

GESIS

ZUMA

53

November 2003

NACHRICHTEN

Zentrum für Umfragen, Methoden und Analysen (ZUMA)

ZUMA ist Mitglied der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS). Die GESIS ist eine Einrichtung der *Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz* (WGL) (www.wgl.de).

Vorsitzender des Trägervereins ZUMA e.V.: Prof. Dr. Jan van Deth

Direktor: Prof. Dr. Peter Ph. Mohler

Hausanschrift

B 2, 1
68 159 Mannheim

Postanschrift

Postfach 12 21 55
68 072 Mannheim

Telefon 0621/1246 - 0

Fax 0621/1246 - 100

Redaktion: 0621/1246 - 228

E-Mail porst@zuma-mannheim.de

Internet www.gesis.org/zuma ZUMA
www.gesis.org/ GESIS

Redaktion: Rolf Porst

ISSN 0941-1670 27. Jahrgang

Die ZUMA-Nachrichten informieren über Forschungsarbeiten und Dienstleistungen des Instituts. Beiträge externer Autoren werden dann aufgenommen, wenn sie aus Kooperationsprojekten mit ZUMA resultieren oder von besonderem wissenschaftlichen Interesse für die Arbeiten im Hause und die Empirische Sozialforschung sind. Alle Beiträge unterliegen einem internen förmlichen Begutachtungsverfahren durch jeweils zwei Gutachter. Der Herausgeber entscheidet auf der Grundlage der Gutachten über die Annahme eines Beitrages. Hinweise zur Manuskriptgestaltung sind auf der letzten Seite des Heftes.

Namentlich gekennzeichnete Beiträge geben die Meinung der Autoren wieder. Der Nachdruck von Beiträgen ist nach Absprache möglich. Die ZUMA-Nachrichten erscheinen im Mai und November eines Jahres. Sie werden Interessenten auf Anforderung kostenlos zugesandt. Die ZUMA-Nachrichten finden Sie auch im WWW:
http://www.gesis.org/publikationen/zeitschriften/zuma_nachrichten/.

Druck: Verlag Pfälzische Post GmbH, Neustadt/Weinstraße.
Gedruckt auf chlorfrei gebleichtem Papier.

© ZUMA, Mannheim, 2003

In eigener Sache..... 5

In Memoriam Prof. Dr. Erwin K. Scheuch..... 7

DISKUSSIONSBEITRAG

Alles Zufall oder? Ein Diskussionsbeitrag zur Qualität von face to face-Umfragen in Deutschland.

Peter Mohler, Achim Koch & Siegfried Gabler..... 10

FORSCHUNGSBERICHTE

Woran bemisst sich eine ‚gute‘ allgemeine Bevölkerungsumfrage? Analysen zu Ausmaß, Bedeutung und zu den Hintergründen von Nonresponse in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS.

Ulrich Schneekloth & Ingo Leven 16

Offline rekrutierte Access Panels: Königsweg der Online-Forschung?

Thorsten Faas 58

Item-Nonresponse in Mitarbeiterbefragungen.

Ingwer Borg & Christian Treder 77

PROJEKTBERICHT

SHARE – Measuring the Ageing Process in Europe.

Axel Börsch-Supan, Hendrik Jürges & Oliver Lipps..... 96

MITTEILUNGEN

„Stellung im Beruf“ als Ersatz für eine Berufsklassifikation zur Ermittlung von sozialem Prestige.

Jürgen H.P. Hoffmeyer-Zlotnik..... 114

Ergebnisse der zweiten Befragung von Nutzern der Mikrozensus-Scientific Use Files.

Paul Lüttinger, Holger Breiholz & Yvonne Lechert..... 128

Allgemeine Bevölkerungsumfrage der Sozialwissenschaften: Daten des ALLBUS 2002 verfügbar.

Michael Blohm..... 136

QUEST – Questionnaire Evaluation Standards. <i>Peter Prüfer & Margrit Rexroth</i>	142
Pretest-Labor bei ZUMA. <i>Peter Prüfer & Margrit Rexroth</i>	143
The International Workshop on Comparative Survey Design and Implementation.....	144
European Centre for Cross-Cultural Surveys Established at ZUMA	145
The International Social Survey Programme.....	146
Scientific Use File des Mikrozensus 2000 für die Wissenschaft zugänglich.....	147
European Social Survey: Erste Länderdaten der Erhebung 2002/2003 verfügbar	149
Call for Papers: Journal of Official Statistics (JOS).....	150
Call for Papers: GOR04. Sixth International GOR Conference. German Online Research '04.....	151

BUCHBESPRECHUNG

Biemer, Paul P. & Lars E. Lyberg: Introduction to Survey Quality.....	153
---	-----

ZUMA-PUBLIKATIONEN

ZUMA-Arbeitsberichte.....	156
ZUMA-Methodenberichte.....	157

ZUMA-VERANSTALTUNGEN 2004

ZUMA-workshops: Erstes Halbjahr 2004	158
--	-----

DURCHWAHL-RUFNUMMERN

IN EIGENER SACHE

Die Nachricht vom Tode von Prof. Erwin K. Scheuch am 12. Oktober 2003 hat bei ZUMA große Betroffenheit hervorgerufen. In seinem Nachruf würdigt MAX KAASE, mit Scheuch über Jahrzehnte professionell und persönlich tief verbunden, dessen Bedeutung für die empirische Sozialforschung im allgemeinen und für GESIS und ZUMA im besonderen.

ZUMA und seine Mitarbeiterinnen und Mitarbeiter nehmen in tiefer Trauer, aber auch in Anerkennung seines umfassenden Lebenswerkes und in großer Dankbarkeit für sein nachhaltiges Engagement für GESIS und ZUMA Abschied von Erwin K. Scheuch.

Es fällt schwer, zum Tagesgeschehen zurückzukehren.

Die Ausschöpfung gilt gemeinhin als eines der wichtigsten Merkmale zur Beurteilung der Qualität repräsentativer allgemeiner Bevölkerungsumfragen. Anhand von Daten aus ALLBUS-Umfragen zeigen ULRICH SCHNEEKLOTH & INGO LEVEN, dass die üblicherweise geforderten hohen Ausschöpfungsquoten bei harter Dokumentation der Bearbeitung faktisch nicht realisierbar sind. Sie gehen deshalb auch der Frage nach, ob und inwieweit eine gemeinhin als „schlecht“ angesehene Ausschöpfung von unter 50% tatsächlich mangelnde Datenqualität nach sich zieht. In einem Diskussionsbeitrag zum Thema vertreten MOHLER, KOCH & GABLER die Ansicht, dass die Qualität von Umfragen – wie bei jedem anderen Produkt auch – von der Qualität ihres Produktionsprozesses abhängig sei, und dass man deshalb viel mehr als bisher auf den Produktionsprozess für das „Produkt Umfrage“ und dessen Verbesserung zu achten habe.

Auch THORSTEN FAAS beschäftigt sich mit der Repräsentativität von Umfragedaten, allerdings bei der speziellen Zielgruppe der Internet-Nutzer. Er vergleicht die Internet-Nutzer aus einem offline rekrutierten Access Panel mit den Internet-Nutzern, die aus einer persönlich-mündlich durchgeführten Repräsentativ-Befragung der deutschen Bevölkerung gewonnen wurden, und kommt zu dem Schluss, dass Online-Umfragen auf der Basis von Access Panels nicht zwangsläufig Repräsentativität für alle Internet-Nutzer beanspruchen können.

INGWER BORG & CHRISTIAN TREDER widmen sich anhand von Daten aus neun großen Mitarbeiterbefragungen dem Problem, ob die Nichtbeantwortung einzelner Items systematisch mit Einstellungen, so wie sie sich aus anderen Items schätzen lassen, zusammenhängt. Damit gehen sie der verbreiteten Praxisvermutung nach, dass Item-Nonresponse in Mitarbeiterbefragungen auf geringes Commitment, niedrigere Zufriedenheit oder ähnliches hindeute. Weiterhin behandeln sie die Frage, ob Item-Nonresponse rein zufällig erfolgt. Ihre Ergebnisse bestätigen beide Hypothesen nicht.

Bei den Mitteilungen möchte ich besonders hinweisen auf den Beitrag von JÜRGEN HOFFMEYER-ZLOTNIK über die „Stellung im Beruf“ als Ersatz für eine Berufsklassifikation zur Ermittlung sozialen Prestiges sowie auf die Ergebnisse der zweiten Befragung von Nutzern der Mikrozensus-Scientific Use Files in dem Beitrag von LÜTTINGER, BREIHZOLZ & LECHERT.

Als neue wissenschaftliche Leiterin für den Bereich „Dauerbeobachtung“ hat URSULA HOFFMANN-LANGE am 1. August 2003 die Arbeit bei ZUMA aufgenommen. Frau HOFFMANN-LANGE wird in dieser Funktion bis 31. Juli 2004 tätig sein, um danach wieder an die Universität Bamberg zurückzukehren, an der sie eine Professur für Politikwissenschaft innehat. ZUMA kennt sie sehr gut, war sie doch von 1975 bis 1980 als Projektleiterin bei ZUMA tätig und auch auf den weiteren Stationen ihres beruflichen Werdeganges (u.a. wissenschaftliche Mitarbeiterin beim ZUMA-Gründer Rudolf Wildenmann oder Leiterin der Abteilung „Jugend und Politik“ am Deutschen Jugendinstitut in München) immer eng mit ZUMA verbunden gewesen.

Zurück bei ZUMA ist auch CAROLINE KRAMER, die nach einem dreijährigen Habilitationsstipendium ihre Arbeit in der Abteilung Sozialindikatoren wieder aufgenommen hat.

Neu bei ZUMA ist WOLFGANG NEUBARTH, der als Nachfolger von MICHAEL BOSNJAK im Forschungsfeld der Internet-Umfragen tätig ist. Schließlich hat auch die Verwaltung Verstärkung bekommen: SIMONE BORDUNE ist seit 1. Oktober für Reisekosten und Hausverwaltung zuständig.

ZUMA verlassen haben BERNHARD CHRISTOPH (Abteilung Sozialindikatoren) und MICHAEL SEE (Information & Kommunikation) .

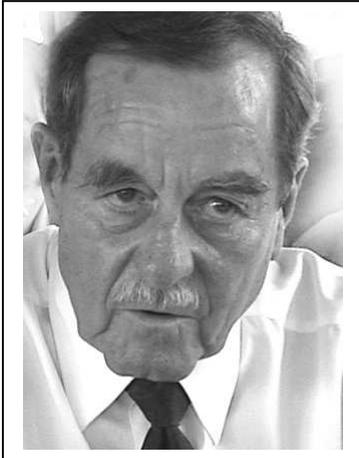
Als Nachfolger von PAUL LÜTTINGER hat ROLF PORST, bei ZUMA ansonsten als Projektberater tätig, mit der vorliegenden Ausgabe die Aufgaben des Redakteurs der ZUMA-Nachrichten übernommen.

Mit dem vorliegenden Heft endet die Tradition, dass der Vorsitzende des ZUMA e.V. auch Herausgeber der ZUMA-Nachrichten ist. Ab Heft 54 wird INGWER BORG – bei ZUMA als wissenschaftlicher Leiter für den Bereich Wissensvermittlung und Beratung zuständig und zudem Professor für angewandte psychologische Methodik an der Universität Gießen – die Herausgeberschaft der ZUMA-Nachrichten für die Jahre 2004 und 2005 übernehmen. Redakteur bleibt weiterhin ROLF PORST.

Jan van Deth

Vorsitzender des ZUMA e.V.

IN MEMORIAM ERWIN K. SCHEUCH



In den ZUMA-Nachrichten 51 vom November 2002 hatte ich Dr. Peter Petersen gedacht, der, bei der Deutschen Forschungsgemeinschaft in vielfältiger Weise für ZUMA engagiert, am 11. Oktober 2002 gestorben war. Gerade ein Jahr später beklagt ZUMA einen weiteren großen Verlust, denn am 12. Oktober 2003 ist Professor Erwin K. Scheuch, emeritierter Ordinarius für Soziologie an der Universität zu Köln, nach schwerer Krankheit im Alter von 75 Jahren in Köln, seiner Geburts- und Heimatstadt, verstorben.

Erwin K. Scheuch war ein großer Sozialwissenschaftler, einer der wenigen deutschen Generalisten der Soziologie und besonders zu charak-

terisieren in seiner Forschung durch die systematische Verbindung von Theorie und Empirie in analytischer Absicht. Hinzu kam seine nie der deutschen Kirchturmperspektive verhaftete Orientierung an Theorie und Methodik international vergleichender Sozialforschung, die ihren Ausdruck nicht nur in zahlreichen hochbeachteten englischsprachigen Veröffentlichungen, sondern auch in einem langjährigen Engagement im Research Committee „Political Sociology“ der International Sociological Association (ISA) und der International Political Science Association (IPSA) gefunden hat.

Hinzu tritt Scheuch's engagiertes Eintreten als Wissenschaftler und Publizist für die deutsche Nachkriegsdemokratie, der Kampf gegen alle rechten und linken Ideologien und in gemeinsam mit seiner Frau veröffentlichten Büchern (Cliques, Klüngel und Karrieren, 1992; Die Spendenkrise. Parteien außer Kontrolle, 2000; Manager im Größenwahn, 2003) die Kritik am Verfall politischer Sitten in Zirkeln der deutschen Elite. Diese Arbeiten zeigen ihn im übrigen gelegentlich als einen großen Polemisiierer, der er auch sein konnte.

All dies – und mehr – war Erwin K. Scheuch. Die Erinnerung an ihn in dieser Veröffentlichung hat jedoch einen natürlichen, besonderen Schwerpunkt. Was war Erwin K. Scheuch für ZUMA? Zur Beantwortung dieser Frage muss ich ein wenig ausholen.

Indem Gerhard Baumert (DIVO), Scheuch und Rudolf Wildenmann an der Kölner Universität die Deutsche Wahlstudie 1961 konzipierten, entstand ein Projekt von beispielhafter Komplexität; darüber ist an anderer Stelle schon viel geschrieben worden. Nicht zuletzt die schwierigen Erfahrungen mit der Umsetzung dieses Vorhabens in empirische Forschung brachten Scheuch und Wildenmann zu der Überzeugung, dass die Empirische Sozialforschung in Deutschland eine Infrastruktur benötige, über die Forschungserfahrungen kumuliert, Datenbestände erhoben und gesichert sowie neue Forschungsmethoden entwickelt und allen Sozialforschern für ihre Arbeit zugänglich gemacht werden könnten.

Dabei konnte Scheuch in Köln zunächst auf die Vorarbeiten von Günther Schmölders als Gründer des Zentralarchivs für Empirische Sozialforschung an der Universität zu Köln (ZA) zurückgreifen. Nach seiner Rückkehr von einem USA-Aufenthalt übernahm er 1964 die Leitung des ZA und engagierte sich in dieser Funktion bis zum Jahr 1993; danach war er bis zu seinem Tode Vorsitzender dessen Trägervereins, der Kölner Gesellschaft für Sozialforschung e.V. Es entsprach seiner Weitsicht, die Gründung des Informationszentrums Sozialwissenschaften (IZ) in Bonn aktiv nicht zuletzt mit dem Ziel angesprochen zu haben, über jährlich durchzuführende Erhebungen einen Überblick über die in Deutschland, Österreich und der Schweiz laufenden Projekte empirischer Sozialforschung zu gewinnen und damit gleichzeitig dem ZA interessante Datenquellen zu erschließen.

Während hier Scheuch Hauptakteur war, setzte vor allem Wildenmann in langjährigen Bemühungen 1974 die Gründung von ZUMA durch, ein Vorhaben, an dessen erfolgreichen Abschluss Scheuch sich aber bereits seit den Anfängen intensiv beteiligte. So war er folgerichtig Gründungsmitglied des ersten Aufsichtsgremiums (Direktorium) von ZUMA, das von 1974 bis 1986 als Hilfseinrichtung der Forschung der DFG bestand, und nach der Zusammenführung 1986 von ZA, IZ und ZUMA in der GESIS als Institut der „Blauen Liste“ (heute „Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz“) bis zu seinem Tode Mitglied des ZUMA-Trägervereins.

ZUMA hat Scheuch viel zu verdanken. Nach der Konsolidierung des Instituts waren er, M. Rainer Lepsius und Rolf Ziegler die verantwortlichen Antragsteller bei der DFG für die Finanzierung der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS); so wurde die Finanzierung der Erhebungen 1980, 1982, 1984 und 1986 sichergestellt, bis danach der ALLBUS ein normaler Teil der GESIS-Haushaltsplanung wurde. Darüber hinaus hat er sich von Anfang an für das in seinem deutschen Erhebungsteil von ZUMA betreute International Social Survey Programme (ISSP) engagiert, ganz

im Sinne seiner Motivation, der international vergleichenden Umfrageforschung entsprechende Daten zu verschaffen. Und last but not least war es Scheuch, gemeinsam mit Wolfgang Zapf und dem Verfasser, der den Antrag verantwortete, der über viele Schritte und schwierige Verhandlungen 1986 zur Gründung der GESIS führte und damit der sozialwissenschaftlichen Infrastruktur in Deutschland eine sichere finanzielle Grundlage sicherte.

Wie die Teilnehmer an den Sitzungen des ZUMA-Trägervereins bezeugen können, war Scheuch für ZUMA ein kreativer Unruhestifter im besten Sinne, engagiert, präzise fragend, immer in Bewegung und nicht selten durchaus aggressiv im Vorbringen seiner Argumente – letzteres vor allem dann, wenn er Grenzüberschreitungen von ZUMA gegenüber dem von ihm geliebten Zentralarchiv witterte. Ohne ihn werden die Sitzungen der ZUMA-Mitgliederversammlung nicht mehr dieselben sein; wir betauern, mit seiner Familie, den großen Verlust, den die deutsche und internationale Sozialwissenschaft und nicht zuletzt ZUMA erlitten hat.

Bremen, im Oktober 2003

Max Kaase

ALLES ZUFALL ODER?

Ein Diskussionsbeitrag zur Qualität von face to face-Umfragen in Deutschland

PETER PH. MOHLER, ACHIM KOCH & SIEGFRIED GABLER

Der European Social Survey (ESS) setzt 70% Ausschöpfung als in Europa erreichbares Ziel an: "...the minimum target response rate – after discounting ineligibles ... – should be 70%. This figure is likely to be exceeded in certain countries and the ESS as a whole would be damaged if major national variations in response rates were to occur." (ESS 2001: 5, vgl. auch ESS 2003: 10). Dieses Ziel wird in Deutschland schon lange kaum noch erreicht. Wenn es aber anderswo in Europa als durchaus überraffbar gilt, ist es wohl an der Zeit, die öffentliche Diskussion über Qualität der Umfrageforschung in Deutschland zu verstärken. ZUMA beteiligt sich seit langem an dieser Diskussion, etwa durch das Symposium "Ausschöpfungen bei Umfragen" im Jahr 1997 (vgl. auch Porst 1996), Beteiligung an der DFG Publikation zur Qualität der Umfrageforschung (Kaase 1999) oder Beiträge in den ZUMA Nachrichten, wie zuletzt Koch (1998), Schräpler (2000), Diekmann/Jann (2001), Koch (2002) oder Schneekloth/Leven (2003) in diesem Heft. In diesem Zusammenhang sei auch auf die umfangreichen Publikationen von Schnell (insbes. Schnell 1997) verwiesen.

Dringlich wird die Diskussion deshalb, weil wir den Eindruck gewonnen haben, dass heutzutage in Deutschland auch qualitätsorientierte Umfragen faktisch kaum eine Ausschöpfungsquote von mehr als 50% erzielen – zumindest dann nicht, wenn das Stichprobenbrutto der Umfragen eindeutig definiert und die Umsetzung dieser Vorgaben im Feld penibel kontrolliert wird. Die oft genannten Gründe für das Absinken der Ausschöpfungsquote – wie schwierigere Erreichbarkeit von Zielpersonen (z.B. aufgrund steigender Frauenerwerbsquoten, der Zunahme von Einpersonenhaushalten u.ä.) oder ein verschlechtertes ‚Umfrageklima‘ (de Heer 1999) – können die schlechten Ausschöpfungsergebnisse nur bedingt erklären. Unsere Erfahrungen mit dem ALLBUS und anderen Studien deuten hier auf die *Form des Produktionsprozesses* als einem zentralen, bislang vernachlässigten Faktor hin. Welcher Produktionsprozess für eine Umfrage gewählt wird, hat offensichtlich nicht nur umfragetechnische, sondern auch betriebswirtschaftliche bzw.

organisatorisch/technische Gründe. Zufällig sind diese nicht, eher kann man sie als ‚strukturelle Bedingungen‘ ansehen. In unserer folgenden Skizze werden wir diesen Gedankengang näher ausführen.

Umfragen, genauer gesagt, repräsentative Bevölkerungsumfragen im ‚Angesicht der zu befragenden Person‘ (face-to-face) entstehen durch einen hochkomplexen Produktionsprozess (Mohler 2000). In diesem Produktionsprozess hat der Zufall nur eine Chance: bei der Auswahl von zu befragenden Personen oder Haushalten, zumindest, wenn die Stichprobe repräsentativ und/oder über Zeit und Raum vergleichbar sein soll (Gabler/Häder 2003, Lohr 1999). Aber auch im Prozess der Auswahl gibt es aufgrund fehlender Listen in der Regel keine uneingeschränkte Zufallsauswahl. Komplexe Auswahlverfahren (geschichtet, mehrstufig) wurden entwickelt, um den Auswahlprozess innerhalb eines gegebenen Auswahlrahmens zu optimieren. Alle anderen Prozesse, beginnend beim Design der Untersuchung, der Fragebogenentwicklung, der Interviewersteuerung, der Dokumentation, der Datenaufnahme bis zur Veröffentlichung der Ergebnisse sind deterministisch festgelegt. Sie folgen wissenschaftlich fundierten Regeln, damit aus der Umfrage ein ‚wissenschaftliches Instrument‘ werden kann, dessen Eigenschaften wohl bekannt sind. Jeder einzelne dieser Prozesse folgt speziellen methodischen Prinzipien, besser gesagt, er sollte solchen Prinzipien folgen. Um Prinzipien und Regeln folgen zu können, müssen diese nicht nur aufgeschrieben, sondern auch wissenschaftlich als tragfähig belegt sein. Dies gilt weitgehend bislang nur für die Teilprozesse Stichprobe (vgl. z.B. Gabler/Hoffmeyer-Zlotnik 1997) und Fragebogenentwicklung (vgl. z.B. Sudman et al. 1996). Für so wichtige Prozesse, wie das Design, die Feldsteuerung oder die Dokumentation, gibt es noch keine durch methodische Diskurse fundierte Regeln. Stattdessen bestimmen Tradition und Common Sense, was wie gemacht wird. Das Resultat sind Schwankungen in der Ausschöpfung von 30 Prozentpunkten und mehr in ein und derselben Gemeinde beim gleichen Institut in verschiedenen Jahren (wie beim ALLBUS z.T. zu beobachten), die sich nicht mehr auf statistischen Zufall zurückführen lassen, sondern auf Mängel im Produktionsablauf hindeuten. Dies führt aber auch zu Unterschieden zwischen Instituten in einem Land (die sattsam bekannte ‚Institutshandschrift‘), zwischen Regionen und Ländern, wie sich zum Beispiel aus den Dokumentationen des ISSP und ESS ablesen lässt (www.issp.org, www.europeansocialsurvey.org).

Nun könnte man fragen, hat die Empirische Sozialforschung damit nicht lange gut gelebt? Was, außer möglicherweise übertriebenem Perfektionismus, sollte einen bewegen, sich den zugegebenermaßen schwierigen Fragen einer fundierten Prozesssteuerung zuzuwenden?

Nun, wer über den eigenen Tellerrand hinausblickt, wird feststellen können, dass in anderen europäischen Ländern wie Dänemark oder Schweden auch streng kontrollierte Erhebungen Ausschöpfungsquoten jenseits der 60%-Marke erreichen. Dabei soll nicht verschwiegen werden, dass es in Europa auch Länder wie die Schweiz gibt, die nicht einmal 50% Ausschöpfung erreichen (http://ess.nsd.uib.no/2003_Fworksummary.jsp).

Wenn also weder die unwilligen Bürger (Survey Climate) noch die verbesserten Kontrollen (wie z.B. bei den im ALLBUS eingesetzten Registerstichproben) hinreichende Gründe für die schlechten Ausschöpfungsergebnisse in Deutschland sind, ist es dann womöglich der zu niedrige Preis, die fehlende Bereitschaft für gute Daten auch gutes Geld zu bezahlen? Diese Frage hat zwei Antworten:

- die eine ist, dass es in Deutschland derzeit weder für ‚Geld noch gute Worte‘ eine hohe, nachvollziehbare Ausschöpfung bei face-zu-face Umfragen gibt, weil die dafür notwendigen Produktionslinien nicht vorhanden sind (s.u.),
- die andere ist, dass es in Deutschland viel zu wenig Erhebungen gibt, die auf eine hohe Qualität oder besser Präzision ihres Erhebungsinstrumentes Wert legen.

Damit fehlt die Nachfrage, welche es den Anbietern erlauben würde, auch die dafür notwendigen Produktionsprozesse aufzulegen. Diesen Teufelskreis gilt es zu durchbrechen, wenn deutsche Umfragen internationalen wissenschaftlichen Standards weiterhin genügen wollen.

Belege für die Bedeutung des Produktionsprozesses zeigt eine noch nicht veröffentlichte Studie in den Niederlanden (vgl. Stoop 2003). Dort wurden Prozesse bewusst verändert mit dem Ziel einer Verbesserung der Ausschöpfung bzw. der Informationen über Nichtteilnehmer. Stoop berichtet für diese landesweite Umfrage eine kontrollierte Ausschöpfung von annähernd 70% im Hauptfeld und weiteren 70% bei einer kleinen Zufallsauswahl von Nichtteilnehmern in der Nachbearbeitung. Die von de Heer (1999) bis dahin berichteten Werte für die Niederlande lagen bei 55% Ausschöpfung. Dies und andere Indikatoren legen die Bedeutung des Erhebungsprozesses für die Ausschöpfungsquote nahe. Je nach Ausgestaltung dieses Prozesses können danach gute Ausschöpfungsquoten sogar unter widrigen umfrage-klimatischen Bedingungen erzielt werden.

Trivial? Die Antwort ist ja und nein zugleich. Einerseits ist es offensichtlich, dass die Qualität eines Produktes wesentlich vom Produktionsprozess abhängt. Weniger offensichtlich ist, dass auch Umfragen Ergebnis eines mit anderen Wirtschaftsgütern vergleichbaren Produktionsprozesses sind (Mohler 2000). Wie die Leadership Group on Quality in the European Statistical System formuliert: "The product is generated by an underlying process. It is unlikely that the product will have good quality if the underlying process is not up to par. Therefore the concept of process quality comes into play." (Lyberg et al.

2001). Deshalb gilt, dass mit einem auf Massenware ausgerichteten Fließband qualitativ hochwertige Waren nicht erzeugt werden können. Die Begrifflichkeit von „Massenware“ und „hochwertige Produkte“ ließe sich natürlich auch auf Umfragen übertragen. Jedoch möchten wir hier anstelle des abwertenden Begriffes „Massenware“ eher von „unterschiedlicher Präzision“ von Erhebungen sprechen. Diese Begrifflichkeit verweist auf andere technische Instrumente bzw. Messgeräte, wo die Präzision sich aus der angezielten Verwendung ergibt. Die dafür notwendigen Produktionsprozesse unterscheiden sich ganz erheblich.

In diesem Sinne wurden bis in die neunziger Jahre auch Umfragen, je nach der angestrebten Präzision getrennt, produziert. Es gab Interviewerstäbe für Marketingerhebungen, Sozialforschung, Gesundheitsforschung usw. Danach fand eine ‚technische Revolution‘, wahrscheinlich angeregt durch die hohen Investitionskosten für computerunterstützte face-to-face Erhebungen (CAPI) statt, indem an die Stelle von speziellen Interviewerstäben (und auch Feldsteuerungen) gemeinsame Stäbe und eine einheitliche Feldsteuerung traten. Von da an wurde ein und derselbe Interviewer für alle möglichen thematische Bereiche, vor allem aber für höchst unterschiedliche Präzision bei der Erfüllung des Stichprobenplanes eingesetzt (von ‚vorgegebener Anschrift einer Person‘ bis zu ‚random irgendwie walk‘). Krasse ‚random-irgendwie-walks‘ laufen darauf hinaus, dass es für die Interviewer einen mehr oder weniger zufallsgesteuerten Pfad durch ein Stadtviertel gibt, der so lange verfolgt wird, bis genügend Interviews erzielt werden. Da durch die Begehungsunterlagen und Begehungsregeln bestimmte Gebiete (z.B. Neubaugebiete, Adressen mit hohen Hausnummern in längeren Straßen, kleine Orte) praktisch keine Auswahlchance haben, verabschiedet man sich faktisch von einer Zufallsstichprobe mit berechenbaren Inklusionswahrscheinlichkeiten. Die Zahl der Nichtkontakte, Verweigerungen etc. bleibt bei diesem Vorgehen unerheblich. Das Verfahren legt den Interviewern nahe, möglichst schnell viele Interviews zu machen und dabei nicht an der Zahl kontaktierter Haushalte zu sparen. Geht man mit einer solchen Strategie an ein Design, in dem die Zielpersonen explizit mit Anschrift vorgegeben werden, wird die Stichprobe nicht nachdrücklich genug bearbeitet. Das Resultat ist eine hohe Zahl von ‚verbrannten‘ Adressen, d.h. eine suboptimale Ausschöpfung (von den weiteren Problemen bei der Durchführung der Interviews soll hier erst gar nicht die Rede sein). Dies bedeutet, dass die ‚technische Revolution‘ der Zusammenlegung von Interviewerstäben zwangsnotwendig zu einer Verringerung der maximal erreichbaren Präzision des wissenschaftlichen Instruments ‚Repräsentative Bevölkerungsumfrage‘ führen musste.

Dies ist aber nur einer von mehreren Strukturbrüchen, welche die Umfrageforschung in den letzten Jahren erlebt hat. Ein weiterer für den Produktionsprozess wichtiger Strukturbruch ist die Gesetzgebung zur Scheinselbständigkeit. Diese erzwingt, dass Interviewer

gleichzeitig für mehrere Erhebungsinstitute tätig sein müssen. Damit hat kein Feldchef mehr alleinigen Zugriff auf bestimmte Interviewer, er kennt deren persönliche Auslastung nicht mehr, die Bindung an das Institut ist, vorsichtig ausgedrückt, schwach. Hinzu kommt eine für hohe Präzisionsleistung eigentlich nicht angemessene Bezahlung. Das führt zu erheblichen Rekrutierungsproblemen bzw. in bestimmten Regionen zu einem Interviewermangel, was unter den gegebenen Bedingungen den Interviewern als selbständigen Unternehmern erlaubt, die weniger gewinnbringenden Studien links liegen zu lassen. Produktionsintensive, sehr präzise Erhebungen werden damit offensichtlich erschwert.

Um einen Weg aus dieser, wie wir meinen, strukturellen Krise der Umfrageforschung zu finden, wäre zuerst einmal eine fundierte öffentliche Diskussion über die für unterschiedliche Präzisionen notwendigen Produktionsprozesse erforderlich. Diese mag vielleicht ergeben, dass die von uns genannten Faktoren nicht die entscheidenden sein mögen. Dafür sind wir sicher offen. Weniger offen wären wir gegenüber entschuldigenden Argumenten, die sinkende Ausschöpfung ergäbe sich aus einer zufälligen Häufung unglücklicher Umstände. In diesem Sinne würden wir die Frage „alles Zufall oder?“ mit einem klaren „Nein“ beantworten.

Literatur

de Heer, W., 1999: International Response Trends: Results of an International Survey. *Journal of Official Statistics*, 15 (2): 129-142.

Diekmann, A./Jann, B., 2001: Anreizformen und Ausschöpfungsquoten bei postalischen Befragungen. Eine Prüfung der Reziprozitätshypothese. *ZUMA-Nachrichten* 48: 18-27.

ESS, 2001, 2003: European Social Survey – Specification for participating countries. http://naticent02.uuhost.uk.uu.net/proj_spec/index.htm

Gabler, S./Häder, S., 2003: Auswahl der Untersuchungseinheiten
http://www.politikon.org/preview.php?sid=a27b9aa5fcd1b95d17babe0abd507705&ilias_url=http%3A%2F%2Fwww.politikon.org/preview/course.php%3Fco_id%3D150%26co_inst%3D1

Gabler, S./Hoffmeyer-Zlotnik, J.H.P. (Hrsg.), 1997: *Stichproben in der Umfragepraxis*. Opladen: Westdeutscher Verlag.

Kaase, M. (Hrsg.), 1999: Qualitätskriterien der Umfrageforschung. Denkschrift der Deutschen Forschungsgemeinschaft. Berlin: Akademie Verlag.

Koch, A., 1998: Wenn "mehr" nicht gleichbedeutend mit "besser" ist: Ausschöpfungsquoten und Stichprobenverzerrungen in Allgemeinen Bevölkerungsumfragen. ZUMA-Nachrichten 42: 66-90.

Koch, A., 2002: 20 Jahre Feldarbeit im ALLBUS. Ein Blick in die Black-Box. ZUMA-Nachrichten 51: 9-38.

Lohr, S., 1999: Sampling: Design and Analysis. Pacific Grove: Duxbury Press.

Lyberg, L. et al., 2001: Summary Report from the Leadership Group (LEG) on Quality. <http://amrads.jrc.it/WPs%20pages/Quality/Documents/LEGsummary.pdf>

Mohler, P. Ph., 2000: Umfrageforschung. S. 310-322 in: B. Held/S. Russ-Mohl (Hrsg.): Qualität durch Kommunikation sichern - Vom Qualitätsmanagement zur Qualitätskultur. Erfahrungsberichte aus Industrie, Dienstleistung und Medienwirtschaft. Frankfurt/M: F.A.Z.-Institut für Management-, Markt- und Medieninformationen.

Porst, R., 1996: Ausschöpfungen bei sozialwissenschaftlichen Umfragen. Die Sicht der Institute. ZUMA-Arbeitsbericht 96/07.

Schneekloth, U./Leven, I., 2003: Woran bemisst sich eine "gute" allgemeine Bevölkerungsumfrage? Analysen zu Ausmaß, Bedeutung und zu den Hintergründen von Nonresponse in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS. ZUMA-Nachrichten 53: 16-57.

Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske + Budrich.

Schräpler, J.-P., 2000: Was kann man am Beispiel des SOEP bezüglich Nonresponse lernen? ZUMA-Nachrichten 46: 117-149.

Stoop, I., 2003: Surveying Nonrespondents. In: *Field Methods* (in press).

Sudman, S./Bradburn, N.M./Schwarz, N., 1996: Thinking about Answers: The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey-Bass.

WORAN BEMISST SICH EINE „GUTE“ ALLGEMEINE BEVÖLKERUNGSUMFRAGE?

Analysen zu Ausmaß, Bedeutung und zu den Hintergründen von Nonresponse in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS

ULRICH SCHNEEKLOTH & INGO LEVEN

Gemeinhin gilt die Ausschöpfung als eines der wichtigsten Merkmale zur Beurteilung der Qualität von repräsentativen allgemeinen Bevölkerungsumfragen. Anhand der Ergebnisse der „best practice“ ALLBUS-Erhebungen der Jahre 1994, 1996 und 2000 lässt sich zeigen, dass die üblicherweise geforderten hohen Ausschöpfungsraten bei harter Dokumentation der Bearbeitung faktisch nicht realisierbar sind. In der Arbeit wird der Frage nachgegangen, inwieweit eine gemeinhin als „schlecht“ angesehene Ausschöpfung auch tatsächlich zu einer mangelhaften Datenqualität führt. Dabei werden drei unterschiedliche Ansätze zur Beurteilung der Ergebnisqualität exemplarisch vorgeführt. In einem ersten Schritt wird ein Abgleich einiger soziodemografischer Randverteilungen der Stichprobe mit dem deutschen Mikrozensus als Referenzstatistik (Nettovalidierung) vorgenommen. In einem zweiten Schritt wird eine (multivariate) Nonresponse-Analyse durchgeführt. In einem dritten Schritt werden schließlich die Effekte der beim ALLBUS üblichen umfangreichen Nachbearbeitung von Adressen, für die bis dahin noch kein Interview realisiert werden konnte, dargestellt. Die Analysen zeigen, dass zufallsbasierte Stichproben auch bei einer Ausschöpfung von weniger als 50 Prozent keine nennenswerten systematischen Verzerrungen aufweisen müssen. Die isolierte Betrachtung der Ausschöpfung erweist sich von daher als ungeeigneter Prädiktor für Ergebnisqualität.

Sample response rates are usually considered to be one of the most important indicators of quality for representative general population surveys. The analysis of the "best practice" ALLBUS surveys of 1994, 1996 and 2000 shows that the high response rates typically called for in this kind of survey stand in conflict with strict documentation requirements for the fielding process. The article discusses whether so-called "poor" response

rates really impair data quality. Three different examples are presented for approaching the problem of judging the quality of random samples. First, we adjust some socio-demographic marginal totals of the random probability sample to those of the German micro-census as a reference statistic (Nettovalidation). Second, we carry out a (multivariate) non-response-analysis. Finally, we present the effects of the post-processing (reworking) of addresses usually employed in ALLBUS surveys for addresses for which interviews have not yet been realized. Our analysis shows that random samples are not necessarily systematically biased, even if the sample response rate is less than 50%. Focusing solely on the sample response rate thus proves to be an inappropriate indicator for data quality.

1. Problemstellung

Repräsentativ angelegte Erhebungen stellen überall dort, wo zum Zwecke der Analyse auf ansonsten nicht verfügbare Daten zurückgegriffen werden muss, den „Königsweg“ der Forschung dar.¹ Im Unterschied zu sonstigen Stichprobenverfahren, die entweder ungeregelt, nach dem Schneeballverfahren oder zum Beispiel auf Basis von vorab definierten Quotenvorgaben organisiert werden, basieren diese Verfahrensmodelle darauf, dass die Auswahl der Zielpersonen streng nach Zufallskriterien vorgenommen wird.² Nur dann, wenn die Auswahlchancen für alle Zielpersonen gleich oder zumindest mathematisch berechenbar sind, ist es zulässig, die Ergebnisse einer Stichprobe anhand von definierten statistischen Regeln als repräsentativ für die Grundgesamtheit zu betrachten.

Repräsentativerhebungen kommen heute in der ganzen Breite der Sozial-, Markt- und Meinungsforschung zum Einsatz. Neben den einschlägigen wissenschaftlichen Erhebungen mit Modellcharakter, wie zum Beispiel dem Sozio-ökonomischen Panel (SOEP) oder der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS), gilt dies auch für einen nicht unerheblichen Teil von Datenbeständen, die offiziell unter dem Titel „Amtliche Statistik“ geführt werden. Verwiesen werden kann in diesem Zusammenhang auf die im jährlichen Turnus erhobene Mikrozensus-Stichprobe des Statistischen Bundes-

1 Nicht umsonst stellt auch die 1999 von Max Kaase herausgegebenen Denkschrift der Deutschen Forschungsgemeinschaft „Qualitätskriterien der Umfrageforschung“ fest dass „Repräsentative Umfragen (...) sich seit dem Zweiten Weltkrieg zum unentbehrlichen Mittel der Datenerhebung über größere Bevölkerungen entwickelt“ haben. (DFG 1999: 12).

2 Ein Überblick zu den gängigen Auswahlverfahren findet sich zum Beispiel bei Schnell/Hill/ Esser (1999: 247ff.)

amtes oder auch seit neuestem auf die von Infratest Sozialforschung für das Statistische Bundesamt durchgeführte Erhebung „Der Arbeitsmarkt in Deutschland“.³

Umso bemerkenswerter ist es, dass Fragestellungen, die sich auf die Qualität von Umfragen beziehen, im wissenschaftlichen Diskurs nach wie vor eher „stiefmütterlich“ behandelt werden und etwas überspritzt formuliert als „Steckenpferd“ eines eher kleineren Zirkels von Methodenforscherinnen und -forschern erscheinen. Aus unserer Sicht gibt es in Deutschland heutzutage nach wie vor keinen wirklich praxistauglichen Konsens darüber, welche Anforderungen an eine definierte Qualität von Umfragen zu stellen sind (Kriterien) bzw. mit welchen Messkonzepten Qualität möglichst kontinuierlich auch empirisch überprüft werden kann (Monitoring). Unser Eindruck ist der, dass nach wie vor zwei Extremstrategien im Rahmen einer „Qualitätssicherung“ in der empirischen Sozialforschung vorherrschen. Entweder man lässt die Dinge laufen und verlässt sich darauf, dass die „durchführenden Instanzen“ (Umfrageinstitute, sonstige private Anbieter, Lehrstühle, Institute etc.) die Dinge schon richtig machen werden, oder aber man formuliert zum Beispiel im Rahmen der Ausschreibung – unabhängig vom Studientyp oder dem inhaltlichen Erkenntnisinteresse – in der Regel „Lehrbuch gestützt“ alle möglichen Maximalanforderungen, um sich dann hinterher doch für das Angebot zu entscheiden, das in einem als finanziell noch verkraftbar betrachteten Rahmen kalkuliert ist. Beide Strategien sind aus unserer Sicht nicht per se zielführend oder qualitätsfördernd. Das eigentliche Dilemma besteht darin, dass es jenseits möglicher Ankündigungen, welche methodischen Standards im Einzelnen sinnhaft sein könnten bzw. welche auch tatsächlich praxistauglich und/oder finanzierbar sind, für den Endnutzer nach wie vor viel zu wenig Kriterien für Ergebnisqualität gibt, mit denen sich die Belastbarkeit der erhobenen Daten einigermaßen verlässlich abschätzen lässt.

1.1 Zur Qualität von Umfragen

Die Denkschrift der Deutschen Forschungsgemeinschaft zu den „Qualitätskriterien der Umfrageforschung“ führt hierzu aus, dass es jenseits der theoretischen Kriterien einer verzerrungsfreien Abbildung der Grundgesamtheit durch die Stichprobe sowie der zuverlässigen Messung der Sachverhalte anhand valider Erhebungsinstrumente praktisch zwei Arten von Indizien gibt, um die Qualität von Umfragen zu bewerten:

- Richtige Ergebnisse sowie
- korrekte Verfahren (DFG 1999: 94).

3 Vergleiche dazu die entsprechenden Darstellungen auf der Web-Site des Statistischen Bundesamtes unter http://www.destatis.de/dt_erheb/arbeitsmarkt/am_einf.htm bzw. zum Mikrozensus unter http://www.destatis.de/micro/d/micro_c1.htm

Die Kriterien mögen auf den ersten Blick trivial erscheinen. Tatsächlich markieren sie aber die beiden zentralen Eckpunkte, an denen sich Umfragen bemessen lassen müssen.

Richtige Ergebnisse sind selbstredend das zentrale Ziel von Forschung. Überall dort, wo keine einfachen Tatbestände, die in der Wirklichkeit mehr oder weniger leicht quantifizierbar sind, sondern komplexere Strukturen gemessen werden, kann man dem erhobenen Ergebnis allerdings nicht immer unmittelbar „ansehen, ob und bis zu welchem Grade es denn „richtig“ oder „falsch“ ist. Häufig hilft hier nur das Vertrauen in die „Regeln der Kunst“, wie es weiter in der zitierten Denkschrift heißt, sprich in die korrekte Anlage und Durchführung der Erhebung. Wir stimmen dieser Einschätzung zu und möchten hierzu ergänzend anmerken, dass es zusätzlich darauf ankommt, für beide vorhin benannten Kriterien zu praktikablen Operationalisierungen zu kommen, die helfen, sowohl ex ante als auch ex post die Qualität einer Umfrage zu beurteilen.

Folgen möchten wir auch der weiteren Argumentation der Denkschrift:

„Diese Vertrauensbildung (in die Belastbarkeit der Ergebnisse – US/IL) über Verfahren ist wichtig und unverzichtbar, aber sie hat ihre Risiken. Bedenklich ist etwa, wenn einzelne Kriterien als Qualitätsmaßstab verabsolutiert werden (beispielsweise eine Stichprobenausschöpfung in einer bestimmten Höhe. (...) Problematisch wird es auch dann, wenn die Orientierung an etablierten Verfahren zu einem methodischen Konservativismus führt.“ (DFG 1999: 94f)

In der Praxis wird die Güte einer Erhebung häufig isoliert anhand der realisierten **Ausschöpfung** beurteilt. Unter Ausschöpfung wird dabei in der Regel das Verhältnis der realisierten Netto-Interviews zu den ursprünglich im Rahmen der Stichprobe vorgegebenen gültigen Bruttofällen verstanden. Der daraus resultierende Quotient hat ohne Zweifel seinen Charme. Drückt man das Ergebnis in Prozent, also im Verhältnis zum Idealwert 100 aus, so ergibt sich eine lineare Größe, der man scheinbar auf den ersten Blick ansieht, ob sie hoch oder niedrig, akzeptabel oder schlecht ist. Was liegt näher, als die Qualität einer Umfrage anhand dieses linearen Kriteriums im Sinne eines Benchmarkings zu messen.

Tatsächlich misst die Ausschöpfung jedoch nur – worauf ebenfalls völlig zu Recht in der Denkschrift der DFG hingewiesen wird – *„wie groß der Spielraum für Selektivität durch Nonresponse ist. Sie besagt nichts über die tatsächliche Selektivität.“ (DFG 1999: 104)* Nonresponse führt nur dann zu einer Verschlechterung der Erwartungstreue, wenn damit systematisch eine Unterausschöpfung von bestimmten Bevölkerungsgruppen verbunden ist. Ein Nonresponse, der mehr oder weniger zufällig verteilt ist, ist hingegen unter Qualitäts Gesichtspunkten betrachtet unerheblich. Die Effekte von Nonresponse müssen dem-

nach erst im Einzelnen überprüft werden, bevor man eine mindere Qualität postuliert. Unter anderem aus diesem Grunde gibt es in der Literatur bisher auch kein allgemeingültiges Kriterium, wie hoch eine Ausschöpfung ausfallen muss, um eine zufallsbasierte Stichprobe als adäquate Abbildung der Grundgesamtheit zu charakterisieren. Eine formale Regel, die besagt, dass eine Stichprobe zum Beispiel zu mindestens 70 Prozent ausgeschöpft sein soll, um Anerkennung zu finden, lässt sich letztendlich weder theoretisch noch empirisch herleiten.⁴

1.2 Empirischer Zugang

Wir möchten mit der vorliegenden Arbeit einen Beitrag dazu leisten, die tatsächlich erreichbaren Anforderungen an die Qualität von Umfragen empirisch zu beleuchten. Einen wichtigen Stellenwert nimmt dabei die Frage ein, inwieweit die Höhe der Ausschöpfung für sich genommen bereits einen hinreichend geeigneten Indikator zur Beurteilung der Qualität einer repräsentativ angelegten Erhebung darstellt. Oder anders formuliert: Kann es gute Qualität auch bei „schlecht“ ausgeschöpften Stichproben geben?

Als empirische Basis haben wir hierfür die von Infratest Sozialforschung zuletzt im Jahr 2000 im Auftrag von ZUMA, Mannheim durchgeführten ALLBUS-Erhebungen herangezogen.⁵ Der ALLBUS eignet sich aus den folgenden Gründen gut für eine methodische Reanalyse.

