/publikationen/73/diw_01.c.970917.de/diw_ssp1527.pdf
- Steverink, N., Westerhof, G. J., Bode, C., & Dittmann-Kohli, F. (2001). The personal experience of aging, individual resources, and subjective well-being. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 56(6), P364–P373. <https://doi.org/10.1093/geronb/56.6.P364>

- Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward. *Statistical Science*, 25(1), 1–21. <https://doi.org/10.1214/09-STS313>
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70. <https://doi.org/10.1177/1094428100310>
- Vogel, C., Klaus, D., Wettstein, M., Simonson, J., & Tesch-Römer, C. (2020). German Ageing Survey (DEAS). In: D. Gu & M. Dupre (Hrsg.) *Encyclopedia of gerontology and population aging*. Cham: Springer. https://doi.org/10.1007/978-3-319-69892-2_1115-1
- Wurm, S., Tesch-Römer, C., & Tomasik, M. J. (2007). Longitudinal findings on aging-related cognitions, control beliefs, and health in later life. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 62(3), P156–P164. <https://doi.org/10.1093/geronb/62.3.P156>