

DIE BERLINER ALTERSSTUDIE

Herausgegeben von
Ulman Lindenberger, Jacqui Smith,
Karl Ulrich Mayer und Paul B. Baltes[†]

Wissenschaftliche Redaktion:
Julia A. M. Delius



Akademie Verlag

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN 978-3-05-004508-5

3., erweiterte Auflage

© Akademie Verlag GmbH, Berlin 2010

Der Akademie Verlag ist ein Unternehmen der R. Oldenbourg-Gruppe.

Das eingesetzte Papier ist alterungsbeständig nach DIN / ISO 9706.

Alle Rechte, insbesondere die der Übersetzung in andere Sprachen, vorbehalten. Kein Teil dieses Buches darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Photokopie, Mikroverfilmung oder irgendein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsmaschinen, verwendbare Sprache übertragen oder übersetzt werden.

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers.

Gesamtgestaltung: K. Groß, J. Metzger; Berlin

Titelphoto: K. Groß, Berlin

Satz und Grafiken: Atelier f:50, Berlin

Druck und Bindung: Druckhaus „Thomas Müntzer“ GmbH, Bad Langensalza

Printed in the Federal Republic of Germany

3. Stichprobenselektivität und Generalisierbarkeit der Ergebnisse in der Berliner Altersstudie

Ulman Lindenberger, Reiner Gilberg, Ulrich Pötter,
Todd D. Little & Paul B. Baltes[†]

Zusammenfassung

Die Stichprobe der Berliner Altersstudie beruht auf Zufallsziehungen aus dem Melderegister des Landeseinwohnermeldeamts von Berlin (West). Es werden fünf aufeinanderfolgende Teilnahmeebenen unterschieden, denen jeweils Stichproben abnehmender Größe zugeordnet sind: (1) Informationen des Landeseinwohnermeldeamts (*Ausgangsstichprobe*, $N=1.908$ oder 100%); (2) Kurzform der Ersterhebung ($N=1.264$; d. h. 66% von $N=1.908$); (3) vollständige Ersterhebung ($N=928$; 49%); (4) Einwilligung in die Intensiverhebung ($N=638$; 33%); (5) vollständige Intensiverhebung (*Intensivstichprobe*, $N=516$; 27%). Die Intensivstichprobe bildet das Kernstück der Berliner Altersstudie. Sie ist nach Alter und Geschlecht geschichtet und entspricht einem Zwölf-Zellen-Design mit jeweils 43 Männern und 43 Frauen in sechs verschiedenen Altersgruppen (70–74, 75–79, 80–84, 85–89, 90–94 und 95+ Jahre). Die Schichtung wurde durch zellenspezifisches, an der Größe des Stichprobenausfalls orientiertes Nachziehen von Adressen erreicht (siehe auch Nuthmann & Wahl, Kapitel 2 in diesem Band). Angesichts der unterschiedlichen Größe der Stichproben und des damit einhergehenden *Stichprobenausfalls* stellt sich die Frage nach der *Stichprobenselektivität*, d. h. nach dem Ausmaß, in dem Merkmale, die den Stichprobenausfall vorhersagen, mit untersuchungsrelevanten Merkmalen zusam-

menhängen. Die Selektivitätsanalysen zeigen, dass die Intensivstichprobe die angestrebte Heterogenität behält. Insbesondere gibt es keine Anhaltspunkte dafür, dass sich die in der Intensivstichprobe beobachteten Zusammenhangsmuster und Varianzen von den auf niedrigeren Teilnahmeebenen beobachteten wesentlich unterscheiden. Die Projektion der insgesamt beobachteten Selektivität auf zentrale Konstrukte der Intensiverhebung ergibt, dass die Stichprobe der Intensiverhebung zwar durchweg positiv selektiert ist, die Größe dieser Selektivitätseffekte jedoch bei keinem der Konstrukte eine halbe Standardabweichung überschreitet¹.

1. Einleitung

Sowohl in der sozialwissenschaftlichen wie auch in der naturwissenschaftlichen Forschung ist man daran interessiert, dass die Gültigkeit beobachteter empirischer Regelmäßigkeiten nicht auf die tatsächlich beobachteten Ereignisse beschränkt bleibt, sondern sich auf einen Raum möglicher, aber nicht durchgeführter Beobachtungen verallgemeinern lässt. Die *Repräsentativität der Messung* bezeichnet mithin das Ausmaß, in dem die in einer Untersuchung gemachten Beobachtungen für andere, nicht beobachtete Ereignisse stehen können².

1 Im vorliegenden Beitrag verweisen die Begriffe Teilnahmeebene und Teilnahmetiefe auf die Menge an Informationen, die der Berliner Altersstudie über eine Person zur Verfügung stehen. Aufgrund dieser formalen Definition ist es durchaus möglich, dass Personen, über die nur Informationen auf der niedrigsten Teilnahmeebene vorliegen, zu keinem Zeitpunkt die Absicht besaßen, an der Studie „teilzunehmen“. Eine differenzierte Betrachtung von „non-response“ und Verweigerung findet sich im Beitrag von Nuthmann und Wahl, Kapitel 2.

2 Die Ausgangsstichprobe von BASE basiert auf einer an bestimmten Kriterien orientierten Zufallsziehung von Adressen aus dem Berliner Landeseinwohnermeldeamt. Diese Form der Zufallsziehung weist den großen Vorteil auf, dass systematische, mit den Fragestellungen der Untersuchung verknüpfte Verzerrungen weitaus weniger wahrscheinlich sind als bei nichtzufälligen Formen der Stichprobengewinnung. Jedoch gäbe es auch bei einer Stichprobenausschöpfung von 100% (d. h. bei vollkommener Abwesenheit von Stichprobenausfall) keine Gewähr dafür, dass die zufällig ausgewählten Personen eine perfekte „Miniatur“ der Berliner Bevölkerung über 70 Jahre darstellten (vgl. Kruskal & Mosteller, 1979a, 1979b, 1979c; Rendtel & Pötter, 1992; Rudinger & Wood, 1990). Auf die Frage nach der Generalisierbarkeit der (weiter auf der nächsten Seite)

Das Hauptanliegen des vorliegenden Kapitels besteht darin, eine wichtige Quelle der Gefährdung von Repräsentativität und Generalisierbarkeit zu dokumentieren und zu analysieren: die mit dem Stichprobenausfall unter Umständen verbundene Stichprobenselektivität. Mit *Stichprobenausfall* bezeichnen wir den Sachverhalt, dass nicht alle Personen, die zur Teilnahme an der Berliner Altersstudie (BASE) aufgefordert wurden, auch tatsächlich das gesamte Erhebungsprotokoll der Berliner Altersstudie durchlaufen haben (zur Beschreibung von Rekrutierung, Feldverlauf und Feldsteuerung siehe auch Kapitel 2 von Nuthmann & Wahl). Dieser Stichprobenausfall kann zu einer *Selektivität* (Verzerrung) der Stichprobe führen, falls Teilnehmer sich von Abbrechern in untersuchungsrelevanten Merkmalen unterscheiden (Glynn, Laird & Rubin, 1993; Kessler, Little & Groves, 1995; Little & Rubin, 1987). Diese mit dem Stichprobenausfall unter Umständen verbundene Selektivität sollte nicht mit Selektivität verwechselt werden, die mit der *Ziehung* der Stichprobe verbunden sein kann („selective sampling“ versus „selective dropout“; vgl. P. B. Baltes, Reese & Nesselroade, 1988). Hinsichtlich der Ziehung ist die in der BASE verwendete Methode der Zufallsziehung aus einer Grundgesamtheit, bei der alle Beobachtungseinheiten die gleiche Wahrscheinlichkeit aufweisen, gezogen zu werden, nach einhelliger Auffassung am besten dazu geeignet, systematische Verzerrungen zu minimieren (Kruskal & Mosteller, 1979a, 1979b, 1979c).

Unabhängig von dem mit der Stichprobenziehung verknüpften Ausmaß an Generalisierbarkeit – das im Fall von BASE als hoch anzusehen ist – besteht das Problem des *Stichprobenausfalls* also darin, dass sich die an den reduzierten Stichproben getroffenen Aussagen nicht mehr ohne weiteres auf die Ausgangsstichprobe übertragen lassen und die Generalisierbarkeit der BASE-Ergebnisse aufgrund dieser der Ziehung nachgelagerten Tatsache mit einschränkenden Kommentaren versehen werden müsste. Die Analyse der Stichprobenselektivität ist geboten, wenn der Stichprobenausfall ein gewisses Maß überschreitet. Ihre Hauptaufgabe besteht darin, Art und Ausmaß der Stichprobenselektivität zu bestimmen und somit einen Beitrag zur angemessenen Interpretation der am Datensatz der Berliner Altersstudie getroffenen Aussagen zu leisten. Darüber hinaus kann die Untersuchung der Stichprobenselektivität

inhaltlich bedeutsame Beiträge zum Verständnis von Morbidität und Mortalität im hohen Alter leisten (vgl. Kruse, Lindenberger & Baltes, 1993). Zum Beispiel ist zu erwarten, dass Personen mit demenziellen Syndromen erhöhte Sterbewahrscheinlichkeiten und erhöhte Ausfallwahrscheinlichkeiten aufweisen.

Im folgenden werden wir zunächst den Begriff der Selektivität näher erläutern. Danach werden wir die Teilnahmeebenen der Berliner Altersstudie sowie die in den Analysen verwendeten statistischen Verfahren und Variablen im Überblick darstellen. Anschließend berichten wir über die Ergebnisse der Selektivitätsanalysen und diskutieren die Frage, in welchem Ausmaß und in welcher Weise die auf der Ebene der BASE-Intensiverhebung gemachten Beobachtungen von der Stichprobenselektivität beeinflusst worden sind.

2. Fragestellung

2.1 Zur Untersuchung von Stichprobenselektivität

In einer Untersuchung wie BASE werden Aussagen über die Zuordnung von Merkmalen zu Personen und Personengruppen getroffen. Diese Aussagen lassen sich, in Abhängigkeit von der Fragestellung, in den meisten Fällen mit Hilfe von statistischen Kennwerten wie Mittelwerten, Häufigkeitsverteilungen (Prävalenzraten), Varianzen und Korrelationen angemessen beschreiben. Einige Beispiele mögen dies veranschaulichen: (a) Wie groß ist das soziale Netzwerk von Frauen und Männern? (Mittelwert); (b) Wie hoch ist der Anteil Dementer bei 95jährigen und älteren Personen? (Prävalenz); (c) Wie groß sind die individuellen Unterschiede in der geistigen (kognitiven) Leistungsfähigkeit bei 70- bis 80jährigen Personen? (Varianz); (d) Wie eng sind sensorische und kognitive Funktionen im hohen Alter miteinander verknüpft? (Korrelation).

Der Geltungsbereich derartiger Aussagen ist zunächst auf jene Personen begrenzt, bei denen die in Frage stehenden Merkmale auch tatsächlich erhoben wurden. Da jedoch nicht alle angeschriebenen Personen die gesamte Untersuchungssequenz der Berliner Altersstudie durchlaufen haben, ergibt sich die Frage, ob man bei einer perfekten Ausschöpfung der Stichprobe zu anderen Aussagen gelangen würde.

am Datensatz der Berliner Altersstudie getroffenen Aussagen kann es somit keine pauschale Antwort geben. Vielmehr variiert die Generalisierbarkeit der Ergebnisse je nach Inhaltsbereich, Aussagentyp und gegenwärtigem Kenntnisstand und wird sich bisweilen auf eine eher kleine (z. B. die Berliner Wohnbevölkerung über 70) oder auf eine eher große Grundgesamtheit (z. B. ältere Menschen in westlichen Industriegesellschaften) beziehen (siehe auch P. B. Baltes et al., Kapitel 1).

Dies wäre vor allem dann der Fall, wenn jene personenbezogenen Merkmale, die den Verbleib bzw. das vorzeitige Ausscheiden aus der Studie vorhersagen, einen systematischen Zusammenhang mit den zu untersuchenden Merkmalen aufweisen würden (Little & Rubin, 1987). So wäre es z. B. möglich, dass Demente eine geringere Wahrscheinlichkeit aufweisen, jenen Teil des Erhebungsprotokolls der Berliner Altersstudie zu erreichen, in dem die klinische Demenzdiagnose vorgenommen wird. In diesem Fall könnten Aussagen zur Demenzprävalenz die Prävalenz in der Ausgangsstichprobe unterschätzen (siehe auch Helmchen et al., Kapitel 8).

Die Untersuchung der Stichprobenselektivität ist demnach eine methodische Vorsichtsmaßnahme, mit der die Wahrscheinlichkeit von falschen Schlussfolgerungen und irreführenden Generalisierungen verringert werden soll. Überschätzen wir das Bildungsniveau, weil Personen mit niedrigem Bildungsniveau weniger häufig teilnehmen als Personen mit hoher Schulbildung? Unterschätzen wir die Variabilität kognitiver Leistungen, weil sowohl sehr leistungsstarke als auch demente Personen weniger häufig teilnehmen als Personen mit durchschnittlichem Leistungsniveau?

Die Analyse der Stichprobenselektivität steht nun vor einem grundlegenden Problem: Um Fragen nach der Nichtzufälligkeit des Stichprobenausfalls perfekt zu beantworten und das Ausmaß der Selektivität präzise zu quantifizieren, bräuchte man genau jene Informationen, die einem fehlen. Man müsste die Merkmalsausprägungen der Nichtteilnehmer kennen – aber dann wären sie keine Nichtteilnehmer mehr. Diese paradoxe Situation hat in der einschlägigen methodologischen Literatur zu der Forderung geführt, von allen Personen, d. h. auch von späteren Nichtteilnehmern, zumindest einige zentrale Angaben oder Basisinformationen zu erhalten (Dalenius, 1988; Esser, Grohmann, Müller & Schäffer, 1989; von Eye, 1989; Herzog & Rogers, 1988; Oh & Scheuren, 1983; Panel on Incomplete Data, 1983; Tennstedt, Dettling & McKinlay, 1992; Weaver, Holmes & Glenn, 1975). Nur dann ist eine immanente empirische Untersuchung der Selektivität – d. h. ohne Bezug auf externe Datenquellen wie z. B. Zensusinformationen – überhaupt möglich.

Für eine zeitlich ausgedehnte Untersuchung wie BASE erweist es sich deshalb als sinnvoll, die Dichotomie Teilnahme versus Nichtteilnahme durch das mehrfach gestufte Konzept der „Teilnahmeebenen“ oder der *Teilnahmetiefe* zu ersetzen. Das Untersuchungsdesign der Berliner Altersstudie war von vornherein auf eine derartige Staffelung der Teilnahmetiefe angelegt (vgl. P. B. Baltes et al., Kapitel 1;

Nuthmann & Wahl, Kapitel 2). Für Selektivitätsanalysen ergibt sich daraus der entscheidende Vorteil, dass bei jedem Übergang von einer Teilnahmeebene zur nächsten die Gruppe der weiterhin Teilnehmenden mit der Gruppe der Teilnahme beendenden Personen auf den bereits gemessenen Merkmalen verglichen werden kann. Auf diese Weise können in einem ersten Schritt Merkmale identifiziert werden, auf denen sich die beiden Gruppen unterscheiden, und es kann in einem zweiten Schritt der Versuch unternommen werden, Schätzwerte für später erhobene Konstrukte zu berechnen, die die Auswirkungen der Stichprobenselektivität berücksichtigen.