Zum Ersten handelt es sich beim ALLBUS dem eigenen Anspruch nach um ein „best practice Verfahrensmodell“, bei dem im zweijährigen Turnus persönlich-mündliche Befragungen auf der Basis von Zufallsstichproben durchgeführt werden. Die Stichprobenziehung erfolgte in den Jahren 1994, 1996 und 2000 anhand einer Einwohnermeldeamtstichprobe (Registerstichprobe). Im Unterschied zum weit verbreiteten und alternativen Random-Route Modell, bei dem der Interviewer ausgehend von einer vorgegebenen und

4 Bei den im Rahmen der scientific community geforderten Schwellenwerten handelt es sich aus unserer Sicht um Faustformeln, die aus den vorhandenen Informationen über früher üblicherweise dokumentierte Ausschöpfungen abgeleitet werden. Schnell, Hill, Esser weisen zum Beispiel aus, dass die „typischen“ Ausschöpfungen in den 70iger Jahren bei 80 Prozent lagen, während seit Beginn der 80iger Jahre eher Werte um die 70 Prozent üblich seien (Schnell/Hill/Esser 1999: 288).

5 Die ALLBUS-Stichproben der Jahre 1994, 1996 und 2000 wurden von Infratest Sozialforschung als Registerstichprobe nach dem jeweils gleichen Verfahrensmodell durchgeführt. Die Daten liegen vor und können über das Zentralarchiv für empirische Sozialforschung (ZA) zum Zwecke wissenschaftlicher Analysen angefordert werden. Die Erhebung von 1998 unterscheidet sich methodisch und wurde von GFM-Getas als Random-Route Befragung durchgeführt. Die Ergebnisse der neuesten Erhebung von 2002, die von Infas ebenfalls auf Basis einer Einwohnermeldeamtstichprobe erstellt wurde, lagen uns noch nicht vor.

zufällig gezogenen Startadresse anhand einer festen Begehungsregel zum Beispiel jeden dritten Haushalt kontaktieren soll (Random-Walk), werden die Adressen in diesem Fall anhand einer ebenfalls zufällig gezogenen Auswahl von Einwohnermeldeämtern bestimmt und danach dem Interviewer vorab fest vorgegeben.⁶ Stichprobenziehung und Ergebnisse der Feldbearbeitung sind in diesem Fall eindeutig getrennt und als mögliche Nettoeffekte im Vergleich zur Bruttostichprobe klar abgrenzbar. Die Feldarbeit erfolgt mit großer Intensität und ist in ihren zentralen Bestandteilen gegenüber dem Auftraggeber transparent und nachvollziehbar angelegt. Die „Regelgebundenheit“ in der Bearbeitung ist damit genuiner Bestandteil des Verfahrensmodells.⁷

Anlage und Prozess der Feldarbeit sind **zum Zweiten** bei den ALLBUS-Untersuchungen umfangreich dokumentiert.⁸ Ein methodisches Kernstück des Verfahrensmodells der Jahre 1994, 1996 und 2000 (Adressen-Random bzw. AR-Modell) bildet die Bereitstellung eines sog. Bruttobandes, in dem alle im Rahmen der Erhebung brutto bei den Einwohnermeldeämtern gezogenen Adresseneinheiten (anonymisiert) aufgeführt und hinsichtlich ihres jeweiligen Bearbeitungsergebnisses nachgewiesen sind. Die vorliegende Messreihe der Jahre 1994, 1996 und 2000 erlaubt von daher einen tatsächlichen Vergleich der Entwicklung der Ausschöpfung, da in diesem Fall die Berechnung der brutto eingesetzten und netto auch tatsächlich realisierten Adressen nach dem gleichen Prinzip erfolgt.

Gerade dieser letzte Aspekt ist unter methodischen Gesichtspunkten zentral. In der Praxis können Ausschöpfungen bei unterschiedlichen Verfahrensmodellen nicht völlig vergleichbar berechnet werden. Tatsächlich macht es einen bedeutenden Unterschied aus, ob die Bruttostichprobe vorab definiert und vorgegeben ist, oder ob der Interviewer zum Beispiel im Rahmen eines Random-Walks im gleichen Schritt aufschreibt, wie viele Haushalte er brutto kontaktiert und wie viele er davon tatsächlich netto befragen konnte (Random-Route Verfahren). Abweichungen bei der Dokumentation von Fehlkontakten können zu deutlichen Veränderungen bezüglich der errechneten Ausschöpfung führen.⁹

6 Hierzu wird im ersten Schritt eine nach regionalen und siedlungsstrukturellen Kriterien geschichtete Zufallsstichprobe von Gemeinden erstellt. Im Anschluss daran werden die so genannten Sample-Points – eine großproportional zur ausgewählten Gemeinde vorgegebene Menge von Zielpersonenadressen – anhand einer Allokationsrechnung ermittelt, die dann in einem zweiten Schritt von den jeweiligen Einwohnermeldeämtern der Gemeinden ebenfalls zufällig aus dem Melderegister gezogen werden. Vergleiche dazu zum Beispiel Koch/Wasmer/Harkness/Scholz (2001).

7 Vergleiche dazu im Einzelnen die Kriterien für Stichprobenqualität in DFG (1999: 100f).

8 Die von ZUMA erstellten Methodenberichte zu den ALLBUS-Untersuchungen stehen im Internet unter <http://www.gesis.org/Dauerbeobachtung/ALLBUS/Bestellen/mberichte.htm> zum Download bereit.

9 Hierbei muss es sich übrigens nicht immer um bewusste Fälschungen handeln. Gerade beim Random-Walk ist es für den Interviewer häufig schwierig zu entscheiden, ob sich hinter einer

Verweisen möchten wir an dieser Stelle auf eine Arbeit von Koch aus dem Jahre 1998, der anhand eines Vergleiches verschiedener allgemeiner Bevölkerungsstichproben zu dem Ergebnis kommt, das ausgewiesenes „mehr“ nicht gleichbedeutend mit „besser“ sein muss (Koch 1998). Hierbei zeigte sich, dass für „hart“ dokumentierte Registerstichproben, wie etwa dem ALLBUS 1994, deutlich niedrigere Ergebnisse im Bereich um die 55 Prozent als für vergleichbare Random-Route Erhebungen mit Ausschöpfungen um die 70 Prozent (Sozialwissenschaften-Bus) bzw. sogar bis zu 80 Prozent (Media-Analyse) angegeben wurden. Ein Vergleich der Randverteilungen der Stichproben mit den Daten des Mikrozensus ergab jedoch, dass die scheinbar „schlechtere“ Registerstichprobe alles in allem betrachtet die beste Übereinstimmung mit der externen Referenzstatistik aufwies.

Koch merkt in diesem Zusammenhang an, dass

„zu fragen ist ..., inwieweit die berichteten Ausschöpfungsergebnisse – insbesondere die hohen Werte von 70 Prozent und mehr – die Realität korrekt wiedergeben. (...) Bei den analysierten Umfragen kann lediglich im ALLBUS 1994 die Ausschöpfung (weitgehend) zuverlässig berechnet werden, weil aufgrund der Ziehung einer Stichprobe von Personenadressen aus Einwohnermelderegistern vor Feldbeginn ein genau definiertes Stichprobenbrutto vorlag (...). Sämtliche andere Umfragen dagegen wurden nach dem ADM-Design durchgeführt. Die faktische Bruttostichprobe wird hier durch die Interviewer in der zweiten und dritten Auswahlstufe gebildet. Die Zuverlässigkeit der Ausschöpfungsbeurteilung hängt davon ab, wie exakt die Interviewer die Vorgaben betrachten.“ (Koch 1998: 81).

Diese Hypothese wird auch durch neuere Befunde zur tatsächlich erreichbaren Ausschöpfung bei allgemeinen Bevölkerungsstichproben erhärtet. So wurden etwa im Rahmen des DEFECT-Projekts unter methodisch kontrollierten Bedingungen bei face-to-face Befragungen mit fest vorgegebenen Haushaltsadressen Ausschöpfungen von 39,3 bzw. 41,5 Prozent berichtet. Hierbei wurden die Adressen von Projektmitarbeiterinnen und -mitarbeitern auf Basis eines Random-Walks aufgeschrieben und den einbezogenen kommerziellen Umfrageinstituten extern vorgegeben. Für eine vergleichbar angelegte Telefon-

Adresse ein Privathaushalt oder zum Beispiel eine Geschäftsadresse verbirgt, ob der Haushalt, bei dem auch bei mehrmaligem Anlauf kein Kontakt herzustellen ist, grundsätzlich unbewohnt ist oder ob dort eine zur Grundgesamtheit gehörige Person lebt, die schwer zu erreichen ist. Dokumentiert der Interviewer die Adresse als nicht zur Grundgesamtheit zugehörig, so gilt der Ausfall als qualitätsneutral und wird bei der Berechnung der Ausschöpfung nicht mit berücksichtigt. Dokumentiert er hingegen einen Ausfallcode wie zum Beispiel „im Haushalt niemanden angetroffen“, so sinkt entsprechend die Ausschöpfung.

erhebung wurde eine Ausschöpfung von 29,4 Prozent und für eine schriftliche Befragung eine Rücklaufquote von 48,7 Prozent ermittelt (Schnell/Kreuter 2000: 96).

Auch die Ausschöpfungen der ALLBUS-Registerstichproben der Jahre 1994 und 1996 haben im Bereich zwischen 53 Prozent und 55 Prozent gelegen – Tendenz fallend. Beim ALLBUS 2000 wurde sogar zum ersten Mal die „50 Prozent-Marke“ unterschritten (vergleiche dazu Übersicht 1), was offenbar von der einschlägigen scientific community als „Alarmsignal“ interpretiert wurde.

Übersicht 1: Anlage und Ergebnisse der register-basierten ALLBUS-Untersuchungen

	ALLBUS 2000		ALLBUS 1996		ALLBUS 1994	
	West	Ost	West	Ost	West	Ost
Grundgesamtheit	Bevölkerung ab 18 Jahren in Privathaushalten in Deutschland		Bevölkerung ab 18 Jahren in Privathaushalten in Deutschland		Bevölkerung ab 18 Jahren in Privathaushalten in Deutschland	
Stichprobe	Einwohnermeldeamtsstichprobe		Einwohnermeldeamtsstichprobe		Einwohnermeldeamtsstichprobe	
Erhebungsmethode	persönlich/mündlich		persönlich/mündlich		persönlich/mündlich	
Geplante durchschnittliche Interviewdauer	50 Minuten		50 Minuten		50 Minuten	
Netto-Fallzahl	2.036	1.102	2.402	1.116	2.342	1.108
Ausschöpfung	46,9	53,7	54,2	54,2	53,2	55,2

Infratest Sozialforschung 2003: eigene Darstellung

Die für den ALLBUS 2000 aufgeführten Gründe für Nonresponse (Übersicht 2) zeigen, wie bei hart bearbeiteten und dokumentierten Verfahrensmodellen üblich, dass rein quantitativ betrachtet weniger die prinzipielle Nichterreichbarkeit von Zielpersonen, sondern vorrangig eine mangelnde Kooperationsbereitschaft („Verweigerung“) zum Nonresponse führt. Es finden sich keinerlei Hinweise dafür, dass die Stichprobenausfälle durch mangelhaften Aufwand in der Feldbearbeitung zustande gekommen sind oder einem mangelndem Engagement der Interviewerinnen und Interviewer geschuldet wären.¹⁰ Wie bereits

¹⁰ Diese Aussage wird übrigens nicht dadurch relativiert, dass auch bei einer Einwohnermeldeamtsstichprobe ein nicht unerheblicher Teil der Adressen als qualitätsneutrale Ausfälle deklariert

analog bezüglich der generellen Höhe der Ausschöpfungsquote erläutert, lässt auch diese Übersicht noch keine gesicherten Rückschlüsse auf mögliche systematische Effekte zu. Entscheidend ist nicht, welcher der von den Interviewern dokumentierten finalen Ausfallgründe am häufigsten auftritt, sondern inwieweit damit Effekte einhergehen, die zu einer Verzerrung der Stichprobe in einem relevanten Ausmaß führen.

Übersicht 2: Gründe für Nonresponse im ALLBUS 2000

	West	Ost
Bruttostichprobe	100	100
./. stichprobenneutrale Ausfälle		
– Adresse falsch	3,5	2,9
– Zielperson verstorben	0,7	0,9
– Zielperson verzogen	8,0	7,6
– Zielperson lebt nicht mehr in Privathaushalt	1,0	0,9
– Adresse nicht abschließend bearbeitet	1,6	0,2
Bereinigte Bruttostichprobe	100	100
./. Systematische Ausfälle		
– Im Haushalt niemanden angetroffen	5,3	1,4
– Zielperson nicht angetroffen	2,1	1,0
– Zielperson nicht befragungsfähig	1,5	1,8
– Zielperson aus Zeitgründen nicht bereit	3,8	1,6
– Zielperson generell nicht bereit	36,8	37,8
– Zielperson spricht nicht hinreichend gut deutsch	1,5	0,1
– Interview nicht korrekt durchgeführt	2,2	2,6
Auswertbare Interviews	46,9	53,7

Koch/Wasmer/Harkness/Scholz (2001: 58)

Wie ist dieser Gesamtbefund nun zu beurteilen? Führen auch schlechter ausgeschöpfte Zufallsstichproben zu akzeptablen Ergebnissen, oder manövrieren wir uns mit dem „best practice“ Modell für persönlich/mündliche Repräsentativerhebungen immer mehr ins Abseits?

werden. Die vorliegenden Erfahrungen zeigen, dass um die 15 Prozent der Zielpersonen, die anhand der Einwohnermelderegister gezogen wurden, nicht (mehr) unter der angegebenen Adresse wohnen bzw. dort nicht mehr erreichbar sind. Beim ALLBUS-Verfahrensmodell werden diese Adressen, um systematische Effekte zu minimieren, durch Reserveadressen, die bei den gleichen Einwohnermeldeämtern vorab mitgezogen wurden, kontrolliert ersetzt.

Im Folgenden werden wir drei verschiedene Wege zur Beurteilung der Ergebnisqualität von Erhebungen exemplarisch ausführen. In einem ersten Zugang werden im Sinne einer Nettovalidierung einige Randverteilungen der ALLBUS-Erhebungen der Jahre 1994, 1996 und 2000 mit den Ergebnissen des Mikrozensus als externer Referenzstatistik abgeglichen. In einem zweiten Zugang wird danach für den ALLBUS 2000 eine Nonresponse-Analyse durchgeführt. Hierbei wird, neben den verfügbaren Merkmalen, die für die Bruttostichprobe von den Einwohnermeldeämtern geliefert wurden, zusätzlich auf externe (kommerziell vertriebene) mikrogeografische Datensätze, mit denen einige prägende Charakteristika der Straßenzüge, aus denen die Bruttoadressen stammen, beschrieben werden können, sowie auf ausgewählte Merkmale einer Befragung der beim ALLBUS 2000 eingesetzten Interviewerinnen und Interviewer zurück gegriffen. Als drittem Zugang werden abschließend die Effekte der für die ALLBUS-Erhebungen typischen intensiven Nachbearbeitungswellen im Vergleich zu den Ergebnissen der ALLBUS-Hauptbearbeitung analysiert.

Wir knüpfen hierbei auch an die zuletzt ebenfalls von Koch im Jahre 2002 vorgelegte Arbeit „20 Jahre Feldarbeit im ALLBUS: ein Blick in die Blackbox“ an, die zusammen mit den bereits erwähnten Methodenberichten einen umfassenden Einblick in die Realisierungsbedingungen der ALLBUS-Untersuchungen gibt (Koch 2002). Im Unterschied zum letztgenannten Artikel, der die Regelgebundenheit des Verfahrensmodells in den Vordergrund stellt, möchten wir uns vorrangig mit dem Aspekt der Ergebnisqualität befassen.

2. Vergleich der Randverteilungen der ALLBUS-Erhebungen 1994, 1996 und 2000 mit dem Mikrozensus

Im folgenden Abschnitt möchten wir die Ergebnisse der ALLBUS-Erhebungen 1994, 1996 und 2000 im Vergleich zu den Ergebnissen des Mikrozensus des Statistischen Bundesamtes betrachten und bewerten. Hierbei gilt es zu überprüfen, ob sich Belege dafür finden lassen, dass schlechter ausgeschöpfte Erhebungen im Vergleich zum Mikrozensus größere Abweichungen aufweisen. Dieses mit dem Begriff der Nettovalidierung zur charakterisierende Verfahren entspricht in etwa dem Vorgehen in der bereits erwähnten Arbeit von Koch (1998), der auf diese Weise verschiedene Zufallsstichproben, für die unterschiedliche Ausschöpfungsquoten berichtet wurden, untersucht hat.¹¹

11 Vergleiche zum Verfahren und zu den Ergebnissen Koch (1998) sowie vorher Hartmann (1990) oder Hartmann/Schimpl-Neimanns (1992).

Wie bereits ausgeführt, ist die Ausschöpfung bei den letzten drei ALLBUS-Erhebungen, die anhand von Einwohnermeldeamtsstichproben gebildet wurden, weiter gesunken. Vor allem in den alten Bundesländern hat sich seit 1994 eine Verringerung der Ausschöpfung um mehr als 6 Prozentpunkte ergeben. Mit 1,5 Prozentpunkten fällt der Rückgang in den neuen Bundesländern weniger dramatisch aus. Um die möglichen Auswirkungen besser unterscheiden zu können, haben wir den Abgleich mit der amtlichen Statistik immer getrennt nach Ost und West vorgenommen. Die untersuchten Merkmale umfassen Alter, Geschlecht, Familienstand, Erwerbsstatus, Schulabschluss und Haushaltsgröße. Weitere Validierungen sind nicht möglich, da der Mikrozensus ansonsten über keine mit dem ALLBUS direkt vergleichbaren Merkmale verfügt. Die Ergebnisse der Validierung haben wir in der folgenden Übersicht 3 zusammengefasst. Zur Veranschaulichung haben wir dabei für jede der abzugleichenden Merkmalsausprägungen alle positiven Abweichungen in den ALLBUS-Untersuchungen, also alle Ausprägungen, die größer als die Referenzergebnisse im Mikrozensus ausfallen, aufaddiert und in einem dadurch komprimierten Gesamtüberblick dargestellt. Die tabellarischen Nachweise mit den Einzelergebnissen sind im Anhang beigefügt.

Insgesamt weisen die ALLBUS-Stichproben eine recht gute Anpassung an die Verteilungen der amtlichen Statistik auf. In den demografischen Grundverteilungen (Alter, Familienstand, Geschlecht sowie Ein- bzw. Mehrpersonenhaushalte) schwanken die Abweichungen in der Regel mit nicht mehr als 2 bis 3 Prozentpunkten um das Referenzergebnis.

Beim höchsten Schulabschluss und beim Erwerbsstatus ergeben sich zum Teil etwas größere Abweichungen, was jedoch auch methodisch verursacht und mit abfragebedingten Artefakten zusammenhängen könnte. Der Mikrozensus weist beim Schulabschluss, insbesondere bei älteren Personen, zu einem nicht unerheblichen Anteil fehlende Angaben auf, die die Gesamtverteilung verzerren. Beim Erwerbsstatus ist davon auszugehen, dass das Berichtswochenkonzept des Mikrozensus, bei der eine konkrete Erwerbstätigkeit in einer definierten Woche abgefragt wird, im Vergleich zum ALLBUS zu systematisch höheren und damit genaueren Angaben führt, da hierbei Teilzeit-Beschäftigung sowie insbesondere geringfügige Beschäftigungsverhältnisse präziser abgebildet werden. Entsprechende methodische Effekte können sich leicht zu Abweichungen in Höhe von mehreren Prozentpunkten addieren.¹²

12 Vergleiche dazu auch die methodischen Anmerkungen von Koch (1998) in den Fußnoten 18 bis 20.

Übersicht 3: Vergleich der Summe der Abweichungen aus ALLBUS und Mikrozensus (Addition der positiven Abweichungen in Prozentpunkten)

	ALLBUS 2000	ALLBUS 96	ALLBUS 94
West			
Ausschöpfungsquote	46,9	54,2	53,2
Geschlecht	0,3	2,2	3,8
Alter	2,6	3,3	2,9
Familienstand	0,6	2,8	1,6
Erwerbsstatus	2,5	1,7	4,4 ¹⁾
Schulabschluss	4,1	7,7	5,7 ¹⁾
Haushaltsgröße	3,0	3,9	1,5
Ost			
Ausschöpfungsquote	53,7	54,2	55,2
Geschlecht	0,1	0,2	0,6
Alter	2,5	3,5	4,5
Familienstand	0,9	5,4	3,7
Erwerbsstatus	1,0	0,7	0,2 ¹⁾
Schulabschluss	3,8	2,4	4,3 ¹⁾
Haushaltsgröße	0,1	3,0	2,9

1) Referenz: Mikrozensus 93

Infratest Sozialforschung, eigene Berechnungen

Insgesamt lässt sich feststellen, dass die sinkende Ausschöpfung bei den letzten ALLBUS-Erhebungen nicht mit einer Zunahme der Abweichungen bei den soziodemographischen Randverteilungen einhergeht. Im Gegenteil! Der ALLBUS 2000 weist sowohl in den alten als auch in den neuen Bundesländern bei den soziodemografischen Grundmerkmalen insgesamt die geringsten Abweichungen auf.

Eine geringfügig andere Tendenz findet sich beim Erwerbsstatus. Hier weicht der ALLBUS 2000 etwas weiter vom Mikrozensus ab, als dies im Westen 1996 bzw. im Osten bei beiden vorherigen Erhebungen der Fall gewesen ist. Im Einzelnen wurden im Westen im Jahr 2000 2,5 Prozentpunkte mehr sog. Nichterwerbspersonen und im Osten 1 Prozentpunkt mehr Arbeitslose als im Mikrozensus ausgewiesen (vergleiche dazu die tabellarische Übersicht im Anhang). Die Abweichungen sind für sich genommen allerdings nicht besonders hoch ausgeprägt.

Weitere kleinere Abweichungen beziehen sich auf das Merkmal Haushaltsgröße, bei dem im ALLBUS 2000 in den alten Bundesländern die Abweichung gegenüber dem ALLBUS 1996 geringer, gegenüber dem ALLBUS 1994 jedoch höher ausfallen. Dies lässt sich auch

beim Merkmal Schulabschluss in den neuen Bundesländern feststellen, wo die Abweichung vom ALLBUS 2000 gegenüber dem ALLBUS 1994 geringer, gegenüber dem ALLBUS 1996 jedoch höher ausfällt. Dahingegen gilt, dass in den alten Bundesländern bei den Merkmalen Alter, Geschlecht, Familienstand und Schulabschluss die Abweichungen beim ALLBUS 2000 geringer als in den vorherigen Studien waren. In den neuen Bundesländern lässt sich dies beim ALLBUS 2000 für die Merkmale Alter, Geschlecht, Familienstand und Haushaltsgröße feststellen.

Anhand des Vergleichs der ALLBUS-Erhebungen 1994, 1996 und 2000 mit der amtlichen Statistik lassen sich demnach keine hinreichend eindeutigen empirischen Belege für die Hypothese finden, dass schlechter ausgeschöpfte Stichproben generell größere Abweichungen nach sich ziehen. Vielmehr zeigt sich, dass der ALLBUS 2000 trotz der relativ geringen Ausschöpfung in sehr weiten Teilen weniger Abweichung zur amtlichen Statistik aufweist als die vorherigen ALLBUS-Untersuchungen der Jahre 1996 und 1994. Auch die Unterscheidung zwischen den alten und neuen Bundesländern hat nicht bestätigen können, dass mit einer besseren Ausschöpfung in den neuen Bundesländern eine bessere Annäherung an die amtliche Statistik verbunden wäre.

Die nach einem hart dokumentierten und registerbasierten Verfahrensmodell durchgeführten ALLBUS-Untersuchungen der Jahre 1996 und 2000 haben demnach den mit dem ALLBUS 1994 im Vergleich zu anderen, besser ausgeschöpften allgemeinen Bevölkerungsumfragen gesetzten Standard bezüglich der Übereinstimmung der soziodemografischen Grundverteilungen mit der amtlichen Statistik nicht nur gehalten, sondern sogar noch leicht verbessert.

3. Effekte von Nonresponse beim ALLBUS 2000

Die Analyse der Randverteilungen bei den ALLBUS-Untersuchungen ermöglicht eine grundsätzliche Verortung der Stichprobe anhand einiger zentraler soziodemografischer Grundparameter. Dies ist insbesondere dann bedeutsam, wenn die Ergebnisse auf die Bevölkerung insgesamt hochgerechnet werden sollen und Quantifizierungen vorgesehen sind. Auf der anderen Seite lässt sich aus einer akzeptablen Abbildung der Gesamtstruktur der Bevölkerung noch nicht hinreichend genau ableiten, dass deshalb auch andere, zum Beispiel komplexere Einstellungsmerkmale in der Stichprobe richtig verteilt sein müssen. Hierfür ist es nötig, zu überprüfen, ob es gelungen ist, im Rahmen der Feldarbeit auch solche Bevölkerungsgruppen zu erreichen, die in einem relevanten Anteil zur Grundge-

samtheit gehören,¹³ dabei jedoch aufgrund bestimmter inhaltlich begründbarer Faktoren schwerer zu befragen sind.

Derartige Fragestellungen lassen sich – mangels verfügbarer externer amtlicher Statistiken – nur auf Basis einer Nonresponse Analyse genauer beurteilen. Ihr Gegenstand besteht, vereinfacht ausgedrückt darin, zu prüfen, inwieweit sich die Verteilung der Respondenten von der Verteilung der Nonrespondenten bezüglich bestimmter vorab festgelegter Merkmale signifikant und bedeutsam unterscheidet, die als typisch für Nonrespondenten angesehen werden und die damit zu einem systematischen "nonresponse-bias" führen können.¹⁴

3.1. Typisierung von Nonresponse

Aus unserer Sicht lassen sich – in Anlehnung an Schnell (1997) – folgende sechs empirisch operationalisierbare Problembereiche, die zu einem stärkeren Nonresponse bei allgemeinen Bevölkerungsumfragen in Privathaushalten führen können, abgrenzen.

- **Unterschiedliche Erreichbarkeit**

Grundsätzlich kann davon ausgegangen werden, dass sich Personen in Privathaushalten je nach Statuspassage („live cycle factors“ – Westerhoven 1978) bzw. Lebensstil hinsichtlich ihrer Antreffbarkeit zu Hause unterscheiden. Typische empirische Indikatoren für „schlechtere“ Antreffbarkeit sind im Rahmen einer Nonresponse Analyse zum Beispiel die Haushaltsgröße (Single-Haushalte) und/oder der Erwerbsstatus (Vollzeit-Erwerbstätig bzw. Studium) bzw. – in Korrelation – die Zugehörigkeit zu bestimmten Altersgruppen, insbesondere jüngere Menschen zwischen Anfang 20 und Mitte 30 Jahren.

13 Diese Bedingung schließt zum Beispiel „Obdachlose“, Personen in Justizvollzugsanstalten, Bewohnerinnen und Bewohner von Pflegeheimen etc., bzw. bei Einwohnermeldeamts basierten Stichproben auch generell Personen, die in Deutschland nicht gemeldet sind, aus. Sie gehören definitionsgemäß nicht zur Grundgesamtheit der „Wohnbevölkerung in Privathaushalten“. Zum „Coverage“-Problem vergleiche Schnell (1991).

14 Als die nach wie vor umfassendste Studie zum Problem des Nonresponse kann für den deutschsprachigen Raum die von Schnell (1997) betrachtet werden. Die von uns gewählte Typisierung der Problembereiche, die mit einem stärkeren Nonresponse verbunden sein können, beruht im Wesentlichen auf den Ausführungen in dieser Arbeit. Ein Überblick findet sich ebenfalls bei Groves (1989: 186ff).

- **Krankheit und Gebrechlichkeit**

Abgesehen von dem Sonderproblem der Nicht-Befragbarkeit von bestimmten Bevölkerungsgruppen in persönlich-mündlichen Erhebungen (Taubstumme, schwerhörige bzw. sprachbehinderte Personen, geistig Behinderte, dementiell Erkrankte etc.) finden sich bei älteren bzw. hochbetagten Menschen deutlich häufiger chronische Erkrankungen, die zu einem Nonresponse führen können. Relevanter empirischer Indikator hierfür ist das Alter, sprich die Zugehörigkeit zur Altersgruppe ab 65 bzw. ab 70 Jahren.

- **Fehlende Kenntnisse der deutschen Sprache**

Fehlende Sprachkenntnisse sind insbesondere für ausländische Mitbürgerinnen und Mitbürger sowie für Zuwanderer eine Hürde, die zu einem verstärkten Nonresponse führen kann. Indikatoren hierfür sind die Nationalität sowie das Herkunftsland.¹⁵

- **Misstrauen aufgrund von Viktimisierungsängsten**

Kriminalitätsfurcht sowie daraus resultierendes Misstrauen und fehlende „Hilfsbereitschaft“ gelten insbesondere in Ballungsräumen bzw. städtischen Kerngebieten als weiterer Erklärungsansatz für Nonresponse. Indikatoren hierfür sind der Siedlungsstrukturtyp bzw. die Wohnquartiersform (Hochhäuser).

- **Umfragen als „Belastung“ bzw. kritische Distanz gegenüber der Datenerhebungen (Schutz der Privatsphäre, Datenschutz, „politisch“ motiviertes Misstrauen)**

Nonresponse als Ausdruck einer Belastung ist dann gegeben, wenn aus der Sicht des Befragten die Teilnahme im Widerspruch zu wichtigen persönlichen Zielen steht oder dadurch Nachteile erwartet werden. Dies ist zum Beispiel dann gegeben, wenn die Zielperson vermutet, dass die Anonymität nicht gewährleistet ist. Entsprechend motiviertes Misstrauen, das sich z. T. natürlich auch auf reale Erfahrungen (als Umfrage getarnte „kalte“ Akquise, Weiterverkauf von Adressen zum Zwecke des gezielten Marketings etc.) stützen kann, wird insbesondere bei bestimmten Bevölkerungsgruppen aus dem intellektuellen bzw. alternativen Milieu vermutet, denen vor dem Hintergrund der Erfahrungen mit der Volkszählung von 1987 nach wie vor eine besondere Distanz gegenüber Datenerhebungen nachgesagt wird. Indikatoren hierfür sind zum Beispiel Milieuvvariablen oder postmaterialistische Präferenzen.

15 Wie bereits erwähnt, gehören nicht in Privathaushalten gemeldete Personen nicht zur Grundgesamtheit von einwohnermeldeamtsbasierten Studien. Das Problem der fehlenden Sprachkenntnisse von zum Beispiel illegal eingewanderten Personen oder von Asylanten entfällt damit an dieser Stelle.

- **Fehlendes Interesse bzw. soziale Distanz**

Eine subjektiv geringere Bedeutsamkeit des Gegenstandes der Untersuchung bzw. ein postuliertes fehlendes Interesse an sozialwissenschaftlichen Themenstellungen aufgrund von Deprivation, fehlender Bildung oder eigener sozialer Distanz stellt eine besondere Variante eines für die professionelle empirische Sozialforschung (angeblich) typischen „Middle-Class-Bias“ dar (Esser/Grohmann/Müller/Schäffer 1989). Typische Indikatoren hierfür sind Bildung, Schichtzugehörigkeit oder auch Interesse an Politik und/oder Gesellschaft.

Im Rahmen einer empirischen Analyse ist demnach zu prüfen, inwieweit sich Respondenten und Nonrespondenten signifikant und bedeutsam in einem oder mehreren der genannten Merkmalsbereiche unterscheiden. Da sich die genannten Bereiche zum Teil überlappen, in Wechselwirkung zueinander stehen oder sich auch gegenseitig aufheben können (Beispiel: bessere Erreichbarkeit von Älteren versus chronische Krankheiten), ist es geboten, dies möglichst im Rahmen einer multivariaten statistischen Analyse zu tun.

3.2 Verfügbare Merkmale

Kernproblem jedweder Nonresponse Analyse ist die praktisch verursachte Schwierigkeit, in der Regel nur für Respondenten, jedoch nicht für Nonrespondenten über geeignete empirische Merkmalsausprägungen zu verfügen.¹⁶ Wir haben in unserer Arbeit einen besonderen (und sicherlich methodisch auch nicht gänzlich unproblematischen) Weg beschritten, um dieses grundsätzliche Problem mit einem alternativen Design zu umgehen.

In Ergänzung zu den im Rahmen der Ziehung der Adressen von den Einwohnermeldeämtern mitgelieferten Merkmale Alter, Geschlecht und Nationalität sowie dem regionalen Schichtungsmerkmal BIK-Siedlungsstrukturtyp (Kerngebiet, Randlage von Stadtregionen, Peripherie, Ländlicher Raum) haben wir die Bruttoadressen zusätzlich mit den extern verfügbaren und kommerziell vertriebenen mikrogeografischen Strukturmerkmalen der Fa. Microm – den so genannten MOSAIC Informationen – angereichert.¹⁷ Hierbei han-

16 Völlig zu Recht stellt auch Schnell fest, dass die (...) „empirische Analyse von individuellem Nonresponse (...) vor dem Problem (steht), dass Nonrespondenten gerade darüber definiert werden, dass für Sie weniger (bzw. i.d.R. sogar überhaupt keine – US/IL) Daten vorliegen, als für den Rest der untersuchten Population.“ (Schnell 1997: 134).

17 Microm GmbH und Sinus Sociovision GmbH: MOSAIC Informationen. Stand 2000. Neuss, Frühjahr 2000. Seit dem Jahr 2002 liegt eine überarbeitete Version vor, mit der die neue Typologie von Sinus Sociovision nun auch für die neuen Bundesländer implementiert ist. Wir beziehen uns noch auf die alte Typologie, da wir die Informationen bereits im Jahre 2001 den

delt es sich um eine Datenbank, die das Wohnumfeld der einzelnen Haushalte in Deutschland beschreibt. In der Basisdatenbank sind Informationen zu folgenden Sachverhalten enthalten (MOSAIC Variablen):

- Art der Bebauung und Quartierstypen,
- Sozioökonomische Merkmale (Status, Ausländeranteil, Anonymitätsbedürfnis, Kreditausfallrisiko),
- Demografie (Alter, Familienstruktur).

Die Informationen sind mit Hilfe eines methodisch leider nicht näher erläuterten mikrogeografischen Segmentierungsverfahrens so verdichtet worden, dass sie als Umfeldcharakteristik prinzipiell jedem Haushalt in Deutschland zugespielt werden können.

Darüber hinaus stehen verschiedene Typologien zur Verfügung. Interessant für unsere Analyse sind die so genannten Sinus-Milieus. Hierbei werden Personengruppen anhand von typischen Einstellungen und Wertorientierungen zu 10 Milieus zusammengefasst, die sich nach der jeweiligen sozialen Lage bzw. Schichtzugehörigkeit (Unterschicht bis Oberschicht) sowie nach ihrer Grundorientierung von „traditionell“ bis „postmodern“ unterscheiden (vergleiche dazu die Kurzcharakterisierung im Anhang). Die Sinus-Typologie ist im Bereich der sozialwissenschaftlichen Forschung bisher eher unüblich und von daher wenig rezipiert. Mit offenbar großem Erfolg wird die Typologie hingegen seit Jahren insbesondere im Bereich der Medienforschung eingesetzt.¹⁸ Der Informationsbestand der Microm-Datenbank wurde mit den Sinus-Milieus zu den MOSAIC Milieus verknüpft. Dazu wurden ebenfalls nicht näher spezifizierte Modelle entwickelt, welche für jedes Haus die statistische Wahrscheinlichkeit berechnen, mit der die Milieus darin vorkommen, und welches davon das dominante Milieu ist. Bis 2002 waren die Sinus-Milieus in Form der MOSAIC Adaption nur für Westdeutschland verfügbar. Aus diesem Grunde können wir die Sinus-Milieus im Rahmen der Nonresponse Analyse auch nur für die alten Bundesländer (inkl. Berlin/West) ausweisen.¹⁹

Natürlich müssen die angereicherten Daten der Fa. Microm vorsichtig beurteilt werden. Sie bilden das Wohnumfeld der Haushalte ab und basieren selber auf diversen aggregier-

ALLBUS-Daten angereichert haben. Vergleiche dazu auch die Informationen auf der Website der Fa. MICROM unter <http://www.microm-online.de>.

18 Bei den Sinus-Milieus handelt es sich um ein urheberrechtlich geschütztes Verfahren der Fa. Sinus Sociovision GmbH. Eine Referenz- und Literaturliste findet sich auf der Homepage von Sinus Sociovision unter <http://www.sinus-milieus.de>.

19 Eine nachträgliche Anreicherung der ALLBUS-Daten mit den neuen MOSAIC Milieus ist nicht möglich, da die Adressen des ALLBUS 2000 aus datenschutzrechtlichen Gründen bereits vernichtet wurden.

ten Datenbeständen, die zu diesem Zwecke zusammengeführt wurden. Wir möchten an dieser Stelle jedoch noch einmal den eigentlichen Zweck dieses Zugangs in Erinnerung rufen. Es geht uns darum, relevante und im ALLBUS ansonsten nicht vorhandene Variablen zu generieren, mit denen sich Respondenten und Nonrespondenten hinsichtlich bestimmter Merkmale beschreiben und ggf. voneinander abgrenzen lassen. Darüber hinaus haben wir dem ALLBUS-Datensatz noch zusätzlich ausgewählte Merkmale einer von uns im Frühjahr 2001 durchgeführten Interviewerbefragung beigelegt.²⁰ Die ausgewählten Variablen bilden näherungsweise die in der Literatur häufig als relevant betrachteten interviewerbezogenen Einflussfaktoren „Persönliche Interviewermerkmale“, „Erfahrung“, „Art der Kontaktstrategie“ sowie „Organisation des Feldeinsatzes“ ab.²¹ Übersicht 4 gibt einen Gesamtüberblick zu den im Rahmen der Nonresponse Analyse verfügbaren Variablen.

Da sich im Rahmen der Analyse keinerlei Hinweise darauf fanden, dass die zugespielten Wohnumfeldinformationen nicht trennscharf sind, spricht aus unserer Sicht nichts dagegen, in diesem Fall diesen ansonsten eher unüblichen und methodisch noch nicht validierten Zugang der Anreicherung einer Stichprobe mit externen Datenbeständen zu wählen.²²

20 Befragt wurden 451 ALLBUS-Interviewer. Mit 392 auswertbaren Fragebögen betrug die Rücklaufquote 87 Prozent.

21 Vergleiche hierzu im Überblick ebenfalls Schnell (1987: 195ff).

22 Eine Korrelationsanalyse der angereicherten MOSAIC Daten mit den Ergebnissen des ALLBUS (Netto-Stichprobe) ergab hoch korrelierte Zusammenhänge zum Beispiel zwischen MOSAIC Lebensform (Single- versus Familiendichte) und den ALLBUS-Daten zur Haushaltsgröße, zum Familienstand und zur Zahl der Kinder im Haushalt. Ebenfalls gut korreliert ist das Merkmal MOSAIC Bebauungstyp mit der ALLBUS-Variable Art des Gebäudes gemäß vorliegender Interviewereinstufung. Da die Interviewereinstufung zur Art des Gebäudes im ALLBUS 2000 für einen Großteil (knapp 40 Prozent) der Nonrespondenten leider nicht vorliegt, konnte auf diese und einige weitere unmittelbar erhobene Wohnquartiersmerkmale nicht zurückgegriffen werden.

Übersicht 4 : Response versus Nonresponse – Verfügbare Merkmale und Operationalisierung

Testhypothese:

Respondenten und Nonrespondenten unterscheiden sich signifikant und hinreichend stark bezüglich einer oder mehrerer der folgenden Dimensionen.

Sample-Point bezogene Faktoren

1. Unterschiedliche Erreichbarkeit

- | | |
|---------------------|---|
| MOSAIC Lebensform | Single-Dichte versus Familien-Dichte im Wohngebiet |
| MOSAIC Bebauungstyp | Wohnblocks/Hochhäuser versus mittelgroße- / Familienwohnhäuser |
| BIK-Typ | Kerngebiet versus Randlage / Peripherie / ländlicher Raum |

2. Gebrechlichkeit / Krankheit

- | | |
|-------|---|
| Alter | Zielpersonen ab 70 Jahre versus sonstige Altersgruppen |
|-------|---|

3. Fehlende Kenntnisse der deutschen Sprache

- | | |
|--------------|-------------------------------------|
| Nationalität | Nicht-Deutsch versus Deutsch |
|--------------|-------------------------------------|

4. Misstrauen aufgrund von Viktimisierungserfahrungen

- | | |
|--|--|
| BIK-Typ | Kerngebiet versus Randlage / Peripherie / ländlicher Raum |
| MOSAIC Bebauungstyp | Wohnblocks/Hochhäuser versus mittelgroße bzw. Familienhäuser |
| 5. Belastung bzw. kritische Schutz der Privatsphäre) | <i>Distanz gegenüber Umfrageforschung (Thematisches Misstrauen /</i> |

- | | |
|------------------------|---|
| Prägendes Sinus-Milieu | Intellektuelles Milieu versus statusorientierte- / adaptive- / sonstige Milieus (nur für die alten Bundesländer verfügbar) |
|------------------------|---|

- | | |
|----------------------|--|
| Anonymitätsbedürfnis | Wohngebiete mit überdurchschnittlichem Anonymitätsbedürfnis versus unterdurchschnittlicher Anonymität |
|----------------------|--|

6. Fehlendes Interesse und/oder generelle soziale Distanz

- | | |
|---------------------|--|
| MOSAIC Status | Wohngebiete mit unterdurchschnittlichem Status versus Gebiete mit überdurchschnittlichem Status |
| MOSAIC Bebauungstyp | Wohnblocks/Hochhäuser versus mittelgroße bzw. Familienhäuser |

Interviewerbezogene Faktoren

1. Persönliche Merkmale

- | | |
|-----------------------|---|
| Intervieweralter | 65 u. älter versus Mittelalt / Jung |
| Schulabschluss | Hochschule/Abitur versus mittlere Reife / Hauptschule |
| Geschlechterkongruenz | Gleich- versus gegengeschlechtliche Interviewsituationen |

2. Erfahrung / Kontaktverhalten

- | | |
|-------------------------|--|
| Interviewererfahrung | 10 u.m. Jahre versus kürzere Erfahrungsjahre |
| Präferiertes Verfahren | AR versus RR / Quote |
| Art der Kontaktaufnahme | Interview ohne Vorankündigung: oft versus gelegentlich/selten |

3. Feldorganisation

- | | |
|-------------------|---|
| Einsatzentfernung | unter 40 km versus 40 km und mehr vom eigenen Wohnort entfernt |
| Arbeitsbelastung | Zeit bis zum ersten realisierten Interview im Point: max. 2 Wochen versus längerer Bearbeitungsdauer |

4. Ergebnisse

Auf Basis der vorhandenen Variablen haben wir 3 Modelle berechnet, mit denen der Einfluss der Merkmale auf die Verteilung der Bruttostichprobe nach Response und Nonresponse geschätzt werden kann. Das von uns gewählte Design einer multivariaten logistischen Regression (Logit-Modell) ermöglicht es, zu überprüfen, inwieweit der Stichproben-Nonresponse beim ALLBUS 2000 auf eine oder mehrere der vorab genannten Stichprobenaufälle zurückzuführen ist.²³ Ergibt sich ein statistisch signifikanter und hinreichend starker Zusammenhang, so muss von einer systematischen Verzerrung, sprich von einem „non response bias“ ausgegangen werden. Lässt sich dies nicht nachweisen, so kann dies aus unserer Sicht als Beleg dafür gewertet werden, dass sich der Stichproben-Nonresponse weitgehend unsystematisch und damit gemäß der Grundlogik einer zufallsbasierten Stichprobe verteilt. Die Ergebnisse der Nonresponse-Analyse sind in der folgenden Übersicht 5 zusammengefasst. Ausgewiesen sind die Regressionskoeffizienten sowie das statistische Signifikanzniveau.²⁴ Als hinreichend gesichert sind grundsätzlich nur statistisch signifikante Ausprägungen zu betrachten. Das Vorzeichen der Regressionskoeffizienten gibt an, inwieweit die Merkmalsausprägung die Wahrscheinlichkeit in Richtung Nonresponse (-) oder Response (ohne Vorzeichen) verändert. Da es sich bei der Mehrzahl der Variablen um kategoriale Merkmale handelt, ist ggf. die Referenzkategorie, auf die sich die Effekte der anderen Ausprägungen im Vergleich beziehen (Dummy bzw. Indikator-Codierung), gleich zu Beginn benannt. Zur Beurteilung der Schätzstärke, sprich der Erklärungskraft des Gesamtmodells, ist am Ende das sog. Pseudo-R² (nach McFadden) ausgewiesen.²⁵ Laut Andreß/Hagenaars/Kühnel deuten Werte unter 0,05 auf einen eher geringen und demnach zu vernachlässigenden Zusammenhang sowie Werte über 0,2 auf einen starken Zusammenhang. Werte über 0,4 sind hingegen nur sehr selten zu erreichen (Andreß/Hagenaars/Kühnel 1997: 288).

23 Mit Hilfe eines multivariaten Logit-Modells lässt sich prüfen, inwieweit bestimmte Variablen, die eines oder mehrere der vorab spezifizierten Konstrukte operationalisieren, als stabile Prädiktoren für Nonresponse zu interpretieren sind. Sofern die Variablenausprägungen signifikant und das Modell auch insgesamt trennscharf genug ist (goodness of fit), kann dies im Unterschied zu einfachen bivariaten signifikanten Abweichungen als systematische Determination interpretiert werden. Zur Notwendigkeit von multivariaten Analysen vergleiche die Arbeit von Rendtel (1995: 198ff).

24 Aus Gründen der Übersichtlichkeit haben wir auf die Ausweisung von Logits verzichtet. Alle Modelle wurden mit der Prozedur NomReg unter SPSS 11.0 berechnet.

25 Das Pseudo-R² gibt an, um wie viel sich der Fehler des Modells unter dem Einfluss der Prädiktoren reduziert (PRE: proportional reduction of error). Es lässt sich demnach analog zum Determinationskoeffizienten der linearen Regression als Grad der Varianzaufklärung interpretieren (Andreß/Hagenaars/Kühnel 1997: 287f).