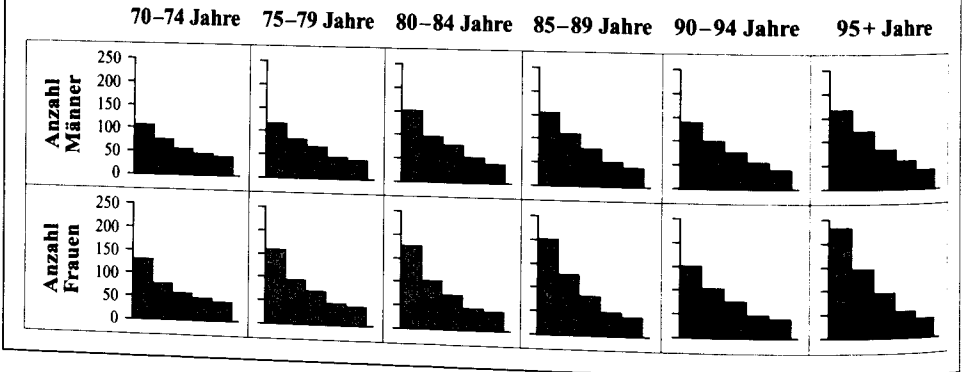
2.2 Definition der Teilnahmeebenen

Wie bereits an anderer Stelle dargestellt, basiert die Ausgangsstichprobe der Berliner Altersstudie auf Adressen, die vom Berliner Landeseinwohnermeldeamt nach dem Zufall ausgewählt und zur Verfügung gestellt wurden. Zusätzlich ist die Stichprobe auf der für die meisten Auswertungen maßgeblichen höchsten Teilnahmeebene, der Intensiverhebung (N=516), nach Alter und Geschlecht geschichtet, d. h., die Ausprägungen der Variablen Geschlecht und Alter sind jeweils gleichverteilt. Somit befinden sich auf der Ebene der Intensiverhebung in sechs aufeinanderfolgenden Altersgruppen (70–74, 75–79, 80–84, 85–89, 90–94 und 95+ Jahre) je 43 Männer und 43 Frauen.

Die Schichtung der Stichprobe auf der Ebene der Intensiverhebung weist gegenüber einer Zufallsverteilung den Vorteil auf, dass Altersunterschiede über den gesamten Altersbereich und in beiden Geschlechtern mit der gleichen Zuverlässigkeit erfasst werden können. Ein Vergleich mit den Erwartungswerten für eine nicht geschichtete Zufallsstichprobe der Berliner Wohnbevölkerung über 70 Jahre mag dies veranschaulichen: Hätte man die Stichprobe nicht nach Alter und Geschlecht geschichtet, sondern statt dessen 516 70jährige und ältere Personen nach dem Zufall aus der Westberliner Wohnbevölkerung gezogen, so hätten sich in einer derartigen Stichprobe zirka 94 70- bis 74jährige Frauen, 42 70- bis 74jährige Männer, 5 Frauen über 94 und lediglich 1 Mann über 94 befunden.

Die Gleichverteilung über Alter und Geschlecht auf der Ebene der Intensiverhebung wurde erreicht, indem für jede der zwölf Zellen des Stichprobensplans eine an den unterschiedlichen Ausschöpfungsraten orientierte Zahl von Adressen eingesetzt wurde. Abbildung 1 veranschaulicht diesen Sachverhalt. In

Abbildung 1: Stichprobenausfall als Funktion von Alter und Geschlecht. Die Säulen innerhalb der zwölf Zellen entsprechen den fünf Teilnahmeebenen: (1) Einwohnermeldeamt ($N=1.908$); (2) Kurzform der Ersterhebung ($N=1.264$); (3) vollständige Ersterhebung ($N=928$); (4) Einwilligung in die Intensiverhebung ($N=638$); (5) vollständige Intensiverhebung ($N=516$). Auf der Ebene der Intensiverhebung befinden sich genau 43 Personen in jeder der zwölf Zellen.



der Abbildung werden fünf verschiedene Teilnahmeebenen unterschieden. *Teilnahmeebene 1* besteht aus allen Personen, an die ein Schreiben verschickt wurde, in dem sie zur Teilnahme an der Studie aufgefordert wurden, abzüglich derjenigen Personen, die sich erst nach Versenden des Anschreibens aus verschiedenen Gründen als nicht zur Grundgesamtheit zugehörig erwiesen ($N=389$; siehe auch Nuthmann & Wahl, Kapitel 2). Die auf dieser Ebene anzutreffende Stichprobe bezeichnen wir als die *Ausgangsstichprobe* ($N=1.908$). Der Ausschöpfungsgrad dieser Stichprobe, die der „erreichten Stichprobe“ im Beitrag von Nuthmann und Wahl (Kapitel 2) entspricht, ist definitionsgemäß 100% (vgl. Fußnote 1).

Teilnahmeebene 2 ($N=1.264$ oder 66% der verifizierten Ausgangsstichprobe von $N=1.908$) bezieht sich auf alle Personen, mit denen die Kurzbefragung durchgeführt wurde. Dies sind zum einen Personen, die alle der ersten 16 Fragen der Ersterhebung beantwortet haben ($N=1.219$; vgl. Nuthmann & Wahl, Kapitel 2), sowie weitere 45 Teilnehmer, für die, zum Teil aufgrund von Angehörigen-Interviews, zumindest Informationen zur Mehrzahl der ersten 16 Fragen der Ersterhebung vorliegen. Voraussetzung für das Erreichen von *Teilnahmeebene 3* ($N=928$; 49% der verifizierten Ausgangsstichprobe von $N=1.908$) ist eine vollständige Ersterhebung (ein im Rahmen von BASE entwickeltes Erhebungsinstrument zu zentralen Konstrukten aller beteiligten Disziplinen; siehe auch P. B. Baltes et al., Kapitel 1; Nuthmann & Wahl, Kapitel 2).

Die vierte und fünfte Teilnahmeebene unterscheiden sich lediglich hinsichtlich der Informationen, die aus dem Intensivprotokoll vorliegen. Alle Personen, die ihre Bereitschaft bekundeten, an der Intensiverhebung teilzunehmen, erreichten *Teilnahmeebene 4* ($N=638$; 33% von $N=1.908$). Manche dieser Personen nahmen aber nicht vollständig an der Intensiverhebung teil, und nur jene Personen, die die gesamte Erhebung durchliefen und für die ein vollständiges Intensivprotokoll vorliegt, befinden sich auf *Teilnahmeebene 5* ($N=516$; 27% von $N=1.908$). Für diese Personen liegt ein außerordentlich umfangreicher multidisziplinärer Datensatz vor (vgl. P. B. Baltes et al., Kapitel 1, für einen Überblick über die erfassten Konstrukte). *Teilnahmeebene 5 bildet das Kernstück der Berliner Altersstudie*. Die Größe des Stichprobenausfalls auf dieser Ebene unterstreicht die Notwendigkeit von Selektivitätsanalysen.

Ein wesentliches Ziel der BASE-Feldsteuerung bestand darin, die Stratifizierung nach Alter und Geschlecht auf Teilnahmeebene 5 zu gewährleisten. Wie aus Abbildung 1 ersichtlich ist, wurde dieses Ziel auch erreicht, denn in jeder der zwölf Zellen des Designs befinden sich 43 Personen auf der fünften Teilnahmeebene. Abweichungen von der Gleichverteilung auf niedrigeren Teilnahmeebenen bringen zum Ausdruck, dass in den verschiedenen Designzellen eine unterschiedlich große Anzahl von Adressen eingesetzt werden musste, um im Ergebnis eine Stratifizierung auf der höchsten Teilnahmeebene zu erreichen. So befinden sich z. B. auf Teilnahmeebene 1

in der Zelle der 95jährigen und älteren Frauen 235 Personen, in der Zelle der 70- bis 74jährigen Männer jedoch nur 109 Personen. Die Gleichverteilung auf der Ebene vollständig abgeschlossener Intensivprotokolle wurde also durch eine an unterschiedlichen Ausschöpfungsraten orientierte Feldsteuerung ermöglicht³.

Die Tatsache eines nach Alter und Geschlecht in seiner Größe variierenden Stichprobenausfalls wirft die Frage auf, ob das Ausmaß der Stichprobenselektivität ebenfalls mit dem Alter und dem Geschlecht zusammenhängt. Zum Beispiel könnte die Stichprobenselektivität bei Hochbetagten größer als bei jungen Alten und bei Frauen größer als bei Männern sein. Aus dieser Überlegung ergibt sich die methodische Anforderung, die Designvariablen Alter und Geschlecht in den Selektivitätsanalysen systematisch zu berücksichtigen.

2.3 Zusammenfassung

Der Geltungsbereich der auf der Ebene der Intensiverhebung getroffenen Aussagen ist zunächst auf jene Personen begrenzt, die das Intensivprotokoll auch tatsächlich durchlaufen haben. Es stellt sich somit die Frage, ob die an diesen Personen gemachten Beobachtungen auch für die Ausgangsstichprobe gültig sind. Die Gültigkeit wäre in dem Maße eingeschränkt, in dem beobachtete oder unbeobachtete personenbezogene Merkmale, die das vorzeitige Ausscheiden aus der Studie vorhersagen, mit den Untersuchungsgegenständen der Intensiverhebung zusammenhängen. Die Tatsache, dass für 73% der Ausgangsstichprobe kein vollständiger Datensatz aus der Intensiverhebung vorliegt, ist für sich genommen noch kein Beweis für das Vorhandensein von *Selektivität*, sondern dokumentiert lediglich den Stichprobenausfall. Von Selektivität kann erst dann die Rede sein, wenn der Ausfall im soeben definierten Sinne nicht zufällig erfolgte. Die im folgenden dargestellten statistischen Verfahren konzentrieren sich auf diese Fragestellung.

3. Methode

Im folgenden geben wir zunächst eine Übersicht über die in den Selektivitätsanalysen verwendeten Variablen (Abschnitt 3.1). Anschließend werden aufgrund früherer gerontologischer Untersuchungen Erwartungen über mögliche Auswirkungen der Stichprobenselektivität zusammengefasst (Abschnitt 3.2). Schließlich geben wir einen formalen Überblick über die zur Analyse der Stichprobenselektivität verwendeten statistischen Verfahren (Abschnitt 3.3).

3.1 Variablenauswahl

Die in den folgenden Analysen verwendeten Variablen werden in Tabelle 1 aufgelistet. Bei der Auswahl der Variablen stand das Bemühen im Vordergrund, das etwaige Vorhandensein von Selektivität möglichst früh (d. h. bereits auf niedrigen Teilnahmeebenen) und möglichst gut zu erfassen.

Auf Teilnahmeebene 1 wurden alle verfügbaren Informationen in die Analysen aufgenommen. Wie aus Tabelle 1 ersichtlich wird, handelt es sich hierbei um Auskünfte des Landeseinwohnermeldeamtes, um die Einschätzung der Wohngegend auf der Grundlage eines von anderen Autoren für das Stadtgebiet von Berlin (West) entwickelten Indexes (Meinlschmidt, Imme & Kramer, 1990) sowie um eine durch die Interviewer erfolgte Wohnquartiersbeschreibung, die die Unterscheidung zulässt, ob Studienteilnehmer selbstständig oder in einer Einrichtung der stationären Altenhilfe leben (Privathaushalt oder Seniorenwohnung versus Seniorenheim, Krankenhaus, Pflegeheim oder Krankenhaus für chronisch Kranke).

Besonders erwähnt werden sollte die Variable Ein-Jahres-Sterblichkeit. Sie basiert auf Mortalitätsinformationen, die nachträglich vom Landeseinwohnermeldeamt zur Verfügung gestellt wurden. Die Ein-Jahres-Sterblichkeit gibt Auskunft darüber, ob eine Person ein Jahr nach Versenden des Anschreibens mit der Aufforderung zur Teilnahme an der Studie noch am Leben war.

Auf den höheren Teilnahmeebenen werden diese Informationen durch eine Reihe weiterer Merkmale ergänzt. Bei der Auswahl dieser Variablen stand das

³ Das an Ausschöpfungsraten orientierte Nachziehen von Adressen ist aus stichprobentheoretischer Sicht problematisch, da die einzelnen Beobachtungen nicht länger als stochastisch unabhängig angesehen werden können. Eine derartige Abhängigkeit der Beobachtungen bedürfte im Rahmen der hier berichteten Selektivitätsanalysen vor allem dann näherer Betrachtung, falls sich später rekrutierte von früher rekrutierten Personen auf den beobachteten Merkmalen unterscheiden würden. Um diese Frage zu untersuchen, wurden die in Tabelle 1 (siehe unten) aufgeführten Variablen mit dem Datum der Versendung des Anschreibens korreliert. Über Alter und Geschlecht hinweg sowie innerhalb der zwölf Zellen des Designs ergaben sich keine starken Beziehungen zwischen dem Datum des Anschreibens und den betrachteten Merkmalen. Dieses Ergebnis deutet darauf hin, dass die durch das an Ausschöpfungsraten orientierte Nachziehen von Adressen entstandene Abhängigkeit der Messungen keinen nennenswerten Einfluss auf die Zusammensetzung der Stichprobe ausgeübt hat.

Tabelle 1: Liste der in den Selektivitätsanalysen verwendeten Variablen¹ (Mittelwerte und in Klammern Standardabweichungen).

	Teilnahmeebene				
	1 N=1,908	2 N=1.264	3 N=928	4 N=638	5 N=516
Einwohnermeldeamt					
Alter bei Anschreiben (in Jahren)	86,1 (8,6)	86,3 (8,7)	86,0 (8,6)	85,5 (8,8)	84,9 (8,7)
Geschlecht (Frauenanteil in %)	55,6	52,6	50,9	48,4	50,0
Ein-Jahres-Sterblichkeit (in %)	13,5	12,0	10,7	9,9	5,6
Verheiratet (in %)	28,6	28,6	28,0	29,5	30,4
Verwitwet (in %)	55,9	56,8	55,3	54,4	53,9
Ledig (in %)	8,2	7,8	9,1	7,8	7,9
Geschieden (in %)	7,2	6,8	7,7	8,3	7,7
Heimbewohner (in %)	14,9	15,1	15,4	16,3	14,5
Sozialindex des Wohngebietes	24,2 (117,8)	24,7 (114,7)	31,4 (110,1)	36,1 (109,6)	40,7 (107,2)
Kurzbefragung					
Beobachtungs-Rating: ADL ²	—	3,9 (1,3)	4,0 (1,3)	4,1 (1,2)	4,2 (1,2)
Beobachtungs-Rating: Demenz	—	4,3 (5,8)	3,7 (5,6)	3,7 (5,5)	3,1 (4,9)
Beobachtungs-Rating: Sensorische Einbußen	—	3,2 (0,8)	3,2 (0,8)	3,2 (0,8)	3,3 (0,8)
Grober Bildungsindex	—	1,9 (0,7)	1,9 (0,7)	1,9 (0,7)	1,9 (0,7)
Lebenszufriedenheit (Item)	—	3,5 (1,1)	3,6 (1,0)	3,6 (1,0)	3,7 (1,0)
Subjektive körperliche Gesundheit (Item)	—	2,7 (1,1)	2,8 (1,1)	2,9 (1,1)	2,9 (1,1)
Subjektive seelisch-geistige Gesundheit (Item)	—	3,3 (1,1)	3,5 (1,0)	3,5 (1,0)	3,5 (1,0)
Ersterhebung					
ADL-Wert	—	—	4,4 (1,3)	4,4 (1,3)	4,5 (1,2)
Subjektive Gehstrecke (in km)	—	—	4,2 (1,4)	4,3 (1,4)	4,4 (1,3)
Body-Mass-Index	—	—	1,2 (0,2)	1,2 (0,2)	1,2 (0,2)
SMIMCO ³	—	—	0,45 (1,3)	0,51 (1,3)	0,59 (1,2)
Zahlen-Buchstaben-Test	—	—	68,1 (28,3)	70,3 (29,1)	73,3 (28,2)
Äquivalenzzinkommen (in DM)	—	—	1,979 (981)	2,020 (1.078)	2,042 (1.037)
Anzahl eng verbundener Personen	—	—	2,1 (3,0)	2,2 (3,3)	2,2 (3,3)
Depressivität	—	—	10,4 (5,8)	10,1 (5,6)	10,0 (5,5)
Ausgeglichenheit	—	—	4,0 (0,9)	4,0 (0,9)	4,1 (0,8)
Zufriedenheit mit dem eigenen Altern	—	—	4,0 (0,8)	4,1 (0,8)	4,1 (0,8)
Lebenszufriedenheit (Skala)	—	—	4,0 (0,8)	4,1 (0,8)	4,1 (0,8)

Intensivprotokoll								
Anzahl der mittel- bis schwergradigen Erkrankungen	—	—	—	—	—	—	—	8,1 (4,0)
ADL/IADL ² (T-Wert)	—	—	—	—	—	—	—	50,0 (10,0)
Sehschärfe (in Snellen-Einheiten)	—	—	—	—	—	—	—	0,32 (0,18)
Gehör (Hörschwelle in Dezibel)	—	—	—	—	—	—	—	54,4 (16,0)
Demenz (klinische Diagnose; Anteil in %)	—	—	—	—	—	—	—	21,1
Kognitive Leistungsfähigkeit (T-Wert)	—	—	—	—	—	—	—	50,0 (10,0)
Bildung (in Jahren)	—	—	—	—	—	—	—	10,8 (2,4)
Freizeitaktivitäten (Anzahl im letzten Jahr)	—	—	—	—	—	—	—	2,9 (2,4)
Größe des sozialen Netzwerks (Anzahl der Personen)	—	—	—	—	—	—	—	10,9 (7,2)
Neurotizismus (T-Wert)	—	—	—	—	—	—	—	50,0 (10,0)
Offenheit (T-Wert)	—	—	—	—	—	—	—	50,0 (10,0)
Depressivität (HAMMD ⁴)	—	—	—	—	—	—	—	5,7 (6,1)
Depression (klinische Diagnose; Anteil in %)	—	—	—	—	—	—	—	25,6

1. Die Variablen werden in den disziplinspezifischen Kapiteln dieses Bandes ausführlich eingeführt (vgl. Steinhagen-Thiessen & Borchelt, Kapitel 7; Helmchen et al., Kapitel 8; Smith & Baltes, Kapitel 9; Mayer & Wagner, Kapitel 10).
 2. ADL/IADL: Activities of Daily Living/Instrumental Activities of Daily Living.
 3. SMMCO: Short Mini Mental Cut-Off; Demenz-Screening.
 4. HAMMD: Hamilton Depression Scale.

Bemühen im Vordergrund, wichtige inhaltliche Bereiche von BASE wie z. B. Gesundheit, funktionelle Kapazität, kognitives Leistungsniveau, Alltagskompetenz, soziales Netzwerk, Bildung und Wohlbefinden möglichst frühzeitig zu erfassen.

3.2 Vorhersagen

Die in diesem Kapitel berichteten Analysen haben vornehmlich explorativ-deskriptiven Charakter. Jedoch lassen sich auf der Grundlage von Selektivitätsanalysen, die im Rahmen anderer gerontologischer Untersuchungen durchgeführt worden sind, einige Erwartungen hinsichtlich der Richtung der Selektivitätseffekte formulieren. In diesen zumeist längsschnittlich angelegten Untersuchungen hat sich gezeigt, dass die Wahrscheinlichkeit der Teilnahme sowie die Teilnahmedauer in der Regel positiv mit einer Vielzahl von Dimensionen korreliert, die sich insgesamt unter dem Begriff „Fitness“ oder „Kompetenz“ zusammenfassen lassen. So sind Personen mit längerer Teilnahmedauer *jünger* (DeMaio, 1980; Hawkins, 1975; Lowe & McCormick, 1955; Mercer & Butler, 1967/68; Weaver et al., 1975), *gesünder* (Goudy, 1976; Hertzog, Schaie & Gribbin, 1978; McC Ardle, Hamagami, Elias & Robbins, 1991; Norris, 1985; Powers & Bultena, 1972; Schaie, Labouvie & Barrett, 1973; Siegler & Botwinick, 1979) und entstammen einer *höheren sozialen Schicht* (Goudy, 1976; Powers & Bultena, 1972; Streib, 1966) als Personen, die ihre Teilnahme von vornherein verweigern oder im Laufe der Untersuchung auscheiden. Außerdem weisen Teilnehmer zumeist ein *höheres kognitives Leistungsniveau* (P. B. Baltes, Schaie & Nardi, 1971; Cooney, Schaie & Willis, 1988; Goudy, 1976; Norris, 1985; Powers & Bultena, 1972; Schaie, Labouvie & Barrett, 1973; Siegler & Botwinick, 1979) sowie eine *niedrigere Sterbewahrscheinlichkeit* auf (Manton & Woodbury, 1983; Powell et al., 1990; Siegler & Botwinick, 1979; Cooney et al., 1988). Schließlich gibt es Hinweise darauf, dass die Teilnahmedauer positiv mit erwünschten Persön-

lichkeitsmerkmalen wie z. B. Flexibilität korreliert (Cooney et al., 1988)⁴.

Die Vermutung liegt nahe, dass die Selektivitätseffekte der BASE-Stichprobe diesem Bild entsprechen sollten. Darüber hinaus weist BASE zumindest zwei Besonderheiten auf, die dafür sprechen, dass die *beobachteten* Selektivitätseffekte *größer* ausfallen dürften als in den meisten anderen Untersuchungen: die Schichtung der Altersvariablen – mit der Konsequenz eines ungewöhnlich hohen Anteils sehr alter Menschen in der Stichprobe – sowie die Zufallsziehung der Adressen durch das Landeseinwohnermeldeamt.

So ist im untersuchten Altersbereich das Lebensalter negativ mit der körperlichen Gesundheit und positiv mit der Sterbewahrscheinlichkeit korreliert (Siegler & Botwinick, 1979). Gesundheitsbezogene und sterblichkeitsbezogene Selektivität sollte in einer nach Alter geschichteten Stichprobe älterer Menschen demnach stärker ausgeprägt sein als in einer Stichprobe älterer Menschen, in der sich das Alter nach dem Zufall verteilt.

Ferner lässt sich argumentieren, dass die Zufallsziehung der Adressen durch das Landeseinwohnermeldeamt mit großer Wahrscheinlichkeit eine Ausgangsstichprobe mit geringem Vorab-Bias auf den vermuteten Selektionsvariablen ergeben hat. In konkreten Worten: Es besteht kein Grund zu der Annahme, dass die vom Landeseinwohnermeldeamt zufällig ausgewählten Personen besonders intelligent, besonders gesund und besonders gebildet waren. Dies steht im Gegensatz zu Untersuchungen, deren Ausgangsstichprobe durch andere Verfahren zustande gekommen ist, beispielsweise über die Mitgliedschaft in einer privaten Krankenversicherung (Schaie, 1983) oder die Zugehörigkeit zur Ärzteschaft (Shock et al., 1984). Die zufällige Ziehung der Ausgangsstichprobe verringert die Wahrscheinlichkeit, dass Selektivität bereits stattgefunden hat, bevor sie überhaupt beobachtet werden konnte. Demnach sollte die Zufallsziehung, ähnlich wie die Schichtung der Altersvariable, das Ausmaß an *beobachtbaren* Selektivitätseffekten eher erhöhen als verringern.

3.3 Methodische Überlegungen

3.3.1 Übersicht über die verwendeten statistischen Verfahren⁵

Im vorliegenden Kapitel wird die Stichprobenselektivität mit drei eng aufeinander bezogenen Verfahren untersucht: (a) logistische Regressionen zur Erfassung von Unterschieden in Mittelwerten und Häufigkeitsverteilungen (Aldrich & Nelson, 1984; Kühnel, Jagodzinski & Terwey, 1989); (b) der Vergleich von Varianz-Kovarianz-Matrizen zur Erfassung von Variabilitäts- und Zusammenhangsunterschieden (Bentler, 1989); (c) die Pearson-Lawleyschen Formeln zur Schätzung von statistischen Kennwerten für die Ausgangsstichprobe auf Merkmalen der Intensiv-erhebung (Lawley, 1943; Neale, 1991). Die gleichzeitige Betrachtung der mit den drei Verfahren erzielten Ergebnisse soll eine möglichst genaue Dokumentation sowie eine relativ umfassende Relevanzabschätzung von Selektivitätseffekten in der Berliner Altersstudie ermöglichen.

Um den Zusammenhang zwischen den drei Verfahren zu verdeutlichen, ist es hilfreich, die bereits oben beschriebene Struktur der Datenbasis zu formalisieren. Zu jeder Teilnehmerebene n wurden verschiedene Personenmerkmale der Studienteilnehmer erhoben, die sich in einem Vektor x_n zusammenfassen lassen. Für Personen der höchsten Teilnehmerebene ($N=516$ von $N=1.908$) liegen beispielsweise Informationen auf x_1 , x_2 , x_3 , und x_5 vor (x_4 kann ausgelassen werden, da hier keine neuen Informationen hinzukommen). Hingegen weisen Studienteilnehmer, die nicht einmal die zweite Teilnehmerebene (Kurzform der Ersterhebung) erreicht haben ($N=644$ von $N=1.908$), lediglich auf dem Vektor x_1 Informationen auf.

Weiterhin sei y_n ein Indikator für den Ausfall der Beobachtung nach der n -ten Teilnehmerebene. Dementsprechend weist y_n den Wert 1 auf, wenn eine Person die Ebene n erreicht hat, nicht aber die Ebene $n+1$. Hätte die Person auch die Ebene $n+1$ erreicht, dann müsste y_n den Wert 0 aufweisen. Für Personen ohne Beobachtungen auf der Ebene n ist y_n nicht definiert.

Die drei verwendeten statistischen Verfahren beruhen nun auf der Annahme, dass sich x_n und y_n als Ziehungen aus einer „Superpopulation“ beschreiben lassen, die durch eine Wahrscheinlichkeitsverteilung

⁴ Zu den Auswirkungen von Selektivität auf Zusammenhangsmuster gibt es bislang auf gerontologischem Gebiet kaum empirische Befunde (siehe aber McArdle et al., 1991). Auch sind bislang in der gerontologischen Literatur kaum Versuche unternommen worden, die Auswirkungen von Selektivität auf später erfasste Variablen mit Hilfe von Schätzverfahren zu quantifizieren (siehe aber auch hier McArdle & Hamagami, 1991; McArdle et al., 1991).

⁵ Die Verfahren werden in den entsprechenden Abschnitten des Ergebnisteils jeweils kurz erläutert. Der methodisch weniger interessierte Leser kann diesen Unterabschnitt auslassen.

charakterisiert ist (Cassel, Särndal & Wretman, 1977). Die Annahme einer Superpopulation erlaubt es, Aspekte der Population durch entsprechende Parameter auszudrücken und klassische statistische Verfahren zu ihrer Schätzung anzuwenden. Insbesondere lassen sich drei verschiedene Fragestellungen formulieren, die unter Zuhilfenahme weiterer Annahmen Aufschluss über die Stichprobenselektivität geben können:

1. *Stichprobenausfall in Abhängigkeit von Mittelwerten und Häufigkeitsverteilungen auf bereits beobachteten Merkmalen.* Kann die Wahrscheinlichkeit für den Ausfall einer Beobachtung durch eine Funktion von Ausprägungen bereits erfasster Merkmale beschrieben werden? Um dies zu untersuchen, wird die Ausfallwahrscheinlichkeit als eine Funktion der bereits erfassten Merkmale beschrieben:

$$E(y_n|x_n) = \Pr(y_n = 1|x_n) \quad (1)$$

2. *Unterschiede in Varianzen und Kovarianzen zwischen Personen unterschiedlicher Teilnahmetiefe.* Unterscheiden sich die weiter teilnehmenden Personen von den ausfallenden Personen hinsichtlich der Varianzen und Kovarianzen der bereits erfassten Merkmale? Gilt:

$$\text{cov}(x_n|y_n=0) \neq \text{cov}(x_n|y_n=1)? \quad (2)$$

3. *Kumulative Auswirkungen des Ausfallprozesses auf Konstrukte der Intensiverhebung.* Gibt es Unterschiede in den Erwartungswerten E (d. h. in den Mittelwerten und Häufigkeitsverteilungen) der Merkmale der Intensiverhebung zwischen Personen, die diese Ebene erreicht haben, und Personen, deren Teilnahme auf einer niedrigeren Ebene endete? Gilt:

$$E(x_5|x_n, y_n=0) \neq E(x_5|x_n)? \quad (3)$$

Gefragt wird also: (1) nach dem Verhältnis des Ausfallindikators zu bereits erhobenen Merkmalen; (2) nach Unterschieden zwischen weiter teilnehmenden und ausfallenden Personen auf den zweiten Momenten (d. h. Varianzen und Kovarianzen) der bereits erfassten Merkmale; und (3) nach der Beziehung von Merkmalen, die auf der höchsten Teilnahmeebene erhoben wurden, zu Merkmalen der vorherigen Teilnahmeebenen und dem Ausfallindikator.

Wenn bei Anwendung der diesen Fragestellungen entsprechenden statistischen Verfahren Unterschiede zwischen weiter teilnehmenden und nicht weiter teilnehmenden Personengruppen auf den untersuchten Merkmalen zutage treten, so deutet dies darauf hin, dass der Stichprobenausfall mit Stichprobenselektivität einherging, deren Stärke dann mit Hilfe statistischer Kennwerte dargestellt werden kann. Werden keine Unterschiede gefunden, so darf ein derartiges

Ergebnis nicht im Sinne eines Nachweises der Abwesenheit von Stichprobenselektivität gedeutet werden. Zum Beispiel kann nicht ausgeschlossen werden, dass einige untersuchungsrelevante Merkmale, die den Stichprobenausfall vorhersagen, von vornherein nicht beobachtet wurden und deswegen auch nicht in die Analyse aufgenommen werden konnten. Dieses Problem ist im vorliegenden Fall insbesondere für die Analyse des Stichprobenausfalls beim Übergang von der ersten zur zweiten Teilnahmeebene von Bedeutung, für den lediglich die bereits auf der ersten Teilnahmeebene erfassten Merkmale zur Verfügung stehen. Weiterhin erfassen die verwendeten Verfahren nicht alle prinzipiell möglichen Formen von Stichprobenselektivität, sondern beschränken sich auf Mittelwerte, Häufigkeitsverteilungen, Varianzen und Kovarianzen. Da sich die Analysen der Berliner Altersstudie ohnehin auf diese Aspekte des Datensatzes konzentrieren, scheint diese Beschränkung auf die ersten und zweiten Momente vertretbar.