Übersicht 5: Nonresponse Analyse: ALLBUS 2000, alte Bundesländer

Logistische Regression: Nonresponse versus Response

	Modell 1 Einfache Merkmale der Melderegister		Modell 2 Erweiterte Merkmale inkl. MOSAIC Daten		Modell 3 Sample-Point- und Interviewmerkmale	
	West	Ost	West	Ost	West	Ost
Geschlecht	,082	,060	-,068	,068	(redund.)	(redund.)
Altersklassen Ref.-Kat.: 70 und älter						
18-29	,453**	,515*	,432**	,500**	,527**	,749**
30-39	,148	,317	,125	,294	,238	,505*
40-49	,269*	,103	,246*	,056	,374**	,144
50-59	,253*	,283	,211	,251	,312*	,422*
60-69	-,001	,237	-,021	,202	,040	,106
Keine Angabe (EWMA)	,572**	-,053	,508*	,021	,170	-,060
Nationalität Ref.-Kat.: Nicht Deutsch						
Deutsch	-,066	-,048	-,130	,016	-,227	-,028
Keine Angabe (EWMA)	,128	,229	,082	,332	,006	,335
BIK-Typ Ref.-Kat.: Kerngebiet						
Randlage Stadtregion	,218*	,350*	,097	,310	-,000	,872**
Peripherie	-,195*	,026	-,336**	-,023	-,347*	,028
Ländlicher Raum	,228**	,643**	,045	,645**	-,109	,816**
MOSAIC Lebensform Ref.-Kat.: überdurchschn. Singles						
leicht überdurchschn. Singles			,064	,457*	,069	,560*
gemischtes Wohngebiet			,039	,295	-,007	,248
leicht überdurchschn. Familien			,064	,345	-,052	,145
überdurchschn. Familien			,093	,001	,114	-,023
Keine Angabe			-,253	(redund.)	,113	(redund.)
Mosaic Bebauungstyp Ref.-Kat.: Wohnblöcke						
1-2 Familienhäuser			,233	,262	,348*	,412*
3-5 Familienhäuser			,036	-,010	,139	,206
6-9 Familienhäuser			-,024	,251	,200	,482*
Keine Angabe			-,150	-,055	-,453	-,421
MOSAIC Status Ref.-Kat.: unterdurchschn. Status						
leicht unterdurchschn. Status			,038	,210	,086	,216
durchschnittl. Status			,044	,100	-,039	,162
leicht überdurchschn. Status			,132	,133	,075	,218
überdurchschn. Status			,059	,338*	-,041	,373
Keine Angabe			(redund.)	(redund.)	(redund.)	(redund.)

Fortsetzung Nonresponse Analyse:						
	Modell 1		Modell 2		Modell 3	
	Einfache Merkmale der Melderegister		Erweiterte Merkmale inkl. MOSAIC Daten		Sample-Point- und Interviewermerkmale	
	West	Ost	West	Ost	West	Ost
Prägendes Sinus-Milieu (West)						
Ref.-Kat.: Intellekt. Milieu						
Etabliertes Milieu			,001	-	,118	-
Postmodernes Milieu			-,096	-	-,178	-
Modernes bürgerl. Milieu			,031	-	,219	-
Adaptives Milieu			,069	-	-,070	-
Statusorientiertes Milieu			,325*	-	,280	-
Trad. bürg. Milieu			,044	-	,122	-
Trad. Arbeitermilieu			-,232	-	-,228	-
Hedonistisches Milieu			,203	-	,236	-
Konsum-materialist. Milieu			,048	-	-,053	-
Keine Angabe			1,471	-	1,611	-
Anonymitätsbedürfnis						
Ref.-Kat.: überdurchschnittl.						
Unterdurchschnittlich			-,054	,150	,136	,134
Leicht unterdurchschnittlich			-,131	,118	,024	,153
Mittlere Anonymität			,080	-,033	,226	-,012
Leicht überdurchschnittl.			-,138	,009	,006	,000
Keine Angabe			-,749	(redund.)	-,618	(redund.)
Alter des Interviewers						
Ref.-Kat.: 65 und älter						
-39					-,231	-,899*
40-54					,119	,220
55-64					,195	-,141
Schulabschluss Interviewer						
Ref.-Kat.: Abitur						
Volksschule					,172	,311
Mittlere Reife					-,014	-,085
Geschlechtkongruenz						
Ref.-Kat.:						
Z-weiblich / I-weiblich						
Z-männlich / I-männlich					-,046	-,123
Z-männlich / I-weiblich					-,120	,095
Z-weiblich / I-männlich					-,220*	-,049
Interviewererfahrung						
Ref.-Kat.: 10 u.m. Jahre						
bis 2 Jahre					-,254	,636
3-5 Jahre					-,089	-,259
6-9 Jahre					,101	-,191
Präferiertes Verfahren						
Ref.-Kat.: Feste Adressen						
Random-Route					-,106	-,477**

Quote					-,036	-,601**
Fortsetzung Nonresponse Analyse:						
	Modell 1		Modell 2		Modell 3	
	Einfache Merkmale der Melderegister		Erweiterte Merkmale inkl. MOSAIC Daten		Sample-Point- und Interviewermerkmale	
	West	Ost	West	Ost	West	Ost
Kontakt ohne Vorankündigung						
Ref.-Kat.: Oft					-,139	-,675**
Gelegentlich					-,401**	-,456**
Selten						
Einsatzradius						
Ref.-Kat.: unter 40 km						
40 bis 60 km					-,143	,519**
60 km und mehr					,253*	,075
Keine Angabe					,389	-1,404
Dauer bis zum Erstinterview						
Ref.-Kat.: unter 14 Tage						
2-4 Wochen					-,085	,347*
5-12 Wochen					-,031	,001
Mehr als 12 Wochen					-,207	,073
Modellkonstante	-,389**	-,265	-,462	-,950	-,045	-,258
Pseudo R2 (Mc-Fadden)	0,010	0,021	0,018	0,034	0,047	,099

Binär-logistische Regression (SPSS 11.0 - NomReg): Nonresponse (0) versus Response (1)

B-Koeffizienten mit * = Sig.-Niveau < 0,05 bzw. mit ** = Sig.-Niveau < 0,01

(redund.) = Freiheitsgrad (bzw. Merkmal) wurde aufgrund von Redundanzen in der Design-Matrix reduziert
Infratest Sozialforschung

Wie nicht anders zu erwarten war, klärt das **erste Modell**, bei dem nur die soziodemografischen Grundinformationen der Einwohnermeldeämter in die Berechnung eingehen, so gut wie keine Varianz auf. Mit einem Pseudo-R² von 0,01 für die alten und 0,02 für die neuen Bundesländer kann es – trotz des signifikanten Einflusses der Variablen *Alter* und *BIK* – im Prinzip vernachlässigt werden. Interessant ist, dass das Merkmal *Nationalität* beim ALLBUS 2000 selber keinen Einfluss auf den Nonresponse hat. Nicht-deutsche Staatsbürger sind im Vergleich zur deutschen Wohnbevölkerung in der Netto-Stichprobe nicht signifikant anders vertreten. Koch weist in diesem Zusammenhang allerdings darauf hin, dass die Ausfälle aufgrund fehlender Sprachkenntnisse durch eine größere Teilnahmereitschaft bei denjenigen Ausländern, die der deutschen Sprache mächtig sind, kompensiert werden (Koch 1997).

Bemerkenswerterweise liefert auch das **zweite Modell**, bei dem die MOSAIC Variablen mit einbezogen sind, mit einem Pseudo-R² von knapp 0,02 für die alten und etwas über 0,03 für die neuen Bundesländer ebenfalls keine befriedigende Schätzstärke. Im Einzelnen lässt sich auch hier wieder der Effekt beim *Alter* nachzeichnen. Senioren ab 70 Jahren

sind im Vergleich zu 18-29jährigen und in den alten Bundesländern auch zu den 40-49jährigen signifikant weniger bei der Gruppe der Respondenten vertreten. Dieser Effekt war für ältere Zielpersonen zu erwarten und spricht, da der ALLBUS als allgemeine Bevölkerungsumfrage nicht gesondert auf Erhebung in Haushalten von hochbetagten Menschen optimiert ist, nicht grundsätzlich gegen die Qualität der Umfrage. Demgegenüber weist die Variable *BIK* nur in den neuen Ländern auf einen etwas stärkeren Nonresponse in Kerngebieten im Vergleich zum ländlichen Raum hin. Andere Indikatoren, wie zum Beispiel der *MOSAIC* *Bebauungstyp* sind hingegen zum Beispiel bezüglich Ausschöpfung in Hochhausgebieten nicht signifikant. Bei den *Sinus-Milieus*, die zum Zeitpunkt der Anreicherung der Daten nur in den alten Ländern verfügbar waren, findet sich beim intellektuellen Milieu nur im Vergleich zum statusorientierten Milieu ein etwas stärkerer Nonresponse. Dieser eher singuläre Effekt lässt sich jedoch kaum als Beleg für eine mögliche Unterausschöpfung von „befragungskritischen“ Zielpersonen aus postmaterialistischen Milieus interpretieren. Gleiches gilt für die zwei weiteren signifikanten Effekte in den neuen Ländern. Die leicht stärkere Ausschöpfung bei Zielpersonen mit *überdurchschnittlichem Status* ist für sich genommen genauso singulär wie die etwas bessere Ausschöpfung von Zielpersonen aus *leicht überdurchschnittlichen Single-Wohngebieten* im Vergleich zu *überdurchschnittlichen Single-Wohngebieten*.

Im **dritten Modell** werden zusätzlich einige relevante Interviewermerkmale mit einbezogen.²⁶ Wir wechseln damit in der Betrachtung die Perspektive. Hatten vorher die Effekte von Nonresponse im Vordergrund gestanden, die zu systematischen Verzerrungen führen können, so geht es jetzt primär um die unterschiedlichen Einflussfaktoren – Sample-Point bezogene und / oder interviewerbezogene – die den beim ALLBUS 2000 festgestellten hohen Nonresponse mit beeinflusst haben.²⁷ Für die alten Bundesländer ergibt sich auch

26 Methodisch muss an dieser Stelle angemerkt werden, dass sich Modell 3 im Vergleich zu den beiden anderen Logit-Modellen auf eine geringere Fallzahl von $n = 3.289$ im Vergleich zu $n = 4.400$ für die alten Länder bzw. $n = 1.529$ im Vergleich zu $n = 2.050$ für die neuen Länder stützt. Die Reduktion der Fallzahl ergibt sich aufgrund der Zuspiegelung der Ergebnisse der Interviewerbefragung, bei der nicht für alle ALLBUS-Interviews gültigen Angaben realisierbar waren. Streng genommen können deshalb auch die Ergebnisse des dritten Modells aufgrund seiner reduzierten Fallzahl nicht unmittelbar mit den beiden anderen Modellen verglichen werden. Da sich jedoch auch dann, wenn man alle drei Modelle mit der gleichen reduzierten Fallzahl berechnet, im Ergebnis keine relevanten Unterschiede ergeben, haben wir aus Gründen der Übersichtlichkeit trotzdem diese Form der vergleichenden Darstellung inklusive der Ergebnisse des dritten Modells gewählt.

27 Die Verteilung der Interviewer auf die einzelnen Sample-Points erfolgt bei allgemeinen Bevölkerungsumfragen nicht systematisch. Eingesetzt werden im Prinzip die CAPI-Interviewer, die vor Ort verfügbar sind. Im Prinzip kann die Verteilung der Interviewer auf die Sample-Points von daher als im weitesten Sinne randomisiert betrachtet werden. Abgewichen wurde hiervon erst im

bei dieser erweiterten Betrachtung mit einem Pseudo- R^2 von knapp unter 0,05 ein nach wie vor schwacher und damit wenig aussagekräftiger Zusammenhang. Für die neuen Bundesländer wächst das Pseudo- R^2 hingegen auf einen akzeptableren Wert von knapp unter 0,1.

Abgesehen von der bereits erwähnten Unterrepräsentierung von *älteren Zielpersonen ab 70 Jahren* erweisen sich von den Sample-Point bezogenen Faktoren jetzt auch *große Wohnblocks* im Vergleich zu 1-2 Familienhäusern etwas schlechter ausgeschöpft. Zumindest in den neuen Bundesländern würde dies, zusammen mit dem hier ebenfalls feststellbaren größeren Nonresponse in *Kerngebieten*, auf mögliche Viktimisierungängste, die zu einer systematisch geringeren Teilnahmebereitschaft führen, hindeuten. Da dieser Effekt jedoch erst im dritten Modell signifikant wird, ist dies vorrangig auf Wechselwirkungen mit interviewerbezogenen Faktoren zurückzuführen. In der eigentlichen Nonresponse-Analyse (Modell 2) war dieses Merkmal für sich genommen nicht signifikant. Es schlägt demnach bei der Beurteilung der Güte der Stichprobe für sich genommen nicht durch.

Bei den interviewerbezogenen Faktoren kommt vor allen Dingen die *Art des Kontaktierens* zum Tragen. Wer eher zurückhaltender ist und nicht unmittelbar und ohne Vorankündigung oder Terminabsprachen versucht, das Interview zu realisieren, ist etwas weniger erfolgreich. Der im Vergleich größere Erfolg von Interviewern, die auch *weiter von ihrem Wohnort* entfernt Befragungen durchführen, hat mit der Feldorganisation zu tun. Im Rahmen der Nachbearbeitung wurden Wechsel mit erfolgreicheren Interviewern vorgenommen, auch wenn diese ggf. dafür weitere Reisen in Kauf nehmen mussten. In den neuen Bundesländern fällt als Merkmal auch die Präferenz von *Studien mit vorgegebenen Adressen* auf. Darüber hinaus waren hier *jüngere Interviewer* unter 40 etwas weniger erfolgreich. Bezüglich des Geschlechts lässt sich für die alten und neuen Bundesländer feststellen, das *männliche Interviewer*, die weibliche Zielpersonen befragen, ebenfalls etwas schlechter abschneiden.

Alle sonstigen Merkmale, wie zum Beispiel Bildungsniveau oder auch Institutserfahrung, sind hingegen nicht signifikant. Insgesamt erhöhen die interviewerbezogenen Faktoren die Schätzstärke des Modells insbesondere in den neuen Bundesländern in einem gewissen Umfang. Dies deutet auf den vergleichsweise stärkeren Einfluss von interviewerbezogenen Faktoren als Ursache für Nonresponse hin. Im Einzelnen scheint es vor allem der Habitus, sprich die konkrete Art und insbesondere die Offensivität beim Kontakten zu

Rahmen der Nachbearbeitung, wo ggf. besonders erfolglose Interviewer durch andere ersetzt wurden. Im Falle eines Interviewerwechsels wurden von uns entweder die Merkmale des zuletzt erfolgreichen Interviewers oder aber, im Falle eines Nonresponses, die Merkmale des Interviewers, der im Rahmen der Hauptbearbeitung zuerst kontaktiert hat, zugespielt.

sein, die den Unterschied ausmacht. Die Erfahrung spielt beim ALLBUS 2000 hingegen eine eher geringere Rolle.²⁸ Alles in allem klärt jedoch auch das dritte Modell insbesondere in den alten Ländern nur relativ wenig Varianz auf.²⁹

Ausschließlich zur Veranschaulichung der Ergebnisse des Logit-Modells möchten wir eine einzelne bivariate Verteilung ausweisen: die Struktur der gültigen Brutto- und die der tatsächlich realisierten Nettostichprobe nach den Sinus-Milieus. Wie aus Übersicht 6 hervorgeht, ergeben sich zwischen Bruttovorgabe und realisierter Nettostruktur auch quantitativ keine wirklich relevanten Unterschiede. Die Abweichungen sind zwar bivariat signifikant, jedoch quantitativ wenig bedeutsam.

Übersicht 6: Verteilung MOSAIC Sinus-Milieus im ALLBUS 2000

Angaben in Prozent	Brutto ¹⁾	Netto
MOSAIC Sinus-Milieus ²⁾		
Etabliertes Milieu	8	8
Intellektuelles Milieu	9	8
Postmodernes Milieu	5	4
Modernes bürgerliches Milieu	7	7
Adaptives Milieu	6	6
Statusorientiertes Milieu	20	23
Traditionelles bürgerliches Milieu	14	13
Traditionelles Arbeitermilieu	5	4
Hedonistisches Milieu	7	8
Konsum-materialistisches Milieu	8	8
Keine Angabe	11	11

1) „Gültiges“ Stichprobenbrutto ohne qualitätsneutrale Ausfälle

2) Angaben sind hier nur für die alten Bundesländer verfügbar

Infratest Sozialforschung, eigene Berechnungen

Definitionsgemäß lassen zufallsbasierte Stichproben mit einer mathematisch berechenbaren Wahrscheinlichkeit Rückschlüsse auf die Grundgesamtheit zu. Diese Aussage darf

28 Schnell weist darauf hin, dass neben der Kontaktstrategie auch der Kontaktzeit eine größere Bedeutung zukommt. Für den anglo-amerikanischen Raum wird die Zeit zwischen 17 und 22 Uhr als beste Zeit für einen ersten Kontakt angegeben. Entsprechende Daten zum Erstkontakt beim ALLBUS 2000 liegen uns allerdings nicht vor, so dass wir dieses Merkmal nicht mit einbeziehen können (Schnell 1997: 228).

29 Dieser Befund deckt sich mit der Interpretation von Schnell, nach dessen Ansicht es bisher nicht gelungen sei, fixe Interviewermerkmale erfolgreich zur Prognose von Ausschöpfungsquoten zu verwenden (Schnell 1997: 195).

jedoch nicht so interpretiert werden, dass man zum Beispiel dann, wenn man eine ganze Reihe von Zufallsstichproben parallel ziehen würde, *ceteris paribus* keine signifikanten Einzelabweichungen bekäme. Insbesondere bei komplexen Stichprobendesigns, wie sie heute in den Sozialwissenschaften üblich sind, ist vielmehr das Gegenteil hochwahrscheinlich, ohne dass damit hinreichend sicher geklärt wäre, dass die inkriminierte Stichprobe in ihren Ergebnissen auch insgesamt „schlechter“ sei. Entscheidend ist nicht, ob einzelne Abweichungen von einer definierten Referenz zum Beispiel nach dem χ^2 -Test, signifikant sind. Es kommt vielmehr darauf an, zu prüfen, ob die Stichprobe ein systematisches Verzerrungsmuster aufweist und – immer vorausgesetzt, dass mögliche einzelne Abweichungen nicht aus zum Beispiel formalen Gründen als nicht mehr hinnehmbar betrachtet werden – wie bedeutsam die Verzerrungen für die zu untersuchenden Fragestellungen sind. Genau aus diesem Grund haben wir eine multivariate Nonresponse-Analyse anhand von Logit-Modellschätzungen durchgeführt, da sich nur auf diese Weise signifikante und bedeutsame Muster eines möglichen „nonresponse bias“ im Zusammenhang überprüfen lassen.³⁰

Auf Basis der durchgeführten Nonresponse-Analyse beim ALLBUS 2000 ergeben sich für die alten Länder – trotz der „geringen“ Ausschöpfung von nur 46,9 Prozent – so gut wie keine und für die neuen Länder – bei einer etwas besseren Ausschöpfung von 53,7 Prozent – nur schwache Anhaltspunkte dafür, dass die eingangs dargestellten Effekte von Nonresponse in irgendeiner relevanten Weise zu systematischen Verzerrungen geführt hätten. Die Gruppen, die üblicherweise einen höheren Nonresponse aufweisen, sind im ALLBUS 2000 hinreichend repräsentiert.

5. Wirkungen der Nachbearbeitung

Eines der Spezifika des ALLBUS-Verfahrensmodells der Jahre 1994, 1996 und 2000 besteht darin, dass im Anschluss an die eigentliche Hauptbearbeitungswelle, die Adressen, die nicht erfolgreich ausgeschöpft werden konnten, in mehreren Wellen nachbearbeitet wurden. Die Nachbearbeitung stützt sich demnach im Grundsatz auf die gleichen Adressen wie die Hauptbearbeitung.³¹

30 Vergleiche hierzu insbesondere Rendtel (1995: 198ff.), der diesen Zugang exemplarisch am Beispiel des Sozio-ökonomischen Panels begründet und angewendet hat.

31 Eine Ausnahme stellt der im Rahmen der Nachbearbeitung vorgenommene Austausch von solchen Befragungsklumpen dar, bei denen in der Hauptbearbeitung aus den 10 vorgegebenen Adressen gar kein oder nur ein Interview realisiert werden konnte. In diesem Fall wurde ein Reserveklumpen eingesetzt. In etwa 15 Prozent der Adressen der Nachbearbeitung wurden neu eingesetzt. 85% waren hingegen bereits im Rahmen der Hauptbearbeitung kontaktiert worden.

Es ist vor allem die Intensität der Nachbearbeitung, die zu dem von Koch (2002) dargestellten Anstieg der Fallpreise von 170 DM beim ALLBUS 1994 auf 250 DM beim ALLBUS 2000 geführt hat.³² Um die Frage der Effektivität der Nachbearbeitung unter qualitativen Gesichtspunkten beurteilen zu können, möchten wir unser Nonresponse-Analysemodell auf den Vergleich der Verteilung der Hauptbearbeitung mit den Effekten der Nachbearbeitung ausdehnen. Testhypothese ist hierbei, dass sich die Verteilung der Hauptbearbeitungswelle signifikant und entsprechend stark anhand der für Nonresponse charakteristischen Merkmale von der im Rahmen der Nachbearbeitung realisierten Verteilung unterscheidet.

5.1 Operationalisierung

Die Operationalisierung der Merkmale erfolgt in diesem Fall anhand von ausgewählten Variablen, die in den ALLBUS-Untersuchungen 1994, 1996 und 2000 gemeinsam verfügbar sind (vergleiche dazu Übersicht 7).

Berechnet wurde für jede der drei genannten ALLBUS-Querschnittserhebungen anhand der jeweils gleichen Variablen ein eigenständiges Logit-Modell. Die Ergebnisdarstellung erfolgt wiederum genauso, wie vorher bei der eigentlichen Nonresponse-Analyse. Ein negatives Vorzeichen (-) bei den Regressionskoeffizienten deutet auf eine stärkere Realisierung im Rahmen der Nachbearbeitung hin. Maßgeblich sind natürlich auch hier nur signifikante Zusammenhänge. Bei der Interpretation der Ergebnisse ist neben der Beurteilung der Stärke des Schätzmodells in diesem Fall auch von Bedeutung, inwieweit sich die möglichen Effekte konstant bei allen 3 ALLBUS-Untersuchungen vergleichbar ausprägen. Die Anlage der Nachbearbeitung als Neukontaktierung bereits vom Interviewer angegangener Adressen setzt ja geradezu voraus, dass sich die Wellen systematisch unterscheiden. Die eigentlich interessante Frage ist, inwieweit sich anhand der Nachbearbeitung bei allen drei ALLBUS-Untersuchungen mit Registerstichprobe ein einheitliches Muster bezüglich solcher Gruppen aufzeigen lässt, die bei einer „normalen“ Bearbeitung schlechter befragbar und damit anfälliger für systematischen Nonresponse sein müssten.

32 Eine weitere Ursache für den deutlich höheren Fallpreis des Jahres 2000 ergibt sich auch dadurch, dass in diesem Fall aufgrund der Umstellung von PAPI (paper pencil assisted personal interviewing) auf CAPI (computer assisted personal interviewing) eine zusätzliche Methodenstudie mit parallelisierten CAPI und PAPI-Interviews pro Point durchgeführt wurde.

Übersicht 7: Wirkungen und Effekte der Nachbearbeitung der ALLBUS-Untersuchungen – Operationalisierung anhand von geeigneten ALLBUS-Variablen (ALLBUS 94, 96, 2000)

Testhypothese:

Die Struktur der Ergebnisse der Hauptbearbeitung unterscheidet sich in einer oder mehrerer der folgenden Dimensionen signifikant und hinreichend stark von den Ergebnissen der Nachbearbeitung

1. <i>Erreichbarkeit</i>	
Haushaltsgröße	Single-Haushalt versus größere Haushalte (2 bzw. 3 Pers. u.m.)
Familienstand	Alleinstehend versus Verheiratet
(Erwerbs-)Status	Vollzeit-Erwerbstätig/Studierende versus TZ-/Nicht Erwerbstätig
BIK-Typ	Kerngebiet versus Randlage / Peripherie / ländlicher Raum
2. <i>Gebrechlichkeit / Krankheit</i>	
Alter	Befragte ab 70 Jahre versus sonstige Altersgruppen
3. <i>Misstrauen aufgrund von Viktimisierungssängsten</i>	
BIK-Typ	Kerngebiet versus Randlage/Peripherie/ländlicher Raum
Wohnstatus	Soz. Wohnungsbau versus Mieter/Eigentümer
4. <i>„Belastung“ bzw. kritische Distanz gegenüber Umfrageforschung (Misstrauen/Schutz der Privatsphäre)</i>	
Links/Rechts	Stark Links / Rechts versus Mitte/eher Links/eher Rechts
Inglehart-Index	Postmaterialisten versus Materialisten/Mischformen
Parteienpräferenz	Verweigerung / Sonstige versus „etablierte“ Parteien
5. <i>Fehlendes Interesse und / oder generelle soziale Distanz</i>	
Soziale Schicht	Unterschicht versus Mittel-/Oberschicht
Pers. wirtschaftliche Lage	Schlecht versus Durchschnitt/Gut
Höchster Schulabschluss	Hauptschule versus Abitur/Mittlere Reife
Interesse an Politik	Gering versus Mittel/Hoch

Infratest Sozialforschung, eigene Darstellung

5.2. Ergebnisse

Die „Extensität“ der Nachbearbeitung hat bei den ALLBUS-Erhebungen von Erhebungsjahr zu Erhebungsjahr zugenommen. Auch hier sticht wieder der ALLBUS 2000 mit einer nach Abschluss der Hauptbearbeitungsphase realisierten Ausschöpfung von nur 30,4 Prozent in den alten und 36,9 Prozent in den neuen Bundesländern, hervor.

Umso erstaunlicher ist es, dass sich in der Modellanalyse insbesondere für die alten Bundesländer wiederum nur sehr schwache Zusammenhänge zwischen den erwartbaren Effekten von systematischem Nonresponse und Realisierung der Interviews im Rahmen der Haupt- oder der Nachbearbeitung zeigen (vergleiche dazu Übersicht 8). Unabhängig von der Höhe der Ausschöpfung erweisen sich die Logit-Modelle für die Jahre 1994, 1996 und 2000 mit einem Pseudo-R² jeweils um die 0,03 als äußerst schwach und wenig trennscharf.

In den neuen Bundesländern ist die Güte der Modelle interessanterweise etwas größer. Bemerkenswerterweise wird in diesem Fall für das Erhebungsjahr 1994, also dem Jahr, in dem nicht mehr als 7 Prozent der Interviews durch die Nachbearbeitung dazu kamen und mit 52,2 Prozent bereits nach Abschluss der Hauptbearbeitung eine Ausschöpfung im Rahmen des damals Erwartbaren realisiert wurde, mit einem Pseudo- R^2 von 0,12 eine erstmals akzeptable Größe erreicht, die auf einen hinreichend starken systematischen Unterschied zwischen Haupt- und Nachbearbeitung hindeutet. Auch in den neuen Bundesländern hat die Bedeutung der Nachbearbeitung für die Qualität der Stichprobe seitdem kontinuierlich und umgekehrt proportional zum gestiegenen Aufwand abgenommen. Beim ALLBUS 2000 findet sich auch für die neuen Länder ein Pseudo- R^2 von nur noch 0,07.

Betrachtet man die Effekte im Einzelnen, so zeigt sich, dass durch die Nachbearbeitung über die Jahre hinweg vergleichbare Verbesserungen ausschließlich im Bereich der BIK-Regionen, sprich bezüglich der Untererfassung von Zielpersonen aus Kerngebieten, erreichbar waren. Dieser Effekt ist allerdings zum Teil auch als Artefakt der Anlage der Nachbearbeitung geschuldet, da es sich bei den Klumpen, die aufgrund mangelhafter Bearbeitung im Rahmen der Nachbearbeitung ausgetauscht und neu in die Stichprobe aufgenommen wurden, vorrangig um solche aus Kerngebieten gehandelt hat. Alle anderen signifikanten Effekte sind hingegen eher uneinheitlich und auf einzelne Erhebungsjahre begrenzt.

Besonders interessant ist, dass die bei der Nonresponse-Analyse festgestellten systematischen Abweichungen beim Alter – sprich die Untererfassung von älteren Zielpersonen ab 70 Jahren – durch die Nachbearbeitung nicht abgemildert wurden. Diesbezüglich hat sich am Ergebnis so gut wie nichts geändert. Ebenfalls nicht signifikant sind die Indikatoren Haushaltsgröße / Single-Haushalte bzw. Erwerbsstatus / Vollzeit-Erwerbstätig. Gemeinhin schlechter erreichbare Gruppen sind durch die Nachbearbeitung in der Stichprobe nicht besser repräsentiert worden. Die Variable Wohntyp / Sozialer Wohnungsbau als möglicher Indikator für Misstrauen und Viktimierungsängste ist über die Jahre hinweg ebenfalls nicht signifikant. Beim ALLBUS 1996 wurden in der Haupterhebung sogar geringfügig mehr Zielpersonen aus dem sozialen Wohnungsbau erreicht als in normalen Mietwohnungen. Auch die Variablen, die einen möglichen Nonresponse bei postmaterialistisch orientierten oder aber aus sonstigen Gründen gegenüber Befragungen misstrauischen Personen abbilden können, sind uneinheitlich. Der „Inglehart-Index“ weist beim ALLBUS 2000 im Rahmen der Nachbearbeitung auf eine geringfügig höhere Befragung von Materialisten im Vergleich zu Postmaterialisten hin. Stabil ist dieses Muster über die Jahre hinweg jedoch nicht gewesen. Die z. T. auch signifikanten Ausprägungen bei der Links-Rechts-Positionierung und bei den Ergebnissen der Sonntagsfrage deuten ebenfalls nicht auf eine Unterrepräsentierung von gegenüber dem Mainstream eher „extremere“ Positionen hin.

Übersicht 8: Unterschiede zwischen Haupt- und Nachbearbeitungsphase bei den ALLBUS-Untersuchungen 1994, 1996 und 2000

Logistische Regressionsanalyse: Nach- versus Hauptbearbeitung

	ALLBUS 2000		ALLBUS 1996		ALLBUS 1994	
	West	Ost	West	Ost	West	Ost
Fallzahl ¹⁾ :	2.036	1.102	2.402	1.116	2.342	1.108
Ausschöpfung insgesamt	46,9	53,7	54,2	54,2	53,2	55,2
Ausschöpfung Hauptbearbeitung	30,4	36,9	41,7	46,1	44,6	52,2
Fallzahlen (in Prozent)						
Hauptbearbeitung	63	66	72	78	79	93
Nachbearbeitung	37	34	28	22	21	7
Prädikatoren						
Haushaltsgröße						
Ref.-Kat: Single						
2 Personen	-,069	,205	-,071	,345	,062	,625
3 u.m. Personen	,088	,469	,225	,609	,395*	,406
Familienstand						
Alleinstehend versus						
Verheiratet	-,102	,144	-,221	,078	,172	-,183
Erwerbsstatus						
Ref.-Kat.: Vollzeit/Student						
Teilzeit	-,182	-,094	-,114	,426	,272	,536
Rentner/Pensionär	-,010	,259	-,218	,640	-,053	,047
Arbeitslos	,105	-,037	-,010	,313	,118	,406
Sonstiges	-,222	,177	,142	,075	-,102	-,178
BIK-Typ						
Ref.-Kat.: Kerngebiet						
Rand von Stadtregionen	,426**	1,056**	,123	-,637*	-,029	1,896
Peripherie	,447**	,804**	,422**	,047	,215	-,494
Ländlicher Raum	,693**	,134	,335**	,466*	,165	,966*
Alter						
Ref.-Kat.: 70 und älter						
18-29	,102	-,112	-,291	,273	,283	,151
30-39	-,153	-,080	-,491	,032	,069	,112
40-49	-,252	-,426	-,279	,309	,181	,135
50-59	,076	-,150	-,409	,048	,180	,158
60-69	-,149	,546*	-,053	,090	-,116	,403
Wohnstatus						
Ref.-Kat.: Sozialer Wohnungsbau						
Sonstige Mietwohnung	-,152	,138	-,386*	-,593**	-,004	,062
Eigentum	-,123	,377	-,166	-,036	-,087	,642
Schicht						
Ref.-Kat.: Untere Schichten						
Mittelschicht	,065	,233	,121	,298	,047	,191
Obere Schichten	,252	1,145*	,330	-,401	-,046	1,246

Weiß nicht	-,276	-,075	-,056	135	-,560*	-,1099*
Fortsetzung der Regressionsanalyse:						
	ALLBUS 2000		ALLBUS 1996		ALLBUS 1994	
	West	Ost	West	Ost	West	Ost
Persönliche wirtschaftliche Lage						
Ref.-Kat.: (Eher) Schlecht						
Teil/Teils	-,206	-,278	-,021	,640*	-,163	,484
(Eher) gut	-,313	-,268	-,236	,756**	-,249	,404
Höchster Schulabschluss						
Ref.-Kat.: Volks-/Hauptschule						
Mittlere Reife	-,060	-,333	-,223	,082	-,078	-,035
FH-Reife	-,126	-,233	-,075	,086	-,201	1,241
Abitur	-,110	-,041	-,219	,483	-,170	-,304
Sonstiges	,093	-,200	,238	,731	,548	-,893
Interesse an Politik						
Ref.-Kat.: Wenig						
Stark	,188	-,275	,264	,221	-,009	,499
Mittel	,039	-,267	,151	,374*	-,156	-,118
Links-Rechts Positionierung						
Ref.-Kat.: Extrem Links/Rechts						
Eher Links	,217	-,060	,066	-,551	-,484	,266
Mitte	,153	-,373	-,066	-,427	-,675**	,493
Eher Rechts	,246	-,522	,111	-,283	-,540*	,752
Keine Angabe	,315	-,200	-,897	-,1902*	-,432	,169
Inglehart-Index						
Ref.-Kat.: Postmaterialisten						
PM-Mischtyp	-,031	-,311	-,068	,659*	,196	,504
M-Mischtyp	-,043	-,237	-,036	,564*	-,016	,794
Materialisten	-,328*	-,572*	,169	,481	,278	,774
Sonntagsfrage						
Ref.-Kat.: Angabe verw./Sonst.						
CDU/CSU	,219	-,572*	-,206	-,669*	,063	,465
SPD	,281	-,433*	,014	-,178	,010	-,254
FDP	,311	-,299	-,441	-,662	,069	,650
Grüne	,464	-,177	-,177	-,851	-,120	,346
Würde nicht wählen	,608*	-,163	,108	-,452	,247	,489
Weiß nicht	,493*	,017	,063	-,396	-,113	,170
Modellkonstante	,183	,920	1,197*	-,028	1,646**	,255
Pseudo R² (Mc-Fadden)	0,028	0,068	0,032	0,091	0,029	0,124
1) Fallzahlen der ALLBUS-Erhebungen. Bedingt durch fehlende Angaben bei den einzelnen Merkmalen wurden die Modelle mit einer geringfügig kleineren Fallzahl berechnet						
Binär-logistische Regression (SPSS 11.0 - NomReg) : Nachbearbeitung (0) versus Hauptbearbeitung (1) B-Koeffizienten mit * = Sig.-Niveau < 0,05 bzw. mit ** = Sig.-Niveau < 0,01						

Zur Veranschaulichung möchten wir auch in diesem Fall wieder einzelne bivariate Verteilungen darstellen. Ausgewählt haben wir die Ergebnisse der Rechts-Links-Positionierung sowie den Inglehart-Index im Vergleich Hauptbearbeitung und Endergebnis.

Übersicht 9: Verteilung der Rechts-Links Einstufung und des Inglehart-Index nach der Hauptbearbeitungswelle und im Endergebnis beim ALLBUS 2000

Angaben in Prozent	West		Ost	
	Hauptbearbeitung	Endergebnis	Hauptbearbeitung	Endergebnis
Rechts-Links Einstufung				
Links (1-2)	4	4	10	9
Eher links (3-4)	24	25	32	30
Mitte (5-6)	40	40	42	44
Eher rechts (7-8)	18	17	7	8
Rechts (9-10)	2	2	1	1
Keine Angabe	12	12	8	8
Inglehart-Index				
Postmaterialisten (P)	28	28	17	16
PM-Mischtyp	28	27	33	34
M-Mischtyp	31	30	33	32
Materialisten (M)	12	13	16	17
Keine Angabe	1	2	1	1

1) Zur Indexbildung vergleiche Codebuch ALLBUS 2000 (ZA-Nr. 3451), Note 2: 494

Infratest Sozialforschung, eigene Berechnungen

Das Ergebnis nach der Hauptbearbeitung unterscheidet sich fast nicht vom Endergebnis.³³ In der Interpretation ergäbe sich kaum ein relevanter Unterschied, wenn man stattdessen nur auf die im Rahmen der Hauptbearbeitung realisierten Fälle zurückgreifen würde.³⁴ Die durchgeführte Analyse zeigt, dass die Abweichungen bei den drei ALLBUS-Untersuchungen von 1994, 1996 und 2000 nicht systematisch verteilt sind. Ein einheitli-

33 Zur relativ geringen Bedeutung von möglichen bivariaten Abweichungen vergleiche die hierzu bereits im letzten Abschnitt gemachten grundsätzlichen Anmerkungen.

34 Es lässt sich übrigens schwer abschätzen, welche Verteilung „wahrer“ ist. Interessanterweise ist die Abweichung im Osten etwas stärker ausgeprägt. Hier hatten wir zwar eine etwas höhere Ausschöpfung, aber trotzdem auch eine leicht stärkere Abweichungssystematik in der Verteilung des Nonresponses. Auch der Vergleich der Haupt- mit der Nachbearbeitung ergab eine leicht stärkere Abweichungssystematik als in den alten Ländern.

ches Muster ist – mit Ausnahme der Verbesserung der Ausschöpfung bei den BIK-Kerngebieten – nicht zu erkennen. Insgesamt gibt es bedingt durch die Nachbearbeitung Effekte. Art und Richtung der Effekte lassen sich jedoch weder hinreichend gut vorhersagen, noch in einen klaren Zusammenhang zu einer besseren Stichprobenqualität bringen. Durch die intensive Nachbearbeitung hat sich demnach vorrangig die Fallzahl, also die Quantität, und weniger die Qualität erhöht. Natürlich ergeben sich in den Strukturen der Haupt- und der Nachbearbeitung gewisse Abweichungen. Diese Unterschiede sind im Vergleich über die Erhebungsjahre jedoch kaum systematisch und für sich genommen relativ gering ausgeprägt. Bemerkenswerterweise sind die Abweichungen auch in diesem Fall in den neuen Ländern, trotz höherer Stichprobenausschöpfung, größer als in den alten Ländern.

6. Fazit und Diskussion

Die 1994, 1996 und 2000 durchgeführten, hart bearbeiteten und dokumentierten registerbasierten Verfahrensmodelle bei den ALLBUS-Erhebungen (best practice) gingen mit einem weiteren Rückgang in der Ausschöpfung – zuletzt beim ALLBUS 2000 auf 46,9 Prozent für die alten und 53,7 Prozent für die neuen Bundesländer – einher. In der vorliegenden Arbeit haben wir die möglichen Qualitätseffekte anhand von drei verschiedenen Zugängen untersucht.

Die im ersten Schritt vorgenommene **Nettovalidierung** der Stichproben am Mikrozensus als Referenzstatistik erbrachte keinerlei neue Hinweise auf mögliche gravierende Verzerrungen. Vielmehr ergab sich eine akzeptable Übereinstimmung der soziodemografischen Randverteilungen mit der amtlichen Statistik. Im Einzelnen erwiesen sich die relativ schlechter ausgeschöpften ALLBUS-Erhebungen in den alten Bundesländern – hierbei insbesondere der ALLBUS 2000 – im Vergleich sogar tendenziell als noch besser angepasst.

Auch die im zweiten Schritt durchgeführte **multivariate Nonresponse-Analyse** von verfügbaren Prädiktoren, die zu einem auch theoretisch begründeten „nonresponse bias“ führen könnten, hat die Hypothese nicht bestätigt, dass die sinkenden Ausschöpfungsraten beim ALLBUS mit einer zunehmenden systematischen Verzerrung einhergehen. Hierbei wurden zusätzlich, neben den von den Einwohnermeldeämtern im Rahmen der Stichprobenziehung mitgelieferten demografischen Informationen, externe und kommerziell vertriebene mikrogeografische Wohnumfeldindikatoren (MOSAIC-Variablen und -Milieudaten) herangezogen. Im Einzelnen ließ sich die für allgemeine Bevölkerungsumfragen mehr oder weniger übliche Unterrepräsentierung von Seniorinnen und Senioren (70 Jahre und älter) feststellen, die vor allem aufgrund der im höheren Lebensalter verstärkt auftretenden Probleme einer Hilfe- oder Pflegebedürftigkeit schwerer zu befragen

sind. Eine etwas stärkere Systematik in der Verteilung des Nonresponses ergab sich – interessanterweise vorrangig bei der relativ besser ausgeschöpften Stichprobe für die neuen Bundesländer – erst bei Einbeziehung weiterer interviewerbezogener Merkmale. Als signifikant erwies sich für die neuen Länder nun auch zusätzlich eine tendenzielle Untererfassung von Zielpersonen aus Kerngebieten von Stadtregionen bzw. aus größeren Wohnblocks oder Hochhäusern. Dieser Effekt deutet für die neuen Länder auf offenbar etwas stärker wirksame mögliche Viktimisierungsängste hin, die die Kooperationsbereitschaft von Zielpersonen bei face-to-face Befragungen im Haushalt einschränken können. Bei den interviewerbezogenen Faktoren erwiesen sich die Merkmale als signifikant, die mit dem „Habitus“ bzw. der Art des Kontaktierens zusammenhängen. Tendenziell erfolgreicher wurden die Points bearbeitet, für die Interviewer eingesetzt werden konnten, die sehr offensiv (zum Beispiel ohne persönliche Vorankündigung oder Terminabsprachen) agieren. Alles in allem ist der Einfluss auch dieser Faktoren auf die Gesamtverteilung des Nonresponses jedoch ebenfalls gering.

Auch die im dritten Schritt vorgenommene **Analyse der Effektivität** der für harte Verfahrensmodelle typischen **intensiven Nachbearbeitung** bestätigte den Gesamtbefund. Es ergaben sich im Vergleich der drei ALLBUS-Untersuchungen keinerlei wirklich aussagekräftige Hinweise darauf, dass die realisierten Nettostichproben trotz der hohen Nonresponse-Raten mit einer systematischen Verzerrung verbunden waren, die erst durch die Nachbearbeitung tendenziell ausgeglichen werden konnten. Im Gegenteil ergab die hierbei analog angewandte multivariate Analyse für die alten Länder so gut wie gar keine, und für die neuen Länder in der Tendenz nur schwache Hinweise auf eine systematisch und nicht zufällig andere Verteilung innerhalb der Haupt- und der Nachbearbeitung. Da die Nachbearbeitung weitestgehend auf den gleichen Fällen basierte, die bereits im Rahmen der Hauptbearbeitung kontaktiert wurden, ohne dabei erfolgreich befragt worden zu sein, ist dies genau das Gegenteil vom dem, was eigentlich intendiert war. Tatsächlich hat die zeit- und auch außerordentlich kostenintensive Nachbearbeitung vorrangig Quantität und weniger Qualität (im Sinne einer Reduzierung möglicher Verzerrungen) gebracht. Das Ergebnis ist eine etwas höhere Ausschöpfungsrate, die letztendlich nur zu einer etwas höheren Fallzahl führt. Letzteres ließe sich, bei deutlich geringerem Aufwand, auch dadurch erreichen, dass die Bruttostichprobe der Hauptbearbeitungswelle von vornherein etwas größer angelegt wird.

Zusammengefasst bestätigen die Analysen, dass die Ausschöpfung für sich genommen kein hinreichend geeigneter Indikator zur Beurteilung der Qualität einer Bevölkerungsumfrage ist. Die Ergebnisse zeigen vielmehr, dass auch „schlecht“ ausgeschöpfte Stichproben offenbar gute Qualität aufweisen können. Wir interpretieren dies als überzeugenden Beleg für die prinzipielle Leistungsfähigkeit von zufallsbasierten Stichprobenverfah-

ren. Letztendlich ist es diesem fundamentalen Prinzip geschuldet, die Auswahl der Zielpersonen streng zufällig und losgelöst von möglichen gezielten Suchstrategien der Interviewerinnen und Interviewer vorzunehmen, dass sich auch der Stichproben-Nonresponse mehr oder weniger zufällig verteilt und nicht zu relevanten systematischen Verzerrungen führt. Voraussetzung hierfür ist, dass die Stichprobe „kunstgerecht“ und professionell gezogen wurde und dass auch die Bearbeitung im Feld sorgfältig genug und mit der hierfür notwendigen Erfahrung vorgenommen wird. Die Ergebnisse unserer Analyse zu den Effekten der Nachbearbeitung zeigen allerdings, dass es hierbei nicht unbedingt darauf ankommt, den Aufwand in der Feldarbeit immer weiter zu erhöhen. Das „Geheimnis“ des Erfolgs liegt offenbar in der soliden Bearbeitung in der Breite, und weniger darin, in diversen Nachbearbeitungswellen immer wieder aufs Neue bestimmte schlecht ausgeschöpfte Sample-Points anzugehen.

In der Konsequenz bedeutet dies, dass im Rahmen der notwendigen Qualitätsdebatte die Bedeutung des Merkmals Stichprobenausschöpfung deutlich relativiert werden sollte. Es ist nicht nur so, dass es keinen Sinn macht, Ausschöpfungen zu vergleichen, ohne die dahinter stehende Art der Definition eines Stichprobenbruttos zu überprüfen. Auch grundsätzlich ist dieses Merkmal für sich genommen als Prädiktor zur Beurteilung der Ergebnisqualität von vergleichbar angelegten Zufallsstichproben wenig aussagekräftig. Aus unserer Sicht wäre es sinnvoller, stattdessen bereits bei der Anlage der Untersuchung mit zu klären, anhand welcher externer Kriterien die Stichprobe validiert werden kann.³⁵ Dies können zum Beispiel die soziodemografischen Verteilungen der amtlichen Statistik sein. Dies kann aber auch anhand von Referenzinstrumenten geschehen, die in entsprechender Form auch in anderen Erhebungen eingesetzt werden. Hilfreich wäre, wenn zumindest größere allgemeine Bevölkerungsumfragen einen bestimmten Apparat von vergleichbaren Indikatoren einsetzen würden, die einerseits in einem darstellbaren Zusammenhang zum Untersuchungsziel (sprich der „abhängigen Variable“) stehen und die andererseits als allgemeine Referenz für andere Erhebungen nutzbar wären. Damit wäre die Möglichkeit eines gezielten Monitorings von repräsentativ angelegten Bevölkerungsstichproben eröffnet. Nonresponse-Studien werden hingegen, aufgrund der damit verbundenen umfangreichen methodischen Voraussetzungen, auch in Zukunft wohl eher die Ausnahme bleiben. Das von uns eingeführte Modell einer (anonymisierten) Verknüpfung von Stichprobendaten und externen Informationen erscheint uns hierfür jedoch hilfreich und weiterführend zu sein.

35 Vergleiche dazu auch die Überlegungen von Christian von der Heyde (von der Heyde 1998).

Abschließend möchten wir noch einmal darauf verweisen, dass die Ergebnisse unserer Untersuchung auf der Re-Analyse der einwohnermeldeamtsbasierten ALLBUS-Erhebungen beruhen. Tatsächlich ist dieses best-practice Verfahrensmodell für die Breite der empirischen Sozialforschung nicht typisch und aufgrund des damit verbundenen Zeit- und Kostenaufwandes auch nicht als grundsätzlicher Standard vorzugeben. Im Bereich der persönlich-mündlichen Erhebungen sind als alternative Modelle verschiedene Formen von Random-Route Verfahren (ADM-Design) eingeführt, die ebenfalls auf einer Zufallsauswahl beruhen, jedoch aufgrund der Integration von Stichprobenbildung und Befragung deutlich forschungsökonomischer realisiert werden können.³⁶ Im Vergleich zum best-practice Modell sind Random-Route Verfahren pragmatischer, insbesondere was die Kontrollierbarkeit der Vorgaben sowie die Intensität, mit der ein Sample-Point ausgeschöpft wird, angeht. Andererseits unterscheiden sie sich jedoch fundamental durch ihren grundsätzlich zufallsbasierten Charakter von unregelmäßigen Verfahren oder auch von Quotenstichproben.³⁷ Unsere Hypothese hierzu lautet, dass auch diese Verfahren, sofern sie ein dokumentiertes Maß an Standards, zum Beispiel hinsichtlich der Zufallsverteilung der Startadressen für den Random-Walk, der Einhaltung der Begehungsvorschriften etc., aufweisen und von einem erfahrenen Interviewerstab bearbeitet werden, zu akzeptabler Qualität führen. Aus unserer Sicht macht es allerdings wenig Sinn, die „Güte“ wie bisher durch überdimensional hohe Ausschöpfungsraten darzustellen, die im Prinzip mit den Ergebnissen von härter angelegten Verfahren nicht mehr vergleichbar sind. Hier ist aus unserer Sicht mehr Mut zum Realismus gefragt.