3.3.2 Selektivität in Abhängigkeit von Altersgruppe und Geschlecht

Die Merkmale Alter und Geschlecht sind für die Berliner Altersstudie von herausragender Bedeutung. Zum einen definieren sie als Schichtungsvariablen den Erhebungsplan von BASE. Zum anderen ist die Beschreibung und Erklärung insbesondere von Alters-, aber auch von Geschlechtsunterschieden über alle Bereiche hinweg ein zentrales Anliegen der gesamten Untersuchung.

Die Validität von Aussagen zu Altersgradienten und Geschlechtsunterschieden wird beeinträchtigt, falls das Ausmaß an Selektivität mit Alter und Geschlecht zusammenhängt. Mit anderen Worten: Wenn die Selektivität der Frauen sich von der der Männer unterscheidet oder wenn die Selektivität mit dem Alter zu- oder abnimmt, dann sind die an der Intensivstichprobe beobachteten Alters- und Geschlechtsunterschiede zumindest teilweise ein Abbild dieser differentiellen Selektivität.

Ein gutes Beispiel für diese Problematik sind Prävalenzschätzungen. So wäre es möglich, dass Selektivitätseffekte zu einer generellen Unterschätzung der Prävalenz einer Krankheit führen. Eine Untersuchung der Selektivität getrennt nach Altersgruppen könnte zeigen, dass das Ausmaß dieser Prävalenzunterschätzung im jüngeren Altersbereich eher gering, im höheren Altersbereich jedoch beträchtlich ist. In diesem Fall würden die Ergebnisse der Selektivitätsanalysen nicht nur die Interpretation der beobach-

teten Prävalenzrate, sondern auch die Interpretation der Alterstrends beeinflussen.

Für die Selektivitätsanalysen ergibt sich daraus die Forderung, etwaige Auswirkungen des Stichprobenausfalls gesondert für Männer und Frauen und getrennt nach Altersgruppen zu untersuchen. Im Idealfall würde man die Selektivität getrennt für jede der zwölf Zellen des Erhebungsplans analysieren (d. h. für Männer und Frauen in je sechs Altersgruppen), um *Interaktionen* zwischen Alter und Geschlecht in ihren Wirkungen auf die Selektivität erfassen zu können. Derartige Analysen, die dem Erhebungsplan der Studie exakt entsprächen, sind jedoch aufgrund niedriger Fallzahlen mit einer deutlichen Reduktion der statistischen Power verbunden. Vornehmlich aus diesem Grund beschränken sich die hier berichteten Ergebnisse auf einen Vergleich der Selektivität zwischen Männern und Frauen über Altersgruppen hinweg sowie zwischen den sechs Altersgruppen (70- bis 74jährige, 75- bis 79jährige usw.) über Männer und Frauen hinweg.

Inhaltliche Aussagen zu Häufigkeiten, Verteilungen und anderen statistischen Kennwerten können sich entweder direkt auf die jeweilige Stichprobe beziehen, oder sie können bei Anwendung von Gewichtungsfunktionen die Tatsache berücksichtigen, dass es in der Westberliner Wohnbevölkerung über 70 im Vergleich zu einer nach Alter und Geschlecht geschichteten Stichprobe relativ mehr Frauen als Männer und mehr „junge“ als „alte“ Alte gibt. Die in diesem Kapitel berichteten Daten beziehen sich, sofern nicht anders vermerkt, auf *ungewichtete* Daten.

4. Ergebnisse

Im folgenden berichten wir die Ergebnisse der Selektivitätsanalysen unter drei verschiedenen Fragestellungen: (1) Mittelwerte und Häufigkeiten; (2) Variabilität und Zusammenhangsmuster; (3) kumulative Auswirkungen der Stichprobenselektivität auf Konstrukte der Intensiverhebung. Wegen seiner herausragenden Bedeutung beginnen wir die Darstellung mit einer gesonderten Betrachtung der Ein-Jahres-Sterblichkeit.

4.1 Ein-Jahres-Sterblichkeit und Teilnahmetiefe

In einer Untersuchung über das hohe Alter sind Zusammenhänge zwischen der Teilnahmewahrscheinlichkeit und der Sterblichkeit von besonderer Bedeutung. Abbildung 2 zeigt die Ein-Jahres-Sterblichkeit als eine Funktion der Teilnahmetiefe. In der Ausgangsstichprobe (Teilnahmeebene 1) beträgt die Ein-Jahres-Sterblichkeit 13,5%, in der Stichprobe der Personen mit vollständiger Intensiverhebung (Teilnahmeebene 5) hingegen weniger als 6%.

Eine direkte Interpretation dieser Werte ist problematisch, weil die Sterbewahrscheinlichkeit für Personen, die höhere Teilnahmeebenen erreicht haben, bereits aufgrund der längeren zeitlichen Ausdehnung der Datenerhebung zwangsläufig niedriger ausfallen muss als für Personen niedriger Teilnahmeebenen. Unterschiede zwischen den Teilnahmeebenen in der Ein-Jahres-Sterblichkeit wären also auch dann zu beobachten, wenn Personen unterschiedlicher Teilnahmeebenen sich in ihrem Sterberisiko nicht von vornherein voneinander unterscheiden würden. Aus diesem Grund ist es notwendig, die empirisch beobachteten Werte mit geschätzten Erwartungswerten in Beziehung zu setzen, die sich nach Berücksichtigung der unterschiedlichen Untersuchungsdauer bei teilnahmeebenen-unabhängiger Sterbewahrscheinlichkeit ergeben würden⁶.

Wie der Vergleich der beobachteten und der unter Annahme der Teilnahmeebenen-Unabhängigkeit erwarteten Werte verdeutlicht (siehe Abb. 2), zeigt sich *auch nach Berücksichtigung der unterschiedlichen Teilnahmedauern, dass Personen, die Teilnahmeebene 5 erreicht haben, weiterhin eine leicht verringerte Ein-Jahres-Sterblichkeit aufweisen*. Dabei beträgt der Unterschied zwischen der beobachteten und der aufgrund der Teilnahmedauer erwarteten Ein-Jahres-Sterblichkeit der Intensivstichprobe 3%. Komplementär zur relativ niedrigen Ein-Jahres-Sterblichkeit der Intensivstichprobe sind die Stichproben der niedrigeren Teilnahmeebenen durch eine gemessen an der Teilnahmedauer leicht erhöhte Ein-Jahres-Sterblichkeit gekennzeichnet.

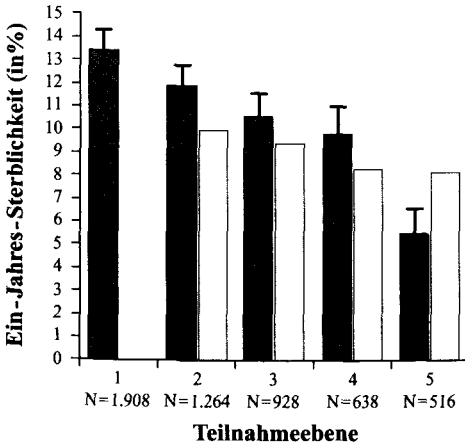
Zusammenfassend lässt sich feststellen, dass die Intensivstichprobe gegenüber der Ausgangsstichprobe

⁶ Die in Abbildung 2 berichteten Schätzwerte der Ein-Jahres-Sterblichkeit bei Berücksichtigung der differentiellen Teilnahmedauer wurden wie folgt berechnet. Für jeden Studienteilnehmer, der zumindest die zweite Teilnahmeebene erreicht hatte, wurde die Teilnahmedauer in Tagen (s) von 365 abgezogen, um den Zeitraum zu bestimmen, in dem Sterblichkeit bei dieser Person auftreten konnte. Bei gegebener Ein-Jahres-Sterblichkeitsrate m in der Ausgangsstichprobe ($m=0,1347$) ließ sich die aufgrund der Teilnahmedauer erwartete Wahrscheinlichkeit Pr einer Person, die verbleibenden Tage des Jahres zu überleben, unter der vereinfachten Annahme eines zeitlich konstanten Sterberisikos durch folgende Gleichung ausdrücken:

$$Pr = (1 - m)^{\lfloor 365 - s \rfloor / 365}$$

Durch Subtraktion der Überlebenswahrscheinlichkeit von eins ergibt sich die gesuchte teilnahmedauerkorrigierte Sterbewahrscheinlichkeit. Die in Abbildung 3 dargestellten Erwartungswerte beruhen auf dem arithmetischen Mittel dieser Wahrscheinlichkeiten.

Abbildung 2: Ein-Jahres-Sterblichkeit in Abhängigkeit von der Teilnahmeebene. Die dunklen Säulen stellen die beobachtete, die hellen Säulen die nach Berücksichtigung der Teilnahmedauer erwartete Wahrscheinlichkeit dar, innerhalb eines Jahres nach Kontakt mit der Studie zu versterben. Die Intensivstichprobe (Teilnahmeebene 5) zeigt gegenüber der Ausgangsstichprobe (Teilnahmeebene 1) eine reduzierte Ein-Jahres-Sterblichkeit, die sich nicht vollständig auf die längere Teilnahmedauer zurückführen lässt.



Anmerkung: Die Fehlerbalken beziehen sich auf den Standardfehler des beobachteten Prozentwerts.

eine deutlich reduzierte Ein-Jahres-Sterblichkeit aufweist. Angesichts der hohen Mortalität alter und sehr alter Menschen ist dieses Ergebnis allein aufgrund der zeitlichen Ausdehnung des Intensivprotokolls (Median=133 Tage) zu erwarten. Eine statistische Auswertung dieser Erwartung ergibt, dass sich die beobachtete Reduktion der Ein-Jahres-Sterblichkeit in der Intensivstichprobe (N=516) gegenüber der Ausgangsstichprobe (N=1.908) in der Tat zum größten Teil, aber nicht vollständig, mit der zeitlichen Ausdehnung der Intensiverhebung aufklären lässt. Personen, die an der Intensiverhebung teilnahmen, wiesen demnach eine überzufällig niedrigere Sterbewahrscheinlichkeit gegenüber der Ausgangsstichprobe auf. Eine (hier nicht mögliche) eingehende inhaltliche Würdi-

gung dieses Befunds, der gut mit den Ergebnissen anderer Untersuchungen übereinstimmt (Cooney et al., 1988; Manton & Woodbury, 1983; Powell et al., 1990; Siegler & Botwinick, 1979), erfordert unter anderem die Identifikation von Variablen, die sowohl mit der Teilnahmetiefe als auch mit der Sterblichkeit in Zusammenhang stehen (vgl. P. B. Baltes et al., 1971; Kruse et al., 1993; Manton & Woodbury, 1983).

4.2 Mittelwerte und Häufigkeitsverteilungen

Zur Beantwortung der Frage nach der Abhängigkeit des Stichprobenausfalls von der Höhe der Ausprägung auf bereits erfassten Merkmalen wurden logistische Regressionen verwendet, deren exponierte Regressionskoeffizienten als „Odds-ratios“ interpretiert werden können. Zum besseren Verständnis der Ergebnisse gehen wir kurz auf die Bedeutung dieses Kennwerts ein.

Bei zweifach gestuften (dichotomen) unabhängigen Variablen kann ein Odds-ratio direkt als Wahrscheinlichkeitsverhältnis der Ausprägungen dieser Variablen interpretiert werden. So bedeutet ein Odds-ratio von 2,0 auf der Variablen Geschlecht mit den Ausprägungen 0 für Männer und 1 für Frauen, dass das Ausfallwahrscheinlichkeitsverhältnis der Frauen doppelt so groß ist wie das der Männer.

Bei stetigen unabhängigen Variablen ist zu beachten, dass die in Odds-ratios ausgedrückten Wahrscheinlichkeitsverhältnisse in den Einheiten der unabhängigen Variablen skaliert sind. Wird also beispielsweise das Alter in Jahren gemessen und der Odds-ratio betrüge 1,1, so wäre die Wahrscheinlichkeit des Nichterreichens der nächsten Teilnahmeebene bei 71jährigen 1,1-mal größer als bei 70jährigen Personen und bei 80jährigen 1,1¹⁰=2,59-mal größer als bei 70jährigen Personen⁷.

4.2.1 Kontrast 1: Nur Teilnahmeebene 1 (N=644) versus höhere Teilnahmeebenen (N=1.264)

Als Kovariablen wurden die auf der Teilnahmeebene 1 zur Verfügung stehenden Variablen Alter, Geschlecht, Familienstand, Sozialindex des Wohngebietes, Wohnsituation sowie die Ein-Jahres-Sterblichkeit in das Modell aufgenommen. Die Variablen Alter und Sozialindex sind intervallskaliert, bei den Variablen Geschlecht, Wohnsituation und Ein-Jahres-Sterblichkeit handelt es sich um dichotome Va-

⁷ Es sei darauf hingewiesen, dass die Ergebnisse der logistischen Regression mit denen der Diskriminanzanalyse übereinstimmen, wenn die unabhängigen Variablen annähernd normal verteilt sind (vgl. Haggstrom, 1983).

Tabelle 2: Ausfallwahrscheinlichkeit von Teilnahmeebene 1 zu Teilnahmeebene 2.

Variable	B	s	p-Wert	Odds-ratio
Alter bei Anschreiben	-0,01	0,01	0,29	0,99
Geschlecht (weiblich)	0,50	0,12	0,00	1,65
Ein-Jahres-Sterblichkeit	0,52	0,15	0,00	1,68
Familienstand ¹				
geschieden	-0,10	0,21	0,65	0,91
verwitwet	-0,25	0,14	0,06	0,78
ledig	-0,09	0,21	0,67	0,91
Heimbewohner	-0,13	0,15	0,40	0,88
Sozialindex des Wohngebietes	-0,01	0,04	0,97	0,99
Konstante	-0,87	0,10	0,00	

Pseudo-R²=0,014¹ Referenzkategorie: verheiratet

riablen mit den Ausprägungen 0 und 1. Die Variable Familienstand wurde in vier Dummyvariablen zerlegt, wobei die Gruppe der Verheirateten als Referenzkategorie gewählt wurde. In Tabelle 2 sind die Koeffizienten (B), deren Standardfehler (s), die zufallskritische Absicherung der Koeffizienten gegen den Erwartungswert Null (p-Werte), die Odds-ratios sowie das Pseudo-R² für den Kontrast 1 wiedergegeben.

Es fällt auf, dass die Erklärungskraft des Gesamtmodells mit einem Pseudo-R² von 0,014 sehr gering ist. Systematische Ausfälle bei den auf dieser Teilnahmestufe zur Verfügung stehenden Variablen erklären also nur einen sehr geringen Teil des Ausfallprozesses. Signifikante Unterschiede zwischen weiter teilnehmenden und die Teilnahme beendenden Personen finden sich nur bei der Variablen Geschlecht und der Ein-Jahres-Sterblichkeit. So ist die Wahrscheinlichkeit, nicht weiter an der Studie teilzunehmen, für Personen, die innerhalb eines Jahres nach dem Anschreiben verstorben sind, um den Faktor 1,68 größer als für Personen, die in diesem Zeitraum nicht verstarben. Für Frauen ist die Ausfallwahrscheinlichkeit um den Faktor 1,65 größer als für Männer. Letzteres bedeutet, dass bei den Frauen eine größere Anzahl von Adressen eingesetzt werden musste, um in der Stichprobe der höchsten Teilnahmeebene, dem Intensivprotokoll, die gleiche Anzahl von Männern und Frauen in den Altersgruppen zu erzielen. Ein systematischer Zusammenhang zwischen Teilnahmewahrscheinlichkeit und dem Familienstand und der Wohnsituation lässt sich auf dieser Teilnahmeebene nicht feststellen. Dies gilt überraschenderweise auch für das Alter.

4.2.2 Kontrast 2: Nur bis Teilnahmeebene 2 (N=336) versus höhere Teilnahmeebenen (N=928)

Zusätzlich zu den unter 4.2.1 aufgeführten Variablen konnten die Merkmale der Teilnahmeebene 2 (Beobachtungs-Rating der ADL, Beobachtungs-Rating der Demenz, Beobachtungs-Rating der sensorischen Einbußen, Bildungsindex, Lebenszufriedenheit, subjektive körperliche Gesundheit, subjektive seelisch-geistige Gesundheit; vgl. Tabelle 1) mitberücksichtigt werden. Das Ergebnis der logistischen Regression ist in Tabelle 3 dargestellt.

Die Einbeziehung weiterer Variablen hat erwartungsgemäß eine im Vergleich zum Modell der ersten Stufe deutlich höhere Erklärungskraft des Modells zur Folge (Pseudo-R²=0,19). Dies geht im wesentlichen auf den Wohngebiets-Sozialindex, das Demenz-Rating aus der begleitenden Beobachtung sowie auf die subjektive seelisch-geistige Gesundheit zurück. Je höher der Wert des Wohngebiets-Sozialindex, d. h. je „besser“ das Wohngebiet, desto niedriger die Ausfallwahrscheinlichkeit (Odds-ratio=0,73; p<0,01). Auch bei der subjektiven seelisch-geistigen Gesundheit sinkt die Ausfallwahrscheinlichkeit mit höheren Werten, also mit zunehmend besserer subjektiver Beurteilung (Odds-ratio=0,70; p<0,01). Für das Demenz-Rating der begleitenden Beobachtung lässt sich eine Zunahme der Ausfallwahrscheinlichkeit mit steigenden Werten feststellen (Odds-ratio=1,09; p<0,01). Personen mit dementiellen Symptomen haben also eine tendenziell größere Wahrscheinlichkeit, nicht weiter an der Studie teilzunehmen. Zusätzlich zu diesen Variablen zeigt sich auch auf dieser Stufe wieder

Tabelle 3: Ausfallwahrscheinlichkeit von Teilnahmeebene 2 zu Teilnahmeebene 3.

Variable	B	s	p-Wert	Odds-ratio
Alter bei Anschreiben	0,02	0,02	0,31	1,02
Geschlecht (weiblich)	0,62	0,30	0,04	1,85
Ein-Jahres-Sterblichkeit	-0,08	0,37	0,81	0,91
Familienstand ¹				
geschieden	-1,14	0,72	0,11	0,32
verwitwet	0,00	0,36	0,99	1,00
ledig	-0,94	0,65	0,15	0,39
Heimbewohner	-0,17	0,33	0,61	0,85
Sozialindex des Wohngebietes	-0,01	0,10	0,00	0,73
Beobachtungs-Rating: ADL	-0,12	0,10	0,20	0,88
Beobachtungs-Rating: Demenz	0,09	0,02	0,00	1,09
Beobachtungs-Rating: Sensorische Einbußen	-0,11	0,16	0,48	0,89
Grober Bildungsindex	-0,44	0,20	0,03	0,64
Lebenszufriedenheit (Item)	-0,09	0,14	0,50	0,91
Subjektive körperliche Gesundheit	-0,08	0,13	0,56	0,93
Subjektive seelisch-geistige Gesundheit	-0,35	0,14	0,01	0,70
Konstante	0,30	0,83	0,72	

Pseudo-R²=0,19¹ Referenzkategorie: verheiratet.

eine größere Ausfallwahrscheinlichkeit für Frauen. Allerdings ist dieser Effekt mit einem größeren Standardfehler behaftet (Odds-ratio=1,85; $p < 0,05$). Auch die Bildung steht in dem aus der Literatur bekannten Zusammenhang mit der Ausfallwahrscheinlichkeit: Mit zunehmendem Bildungsstand sinkt die Ausfallwahrscheinlichkeit (Odds-ratio=0,64; $p < 0,05$). Auf den anderen verwendeten Variablen ließen sich dagegen keine Selektivitätseffekte nachweisen.

4.2.3 Kontrast 3: Nur bis Teilnahmeebene 3 (N=290) versus höhere Teilnahmeebenen (N=638)

Auf dieser Stufe wurden zusätzlich zu den unter Abschnitt 4.2.2 aufgeführten Merkmalen zentrale Variablen aus der Ersterhebung mit in das Modell aufgenommen (vgl. Tabelle 1). Gegenüber dem Modell der zweiten Stufe ist die Erklärungskraft des Modells mit einem Pseudo-R² von 0,05 auf dieser Stufe deutlich niedriger. Die Wahrscheinlichkeit, nach der Ersterhebung nicht weiter teilzunehmen, lässt sich demnach nur in sehr geringem Umfang systematisch auf die hier verwendeten Variablen beziehen. Wie Tabelle 4 verdeutlicht, sind dabei die Wohnsituation und die Lebenszufriedenheit von Bedeutung.

So ist die Ausfallwahrscheinlichkeit von Personen, die in einem Privathaushalt leben, im Vergleich zu Personen in stationären Altenhilfeeinrichtungen auf dieser Stufe etwa doppelt so groß (Odds-ratio=0,49; $p < 0,05$). Schließlich nimmt die Ausfallwahrscheinlichkeit mit höheren Werten auf der Lebenszufriedenheitsskala erwartungsgemäß ab (Odds-ratio=0,70; $p < 0,05$). Keine der anderen in dem Modell verwendeten Variablen weist einen signifikanten Zusammenhang mit der Teilnahmewahrscheinlichkeit auf dieser Teilnahmeebene auf.

4.2.4 Kontrast 4: Abbruch im Laufe der Intensiverhebung (N=122) versus höchste Teilnahmeebene (N=516)

Auf dieser Stufe wurden keine zusätzlich zu den unter 4.2.3 aufgeführten Variablen verwendet. Mit einem Pseudo-R² von 0,16 beschreibt das Modell den Ausfallprozess während der Intensiverhebung im Vergleich zur dritten Stufe wesentlich besser. Aus Tabelle 5 lässt sich ersehen, dass auf dieser Stufe die Ein-Jahres-Sterblichkeit den stärksten Effekt hat (Odds-ratio=7,07; $p < 0,01$), was, wie oben bereits ausgeführt, angesichts der Dauer der Intensiverhebung durchaus plausibel ist.

Neben der Ein-Jahres-Sterblichkeit erweist sich der Sozialindex des Wohngebietes auf dieser Stufe erneut als bedeutsam (Odds-ratio=0,80; $p < 0,05$), wobei auch hier wiederum Personen aus besseren Wohngebieten eine niedrigere Ausfallwahrscheinlichkeit haben. Ebenfalls von Bedeutung sind die Effekte des Body-Mass-Index und des Zahlen-Buchstaben-Tests. Je besser die kognitiven Fähigkeiten, gemessen mit Hilfe des Zahlen-Buchstaben-Tests, desto geringer ist die Wahrscheinlichkeit, die Intensiverhebung nicht zu beenden (Odds-ratio=0,99; $p < 0,05$). Auch für den Body-Mass-Index, einem Indikator körperlicher Gesundheit, gilt, dass mit dessen Abnahme die Ausfallwahrscheinlichkeit steigt

(Odds-ratio=0,17; $p < 0,05$). Für alle anderen hier verwendeten Merkmale ist kein systematischer Zusammenhang mit der Ausfallwahrscheinlichkeit während der Intensiverhebung festzustellen.

4.2.5 Zusammenfassung der Analysen zu Mittelwerten und Häufigkeitsverteilungen

Zusammenfassend lässt sich festhalten, dass über alle Teilnahmeebenen der Studie hinweg die Teilnahme-wahrscheinlichkeit in einem relativ schwachen Zusammenhang mit den beobachteten Merkmalen steht. Die deutlichste Selektivität zeigt sich für Personen.

Tabelle 4: Ausfallwahrscheinlichkeit von Teilnahmeebene 3 zu Teilnahmeebene 4.

Variable	B	s	p-Wert	Odds-ratio
Alter bei Anschreiben	0,01	0,02	0,61	1,01
Geschlecht (weiblich)	-0,05	0,23	0,82	0,95
Ein-Jahres-Sterblichkeit	0,15	0,30	0,61	1,16
Familienstand ¹				
geschieden	-0,36	0,43	0,41	0,70
verwitwet	0,00	0,25	1,00	1,00
ledig	0,41	0,38	0,28	1,51
Heimbewohner	-0,71	0,32	0,03	0,49
Sozialindex des Wohngebietes	-0,08	0,08	0,30	0,92
Beobachtungs-Rating: ADL	0,01	0,10	0,96	1,01
Beobachtungs-Rating: Demenz	-0,01	0,02	0,52	0,98
Beobachtungs-Rating: Sensorische Einbußen	-0,14	0,12	0,26	0,87
Grober Bildungsindex	0,06	0,14	0,68	1,06
Lebenszufriedenheit (Item)	0,22	0,12	0,07	1,25
Subjektive körperliche Gesundheit	-0,10	0,10	0,34	0,90
Subjektive seelisch-geistige Gesundheit	0,06	0,11	0,55	1,07
ADL-Wert	0,08	0,09	0,36	1,09
Subjektive Gehstrecke	-0,03	0,10	0,80	0,97
Body-Mass-Index	0,38	0,49	0,43	1,45
SMMCO	-0,19	0,10	0,06	0,83
Zahlen-Buchstaben-Test	-0,01	0,01	0,31	0,99
Äquivalenzeinkommen	0,00	0,00	0,66	1,00
Anzahl eng verbundener Personen	-0,02	0,04	0,51	0,98
Depressivität	-0,01	0,02	0,55	0,99
Ausgeglichenheit	-0,20	0,13	0,14	0,82
Zufriedenheit mit dem eigenen Altern	-0,16	0,16	0,31	0,85
Lebenszufriedenheit (Skala)	-0,35	0,16	0,03	0,70
Konstante	1,40	1,22	0,25	

Pseudo- $R^2 = 0,05$

¹ Referenzkategorie: verheiratet.

die innerhalb eines Jahres nach dem Anschreiben verstorben sind. Ausfälle im Verlauf der Intensiverhebung werden in erheblichem Maße durch dieses Merkmal bestimmt. Dies kann zum Teil, jedoch nicht vollständig, auf die zeitliche Ausdehnung der Intensiverhebung zurückgeführt werden. Weitere signifikante Effekte auf die Teilnahmewahrscheinlichkeit fanden sich beim Sozialindex des Wohngebietes auf der zweiten und vierten Teilnahmeebene. Personen aus besseren Wohngebieten sind demnach in der Stichprobe, für die eine vollständige Intensiverhebung vorliegt, leicht überrepräsentiert. Gemeinsam mit dem positiven Zusammenhang zwischen Bildung und Teilnahmewahrscheinlichkeit deutet dies auf eine Überprä-

sentierung von Personen aus höheren Schichten hin. Allerdings ist der Bildungseffekt nur auf der zweiten Teilnahmeebene bedeutungsvoll und überdies mit einem relativ großen Standardfehler behaftet.

Auf der dritten Teilnahmeebene konnte demgegenüber kein Effekt der Bildung und auch kein Effekt des Einkommens auf die Teilnahmewahrscheinlichkeit festgestellt werden, was der These eines eindeutig positiven Selektivitätseffekts nach sozialer Schicht entgegensteht. Die Effekte des Demenz-Ratings und der subjektiven seelisch-geistigen Gesundheit auf der zweiten Teilnahmeebene, wie auch der Lebenszufriedenheit auf der dritten sowie des Body-Mass-Index und des Zahlen-Buchstaben-Tests

Tabelle 5: Ausfallwahrscheinlichkeit von Teilnahmeebene 4 zu Teilnahmeebene 5.

Variable	B	s	p-Wert	Odds-ratio
Alter bei Anschreiben	0,03	0,02	0,25	1,03
Geschlecht (weiblich)	-0,35	0,33	0,29	0,70
Ein-Jahres-Sterblichkeit	1,96	0,39	0,00	7,07
Familienstand ¹				
geschieden	0,57	0,52	0,27	1,77
verwitwet	-0,10	0,36	0,78	0,90
ledig	0,47	0,58	0,42	1,60
Heimbewohner	0,03	0,41	0,94	1,03
Sozialindex des Wohngebietes	-0,23	0,11	0,04	0,80
Beobachtungs-Rating: ADL	0,09	0,14	0,54	1,09
Beobachtungs-Rating: Demenz	0,03	0,03	0,19	1,04
Beobachtungs-Rating: Sensorische Einbußen	-0,08	0,18	0,64	0,92
Grober Bildungsindex	-0,13	0,20	0,51	0,88
Lebenszufriedenheit (Item)	-0,07	0,17	0,68	0,93
Subjektive körperliche Gesundheit	0,05	0,14	0,73	1,05
Subjektive seelisch-geistige Gesundheit	-0,12	0,15	0,42	0,89
ADL-Wert	0,19	0,13	0,16	1,21
Subjektive Gehstrecke	-0,02	0,15	0,90	0,98
Body-Mass-Index	-1,77	0,77	0,02	0,17
SMMCO	0,03	0,13	0,82	1,03
Zahlen-Buchstaben-Test	-0,01	0,01	0,02	0,99
Äquivalenzeinkommen	0,00	0,00	0,51	1,00
Anzahl eng verbundener Personen	0,08	0,04	0,06	1,08
Depressivität	-0,02	0,03	0,44	0,98
Ausgeglichenheit	0,02	0,19	0,92	1,02
Zufriedenheit mit dem eigenen Altern	-0,32	0,23	0,16	0,73
Lebenszufriedenheit (Skala)	-0,26	0,23	0,27	0,77
Konstante	3,61	1,89	0,06	

Pseudo-R²=0,16

¹ Referenzkategorie: verheiratet.

auf der vierten Teilnahmeebene, deuten auf eine Überrepräsentierung von Personen mit höherem kognitivem Leistungsniveau, von gesünderen und zufriedeneren Personen hin.

Es muss jedoch einschränkend hinzugefügt werden, dass von der Vielzahl der verwendeten Merkmale nur wenige in einem statistisch signifikanten Zusammenhang mit der Teilnahmewahrscheinlichkeit stehen. Dies gilt insbesondere auch für die Gesundheitsmaße, von denen nur der Indikator des Body-Mass-Index auf der letzten Teilnahmestufe einen erwähnenswerten Einfluss auf die Teilnahmewahrscheinlichkeit ausübt, so dass nicht von einer generellen Überrepräsentierung gesünderer Personen ausgegangen werden kann. Vielmehr zeigt sich auf der dritten Teilnahmeebene eine größere Teilnahmewahrscheinlichkeit von Personen in stationären Altenhilfeeinrichtungen, was einem positiven Selektivitätseffekt des Gesundheitszustandes entgegensteht. Überraschenderweise konnte auf keiner der Teilnahmeebenen ein linearer, von den in die Modelle einbezogenen Variablen unabhängiger Alterseffekt nachgewiesen werden. Die Geschlechtseffekte auf der ersten und zweiten Teilnahmestufe weisen in die aus der Literatur bekannte Richtung einer höheren Ausfallwahrscheinlichkeit von Frauen. Aufgrund der proportional nach dem Geschlecht stratifizierten Stichprobe der Intensiverhebung hat dies jedoch keinen Einfluss auf die Zusammensetzung dieser Stichprobe.