Wir möchten uns deshalb dafür aussprechen, den faktischen „Pluralismus der Verfahrensmodelle“ auch in Zukunft beizubehalten. Dies sollte jedoch nicht als Plädoyer für einen möglichen „Methoden-laissez faire“ verstanden werden. Unsere Empfehlung geht in diesem Fall eindeutig dahin, am Zufallscharakter der Stichprobenbildung (probability sample) sowie an den hierbei erforderlichen Standards der Dokumentation festzuhalten

36 Typischerweise wird bei den heute gängigen Formen von Random-Route Erhebungen darauf verzichtet, die Adressen in einem Vorlauf erst dokumentieren zu lassen und diese danach den Interviewern extern vorzugeben (Adressenvorlauf). Die Integration von Stichprobenbildung und Befragung im Rahmen eines einheitlichen Random-Walks führt in diesem Fall zu spürbaren Kostenvorteilen, ohne dass der Charakter einer Zufallsauswahl dadurch grundsätzlich in Frage gestellt wäre.

37 Zu den Problemen von Quotenerhebungen vergleiche zum Beispiel die nach wie vor äußerst instruktive empirische Analyse von Hoag (1986): „Der Bekanntenkreis als Universum: Das Quoten-Verfahren der Shell-Studie“.

und auf Quoten-Erhebungen oder sonstige Verfahren nur in begründeten besonderen Fällen zurückzugreifen.³⁸

Korrespondenzadressen

Ulrich Schneekloth
Infratest Sozialforschung
Landsberger Straße 338
80687 München
email: Ulrich.Schneekloth@nfoeurope.com

Ingo Leven
Infratest Sozialforschung
Landsberger Straße 338
80687 München
email: Ingo.Leven@nfoeurope.com

Literatur

Andreß, H.J./Hagenaars, J.A./Kühnel, S., 1997: Analyse von Tabellen und kategorialen Daten. Berlin.

Deutsche Forschungsgemeinschaft (DFG), 1999: Qualitätskriterien der Umfrageforschung. Berlin.

Esser, H./Grohmann, H./Müller, W./Schäffer, K.-A., 1989: Mikrozensus im Wandel. Wiesbaden.

Groves, R.M., 1989: Survey Errors and Survey Costs. New York.

Hartmann, P.H., 1990: Wie repräsentativ sind Bevölkerungsumfragen. Ein Vergleich des ALLBUS und des Mikrozensus. ZUMA-Nachrichten 26: 7-30.

Hartmann, P.H./Schimpl-Neimanns, B., 1992: Sind Sozialstrukturanalysen mit Umfragedaten möglich? Analysen zur Repräsentativität einer Sozialforschungsumfrage. Kölner Zeitschrift für Soziologie und Sozialpsychologie 49: 315-340.

Heyde von der, Chr., 1998: Qualitätsmaße und Qualitätskosten. München.

38 Quoten-Erhebungen sind aus unserer Sicht bei allgemeinen Bevölkerungsumfragen grundsätzlich dann eine Alternative, wenn die Grundgesamtheit nur eine sehr kleine Teilgruppe der Bevölkerung umfasst, die bei zufallsbasierten Erhebungen nur mit sehr hohem Risiko und entsprechend hohen Kosten abgebildet werden können. Weitere Gründe stellen natürlich auch externe Rahmenbedingungen dar, zum Beispiel wenn nur sehr wenig Feldzeit für die Durchführung der Erhebung eingeräumt werden kann. Dies ändert jedoch nichts an dem grundsätzlichen Dilemma, dass bei einer quotenbasierten Erhebung auch bei optimaler Aussteuerung der vorgegebenen Zellen kein wissenschaftlich haltbarer Rückschluss zulässig ist, ob damit auch die nicht vorgegebenen Merkmalsverteilungen richtig abgebildet sind.

Hoag, W., 1986: Der Bekanntenkreis als Universum: Das Quotenverfahren der Shell-Studie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38: 123-132.

Koch, A./Gabler, S./Braun, M., 1994: Konzeption und Durchführung der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1994. ZUMA-Arbeitsbericht 1994/11. Mannheim.

Koch, A., 1997: Teilnahmeverhalten beim ALLBUS 1994. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 49: 98-122.

Koch, A., 1998: Wenn „mehr“ nicht gleichbedeutend mit „besser“ ist: Ausschöpfungsquoten und Stichprobenverzerrungen in allgemeinen Bevölkerungsumfragen. *ZUMA-Nachrichten* 42: 66-93.

Koch, A./Wasmer, M./Harkness, J./Scholz, E., 2001: Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 2000. Mannheim.

Koch, A., 2002: 20 Jahre Feldarbeit im ALLBUS: ein Blick in die Blackbox; in: *ZUMA-Nachrichten* 51: 9-37.

Microm GmbH und Sinus Sociovision GmbH, 2000: MOSAIC Informationen. Stand 2000. Neuss, Frühjahr 2000.

Rendtel, U., 1995: Lebenslagen im Wandel: Panelfälle und Panelrepräsentativität. Frankfurt a. M.

Schnell, R., 1991: Wer ist das Volk? Zur faktischen Grundgesamtheit bei allgemeinen Bevölkerungsumfragen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 1: 106-137.

Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen.

Schnell, R./Hill, P. B./Esser, E., 1999: Methoden der empirischen Sozialforschung – 6. völlig überarbeitete und erweiterte Auflage. München.

Schnell, R./Kreuter, F., 2000: Das DEFECT-Projekt: Sampling Error und Nonsampling-Errors in komplexen Bevölkerungsstudien. *ZUMA-Nachrichten* 47: 89-102.

Statisches Bundesamt, 1998: Bevölkerung und Erwerbstätigkeit. Fachserie 1 Reihe 4.1.2. Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen. Stuttgart.

Statisches Bundesamt, 2001a: Bevölkerung und Erwerbstätigkeit. Fachserie 1 Reihe 3. Haushalte und Familien. Stuttgart.

Statisches Bundesamt, 2001b: Bevölkerung und Erwerbstätigkeit. Fachserie 1 Reihe 4.1.2. Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen. Stuttgart.

Wasmer, M./Koch, A./Harkness, J./Gabler, S., 1997: Konzeption und Durchführung der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1996. ZUMA-Arbeitsbericht 1996/08. Mannheim.

Westerhoven, E. van, 1978: Coverage Non-Response: Does it pay? A study of refusers and absentees. *Journal of the Market Research Society*, 20, 4: 245-247.

Anhang

**Tabelle A1: Differenz zum Mikrozensus: Merkmal Altersklassen
(Abweichung in Prozentpunkten)**

	MZ 1994	ALLBUS 1994	MZ 1996	ALLBUS 1996	MZ 2000	ALLBUS 2000
West						
20 bis 29 Jahre	19,0	-0,1	17,4	0,8	14,7	-0,2
30 bis 39 Jahre	19,6	2,2	20,5	1,4	20,9	-0,8
40 bis 49 Jahre	16,3	0,6	17,2	0,6	18,1	1,7
50 bis 59 Jahre	18,5	0,0	17,6	0,4	16,2	-0,4
60 bis 69 Jahre	13,4	0,1	13,8	-0,1	15,6	0,8
≥ 70 Jahre	13,3	-2,8	13,6	-3,3	14,6	-1,2
Ost						
20 bis 29 Jahre	16,8	-1,6	15,4	-2,7	14,5	-1,4
30 bis 39 Jahre	20,7	-1,9	20,4	-0,2	19,0	1,5
40 bis 49 Jahre	16,6	1,8	17,8	1,3	19,5	0,3
50 bis 59 Jahre	19,9	0,8	18,6	2,0	16,1	-1,0
60 bis 69 Jahre	13,9	1,9	15,1	0,2	17,0	0,6
≥ 70 Jahre	12,2	-1,0	12,6	-0,4	13,9	0,1

Mikrozensus 1994, 1996 und 2000: Statistisches Bundesamt 2001a: 240

**Tabelle A2: Differenz zum Mikrozensus: Merkmal Geschlecht
(Abweichung in Prozentpunkten)**

	MZ 1994	ALLBUS 1994	MZ 1996	ALLBUS 1996	MZ 2000	ALLBUS 2000
West						
Männlich	47,8	3,8	47,9	2,2	48,0	0,3
Weiblich	52,2	-3,8	52,1	-2,2	52,0	-0,3
Ost						
Männlich	47,0	0,6	47,3	0,2	47,8	-0,1
Weiblich	53,0	-0,6	52,7	-0,2	52,2	0,1

Mikrozensus 1994, 1996 und 2000: Statistisches Bundesamt 2001a: 252-254

**Tabelle A3: Differenz zum Mikrozensus: Merkmal Familienstand
(Abweichung in Prozentpunkten)**

	MZ 1994	ALLBUS 1994	MZ 1996	ALLBUS 1996	MZ 2000	ALLBUS 2000
West						
Ledig	22,2	-0,5	22,1	-0,5	22,0	0,2
Verheiratet	62,8	1,7	62,9	2,8	62,7	0,4
Verwitwet/geschieden	15,0	-1,2	15,0	-2,3	15,3	-0,6
Ost						
Ledig	17,4	-3,1	18,9	-5,5	22,2	-0,3
Verheiratet	65,2	3,7	63,7	5,4	60,1	-0,6
Verwitwet/geschieden	17,4	-0,6	17,4	0,1	17,7	0,9

Mikrozensus 1994, 1996 und 2000: Statistisches Bundesamt 2001a: 240

**Tabelle A4: Differenz zum Mikrozensus: Merkmal Erwerbsstatus
(Abweichung in Prozentpunkten)**

	MZ 1993	ALLBUS 1994	MZ 1996	ALLBUS 1996	MZ 2000	ALLBUS 2000
West						
Erwerbstätige	52,2	4,4	54,1	1,7	54,9	-1,6
Erwerbslose			4,8	-1,2	4,1	-0,9
Nichterwerbspersonen	47,8	-4,4	41,1	-0,5	41,0	2,5
Ost						
Erwerbstätige	52,9	0,2	53,2	0,5	51,5	-1,0
Erwerbslose			11,5	0,2	11,6	1,0
Nichterwerbspersonen	47,1	-0,2	35,3	-0,7	36,9	0,0

Mikrozensus 2000: Statistisches Bundesamt 2001b: 160 (West) / 292 (Ost)

Mikrozensus 1996: Statistisches Bundesamt 1998: 139 (West) / 250 (Ost),

Mikrozensus 1993: Koch 1998: 76

**Tabelle A5: Differenz zum Mikrozensus: Merkmal Schulabschluss
(Abweichung in Prozentpunkten)**

	MZ 1993	ALLBUS 1994	MZ 1996	ALLBUS 1996	MZ 2000	ALLBUS 2000
West						
Haupt-/Volksschulabschluss	59,6	-5,7	58,6	-7,7	54,5	-4,1
Bis Realschulabschluss	23,5	4,2	21,8	3,3	23,5	0,5
Fachhoch-/Hochschulreife ¹⁾	16,9	1,5	19,6	4,4	22,0	3,6
Ost						
Bis Realschulabschluss	86,0	-4,3	84,8	-2,4	77,5	3,8
Fachhoch-/Hochschulreife ¹⁾	13,9	4,3	15,2	2,4	22,5	-3,8

1) Beim Mikrozensus 1993 ist die Fachhochschulreife in der Kategorie „Bis Realschulabschluss“ mit enthalten.

Mikrozensus 2000: Statistisches Bundesamt 2001b: 160 (West) / 292 (Ost)

Mikrozensus 1996: Statistisches Bundesamt 1998: 142 (West), 254 (Ost)

Mikrozensus 1993: Koch 1998: 76

Tabelle A6: Differenz zum Mikrozensus: Merkmal Haushaltsgröße (Abweichung in Prozentpunkten)

	MZ 1994	ALLBUS 1994	MZ 1996	ALLBUS 1996	MZ 2000	ALLBUS 2000
West						
Einpersonenhaushalt	20,3	-1,5	20,9	-3,9	21,2	-3,0
Mehrpersonenhaushalt	79,7	1,5	79,1	3,9	78,8	3,0
Ost						
Einpersonenhaushalt	17,0	-2,9	17,5	-3,0	19,6	0,1
Mehrpersonenhaushalt	83,0	2,9	82,5	3,0	80,4	-0,1

Mikrozensus 1994, 1996, 2000: Statistisches Bundesamt 2001a: 251

SINUS SOCIOVISION

4.1.4 Kurzcharakteristik der Sinus-Milieus 2000 (Westdeutschland)

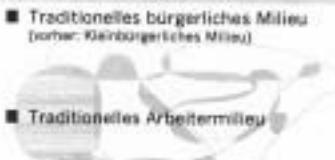
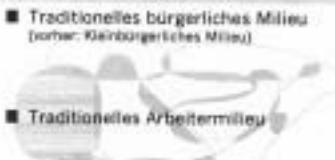
Gesellschaftliche Leitmilieus

 <p>■ Etabliertes Milieu (vorher: Konservativ-bürokratisches Milieu)</p>	10% >	Die erfolgsorientierte Konsum-Elite unserer Gesellschaft mit ausgeprägten Exklusivitätsansprüchen
 <p>■ Intellektuelles Milieu (vorher: Liberal-intellektuelles Milieu)</p>	10% >	Die aufgeklärte, postmateriell orientierte Werte-Avantgarde unserer Gesellschaft
 <p>■ Postmodernes Milieu</p>	6% >	Die extrem individualistische, „multi-optionale“ Life Style-Avantgarde unserer Gesellschaft

Moderner Mainstream

 <p>■ Adaptives Milieu (vorher: Modernes Arbeitermilieu)</p>	8% >	Der gut ausgebildete, mobile und pragmatische Mainstream der jungen modernen Mitte
 <p>■ Statusorientiertes Milieu (vorher: Aufstiegsorientiertes Milieu)</p>	18% >	Die beruflich und sozial aufstrebende untere Mitte – die Erfolgsinsignien unserer Konsumgesellschaft im Blick
 <p>■ Modernes bürgerliches Milieu</p>	8% >	Die konventionelle neue Mitte, die nach einem harmonischen, behüteten Leben in gesicherten Verhältnissen strebt

Traditioneller Mainstream

 <p>■ Traditionelles bürgerliches Milieu (vorher: Kleinbürgerliches Milieu)</p>	14% >	Die Sicherheits- und Status orientierte Kriegsgeneration, die an den traditionellen Werten wie Pflicht und Ordnung festhält
 <p>■ Traditionelles Arbeitermilieu</p>	4% >	Die an den Notwendigkeiten des Lebens ausgerichtete traditionelle Arbeiterkultur der Eckkneipen, Kleintierzüchter und Schützenvereine

Moderne Unterschicht

 <p>■ Konsum-materialistisches Milieu (vorher: Traditionelles Arbeitermilieu)</p>	11% >	Die stark hedonistisch geprägte Unterschicht, die Anschluss halten will an die Konsum-Standards der breiten Mitte
 <p>■ Hedonistisches Milieu</p>	12% >	Die unangepasste junge Unterschicht, die Spass haben will und sich den Konventionen und Verhaltenserwartungen der Leistungsgesellschaft verweigert

OFFLINE REKRUTIERTER ACCESS PANELS: KÖNIGSWEG DER ONLINE-FORSCHUNG?

THORSTEN FAAS

Der Beitrag beschäftigt sich mit der Frage der Repräsentativität von offline rekrutierten Access Panels. Nach idealtypischer Vorstellung sollten solche Panels für die Internet-Nutzer repräsentative Ergebnisse liefern. Verglichen wird eine Umfrage, die auf einem Access Panel basiert, mit einer persönlichen Repräsentativ-Befragung der deutschen Bevölkerung, die zeitgleich durchgeführt wurde und aus der für die Zwecke des Vergleichs nur die Internet-Nutzer ausgewählt wurden. In beiden Fällen sollte es sich um repräsentative Befragungen deutscher Internet-Nutzer handeln. Tatsächlich weichen die Ergebnisse der beiden Umfragen aber (teilweise deutlich) voneinander ab. Dies betrifft sowohl soziodemografische Variablen, aber auch Fragen der Internet-Nutzung sowie politische Einstellungen. Praktisch bedeuten die Ergebnisse, dass Online-Umfragen auf Basis von Access Panels *nicht* zwangsläufig Ergebnisse liefern, die Repräsentativität für alle Internet-Nutzer für sich beanspruchen können. Ihre Einsatzmöglichkeit in der wissenschaftlichen Praxis (und darüber hinaus) schränkt dies deutlich ein.

The article discusses whether online access panels whose members were pre-recruited in offline surveys can yield representative results. In an ideal world of sampling, this should be the case. Two surveys are compared to test this proposition. One was based on an online access panel, the other – conducted at the same time – was a representative survey of the German population. Only Internet users were taken from the second survey to make comparison possible. Both samples should arguably be representative for German Internet users. However, the data show that the two surveys differ considerably. Differences were found for sociodemographic variables, for questions on Internet usage and also for political attitudes. The results indicate that online surveys based on access panels do not necessarily yield representative results for all Internet users, a fact that considerably affects their power for academic analysis (and beyond).

1. Einleitung¹

Online-Umfragen² werden mit zahlreichen Vorschußlorbeeren bedacht: Zeit und Geld sollen sie sparen, zudem sind sie – rein technisch betrachtet – für nahezu jedermann realisierbar. Von einer „Demokratisierung der Umfrageforschung“ ist zu lesen. Dies kann allerdings nicht darüber hinwegtäuschen, dass mit Online-Umfragen auch zahlreiche Probleme verbunden sind. Insofern ist Couper (2000: 464) zuzustimmen, wenn er sagt: „Clearly, we stand at the threshold of a new era of survey research, but how this will play is not yet clear“.

Die größte Herausforderung stellt die Repräsentativität von Online-Befragungen dar: Bei aller technischen Einfachheit der Programmierung eines web-tauglichen Fragebogens erweist sich die konkrete Umsetzung einer *repräsentativen* Online-Umfrage als schwierig: Nach neuesten Ergebnissen liegt die aktuelle Internet-Penetration in Deutschland nur bei rund 50 Prozent, das heisst, dass bisher nur etwa die Hälfte der deutschen Bevölkerung Zugang zum World Wide Web hat ([N]Onliner-Atlas 2003), was für bevölkerungsrepräsentative Umfragen zwangsläufig zur Folge hat, dass es gravierende systematische Unterschiede zwischen Auswahl- und Zielgesamtheit gibt. Dies spiegeln auch die vom „Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute“ (ADM) und anderen herausgegebenen „Standards zur Qualitätssicherung für Online-Befragungen“ wider, in denen es konsequenterweise heißt: „Repräsentativität kann bei Online-Befragungen in der Regel *nur hinsichtlich der Internet-Nutzer insgesamt ...* erreicht werden“ (ADM et al. 2001; Hervorhebung durch den Verfasser). Doch selbst mit dieser Einschränkung sind die praktischen Probleme bei der Durchführung repräsentativer Online-Umfragen nicht behoben. Vielmehr stellt sich die Frage, in welcher Weise Internet-Nutzer ausgewählt werden können, damit tatsächlich für sie repräsentative Aussagen möglich werden (zur Problematik der Stichprobenziehung im Internet siehe unter anderem Couper 2000; Wildner/Conklin 2001; Kemmerzell/Heckel 2001; Hauptmanns/Lander 2003).³ In der Offline-Welt haben sich dazu für (persönliche) repräsentative Bevölkerungsumfragen zwei mögliche Wege der zufälligen Probandenauswahl etabliert: Einerseits macht man sich den Umstand zunutze, dass in Form der Einwohnermeldeamtsregister eine (nahezu) vollständige Liste möglicher Zielpersonen vorliegt und wählt (nach einer vorherigen Zufallsauswahl von

1 Mein Dank gilt Harald Schoen sowie den beiden anonymen Gutachtern für wertvolle Kommentare und Anregungen. Ebenso möchte ich Christian Holst und Uwe Reising für die mir zur Verfügung gestellten Informationen zum Access Panel von INRA Deutschland danken.

2 Eine gute Einführungsdarstellung zum Thema Online-Umfragen findet sich in Schonlau et al. (2002).

3 Auf Selbstselektion beruhende Internet-Umfragen bleiben hier ausgeblendet, siehe hierzu u.a. Bandilla/Bosnjak (2000), Faas (2003), Faas/Rattinger (2003), Hauptmanns (1999).

Einwohnermeldeämtern) zufällig Probanden aus diesen Registern aus. Andererseits wählt man zunächst zufällig Sample Points aus, ermittelt dann innerhalb dieser Punkte mittels vorab definierter Behebungsregeln Haushalte und findet schließlich innerhalb des Haushalts über einen entsprechenden Zufallsschlüssel die Zielperson.⁴ Analoges existiert in der Online-Welt nicht: Es gibt weder eine vollständige Liste aller Internet-Nutzer noch ein dem Random-Route-Ansatz vergleichbares Verfahren noch eine Möglichkeit, Email-Adressen zufällig zu generieren. Daraus folgern die Standards zur Qualitätssicherung: „Online-Befragungen, die für die Zielgruppe der Internet-Nutzer insgesamt Anspruch auf Repräsentativität erheben, sollten gegenwärtig auf der Grundlage einer vorherigen Offline-Auswahl beziehungsweise Offline-Rekrutierung mittels geeigneter Screening-Techniken durchgeführt werden (...)“ (ADM et al. 2001). Für die Durchführung von Online-Stichproben müssen also zunächst *offline* Zielpersonen erfasst werden. Konkret bedeutet dies, dass in persönlichen oder telefonischen bevölkerungsrepräsentativen Umfragen erhoben wird, ob bei den Befragten a) ein Computer mit Internet-Anschluss und b) die Bereitschaft, zukünftig an Online-Umfragen teilzunehmen, vorhanden ist. Ist beides der Fall, wird der Nutzer in ein sogenanntes Access Panel – also einen Pool von Befragungswilligen – aufgenommen, aus dem im Falle einer konkreten Online-Umfrage dann zufällig Personen ausgewählt und zur Teilnahme an einer Umfrage eingeladen werden.

Ein zusätzlicher Punkt, der dabei allerdings zu berücksichtigen ist (und der ebenfalls gegenüber traditionellen Offline-Umfragen die Durchführung von Online-Umfragen erheblich erschwert), ist die Dynamik der Grundgesamtheit. Im Gegensatz zur grundsätzlich stabilen Grundgesamtheit bei Offline-Umfragen befindet sich die Grundgesamtheit der Internet-Nutzer im Zustand permanenten Wandels, genauer gesagt: permanenter Ausdehnung. Nach den Ergebnissen der ARD/ZDF-Online-Studien etwa ist die Zahl der Internet-Nutzer in Deutschland von 6,6 Millionen im Jahre 1998 über 11,2 Millionen 1999, 18,3 Millionen 2000, 24,8 Millionen 2001 auf 28,3 Millionen im Jahr 2002 gestiegen (van Eimeren et al. 2002). Damit verändert sich zwangsläufig auch die (Sozial-) Struktur der Internet-Nutzer. Für den Aufbau und die Pflege von Access Panels folgt daraus zumindest die Notwendigkeit, kontinuierlich neue Teilnehmer zu rekrutieren. Ergänzend ist im Vorfeld der Stichprobenziehung durch entsprechende Schichtungsvor-

4 Bei Telefonumfragen findet sich eine Hybridform aus diesen beiden: Zunächst gewinnt man hier aus Datenbanken die Zahlenblöcke, in denen Telefonnummern liegen (was grob dem Melderegister entspricht), bildet anschließend alle theoretisch möglichen Telefonnummern in diesen Blöcken und wählt zufällig daraus Telefonnummern aus. Damit können auch nicht gelistete Nummern erreicht werden (sog. Gabler/Häder-Verfahren, siehe etwa Gabler/Häder 1997). Innerhalb eines kontaktierten Haushalte wird schließlich per Zufallsschlüssel eine Zielpersonen ausgewählt (was eher dem Random-Route-Ansatz entspricht).

gaben oder alternativ nach durchgeführter Umfrage durch entsprechende Gewichtungsmassnahmen sicherzustellen, dass eine gezogene Stichprobe zumindest hinsichtlich einiger sozialstruktureller Größen die (etwa aus Offline-Umfragen bekannten) aktuellen Referenzgrößen der Internet-Nutzer widerspiegelt.

Letztlich, so die Idealvorstellung, können dann repräsentative Aussagen für die Internet-Nutzer getroffen werden, da es sich um eine mehrstufige, ggf. geschichtete Zufallsauswahl handelt: Zunächst werden zufällig mit den etablierten Verfahren Zielpersonen für persönliche oder telefonische bevölkerungsrepräsentative Umfragen ausgewählt. Darunter sollte sich ebenfalls ein Zufallssample der Internet-Nutzer befinden, die wiederum eingeladen werden, Mitglied des Access Panels zu werden. Schließlich werden aus diesem Access Panel zufällig Personen zur Teilnahme an einer konkreten Umfrage ausgewählt. Inwieweit sich dieses idealtypische Verfahren allerdings in der Praxis bewährt, ist bisher kaum bekannt. Eine gewisse Skepsis jedenfalls scheint angebracht: Denn an jeder Stufe der mehrstufigen zufälligen Auswahl können auch systematische Fehler wirken, zu deren Quellen beispielsweise niedrige Ausschöpfungsquoten bei den zugrundeliegenden bevölkerungsrepräsentativen Umfragen, aber vor allem auch systematische Unterschiede in der Bereitschaft, sich für Online-Umfragen zur Verfügung zu stellen (und dann auch tatsächlich teilzunehmen), zählen.⁵ Schließlich kommt das für Panels typische Problem des Conditioning hinzu: Da die Mitglieder des Panels mehrfach an Befragungen teilnehmen, können sich leicht Lerneffekte oder bestimmte Antwortstrategien entwickeln.

Ob sich die idealtypische Repräsentativität von Access Panels also auch praktisch manifestiert, ist unklar. An dieser Stelle soll der vorliegende Beitrag ansetzen, indem Ergebnisse auf Basis eines Access Panels mit den auf die erfassten Internet-Nutzer beschränkten Ergebnissen einer parallel durchgeführten bevölkerungsrepräsentativen Umfrage verglichen werden. In beiden Fällen sollte es sich bei Gültigkeit der Verfahren um repräsentative Zufallsstichproben der Internet-Nutzer handeln. Abweichungen der Ergebnisse zwischen den beiden Stichproben sollten demnach lediglich zufallsbedingt auftreten. Ob dies tatsächlich der Fall ist, wird im Folgenden zu prüfen sein, wobei sowohl sozialstrukturelle Variablen als auch Fragen zur Internet-Nutzung sowie substanzielle Fragen der Politikwissenschaft für den Vergleich herangezogen werden.

5 Nach den Ergebnissen von Kemmerzell/Heckel (2001) waren etwa von den im Rahmen einer bevölkerungsrepräsentativen Umfrage identifizierten Internet-Nutzern nur 25 Prozent bereit, ihre Email-Adresse für zukünftige Online-Umfragen zur Verfügung zu stellen.

2. Datengrundlage: Zwei Umfragen anlässlich der Bundestagswahl 2002

Zwei anlässlich der Bundestagswahl 2002 durchgeführte Umfragen liegen der Untersuchung zugrunde, bei beiden sollte es sich im Idealfall um repräsentative Zufallsamples der deutschen Internet-Nutzer ab 16 Jahren handeln.⁶ Die erste Erhebung ist eine repräsentative, Internet-basierte Online-Umfrage unter 1.165 deutschen Internet-Nutzern. Diese Nutzer wurden zufällig aus dem (offline rekrutierten) Access Panel des Meinungsforschungsinstituts INRA ausgewählt und im Zeitraum vom 13. September bis zum 1. Oktober 2002 befragt.⁷ Die Ausschöpfungsquote lag bei 73,5 Prozent.⁸ Hier sollte es sich um eine repräsentative Stichprobe der deutschen Internet-Nutzer, die das Internet *privat* nutzen, ab 16 Jahren handeln, da es sich wie oben skizziert um eine mehrstufige Zufallsauswahl handelt. Das Verfahren entspricht damit den „Standards zur Qualitätssicherung für Online-Befragungen“ (ADM et al. 2001).⁹

Demgegenüber handelt es sich bei der zweiten Umfrage um eine repräsentative Bevölkerungsumfrage (mit der *gesamten* deutschen Bevölkerung ab 16 Jahren als Grundgesamtheit), in deren Rahmen im Zeitraum vom 12. August bis zum 8. November 2002 3.263 Personen persönlich befragt wurden.¹⁰ Diese Personen wurden zufällig – über Sample

6 Die Daten stammen aus dem DFG-geförderten Projekt „Politische Einstellungen, politische Partizipation und Wählerverhalten im vereinigten Deutschland: Ein Langfristvergleich von Determinanten und Konsequenzen“ unter Leitung von Prof. Dr. Jürgen W. Falter, Prof. Dr. Oscar W. Gabriel und Prof. Dr. Hans Rattinger, denen ich für die Bereitstellung der Daten herzlich danke.

7 Das verwendete Access Panel ist 1999 geschaffen worden, die Rekrutierung erfolgte seit dem kontinuierlich im Rahmen von Telefon- und persönlichen Zufallsstichproben. Die Erfolgsquote bei der Rekrutierung lag bei etwa 20 Prozent. Die Größe des Access Panels lag zum Zeitpunkt der Studie bei knapp 4000 Teilnehmern. Aus diesem Pool erfolgte die zufällige Auswahl geschichtet nach Geschlecht, Haushaltsgröße und Bildung, deren Verteilung in der Grundgesamtheit der Internet-Nutzer zuvor auf Grundlage der Ergebnisse aller im letzten Quartal durchgeführten bevölkerungsrepräsentativen Mehrthemenumfragen ermittelt worden war.

8 Vor und nach der Wahl wurden jeweils 793 Panel-Teilnehmer eingesetzt. Von den Vorwahl-Teilnehmern beteiligten sich 598 (Ausschöpfung 75,4 Prozent), von den Nachwahlteilnehmern 567 (Ausschöpfung: 71,5%).

9 Konkret erfolgt die Rekrutierung von Panelisten unter allen Teilnehmern von persönlichen und telefonischen bevölkerungsrepräsentativen Random-Studien des Instituts, die die Merkmale „Computer im eigenen Haushalt“ und „persönliche Internet-Nutzung“ aufweisen.

10 Die Feldzeit der repräsentativen Bevölkerungsumfrage ist damit länger als die Feldzeit der Online-Untersuchung. Diese unterschiedlichen Feldzeiten sind allerdings insofern für die Analyse unproblematisch, als sich die Sozialstruktur und das Antwortverhalten der Befragten der repräsentativen Bevölkerungsumfrage innerhalb der Feldzeit nicht (signifikant) unterscheiden. Vergleicht man innerhalb dieser Befragung die Befragten, die im exakt parallelen Zeitraum

Points, Random Route und *last-birthday*-Methode – ausgewählt.¹¹ Darunter befanden sich auch 1.076 Personen, die auf die Frage „Nutzen Sie persönlich das Internet“ mit „Ja“¹² geantwortet haben. Aufgrund der Tatsache, dass es sich bei der Stichprobe insgesamt um ein repräsentatives Sample der deutschen Bevölkerung ab 16 Jahren handelt, sollte es sich folglich bei diesen 1.076 Personen um ein zufälliges Sample der deutschen Internet-Nutzer ab 16 Jahren handeln. Um die Grundgesamtheit im Vergleich zur ersten Umfrage so weit wie möglich zur Deckung zu bringen, wurde aus dieser Stichprobe ein zweiter Vergleichsmaßstab abgeleitet: Hierzu wurden jene Internet-Nutzer, die das Internet ausschließlich am Arbeitsplatz nutzen, ausgeschlossen. Damit verbleiben noch 950 Internet-Nutzer, die das Internet *privat* nutzen. Im Folgenden wird die auf dem Access Panel basierende Stichprobe der Einfachheit halber als „Online-Befragung“ bezeichnet, die auf der bevölkerungsrepräsentativen, persönlichen Befragung beruhende dagegen kurz als „persönliche Befragung“, wobei hier nochmals zwischen der „erweiterten“ und der (um die das Internet ausschließlich beruflich Nutzenden) „verkürzten“ Version zu unterscheiden sein wird. Zusätzlich wurden die Stichproben an die zum Zeitpunkt der Feldarbeit aktuelle Sozialstruktur der Internet-Nutzer angepasst, wobei die Merkmale Haushaltsgröße, Geschlecht, Alter und (formale) Bildung herangezogen wurden.¹³ Es gilt nun, diese Stichproben der deutschen Internet-Nutzer zu vergleichen, sowohl im ungewichteten als auch im gewichteten Zustand. Der Vergleich umfasst drei Schritte: In einem ersten Schritt werden sozialstrukturelle Merkmale (Geschlecht, Alter, Bildung) verglichen, im zweiten Schritt dann Details zur Internet-Nutzung, im dritten Schritt schließlich substantielle Variablen der Politikwissenschaft.

zur Online-Befragung befragt wurden, mit jenen, die außerhalb dieses Zeitraums befragt wurden, so zeigen sich hinsichtlich Geschlecht, Alter, Bildung, aber auch hinsichtlich des Wahlverhaltens keine signifikanten Unterschiede.

11 Die Ausschöpfungsquote dieser Umfrage lag bei 63,8 Prozent.

12 Die Antwortoptionen waren insofern differenzierter, als die Befragten zwischen „Ja, nur zu Hause“, „Ja, nur am Arbeitsplatz“, „Ja, überwiegend zu Hause, aber auch am Arbeitsplatz“ sowie „Ja, überwiegend am Arbeitsplatz, aber auch zu Hause“ wählen konnten.

13 Die Struktur der Grundgesamtheit privater Internet-Nutzer in Deutschland ab 16 Jahren wurde, wie schon oben kurz skizziert, auf der Grundlage der sozialstrukturell gewichteten Ergebnisse aller im zurückliegenden Quartal von INRA Deutschland durchgeführten bevölkerungsrepräsentativen Mehrthemenumfragen ermittelt. Folgende Klasseneinteilung wurde bei der Gewichtung verwendet: Haushaltsgröße – 1, 2, 3, 4 und mehr Personen; Alter – 16 bis 24 Jahre, 25 bis 34 Jahre, 35 bis 44 Jahre, über 44 Jahre; Bildung – Volks-/Hauptschule, Real-/Fach-/Handelsschule; Oberschule/Gymnasium; Universität/Hoch-/Fachhochschule. Die Anpassung dieser Verteilungen erfolgte – getrennt für Männer und Frauen – univariat. Zusätzlich wurde die bewusste Überrepräsentation Ostdeutschlands in der persönlichen Befragung durch ein entsprechendes Ost-West-Gewicht (für die Gesamtbevölkerung) korrigiert.

Tabelle 1: Details zu den zwei im Umfeld der Bundestagswahl 2002 durchgeführten Umfragen

	Persönliche Befragung	Online-Befragung
Feldzeit	12. August – 8. November 2002	13. September – 1. Oktober 2002
Teilnehmer	3.263	1.165
Davon Internet-Nutzer	1.076	1.165
Davon <i>private</i> Internet-Nutzer	950	1.165
Rekrutierung	Zufällige Auswahl über Sample Points, Random Route und <i>last-birthday</i>	Zufällige Auswahl aus einem Access Panel
Befragungsmodus	PAPI*	CASI*
Ausschöpfung	63,8 Prozent	73,5 Prozent

* PAPI = Paper and Pencil Interview, CASI = Computer Assisted Self-Administered Interview

3. Sozialstrukturelle Variablen

Betrachtet man die Verteilungen sozialstruktureller Variablen in beiden Stichproben (siehe Tabelle 2), so zeigen sich zwar keine signifikanten Unterschiede¹⁴ hinsichtlich des Geschlechts – in der Internet-Stichprobe sind 58,8 Prozent der Befragten männlichen Geschlechts, in der persönlichen Befragung rund zwei Prozentpunkte weniger –, dies gilt aber nicht für die Alters- und Bildungsverteilungen: Die Befragten der Internet-Stichprobe sind mit einem mittleren Alter von 36,7 Jahren signifikant jünger als die persönlich befragten Internet-Nutzer, wobei insbesondere die jüngste Altersklasse mit 21,6 Prozent zu Lasten der Altersgruppe über 44 Jahren deutlich stärker besetzt ist.

Dies scheint auch die Verteilung der Bildungsabschlüsse zu beeinflussen: In der Online-Umfrage finden sich überdurchschnittlich viele Schüler. Alleine auf das Alter lassen sich die Bildungsunterschiede aber nicht zurückführen, da sich darüber hinaus zeigt, dass in der Online-Umfrage (selbst nach Kontrolle für die überdurchschnittlich hohe Zahl von Schülern) weniger Personen einen Volks-/Hauptschulabschluss haben. Insgesamt deuten sich damit bereits Muster an: Die Mitglieder des Access Panels sind jünger und höher gebildet. Bedenkt man, dass diese Gruppen überdurchschnittlich stark im Internet vertreten sind, so könnte dies ein Anzeichen dafür sein, dass im Access Panel vor allem „Heavy Users“ vertreten sind. Dies kann aber anhand weiterer Variablen genauer geprüft werden.

14 Die durchgeführten Signifikanztests erfassen dabei nicht mögliche Design-Effekte, dies ist bei der Interpretation zu berücksichtigen.

Tabelle 2: Sozialstrukturelle Zusammensetzung der beiden Stichproben, in Prozent (ungewichtete Daten)

	Online-Befragung	Persönliche Befragung (erweitert)	Persönliche Befragung (verkürzt)
Geschlecht			
Männer	58,8	56,3	56,7
Frauen	41,2	43,7	43,3
Summe	100,0	100,0	100,0
Cramers V		0,025 (n.s.)	0,021 (n.s.)
Alter			
16-24	21,6	16,3	16,2
25-34	23,3	24,2	24,9
35-44	28,8	26,7	26,8
45-59	20,3	23,3	22,2
ab 60	5,9	9,5	9,9
Summe	100,0	100,0	100,0
Cramers V		0,097 (***)	0,099 (***)
Mittleres Alter (in Jahren)	36,7	39,1	39,1
Bildung			
Noch Schüler	6,9	4,4	4,4
Volks-, Hauptschulabschluss	13,0	17,5	17,7
Mittlere Reife	36,1	33,6	34,0
Abitur	43,9	44,4	43,9
Summe	100,0	100,0	100,0
Cramers V		0,080 (**)	0,081 (**)
N	1165	1076	949

Bei der Stichprobenziehung der persönlichen Umfrage wurde Ostdeutschland überrepräsentiert. Dies wurde durch ein Ost-West-Gewicht korrigiert.

Die Signifikanztests vergleichen jeweils die Verteilungen aus der persönlichen Befragung mit der Verteilung aus der Online-Erhebung; n.s. = nicht signifikant; * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

4. Internet-Nutzung

Wie Tabelle 3 zeigt, ergeben sich in der Tat deutliche Unterschiede hinsichtlich des Nutzungsverhaltens des Internets zwischen den beiden Stichproben, die sich auch nicht dadurch ändern, dass man die persönliche Befragung auf die *privaten* Internet-Nutzer beschränkt. Diese Unterschiede erstrecken sich auf Ort, Häufigkeit und Beginn der Internet-Nutzung, wobei generell gilt, dass im Access Panel häufige, erfahrene Internet-Nutzer überproportional stark vertreten sind.

Die schwächsten (aber dennoch signifikanten) Unterschiede ergeben sich für den Ort der Internet-Nutzung: In der Online-Stichprobe finden sich mehr User, die sowohl privat als auch beruflich das Internet nutzen, während in der persönlichen Befragung die ausschließlich private Nutzung verbreiteter ist. Die Unterschiede zwischen den beiden Verfahren schwächen sich ab, wenn man aus der persönlichen Befragung jene Nutzer, die das Internet ausschließlich am Arbeitsplatz nutzen, ausschließt; völlig verschwinden die Unterschiede aber auch dann nicht, vielmehr bleibt die Tendenz der Verzerrung bestehen. Daran vermag auch die verwendete Gewichtung nichts zu ändern.

In ihrer Größenordnung deutlicher werden die Unterschiede, wenn man Häufigkeit und Beginn der Internet-Nutzung betrachtet. In der Online-Befragung nutzen rund 75 Prozent der Befragten das Internet (fast) jeden Tag, während es in der Vergleichsstichprobe nur rund 35 Prozent tun. Umgekehrt nutzen nur rund 4 Prozent der Teilnehmer der Online-Umfrage das Internet höchstens „ein paar Mal im Monat“, dagegen sind dies 15 Prozent in der persönlichen Befragung, wobei diese Unterschiede signifikant sind und weder von der gewählten Grundgesamtheit der persönlichen Befragung noch von der eingesetzten Gewichtung in bemerkenswertem Maße tangiert werden.

Diese systematischen Verzerrungen finden auch in der Frage, seit wann das Internet genutzt wird, ihren Niederschlag. Knapp die Hälfte der Mitglieder des Access Panels nutzt das Internet schon seit 1998 oder früher, verglichen mit weniger als 30 Prozent in der persönlichen Befragung. Umgekehrt nutzt nur rund ein Prozent der Access Panelisten das Internet erst seit 2002, während rund 8 Prozent der Befragten der persönlichen Befragung erst im Jahr 2002 zur Netzgemeinde gestoßen sind. Offensichtlich führt also die Unerfahrenheit bei neuen Internet-Nutzern systematisch dazu, sich zunächst *nicht* dazu bereit zu erklären, an Online-Umfragen teilzunehmen, was letztlich bedeutet, dass Access Panels immer das Manko eines Time-Lags anhaftet.

Tabelle 3: Ort, Häufigkeit und Beginn der Internet-Nutzung, in Prozent (ungewichtete und gewichtete Daten)

	Online-Befragung		Persönliche Befragung (erweitert)		Persönliche Befragung (verkürzt)	
	Ungewichtet	Gewichtet	Ungewichtet	Gewichtet	Ungewichtet	Gewichtet
Ort der Internet-Nutzung						
Ja, nur zu Hause	48,6	50,0	54,3	56,9	61,5	63,5
Ja, nur am Arbeitsplatz	0,9	1,2	11,7	10,4		
Ja, überwiegend zu Hause, aber auch am Arbeitsplatz	35,7	35,3	22,4	22,1	25,4	24,7
Ja, überwiegend am Arbeitsplatz, aber auch zu Hause	14,8	13,6	11,6	10,6	13,2	11,8
Summe	100,0	100,0	100,0	100,0	100,0	100,0
Cramers V			0,259 (***)	0,238 (***)	0,143 (***)	0,153 (***)
Häufigkeit der Internet-Nutzung						
Mehrmals täglich	39,1	38,3	18,4	17,2	19,1	17,6
Jeden Tag oder fast jeden Tag	36,1	35,8	28,3	28,0	27,5	27,5
Ein paar Mal in der Woche	21,0	21,7	38,0	38,3	38,6	38,9
Ein paar Mal im Monat	3,0	3,5	12,8	13,6	12,5	13,3
Seltener	0,8	0,8	2,6	2,9	2,3	2,6
Summe	100,0	100,0	100,0	100,0	100,0	100,0
Cramers V			0,319 (***)	0,322 (***)	0,315 (***)	0,319 (***)
Erste Internet-Nutzung						
2002	1,2	1,2	8,0	7,6	8,4	8,0
2001	7,8	7,1	18,0	20,5	17,6	19,9
2000	21,9	23,9	26,2	26,3	25,7	25,9
1999	21,3	21,4	18,6	18,3	18,8	18,6
1998 oder früher	47,9	46,5	29,2	27,4	29,5	27,6
Summe	100,0	100,0	100,0	100,0	100,0	100,0
Cramers V			0,259 (***)	0,288 (***)	0,269 (***)	0,289 (***)
N	1165	1160	1076	1076	949	949

Anmerkungen: siehe Tabelle 2.

Die bisher festgestellten Unterschiede zwischen den beiden Stichproben sind per se schon bemerkenswert und zeigen, dass Access Panels nicht zwangsläufig – wie es die ADM-Empfehlungen annehmen – „Repräsentativität ... bei Online-Befragungen ... hinsichtlich der Internet-Nutzer insgesamt“ (ADM et al. 2001) erreichen können. Bestimmte Gruppen innerhalb der Internet-Nutzer – nämlich neue, unerfahrene Nutzer – sind systematisch unterrepräsentiert. Wie relevant sind aber diese Unterschiede? Inwieweit schlagen sich diese Unterschiede auch in den substantziellen Ergebnissen nieder? Dies soll nun abschließend erörtert werden.

5. Substanzielle Variablen

Aus der Perspektive der Forschung ist natürlich die Frage nach substantziellen Unterschieden zwischen den beiden Stichproben zentral. Hängen die substantziellen Schlüsse, die aufgrund der Daten gezogen werden, von der gewählten Methodik ab? Wäre dem so, würde das erhebliche Zweifel an der Verwendbarkeit von auf Access Panels basierenden Daten hegen.

Betrachtet man zunächst ein einfaches, aber für die empirische Wahlforschung zentrales Element, nämlich die Verteilung der Wahlabsicht, so ergeben sich deutliche Unterschiede zwischen den Verteilungen in den beiden Samples und dem tatsächlichen Wahlergebnis vom 22. September 2002. Bevölkerungsrepräsentativ sind also beide Stichproben nicht (siehe hierzu auch Faas 2003; Faas/Rattinger 2003). Dies ist im hiesigen Zusammenhang aber nur von sekundärem Interesse, vielmehr geht es an dieser Stelle darum, Unterschiede *zwischen* den beiden Stichproben zu analysieren. Auch hier treten deutliche Unterschiede hervor (siehe Tabelle 4): Das Access Panel beinhaltet offensichtlich einen deutlichen Bias zu Lasten der Union: Sie erhält (ungewichtet) nur 24,0 Prozent der Zweitstimmen, während sie in der persönlichen Befragung 31,6 Prozent (von allen Nutzern) bzw. 32,4 Prozent (von den ausschließlich privaten Nutzern) erhält. Davon profitieren vor allem die SPD, die FDP sowie die sonstigen Parteien.¹⁵

15 Die Tatsache, dass die sonstigen Parteien in der CASI-Umfrage aufbauend auf dem Access Panel im Vergleich zur persönlichen Befragung deutlich besser abschneiden, ist auch mit einer anderen Erklärung als Sample-Unterschieden vereinbar: In der persönlichen Interview-Situation könnte die Hemmschwelle, eine nicht-etablierte („sonstige“) Partei zu nennen, aufgrund sozialer Erwünschtheit deutlich höher liegen als in der anonymen CASI-Befragung.