4.3 Variabilität und Zusammenhangsmuster

Ein Vergleich der Varianzen und Kovarianzen kann weitere Aufschlüsse über die Stichprobenselektivität geben. So wäre es denkbar, dass die logistischen Regressionen zu dem Ergebnis führen, dass Personen mit niedrigen oder hohen Werten auf gewissen Variablen eine niedrigere Wahrscheinlichkeit aufweisen, die nächste Teilnahmeebene zu erreichen. Dieser Selektivitätseffekt auf der Ebene der Mittelwerte bedeutet jedoch nicht zwangsläufig, dass sich die beiden Gruppen in ihrer Variabilität und in ihrem Zusammenhangsmuster voneinander unterscheiden.

Auf der anderen Seite ist es auch möglich, dass die Auswirkungen des Stichprobenausfalls deutlicher in den Varianzen und Kovarianzen als in den Mittelwerten zutage treten. Wenn z. B. besonders gesunde

und besonders kranke Personen aus verschiedenen Gründen eine gegenüber durchschnittlich gesunden Personen verringerte Wahrscheinlichkeit aufweisen, die nächste Teilnahmeebene zu erreichen, dann könnte dies zur Folge haben, dass sich die Abbrecher im Mittelwert auf einer gesundheitsbezogenen Variablen nicht von der ursprünglichen Stichprobe unterscheiden würden. Ein derartiger die Variabilität betreffender Selektivitätseffekt ließe sich mit Hilfe linearer logistischer Regressionen kaum nachweisen (d. h., das LOGIT-Modell wäre fehlspezifiziert). Die Uneinheitlichkeit des Ausfallprozesses sollte aber bei den Abbrechern zu einer Erhöhung der Varianz auf den gesundheitsbezogenen Variablen führen, so dass die Varianz von Abbrechern und weiterhin Teilnehmenden durch zwei unterschiedliche Parameter geschätzt werden müsste. Der Vergleich der Varianz-Kovarianz-Matrizen erlaubt demnach zum einen die Untersuchung der Stichprobenselektivität auf der Ebene der Varianzen und Kovarianzen und zum anderen die Überprüfung von statistischen Annahmen der logistischen Regression⁸.

Ähnlich wie bei den logistischen Regressionen wurde beim Vergleich der Varianzen und Kovarianzen überprüft, ob sich die Gruppe der weiterhin Teilnehmenden von der Gruppe der die Teilnahme vorzeitig beendenden Personen unterscheidet. Zu diesem Zweck wurden drei Kontraste definiert. Mit dem ersten Kontrast wurde überprüft, ob sich Personen, für die lediglich Informationen der Teilnahmeebene 1 vorliegen (N=644) von Personen unterscheiden, für die zusätzliche Informationen, also zumindest Informationen aus Teilnahmeebene 2, vorhanden sind (N=1.264). Der zweite Kontrast erfasst Unterschiede zwischen Personen, für die Informationen aus Teilnahmeebene 2, aber nicht aus Ebene 3 vorliegen (N=336), und Personen, für die zumindest eine vollständige Ersterhebung vorhanden ist (Teilnahmeebene 3 und höher; N=928). Schließlich dient der dritte Kontrast dem Vergleich zwischen Personen mit Ersterhebung, aber ohne vollständigem Intensivprotokoll (N=412) und Personen, die die Intensiverhebung vollständig durchlaufen haben (N=516). Für Personen, die ihre Bereitschaft zur Teilnahme an der Intensiverhebung bekundeten, aber keinen vollständigen Datensatz aus der Intensiverhebung aufweisen, wurde – im Gegensatz zu den logistischen Regressionen – wegen geringer Fallzahl (N=122) kein gesonderter Kontrast definiert.

8 Der Vergleich der Varianz-Kovarianz-Matrizen wurde mit dem für Strukturgleichungsmodelle entwickelten Statistikprogramm EQS durchgeführt (Bentler, 1989). Der Hauptvorteil dieses Programms gegenüber Standardverfahren (z. B. der Box-M-Test in der Prozedur von SPSSX) besteht darin, dass für einzelne Elemente oder auch für vom Anwender bestimmte Gruppen von Elementen der zu vergleichenden Matrizen Signifikanztests durchgeführt werden können. Ein Ausdruck der verwendeten Programme wird von den Autoren auf Anfrage zur Verfügung gestellt.

Auf eine vollständige Auflistung der statistischen Kennwerte wurde aus Gründen der Übersichtlichkeit verzichtet. Statt dessen werden Gruppenunterschiede auf einzelnen Variablen nur dann berichtet, wenn deren statistische Signifikanz auf einem α -Niveau von $p=0,01$ abgesichert werden konnte. Dichotome Variablen wie Familienstand wurden nicht berücksichtigt, da bei derartigen Merkmalen die Varianzunterschiede und die bereits zuvor geprüften Mittelwertsunterschiede statistisch voneinander abhängig sind. Als globales Maß der Anpassungsgüte des Modells, das die Gleichheit der Varianzen und Kovarianzen zwischen den Vergleichsgruppen zum Ausdruck bringt, wird der *Comparative Fit Index* aufgeführt (CFI; vgl. Bentler, 1989). Werte über 0,95 auf diesem Index gelten als ein Hinweis auf die hohe Anpassungsgüte des Modells.

4.3.1 Kontrast 1: Nur Teilnahmeebene 1 (N=644) versus höhere Teilnahmeebenen (N=1.264)

Überprüft wurde die Unterschiedlichkeit der Varianzen und Kovarianzen von Alter und Sozialindex (die einzigen nicht-dichotomen Merkmale auf dieser Ebene). Zu diesem Zweck wurde ein Modell spezifiziert, in dem die Varianzen und Kovarianzen in den beiden Gruppen mit jeweils *demselben* Parameter geschätzt wurden. Anschließend wurde überprüft, ob das Schätzen durch zwei *verschiedene* Parameter zu einer signifikanten Verbesserung der Anpassungsgüte des Modells führt. Dies war nicht der Fall ($\chi^2=5,25$; $df=3$; $p=0,15$). Die Anpassungsgüte des Modells war hoch (CFI=1,000).

4.3.2 Kontrast 2: Nur bis Teilnahmeebene 2 (N=336) versus höhere Teilnahmeebenen (N=928)

Überprüft wurde die Unterschiedlichkeit der Varianzen und Kovarianzen von Alter, Geschlecht, Bildung (Bildungsindex der Kurzbefragung), Lebenszufriedenheit (Item), subjektiver körperlicher sowie seelisch-geistiger Gesundheit. Es ergaben sich keine signifikanten Unterschiede zwischen den Gruppen. Die Anpassungsgüte des Modells war hoch (CFI=0,979).

4.3.3 Kontrast 3: Nur bis Teilnahmeebene 3 (N=412) versus höchste Teilnahmeebene (N=516)

Zusätzlich zu den unter 4.3.2 aufgeführten Variablen wurden die Variablen subjektive Gehstrecke, Äqui-

valenzeinkommen, Anzahl eng verbundener Personen, ADL-Wert, Body-Mass-Index, emotionale Ausgeglichenheit, Zufriedenheit mit dem eigenen Altern, Lebenszufriedenheit (Skala), Kurzform der Mini Mental State Examination (SMMCO), der Zahlen-Buchstaben-Test sowie der Depressivitäts-Score berücksichtigt (siehe auch Tabelle 1).

Der Overall-Test über alle 17 Varianzen war statistisch signifikant ($\chi^2=62,13$; $df=17$; $p<0,01$). Eine Inspektion der einzelnen Varianzen ergab, dass dieses signifikante Gesamtergebnis auf zwei Parameter zurückging: Erstens wies der Sozialindex des Wohngebietes in der Intensivstichprobe eine niedrigere Varianz auf als in der Gruppe der Personen, die nur bis Teilnahmeebene 3 gelangten ($s=1,07$ versus $s=1,15$; $\chi^2=26,24$; $df=1$; $p<0,001$). Zweitens wies das Äquivalenzeinkommen in der Intensivstichprobe eine höhere Varianz auf als in der Nur-Teilnahmeebene-3-Gruppe ($s=1.037$ versus $s=898$; $\chi^2=10,23$; $df=1$; $p<0,001$).

Bei den Kovarianzen ergab sich für den Overall-Test ebenfalls ein signifikanter Wert ($\chi^2=204,45$; $df=136$; $p<0,01$). Allerdings erwiesen sich von den 136 Einzelvergleichen lediglich drei als signifikant. Dabei handelte es sich um den Zusammenhang zwischen subjektiver körperlicher Gesundheit und Ausgeglichenheit ($\chi^2=11,38$; $df=1$; $p<0,01$), zwischen subjektiver seelisch-geistiger Gesundheit und Lebenszufriedenheit (Skala; $\chi^2=7,36$; $df=1$; $p<0,01$) sowie zwischen subjektiver Gehstrecke und Depressivitäts-Score ($\chi^2=9,73$; $df=1$; $p<0,01$). In allen drei Fällen war der Zusammenhang in der Intensivstichprobe stärker ausgeprägt als in der Nur-Teilnahmeebene-3-Gruppe.

Bei diesem Kontrast erbrachten also fünf von insgesamt 153 Vergleichen Hinweise auf das Vorhandensein von Gruppenunterschieden. Diese Hinweise sollten jedoch nicht überbewertet werden, da sich bei der großen Zahl der durchgeführten Vergleiche solche Unterschiede auch zufällig ergeben können. Wie bei den vorangegangenen Kontrasten überwiegt also auch hier der Eindruck einer großen Ähnlichkeit der Varianzen und Kovarianzen der beiden Vergleichsgruppen. Dies kommt erneut in der hohen Anpassungsgüte des Gleichheitsmodells zum Ausdruck (CFI=0,983).

4.4 Kumulative Auswirkungen der Selektivität auf ausgewählte Konstrukte der Intensiverhebung

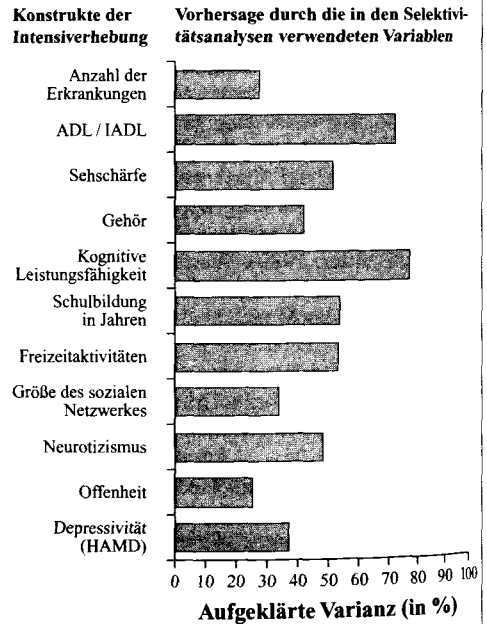
Die von Pearson (1903) und Lawley (1943) entwickelten Formeln zur Selektivität erlauben es, mit Hilfe

von linearen Regressionen statistische Kennwerte (Mittelwerte, Varianzen und Kovarianzen) für die Ausgangsstichprobe zu schätzen, die die beobachtete Selektivität berücksichtigen (Meredith, 1964; Muthén, Kaplan & Hollis, 1987; Smith, Holt & Smith, 1989). Es wird unterschieden zwischen unabhängigen Variablen, die auch als Selektionsvariablen bezeichnet werden, und abhängigen Variablen (d. h. Variablen, auf denen lediglich die weiterhin teilnehmenden Personen Werte aufweisen). Auf der Grundlage des linearen Zusammenhangs zwischen Selektions- und abhängigen Variablen sowie der Mittelwertsdifferenz zwischen der Ausgangsstichprobe und der selektierten Teilstichprobe auf den Selektionsvariablen werden die Mittelwerte, Varianzen und Kovarianzen der Ausgangsstichprobe auf den abhängigen Variablen geschätzt.

Das Pearson-Lawleysche Verfahren benutzt Antworten auf die zuvor gestellten Fragen über den Zusammenhang von Erwartungswerten (Mittelwerten, Häufigkeiten), Varianzen und Kovarianzen zusammen mit (empirisch nicht testbaren) Annahmen über die Linearität der Regression der abhängigen Variablen auf die Selektionsvariablen sowie über die Konstanz der bedingten Varianzen (Homoskedastizität). Unter diesen Annahmen erlaubt das Verfahren die Projektion von Selektivitätseffekten auf Konstrukte des Intensivprotokolls und führt somit zu einer direkten Abschätzung der Auswirkung von Selektivitätseffekten auf die zentralen Konstrukte der Berliner Altersstudie. Bei dieser Projektion wird (im Rahmen des linearen Modells) die gesamte verfügbare Information optimal genutzt.

Bereits Aitkin (1934) und Lawley (1943) haben darauf hingewiesen, dass die Formeln *mehrmals nacheinander* angewendet werden können. Für die Selektivitätsanalysen von BASE bedeutet dies, dass zunächst die auf der Teilnahmeebene 1 vorliegenden Variablen genutzt werden, um die Ausprägungen der Ausgangsstichprobe auf Variablen der Teilnahmeebene 2 zu schätzen. Anschließend dienen die Variablen der Teilnahmeebenen 1 und 2 (d. h. die beobachteten Werte von Ebene 1 und die geschätzten Werte von Ebene 2) als Selektionsvariablen, und die Variablen der Teilnahmeebene 3 fungieren als abhängige Variablen. Schließlich werden die Variablen der ersten drei Teilnahmeebenen zu Selektionsvariablen, mit denen sich die Ausprägungen der Ausgangsstichprobe auf den Konstrukten des Intensivprotokolls schätzen lassen⁹.

Abbildung 3: Die gemeinsame Vorhersagekraft der für die Selektivitätsanalysen genutzten Variablen in der Intensivstichprobe (N=516).

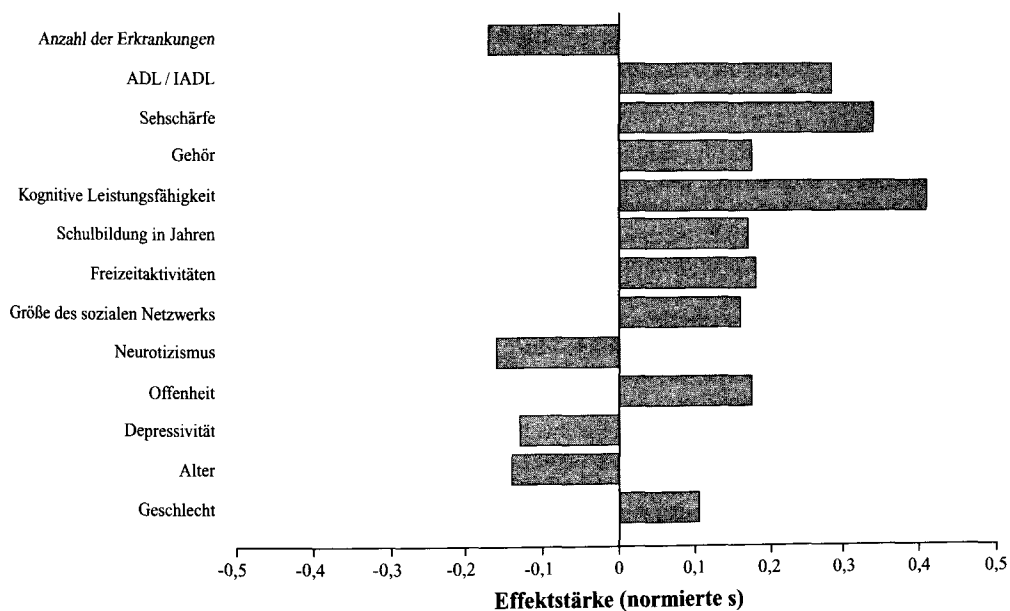


Anmerkung: Die Abbildung veranschaulicht das Ausmaß, in dem die in den Selektivitätsanalysen berücksichtigten Variablen individuelle Unterschiede auf zentralen Konstrukten der Intensivhebung vorherzusagen vermögen. Dargestellt ist also *nicht* das Ausmaß der Selektivitätseffekte, sondern der Zusammenhang zwischen Selektionsvariablen und Konstrukten der Intensivhebung.

Bei der Interpretation der mit diesem Verfahren erzielten Ergebnisse sollte bedacht werden, dass die berechneten Schätzwerte um so aussagekräftiger sind, je höher die Merkmale niedrigerer Teilnahmeebenen (d. h. die Selektionsvariablen) mit den Merkmalen der jeweiligen höheren Teilnahmeebene (d. h. den unabhängigen Variablen) zusammenhängen. Abhängige Variablen, für die es auf niedrigeren Teilnahmeebenen keine „Vorläufer“ gibt, können nicht korrigiert werden und behalten zwangsläufig den Mittelwert, den sie in der selektierten Teilstichprobe aufweisen. Diese Eigenschaft des Verfahrens reflektiert das (eingangs erwähnte) Paradox bei der Analyse von Selektivität; nur Münchhausen konnte sich an seinen eigenen Haaren aus dem Sumpf ziehen.

⁹ Für die Berechnung der Schätzwerte mit Hilfe der Pearson-Lawleyschen Formeln wurde das Statistikprogramm Mx (Neale, 1991) verwendet. Ein Ausdruck der verwendeten Programme wird von den Autoren auf Anfrage zur Verfügung gestellt.

Abbildung 4: Diskrepanz zwischen den beobachteten und den aufgrund der Pearson-Lawleyschen Formeln geschätzten, selektivitätskorrigierten Mittelwerten von zentralen Konstrukten der Intensiv-erhebung. Die Größe der Diskrepanz ist in normierten Standardabweichungen (s) dargestellt.



Aus diesem Grund wurde vorab überprüft, wie gut die im Pearson-Lawleyschen Schätzverfahren berücksichtigten Selektionsvariablen zur Varianzaufklärung in den ausgewählten Konstrukten der Intensiv-erhebung beitragen. Für diese Analyse wurden keine Werte geschätzt; sie beruht also auf der Intensivstichprobe ($N=516$).

Wie sich aus Abbildung 3 entnehmen lässt, war die Varianzaufklärung in den Konstrukten der Intensiv-erhebung generell recht hoch. Der niedrigste Wert (28% aufgeklärte Varianz) ergab sich für die Anzahl der mittel- bis schwergradigen Erkrankungen, der höchste Wert für die allgemeine kognitive Leistungsfähigkeit (78%). Bei einer (hier nicht im einzelnen berichteten) Inspektion der Korrelationen und Semipartialkoeffizienten zeigt sich, dass es jeweils die als „Vorläufer“ konzipierten Variablen sind, die in erster Linie zur Varianzaufklärung in den ihnen zugeordneten Konstrukten beitragen. Besonders hervorzuheben sind die im Rahmen der begleitenden Beobachtung erhobenen Ratings zu Demenz, Sensorik und ADL, die bereits auf Teilnahmeebene 2 erhoben wurden und wesentlich zur Varianzaufklärung in den entsprechenden Konstrukten der Intensiv-erhebung beitragen.

Die prädiktive Kraft der Selektionsvariablen reicht also durchaus aus, um – bei vorhandener Selektivität – zu Parameterschätzungen für die Ausgangsstichprobe auf den Konstrukten der Intensiv-erhebung zu gelangen, die deutlich von den für die Intensivstichprobe beobachteten Werten abweichen.

4.4.1 Auswirkungen der Selektivität auf die Gesamtstichprobe

Abbildung 4 zeigt die Diskrepanz zwischen den für die Ausgangsstichprobe geschätzten Mittelwerten und den beobachteten Mittelwerten der Intensiv-erhebung. Die Metrik des hierfür verwendeten Maßes ist die normierte Standardabweichung E_s ([Mittelwert der Ausgangsstichprobe minus Mittelwert der selektierten Stichprobe] / Standardabweichung der Ausgangsstichprobe; vgl. Hedges & Olkin, 1985). Sie gibt an, wie weit der Mittelwert der selektierten Stichprobe vom Mittelwert der Ausgangsstichprobe entfernt ist. Ein Wert von $E_s=0,5$ würde demnach bedeuten, dass sich der Mittelwert der selektierten Stichprobe eine halbe Standardabweichung über dem der Ausgangsstichprobe befindet. Positive Werte ste-

hen somit für Fälle, in denen die beobachteten Mittelwerte über den für die Ausgangsstichprobe geschätzten Werten liegen. Die Variablen Alter und Geschlecht (0=Männer, 1=Frauen) sind ebenfalls aufgeführt, da einige der gezeigten Effekte mit den Designvariablen zusammenhängen könnten. Im Fall von Alter und Geschlecht konnten die Effektstärken direkt berechnet werden, da diese Merkmale für alle Studienteilnehmer bekannt sind.

Das Muster der Effekte bestätigt das aus der Literatur bekannte Bild: Personen mit höherer kognitiver Leistungsfähigkeit, besserem Seh- und Hörvermögen, selbständigerer Lebensführung (ADL/IADL), einer niedrigeren Anzahl von Erkrankungen sowie höherer Schulbildung weisen eine erhöhte Wahrscheinlichkeit auf, die Intensiverhebung vollständig zu durchlaufen. Auch die Ergebnisse für weniger häufig untersuchte Bereiche wie Größe des sozialen Netzwerks, Ausmaß an Freizeitaktivitäten, Neurotizismus, Offenheit und Depressivität weisen in diese Richtung.

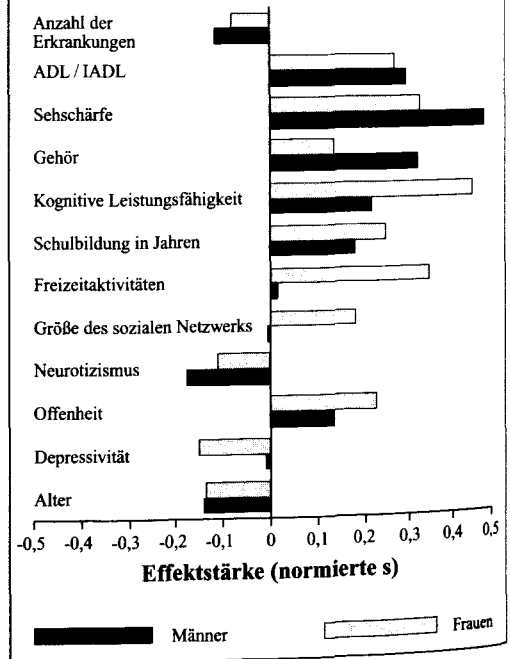
Alle Effekte sind statistisch signifikant ($p < 0,01$). Zu fragen ist nach der inhaltlichen Bedeutsamkeit der einzelnen Effektgrößen. Unter der Annahme der Normalverteilung bedeutet eine Effektstärke von 0,14 (die kleinste Effektstärke in Abb. 4 für Depressivität), dass sich 56% der Beobachtungen der Intensiverhebung – statt der nach dem Zufall erwarteten 50% – über dem Mittelwert der Ausgangsstichprobe befinden. Bei einer Effektstärke von 0,42 (die größte Effektstärke für kognitive Leistungsfähigkeit) sind es 66%. Die Effektstärken stehen also für relativ kleine Abweichungen vom geschätzten Erwartungswert.

Nicht aufgeführt in Abbildung 4 sind die geschätzten Häufigkeitsverteilungen für die klinischen Diagnosen Demenz und Depression. Für die Demenz ergab sich ein Erwartungswert von 30%, der deutlich über dem beobachteten Wert von 21% liegt. Bei der Depression kam es zu keinen wesentlichen Verschiebungen (geschätzter Wert für die Ausgangsstichprobe: 27%; beobachteter Wert: 26%).

4.4.2 Auswirkungen der Selektivität auf Geschlechtsunterschiede

Abbildung 5 bezieht sich auf die Frage möglicher Geschlechtsunterschiede im Ausmaß der Selektivität. Am deutlichsten sind Geschlechtsunterschiede auf den Konstrukten Freizeitaktivitäten (Männer: 0,02; Frauen: 0,35), kognitive Leistungsfähigkeit (Männer: 0,22; Frauen: 0,45); Netzwerkgröße (Männer: -0,01; Frauen: 0,19). Diese Ergebnisse legen nahe, dass die Frauen der Intensiverhebung in Bezug auf diese drei

Abbildung 5: Diskrepanz zwischen den beobachteten und den aufgrund der Pearson-Lawleyschen Formeln geschätzten, selektivitätskorrigierten Mittelwerten von zentralen Konstrukten der Intensiverhebung, getrennt für Männer und Frauen. Die Größe der Diskrepanz ist in normierten Standardabweichungen (s) dargestellt.

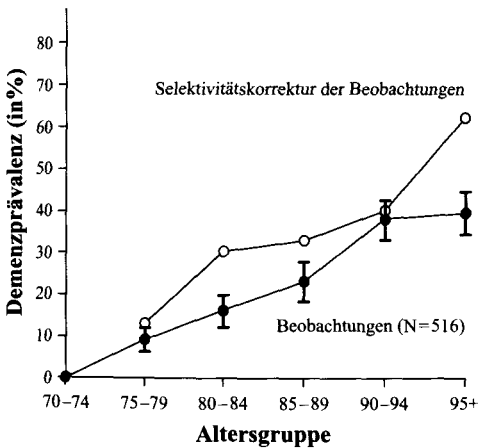


Konstrukte eine stärker positiv selektierte Stichprobe darstellen als die Männer.

4.4.3 Auswirkungen der Selektivität auf Altersunterschiede

Eine Inspektion der nach Altersgruppen getrennt berechneten Effektstärken ergab zumeist keine deutlichen Hinweise auf altersdifferentielle Selektivität. Eine Ausnahme bildet die klinische Diagnose der Demenz (siehe auch Helmchen et al., Kapitel 7). Abbildung 6 zeigt die beobachteten und geschätzten Werte für die Demenzprävalenz. Betrachtet man die beobachteten Werte, so deuten sie darauf hin, dass die Prävalenz in den beiden höchsten Altersgruppen nicht weiter ansteigt. Die geschätzten Werte hinge-

Abbildung 6: Mögliche Auswirkungen der Selektivität auf Altersunterschiede der Demenzprävalenz in der Berliner Altersstudie.



Anmerkung: Die ausgefüllten Kreise bezeichnen die auf der Ebene der Intensivstichprobe (N=516) beobachtete Häufigkeit dementieller Syndrome (siehe auch Helmchen et al., Kapitel 7). Die offenen Kreise bezeichnen die nach Berücksichtigung der Stichprobenselektivität geschätzten Häufigkeiten. Die Fehlerbalken beziehen sich auf die Standardfehler der beobachteten Prozentwerte.

gen stützen die Annahme, dass die Demenzprävalenz auch nach dem 90. Lebensjahr weiter zunimmt. In diesem Fall führt die Berücksichtigung der Selektivität also zu unterschiedlichen inhaltlichen Aussagen. Eine Inspektion der Regressionskoeffizienten zeigt, dass auf Teilnahmeebene 2 das Beobachtungs-Rating der Demenz und auf Teilnahmeebene 3 der SMMCO sowie der Zahlen-Buchstaben-Test die größten Beiträge zur Vorhersage der geschätzten Werte leisteten.

5. Zusammenfassung der Ergebnisse und abschließende Interpretation

Bevor wir die Ergebnisse der Selektivitätsanalysen zusammenfassen und kommentieren, wollen wir noch einmal ausdrücklich auf die Grenzen derartiger Analysen hinweisen. Selektivitätsanalysen machen Aussagen zum Ausmaß der beobachteten und nicht zum Ausmaß der prinzipiell denkbaren Selektivität. Sie versuchen, vorhandene Daten aufeinander zu beziehen und die in ihnen verborgene Information

optimal auszuschöpfen, können jedoch auf Fragen, deren Beantwortung zusätzliche Informationen erfordert, keine abschließende Antwort geben. In diesem Zusammenhang ist insbesondere der Übergang von der ersten (N=1.908) zur zweiten (N=1.264) Teilnahmeebene zu erwähnen, der nur durch wenige Variablen dokumentiert ist. Die Frage, ob beim Vorhandensein weiterer Variablen auf der ersten Teilnahmeebene auch für diesen Übergang (und somit insgesamt) ein größeres Ausmaß an Selektivität beobachtet worden wäre, lässt sich nicht schlüssig beantworten.

Im Bewusstsein dieser in der Sache liegenden Grenzen lassen sich die Ergebnisse der Selektivitätsanalysen wie folgt zusammenfassen:

1. Die Selektivitätsanalysen bieten keine Anhaltspunkte dafür, dass sich Zusammenhangsmuster und Variabilität der Merkmale der Intensivstichprobe (N=516) von den Stichproben niedrigerer Teilnahmeebenen wesentlich unterscheiden. Insbesondere gibt es mit einer Ausnahme – einer leichten Abnahme der Varianz des Wohngebiets-Sozialindex – keinen Hinweis auf eine Reduktion der Varianzen. Die Vielfältigkeit der Altersformen und Lebenslagen sowie die Struktur der Merkmalszusammenhänge bleiben demnach in der Intensivstichprobe weitgehend erhalten. Unter dem Gesichtspunkt der Selektivität scheint die Generalisierbarkeit von Aussagen über Merkmalszusammenhänge also zulässig zu sein. Dieses Ergebnis ist insbesondere für Analysen wichtig, bei denen Fragen nach Zusammenhängen zwischen verschiedenen Funktionsbereichen im Vordergrund stehen.

2. Im Gegensatz zu den Varianzen und Kovarianzen gibt es bei den Mittelwerten und Häufigkeitsverteilungen einiger Merkmale Hinweise auf Selektivitätseffekte. Gemessen an der in den Analysen verwendeten Anzahl untersuchungsrelevanter Merkmale ist die Anzahl dieser Effekte jedoch eher gering. Über alle Stufen der Teilnahme betrachtet gibt es Hinweise auf eine positive Selektion nach sozialer Schicht, subjektiver Beurteilung der seelisch-geistigen Gesundheit, Lebenszufriedenheit sowie kognitiver Leistungsfähigkeit und Demenz. Bei den Gesundheitsmaßen finden sich, abgesehen vom Body-Mass-Index auf der letzten Teilnahmeebene, keine Hinweise auf systematische Ausfälle. Die höhere Teilnahmewahrscheinlichkeit von Heimbewohnern, die wahrscheinlich auf die bessere Erreichbarkeit dieser Gruppe zurückgeführt werden kann, ist ein Hinweis auf gegenläufige Selektivitätseffekte. Möglicherweise wird eine positive Selektion hinsichtlich Gesundheit und kognitiver

Leistungsfähigkeit durch die relativ hohe Teilnahmerate von Personen in stationären Altenhilfeeinrichtungen teilweise ausgeglichen.