Tabelle 4: Verteilung der Wahlabsicht in beiden Stichproben, in Prozent (ungewichtete und gewichtete Daten)

	Online-Befragung		Persönliche Befragung (erweitert)		Persönliche Befragung (verkürzt)	
	Ungewichte f	Gewichte f	Ungewichte f	Gewichte f	Ungewichte f	Gewichte f
Wahlverhalten						
SPD	41,1	39,7	37,7	35,3	37,9	35,5
CDU/CSU	24,0	25,6	31,6	35,0	32,4	35,2
Grüne	14,1	14,3	14,6	14,2	14,0	13,9
FDP	11,4	11,0	9,0	8,9	8,9	8,9
PDS	4,7	4,7	5,3	5,0	5,2	4,9
Sonstige	4,7	4,7	1,8	1,7	1,6	1,5
Summe	100,0	100,0	100,0	100,0	100,0	100,0
Cramers V			0,117 (***)	0,127 (***)	0,123 (***)	0,130 (***)
N	1084	1089	835	824	744	732

Anmerkungen siehe Tabelle 2.

Die verwendeten Gewichte verändern die Stimmenanteile durchaus, ohne allerdings dazu zu führen, dass die Unterschiede zwischen den Stichproben geringer werden – im Gegenteil, Cramers V steigt sogar minimal an. Substanziell steigt der Anteil der Union durch die Gewichtung an, während der SPD-Anteil sinkt. Die Anteile der übrigen Parteien bleiben von der Gewichtung nahezu unberührt.

Um das Bild abschließend noch zu verbreitern, kann man sich den Umstand zunutze machen, dass in der Umfrage auch zahlreiche Items enthalten waren, die allesamt als Antwortoptionen Fünfer-Skalen vorsehen. Herangezogen werden dazu sechs Items zu Fragen der allgemeinen und individuellen Wirtschaftslage, fünfzehn Items zu Fragen der Zufriedenheit mit der Demokratie und ihren Akteuren, dreizehn Items zu Fragen des Institutionenvertrauens sowie schließlich sechzehn Items zu Fragen des politischen Interesses und Engagements (Details zu diesen Fragen finden sich im Anhang). Für jedes dieser insgesamt 50 Items wurden drei Dinge geprüft: Erstens, ob sich die Verteilungen

zwischen den beiden Stichproben¹⁶ insgesamt (gemessen mit einem Chi-Quadrat-Unabhängigkeitstest) signifikant unterscheiden; zweitens, ob sich die für beide Stichproben resultierenden Mittelwerte signifikant unterscheiden und schließlich drittens, ob der Mittelwert der auf dem Access Panel beruhenden Stichprobe größer als der Vergleichswert der persönlichen Befragung ist. Im Idealfall (keine Unterschiede zwischen den beiden Stichproben) sollten sich dabei keine signifikanten Unterschiede – weder hinsichtlich der Verteilung insgesamt noch hinsichtlich des Mittelwerts – ergeben; der Anteil der Mittelwerte, die im Access Panel größer sind als in der Vergleichsstichprobe, sollte bei ungefähr 50 Prozent liegen, in diesem Fall würden die Unterschiede nämlich sowohl nach oben als auch nach unten abweichen.

Tabelle 5: Verteilungs- und Mittelwertunterschiede bei 50 Einzelitems zwischen beiden Stichproben, in Prozent (ungewichtete und gewichtete Daten)

Thema	Item- zahl		Anteil signifikant verschiedener Verteilungen	Anteil signifikant verschiedene r Mittelwerte	Anteil mit größerm Mittelwert im Access Panel
Wirtschaftslage	6	Ungewichtet	100	0	50
		Gewichtet	83	17	50
Verdrossenheit/ Unzufriedenheit	15	Ungewichtet	87	53	73
		Gewichtet	87	53	87
Institutionenvertrauen	13	Ungewichtet	69	54	8
		Gewichtet	85	77	8
Politische Involvierung/ Interesse	16	Ungewichtet	94	38	56
		Gewichtet	88	44	56
Insgesamt	50	Ungewichtet	86	42	
		Gewichtet	86	52	

Hohe Werte bei den Wirtschaftsfragen signalisieren eine positive Bewertung der Wirtschaftslage, bei den Unzufriedenheitsfragen ein hohes Maß an Verdrossenheit, bei den Fragen nach dem Institutionenvertrauen in hohes Maß an Vertrauen und schließlich bei den Fragen nach der politischen Involvierung ein hohes Maß an Engagement. Entsprechend wurden alle Items vor der Analyse gleich gepolt.

Verglichen wird die auf dem Access Panel basierende Stichprobe mit der Teilstichprobe der *privaten* Internet-Nutzer aus der persönlichen Befragung.

16 Um eine möglichst hohe Konkordanz zwischen den beiden Stichproben herzustellen, wurden dabei aus der persönlichen Befragung ausschließlich die Befragten mit privater Nutzung herangezogen.

Die tatsächlichen Ergebnisse weichen von diesem Idealmuster allerdings deutlich ab (siehe Tabelle 5). Bei den insgesamt fünfzig Items unterscheiden sich in 86 Prozent der Fälle die Verteilungen signifikant, in 42 Prozent auch die Mittelwerte; nach Gewichtung steigt dieser Wert sogar auf 52 Prozent, während der erste unverändert hoch bei 86 Prozent bleibt. Betrachtet man nun die einzelnen thematischen Blöcke, so zeigt sich zunächst hinsichtlich der Fragen zur allgemeinen und individuellen Wirtschaftslage, dass sich nahezu alle Verteilungen signifikant unterscheiden, während sich die Mittelwerte praktisch nicht signifikant unterscheiden. Zudem gibt es auch keine systematischen Verzerrungen der Art, dass die Mittelwerte in einer Stichprobe *immer* höher sind als in der Vergleichsstichprobe. Dies gilt für ungewichtete und gewichtete Daten.

Auch im Falle der Unzufriedenheitsitems unterscheiden sich nahezu alle Verteilungen signifikant, zudem über die Hälfte der Mittelwerte, die – im Gegensatz zu den Fragen nach der Wirtschaftslage – auch deutlich in eine Richtung verzerrt sind: Die (mittlere) Unzufriedenheit ist gemessen auf Basis des Access Panels in 73 Prozent (ungewichtet) beziehungsweise sogar in 87 Prozent der Items (gewichtet) höher als in der Vergleichsstichprobe. Wir sehen also einen deutlichen Bias hin zur Unzufriedenheit in der Online-Stichprobe.

Ein ähnliches Bild ergibt sich für Items, die das Vertrauen in verschiedene Institutionen des politischen Systems messen. Auch hier treten zunächst mehrheitlich signifikante Unterschiede sowohl hinsichtlich der Verteilungen insgesamt als auch hinsichtlich der Mittelwerte zu Tage – im gewichteten Fall sogar noch stärker als im ungewichteten. Zudem haben diese Abweichungen auch eine eindeutige Richtung: Das Vertrauen in der Online-Umfrage ist fast durchweg niedriger. Etwas positiver ist das Bild, das sich bezüglich der politischen Involvierung ergibt. Zwar unterscheiden sich auch hier fast durchweg alle Verteilungen signifikant. Für die Mittelwerte gilt dies aber in geringerem Maße, zudem ist praktisch keine systematische Verzerrung der Mittelwerte in eine Richtung erkennbar.

6. Zusammenfassung und Ausblick

Sozialstruktur, Internet-Nutzungsverhalten ebenso wie substantielle Variablen der Politikwissenschaft unterscheiden sich signifikant zwischen den beiden hier verglichenen Stichproben. Offensichtlich handelt es sich also nicht in beiden Fällen um Stichproben, die Repräsentativität für *alle* Internet-Nutzer beanspruchen können.

Dabei liegt die Vermutung nahe, dass angesichts der Mehrstufigkeit des Auswahlverfahrens die Access Panel basierte Stichprobe diejenige ist, die weniger repräsentativ ist. Zu

vermuten ist dies dabei nicht aufgrund der Mehrstufigkeit per se, sondern vielmehr aufgrund der plausibel erscheinenden These, dass an jeder Stufe *systematische* Verzerrungen auftreten. Dies gilt besonders für die erste Stufe – die Bereitschaft, sich für das Access Panel zur Verfügung zu stellen. Die Betrachtung der Variablen bezüglich der Internet-Nutzung zeigte hier deutlich, dass die Mitglieder des Panels besonders erfahrene, häufige Nutzer sind. Unerfahrene (und seltenere) Nutzer werden offensichtlich systematisch benachteiligt, was – wie gezeigt wurde – auch die Ergebnisse der substanziellen Variablen beeinflusst. Diese Unterschiede können auch nicht durch (sozialstrukturelle) Gewichtungungen gemildert werden.

Praktisch bedeuten die hier präsentierten Ergebnisse, dass Online-Umfragen auf Basis von Access Panels *nicht* zwangsläufig Ergebnisse liefern, die Repräsentativität für alle Internet-Nutzer für sich beanspruchen können. Ihre Einsatzmöglichkeit in der wissenschaftlichen Praxis schränkt dies deutlich ein.

Korrespondenzadresse

*Thorsten Faas, M.Sc. (LSE)
Institut für Politikwissenschaft
Universität Duisburg-Essen – Campus Duisburg
Lotharstraße 65
D – 47048 Duisburg
email: thorsten.faas@uni-duisburg.de*

Literatur

ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V./ASI Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V./BVM Berufsverband Deutscher Markt- und Sozialforscher e.V./D.G.O.F. Deutsche Gesellschaft für Online-Forschung e.V., 2001: Standards zur Qualitätssicherung für Online-Befragungen. Frankfurt am Main: ADM.

Bandilla, W./Bosnjak, M., 2000: Online Surveys als Herausforderung für die Umfrageforschung: Chancen und Probleme. S. 9-28 in: P. Mohler/P. Lüttinger (Hrsg.), Querschnitt: Festschrift für Max Kaase. Mannheim: ZUMA.

Couper, M.P., 2000: Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly* 64: 464-494.

Faas, T., 2003: Umfragen im Umfeld der Bundestagswahl 2002: Offline und Online im Vergleich. *ZA-Informationen* 52: 120-135.

Faas, T./Rattinger, H., 2003: Drei Umfragen, ein Ergebnis? Ergebnisse von Offline- und Online-Umfragen anlässlich der Bundestagswahl 2002 im Vergleich. Konferenzpapier anlässlich der Jahrestagung des DVPW-Arbeitskreises „Wahlen und politische Einstellungen“ in Stuttgart, 5./6. Juni 2003.

Gabler, S./Häder, S., 1997: Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland. ZUMA-Nachrichten 41: 7-19.

Hauptmanns, P., 1999: Grenzen und Chancen von quantitativen Befragungen mit Hilfe des Internet. S. 21-37 in: B. Batinić/L. Gräf/A. Werner/W. Bandilla (Hrsg.), *Online Research*, Göttingen: Hogrefe.

Hauptmanns, P./Lander, B., 2003: Zur Problematik von Internet-Stichproben. S. 27-40 in: A. Theobald/M. Dreyer/T. Starsetzki (Hrsg.), *Online-Marktforschung: Theoretische Grundlagen und praktische Erfahrungen*, 2. Auflage, Wiesbaden: Gabler.

[N]Onliner-Atlas 2003. Eine Topographie des digitalen Grabens durch Deutschland. Gemeinschaftsprojekt der Initiative D21 und TNS Emnid, siehe auch www.initiated21.de.

Kemmerzell, P./Heckel, C., 2001: Grundgesamtheit und Stichprobe bei Online-Befragungen, repräsentativ zu was? *Planung&Analyse* 4: 52-60.

Schonlau, M./Fricker, R.D. Jr./Elliott, M.N., 2002: *Conducting Research Surveys via E-mail and the Web*. Santa Monica, CA: RAND.

Van Eimeren, B./Gerhard, H./Frees, B., 2002: ARD/ZDF-Online-Studien 2002: Entwicklung der Onlinenutzung in Deutschland: Mehr Routine, weniger Entdeckerfreude. *Media Perspektiven* 34: 346-362.

Wildner, R./Conklin, M., 2001: Stichprobenbildung für Marktforschung im Internet. *Planung&Analyse* 2: 18-27.

Anhang: Verwendete Items

Wirtschaftslage

Allgemeine Wirtschaftslage

- „Wie beurteilen Sie ganz allgemein die heutige wirtschaftliche Lage in den alten (neuen) Bundesländern? Sehr gut, gut, teils gut / teils schlecht, schlecht oder sehr schlecht?“
- „Ist die allgemeine wirtschaftliche Lage in den alten (neuen) Bundesländern in den letzten ein bis zwei Jahren: wesentlich besser geworden, etwas besser, gleich geblieben, etwas schlechter oder wesentlich schlechter geworden?“
- „Und wie wird es in einem Jahr sein? Erwarten Sie, dass die wirtschaftliche Lage in den alten (neuen) Bundesländern im allgemeinen dann: wesentlich besser, etwas besser, gleich geblieben, etwas schlechter oder wesentlich schlechter sein wird?“

Individuelle Wirtschaftslage

- „Wie beurteilen Sie zur Zeit Ihre eigene wirtschaftliche Lage? Würden Sie sagen, Ihre eigene wirtschaftliche Lage ist: sehr gut, gut, teils gut / teils schlecht, schlecht oder sehr schlecht?“
- „Ist Ihre eigene wirtschaftliche Lage in den letzten ein bis zwei Jahren: wesentlich besser geworden, etwas besser geworden, gleich geblieben, etwas schlechter geworden oder wesentlich schlechter geworden?“
- „Was glauben Sie, wie wird Ihre eigene wirtschaftliche Lage in einem Jahr sein? Erwarten Sie, dass Ihre eigene wirtschaftliche Lage dann wesentlich besser, etwas besser, gleich geblieben, etwas schlechter oder wesentlich schlechter sein wird?“

Politikverdrossenheit

Demokratiezufriedenheit

- „Wie zufrieden oder unzufrieden sind Sie – alles in allem – mit der Demokratie, so wie sie in Deutschland besteht?“

(Un-)Zufriedenheit mit Parteien

Ich lese Ihnen jetzt einige Aussagen über die Parteien in Deutschland vor. Bitte sagen Sie mir zu jeder Aussage anhand dieser Skala von - 2 bis + 2, ob sie Ihrer Meinung nach zutrifft oder nicht. Der Wert - 2 bedeutet, dass sie überhaupt nicht zutrifft, + 2 bedeutet, dass sie voll und ganz zutrifft. Mit den Werten dazwischen können Sie Ihre Meinung abstimmen.

- „Die Parteien wollen nur die Stimmen der Wähler, ihre Ansichten interessieren sie nicht.“
- „Die Parteien betrachten den Staat als Selbstbedienungsladen.“
- „Die meisten Parteipolitiker sind vertrauenswürdige und ehrliche Menschen.“
- „Ohne gute Beziehungen zu den Parteien kann der Bürger heute überhaupt nichts mehr erreichen.“
- „Auch einfachen Parteimitgliedern ist es möglich, ihre Vorstellungen in den Parteien einzubringen.“

- „Den Parteien geht es nur um die Macht.“
- „Die Parteien unterscheiden sich in ihren Zielen so sehr, dass der Bürger klare Alternativen hat.“
- „Die Parteien üben in der Gesellschaft zuviel Einfluss aus.“
- „Ohne Berufspolitiker in den Parteien würde unser Land schlechter regiert werden.“
- „Die meisten Parteien und Politiker sind korrupt.“

Systemresponsivität

Wir haben hier eine Reihe von häufig gehörten Meinungen über die Politik und die Gesellschaft zusammengestellt. Sagen Sie mir bitte, ob Sie diesen Meinungen zustimmen oder nicht. Der Wert - 2 bedeutet, dass Sie dieser Meinung überhaupt nicht zustimmen, der Wert + 2 bedeutet, dass Sie ihr voll und ganz zustimmen. Mit den Werten dazwischen können Sie ihre Meinung abstufen.

- „Politiker kümmern sich darum, was einfache Leute denken.“
- „Leute wie ich haben keinen Einfluss auf die Regierung.“
- „Die Bundestagsabgeordneten bemühen sich um einen engen Kontakt zur Bevölkerung.“
- „Die Bürger haben kaum Möglichkeiten, auf die Politik Einfluss zu nehmen.“

Institutionenvertrauen

Ich lese Ihnen jetzt eine Reihe von öffentlichen Einrichtungen vor. Sagen Sie mir bitte anhand dieser Liste bei jeder, ob Sie ihr vertrauen oder nicht. Nennen Sie jeweils den Skalenwert.

- „Wie ist das mit dem Bundestag?“
- „...dem Bundesverfassungsgericht?“
- „...der Bundesregierung?“
- „...den Gerichten?“
- „...der Polizei?“
- „...der Verwaltung?“
- „...den Kirchen?“
- „...den Parteien?“
- „...der Bundeswehr?“
- „...den Gewerkschaften?“
- „...den Wirtschafts- und Arbeitgeberverbänden?“
- „...den Umweltschutzgruppen?“
- „...den Abgeordneten des Deutschen Bundestages?“

Involvierung

Formen politischer Partizipation

Wenn Sie in einer für Sie wichtigen Sache politischen Einfluss nehmen und Ihren Standpunkt zur Geltung bringen wollten, welche der folgenden Dinge würden Sie dann tun? Bitte sagen Sie es mir anhand dieser Skala. -2 bedeutet, würde ich bestimmt nicht tun, +2 bedeutet, würde ich bestimmt tun. Mit den Werten dazwischen können Sie Ihre Meinung abstufen.

- „In einer Bürgerinitiative mitarbeiten.“
- „An einer genehmigten Demonstration teilnehmen.“
- „An einer Demonstration teilnehmen, auch wenn mit Gewalt gerechnet werden muss.“
- „Den Straßenverkehr blockieren.“
- „Versuchen, von einer Partei Unterstützung zu bekommen.“

Politisches Interesse, Wahlkampfinteresse und -aktivitäten

- „Wie stark interessieren Sie sich für Politik? Würden Sie sagen sehr stark, ziemlich stark, mittelmäßig, weniger stark, überhaupt nicht.“
- „Nun noch einmal zur Bundestagswahl 2002: Wie intensiv verfolgen Sie den Wahlkampf? Sehr intensiv, intensiv, teils/teils, weniger intensiv, gar nicht?“

Man kann sich ja auf verschiedene Weise mit dem Wahlkampf beschäftigen. Sagen Sie mir bitte, wie oft Sie Folgendes im laufenden Wahlkampf machen: sehr oft, oft, manchmal, selten oder nie?

- „Über die Wahl mit Freunden, Verwandten oder Arbeitskollegen sprechen.“
- „Wahlkampfveranstaltungen besuchen.“
- „Internetseiten über die Wahl oder den Wahlkampf ansehen.“

Politische Internet-Nutzung

Wir würden Sie gerne noch etwas genauer zur Nutzung des Internets befragen. Man kann das Internet ja für die verschiedensten Dinge nutzen, auch für verschiedene politische Dinge. Sagen Sie mir bitte anhand dieser Liste, ob Sie das Internet zu den folgenden Zwecken sehr oft, oft, manchmal, selten oder nie nutzen.

- „Ganz allgemein um politische Informationen im Netz zu suchen.“
- „Um Informationen auf Webseiten von politischen Einrichtungen, also beispielsweise von Parteien, Bürgerinitiativen, Behörden, Abgeordneten usw., zu suchen.“
- „Um politische Einrichtungen oder Personen online zu kontaktieren, zum Beispiel indem man eine E-mail an sie schickt.“

Interne Efficacy

Wir haben hier eine Reihe von häufig gehörten Meinungen über die Politik und die Gesellschaft zusammengestellt. Sagen Sie mir bitte, ob Sie diesen Meinungen zustimmen oder nicht. Der Wert - 2 bedeutet, dass Sie dieser Meinung überhaupt nicht zustimmen, der Wert + 2 bedeutet, dass Sie ihr voll und ganz zustimmen. Mit den Werten dazwischen können Sie ihre Meinung abstufen.

- „Die ganze Politik ist so kompliziert, dass jemand wie ich nicht versteht, was vorgeht.“
- „Ich traue mir zu, in einer Gruppe, die sich mit politischen Fragen befasst, eine aktive Rolle zu übernehmen.“
- „Wichtige politische Fragen kann ich gut verstehen und einschätzen.“

ITEM-NONRESPONSE IN MITARBEITERBEFRAGUNGEN

INGWER BORG & CHRISTIAN TREDER

An insgesamt neun großen Mitarbeiterbefragungen, die in drei verschiedenen Modi erhoben wurden, wird hier untersucht, ob die Nichtbeantwortung einzelner Items systematisch mit Einstellungen, so wie sich diese aus anderen Items schätzen lassen, zusammen hängt. Damit wird der verbreiteten Vermutung nachgegangen, dass Item-Nonresponses (ebenso wie Unit-Nonresponses) in Mitarbeiterbefragungen auf geringes Commitment, niedrige Zufriedenheit oder ähnliches hindeuten. Ebenso wird untersucht, ob Item-Nonresponses rein zufällig erfolgen. Die Ergebnisse bestätigen beide Hypothesen nicht. Es scheint eher so zu sein, dass der Hauptgrund für Item-Nonresponses darin liegt, dass die Befragten keine Meinung zum jeweiligen Befragungsgegenstand haben oder herleiten können oder sich ihrer Antwort nicht sicher genug sind.

It is often assumed in employee surveys that item nonresponse, just like unit nonresponse, is an indicator of low commitment, low satisfaction, or similar attitudes. Nine large employee surveys using three different modes of data collection are used to study whether item nonresponse behavior correlates systematically with attitudes measured by other items in the questionnaire. In addition, we investigate whether item nonresponse can be explained as simple errors. Our findings do not support either hypothesis. Instead, it seems that item nonresponse occurs primarily because the respondents do not have an opinion or cannot easily generate one on the subject of the item, or because they feel too uncertain about their answers.

1. Einleitung: Unit- und Item-Nonresponse

Sinkende Ausschöpfungsquoten werden heute in der Umfrageforschung als besondere Herausforderung betrachtet. Es gibt zahlreiche Vorschläge und Empfehlungen für Maßnahmen, die die Beteiligungsbereitschaft erhöhen (Dillman 2000; Porst 2001), aber darüber, ob und ggf. wie sich Nichtantworter von Antwortern unterscheiden, ist relativ wenig bekannt.

Rogelberg et al. (im Druck) berichtet dazu eine interessante Untersuchung. Von den insgesamt 80% Nichtantwortern einer Umfrage bei Studenten einer amerikanischen Universität gaben 83% in einer Vorbefragung an, dass sie sich an dieser Umfrage beteiligen werden oder dass sie dies jedenfalls nicht kategorisch ausschließen. Sie werden von Rogelberg et al. (im Druck) als „passive“ Nichtantworter bezeichnet. Nur die verbleibenden 17% der 80% Nichtantworter lehnten eine Beteiligung eindeutig („aktiv“) ab. Die Studie zeigt, dass sich nur die aktiven Nichtantworter von den Antwortern unterscheiden: Sie sind weniger gewissenhaft, weniger sozial verträglich, unzufriedener mit ihrer Universität und mit ihr emotional weniger verbunden als der Rest. Überträgt man diesen Befund auf Umfragen allgemein, dann kann man vermuten, dass dann, wenn die Mehrzahl der Nichtantworter passiv ist, eigentlich kein Grund besteht anzunehmen, dass sie sich systematisch von den Antwortern unterscheiden.

Dass die meisten Nichtantworter eher passiv sind oder jedenfalls keine besondere Beziehung zu den Inhalten des Fragebogens haben, zeigt sich auch im Vergleich von Personen aus früheren und späteren Wellen oder Abschnitten von Umfragen, die selten größere Unterschiede bei Meinungen und Einstellungen zeigen (Schwirian/Blaine 1966; Borg 2001).

Passives Nichtantworten spielt jedoch bei der Deutung kleiner Ausschöpfungsquoten in der Praxis selten eine Rolle. Das gilt insbesondere auch für Organisationsbefragungen, wo geringe Ausschöpfung in der Regel direkt als Evidenz für Unzufriedenheit gedeutet wird (z.B. Bauer 1990; Dalecki et al. 1993; Green 1991; Pearl/Fairley 1985)¹. Nachbefragungen (Borg 2003b) zeigen dagegen, dass es für eine Nichtteilnahme viele Gründe geben kann. Die wenigsten deuten auf aktiven Widerstand hin, sondern mehr auf Passivität („keine Zeit“, „vergessen“, „zu viel Arbeit“, „keine Lust“, „einer mehr oder weniger macht nichts aus“, „bin zu neu hier“). Dazu kommen noch allgemeine Dispositionen wie Persönlichkeit, Werte und Einstellungen und damit korrelierende demografische Merkmale wie Alter und soziale Schicht (Diekmann 1995). Relevant für die Rücklaufquote von Mitarbeiterbefragungen (MABs) sind zudem immer auch rein organisatorische und technische Gründe, insbesondere die gewählte Erhebungsmethode, die konkrete Durchführung der Erhebung, die Informationskampagne und das Commitment der Vorgesetzten zur Befragung und ihren Folgeaktionen.

Es liegt nahe, derartige Überlegungen zur „unit nonresponse“ auch auf das Nichtbeantworten einzelner Items zu übertragen, so weit diese nicht schlicht Fehler sind oder vom

1 Der umgekehrte Fall kommt gelegentlich auch vor. So interpretierte der Vorstandsvorsitzende eines Großunternehmens die geringe Beteiligung der Mitarbeiter an einer Mitarbeiterbefragung ohne Umschweife als Beleg dafür, dass die Mitarbeiter sehr zufrieden sind: „Sie hatten keinen Anlass, sich an der MAB zu beteiligen, weil die Dinge so, wie sie sind, in Ordnung sind“ (Borg 2003b).

Forscher absichtlich geplant sind durch das Setzen entsprechender Filter, die Verwendung von „Weiß nicht“- oder „Trifft nicht zu“-Kategorien oder Ähnlichem. Bei genauerer Betrachtung finde man u.a.,

- dass Items, die früher im Fragebogen platziert sind, eher beantwortet werden (Dickinson/Kirzner 1985; Guadagnoli/Cleary 1992);
- dass INRs (=Item-Nonresponses) vermehrt nach Filterverzweigung auftreten (Messmer/Seymour 1982);
- dass INRs bei kognitiv schwierigeren Items (z.B. Rankings langer Listen, Erinnern früherer Ereignisse) häufiger sind (Jones/Lang 1983; Tourangeau/Smith 1996);
- dass INRs bei vermutlich eher heiklen Themen wie Fragen zu Sex, sozialer Vereinigung, Depression oder Gesundheit wahrscheinlicher sind (Tourangeau/Smith 1996; Colsher/Wallace 1989);
- dass INRs eher dann auftreten, wenn die wahre Meinung „negativ“ ist und dem Befragten dies sozial unerwünscht erscheint (Johanson et al. 1993);
- dass INRs aus Mangel an Interesse an der Frage selbst entstehen (Coombs/Coombs 1976);
- dass der Modus der Datenerhebung sich auf die INRs auswirkt: Face-to-face Interviews haben weniger INRs als Selbstausfüller (DeLeeuw et al. 1989), wobei computerunterstützte Selbstausfüller einen Vorteil haben gegenüber traditionellen Selbstausfüllern (Tourangeau/Smith 1996).

Sogar das Antwortformat kann einen Effekt haben. So führen 5-stufige Antwortskalen zu weniger INRs als 11-stufige (Leigh/Martin 1987). Allgemeine Dispositionen wie z.B. die Persönlichkeitseigenschaften Gewissenhaftigkeit und soziale Verträglichkeit können diese Variablen noch moderieren (Omura 1983). Hinzu kommen noch demografische Merkmale wie vor allem Bildung und Alter (Guadagnoli/Cleary 1992; Colsher/Wallace 1989; Craig/MacCann 1978), die recht offensichtliche Zusammenhänge mit diesen psychologischen Variablen haben.

Diese Gründe kann man sortieren in *technische* (Equipment-Versagen, Interviewerfehler, Out-of-Range-Antworten, Übersehen, Überblättern u.ä., aber auch einfache Designeffekte; Hoox/DeLeeuw 1998), *kognitive* („kann nicht“) und *motivationale* („will nicht“), wobei diese noch interagieren können (Leigh/Martin 1987). In der Praxis von Mitarbeiterbefragungen wird i.d.R. nur der letztere unterstellt: „Der Befragte will nicht antworten“. Tatsächlich zeigt Borg (1991), dass Personen, die auf eines oder mehrere *demografische*

Items keine Antwort geben, signifikant unzufriedener sind und sich allgemein negativer äußern als Personen, die die demografischen Items ausfüllen. Die Erklärung dafür ist, dass sie sich durch das Nichtantworten auf diese Fragen nicht identifizierbar machen wollen, um so das Risiko, entdeckt zu werden, auszuschalten oder zu verringern. Die Unterschiede sind signifikant, aber klein und es bleibt daher die Frage, ob nicht die meisten INRs eher aus einfacheren Gründen, d.h. insbesondere deshalb, weil eine Antwort unter den Bedingungen knapper Zeit und relativ geringer Motivation zu schwer fällt, entstehen. Umgekehrt könnte man auch fragen, ob Personen, die sich z.B. als allgemein unzufriedener bezeichnen, vermehrt INRs aufweisen oder ob Personen, die ein relativ geringes Commitment zur Organisation haben, mehr Items unbeantwortet lassen, weil sie weniger involviert sind und weniger bereit sind, freiwillig Denkarbeit zu leisten. Ein anderes Motiv für mehr INRs könnte die Einstellung zur Mitarbeiterbefragung selbst sein: Ist sie weniger positiv, könnten die Befragten eher dazu neigen, Items auszulassen, weil die Umfrage ihrer Meinung nach ohnehin keine positive Wirkung verspricht.

Im Folgenden werden wir diesen Fragen in neun großen MABs nachgehen. Dabei interessiert vor allem, ob INRs nicht ähnlich wie bei den Unit-Nonresponses eher „profane“ Gründe haben als die in der Interpretation so gerne herangezogenen Motive. Kann man also INRs in MABs vor allem als Folge von Nicht-Wissen oder Unsicherheit bei der Abgabe einer Antwort deuten und nicht so sehr als absichtliches Unterdrücken einer Meinung?

2. Methode

2.1 Datenbasis

Basis der folgenden Untersuchungen sind die Daten aus neun Mitarbeiterbefragungen (MABs), die, bis auf eine Ausnahme, alle als Vollbefragungen durchgeführt wurden. Die MABs wurden in drei Unternehmen durchgeführt.

- Unternehmen A ist ein weltweiter Marktführer in der IT-Branche. Seine Mitarbeiter sind überwiegend hochqualifizierte Akademiker, die in einer sehr flachen, projektbezogenen Organisation arbeiten. Befragt wurden in 1998 insgesamt ca. 14.000, in 2000 ca. 22.000 und in 2002 ca. 29.000 Mitarbeiter.
- Unternehmen B ist ein Produktionsbetrieb der Automobilindustrie. Etwa 80% seiner Mitarbeiter sind Arbeiter, die im Dreischichtbetrieb arbeiten. Im Unternehmen B wurden insgesamt vier MABs in den Jahren 1996 bis 1999 durchgeführt. Dabei wurden zwischen 28.500 und 29.500 Mitarbeitern befragt.

- Unternehmen C ist ein Großkonzern der Logistik-Branche, der stark differenziert ist in unterschiedliche Geschäftsbereiche mit den verschiedensten Hierarchiestufen, Laufbahnen, Beschäftigungsverhältnissen und Arbeitszeitmodellen. MABs wurden hier in 1995 bei 14.500 Personen (einzige echte Stichprobe) und in 1998 bei 222.000 Personen durchgeführt.

2.2 Methode

Alle Befragungen wurden als Selbstausfüller entweder im Papier-und-Bleistift- oder im Online-Modus durchgeführt. Die durchschnittliche Bearbeitungszeit betrug in allen Fällen zwischen 25 und 40 Minuten.

Alle Befragungen waren methodisch sehr ähnlich aufgebaut. So waren alle Items geschlossen formuliert mit einer fünfstufigen Antwortskala im Likert-Format mit den Kategorien „stimme voll zu“, „stimme zu“, „teils/teils“, „stimme nicht zu“ und „stimme überhaupt nicht zu“. Eine „Weiß nicht“-Kategorie wurde nicht angeboten, weil die Items so konstruiert waren, dass sie von allen Mitarbeitern – wenn auch möglicherweise nur mit eingeschränkter subjektiver Urteilssicherheit – beantwortet werden konnten. Die Befragten wurden im Instruktionsteil der Fragebögen wie folgt dazu ermutigt, alle Items zu beantworten: „In diesem Fragebogen geht es um Ihre Meinung. Also darum, wie Sie persönlich die Dinge sehen. Ihre Antworten sind dann ‚richtig‘, wenn sie Ihre Meinung richtig ausdrücken. Lesen Sie jede Frage sorgfältig durch und antworten Sie dann zügig. Ihr erster Eindruck ist meist auch der treffende. Lassen Sie nach Möglichkeit keine Frage aus, es sei denn, Sie haben zu einer Frage wirklich keine Meinung. Wie Vortests gezeigt haben, ist dieser Fall für die meisten Mitarbeiter der ABC AG aber sehr unwahrscheinlich.“

Inhaltlich adressierten die Befragungen die typischen Themengebiete der Arbeitszufriedenheitsforschung (Neuberger 1974; Spector 1998), also Wahrnehmungen und Einschätzungen der Arbeitsplatzbedingungen, der Ziele und Aufgaben, der Entwicklungsmöglichkeiten, der Vorgesetzten, der Bezahlung und der Sozialleistungen, der Kollegen, der Zusammenarbeit untereinander, der Information und der Kommunikation im Unternehmen sowie dem Commitment zum Unternehmen. Beispiele für Items hierzu sind: „Alles in allem bin ich mit meinem Vorgesetzten zufrieden“, „Meine Arbeitstätigkeit macht mir Spaß“ und „Die Zusammenarbeit untereinander in unserer Arbeitsgruppe ist gut“.

Zusätzlich wurden verschiedene Items verwendet zu Themen wie der Kundenorientierung, der Qualität der Produkte oder der Strategie des Unternehmens. Beispiele hierfür sind die Items „Ich kenne die Bedürfnisse unserer Kunden“, „Bei gewissen Veränderungen der Bedingungen meines Arbeitsumfelds könnte ich viel produktiver sein“ und „Ich

kenne die Strategie des Unternehmens so gut, dass ich sie einem neuen Kollegen erklären könnte“.

Die Items wurden in den Fragebögen in Blöcke thematischer Verwandtschaft gruppiert, wie dies im MAB-Kontext üblich ist (Borg 2003b). Diese Themenblöcke waren in den Firmen A und C die üblichen psychologischen Kategorien wie Arbeitsplatzbedingungen, Tätigkeit, Vorgesetzter usw., d.h. Kategorien, die den empirisch-statistischen Faktoren der Arbeitszufriedenheit und somit den Denkkategorien des Individuums entsprechen. Im Unternehmen B wurden die Items sortiert in Blöcke, die den Hauptgesichtspunkten eines von der European Foundation for Quality Management inspirierten Unternehmensleitbilds mit Kategorien wie „Offene Kommunikation“, „Produktivität“ und „Kosten“ entsprachen (Zimmermann/Jordan 2003).

Insgesamt lag die Zahl der Items, die allen Mitarbeitern präsentiert wurde, zwischen 56 und 104 (Tabelle 1). Items, die sich nur an spezielle Teilgruppen (z.B. Führungskräfte) richteten, werden hier nicht berücksichtigt.

Hinzu kamen jeweils noch einige demografische Items. In der Online-Befragung im Unternehmen A wurden diese Items „vom Computer“ ausgefüllt. Sie konnten vom Befragten nicht verändert werden. Im Unternehmen B und in der MAB C1998 wurden den Befragten im Rahmen ihres Fragebogenpakets selbstklebende Etiketten mit einigen wichtigen demografischen Daten zugestellt. Diese Informationen waren zuvor ebenfalls aus dem Personalinformationssystem gezogen worden. Die Befragten wurden gebeten, diesen Aufkleber in ein entsprechendes Feld des Papierfragebogens einzukleben („Etikettenmethode“, siehe Borg 2003b). Dies wurde von praktisch allen Personen, die sich an den Befragungen beteiligten, auch realisiert. In der Stichprobenbefragung C1995 wurden die demografischen Items als Selbstaufüller erhoben. Vorgegeben wurde hierfür ein relativ grobes Schema von Kategorien zum Ankreuzen (Hierarchieebene, Name der Organisationseinheit). Die Item-Nonresponses lagen dabei unter 4%.

3. Ergebnisse

3.1 Das Ausmaß von Item-Nonresponse

Tabelle 1 zeigt die Beteiligungsquoten aller Mitarbeiterbefragungen. Sie liegen zwischen 50% und 96%, mit dem üblichen Moduseffekt einer sehr hohen Beteiligung bei der Wahllokalermittlung (einer besonderen Form der „Gruppenerhebung“), mittlerer Beteiligung bei einer Online-MAB und relativ geringer Beteiligung bei einer postalischen Befragung.

Tabelle 1: Beteiligungsquoten und Item-Nonresponseanteile in neun MABs

Befragung	Beteiligung (%)	Modus	Anzahl Items	Anzahl Einst. Items	INR gesamt (%)	INR Einstellung (%)	Ausgel. Items gesamt ¹ (%)	Ausgel. Einst. Items ² (%)	Ausgel. Items gesamt ¹ FK (%)	Ausgel. Einst. Items ² FK (%)
A1998	89	Online	93	34	70	27	11	13	8	10
A2000	80	Online	94	34	57	30	7	12	5	9
A2002	79	Online	89	32	49	19	5	11	4	9
B1996	94	Gruppe	104	50	35	20	11	17	11	22
B1997	96	Gruppe	77	35	32	17	14	21	5	9
B1998	96	Gruppe	80	35	32	19	15	23	6	8
B1999	95	Gruppe	77	38	32	20	15	23	5	9
C1995	61	Post	110	47	78	14	3	11	2	8
C1998	50	Post	56	31	17	11	8	11	7	9

Anmerkung: ¹Die Auswertung bezieht sich auf die Gruppe der Item-Nonresponder bzw. der Führungskräfte unter den Item-Nonrespondern. ²Die Auswertung bezieht sich auf Auslassungen bei Einstellungsitems zu Facetten der Arbeitszufriedenheit innerhalb der Gruppe der Item-Nonresponder bzw. der Führungskräfte in dieser Gruppe.

Tabelle 1 zeigt weiter einige summarische Informationen über den Umfang von Item-Nonresponses. Der Anteil der Personen, die *mindestens ein inhaltliches Item unbeantwortet lassen* (im Folgenden als Item-Nonresponder oder INR bezeichnet) ist sehr unterschiedlich in den verschiedenen MABs. Er liegt zwischen 17% und 78%. Einen Hinweis auf den Grund für diese Unterschiede erkennt man, wenn man sich nur auf die Items beschränkt, die Einstellungen messen: Für Einstellungsitems fällt der Anteil der INR beträchtlich auf Werte zwischen 11% und 27%. Ausgelassen werden also eher Items, in denen es nicht darum geht, wie positiv oder wie negativ der Befragte selbst zum Gegenstand der Frage steht, sondern solche Items, wo es um Wissen, um Fakten und um (emotional) „kalte“ Urteile geht.

Tabelle 1 zeigt weiter, dass Personen, die Items auslassen, trotzdem immer noch ca. 80%-90% aller Items beantworten. Falls eine Person jedoch ein Item auslässt, das sich auf eine Facette der Arbeitszufriedenheit bezieht, neigt sie dazu, dann auch gleich mehr Items auszulassen als dies bei Nichteinstellungsitems der Fall ist. Das ist nicht überraschend, weil die Einstellungsitems im Kontext der Arbeitszufriedenheit i.d.R. deutlich positiv

untereinander korreliert sind (Borg 2003a) und sich zudem noch in wenige Cluster/Faktoren gruppieren, während andere Items eher isolierte Einzelinhalte adressieren.

Schließlich zeigt Tabelle 1, dass Führungskräfte (FK), die ein oder mehr Items auslassen, i. allg. deutlich weniger Items auslassen als Nichtführungskräfte.

3.2 Item-Nonresponse bei verschiedenen Inhalten

Sind die Item-Nonresponses zufällig über Befragungsthemen verteilt? Abbildung 1 gibt hierauf eine Antwort. Gezeigt sind für die Befragungen A1998, B1999 und C1998 die Auslassungen bei vergleichbaren Einstellungsitems, die hier den typischen Kategorien der Arbeitszufriedenheit zugeordnet sind. Die Streuungsmarkierungen über den einzelnen Balken markieren die Auslassungshäufigkeiten der am häufigsten bzw. der am seltensten ausgelassenen Frage im jeweiligen Themenblock (falls dieser mindestens 2 Items aufweist).

Abbildung 1 zeigt, dass die Item-Nonresponses thematisch nicht zufällig verteilt sind. Die Abweichung der empirischen Verteilungen von einer Rechteckverteilung ist in jedem der drei Fälle signifikant². Besonders deutlich ist dies für Unternehmen A, wo der relativ hohe Anteil an Auslassungen bei den Items zu den höheren Führungskräften auffällt.

Die Streuungsmargen in den Balken zeigen, dass z.T. auch innerhalb der Themenblöcke einzelne Items überzufällig häufig ausgelassen werden. Die Auslassungsquoten sind also auch innerhalb der Blöcke nicht rechteckig verteilt³.

Aufschluss darüber, aus welchem Themenbereich besonders viele der ausgelassenen Items stammen, zeigt Tabelle 4⁴. Die fünf Items, die in jeder dieser MABs am häufigsten ausgelassen wurden, wurden den in der Tabelle gezeigten Themenkategorien zugeordnet. Die Themenkategorien sind hier so geordnet, wie dies in MABs üblich ist (Borg 2003b) und wie sie in den Befragungen in den Firmen A und C verwendet wurden⁵. Beschränkt man sich nur auf Einstellungsitems zu verschiedenen Facetten der Arbeitszufriedenheit, dann

2 $\chi^2(8, N=14640)=15597.02, p<.01$ (A1998); $\chi^2(8, N=39165)=1562.62, p<.01$ (B1999); $\chi^2(8, N=42697)=27782.68, p<.01$ (C1998).

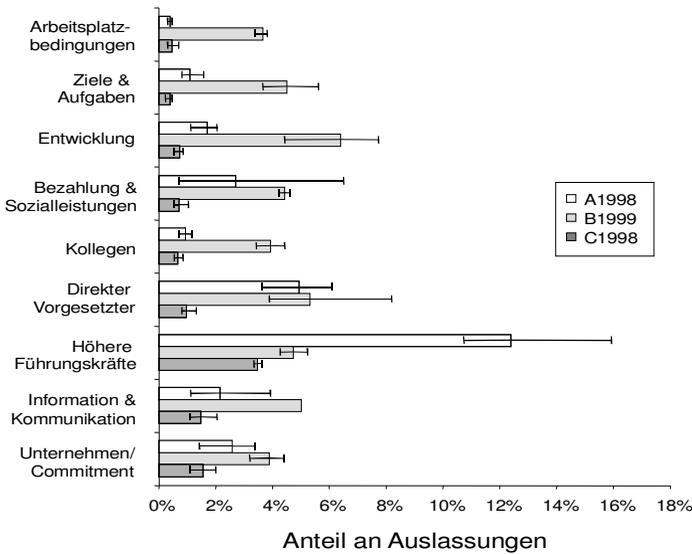
3 $\chi^2(33, N=14640)=16664.53, p<.01$ (A1998); $\chi^2(30, N=39165)=2880.30, p<.01$ (B1999); $\chi^2(32, N=42697)=29260.89, p<.01$ (C1998).

4 Um gleiche Gewichtung der verschiedenen Unternehmen zu bewahren, wurden von jedem Unternehmen gleich viele Befragungen einbezogen. Da für C nur zwei Befragungen vorlagen, wurden hierfür neben C1995 und C1998 die Datensätze A1998, A2000, B1998 und B1999 gewählt.

5 Die Zuordnung der Items zu den Themenkategorien ergab sich für die Befragungen A und C automatisch aus den Überschriften der Itemblöcke im Fragebogen, in denen das Item platziert war. Für Befragung B erfolgte die Zuordnung nachträglich aufgrund des Iteminhalts.

zeigt Tabelle 2, dass Auslassungen vor allem bei den Themen Höhere Führungskräfte und bei Entwicklung/Aufstieg zu finden sind. Betrachtet man hingegen alle inhaltlichen Items (alle Einstellungsfragen und zusätzlich alle Fragen zu Meinungen und Fakten), dann erweisen sich die Items zu den Geschäftspartnern/Lieferanten und vor allem die Items, die Fragen zur letzten Mitarbeiterbefragung stellen, als Hauptquellen für Missings⁶.

Abbildung 1: Anteil an Auslassungen bei Items zu Einstellungen zu verschiedenen Aspekten der Arbeitszufriedenheit



6 Für die Items, die nach Meinungen zur letzten Mitarbeiterbefragung fragten, wurde in den Fragebögen vermerkt, dass sich der Befragte nur dann äußern soll, wenn er bei der letzten MAB und ihren Folgeprozessen beteiligt war. Ein Hinweis ähnlicher Art wurde bei den Items zu anderen Themen nicht gegeben. Im Gegenteil: Im Einleitungsteil der Fragebögen wurde darauf verwiesen, dass „Vortests“ gezeigt hätten, dass es „eher unwahrscheinlich“ ist, dass der Befragte „keine Meinung“ zu einer Frage hat. Eine „Weiß-nicht“-Kategorie wurde nicht angeboten, weil sie das Ausweichen in diese Form des Nichtantwortens fördert (Dillman 2000). Das verhindert die Einbindung der Mitarbeiter in den intendierten Veränderungsprozess und ist daher für MABs kontraproduktiv.

Tabelle 2: Verteilung der 30 am häufigsten ausgelassenen Fragen über verschiedene Inhalte auf der Basis von 6 MABs

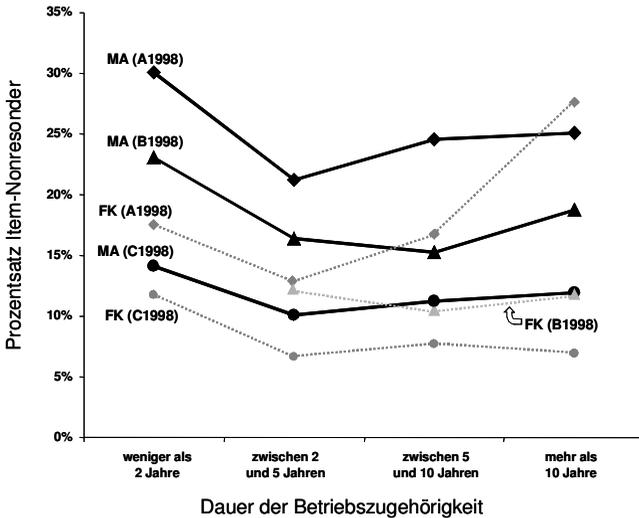
Thema	Einstellungen zu Aspekten der Arbeitszufriedenheit		Alle inhaltlichen Items	
	Anzahl	%	Anzahl	%
Arbeitsplatzbedingungen	2	7	0	0
Ziele & Aufgaben	0	0	0	0
Entwicklung	8	27	3	10
Bezahlung & Sozialleistungen	1	3	0	0
Kollegen	0	0	1	3
Direkter Vorgesetzter	1	3	2	7
Höhere Führungskräfte	17	57	5	17
Information & Kommunikation	1	3	0	0
Unternehmen/Commitment	0	0	0	0
Produktivität	-	-	1	3
Geschäftspartner/Lieferanten	-	-	6	20
Unternehmenskultur	-	-	2	7
Zur letzten Mitarbeiterbefragung	-	-	8	27
Aktuelle interne Themen	-	-	2	7
	30	100	30	100

Man kann zusätzlich noch den Mustern der Auslassungen nachgehen. Für A1998 findet man dabei, dass bei den Items zum Thema Höhere Führungskräfte in 1139 (=9,6%) der Fälle nicht nur einzelne Items, sondern gleich alle Items unbeantwortet blieben. Dieses Muster bestätigt sich in den Wiederholungsbefragungen in diesem Unternehmen in 2000 (8,9% vollständige Auslassungen) und in 2002 (7,5%). Eine zweite Gruppe von 3,1% der Befragten (A1998) lässt dagegen nur ein bestimmtes Item aus, das sich auch in den anderen MABs in gleicher Weise als bevorzugte Quelle für einzelne Missings bestätigt. Dieses Item lautet: „Diese Führungskraft unternimmt wirklich etwas, wenn von den Mitarbeiterinnen und Mitarbeitern Vorschläge oder Kritik kommen.“ Ähnliche Musteranalysen zeigen auch für die anderen Datensätze, dass die Auslassungen deutlich nicht-zufällig erfolgen.

3.3 Demografische Charakteristika

Im Folgenden beleuchten wir den Zusammenhang zwischen Item-Nonresponse und demografischen Charakteristika der Befragten in den drei MABs aus dem Jahr 1998. Betrachtet werden hier zwei demografische Variablen, die Standard sind für Mitarbeiterbefragungen: Dauer der Betriebszugehörigkeit und Position der Person in der Hierarchie (1=„Top-Management“, 2=„Senior Management“ usw. bis zu Mitarbeitern der Basis). Der Zusammenhang zwischen diesen demografischen Variablen beträgt $\tau_b = -.22$ (B1998) bzw. $\tau_b = -.19$ (A1998 und C1998). Aufstieg in der Hierarchie und Betriebszugehörigkeit korrelieren also positiv.

Abbildung 2: Anteil an Item-Nonrespondern bei Personen mit verschiedener Betriebszugehörigkeit und Position



In Abbildung 2 ist der Anteil an Item-Nonrespondern in Abhängigkeit von der Dauer der Betriebszugehörigkeit und der Position abgetragen. Der Prozentsatz ist hier immer bezogen auf die Personengruppe innerhalb der einzelnen Merkmalskombinationen. Hinsichtlich der Position wird lediglich zwischen Führungskraft (FK) und Mitarbeiter der Basis (MA) unterschieden. Für jedes der drei Unternehmen zeigt sich, dass Führungskräfte seltener Item-Nonresponder sind als Mitarbeiter der Basis (wie schon in Tabelle 1). Für die Betriebszugehörigkeit erkennt man tendenziell einen U-förmigen Trend: Neue Mitar-

beiter haben die relativ höchsten Auslassungsquoten; dann fallen die Antwortquoten für die Mitarbeiter mit einer Betriebszugehörigkeit von 2-5 Jahren; danach steigen die Auslassungen tendenziell wieder an.

Der Zusammenhang zwischen der Variable FK/MA und der dichotomen Variable INR-Einstellung (1=Person lässt ein oder mehr Items aus, 0=Person lässt kein Item aus) in den neun Befragungen liegt zwischen $r=-0,06$ und $r=0,00$. Die Korrelation zwischen FK/MA und der Anzahl an Auslassungen pro Person („Missings Einstellung“⁷) liegt im selben Wertebereich. Die Werte sind für die Betriebszugehörigkeit ähnlich: Die Korrelation mit INR-Einstellung liegt zwischen $\tau_b=-,08$ und $\tau_b=,01$. Nicht-lineare Trends zeigen sich nur schwach: Betriebszugehörigkeit und Missings Einstellung korrelieren beispielsweise nur mit $\eta=0,14$, wobei sich allerdings die relativ geringen N 's der Extremgruppen dämpfend auf den Koeffizienten auswirken.

3.4 INR, Commitment und Zufriedenheit

Im Folgenden wird die Beziehung von Item-Nonresponses zu allgemeinen Arbeitseinstellungen untersucht. Betrachtet werden hier das Commitment des Befragten zur Organisation, seine Arbeitszufriedenheit und seine Fluktuationsneigung.

Für die Befragung A2000 wurde der von Borg (2001) beschriebene Commitment-Index verwendet. Für die anderen Befragungen wurden ähnliche Indices konstruiert. Sie enthielten Items wie „Ich habe volles Vertrauen in den Vorstand“, „Ich stehe voll hinter der Strategie der Firma“ oder „Ich bin stolz, für die ABC AG zu arbeiten“. Cronbach's Alpha liegt für die drei Indices zwischen 0,83 und 0,85.

Ein Index für die allgemeine Arbeitszufriedenheit (aAZ) wurde berechnet durch Mittelung der Urteile zu den verschiedenen Facetten der Arbeitszufriedenheit, d.h. konkret durch Mittelung der Ratings bei den Items „Alles in allem bin ich zufrieden mit...[X]“, wobei X hier der Platzhalter ist für die verschiedenen Themenbereiche Arbeitsplatzbedingungen, Arbeitstätigkeit, Vorgesetzter, Kollegen usw. (vgl. Borg 2001). Cronbach's Alpha liegt für diese so berechneten summativen Skalen zwischen 0,85 und 0,91.

Die Fluktuationsneigung wurde mit dem üblichen „single item“ gemessen: „Ich plane ernsthaft, das Unternehmen in den kommenden 12 Monaten zu verlassen“.

7 Die Variable „Missings Einstellung“ gibt die Anzahl der ausgelassenen Einstellungitems in den Arbeitszufriedenheitsthemen pro Person an.

Tabelle 3 zeigt die linearen Korrelationen von Commitment, allgemeiner Arbeitszufriedenheit und Fluktuationsneigung mit den Auslassungsquoten bei Einstellungs- und bei anderen Items. Die Korrelationen sind äußerst klein. Allerdings sind alle Korrelationen für Commitment und für Arbeitszufriedenheit positiv, ein so nicht erwarteter Zusammenhang. Der nichtlineare Zusammenhang zwischen Missings Einstellung und Commitment liegt zwischen $\eta=.05$ und $\eta=.11$, zwischen Missings Einstellung und aAZ bei $\eta=.06$ bis $\eta=.11$. Die Korrelation zwischen Missings Einstellung und Fluktuationsneigung beträgt $\eta=.05$ bis $\eta=.08$.

Tabelle 3: Zusammenhänge zwischen Item-Nonresponse (INR-Einstellung bzw. Missing Einstellung) und Commitment, allgemeiner Arbeitszufriedenheit und Fluktuationsneigung

Befragung	Indikator	Commitment	aAZ	Fluktuationsneigung
A1998	INR Einstellung	.02	.06	.06
	Missings Einstellung	.01	.05	.06
B1998	INR Einstellung	.00	.03	-
	Missings Einstellung	.00	.02	-
B1996	INR Einstellung	.05	.05	-.03
	Missings Einstellung	.09	.06	-.05
C1995	INR Einstellung	.03	.04	-.01
	Missings Einstellung	.03	.03	.00

Legende: INR-Einstellung ist 1, wenn die Person ein oder mehr Items ausgelassen hat; 0 sonst. „Missing Einstellung“ ist die Zahl der Auslassungen pro Person.

3.5 INR und Einstellungen zu MABs

Eine mögliche Erklärung für Item-Nonresponse ist die, dass sie von Einstellungen zu und Erfahrungen mit Mitarbeiterbefragungen abhängt. Tabelle 4 zeigt hierzu die Korrelationen entsprechender Items in den verschiedenen MABs mit den Auslassungen bei anderen Items. Verwendet werden hier die drei Items: Befürwortung von Umfragen im eigenen Unternehmen; Zufriedenheit mit dem zu bearbeitenden Fragebogen; und Beurteilung der Nützlichkeit einer MAB. Wie man sieht, besteht ein systematischer, aber nur sehr schwacher Zusammenhang von positiver MAB-Einstellung und –Erfahrung mit der Tendenz, auf inhaltliche Fragen eher vollständig zu antworten.

Tabelle 4: Korrelationen zwischen Items zu Mitarbeiterbefragungen und INR bzw. Anzahl der Auslassungen

Item	Befragung	INR Einstellung	Ausgelassene Einstellungs- items	INR gesamt	Ausgelassene Items gesamt
Befürwortung von MABs	A2000	-.03	-.03	.01	-.03
	A1998	-.05	-.06	-.02	-.04
	B1996	-.05	-.04	-.05	-.05
	C1995	-.04	-.01	-.00	-.01
	C1995	-.02	-.00	.01	-.00
	C1995	-.03	-.00	-.00	-.01
Zufriedenheit mit Fragebogen	A1998	-.01	-.00	.03	.01
	A1998	-.05	-.07	-.04	-.06
	C1995	-.03	-.02	-.02	-.03
Nützlichkeit Einer MAB	A2000	-.12	-.11	-.06	-.09
	A2000	-.02	-.04	-.01	-.03
	A1998	-.03	-.03	.04	.01
	A1998	-.04	-.05	-.01	-.04

3.6 INR und Bearbeitungszeitpunkt

Man kann weiter vermuten, dass Item-Nonresponse mit dem Zeitpunkt der Abgabe des Fragebogens korreliert, weil Spätantworter allgemein weniger motiviert sein könnten als Frühantworter. Um diese Frage zu beantworten, werden die Daten aus den Online-Befragungen A2000 und A2002 herangezogen, in denen Abgabedatum und Abgabezeitpunkt exakt erfasst wurden. Die Korrelation zwischen dem Halbtage (A2000) bzw. dem Tag der Abgabe⁸ (A2002) und INR ist $r=0,03$ (A2002) bzw. $r=0,00$ (A2000). Die Korrelation zwischen dem Abgabezeitpunkt und der Anzahl der Auslassungen bei Einstellungsitems ist $r=0,03$ (A2002) bzw. $r=0,04$ (A2000). Abgabezeitpunkt und Item-Nonresponse sind also praktisch linear unkorreliert.

Bei genauerer Betrachtung zeigen sich Nichtlinearitäten, die jedoch keiner einfachen Funktion über der Zeitachse folgen und zudem auf sehr verschiedenen N 's pro Zeitabschnitt beruhen. Die A2002-Studie zeigt vor allem am ersten Tag der MAB einen relativ geringen Anteil an Nonrespondern. Danach pendelt dieser Anteil um knapp 20%, unabhängig vom Tag der Abgabe, mit leicht zunehmender Tendenz.

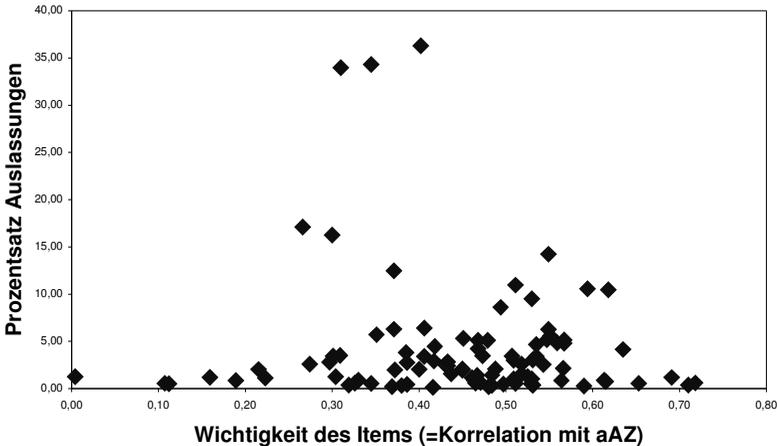
⁸ Normalisiert jeweils auf die Zeitzone, in der sich der Befragte befand. Siehe dazu genauer Borg/Tuten (2003).

Für A2000 ist diese Prüfung durch technische Eigenschaften der Erhebung erschwert: An Wochenenden und während einer virusbedingten Systemabschaltung (2 Tage) konnten keine Fragebogen abgegeben werden. Hier zeigt sich aber auch, dass der Anteil der Non-responder am Anfang der Befragung und nach jeder Unterbrechung zunächst relativ klein ist und dann tendenziell ansteigt bis zur nächsten Unterbrechung.

3.7 INR und Wichtigkeit des Items

Abschließend wird der Zusammenhang zwischen Item-Nonresponses und der Wichtigkeit von Items für den Befragten betrachtet. Damit wird die Vermutung überprüft, dass wichtige Items eher weniger häufig zu Auslassungen führen als unwichtige. Die Wichtigkeit des Items wird hier gemessen als Korrelation des Items mit der allgemeinen Arbeitszufriedenheit (=aAZ). Items werden hier also als „wichtig“ interpretiert, wenn sie sich zur Erklärung der allgemeinen Arbeitszufriedenheit der Befragten eignen. Neben- und Einzelthemen sind in diesem Sinn nicht so wichtig. Der Zusammenhang zwischen der Wichtigkeit des Items und seiner Auslassungsquote liegt zwischen $r=0,19$ (B1999) und $r=-0,27$ (A1998). Abbildung 3 zeigt beispielhaft das Streudiagramm des Zusammenhangs zwischen Auslassungsquote und Wichtigkeit der Items für A2000. Das Diagramm ist typisch für alle Befragungen, wobei in vielen Befragungen die Datenpunkte eher noch deutlicher einen nahezu horizontalen Regressionstrend bilden.

Abbildung 3: Streudiagramm der Wichtigkeit des Items und den Auslassungen für dieses Item, für alle Items in A2000



Auffällig in Abbildung 3 sind drei Ausreißer am oberen Rand. Diese Punkte entsprechen den drei Fragen, die zum Thema „Letzte Mitarbeiterbefragung“ gestellt wurden. Auch in einigen anderen Befragungen kann man einzelne Ausreißer identifizieren. Bereinigt man die Korrelationen um diese Werte, so liegen die resultierenden Zusammenhänge zwischen $r = +0,19$ (B1999) und $r = -0,19$ (B1996).

4. Diskussion

Die obigen Analysen haben unserer Meinung nach eines recht deutlich gezeigt: Die landläufige Vorstellung, dass Item-Nonresponses (in Mitarbeiterbefragungen) vor allem dadurch entstehen, dass der Befragte zwar eine Antwort hat, diese aber aus bestimmten Motiven, einer allgemeinen Stimmung oder aus bestimmten Dispositionen heraus nicht berichtet (weil er nicht „will“ oder weil er zu unengagiert ist), die Ausnahme zu sein scheint. Vielmehr zeigen die Verteilungen der Missings über die verschiedenen Themenbereiche, dass Auslassungen vor allem dort vorkommen, wo der normale Mitarbeiter i. allg. weniger gut Bescheid wissen dürfte. Diese Deutung wird auch dadurch gestützt, dass Missings seltener bei Fragen auftreten, die mit Arbeitsaspekten zusammenhängen, die als wichtige Bausteine der Arbeitszufriedenheit bekannt sind und die damit aus der unmittelbaren Umwelt des individuellen Befragten stammen. Auch die Korrelationen von Item-Nonresponses mit dem Zeitpunkt des Ausfüllens des Fragebogens (Früh- vs. Spätantworter), mit der Wichtigkeit des Items für die Arbeitszufriedenheit, mit Fluktuationsneigung, mit Commitment zum Unternehmen und mit Arbeitszufriedenheit sind ausnahmslos extrem klein.

Die naheliegende Annahme, dass z.B. ein größeres Commitment zu weniger Auslassungen führt, bestätigt sich hier jedenfalls nicht. Tendenziell ist sogar eher das Gegenteil zu beobachten, nämlich leicht positive Korrelationen zwischen Missings und Commitment. Folgt man der obigen Interpretation, dass der Hauptgrund für Nichtantworten Nichtwissen oder Unsicherheit ist, dann kann man das so deuten, dass mehr Commitment dazu führt, dass die Befragten genauer nachdenken und lieber keine Antworten geben als „schlechte“.

Ebenfalls konsistent mit unserer Interpretation ist der Befund, dass Auslassungen bei neuen Mitarbeitern häufiger vorkommen. Diese sind eben in vieler Hinsicht noch nicht so gut informiert oder jedenfalls unsicherer. Bei älteren Mitarbeitern zeigt sich ebenfalls ein leichter Anstieg, vielleicht deshalb, weil hier das Interesse nachlässt, sich mit vielen Themen zu beschäftigen. Der Mitarbeiter beschränkt sich dann eher auf seine Welt und kümmert sich um den Rest nicht weiter.

Wenn wir an die Befunde von Rogelberg et al. (im Druck) zur Unit-Nonresponse anknüpfen, der zeigt, dass der weitaus größte Teil der Unit-Nonresponder lediglich passive Verweigerer sind, dann ist unsere Interpretation für Item-Nonresponder ähnlich profan: Der Großteil der Nichtantworter ähnelt den Antwortern in den hier betrachteten Charakteristika stark. Das heißt nicht, dass nicht auch das von seiner Psychologie her viel interessantere bewusste Verschweigen einer an sich klaren, aber negativ-kritischen Meinung vorkommen kann. Borg (1991) hat beispielsweise nachgewiesen, dass Auslassungen bei demografischen Items einhergehen mit geringerer Zufriedenheit und negativeren Einstellungsäußerungen. Die Effekte sind allerdings nicht sonderlich groß. Wenn man also Item-Nonresponses vermeiden will, dann empfiehlt sich ein sorgfältiges Pretesting der Items daraufhin, ob die Befragten auf diese Items leicht antworten können. Schwierige Items sollte man in Selbstausfüllern möglichst reduzieren. Heikle Fragen (hier z.B. Fragen zum Vorgesetzten) kann man durch gute und glaubhafte Maßnahmen zur Sicherung der Anonymität entschärfen, was in den hier berichteten Umfragen offensichtlich weitgehend gelungen ist.

Zudem legen die hier gezeigten Befunde nahe, dass man u.U. nicht für alle Zielpersonen einer Mitarbeiterbefragung alle Items verwenden sollte oder diese jedenfalls in jeweils anderer Weise formulieren könnte. So könnte man z.B. die Fragen zu den höheren Führungskräften für Mitarbeiter der Basis eher affektiv akzentuieren, für Führungskräfte dagegen eher kognitiv-faktisch. Damit hätten dann die verschiedenen Schichten die Möglichkeit, sich zu diesen Inhalten so zu äußern, wie es ihren jeweiligen Wissens- und Verhaltenskategorien entspricht. Das sollte nicht nur die Zahl der Nichtantworten reduzieren, sondern auch Reliabilität und Validität der Antworten erhöhen.

Korrespondenzadressen

Prof. Dr. Ingwer Borg
ZUMA
Postfach 12 21 55
D – 68072 Mannheim
email: borg@zuma-mannheim.de

Christian Treder
Human Resources Consulting
Nymphenburgerstraße 148
D – 80634 München
email: Christian.Treder@hrc-online.de

Literatur

- Bauer, J., 1990: Mitarbeiterbefragung bei Karstadt: Ein Instrument zur Organisationstherapie. S. 131-145 in: M. Domsch/A. Schneble (Hrsg.), Mitarbeiterbefragungen. Heidelberg: Physica.
- Borg, I., 1991: Sind Personen, die sich in Umfragen nicht identifizierbar machen, besonders kritisch? Zeitschrift für Sozialpsychologie 22: 123-129.
- Borg, I., 2001: Früh- versus Spätantworter. ZUMA-Nachrichten 47: 7-19.
- Borg, I., 2003a: Affektiver Halo in Mitarbeiterbefragungen. Zeitschrift für Arbeits- und Organisationspsychologie 47: 1-11.
- Borg, I., 2003b: Führungsinstrument Mitarbeiterbefragung. Göttingen: Hogrefe. 3. Auflage.
- Borg, I./Tuten, T., 2003: Early versus later respondents in Intranet-based, organizational surveys. Journal of Behavioral and Applied Management 4 (1): 134-147.
- Colsher, P.L./Wallace, R.B., 1989: Data quality and age: Health and psychobehavioral correlates of item nonresponse and inconsistent responses. Journal of Gerontology 44: 45-52.
- Coombs, C.H./Coombs, L.C., 1976: "Don't know": Item ambiguity or response uncertainty? Public Opinion Quarterly 40: 497-514.
- Craig, C.S./McCann, J.M., 1978: Item nonresponse in mail surveys: Extent and correlates. Journal of Marketing Research 15: 285-289.
- Dalecki, M.G./Whitehead, J.C./Blomquist, G.C., 1993: Sample nonresponse bias and aggregate benefits in contingent valuation: An examination of early, late and nonrespondents. Journal of Environmental Management 38 : 133-143.
- De Leeuw, E.D./Hox, J.J./Van der Zouwen, J., 1989: Data quality in mail surveys as compared to face to face and telephone interviews: A meta-analysis of the research literature. Kwantitatieve Methoden 31: 123-146.
- Dickinson, J.R./Kirzner, E., 1985: Questionnaire item omission as a function of within-group question position. Journal of Business Research 13: 71-75.
- Diekmann, A., 1995: Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen. Hamburg: Rowohlt.
- Dillman, D., 2000: Mail and Internet surveys. New York: Wiley.
- Graham, J.W./Donaldson, S.I., 1993: Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data. Journal of Applied Psychology 78: 119-128.

- Green, K.E., 1991: Reluctant respondents: Differences between early, late, and nonresponders to a mail survey. *Journal of Experimental Education* 59: 268-276.
- Guadagnoli, E./Cleary, P.D., 1992: Age-related item nonresponse in surveys of recently discharged patients. *Journal of Gerontology* 47: 206-212.
- Hoox, J./De Leeuw, E., 1998 : Patterns of item nonresponse. Vortragspapier zur RC33, Köln. März.
- Johanson, G.A./Gips, C.J./Rich, C.E., 1993: "If you can't say something nice": A variation on the social desirability response set. *Evaluation Review* 17: 116-122.
- Jones, W.H./Lang, J.R., 1983: Reducing item omissions in mail surveys. *Political Methodology* 9: 285-293.
- Leigh, J.H./Martin, C.R., 1987: "Don't know" item nonresponse in a telephone survey: Effects of question form and respondent characteristics. *Journal of Marketing Research* 24: 418-424.
- Messmer, D.J./Seymour, D.T., 1982 : The effects of branching on item nonresponse. *Public Opinion Quarterly* 46: 270-277.
- Neuberger, O., 1974: Messung der Arbeitszufriedenheit. Stuttgart: Kohlhammer.
- Omura, G.S., 1983: Correlates of item nonresponse. *Journal of the Marketing Research* 25: 321-330.
- Pearl, D./Fairley, D., 1985: Testing for the potential for nonresponse bias in sample surveys. *Public Opinion Quarterly* 49: 553-560.
- Porst, R., 2001: Wie man die Rücklaufquote bei postalischen Befragungen erhöht. ZUMA How-To Reihe, Nr. 9.
- Rogelberg, S.G./Conway, J.M./Sederburg, M.E./Spitzmüller, C./Aziz, S./Knight, W.E., (im Druck): Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology*.
- Schwirian, K.P./Blaine, H.R., 1966: Questionnaire-return bias in the study of blue collar workers. *Public Opinion Quarterly* 30: 656-663.
- Spector, P. E., 1998: Job satisfaction: Application, assessment, causes, and consequences. Thousand Oaks: Sage.
- Tourangeau, R./Smith, T.W., 1996: Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly* 60: 275-304.
- Zimmermann, M./Jordan, R., 2003: Online-Mitarbeiterbefragungen. S. 91-114 in: A. Theobald/M. Dreyer/Th. Starsetzki (Hrsg.), *Online-Marktforschung: Theoretische Grundlagen und praktische Erfahrungen*. Wiesbaden: Gabler.

SHARE¹ – MEASURING THE AGEING PROCESS IN EUROPE

AXEL BÖRSCH-SUPAN, HENDRIK JÜRGES & OLIVER LIPPS

The purpose of the paper is to present SHARE, a EU-sponsored project that will build up a Survey of Health, Aging and Retirement in Europe. We detail objectives, milestones, and current status. A special focus is the description of the cross-national and interdisciplinary nature of the project which introduces considerable complexity in light of the many different options and restrictions in the participating countries.

Ageing is one of the largest social and economic challenges of the 21st century in Europe. SHARE will be a fundamental resource to help mastering this unprecedented challenge. The main aim of SHARE is to create a pan-European interdisciplinary panel data set covering persons aged 50 and over. Scientists from some 15 countries work on feasibility studies, experiments, and instrument development, culminating in a survey of about 22.000 individuals. Project participants are currently eleven countries ranging from Scandinavia (Sweden, Denmark), Western and Central Europe (France, Belgium, The Netherlands, Germany, Switzerland, Austria) to the Mediterranean (Spain, Italy, Greece). SHARE will be based on best practice technologies in the participating countries. The survey will follow a common set-up across all countries with the goal of collecting data that are strictly comparable to allow cross-country research. Hence, one of the most difficult tasks consist in taking into account differences in language, culture and institutions. Other difficult tasks are of a more technical nature such as developing country-specific feasible sample designs and making use of suitable sampling frames that are already available.

The main objective of SHARE is to provide a fundamental knowledge base for science and public policy in order to understand and to master the challenges posed by population

1 A Survey on Health, Aging and Retirement in Europe. The paper is basically an excerpt of the mea discussion paper 32-03: http://www.mea.uni-mannheim.de/mea_neu/pages/files/nopage_pubs/k3pzhnwk1t1zjymk_dp32.pdf
Acknowledgements: We are grateful for financial support from the European Union DG RESEARCH under the 5th framework program and from the U.S. National Institute on Aging (NIA).

ageing. Due to the dynamic character of ageing, both on an aggregate and an individual level, SHARE is being designed in a longitudinal way. Specifically, SHARE aims to create, evaluate and analyse a large-scale pan-European and interdisciplinary household panel as a new state of the art survey. The collected data include information on economics, physical and mental health, and social support networks. The multidisciplinary nature of the data will provide new insights in the complex interactions between economic, health, psychological and social factors determining the quality of life of the elderly in Europe.

1. Introduction

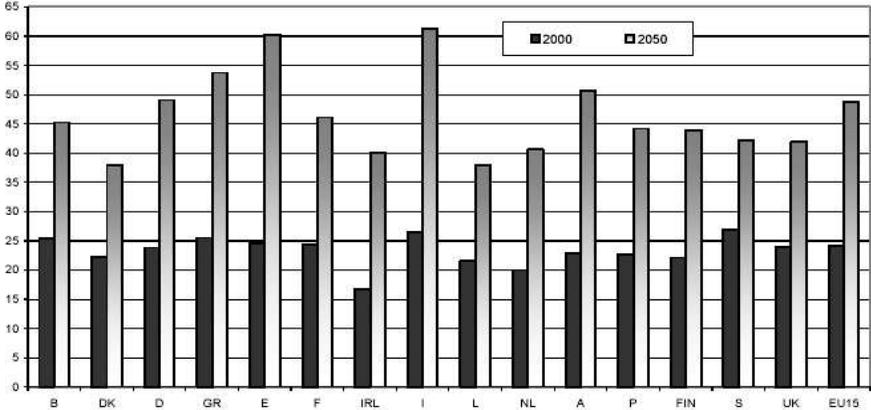
1.1 Background

Ageing is one of the greatest social and economic challenges of the 21st century in Europe. Of the world regions, Europe has the highest proportion of population aged 65 or over, with Italy the leading country. Outside Europe, only Japan has a similar age structure. In Europe, the ratio of persons aged over 65 as a percentage of the working age population 20-64 (the dependency ratio), is expected to increase from about 24% in 2000 to 38% in 2025, and to 49% in 2050, see Figure 1. This increase of the dependency ratio in itself places a heavy financial burden on society through pay-as-you-go financed pension, health and long-term care systems.

In addition, Europeans retire much earlier than inhabitants of other developed countries: for instance, in Belgium only a quarter of all males are still in the labour force at age 55-64, (see Figure 2, compared to three quarters in Japan (U.S. National Academy of Sciences 2001). This typically European combination of an ageing population and retirement at ever earlier ages with relatively generous benefits puts very severe strains on our capacity to care for the elderly in the future.

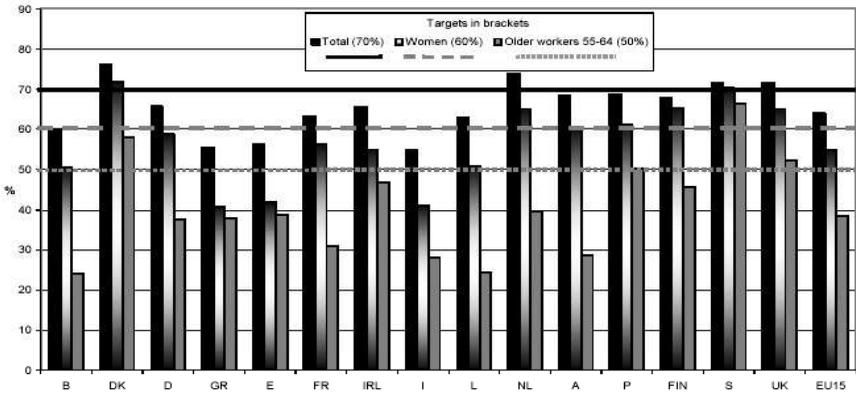
There are other pressures on the social security and welfare system as well. Older workers are more likely to be in disability programmes and generally the health care cost of the elderly is substantially higher per capita than of the non-elderly. A special aspect of ageing lies in the increasing number of the oldest old, a population segment with a high prevalence of long-term care needs (Suzman et al. 1992). The ageing of society will cause the number of people on disability or consuming health care to go up.

Figure 1: Old-Age Dependency Ratio (persons aged 65 and over divided by persons 20-64), 2000-2050 ²



Source: EUROSTAT Population projections - Baseline scenario

Figure 2: Share of Men Aged 55-64 Still in the Labour Force, 2001



Source: EU-LFS, EUROSTAT

² Taken from: www.europa.eu.int/comm/employment_social/soc-prot/pensions/index_en.htm

Thus, everything else equal, ageing places a much higher burden on the sustainability of income maintenance systems in Europe than elsewhere in the world, and European public policy – pension policy, health care policy, labour market policy – is challenged in particular. Public policy plays an important role in explaining the differences in health care utilisation or disability insurance across countries (Aarts et al. 1996), public policy appears to be a major factor in explaining the low retirement ages in Europe (Gruber/Wise 1999), and public policy has strongly shaped savings and wealth patterns across Europe, Japan and the United States (Börsch-Supan 2003).

1.2 Prepared policy making

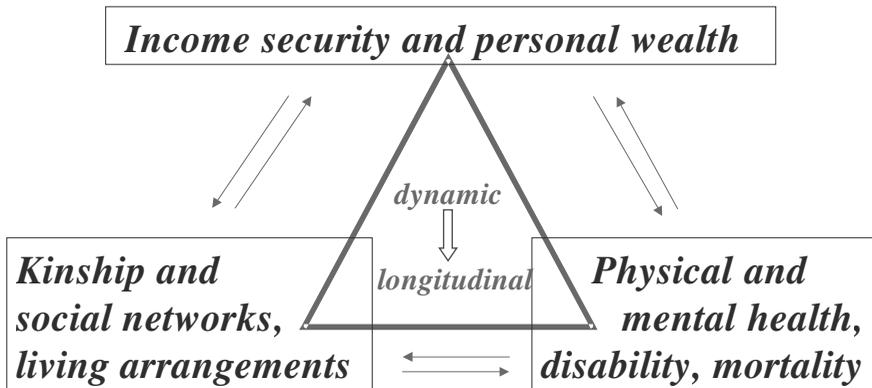
To deal with the challenges for public policy posed by the ageing of European societies, one needs to understand the complex interactions among economic, health, psychological and social factors that determine the quality of life of the elderly. These interactions are symbolised by the arrows in Figure 3 which link the three corners of the triangle, each representing the three broad fields of economics, health and social networks.

Income and wealth strongly affect health and well-being of the elderly. For example, it is well-known that wealthier persons live considerably longer than poorer persons. The direction of causality, however, is not well understood. Wealthier people may be able to afford more health care and thus remain longer healthy, once older. On the other hand, less healthy people may have been hampered in their earnings ability and career chances, ending up as less wealthy elderly. A better understanding of what causes what under which circumstances will permit us to better target our policy actions.

Another bidirectional link is between health and family/social networks. A “healthy” social environment keeps elderly longer healthy, In turn, health shocks such as a stroke often precipitate a change in living arrangements such as a move to children or into a nursing home. Again, understanding the linkages is important in times of population ageing when the potential supply of family help (the number of children per elderly) will decline and the demand for state-provided help will increase, straining the financial abilities of the EU member states.

The triangle is closed by interactions between income security and social environment. On the one hand, a well working social network is a resource also in an economic sense, providing money and in-kind support for the less well-to-do elderly. In turn, income and wealth position are strong determinants of where the elderly will live. Also these linkages are strongly affected by public policy such as income maintenance programs, old-age and disability pensions.

Figure 3: Interactions among economic, health, and social factors in the well-being of the elderly



The linkages are dynamic because the elderly age individually (ageing is a process over time, not a state in time). An analysis of the linkages in Figure 3 therefore requires a longitudinal point of view, symbolised in the interior of the triangle. Moreover, the institutional frame is changing over time since we observe how the EU member states go through the demographic ageing process and adapt their pension systems, restructure health care policies and labour market regulations accordingly.

An understanding of these dynamic linkages and how they are affected by community and national policies requires multidisciplinary data and research on ageing. In this respect, however, Europe is ill-equipped. While some member states have collected data in specific disciplines at various points in time, there is no Europe-wide longitudinal and multifaceted knowledge base for this crucial challenge of our new century. Analysing cross-nationally comparable data provides a large added value to the EU because analysing data on a pan-European level is worth more than the sum of its national parts. The main reasons are:

First, matters of economic and social policy are increasingly Community matters, due to the increasing personal and capital mobility, precipitating common policies and regulations such as the pension directive. The gradually increasing importance of the method of open co-ordination requires indicators based on reliable and comparable data such as collected in SHARE to assess and guide Community policy.

Second, the diversity in institutional histories, policies, and cultural norms, represents a unique living laboratory in which the various determinants of the current economic, health and socio-psychological conditions can be understood much easier than in the more homogeneous environment of a single country. A large added value provided by SHARE both to science and to society is to exploit this living European laboratory for the analysis of the elderly's quality of life. The insights gained from analysing and comparing the diversity of experiences will help both a supranational body like the EU and its member countries to prepare more effectively for the continuing changes in age demographics in the future.

2. Innovation

The innovation of the SHARE project lies in its multidimensional design which combines interdisciplinarity, cross-national comparability, and longitudinality. Never before has a team from such diverse disciplines collected longitudinal data involving so many countries.

In order to study the quality of the life of the elderly and how it is affected by the population ageing process and by the various social and economic policies in Europe, one needs multidisciplinary, longitudinal and internationally comparable data:

Multidisciplinary data

- One needs *multidisciplinary data*, for the simple reason that many societal aspects of ageing have a multidisciplinary character (e.g. retirement and health, or financial and health factors determining inflow in disability insurance programmes). To deal with the challenges for public policy posed by the ageing of European societies, one needs to understand the complex interactions among economic, health, psychological and social factors that determine the quality of life of the elderly, and in particular the mechanisms through which policy measures such as pension reform, health care reorganisation and labour market restructuring affect elderly citizens.

Longitudinal data

- One needs *longitudinal data*, because many events associated with ageing are dynamic in nature. For instance, current pensions or social security benefits will usually depend on one's earnings history; current health is partly determined by past behaviour and past health events. Without longitudinal data one cannot distinguish between age and cohort effects. That is, if we observe differences in, for example, health, income or wealth between individuals of different age, we cannot ascertain if the difference is simply due to age or due to the fact that the younger person is on a different trajectory than the older person, because of the different life experiences associated with different cohorts.

Internationally comparative data

- One needs *internationally comparative data* to exploit the rich variety in policies, institutions and other factors across European countries. The impact of public policy can only be understood if we observe one policy in contrast to other policies. Many of the policies that one might want to consider to address future public policy challenges resulting from an ageing population, have already been implemented in some form in at least one of the European countries. Exploiting the variation in institutions across European countries creates a unique *laboratory* in which to study the effects of institutions on societal processes (Gruber/Wise 1999).

The unique and innovative feature of SHARE lies in the *combination* of these three features. We have interdisciplinary data sets in some countries, notably the English Longitudinal Survey on Ageing (ELSA)³ and the Health and Retirement Study (HRS)⁴ in the United States, the German “Alterssurvey” and the Italian Longitudinal Survey on Ageing. We also have cross-national data sets on single issues, notably the European Community Household Panel (ECHP), its successor, the Survey of Income and Living Conditions (SILC), the European Social Survey (ESS), and the various health surveys collected by the WHO. Some of these data sets are longitudinal (ELSA, HRS and ECHP). The combination of interdisciplinarity and longitudinality has made ELSA and HRS role-models for SHARE.

The cross of longitudinality, genuine interdisciplinarity, and a truly cross-national design, however, has not been attempted before. In addition, SHARE is designed to meet all country specific institutional and linguistic requirements in a single common design.

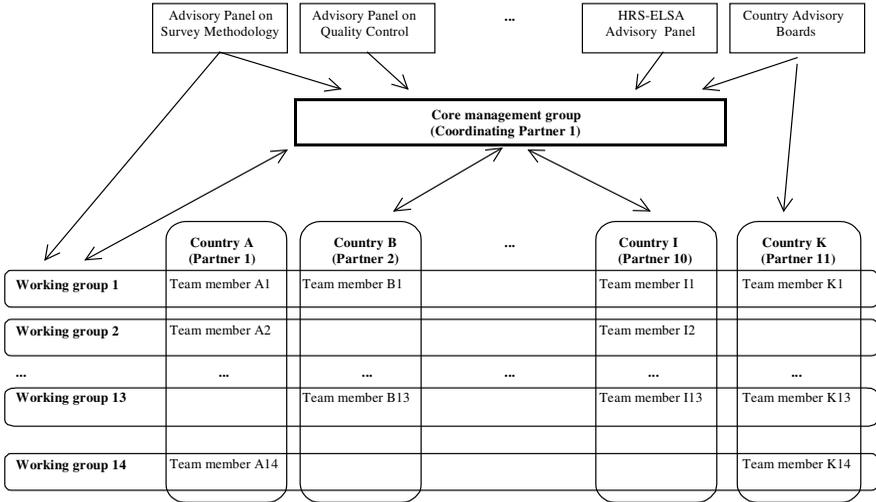
3. Participating Partners and Organisational Structure

Researchers involved in SHARE are organised in multidisciplinary country teams (CT) and cross-national working groups (WG) in a “matrix organisation”, assisted by a number of expert support and advisory teams. Each researcher belongs to both a country team and a working group. The organisational structure is summarised in Figure 4:

3 <http://www.natcen.ac.uk/elsa/>

4 <http://hrsonline.isr.umich.edu/>

Figure 4: Matrix structure of SHARE working groups and country teams



Team member A1 comes from country A and is an expert in field 1. This team member therefore is a member of country team A and working group 1. Ideally, each country is represented in all working groups. This is not always feasible, explaining some empty cells in the matrix. Multidisciplinary country team A consists of researchers A1, A2, ..., and A14. In turn, the cross-national working group 1 consists of researchers A1, B1, ..., I1 and K1.

In addition to the matrix of country teams and working groups, the SHARE structure involves a core management group, advisory boards and support groups. The division of labour is as follows:

A **core management group** supervises the entire project. The overall direction of the project will be carried out by the co-ordinator in collaboration with the core management group which consists of internationally-respected senior experts in their fields. The core management group settles potential disagreements among country teams and working groups or between a country team and a working group.

The co-ordinator is aided by a **co-ordination team**. It ensures overall quality and cross-national comparability, co-ordinates the development of the questionnaire modules, and co-ordinates and participates in all crucial negotiations.

The *country teams* are responsible to conduct the project in all of its phases in their respective countries. In particular, they negotiate with the survey agencies to conduct the national survey, manage the translations, participate in the training process to motivate the interviewers and oversee the fieldwork. The country teams are also responsible to make sure that the survey does justice to country-specific institutions (such as health care and pension system) and follows country-specific legal requirements (such as data confidentiality).

The task-oriented cross-national *working groups* consist of those members in each country team who are specialists in the field of the working group. The working groups design the questionnaire modules, conduct response analyses during the development process and modify the questionnaires accordingly. The working group leaders (WGL) are leading specialists in their fields. The composition and leadership of each working group is determined by the co-ordinator. Eleven working groups will produce the questionnaire design and write up subject-specific parts of the final report on a design of SHARE. In addition, three working groups address methodological issues.

In order to draw from the best experience available, several *advisory institutions and review panels* have been set up. There are several ad hoc advisory panels on overarching issues such as survey methodology, quality control, and data management and dissemination. Furthermore, SHARE is supported by an advisory group consisting of leading researchers of the US HRS and the UK ELSA.

4. Questionnaire Content and Questionnaire Design

Data to be collected will include *health variables* (e.g. self-reported health, physical functioning, cognitive functioning, health behaviour, use of health care facilities), *psychological variables* (e.g. psychological health, well-being, life satisfaction, control beliefs), *economic variables* (e.g. current work activity, job characteristics, job flexibility, opportunities to work past retirement age, employment history, pension rights, sources and composition of current income, wealth and consumption, housing, education), *social support variables* (e.g. assistance within families, transfers of income and assets, social networks, volunteer activities, time use).

All data will be collected by face-to-face, computer-aided personal interviews (CAPI), supplemented by a self-completion (“drop off”) paper and pencil section. The generic survey instrument is written in English as a computer program in the Blaise language. In each country or region, the English text is replaced with text in its own language. All texts are stored in a data base that can be accessed for translation and editing by a “language management utility” (LMU).

4.1 Respondents

Respondents are all household members aged 50 and over, plus their spouses, independent of age. Example: Anna is 52 years old. She lives together with her husband Bert of age 49 and her daughter Cecilia (age 17). In the same household lives also Bert's mother Dorothy who is of age 70. SHARE will interview Anna, Bert and Dorothy.

In order to save time and avoid duplications, some parts of the questionnaire need only be answered by one respondent in a household or couple, respectively. Questions on housing and housing finances should be answered by the household member who is most knowledgeable in housing matters ("housing respondent"). Questions about finances need be answered by one person in a couple only, again preferably by the partner who is most knowledgeable ("financial respondent"). If a couple keeps its finances completely separate, each partner will be treated as a separate financial unit and each will answer his/her own questions on finances.

A single-person interview is designed to take 80 minutes while the interview length for a couple is about 120 minutes.

4.2 Description of modules

In the following, each module of the questionnaire is described, in the order in which it appears in the questionnaire. The current version of the questionnaire is available on www.SHARE-project.org.

Coverscreen: The interview starts with a "coverscreen" that provides an introduction to the study and contains the statement of confidentiality. The coverscreen collects basic demographic information about everyone who currently lives in the household (name, gender, birth year and month, relationship to informant, and whether married or living with someone as married). It establishes whether household members are eligible for a SHARE interview and who is going to be the housing, financial, and family respondent. This section only needs to be completed by one person in each household, the "informant".

Demographics: This module collects details about each respondent's marital status, country of birth, education, and occupation. It also collects selected details about parents such as their last occupation, health status, and frequency of contact.

Physical Health: This module covers many different aspects of people's health; self-reported general health, longstanding illness or disability, eyesight and hearing, specific diagnoses and symptoms, pain, and difficulties with a range of activities of daily living.

Behavioural Risks: This module collects information on health behaviours such as smoking, alcohol use, and physical activities.

Cognitive function: This module contains subjective and objective measures of four aspects of the respondents' cognitive functioning: literacy, numeracy, memory, and verbal fluency.

Mental Health: This module asks how the respondents' view their lives and collect information about emotional problems.

Health Care: This module asks about recent doctor visits and hospital stays. It also contains questions about the respondents' level of health insurance.

Employment and Pensions: This module collects information about the respondents' current work activities, their income from work and other sources, and any current or past pensions that they may be entitled to. For respondents who have retired and are receiving a pension, we ask about the number and kind of pensions and how much they receive.

Grip Strength: This type of physical measurement involves recording the respondents' maximum handgrip strength with the aid of a dynamometer.

Walking Speed: This type of physical measurement involves asking the respondents to walk a certain distance and measuring the time it takes for the respondents to complete this activity. Only persons of age 65+ conditional of previously asked physical health, or persons of age 75+, respectively, are asked to perform this test.

Children: This module collects information about the respondents' children.

Social Support: This module collects information about any help the respondents might receive from family and other people not living in the household and how household members help others. Questions on most kinds of help received by members of a couple are asked of the "family respondent".

Financial Transfers: This module asks the "financial respondent" about any regular financial transfers and payments the respondents may have given or received from non-household members. It also asks about inheritances.

Housing: This module collects information about the respondents' current housing situation, including the size and quality of the accommodation. Owners are asked about the value of their property and, depending on the individuals' tenure, questions are asked about mortgages and rent payments. The section on housing is asked of one person per household, regardless of how many people are eligible for the interview.

Household Income: This module collects summary measures of the household income from various sources.

Consumption: This module asks about various types of household expenditures, e.g. on food, fuel, electricity, and telephone. It is answered by the “housing respondent”.

Assets: This module asks about the amount of financial and non-financial assets held in various forms and income from these assets. This section will be completed by one person in each financial unit (the “financial respondent”). A financial unit is defined as either a single person or a couple, so in most couples only one of them will complete the sections on assets on behalf of both of them.

Expectations: This module explores people’s expectations, the level of certainty they feel about the future, and how they value risk and make financial decisions within their household.

Interviewer Observations: This module concerns the interviewing experience and should be answered by the interviewer as soon as possible after the interview. These questions are important in understanding the circumstances surrounding the interview and can sometimes help researchers clarify any confusing or conflicting information. Included are e.g. information of background interview characteristics, third persons present, time and day, atmosphere, area, housing, household characteristics, etc.

5. Fieldwork Procedures

It is crucial in SHARE to ensure consistency of methods and fieldwork procedures across countries in order to obtain a genuinely comparable cross-national survey of high quality. This section summarises the main elements by which SHARE enforces cross-national comparability and high quality standards.⁵

5.1 Objectives

The first objective of tight fieldwork procedures is to achieve high data quality, such as high response and low non-contact rates. For this reason, SHARE has selected most reputable survey agencies capable of carrying out data collection for this complex study. In all countries, the agencies must sign a common standard contract, along with country specific specifications. In order to provide common standards, our second and equally important objective, a member of the SHARE co-ordination team and a member of the

5 See Lipps (2002) for details.

working group on cross-national survey design must be involved in all the crucial negotiations, and detailed written standards have to be adhered to.