Überraschenderweise besteht zwischen dem Alter und der Teilnahmewahrscheinlichkeit auf keiner der Teilnahmeebenen ein starker Zusammenhang. Dieses Ergebnis scheint im Widerspruch zu dem aus Abbildung 1 ersichtlichen Muster eines mit der Altersgruppe variierenden Stichprobenausfalls zu stehen. Es ist jedoch zu bedenken, dass mit den hier berichteten logistischen Regressionen lediglich lineare Alterseffekte erfasst wurden (d. h. eine lineare Zu- oder Abnahme des Stichprobenausfalls mit dem Alter). Detailliertere Analysen der in Abbildung 1 dargestellten Daten ergaben Hinweise auf *nichtlineare* Zusammenhänge (Gilberg & Pötter, 1994). So ist z. B. der Stichprobenausfall in der Altersgruppe der 85- bis 89jährigen stärker ausgeprägt als in den beiden benachbarten Altersgruppen.

Deutlichere Selektivitätseffekte finden sich demgegenüber beim Geschlecht. Aufgrund der durch den Erhebungsplan garantierten Gleichverteilung von Männern und Frauen hat die niedrigere Teilnahmewahrscheinlichkeit von Frauen jedoch keinen Einfluss auf das Zahlenverhältnis von Männern zu Frauen in der Intensivstichprobe.

3. *Die Teilnehmer der BASE-Intensivstichprobe weisen gegenüber der Ausgangsstichprobe eine deutlich verringerte Mortalität auf* (siehe Abb. 2). Dieser Befund lässt sich zu einem großen Teil, aber nicht vollständig, auf die zeitliche Ausdehnung des Intensivprotokolls zurückführen. Die weiterführende Interpretation dieser mortalitätsbezogenen Stichprobenselektivität erfordert die Identifikation von Einflussgrößen, die sowohl die Teilnahmetiefe als auch die Sterberate vorhersagen. Dabei ist zu bedenken, dass sich im untersuchten Altersbereich bereits die Zusammensetzung der Grundgesamtheit mit dem Alter so sehr verändert, dass die Annahme einer über das Alter hinweg einheitlichen Ausgangsstichprobe eigentlich nicht zu halten ist. So erreichen in der Bundesrepublik ungefähr 81% der Frauen und 66% der Männer das Alter 70, jedoch nur etwa 18% der Frauen und 7% der Männer das Alter 90 (Statistisches Bundesamt, 1993). Es ist bekannt, dass individuelle Unterschiede in der Langlebigkeit mit persönlichen Merkmalen wie Gesundheitszustand, kognitivem Leistungsniveau und Wohlbefinden korrelieren (Siegler & Botwinick, 1979). Mortalität ist somit ein selektiver Prozess, und altersbezogene Auswertungen des BASE-Datensatzes

vermengen (konfundieren) den Alterungsprozess der Überlebenden und die selektive Wirkung der Mortalität (vgl. Kruse et al., 1993; Mulder, 1993). Die Beziehungen zwischen diesen Prozessen sollen in weiterführenden Analysen, die auch längsschnittliche Daten sowie Sterblichkeitsinformationen über mehrere Jahre berücksichtigen, näher untersucht werden.

4. *Die kumulative Analyse von Selektivität und ihre Projektion auf die Konstrukte der Intensivhebung mit Hilfe der Pearson-Lawleyschen Formeln ergibt, dass die Intensivstichprobe in allen in Betracht genommenen Bereichen positiv selektiert ist.* Das Ausmaß dieser Selektivitätseffekte überschreitet jedoch bei keinem der Konstrukte eine halbe Standardabweichung und variiert nur unwesentlich mit dem Alter. Für die Interpretation der Ergebnisse anderer Kapitel bedeutet dies, dass die auf der Ebene des Intensivprotokolls empirisch vorgefundenen Werte das Altsein und Altwerden tendenziell in einem etwas zu günstigen Licht erscheinen lassen.

Eine bedeutsame Abweichung von diesem Bild einer eher gering ausgeprägten und relativ alterskonstanten Selektivität bildet die Demenzprävalenz. Hier ergibt sich eine relativ deutliche Abweichung zwischen dem in der Intensivstichprobe beobachteten (21%) und dem mit Hilfe der Pearson-Lawleyschen Formeln geschätzten (30%) Wert. Offensichtlich führt der Stichprobenausfall zu einer Unterschätzung der Demenzprävalenz in der Ausgangsstichprobe. Diese Unterschätzung ist bei den 95jährigen und Älteren besonders stark ausgeprägt (40% versus 62%).

Leider ist es kaum möglich, das Ausmaß der Selektivität in der Berliner Altersstudie direkt mit den Ergebnissen anderer Untersuchungen zu vergleichen. Zum einen wird ein derartiger Vergleich durch zahlreiche methodische und inhaltliche Unterschiede zwischen den Studien erschwert. Zum anderen sind Angaben zur Selektivität in vielen Fällen lückenhaft und unvollständig, so dass über die Größe der Effekte lediglich Mutmaßungen angestellt werden können. Insgesamt ergibt sich jedoch der Eindruck, dass sich das Ausmaß der Selektivitätseffekte der Berliner Altersstudie eher im Rahmen des Üblichen bewegt. Angesichts der Größe des Stichprobenausfalls (siehe Abb. 1) und des hohen Altersdurchschnitts der Teilnehmer ist dies ein durchaus zufriedenstellendes Ergebnis.

Literaturverzeichnis

- Aitkin, A. C. (1934). Note on selection from a multivariate normal population. *Proceedings of the Edinburgh Mathematical Society*, 4, 106–110.
- Aldrich, J. H. & Nelson, F. D. (1984). *Linear probability, logit and probit models*. Beverly Hills, CA: Sage.
- Baltes, P. B., Mayer, K. U., Helmchen, H. & Steinhagen-Thiessen, E. (1993). The Berlin Aging Study (BASE): Overview and design. *Ageing and Society*, 13, 483–515.
- Baltes, P. B., Reese, H. W. & Nesselroade, J. R. (1988). *Life-span developmental psychology: Introduction to research methods*. Hillsdale, NJ: Erlbaum.
- Baltes, P. B., Schaie, K. W. & Nardi, A. H. (1971). Age and experimental mortality in a seven-year longitudinal study of cognitive behavior. *Developmental Psychology*, 5, 18–26.
- Bentler, P. M. (1989). *EQS: Structural equations manual (Version 3.0)*. Los Angeles, CA: BMDP Statistical Software, Inc.
- Cassel, C. M., Särndal, C. E. & Wretman, J. H. (1977). *Foundations of inference in survey sampling*. New York: Wiley.
- Cooney, T. M., Schaie, K. W. & Willis, S. K. (1988). The relationships between prior functioning on cognitive and personality dimensions and subject attrition in longitudinal research. *Journal of Gerontology: Psychological Sciences*, 43, P12–P17.
- Dalenius, T. (1988). A first course in survey sampling. In P. Krishnaiah & C. R. Rao (Hrsg.), *Handbook of statistics: Sampling* (Bd. 6, S. 15–46). Oxford: North-Holland.
- DeMaio, T. J. (1980). Refusals: Who, where and why. *Public Opinion Quarterly*, 44, 223–233.
- Esser, H., Grohmann, H., Müller, W. & Schäffer, K. H. (1989). *Mikrozensus im Wandel*. Stuttgart: Metzler-Poeschel.
- Eye, A. von (1989). Zero-missing non-existing data: Missing data problems in longitudinal research and categorical data solutions. In M. Brambring, F. Lösel & H. Skowronek (Hrsg.), *Children at risk: Assessment, longitudinal research, and intervention* (S. 336–355). Berlin: de Gruyter.
- Gilberg, R. & Pötter, U. (1994). *Poststratifizierungsgewichte für die BASE-Studie*. Unveröffentlichtes Manuskript, Max-Planck-Institut für Bildungsforschung, Berlin.
- Glynn, R. J., Laird, N. M. & Rubin, D. B. (1993). Multiple imputation in mixture models for nonignorable nonresponse with follow-ups. *Journal of the American Statistical Association*, 88 (423), 984–597.
- Goudy, W. J. (1976). Nonresponse effects on relationships between variables. *Public Opinion Quarterly*, 40, 360–369.
- Haggstrom, G. W. (1983). Logistic regression and discriminant analysis by ordinary least squares. *Journal of Business and Economic Statistics*, 1, 229–238.
- Hawkins, D. F. (1975). Estimation of nonresponse bias. *Sociological Methods and Research*, 3, 461–488.
- Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hertzog, C., Schaie, K. W. & Gribbin, K. (1978). Cardiovascular disease and changes in intellectual functioning from middle to old age. *Journal of Gerontology*, 33, 872–883.
- Herzog, A. R. & Rodgers, W. L. (1988). Age and response rates to interview sample surveys. *Journal of Gerontology: Social Sciences*, 43, S200–S205.
- Kessler, R. C., Little, R. J. & Groves, R. M. (1995). Advances in strategies for minimizing and adjusting for survey non-response. *Epidemiologic Reviews*, 17, 192–204.
- Kruse, A., Lindenberger, U. & Baltes, P. B. (1993). Longitudinal research on human aging: The power of combining real-time, microgenetic, and simulation approaches. In D. Magnusson & P. Casaer (Hrsg.), *Longitudinal research on individual development: Present status and future perspectives* (S. 153–193). Cambridge: Cambridge University Press.
- Kruskal, W. & Mosteller, F. (1979a). Representative sampling: I. Non-scientific literature. *International Statistical Review*, 47, 13–24.
- Kruskal, W. & Mosteller, F. (1979b). Representative sampling: II. Scientific literature, excluding statistics. *International Statistical Review*, 47, 111–127.
- Kruskal, W. & Mosteller, F. (1979c). Representative sampling: III. The current statistical literature. *International Statistical Review*, 47, 245–265.
- Kühnel, S. M., Jagodzinski, W. & Terwey, M. (1989). Teilnehmen oder boykottieren: Ein Anwendungsbeispiel der binären logistischen Regression mit SPSSX. *ZA-Information*, 25, 44–75.
- Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh*, 62, 28–30.
- Little, R. & Rubin, D. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Lowe, F. E. & McCormick, T. C. (1955). Some survey sampling biases. *Public Opinion Quarterly*, 19, 303–315.

- Manton, K. G. & Woodbury, M. A. (1983). A mathematical model of the physiological dynamics of aging and correlated mortality selection: II. Application to Duke Longitudinal Study. *Journal of Gerontology*, 38, 406–413.
- McArdle, J. J. & Hamagami, F. (1991). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. In L. M. Collins & J. L. Horn (Hrsg.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (S. 276–298). Washington, DC: American Psychological Association.
- McArdle, J. J., Hamagami, F., Elias, M. F. & Robbins, M. A. (1991). Structural modeling of mixed longitudinal and cross-sectional data. *Experimental Aging Research*, 17, 29–51.
- Meinlschmidt, G., Imme, U. & Kramer, R. (1990). *Sozialstrukturatlas Berlin (West): Eine statistisch-methodische Analyse mit Hilfe der Faktorenanalyse*. Berlin: Senatsverwaltung für Gesundheit und Soziales.
- Mercer, J. R. & Butler, E. W. (1967/68). Disengagement of the aged population and response differentials in survey research. *Social Forces*, 46, 89–96.
- Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, 29, 177–185.
- Mulder, P. G. H. (1993). The simultaneous processes of ageing and mortality. *Statistica Neerlandica*, 47, 253–267.
- Muthén, B. O., Kaplan, D. & Hollis, M. (1987). On structural equation modeling with data that are not completely missing at random. *Psychometrika*, 52, 431–462.
- Neale, M. C. (1991). *Mx: Statistical modelling*. Box 3MCV, Richmond, CA 23298: Department of Human Genetics.
- Norris, F. H. (1985). Characteristics of nonrespondents over five waves of a panel study. *Journal of Gerontology*, 40, 627–636.
- Oh, H. L. & Scheuren, F. J. (1983). Weighting adjustment for unit non-response. In W. G. Madow, I. Olkin & D. B. Rubin (Hrsg.), *Incomplete data in sample surveys: Theory and bibliographies* (Bd. 2, S. 143–184). New York: Academic Press.
- Panel on Incomplete Data (1983). Part I: Report. In W. G. Madow, I. Olkin & D. B. Rubin (Hrsg.), *Incomplete data in sample surveys: Report and case studies* (Bd. 1, S. 3–106). New York: Academic Press.
- Pearson, K. (1903). Mathematical contributions to the theory of evolution: XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society of London (Series A)*, 200, 1–66.
- Powell, D. A., Furchtgott, E., Henderson, M., Prescott, L., Mitchell, A., Hartis, P., Valentine, J. D. & Milligan, W. L. (1990). Some determinants of attrition in prospective studies on aging. *Experimental Aging Research*, 16, 17–24.
- Powers, E. A. & Bultena, G. L. (1972). Characteristics of deceased dropouts in longitudinal research. *Journal of Gerontology*, 27, 350–353.
- Rendtel, U. & Pötter, U. (1992). *Über Sinn und Unsinn von Repräsentativitätsstudien* (DIW-Diskussionspapier Nr. 61). Berlin: Deutsches Institut für Wirtschaftsforschung.
- Rudinger, G. & Wood, P. K. (1990). N's, times and number of variables in longitudinal research. In D. Magnusson & R. L. Bergman (Hrsg.), *Data quality in longitudinal research* (S. 157–180). Cambridge: Cambridge University Press.
- Schaie, K. W. (1983). The Seattle Longitudinal Study: A twenty-one year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Hrsg.), *Longitudinal studies of adult psychological development* (S. 64–135). New York: Guilford.
- Schaie, K. W., Labouvie, G. V. & Barrett, T. J. (1973). Selective attrition effects in a fourteen-year study of adult intelligence. *Journal of Gerontology*, 28, 328–334.
- Shock, N. W., Greulich, R. C., Costa, P. T., Jr., Andres, R., Lakatta, E. G., Arenberg, D. & Tobin, J. D. (1984). *Normal human aging: The Baltimore Longitudinal Study on Aging* (NIH Publication Nr. 84-2450). Washington, DC: Government Printing Office.
- Siegler, I. C. & Botwinick, J. (1979). A long-term longitudinal study of intellectual ability of older adults: The matter of selective subject attrition. *Journal of Gerontology*, 34, 242–245.
- Smith, C. J., Holt, D. & Smith, T. M. F. (1989). *Analysis of complex surveys*. New York: Wiley.
- Statistisches Bundesamt (1993). *Statistisches Jahrbuch 1993*. Wiesbaden: Metzler-Poeschel.
- Streib, G. F. (1966). Participants and dropouts in a longitudinal study. *Journal of Gerontology*, 21, 200–209.
- Tennstedt, S. L., Dettling, U. & McKinlay, J. B. (1992). Refusal rates in a longitudinal study of older people: Implications for field methods. *Journal of Gerontology: Social Sciences*, 47, S313–S318.
- Weaver, C. N., Holmes, S. L. & Glenn, N. D. (1975). Some characteristics of inaccessible respondents in a telephone survey. *Journal of Applied Psychology*, 60, 260–262.