5.2 Keeping track of contacts and non-response

Contact data which are collected at all interactions with respondents, informants, and gatekeepers are sent to SHARE. Contacts, response and non-response outcomes are recorded, calculated and keyed according to a pre-specified standard format, which includes at least the mutually exclusive categories listed below, which are part of the sample management system provided by SHARE:

- Number of total issued and contacted addresses (or other sample units) and mode, time and date of contact and – if applicable – date of appointments for the interview
- Mode, time, and date of all contact attempts. Further details of the attempts must be delivered to the survey agency, including observable area, stratum, dwelling and housing conditions, information about moving or deceased, where possible
- Number, time, and date of household and “target respondents” refusal (if applicable) classified into standard categories (including where possible details of gender, age-bands)
- Number of respondents who are too ill or otherwise incapable (e.g. language problems) or not available, split into temporarily and permanently, if possible
- Number, time, and date of achieved interviews, started and still to be completed, and started but not to be completed interviews
- Number, time, and date of collected drop off questionnaires

These data are used to compute the following key statistics:

- Household – non-response
- Person–non-response (unit-non-response, by a set of pre-specified reasons, see below)
- Break-off during the interview by specific persons
- Item–non-response by person

In addition, reports are regularly submitted on verification efforts, plus regular frequency lists of key variables. The leader of the country teams (CTL) review the timing, breakout and frequency of the reports together with the SHARE co-ordinator.

Appropriate measures must be taken in case of a large discrepancy between current and prospected statistics of key variables.

5.3 Probability samples

Samples for the pre-test and the main survey⁶ are full probability samples. The sampling frames will differ according to availability in different countries. It is the responsibility of the CTL to construct together with the survey agency a sample design that is at the same time suitable for this country and compatible with all other SHARE sampling designs. All country-specific sampling procedures and the sampling process has to be approved by the SHARE co-ordination team and the SHARE working group on cross-national survey design. The addresses used in the main test survey will remain (co-)property of SHARE, such that re-interviewing in a future wave is feasible.

Quota sampling is not permissible for the pre-test and the main test survey. No oversampling by age or other socio-demographic characteristics is planned. The sampling frame (if existing and generally accepted) or sampling units at different stages, including the degree of clustering and the data base used for the selection of communities as well as stratification factors applied to the sampling frame, will be described in detail in the final report, detailing the following:

- The process of the household selection from a multi-household (or multi-individual) address has to be spelled out in detail by the agency and agreed in advance, before signing the contract.
- The selection probabilities of every sample household and every sample member must be estimated and recorded after the survey.
- The remaining systematic non-coverage problems (telephone sample coverage, language minorities, other impairments, e.g. a high rate of illiteracy) must be recorded.

5.4 Interviewer training

Training is the key to a successful survey. Hence, SHARE pays a lot of attention to interviewer training. This includes both technical aspects and motivation. The interviewers are trained personally by the survey agency and the CTL, who in turn is trained using the “train-the-trainer” materials. Participation of the CTL at all training meetings is crucial for the motivation of the interviewers and the quality of the content.

5.5 Fieldwork monitoring

SHARE will closely monitor the fieldwork progress during the pre-test and main test survey. This includes producing a weekly report on response rates. All survey agency must accept quality control back-checks (e.g. contacting interviewed households by SHARE to

6 See Stage 3: Development of multi-language sampling frame and Stage 4: Running the prototype “main test survey”.

ensure that the interview actually took place, acceptance of visits by CTL/co-ordinator, acceptance of code of ethics). The survey agency sends the raw data on a weekly basis directly to SHARE (i.e. without editing), by electronic means. In case of interviews with errors, these may be sent back to the field for correction.

5.6 Debriefing sessions

After each survey, survey agencies hold “debriefing” meetings with their interviewers, the CTL and, possibly, members of the SHARE co-ordination team in which interviewers report on their experiences during the fieldwork. The debriefing meetings after the UK and the all-country pilots⁷ were very successful in showing where such a complex survey needs improvement and revision to become efficient for the interviewer and pleasant for the respondents.

6. The Development Process

Core of the SHARE development process is the iteration between *questionnaire development* and *data collection*. Point of departure was the US HRS (Health and Retirement Survey), the UK ELSA (English Longitudinal Survey of Ageing) and similar other survey instruments (e.g., in Germany, Italy and Sweden) which have addressed relevant questions.

The development process is taking place in four stages:

Stage 1: Initial questionnaire design in English language

In the first stage, completed by now, the working groups produced an English-language draft questionnaire. The entire group met in plenary sessions during this process to test ideas and ensured that the proposed questions are likely to be viable in all participating countries.

The first stage culminated in an English-language pilot which took place in the UK in September 2002. The main purpose of this pilot was to test the feasibility of the survey instrument and the CAPI program. It was based on a quota sample: 40 households had at least one respondent aged 50-70, 40 households had at least one respondent aged 71-85, and 10 households had at least one respondent aged 86+. 30 households contained at least one respondent who was working. Single/couple or composite households were equally frequent. The pilot was conducted by the National Centre for Social Research (London) which has also conducted the first wave of ELSA. The aim of this pilot was to test the

7 See Stage 2: Development of multi-language instrument.

English language questionnaire, explore its length and item non-response frequencies for various household types, and collect interviewer feedback.

Stage 2: Development of multi-language instrument

Starting from the lessons from this UK-pilot, the English-language questionnaire was thoroughly revised. The next stage consisted of an array of cognitive interviews in selected countries based on the English-language questionnaire in order to test the international feasibility of the generic instrument. After an additional round of revisions, a translation tool (the “language management utility”, LMU) was developed to enforce the comparability of all national translations with the generic English-language questionnaire. This tool is also used to keep track of necessary further adaptations to each country’s institutions and circumstances.

The translation tool and the translated questionnaires were tested in two countries, Germany and Italy. These test runs resulted in another round of improvements of tools and instrument, before the English version was translated in all SHARE languages. Languages include language variants which are treated separately, such as Belgian French and Swiss German.

The second stage culminated in a first *pilot* simultaneously in all SHARE countries, using quota samples ($n = 50$ households, some 75 persons) similar to the UK-Pilot in stage one. These interviews were conducted in June 2003 and aimed at testing whether the questions are understood and answered as intended in each country, along with measuring the duration of the different modules. Overall, this June Pilot was a success as it showed the general feasibility of such an ambitious survey, in all of the SHARE countries. More refinement will be done based on the experiences drawn, to be prepared for a pure random sample.

Stage 3: Development of multi-language sampling frame

In the next stage, after further refinements of the instrument, the full questionnaire using random samples ($n = 100$ primary respondents per country plus their spouses) will be fielded in January/February 2003. Aim is to allow predictions to be made of the reliability and validity of the full questionnaire, including more “problematic” respondents than are to be expected using a quota sample. In addition, this *pre-test* should also test the country-specific procedures to achieve a probability sample, as well as a sample management system.

An extensive statistical analysis of the pilot results will be performed to assess the reliability and validity of the questions. Using data from the testing interviews, the pilot

results and past data, these will suggest improvements to questions, and assist in the design of the final questionnaire.

Stage 4: Running the prototype “main test survey”

The last stage will consist of a medium-scale survey of this final questionnaire (n = 1000-1500 primary respondents per country plus their spouses, totalling some 25.000 respondents), scheduled for Spring/Summer 2004. This stage will be the essential step to demonstrate the feasibility and the usefulness of SHARE, in that it permits substantive data analysis addressing the main questions of interest. This “main test survey” will deliver a prototype for the planned multi-year panel, and should serve as a demonstration object to the European Commission.

Contact Addresses

*Prof. Axel Börsch-Supan, Ph.D.
Mannheim Research Institute for
the Economics of Aging (MEA)
University of Mannheim
L 13, 17
D-68131 Mannheim, Germany
email: axel@boersch-supan.de*

*Dr. Hendrik Jürges
Mannheim Research Institute for
the Economics of Aging (MEA)
University of Mannheim
L 13, 17
D-68131 Mannheim, Germany
email: juerges@mea.uni-mannheim.de*

*Dr.-Ing. Oliver Lipps
Mannheim Research Institute for
the Economics of Aging (MEA)
University of Mannheim
L 13, 17
D-68131 Mannheim, Germany
email: lipps@mea.uni-mannheim.de*

References

Aarts, L.J.M./R.V. Burkhauser, Ph.R. de Jong (eds.), 1996: *Curing the Dutch Disease, An International Perspective on Disability Policy Reform*. Aldershot: Avebury.

Börsch-Supan, A. (ed.), 2003: *Life-Cycle Savings and Public Policy*. Academic Press.

European Commission, 2000: *The Contribution of Public Finances to Growth and Employment: Improving Quality and Sustainability*. Communication from the European Commission to the Council and the European Parliament.

European Commission, 2003: *Joint Commission/Council Report on adequate and sustainable pensions*.

Gruber, J./Wise, D.A. (eds.), 1999: *Social Security and Retirement around the World*. Chicago University Press.

Lipps, O., 2002: *Guidelines for survey agency selection*. Unpublished manuscript. Mannheim.

SRC, 2003: *Interviewer Project Manual for the Survey of Health, Ageing, and Retirement in Europe*. Unpublished Interviewer manual prepared by the staff of the Survey Research Center at for the SHARE June Pilot. University of Michigan's Institute for Survey Research.

Suzman, R.M./ Willis, D.P./ Manton, K.G. (eds.), 1992: *The Oldest Old*. Oxford: Oxford University Press.

United Nations, 2002: *Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat. World Population Prospects*. <http://esa.un.org/unpp>

U.S. National Academy of Sciences (USA), 2001: *A Research Agenda and New Data for An Aging World*, http://www7.nationalacademies.org/cpop/New_data_aging_world.html

„STELLUNG IM BERUF“ ALS ERSATZ FÜR EINE BERUFSSKLASSIFIKATION ZUR ERMITTLUNG VON SOZIALEM PRESTIGE

JÜRGEN H.P. HOFFMEYER-ZLOTNIK

1. Vorbemerkung

Das vorzustellende Instrument dient der Bestimmung des sozialen Prestiges über die in den Demographischen Standards (StaBA 1999) enthaltene Variable „Stellung im Beruf“ (vgl. Hoffmeyer-Zlotnik 1993, 1998). Die Klassifikation des sozialen Prestiges geschieht in fünf Kategorien. Sofern eine Klassifikation von sozialem Prestige in groben Kategorien zur Bestimmung des sozio-ökonomischen Status ausreicht, kann dieses Instrument ein aufwendiges Erfassen von Beruf oder beruflicher Tätigkeit ersetzen. Denn eine Vielzahl von Studien erfasst den Beruf offen, klassifiziert die Antworten nach der „International Standard Classification of Occupations“ oder nach der „Klassifikation der Berufe“ des Statistischen Bundesamtes (1992) und überträgt die Berufsklassifikationen in einen Prestige-, Status- oder Klassenindex (siehe Hoffmeyer-Zlotnik/Geis 2003). Zur Bestimmung des sozio-ökonomischen Status würde es aber reichen, neben Bildung und Einkommen eine ebenso grobe Variable zum beruflichen Prestige zur Verfügung zu haben. In diesen Fällen könnte der hier vorzustellende Index die aufwendige und teure Erfassung und Kodierung offener Berufsangaben ersetzen.

2. Die Ausgangssituation zur Bestimmung von sozio-ökonomischem Status

Zur empirischen Messung des sozio-ökonomischen Status werden in der Regel drei Indikatoren genutzt: „Bildung“, „Beruf“ und „Einkommen“. Hierbei wird „Bildung“ in der Regel über den höchsten allgemeinbildenden Schulabschluss erfasst, „Beruf“ über die berufliche Tätigkeit, die Stellung im Beruf, die Branche oder über berufliches Prestige bzw. beruflichen Status; „Einkommen“ wird entweder monetär erfasst, bezogen auf die

Zielperson oder einen Haushalt, in Bezug auf alle Verdienere oder alle Haushaltsmitglieder, oder alternativ über den Besitz langfristiger Konsumgüter.

Eine Alternative zu einem Index, basierend auf „Bildung“, „Beruf“ und „Einkommen“, stellen Indices dar, die allein auf dem Beruf aufbauen und hierüber „Prestige“, „Status“ oder „Klassenlage“ ausweisen, dabei aber jeweils implizit über den Beruf auf weitere Variablen rückschließen. Denn „Bildung“ und „Einkommen“ bereiten Probleme bei der Einordnung. Seit über Schulreformen, die zunächst höhere Bildung für alle Bevölkerungsgruppen zugänglich machen und später den Basisabschluss anheben wollten, das Bildungsniveau sich verschoben hat und sich das Gymnasium von der „Eliteschule“ zu einer „Volksschule“ entwickelt hat, hat sich das Verhältnis von „Bildung“ zu „Beruf“ nachhaltig verändert. Über die Bildungsexpansion und ein Überangebot an Ausbildungsnachfragern, bei einer sinkenden Anzahl von Ausbildungs- und Arbeitsplätzen, ist höhere Bildung heute weniger als Status-Plus zu sehen, sondern bietet lediglich eine bessere Ausgangssituation im Verteilungskampf um die knapper werdenden Güter „Ausbildung“ und „Arbeit“. Ebenfalls finden für die breite Masse der Erwerbstätigen, bedingt durch eine lang anhaltende Stagnation bei den Lohn- und Gehaltssteigerungen, reale Einkommensverluste statt. Steigende Bildungsabschlüsse und stagnierende bis sinkende Einkommen dokumentieren in einem Index zum sozio-ökonomischen Status, je nach der Gewichtung der Variablen, eine positive oder negative soziale Mobilität, die wahrscheinlich in beiden Fällen nur schwer zu interpretieren ist. Hinzu kommen bei der Einkommensvariable Messprobleme, denn die Einkommensbezieher benennen in allgemeinen Bevölkerungsumfragen ihr Einkommen nur in Teilmengen (vgl. Hoffmeyer-Zlotnik/ Warner 1998), in der Regel bestehend aus der Haupteinkommensquelle (Arbeitseinkommen eines Durchschnittsmonats, Rente, Pension oder Stipendium). Andere Einkommen (wie 13. Monatsgehalt, Steuerrückzahlungen, staatliche Transferleistungen, Einkommen aus Vermögen oder aus Verpachtung und Vermietung oder gar Deputate oder private Transferleistungen) werden, oft unbewusst, unterschlagen. Damit erfährt die Einkommensangabe des Individualeinkommens in der Regel eine Unterschätzung von einem Drittel, die des Haushaltseinkommens, je nach Nähe des Schätzenden zum Hauptverdiener, um bis zu weiteren 20%. Dieses lässt es geraten erscheinen, „Einkommen“ als Einzelvariable bei der Bestimmung des sozio-ökonomischen Status sehr grob zu klassifizieren und nicht zu hoch zu bewerten.

Der „Beruf“ als zentrales sozialstatistisches Hintergrundmerkmal basiert auf der impliziten Annahme, die Verortung einer Person im sozialen Gefüge einer Gesellschaft sei vor allem durch die in einer arbeitsteiligen Gesellschaft spezifische Erwerbstätigkeit gegeben (Mayer 1979: 81). Bildung und Einkommen stehen mit der Variable der beruflichen Position in engem Zusammenhang. „Beruf“ als Einzelvariable enthält quasi „Bildung“ und

„Einkommen“, denn jeder Beruf ist über die diesen voraussetzende Ausbildung an ein bestimmtes Bildungsniveau geknüpft und jeder Beruf ist mit einer durchschnittlichen Höhe an Erwerbseinkommen verbunden, das zwar eine große Spannweite aufweist, aber dennoch als Richtwert dienen kann. Unter dieser Voraussetzung macht die Variable „Beruf“ als die zentrale Variable bei der Ermittlung des sozio-ökonomischen Status Sinn.

Aber auch die Definition von „Beruf“ ist nicht unproblematisch. Es gibt einerseits Unterschiede von Kultur zu Kultur, bedingt durch eine unterschiedliche Organisation der Arbeit, wodurch „nominell identische Berufe nicht notwendig die selben Tätigkeiten umfassen“, und andererseits können „Berufe .. auch immer spezifischer definiert werden“ (vgl. Treiman 1979: 126f). Daher ist es bei der Erfassung der Berufe wichtig, eine Skala zu nutzen, die die nationalen Berufsstatushierarchien angemessen repräsentiert. Solch eine Standardklassifikation der Berufe liefert zum Beispiel die „International Standard Classification of Occupations“ (ISCO). Eine nationale Alternative stellt die „Klassifikation der Berufe“ des Statistischen Bundesamtes (1992) dar, die zwar für die Bestimmung von Prestige und Status weniger gut geeignet ist als die ISCO-Klassifikation, die aber dennoch eine gute Ausgangsposition für berufsclassifizierende Indices darstellt.

3. Klassifikation der Berufe und darauf aufsetzende Prestige-, Status- und Klassen-Indices

Während die Klassifikation des Statistischen Bundesamtes in der untersten Stufe einen Katalog von nahezu 30.000 Berufsbezeichnungen darstellt, ist die International Standard Classification of Occupations (ISCO) stärker strukturell gegliedert. Die Gliederung der ISCO-68 geht über eine sektorale Gliederung nach Beschäftigtensektoren in zwei Schritten zu einer fach- oder branchenbezogenen Gliederung, die schließlich in der vierten und fünften Stelle durch eine Klassifikation der Tätigkeit präzisiert wird (ILO 1968). Die ISCO-88 (siehe: ILO 1990), als Revision der ISCO-68, hat als neue Bestimmungsfaktoren zur Konkretisierung der Gleichheit bzw. der Verwandtschaft von beruflichen Tätigkeiten die Merkmale „skill level“ und „skill specialisation“ eingeführt (vgl. Statistisches Bundesamt 1992: 12). „Skill level“ stellt hierbei im Sinne eines Anforderungsniveaus ein berufs- und arbeitsplatzbezogenes Merkmal dar. Das „skill level“-Konzept unterscheidet, in Anlehnung an die „International Standard Classification of Education (ISCED)“, nach vier relativ groben Kategorien von Bildungsqualifikation bzw. von Anforderungen. Damit wird die berufliche Tätigkeit mit einem bestimmten Kenntnis- und Fertigkeiteniveau verbunden (vgl. ILO 1990: 2-3). „Skill specialisation“ ist ein berufsfachliches Kriterium, welches einerseits auf die Art der ausgeübten Tätigkeit innerhalb einer gegebenen Qualifikationsebene abhebt und andererseits versucht, Merkmale wie verarbeitetes Material,

Berufsmilieu, Art der produzierten Güter und Dienstleistungen, Besonderheiten der Fertigungsverfahren etc. in die Berufsklassifikation mit einzubeziehen.

Im Gegensatz zur ISCO-68 geht die ISCO-88 nicht auf die Ebene der „beruflichen Tätigkeit“ („occupations“) hinunter. Auch wenn in der Regel mehr als eine „berufliche Tätigkeit“ einer „Berufsgruppe“ („unit group“) zuzuordnen sind, so erscheint die Ebene der „unit groups“ den Autoren der ISCO-88 (ILO 1990: 4) die sinnvollere und aussagekräftigere Ebene, denn: auf unterschiedlichen nationalen Ebenen müssen die „occupations“ voneinander abweichen „depend on the size of the economy and the level of economic development, the level and type of technology, work organisation and historical circumstances“ (ILO 1990: 4). Durch das Abrücken von der eine Tätigkeit klassifizierenden Ebene der „occupations“ in der ISCO-68 und einem Hinwenden zu einer auf das Anforderungsniveau für Beruf abgestellten Ebene von „skill specialisation“ in der ISCO-88, dürfte jede nationale Gesellschaft, egal welcher Entwicklungsstand und/oder welche Auffassung von Arbeit vorherrscht, die ISCO-88 problemlos anwenden können. Und die ISCO-88 ist eine ideale Basis für Prestige, Staus und Klassenlage.

Als Instrument der Standardisierung von gesellschaftlichen Hierarchien wurde ab Mitte der 70-er Jahre eine standardisierte berufliche Prestigeskala (SIOPS) entwickelt (Treiman 1975, 1977), die auf den „unit groups“ der ISCO-68 aufbaute. In den 90-er Jahren wurde die internationale Berufsprestige-Skala von Treiman auf die ISCO-88-Codierung umgestellt (vgl. Hoffmeyer-Zlotnik/Geis 2003). Soweit möglich, wurden auf der Basis der 390 „unit-groups“ die vorhandenen eins-zu-eins-Übertragungen von ISCO-68 zu ISCO-88 aus dem Handbuch übernommen. Dort, wo eine eins-zu-eins-Übertragung nicht möglich war, wurde ein mittlerer Wert aller zu einer „unit-group“ zählenden Berufe ermittelt und den ISCO-88-Werten in gleicher Logik zugeordnet, wie dieses zuvor schon mit ISCO-68 geschehen war (Ganzeboom et al. 1996).

Neben der Prestige-Skala von Treiman ist 1992 eine weitere, international messende Bewertungs-Skala von gesellschaftlicher Ordnungsstruktur publiziert worden: der „Standard International Socio-Economic Index of Occupational Status“ (ISEI-Skala) von Ganzeboom, de Graaf, Treiman und de Leeuw (1992). Hierbei geht es nicht um berufliches Prestige, sondern um den sozio-ökonomischen Status. Die Überlegung zu dieser Skala geht davon aus, dass jede berufliche Tätigkeit einen bestimmten Bildungsgrad erfordert und durch ein bestimmtes Lohnniveau entlohnt wird.

Als dritter Typ ist der Index der „Klassenlage“ von Erikson, Goldthorpe und Portocarero (EGP) zu sehen. Die „Klassenlage“ ist im Gegensatz zu Prestige und sozio-ökonomischem Status kein hierarchisches sondern ein kategoriales Differenzierungsschema, das in die drei Obergruppen „Arbeitgeber“, „Selbständige“ und „Arbeitnehmer“ unterschei-

det. Die „Klassenlage“ verbindet Informationen über Beschäftigten-Status und berufliche Tätigkeit. Die Konzeption der Klassenlage ist als Resultat von Marktlage und Arbeitssituation, in der sich eine beschäftigte Person befindet, zu sehen (Erikson/Goldthorpe/Portocarero 1979; Erikson/Goldthorpe 1992).

Einen Vergleich der unterschiedlichen Prestige- und Status-Skalen bietet Wolf (1995; 1998), nähere Einzelheiten zu den drei angesprochenen Skalen beschreiben Ganzeboom und Treiman (2003) und in einem kurzen Überblick Hoffmeyer-Zlotnik/Geis (2003).

4. Die Variable „Stellung im Beruf“

Die ISCO-basierten Skalen zu Prestige, Status oder Klassenlage erfordern, wie oben dargestellt, eine offene Erfassung und aufwendige Codierung der Berufe. Da die Berufsklassifikation nach ISCO bis heute nur zu einem Teil computerunterstützt ablaufen kann, der andere Teil aber immer noch durch trainierte Spezialisten erfolgen muss, erfordert diese hohe Ressourcen an Zeit und Geld. Will man mit den Skalen für Prestige, Status oder Klassenlage arbeiten, so ist die ISCO-Codierung die Voraussetzung. Erst auf der ISCO-Verdodung aufbauend, lassen sich die gewünschten Skalen maschinell umsetzen. Deutlich weniger aufwendig ist die Erhebung der „Stellung im Beruf“ und eine hierauf basierte Prestige-Skala.

Das Erfassen der „Stellung im Beruf“ erfolgt über eine Liste der Klassifikation der beruflichen Stellung nach der Mikrozensus-Zusatzbefragung von 1971 (Tabelle 1), jene sehr detailliert abfragende Liste, die das Zentrum für Umfragen, Methoden und Analysen (ZUMA) seit 1976 als unverzichtbaren Bestandteil der Standarddemographie betrachtet (vgl. Pappi 1979: 280f) und die hierüber Bestandteil der „Demographischen Standards“ von ASI, ADM und Statistischem Bundesamt geworden ist (siehe: Statistisches Bundesamt 1999). Was diese Klassifikation der „beruflichen Stellung“ auszeichnet ist, dass über die zur Statusbestimmung irrelevanten versicherungstechnischen Kategorien („Selbständiger“, „Beamter“, „Angestellter“, „Arbeiter“) hinausgehend, pro Gruppe eine Differenzierung vorgenommen wird, bei den Selbständigen nach der Betriebsgröße und bei den abhängig Beschäftigten nach unterschiedlichen Tätigkeitsmerkmalen.

Die Selbständigen, unterteilt in die Gruppen „selbständige Landwirte“, „akademische freie Berufe“ und „Selbständige in Handel, Gewerbe, Industrie, Dienstleistungen“ werden untergliedert nach der Größe des Betriebes. Hierbei werden die landwirtschaftlichen Betriebe nach der Hektar-Größe der Nutzfläche, die beiden anderen Gruppen nach der Anzahl der Mitarbeiter gegliedert. Die Beamten werden nach dem Dienstrecht (einfacher,

Tabelle 1: Klassifikation der beruflichen Stellung

<p>Selbständige Landwirte mit landwirtschaftlich genutzter Fläche von...</p> <p>10: unter 10 ha 11: 10 ha bis unter 20 ha 12: 20 ha bis unter 50 ha 13: 50 ha und mehr</p> <p>Akademische freie Berufe (z.B. Arzt, Anwalt mit eigener Praxis)</p> <p>15: 1 Mitarbeiter oder allein 16: 2 bis 9 Mitarbeiter 17: 10 Mitarbeiter und mehr</p> <p>Selbständige in Handel, Gewerbe, Industrie, Dienstleistungen, u.a.</p> <p>21: 1 Mitarbeiter oder allein 22: 2 bis 9 Mitarbeiter 23: 10 bis 49 Mitarbeiter 24: 50 Mitarbeiter und mehr</p> <p>Beamte/Richter/Berufssoldaten</p> <p>40: Beamte im einfachen Dienst (bis einschließlich Oberamtsmeister) 41: Beamte im mittleren Dienst (vom Assistenten bis einschl. Hauptsekretär/Amtsinspektor) 42: Beamte im gehobenen Dienst (vom Inspektor bis einschl. Oberamtsmann/Oberamtsrat) 43: Beamte im höheren Dienst, Richter (vom Regierungsrat aufwärts)</p>	<p>Angestellte</p> <p>50: Industrie- und Werkmeister im Angestelltenverhältnis 51: Angestellte mit einfacher Tätigkeit (z.B. Verkäufer, Kontorist, Stenotypistin) 52: Angestellte, die schwierige Aufgaben nach allgemeiner Anweisung selbständig erledigen (z.B. Sachbearbeiter, Buchhalter, technischer Zeichner) 53: Angestellte, die selbständige Leistungen in verantwortungsvoller Tätigkeit erbringen oder begrenzte Verantwortung für die Tätigkeit anderer tragen (z.B. wiss. Mitarbeiter, Prokurist, Abteilungsleiter) 54: Angestellte mit umfassenden Führungsaufgaben und Entscheidungsbefugnissen (z.B. Direktor, Geschäftsführer, Vorstand größerer Betriebe und Verbände)</p> <p>Arbeiter</p> <p>60: ungelernete Arbeiter 61: angelernte Arbeiter 62: gelernte und Facharbeiter 63: Vorarbeiter und Kolonnenführer 64: Meister/Poliere</p> <p>Mithelfende Familienangehörige</p> <p>30: Mithelfende Familienangehörige</p>
--	--

mittlerer, gehobener, höherer Dienst) unterschieden. Über die dienstrechtliche Gliederung wird auch die Tätigkeit und die damit verbundene Autonomie des Handelns definiert. Die Arbeiter werden entsprechend ihrer Ausbildung und damit nach einer Hierarchie ihrer Einsetzbarkeit und Verantwortlichkeit für die Tätigkeit in „Ungelernte“, „Angelernte“, „Facharbeiter“ und „Meister/Poliere“ gruppiert. Die Angestellten werden nach der Differenziertheit der Tätigkeit und nach einer damit verbundenen Verantwortung für das eigene

Handeln klassifiziert. Hierbei reicht die Handlungsautonomie von der „einfachen Tätigkeit“ über „schwierige Aufgaben nach allgemeiner Anweisung selbständig erledigen“ und „selbständige Leistungen in verantwortungsvoller Tätigkeit erbringen oder begrenzte Verantwortung für die Tätigkeit anderer tragen“ bis zu diejenigen, die „umfassende Führungsaufgaben und Entscheidungsbefugnisse“ innehaben.

5. Die Skala „Autonomie des Handelns“

Für jede der drei Gruppen der abhängig Beschäftigten (Beamte, Arbeiter, Angestellte) ergibt sich eine hierarchische Untergliederung, die in eine Skalierung nach unterschiedlichen Stufen der „Autonomie des Handelns“ mündet. Vereint man die drei Einzelskalen zu einer übergeordneten Skala nach der Differenziertheit des Handelns und der damit verbundenen eigenen Verantwortung, so erhält man eine gemeinsame Skala der „Autonomie des Handelns“, beginnend mit niedriger Handlungsautonomie bei einfacher Tätigkeit, bis hin zu hoher Handlungsautonomie bei umfassenden Führungsaufgaben.

In einem zweiten Schritt ist diese Skala der „Autonomie des Handelns“ der abhängig Beschäftigten zu erweitern um die Gruppen der Selbständigen, deren Handlungsautonomie sich in Produktion und Dienstleistung sowie im akademisch freien Bereich in der Regel aus der Größe des Betriebes ergibt. Und nicht nur bei den „akademischen freien Berufen“ und bei den „Selbständigen in Handel, Gewerbe, Industrie, Dienstleistungen“ ist die Betriebsgröße das entscheidende Kriterium für die „Autonomie des Handelns“, sondern auch bei den Landwirten, obwohl hier nicht die Anzahl der Mitarbeiter ausschlaggebend ist. Bei den Landwirten ist ein weiteres zentrales Kriterium der Grad der Spezialisierung, was in der Abfrage jedoch nicht erfasst wird. Damit wird es schwer, die Landwirte exakt zu klassifizieren.

In einem dritten Schritt werden die „mithelfenden Familienangehörigen“ mit einem mittleren Wert in die Skala eingefügt. Das Problem bei dieser Gruppe ist einerseits die sehr hohe Heterogenität der Tätigkeit, aber andererseits deren Verantwortung für die Tätigkeit, da sie als Familienangehörige des Unternehmers an dessen Risiko teilhaben.

Tabelle 2 zeigt das Resultat der Einordnung der drei Gruppen der abhängig Beschäftigten, der Selbständigen und der mithelfenden Familienangehörigen in eine Skala der „Autonomie des Handelns“. Hier fallen bei den abhängig Beschäftigten die un- und angelernten Arbeiter (Code: 60, 61) in die unterste Kategorie, Personen, die eine Ausbildung haben und einfache Tätigkeiten verrichten (Code: 40, 51, 62), in Kategorie 2, Personen, die schwierige Aufgaben nach allgemeiner Anweisung selbständig erledigen (Code: 41, 52, 63), in Kategorie 3, Personen, die selbständige Leistungen in verantwortungsvoller Tätig-

keit erbringen oder begrenzte Verantwortung für die Tätigkeit anderer tragen (Code: 42, 50, 53, 64), in Kategorie 4 und Personen mit umfassenden Führungsaufgaben und Entscheidungsbefugnissen (Code: 43, 54), in Kategorie 5. Die Landwirte verteilen sich je nach Hofgröße über die Kategorien 2 bis 4, Selbständige und Freiberufler beginnen bei der mittleren Autonomie, in der Kategorie 3. Die mithelfenden Familienangehörigen werden als Gruppe geschlossen der mittleren Kategorie zugeordnet.

Tabelle 2 stellt damit gleichzeitig die Recodieranweisung zur Erstellung der Variable „Autonomie des Handelns“ bzw. „berufliches Prestige“ dar.

Tabelle 2: „Stellung im Beruf“ nach „Autonomie des Handelns“

Wert „Autonomie des Handelns“	Code „Stellung im Beruf“		
1 – niedrig	60, 61		
2	40, 51, 62	10, 11	
3	41, 52, 63	12, 15, 21	30
4	42, 50, 53, 64	13, 16, 22	
5 – hoch	43, 54	17, 23, 24	

In der Anwendung empfiehlt es sich, die aktuelle oder die letzte Erwerbstätigkeit zu erfassen, soweit diese für den sozio-ökonomischen Status (SES) ausschlaggebend ist bzw. war. Ein Problem stellt die Tatsache dar, dass in einem Haushalt bei mehreren Erwerbstätigen oft Tätigkeiten mit unterschiedlichen Prestigewerten ausgeübt werden. Hier sollte man, sofern Daten über alle Erwerbstätigen vorliegen, die Person mit dem höchsten Prestigewert als die den sozio-ökonomischen Status des Haushalts bestimmende betrachten. Weder ein Addieren von zwei Werten noch ein Mitteln, auch bei Personen, die zwei Tätigkeiten gleichzeitig ausüben, ist möglich oder sinnvoll. In jedem Fall ist der den Lebensstil beeinflussende Wert der höhere von zwei möglichen.

6. Überprüfen der Güte der „Autonomie des Handelns“

Zur Überprüfung der Güte des Index „Autonomie des Handelns“ wird die ISCO-basierte Prestige-Skala von Treiman herangezogen. Erste Voraussetzung für eine Reliabilitätsprüfung der „Autonomie des Handelns“ im Vergleich zur Treiman-Prestige-Skala ist eine Transformation der Treiman-Skala auf eine 5-er Skala von „1 = niedrig“ bis „5 = hoch“, damit beide Skalen den gleichen Wertebereich abbilden. Diese Transformation der Trei-

man-Skala, die in der Bundesrepublik Deutschland bei „14 = niedrig“ beginnt und bei „78 = hoch“ endet, hat nach inhaltlich sinnvollen „Schnitten“ zu erfolgen (siehe Tabelle 3).

Tabelle 3: Transformation des Treiman-Prestige-Skala in eine 5-Punkte-Skala

Wert „Prestige“	Wertebereich „Treiman-Prestige-Score“ (anhand der Daten des Allbus 84 überprüft)
1 – niedrig	14 bis 32
2	33 bis 41
3	42 bis 50
4	51 bis 63
5 – hoch	64 bis 78

Die Schnittstellen sind nach Plausibilität festgesetzt worden: Gruppe 1 mit dem Wertebereich 14 bis 32 bezeichnet vor allem niedrig bewertete manuelle Tätigkeiten mit sehr geringer Handlungsautonomie. Gruppe 2 mit den Wertebereichen 33 bis 41 umfasst abhängige Beschäftigung mit geringer Handlungsautonomie, vor allem in der Produktion, aber auch im Dienstleistungsbereich, sowie das ganze traditionelle Handwerk von der Nahrungsmittelproduktion bis zur Stoff-, Holz- und Metallverarbeitung. Gruppe 3 mit dem Wertebereich 42 bis 50 umfasst Tätigkeiten, die einen mittleren Ausbildungsabschluss voraussetzen und die in einem beschränkten Maße verantwortungsvolles Handeln erfordern, wie „Hebamme“, „medizinisch-technische Assistentin“, „Techniker“, „Mechaniker“, „Supervisor in der Produktion“ oder „Makler“. Gruppe 4 mit dem Wertebereich 51 bis 63 umfasst Tätigkeiten in abhängiger oder selbständiger Position, die eine Fachhoch- oder Hochschulbildung voraussetzen, die aber noch kein hohes Prestige aufweisen. Ab dem Wert 64 beginnen die freiberuflichen Akademiker und die leitenden Tätigkeiten der Gruppe 5, die ein hohes Ansehen wie z.B. „Arzt“, „Apotheker“, „Jurist“ und/oder als Personen in leitender Funktion einen größeren Büroraum haben wie z.B. „Produktionsleiter“.

Der für den Vergleich ausgewählte Datensatz ist der ALLBUS 1984, ein Datensatz, der für die Grundgesamtheit der wahlberechtigten Bevölkerung in Privathaushalten der Bundesrepublik Deutschland (zum Stand von 1984) steht. Dieser Datensatz wurde gewählt, da er nicht nur eine ISCO-Verdodung für die berufliche Tätigkeit der Befragten aufweist, sondern die Berufsangaben auch in Verschriftung vorliegen. Lässt man in diesem Datensatz sämtliche Fehler der Zuordnung von Befragten zu Kategorien zunächst unberücksichtigt,

dann ergibt sich zwischen „Berufs-Prestige“ und „Autonomie des Handelns“ eine Korrelation von (Pearson's) $r = .5949$.

In einem weiteren Schritt müssen die Zuordnungsfehler von „beruflicher Tätigkeit“ zu „Stellung im Beruf“ im Datensatz bereinigt werden. Solcherart Zuordnungsfehler gibt es in jedem Datensatz. Sie basieren in der Regel auf einem Missverstehen der Aufgabenstellung durch die Befragungsperson, auf kognitiven Prozessen, die zu einer verzerrten Wahrnehmung führen, oder auf einer subjektiven Interpretation von Sachverhalten. Vor der Fehlerbereinigung steht allerdings die Definition eines Fehlers. Hierzu wurde zu jeder Kategorie der „Stellung im Beruf“ der größtmögliche Wertebereich abgegrenzt, den diese Kategorie im Bereich des „Berufsprestige“ abdecken kann (Tabelle 4). Z.B. kann die (Ober-) Kategorie „Landwirt“ den Wertebereich von 14: „land- und tierwirtschaftliche Arbeitskraft“ bis 55: „Landwirte im Spezialbetrieb“ abdecken; unterhalb des Code 14 existieren für das Spektrum der beruflichen Tätigkeit in der Bundesrepublik keine Codes und oberhalb des Code 55 gibt es keine Prestigewerte mehr für die Gruppe der „Landwirte“.

Ein Blick in die konkreten Daten (des ALLBUS 1984) zeigt, dass für die meisten Kategorien der „Stellung im Beruf“ nicht der volle Wertebereich des „Berufsprestige“ gelten kann. Im Vergleich beider Variablen wird deutlich, dass pro Kategorie der Variablen „Stellung im Beruf“ nur ein eingeschränkter Wertebereich auf der „Prestige“-Skala betrachtet werden darf (siehe Tabelle 4). Beispiele einer falschen Selbstzuordnung sind: „Gymnasiallehrer“, die sich den „Stellung“-Kategorien 40 und 41 zuordnen, „Soldat“ unter Kategorie 43 zugeordnet, „technischer Zeichner“ unter Kategorie 51 oder unter Kategorie 61 zugeordnet, „Buchhalter“ und „Sekretärin“ unter Kategorie 54 sowie „Bauarbeiter“ unter Kategorie 64 zugeordnet.

Eine weitere Fehlerquelle liegt im internationalen Anspruch, den die Treiman-Prestige-Skala erhebt. Über die internationale Verortung geschehen Zuordnungen, die so für die Kultur der Bundesrepublik Deutschland nicht haltbar sind: Im Bereich der Selbständigen in Handel, Gewerbe und Dienstleistung treten die sehr niedrig bewerteten Berufe „Verkaufshilfskraft“, „Wäscher“, „Schuhinstandsetzer“, „Werkzeugschärfer“ und „Motorfahrzeugfahrer“ auf, alles Tätigkeiten, die in einem Entwicklungsland einen anderen Stellenwert und ein anderes Prestige haben als in Mitteleuropa, aber mit der identischen Skala in den unterschiedlichen gesellschaftlichen Systemen verortet werden.

Nach der so vorgenommenen Edition der Daten ist die Korrelation zwischen den entsprechend Tabelle 3 und Tabelle 4 recodierten Variablen noch einmal berechnet worden. Die Korrelation zwischen „Autonomie des Handelns“ und bereinigtem „Berufs-Prestige“ beträgt nun (Pearson's) $r = .7860$.

Tabelle 4: „Autonomie des Handelns“ und „Prestige-Bereiche“; die sich gegenseitig ausschließen

Gruppen nach „Stellung im Beruf“	Bereich möglichen Berufs-Prestiges Prestige-Werte zwischen
10, 11, 12, 13	14 bis 55
15, 16, 17	33 bis 78
21	20 bis 78
22, 23	33 bis 78
24	51 bis 78
30	33 bis 78
40, 41	14 bis 50
42, 43	33 bis 78
50	33 bis 78
51	14 bis 41
52	14 bis 50
53	42 bis 78
54	51 bis 78
60, 61	14 bis 41
62	14 bis 50
63	33 bis 78
64	42 bis 78

Wolf (1995; 1998) hat einen weit umfangreicheren Vergleich über eine größere Auswahl sowohl ISCO-basierter als auch Stellung-im-Beruf-basierter Indices vorgenommen, die die Aussagekraft der hier vorgestellten Skala zur „Autonomie des Handelns“ eindrucksvoll bestätigen.

7. Fazit

Die Variable der „Stellung im Beruf“ mit den der Mikrozensus-Zusatzerhebung von 1971 entlehnten 26 Kategorien, recodiert zur „Autonomie des Handelns“, stellt einen vergleichbar guten Schicht-Indikator als Prestige-Index oder eine Variable für den sozio-ökonomischen Status (SES) dar. Die Übereinstimmung von „Prestige“ und „Autonomie“ von $r = .79$ spricht für sich. Die Abfrage der „Stellung im Beruf“ ist, im Gegensatz zur beruflichen Tätigkeit, wenig aufwendig, die Vercodung geschieht im Feld und eine Recodierung zum Index „Autonomie des Handelns“ besteht aus fünf Zeilen. Die Fehleranfäll-

lichkeit im Feld, bedingt durch ein falsches sich Zuordnen von Befragungspersonen, lässt sich über ein entsprechendes Listenlayout oder über eine zweistufige Abfrage reduzieren. Die zweistufige Abfrage erfragt zunächst die Oberkategorien und legt in einem zweiten Schritt, pro Oberkategorie, jeweils eine separate Liste oder Karte für die der gewählten Oberkategorie entsprechenden Unterkategorien (der Handlungsautonomie) vor. Die aufwendige Abfrage und teure Vercodung der „beruflichen Tätigkeit“ lässt sich mit der „Autonomie des Handelns“ im Regelfall vermeiden, sofern es nur um einen Schicht-Index geht. Ein Erfassen und Vercoden der „beruflichen Tätigkeit“ könnte damit auf spezielle Fragestellungen beschränkt bleiben, die mehr als eine grobe Verortung von Befragungspersonen in der gesellschaftlichen Hierarchie erfordern. Die Mehrheit der Studien wird jedoch mit einem Prestige-Index, bzw. einer SES-Variable, auf einer 5-Punkte-Skala messend, wie sie die „Autonomie des Handelns“ darstellt, hinreichend arbeiten können.

Korrespondenzadresse

PD Dr. Jürgen H.P. Hoffmeyer-Zlotnik
ZUMA
Postfach 12 21 55
D – 68072 Mannheim
email: hoffmeyer-zlotnik@zuma-mannheim.de

Literatur

- Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS, 1984: Codebuch (ZA-Nr.1340), herausgegeben vom Zentralarchiv für empirische Sozialforschung. Köln: Zentralarchiv für empirische Sozialforschung.
- Duncan, O.D., 1961: A Socioeconomic Index for all Occupations. S. 109-138 in: Reiss, A.J. jr. (Ed.), Occupations and Social Status. New York: Free Press of Glencoe.
- Erikson, R./Goldthorpe, J.H., 1992: The Constant Flux. A Study of Class Mobility in Industrial Societies. Oxford: Clarendon Press.
- Erikson, R./Goldthorpe, J.H./Portocarero, L., 1979: Intergenerational Class Mobility in Three Western European Societies. British Journal of Sociology 30: 415-441.
- Ganzeboom, H.B.G./de Graaf, P.M./Treiman, D.J./de Leeuw, J., 1992: A Standard International Socio-Economic Index of Occupational Status. Social Science Research 21: 1-56.

Ganzeboom, H.B.G./Treiman, D.J., 2003: Three Internationally Standardised Measures for Comparative Research on Occupational Status. S.159-193 in: Hoffmeyer-Zlotnik, J.H.P./Wolf, Ch. (Eds.), *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables*. New York: Kluwer Academic / Plenum Publishers.

Ganzeboom, H.B.G./Treiman, D.J., 1996: Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research* 25: 201-239.

Hoffmeyer-Zlotnik, J.H.P., 1998: „Beruf“ und „Stellung im Beruf“ als Indikatoren für soziale Schichtung. *RKI-Schriften* 1: 54-64.

Hoffmeyer-Zlotnik, J.H.P., 1993: Operationalisierung von „Beruf“ als zentrale Variable zur Messung von sozio-ökonomischem Status. *ZUMA-Nachrichten* 32: 135-141.

Hoffmeyer-Zlotnik, J.H.P./Geis, A.J., 2003: Berufsklassifikation und Messung des beruflichen Status/Prestige. *ZUMA-Nachrichten* 52: 125-138.

Hoffmeyer-Zlotnik, J.H.P./Warner, U., 1998: Die Messung von Einkommen im nationalen und internationalen Vergleich. *ZUMA-Nachrichten* 42: 30-65.

International Labour Office (Ed.), 1968: *International Standard Classification of Occupations*. Revised Edition 1968. Geneva: ILO.

International Labour Office (Ed.), 1990: *ISCO-88. International Standard Classification of Occupations*. Geneva: ILO.

Mayer, K.U., 1979: Berufliche Tätigkeit, berufliche Stellung und beruflicher Status - empirische Vergleiche zum Klassifikationsproblem. S. 79-123 in: Pappi, F.U. (Hrsg.): *Sozialstrukturanalysen mit Umfragedaten*. Königstein/Ts: Athenäum.

Pappi, F.U. (Hrsg.), 1979: *Sozialstrukturanalysen mit Umfragedaten*. Königstein/Ts: Athenäum.

Statistisches Bundesamt (Hrsg.), 1999: *Demographische Standards. Eine gemeinsame Empfehlung des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (ADM), der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute (ASI) und des Statistischen Bundesamtes*. Ausgabe 1999. Reihe: Methoden – Verfahren – Entwicklungen. Materialien und Berichte. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt (Hrsg.), 1992: *Klassifizierung der Berufe - Systematisches und alphabetisches Verzeichnis der Berufsbenennungen*. Stuttgart: Metzler & Poeschel.

Treiman, D.J., 1979: Probleme der Begriffsbildung und Operationalisierung in der international vergleichenden Mobilitätsforschung. S. 124-167 in: Pappi, F.U. (Hrsg.): *Sozialstrukturanalysen mit Umfragedaten*. Königstein/Ts.: Athenäum.

Treiman, D.J., 1977: Occupational Prestige in Comparative Perspective. New York: Academic Press.

Treiman, D.J., 1975: Problems of Concept and Measurement in the Comparative Study of Occupational Mobility. *Social Science Research* 4: 183-230.

Wolf, Ch., 1998: Zur Messung des sozialen Status in epidemiologischen Studien: Ein Vergleich unterschiedlicher Ansätze. *RKI-Schriften* 1: 75-86.

Wolf, Ch., 1995: Sozio-ökonomischer Status und berufliches Prestige. *ZUMA-Nachrichten* 37: 102-136.

ERGEBNISSE DER ZWEITEN BEFRAGUNG VON NUTZERN DER MIKROZENSUS- SCIENTIFIC USE FILES

PAUL LÜTTINGER, HOLGER BREIHZ¹ & YVONNE LECHERT

1. Einleitung

Im Mai 2003 führten das Statistische Bundesamt und ZUMA gemeinsam die zweite Befragung unter den Nutzern der Mikrozensus Scientific Use Files durch. Bei der ersten Nutzerbefragung vor annähernd fünf Jahren² stand alleine der Mikrozensus von 1995 als Scientific Use File zu reduzierten Kosten für die Wissenschaft zur Verfügung. Mittlerweile, d.h. bei der Befragung im Mai 2003, sind es sieben Mikrozensen, die als Scientific Use File zu einem geringen einheitlichen Grundpreis von 65 € verfügbar sind. Dementsprechend ist die Zahl der Nutzer von Scientific Use Files in den letzten Jahren sprunghaft gestiegen. Rund 130 Institute arbeiten zwischenzeitlich in den verschiedensten Forschungskontexten damit. Über 100 wissenschaftliche Beiträge³ wurden auf der Basis dieser Daten bereits veröffentlicht und viele weitere Beiträge sind in Arbeit. Dementsprechend gewachsen ist auch die Erfahrung der Nutzer im Umgang mit den Daten, so dass eine weitere Umfrage gerechtfertigt erschien. Zudem ist der Start der Service- und Forschungsdatenzentren⁴ ein geeigneter Zeitpunkt, um eine verlässliche Bewertung des

1 Holger Breiholz ist Referatsleiter im Statistischen Bundesamt, Gruppe IXB Mikrozensus, Arbeitskräfteerhebungsprobe, Haushalt und Familie.

2 Vgl. Köhler/Schimml-Neimanns/Schwarz, 2000: Pilotprojekt zur Erleichterung der Nutzungsmöglichkeiten von faktisch anonymisierten Mikrodaten. *Wirtschaft und Statistik* (1): 30-37.

3 Soweit uns bekannt. Die Publikationen sind dokumentiert und unter http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Publikationen/MZ_Literatur.htm abrufbar.

4 Diese Zentren sind das „German Microdata Lab“ („Servicezentrum für Mikrodaten“) der GESIS bei ZUMA, das Internationale Datenservicezentrum am Forschungsinstitut zur Zukunft der Arbeit (IZA, Bonn), und die Forschungsdatenzentren des Statistischen Bundesamtes (Wiesbaden, Bonn), der Statistischen Landesämter und der Bundesanstalt für Arbeit (Nürnberg). Die Einrichtung dieser Zentren geht auf Empfehlungen der Kommission zur Verbesserung der informatio-

gegenwärtigen Daten- und Serviceangebotes zu erhalten, anhand derer die zukünftige Entwicklung zu beurteilen ist. Folgende Bereiche standen im Mittelpunkt der Befragung:

- Scientific Use Files, mit denen gearbeitet wurde
- Kenntnis und Bewertung der Informationsmaterialien des Statistischen Bundesamtes
- Bewertung der Beratung durch das Statistische Bundesamt und ZUMA
- Kenntnis und Bewertung des WWW-Angebotes von ZUMA
- Nützlichkeit der Daten für die Forschungsvorhaben
- Probleme bei der Auswertung der Scientific Use Files
- Interesse an Veranstaltungen (Workshops, Konferenzen)
- Interesse an weiteren Scientific Use Files
- Kommentare und Anregungen

Insgesamt wurden 130 Einrichtungen/Projektleiter im Mai 2003 angeschrieben und gebeten, den Fragebogen zu beantworten. Zudem wurden die Projektleiter gebeten, den Fragebogen auch an andere Personen, die ebenfalls mit den Daten gearbeitet haben, weiterzugeben, d.h. es wurde die Möglichkeit eingeräumt, mehr als einen Antwortbogen zurückzusenden.⁵ 65 Fragebogen gingen letztlich in die Auswertung ein.

2. Ergebnisse

Information und Beratung

Wie in der Einleitung erwähnt ist im Vergleich zur ersten Befragung vor fünf Jahren eine enorme Steigerung der genutzten Files festzustellen. Die 65 Befragten nutzen insgesamt 207 Datensätze, wobei ein deutlicher Schwerpunkt bei den Mikrozensen von 1995-1997⁶ liegt, auf die annähernd 60 Prozent der Nennungen entfallen.

nellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) zurück, die vom BMBF eingesetzt wurde. Neben Eigenleistungen der statistischen Ämter und Institute erfolgt die Finanzierung der Zentren durch das BMBF. Siehe hierzu auch Lüttinger/Schimpl-Neimanns/Wirth/Papastefanou 2003: Mikrodaten (German Microdata Lab): Das Servicezentrum für amtliche Mikrodaten bei ZUMA. ZUMA-Nachrichten 52: 153-172.

- 5 Dies macht die Ermittlung der Responserate sehr schwierig, zumal die rückläufigen Fragebögen auch keine Absenderadresse enthielten. Bei 66 zurückgesendeten Fragebögen dürfte die Responserate zwischen circa 50 Prozent (wenn man die 130 Einrichtungen als Basis nimmt) und 17 Prozent liegen (wenn man davon ausgeht, dass pro Institut im Schnitt etwa drei Personen mit den Daten gearbeitet haben).
- 6 Der Mikrozensus-Scientific Use File 1998 wurde erst im Januar 2003 zugänglich. Die Bereitstellung des Mikrozensus-Scientific Use Files 2000 erfolgte nach Durchführung der Nutzerbefragung im September 2003.

Die aktive Informationspolitik von ZUMA und dem Statistischen Bundesamt, d.h. die kontinuierliche Berichterstattung über diese neuen Entwicklungen z.B. in Publikationen (Wirtschaft und Statistik, ZUMA-Nachrichten) und den WEB-Seiten der beiden Institute führte dazu, dass der überwiegende Teil der Nutzer durch ZUMA (47,4 Prozent) oder das Statistische Bundesamt (35,9 Prozent) auf die Existenz von Scientific Use Files aufmerksam gemacht wurden. Die Information durch Kollegen spielt gegenwärtig noch kaum eine Rolle, was sich aber sicherlich mit zunehmender Verbreitung der Daten zukünftig ändern wird.

Das Statistische Bundesamt stellt zusammen mit den Daten auch zahlreiche Informationsmaterialien bereit. Diese Materialien werden überwiegend positiv beurteilt: Annähernd vier von fünf Antwortenden bewerten sie als sehr hilfreich bzw. hilfreich. Die wenigen Verbesserungsvorschläge dazu beziehen sich vorwiegend auf den Informationsgehalt der Daten und die verwendeten Konzepte: So werden insbesondere genauere Informationen zu den erhobenen Variablen und verwendeten Haushaltskonzepten, sowie zu generierten oder veränderten Variablen gewünscht. Weitere Wünsche betrafen die Bereitstellung von SPSS-Routinen, Syntaxvorlagen bzw. Beispiele für Berechnungen.

Trotz der positiven Bewertung der Informationsmaterialien haben über die Hälfte der Nutzer noch eine zusätzliche Beratung in Anspruch genommen und dies unabhängig davon, wie das Informationsmaterial vom jeweiligen Nutzer bewertet wird. Der Informationsbedarf der Nutzer geht also deutlich über die bereitgestellten Materialien hinaus. Zur Beratung wandten sich 54 Prozent der Nutzer an ZUMA und 42,9 Prozent an das Statistische Bundesamt. Im Vordergrund standen dabei Fragen zur Gewichtung und Hochrechnung der Daten, Probleme bei der Datenaufbereitung und beim Datenmanagement, zu Klassifikationen und der Vergleichbarkeit der Daten.

SPSS ist nach wie vor das mit zwei Dritteln am häufigste genutzte Programm bei der Auswertung der Daten; ein Fünftel nutzt mittlerweile auch Stata während die ebenfalls genannten Programme TDA und SAS nur eine untergeordnete Rolle spielen.

Kenntnis und Bewertung des WWW-Angebotes von ZUMA

Gegenüber der Befragung von 1998 wurde die Bewertung des WWW-Angebotes der Abteilung Mikrodaten⁷ bei ZUMA neu aufgenommen. Dieses Angebot, das in den letzten Jahren erheblich ausgeweitet wurde, kennen 84,6 Prozent (n=55) der Nutzer. Der Service

7 Mit der Etablierung des „Servicezentrums für Mikrodaten“ bei ZUMA wurden die beiden Abteilungen „Mikrodaten“ und „Einkommens- und Verbrauchsstichprobe“ unter dem Namen „German Microdata Lab“ zusammengefasst.

der Abteilung wird zudem sehr positiv beurteilt: Insgesamt 96,3 Prozent beurteilen die WWW Seiten insgesamt als sehr hilfreich (52,7 Prozent) bzw. hilfreich (43,6 Prozent).

Am häufigsten benutzt wurden die Datendokumentationen (22,9 Prozent), die Arbeits- und Methodenberichte (20,5 Prozent), SPSS-Setups (20,5 Prozent) sowie die Informationen zum Datenzugang (17,5 Prozent). Mikrodatentools und FAQ-Liste werden dagegen nicht so häufig von den Nutzern nachgefragt.

Die von ZUMA im WWW bereitgestellten Informationsmaterialien wurden mehrheitlich als „sehr nützlich“ bzw. „nützlich“ bewertet. Insbesondere die SPSS-Setups (68,4 Prozent „sehr nützlich“) aber auch die Informationen zum Datenzugang (100 Prozent „sehr nützlich/nützlich“) oder die Arbeits- und Methodenberichte (91,4 Prozent „sehr nützlich/nützlich“) werden äußerst positiv beurteilt. Von den fünf kritischen Äußerungen, die zu den Materialien gemacht werden, beziehen sich zwei auf die FAQ-Liste („redundant zum Rest-Angebot“) sowie zwei auf die SPSS-Setups (zu lange labels/funktionierte nicht).

Dass dieser Service der Abteilung Mikrodaten bislang die Nutzerbedürfnisse nahezu voll und ganz erfüllt wird auch darin ersichtlich, dass nur fünf Befragte über das bestehende Angebot hinausgehende Wünsche haben, so z.B. SAS-Setups oder Stata-Setups. Dieser Wunsch wurde bereits aufgegriffen, da seit dem Mikrozensus 1998 auch SAS- und Stata-Setups erstellt werden.

Probleme beim Arbeiten mit den Daten und Erreichen des Analyseziels

Probleme beim Arbeiten mit den Scientific Use Files hatte über die Hälfte der Nutzer (57,6 Prozent). Am häufigsten wurde als Ursache „fehlende Informationen zu den Daten“ genannt (38,2 Prozent), gefolgt von „fehlender Praxis“ (26,5 Prozent) und „Handhabbarkeit“ (17,6 Prozent). Während bei der ersten Nutzerbefragung auf die explizite Frage nach technischen Problemen beim Arbeiten mit dem Datensatz noch ein Viertel der Forscher angab diese gehabt zu haben und dabei vor allem lange Rechenlaufzeiten als problematisch bewertet wurden, spielen diese wohl aufgrund der zunehmend besseren Hard- und Softwareausstattung mittlerweile keine Rolle mehr.

Ihr Analyseziel konnten 35,9 Prozent der Nutzer realisieren und 60,9 Prozent gaben an, dies zumindest noch teilweise realisiert zu haben. Nur 3,1 Prozent dagegen konnten ihr Analyseziel nicht erreichen. Die häufigste Ursache für die mangelnde Erreichung des Analyseziels wird, wie schon in der ersten Nutzerbefragung, im Fehlen von Merkmalen gesehen (64,7 Prozent). Dies ist zum einen darin begründet, dass nur die vom Nutzer angeforderten Merkmale und nicht das gesamte File zur Verfügung gestellt werden. Zum anderen werden von den Nutzern Merkmale gefordert (siehe hierzu auch den Punkt „Verbesserungsvorschläge“), die nicht im Erhebungsprogramm des Mikrozensus enthalten

sind. Diese Problematik eines auf spezifische Forschungszwecke „maßgeschneiderten“ Datensatzes wird sich aufgrund des gesetzlichen Rahmens, innerhalb dessen sich der Mikrozensus bewegt, sicherlich nicht zufriedenstellend für die Anwender lösen lassen.

Rechtliche Einschränkungen

Ein weiterer Problembereich sind die mit der faktischen Anonymisierung des Mikrozensus verbundenen rechtlichen Einschränkungen. Erfreulicherweise stellen diese Einschränkungen für 60,9 Prozent der Nutzer keine Behinderung dar. Für 39,1 Prozent ergaben sich daraus allerdings Probleme. Am häufigsten wurde, wie auch schon in der ersten Befragung 1998, die Beschränkung auf Projekte und Fristen (36,7 Prozent) und die fehlende Regionalisierung der Daten (20,0 Prozent) als problematisch genannt. In diesem Zusammenhang hat ZUMA in Kooperation mit der amtlichen Statistik und dem Bundesamt für Bauwesen und Raumordnung (BBR) basierend auf den Empfehlungen des Anonymisierungsprojektes (Müller et al. 1991)⁸ ein Konzept für die Operationalisierung eines MZ-Regionalfiles entwickelt (Wirth 2000)⁹ und mit Regionalforschern diskutiert. Gegenwärtig sind noch Fragen offen, die neben dem Stichprobenfehler die Veröffentlichung von Ergebnissen auf regionaler Ebene betreffen. ZUMA wird sich zusammen mit den Statistischen Ämtern dafür einsetzen, dass diese Fragen baldmöglichst geklärt werden, so dass Regionaldaten des Mikrozensus als Scientific Use File bereitgestellt werden können.

Veranstaltungen (Workshops, Konferenzen)

ZUMA und das Statistische Bundesamt bieten für die Nutzer von amtlichen Mikrodaten seit Jahren gemeinsam Workshops und Konferenzen an. Dieses Konzept des Angebotes verschiedener Plattformen des Austauschs und der Ausbildung wird von den Befragten eindrucksvoll bestätigt. Das Interesse an solchen Workshops und Nutzerkonferenzen zum Mikrozensus ist weiterhin beträchtlich: etwa drei Viertel sind an einer Teilnahme an einer der beiden Veranstaltungen bzw. an beiden interessiert. Am häufigsten werden thematische Veranstaltungen gewünscht (33,7 Prozent), es besteht jedoch auch Interesse an Veranstaltungen, die dem Austausch der Nutzer untereinander dienen (29,2 Prozent), anwendungsbezogene Übungen mit den Daten ermöglichen (23,6 Prozent) sowie allgemeine Informationen zum Mikrozensus und Einführungen zu den Daten vermitteln (13,5 Prozent).

8 Müller, W./Blien, U./Knoche, P./Wirth, H., 1991: Die faktische Anonymität von Mikrodaten. Forum der Bundesstatistik, Band 9, hrsg. v. Statistischen Bundesamt. Stuttgart: Metzler-Poeschel.

9 Wirth, H., 2000: Konzeption eines Mikrozensus-Regionalfiles. Expertengespräch Mikrozensus-Regionalfile. (Ms.) Mannheim: ZUMA.

Publikationen und Forschungsschwerpunkte

Erfreulicherweise publizieren die meisten Forscher ihre anhand der Mikrozensusdaten gefundenen Ergebnisse in einem breiten Spektrum an wissenschaftlichen Publikationen (84,6 Prozent) und tragen damit ganz wesentlich zu einer weiteren Verbreitung der Daten bzw. der Kenntniss darüber bei. Dabei handelt es sich um etwa gleich viele Zeitschriftenartikel (21,7 Prozent) und Arbeitspapiere/graue Literatur (23,3 Prozent). 17,1 Prozent der Befragten nutzten die Daten für Expertisen oder Projektberichte, 16,3 Prozent für Beiträge in einem Sammelband, 11,6 Prozent für Diplomarbeiten, Dissertationen oder Habilitationen und 10,1 Prozent für ein eigenes Buch.

Der Charakter des Mikrozensus als Mehrthemenumfrage spiegelt sich auch in einem breiten Spektrum an Forschungsgebieten wider, in denen die Daten genutzt werden. Einen Schwerpunkt bilden dabei die Bereiche „Arbeitsmarkt“ (23,5 Prozent), „Bildungsforschung“ (9,4 Prozent) und „Demografie“ (8,2 Prozent). Die weiteren Nennungen verteilen sich auf bekannte Felder wie Familienforschung, Wirtschaftsforschung, Medizinsoziologie aber auch methodisch/statistische Forschung, Sozialpolitik oder Umwelt und Verkehr.

Weiterer Bedarf und zukünftige Forschungsprojekte

Mit Ausnahme von drei Befragten äußern die Nutzer ein Interesse an weiteren Mikrozensus-Scientific Use Files. Im Vordergrund stehen dabei mit Abstand aktuelle Daten (46,7 Prozent), wobei vor allem eine möglichst zeitnahe Bereitstellung gewünscht wird. Daneben besteht in geringerem Maße Interesse an den bislang noch fehlenden Scientific Use Files der 90er Jahre (1990; 1992; 1994; 1999), sowie an Scientific Use Files für Mikrozensen ab den 60er Jahren. Das Interesse an älteren Mikrozensusdaten begründet sich in der Regel auf einer Erweiterung der Datenbasis - im gleichen Forschungsbereich - in die Vergangenheit. So lautete die Begründung eines Befragten, der im Gesundheitsbereich forscht: „alle Mikrozensen mit speziellen Daten zur Gesundheit“.

Die zukünftigen Forschungsprojekte decken sich erwartungsgemäß weitgehend mit den gegenwärtig bearbeiteten Fragestellungen. Schwerpunkte bilden auch hier Arbeitsmarkt (26,8 Prozent) und Haushalts- und Familienforschung (14,1 Prozent). Daneben ist eine Vielzahl an weiteren Forschungsprojekten aus den verschiedensten Bereichen genannt.

Anregungen und Kritik

32 Befragte haben z.T. sehr detailliert die Frage nach Verbesserungsvorschlägen und Anregungen beantwortet. Obwohl von vielen Befragten nochmals ausdrücklich das Angebot und der Service gelobt wurde, gab es doch auch sehr deutliche Kritik, die sich vor allem auf zwei Bereiche bezieht:

Ein wesentlicher Kritikpunkt bei den Befragten ist die mangelnde Aktualität der Scientific Use Files bzw. die lange Zeitdauer zwischen Erhebung und Bereitstellung der Daten. („Für die Aktualität der Forschungsergebnisse wäre es erforderlich, wenn die Files zeitnäher zur Verfügung gestellt würden“; „Verzögerungen von bis zu 5 Jahren zwischen Erhebung und Bereitstellung der Daten schränken Einsetzbarkeit des MZ für viele Fragestellungen ein. Denn mehr denn je ist heute Aktualität gefragt“).

Die Befragten kritisieren dies weiterhin deutlich, dass für die eigene Fragestellung wichtige Variablen oder Merkmale nicht erhoben werden, so z.B. im Einkommensbereich das Einkommen von Landwirten, die Erhebung des Bruttoerwerbseinkommen oder im familiensoziologischen Kontext z.B. die Erfassung der Kinder, die nicht mehr im Haushalt leben.

3. Zusammenfassung und Ausblick

Mit dem verbesserten Zugang zu amtlichen Mikrodaten in Form von kostengünstigen Scientific Use Files ist die Basis für eine empirische Analyse der Struktur und des Wandels unserer Gesellschaft um ein bedeutendes Element erweitert worden. Dies belegen die sichtlich gestiegenen Zahlen von genutzten Scientific Use Files und auch Institutionen, die mit diesen Daten arbeiten, sowie die erfreulich hohe Zahl von Publikationen mit diesen Daten.

Bislang noch nicht zufriedenstellend gelöst ist eine merkliche Verkürzung der Zeitdauer von der Erhebung des Mikrozensus bis zur Bereitstellung der Daten für die Wissenschaft. Hier bahnt sich jedoch eine Verbesserung der Situation durch die neu gegründeten Forschungsdatenzentren des Bundes und der Länder an, die ein „zeitnahes und nutzungsgerechtes“¹⁰ Angebot an Mikrodaten anstreben.

Auch der in verschiedenen Zusammenhängen geäußerte Wunsch nach mehr und besseren Informationen zu den Daten sollte mit den gegenwärtig laufenden Projekten zu elektronisch verfügbaren Metainformationssystemen (so z.B. das Projekt MISSY¹¹ im German Microdata Lab von ZUMA) in naher Zukunft erfüllt werden können.

10 Siehe <http://www.forschungsdatennetzwerk.de/kurzbeschreibung.asp>.

11 Im Projekt MISSY werden datenrelevante Informationen, die für die Analyse von amtlichen Mikrodaten unabdingbar sind, in systematischer Weise online zur Verfügung gestellt. Dadurch wird die Verwendung von Mikrodaten für empirisch arbeitende Wirtschafts- und Sozialforscher deutlich erleichtert. Fünf Teilaspekte stehen dabei im Vordergrund: Sammlung und Zusammenführung von datenrelevanten Informationen, Erweiterung der Zugangsmöglichkeiten, effiziente Nutzung der Daten mehrerer Erhebungszeitpunkte, Steigerung der Effektivität bei der Informa-

Was den Wunsch nach zusätzlichen Daten und bestimmten Variablen aus dem jeweiligen Forschungskontext der Forscher betrifft, gestaltet sich die Situation etwas schwieriger, da die Inhalte des Mikrozensus per Gesetz und Verordnungen geregelt sind. Gegenwärtig ist ein neues Mikrozensus-Gesetz für die Erhebungen ab 2005 in Vorbereitung. In diesem Zusammenhang bietet sich für die amtliche Statistik die Gelegenheit, neue Merkmale in das Erhebungsprogramm aufzunehmen, die von den Nutzern der Scientific Use Files dringend benötigt werden. Eventuell zeichnet sich darüber hinaus eine weitere positive Entwicklung ab, da der Statistische Beirat des Statistischen Bundesamtes im Juni 2003 die Bundesregierung gebeten hat, das Statistische System flexibler zu gestalten, d.h. dass „statistische Erhebungen entsprechend den Informationserfordernissen geändert werden können, ohne dass dafür – wie bisher – zeitaufwändige Gesetzgebungsverfahren eingeleitet werden müssen“¹². Dies wäre zumindest die notwendige Voraussetzung für eine flexiblere Gestaltung der Fragenprogramme entsprechend dem Informationsbedarf der empirischen Sozial- und Wirtschaftswissenschaften.

Korrespondenzadresse

*Dr. Paul Lüttinger
ZUMA
Postfach 12 21 55
D-68072 Mannheim
Email: luettinger@zuma-mannheim.de*

*Holger Breiholz, Dipl.-Volkswirt
Statistisches Bundesamt, Zweigstelle Bonn
Gruppe IXB – Mikrozensus
Graurheindorferstr. 19
D-53117 Bonn
Email: holger.breiholz@destatis.de*

tionsabfrage, Dokumentation und Archivierung älterer Daten (http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Daten/MISSY/MISSY_index.htm).
12 <http://www.destatis.de/presse/deutsch/pm2003/p2470530.htm>.

ALLGEMEINE BEVÖLKERUNGSUMFRAGE DER SOZIALWISSENSCHAFTEN: DATEN DES ALLBUS 2002 VERFÜGBAR

MICHAEL BLOHM

Im Jahr 2002 wurde die allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) zum dreizehnten Mal durchgeführt. In computerunterstützten persönlichen Interviews (CAPI) wurden in Westdeutschland ca. 1900 und in Ostdeutschland ca. 900 Personen befragt. Für die Stichprobenbildung wurden wie in den Jahren 1994, 1996 und 2000 Adressen aus den Einwohnermeldeämtern gezogen. Die Grundgesamtheit der Umfrage bildete die erwachsene Wohnbevölkerung in Privathaushalten. Auch ausländische Personen wurden befragt, wenn das Interview in deutscher Sprache durchgeführt werden konnte.

Inhaltlicher Schwerpunkt des ALLBUS 2002 ist das Thema "Religion, Weltanschauung und Werte". Das Frageprogramm besteht zum Teil aus der Replikation einschlägiger Fragen der ALLBUS-Erhebungen 1982 und 1992. Um den neueren Entwicklungen der Religionssoziologie Rechnung zu tragen, wurden neue Fragen, unter anderem zu außerchristlicher Religiosität und zu Glaubensinhalten, aufgenommen. Zur Vorbereitung des Religionsmoduls im ALLBUS 2002 war ein Symposium mit ausgewiesenen Experten auf dem Gebiet der Religionssoziologie durchgeführt worden. Die Erweiterungen des Religionsmoduls resultieren aus den Diskussionen dieses Symposium sowie auf Vorschlägen weiterer externer Experten.

Für den Bereich Religion und Weltanschauung sind im ALLBUS 2002 unter anderem folgende Aspekte erfragt worden:

- Erfahrungen mit und Meinung zu verschiedenen Formen von Glauben, Paraglauben und Aberglauben; Gottesglaube und Lebenssinn; Nachdenken über metaphysische Fragen; Häufigkeit von Meditation.
- Einstellungen zu Religiosität; religiöse Toleranz; religiöse Indifferenz; Einstellung zur kirchlichen Trauung; Selbsteinstufung der Religiosität; Glaubensinhalte; Religion vs. Wissenschaft; Sündenlehre; eigenes Bekehrungserlebnis.

- Teilnahme an kirchlichen Veranstaltungen und am Gemeindeleben; Vertrauen in die evangelische bzw. die katholische Kirche.
- Empfang von Sakramenten; Gebetshäufigkeit; Kirchengangshäufigkeit; Konfession; frühere Konfession; Kirchengangshäufigkeit als Kind; Angehörigkeit zu einer nicht-christlichen Religionsgemeinschaft; Konfession und Kirchengangshäufigkeit der Kinder und der Eltern; Taufe der Kinder.

Da mehrere Indikatoren zur Religiosität für drei Generationen vorliegen, ist es mit dem ALLBUS 2002 möglich, Analysen zur Tradierung religiöser Verhaltensweisen durchzuführen, beispielsweise zur Kirchengangshäufigkeit, die für den Befragten, dessen Eltern und dessen Kinder erfragt wurde.¹ In den Abbildungen 1 und 2 sind die Anteile derjenigen abgetragen, die nie zur Kirche gehen, getrennt nach Generation und Geburtskohorte des Befragten. In Westdeutschland nimmt im allgemeinen der Anteil derjenigen, die nie zur Kirche gehen, von Generation zu Generation zu. Auch für die Geburtskohorten der Befragten ist zu erkennen, dass der Anteil derjenigen, die nicht zur Kirche gehen, zunimmt, je jünger die Kohorten sind.

Für Ostdeutschland stellt sich die Situation anders dar. Generell ist der Anteil derjenigen, die nicht in die Kirche gehen, höher als in Westdeutschland. Auch sind die Unterschiede zwischen Befragten- und Elterngeneration größer als in Westdeutschland und zum Teil auch gegenläufig. Befragte, die vor 1950 geboren wurden, gehen – gegenwärtig – deutlich weniger in die Kirche als ihre Eltern. Befragte, welche nach 1959 geboren wurden, gehen – gegenwärtig – häufiger in die Kirche als ihre Eltern zum Zeitpunkt, als die Befragten 15 Jahre alt waren. Hier kommt zum Tragen, dass diese Befragten mit ihren Eltern in der ehemaligen DDR lebten. In der zweiten Generationenfolge gehen die Kinder aller Befragten(-kohorten) weniger häufig zur Kirche als die Befragten selbst, wobei die Zunahme derer, die nicht in die Kirche gehen, in Ostdeutschland größer als in Westdeutschland ist.

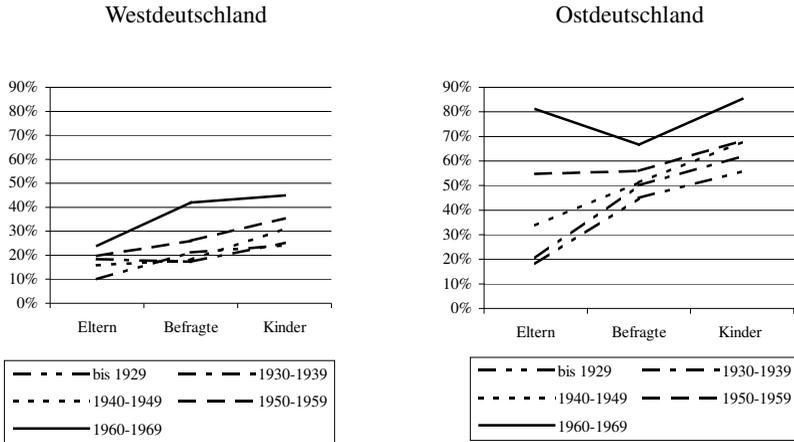
Neben Religion und Weltanschauung waren auch Werte Schwerpunkt im ALLBUS 2002. Zur Messung von Wertorientierungen wurden drei Instrumente eingesetzt. Neben politischen Zielen (Inglehart) wurden individuelle Wertorientierungen mit einem leicht modifizierten Instrument von Klages erfasst. Zur Messung kollektiver Wertorientierungen wurde eine, um ein Item erweiterte, Skala von Westle und Roßteutscher verwendet.² Es wurde

1 Für die Eltern wurde nach der Kirchengangshäufigkeit gefragt, als die Befragten 15 Jahre alt waren. Für die Befragten und deren Kinder wurde die gegenwärtige Kirchengangshäufigkeit erfragt.

2 Informationen zu und Analysen mit der Skala von Klages finden sich beispielsweise in: Klages, H./Gensicke, T., 1999: Wertewandel und Bürgerliches Engagement an der Schwelle zum

gefragt, wie gern Personen in Gesellschaften leben, in denen verschiedenen Werte wichtig sind.³

Abbildungen 1 und 2: Anteil derjenigen, die nie zur Kirche gehen, getrennt nach Generation und Geburtskohorte der Befragten⁴

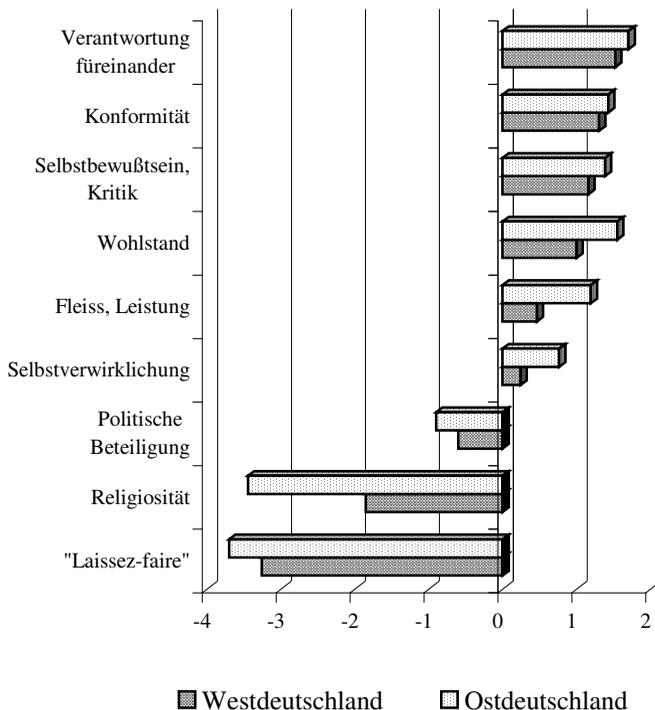


Quelle: ALLBUS 2002, eigene Berechnungen

21. Jahrhundert. Speyerer Forschungsberichte 193. Speyer: Forschungsinstitut für öffentliche Verwaltung. Herbert, W., 1991: Wandel und Konstanz von Wertstrukturen. Speyerer Forschungsberichte 101. Speyer: Forschungsinstitut für öffentliche Verwaltung. Für die Skala von Westle und Roßteutscher in: Roßteutscher, S., 1997: Consensus and Conflict. Value Collectives and Social Conflict in Contemporary German Society. Dissertation: European University Institute, Florence. Westle, B./Roßteutscher, S., 1992: Projekt Politische Kulturen im geeinten Deutschland. Dokumentation zur Konzeption, Operationalisierung und zum Test des standardisierten Erhebungsinstruments. Mannheim: Manuskript.

- 3 Die Frageformulierung lautet: „Ich lese Ihnen nun Beschreibungen vor, wie Gesellschaften aussehen könnten. Bitte sagen Sie mir für jede der folgenden Beschreibungen, wie gerne Sie in einer solchen Gesellschaft leben möchten. Eine Gesellschaft die Wert darauf legt, dass die Menschen {Beschreibung Gesellschaft}.“ Die Befragten konnten ihre Antwort auf einer Skala von 1 ‚überhaupt nicht gerne‘ bis 10 ‚sehr gerne‘ einstufen.
- 4 In die Berechnungen gingen nur Befragte mit deutscher Staatsangehörigkeit ein, die Kinder haben, welche vor 1984 geboren wurden. Für die Eltern wurde der Anteil der Nicht-Kirchgänger zwischen Vater und Mutter gemittelt. Auch bei den Kindern wurde bei mehr als einem Kind pro Befragter der Anteil der Nicht-Kirchgänger gemittelt. Es wurden nur erwachsene Kinder (vor 1984 geboren) berücksichtigt.

Abbildung 3: Relative Wichtigkeiten kollektiver Wertorientierungen für West- und Ostdeutschland⁵



Quelle: ALLBUS 2002, eigene Berechnungen

5 Die relativen Wichtigkeiten ergeben sich, wenn die Antworten zu den einzelnen Wertorientierungen um die Zustimmungstendenz bereinigt werden. Z.B. relative Wichtigkeit Item1 = $\frac{\text{Item1} - (\text{Item1} + \text{Item2} + \dots + \text{Item9})}{9}$.

Wie Abbildung 3 zeigt, ist die Rangfolge der kollektiven Wertorientierungen in West- und Ostdeutschland – mit einer Ausnahme – die Gleiche. Am liebsten möchten die Befragten in einer Gesellschaften leben, in der die Menschen füreinander Verantwortung tragen. Am wenigsten gern in einer Gesellschaft, in der jeder machen kann, was er will. West- und Ostdeutsche unterscheiden sich in der Präferenz, in einer Gesellschaft zu leben, in der auf wirtschaftliche Sicherheit und Wohlstand Wert gelegt wird. Für die Ostdeutschen ist dies die zweitwichtigste Wertorientierung, in Westdeutschland nur die viertwichtigste.

Mit dem gleichzeitigen Einsatz dreier Werte-Instrumente ermöglicht der ALLBUS 2002 nicht nur Analysen zum Einfluss unterschiedlicher Wertorientierungen auf Variablen und Variablenbeziehungen, sondern auch zum Zusammenhang zwischen kollektiven und individuellen Wertorientierungen sowie politischen Zielen.

Zur Abrundung des Frageprogramms wurden Fragen aus den bisherigen ALLBUS-Erhebungen repliziert, die in Verbindung mit den Schwerpunktthemen von besonderem Interesse sind. Dies sind unter anderem Fragen aus den Bereichen:

- Einstellungen zu Politik und Wirtschaft: Politische Partizipation; Parteipräferenz; Vertrauen in öffentliche Einrichtungen und Organisationen; Wahrscheinlichkeiten, diverse Parteien zu wählen; politisches Interesse; ehrenamtliche Tätigkeiten; Demokratiezufriedenheit in Deutschland; Beurteilung der derzeitigen und zukünftigen Wirtschaftslage.
- Moralische Bewertung abweichenden Verhaltens.
- Einstellungen zu Familie, Ehe und Partnerschaft.
- Einstellungen und Kontakte zu in Deutschland lebenden Ausländern.
- Subjektive Befindlichkeiten: allgemeines Vertrauen zu Mitmenschen und zu Politikern; sozialer Pessimismus und Zukunftsorientierung (Anomia); Nationalstolz; gerechter Anteil am Lebensstandard; Autoritarismus.

Wie in jeder Erhebung wurden auch im Jahr 2002 demographische Informationen detailliert erhoben. Außerdem finden sich in den Daten des ALLBUS 2002 auch Informationen zur Interviewdurchführung, zu den Merkmalen der Interviewer, zur Interviewsituation, zu Erfahrungen mit Umfragen und Einstellungen des Befragten gegenüber Umfragen.

Den gesamten Fragebogen zum ALLBUS 2002 können Sie sich unter <http://www.gesis.org/Dauerbeobachtung/Allbus/Inhalte/Fragenprogramm/frabo2002.htm> herunterladen. Aktuelle Informationen zum ALLBUS Programm finden Sie unter <http://www.gesis.org/Dauerbeobachtung/Allbus/index.htm>.

ISSP-Erhebung für Deutschland

Im Rahmen der ALLBUS-Befragung für Deutschland wurde auch wieder der deutsche Part des International Social Survey (ISSP) durchgeführt. Zwei ISSP-Module – nämlich zum Thema „Soziale Beziehungen und Hilfeleistungen“ (2001) und zum Thema „Familie in Deutschland“ (2002) wurden im Split bei jeweils der Hälfte der Befragten erfasst. Die Module waren als schriftlicher Selbstausfüllbogen konzipiert, den die Befragten im Anschluss an das ALLBUS-Interview bearbeiteten. Dem internationalen Projektverbund ISSP gehören mittlerweile 39 Länder aus fünf Kontinenten an (www.issp.org). Die beiden mit dem ALLBUS 2002 erhobenen Module stellen (modifizierte) Replikationen früherer ISSP-Erhebungen dar. Das Thema "Soziale Beziehungen und Hilfeleistungen" wurde bereits 1986 erhoben, "Familie in Deutschland" schon zweimal, 1988 und 1994. Das Frageprogramm "Soziale Beziehungen und Hilfeleistungen" umfasst unter anderem Fragen zu folgenden Themen:

- Kontakt mit Familie, Freunden und Bekannten
- Mitgliedschaft in Gruppen und Vereinen
- Solidarität in der Gesellschaft
- Lebenszufriedenheit
- Empfang von und eigene Hilfeleistungen
- „Personal efficacy“ in der Gesellschaft

Das Modul „Familie in Deutschland“ enthält unter anderem Fragen zu den Bereichen:

- Konflikt zwischen Beruf und Familie
- Arbeitsteilung und Entscheidungsfindung im Haushalt
- Rolle von Männern und Frauen in der Familie
- Lebenszufriedenheit

Informationen zum ALLBUS 2002 erteilt Michael Blohm bei ZUMA (Tel.: 0221/1246-276; email: blohm@zuma-mannheim.de). Ansprechpartnerin für das ISSP bei ZUMA ist Janet Harkness (Tel.: 0221/1246-284; email: harkness@zuma-mannheim.de).

Die Daten und das Codebuch (incl. der deutschen Daten der beiden ISSP-Module) sind seit Herbst 2003 erhältlich. Sie können sich diese – ebenso wie die Daten und Informationen zu den früheren Erhebungen – im ALLBUS WWW-Angebot kostenfrei unter der Adresse: <http://www.gesis.org/Datenservice/ALLBUS/index.htm> herunterladen. Es besteht auch die Möglichkeit, sämtliche Daten und Codebücher der ALLBUS-Erhebungen auf CD-ROM gegen eine geringe Gebühr beim Zentralarchiv für Empirische Sozialforschung (Ansprechpartner: Dr. Michael Terwey) zu beziehen.

Dr. Michael Terwey, Tel.: 0221/47694-22; email: terwey@za.uni-koeln.de

Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln; Postfach 410960; D-50869 Köln; Fax: 0221/47694-44.

QUEST – QUESTIONNAIRE EVALUATION STANDARDS

PETER PRÜFER & MARGRIT REXROTH

Die Internationale Pretestgruppe QUEST (**Q**uestionnaire **E**valuation **S**tandards) tagte vom 21. Oktober bis 23. Oktober 2003 bei ZUMA.

Die Gruppe besteht aus „aktiven Praktikern“, die allesamt umfangreiche Erfahrungen in der Durchführung von Pretests haben. Die Arbeitsgruppe trifft sich alle zwei Jahre zum Austausch von Forschungsergebnissen mit dem Ziel, über effektive Methoden und Techniken im Pretestbereich zu diskutieren und international verbindliche Standardprozeduren zu verabschieden. Um die Effektivität auch der Arbeitsgruppe selbst zu sichern, ist sowohl die Anzahl der Teilnehmer als auch der durch sie vertretenen Einrichtungen bewusst beschränkt.

An QUEST beteiligt sind derzeit 14 Institutionen aus insgesamt 9 Ländern: Deutschland, England, Finnland, Kanada, Neuseeland, den Niederlanden, Norwegen, Schweden und den U.S.A.

Der Arbeitsgruppe gehören so renommierte Umfrageforscher wie Jack Fowler (Universität Boston), Jennifer Rothgeb und Terry DeMaio (U.S. Bureau of Census, Washington), James L. Esposito (Bureau of Labor Statistics, Washington) an.

ZUMA ist seit 1997 Mitglied von QUEST. Seit dieser Zeit trafen sich die Mitglieder von QUEST zu folgenden Konferenzen:

- Oktober 1997: Statistics Sweden, Örebro, Schweden
- April 1999: The Office of National Statistics (ONS), London, Großbritannien
- Oktober 2001: U.S. Bureau of the Census, Washington, U.S.A.
- Oktober 2003: ZUMA, Mannheim, Deutschland

Die Vorträge, die im Rahmen der QUEST-Konferenzen gehalten wurden, sind jeweils in einem Proceedings-Band veröffentlicht. Die Vorträge der diesjährigen Konferenz bei ZUMA sollen im Rahmen eines ZUMA Specials veröffentlicht werden.

Fragen zu QUEST beantworten Peter Prüfer (Tel.: 0621/1246-227; email: pruefer@zuma-mannheim.de) und Margrit Rexroth (Tel.: 0621/1246-230; email: rexroth@zuma-mannheim.de).

PRETESTLABOR BEI ZUMA

PETER PRÜFER & MARGRIT REXROTH

Die Durchführung von Pretests ist seit vielen Jahren ein fester und häufig genutzter Bestandteil des Dienstleistungsangebotes von ZUMA.

Seit etwa 5 Jahren fokussiert sich das Pretest-Angebot auf die Durchführung kognitiver Pretests, mit denen in einem frühen Stadium der Fragebogenentwicklung unter Einsatz kognitiver Techniken vor allem das Fragenverständnis, das als eine der wichtigsten Ursachen von „survey errors“ gilt, untersucht wird. Mit Techniken wie Probing, Paraphrasing, Confidence Rating etc. (vgl. Prüfer, P./Rexroth, M., 1996: Verfahren zur Evaluation von Survey-Fragen: ein Überblick. ZUMA-Arbeitsbericht Nr. 96/05) wird gezielt und aktiv überprüft, wie Befragungspersonen Fragen oder Begriffe verstehen.

Zur optimalen Durchführung kognitiver Interviews richtet ZUMA zur Zeit ein Pretestlabor ein, in dem diese Interviews mittels Audio-/Video-Technik aufgezeichnet werden können. Dem Auftraggeber bietet sich somit die Möglichkeit, den Verlauf der Interviews mittels Videoaufzeichnung nachzuvollziehen und gegebenenfalls auch selbst zu analysieren.

Mit Jahresbeginn 2004 bietet ZUMA diese neue Dienstleistung für alle diejenigen an, die im Bereich der empirischen Sozialforschung Umfragen durchführen. Zuständig für das kognitive Labor und die Durchführung kognitiver Pretests sind die Mitarbeiter des Arbeitsbereichs „Pretesting“, Peter Prüfer (Tel.: 0621/1246-227, email: pruefer@zumamannheim.de) und Margrit Rexroth (Tel.: 0621/1246-230, email: rexroth@zumamannheim.de).

THE INTERNATIONAL WORKSHOP ON COMPARATIVE SURVEY DESIGN AND IMPLEMENTATION

In March 2003, an international group of some 35 researchers followed an invitation from ZUMA to attend the founding meeting of the International Workshop on Comparative Survey Design and Implementation – *CSDI*.

The main aims of *CSDI* as a working group are

- to establish an international and cross-disciplinary network of researchers committed to addressing the growing need for sound and up-to-date methodology in comparative survey research
- to set up and foster active ‘task forces’ on key methodological issues for comparative research
- to provide a forum and platform for informed discussion
- to establish a framework for defining standards for cross-national survey implementations
- to organise annual meetings at which research findings and task force activities can be presented and contacts of various kinds facilitated across disciplines and regions
- to seek to secure funding for joint projects to investigate or develop tools, procedures and guidelines for essential components of comparative survey research
- to co-ordinate and intensify knowledge transfer within and beyond *CSDI*

The initiative to form *CSDI* is a direct outcome of the 4th *ZUMA Symposium on Cross-Cultural Survey Methods*, held near Mannheim in spring of 2002. It stems from ZUMA’s own long-standing involvement in cross-national survey research and from ongoing co-operation between researchers who have attended one or more of the ZUMA symposia.

The second *CSDI* workshop will be held in Paris in spring 2004, with INED acting as local co-ordinator; the third will be held in Madrid in spring 2005, with ASEP as the local host. Participant numbers will be restricted in order to maintain the workshop character and output orientation of the meetings. Those attending will present on topics that can contribute to enhancing the quality of cross-national / cross-cultural survey research.

ZUMA has set up a *CSDI* website //www.csd-workshop.org // and runs a list server open to anyone interested in comparative survey research issues.

Contact person: Dr. Janet Harkness, email: harkness@zuma-mannheim.de

EUROPEAN CENTRE FOR CROSS-CULTURAL SURVEYS ESTABLISHED AT ZUMA

In principal set up to serve the German scientific community, ZUMA has for many years also been involved in cross-national survey research – as primary researchers, members of cross-national projects and programmes, survey designers, and, to an ever-growing extent, as consultants. ZUMA has thus a substantial track record in cross-national and cross-cultural survey and survey-related projects, including participation in the *International Social Survey Programme* (ISSP) since 1985 and in the *European Social Survey* (ESS) since 2001 (as members of the *Core Co-ordinating Team*); as consultants in *SHARE* (2002-), an EU-funded pilot study on health and ageing in Europe; as participants in the *International Mental Health Stigma Study* (IMHSS, 2003/4-2007); in *Pioneur*, a EU-funded project on EU-nationals working abroad (2002-2005); in *REAP*, an EU-funded five country project on the Environment (1993-1995); and in *SPQR*, an EU-funded five country project (1995-1997), in which ZUMA's contribution focused on demographic variables in cross-national social surveys.

In establishing the *European Centre for Cross-Cultural Surveys* (www.europeancentre.de), ZUMA has created a focal point for research activities and work teams hitherto located across different divisions within ZUMA. Headed by Professor Peter Ph. Mohler, Director of ZUMA, the new Centre will be the hub of ZUMA's current and future cross-cultural research activities. As a European Centre, it is the first to focus entirely on methodological issues in cross-cultural, multinational survey research, spanning both quantitative and non-quantitative/qualitative research.

The *European Centre* is the most recent development to reflect ZUMA's commitment to improving methods and standards in cross-cultural research within and across countries. For over a decade, for example, *ZUMA Symposia on Cross-Cultural Survey Methods* have served as an innovative forum for international and pan-European discussion of key methodological issues in multinational survey research. In spring 2003, ZUMA also initiated the *International Workshop on Comparative Survey Design and Implementation* (CSDI; <http://www.csdi-workshop.org/>), intended as platform and forum for a cross-disciplinary and international network of researchers committed to improving quality in areas of central importance for cross-cultural survey research.

THE INTERNATIONAL SOCIAL SURVEY PROGRAMME

The International Social Survey Programme (ISSP) is a collaborative programme of social science survey research conducted on five continents in currently 39 member countries – and soon approaching its twentieth year. Each year a module is finalised on a topic of importance for social science research. To date the ISSP has fielded 8 different modules on topics which all have been replicated at least once: *Role of Government*; *Social Networks*; *Social Inequality*; *Work Orientations*, *Religion*; *Family and Changing Gender Roles*; *National Identity* and *Environment*. A new module on *Citizenship* will be fielded by most ISSP members in 2004, and a further new module, *Leisure and Sports*, is in the planning for 2007.

More details on modules and the ISSP itself are available on the ISSP website (<http://www.issp.org>) as well as on GESIS pages. ZUMA pages cover details of German implementations and information on ZUMA's roles and activities in the ISSP (http://www.gesis.org/en/social_monitoring/issp/index.htm). The Central Archive is the official data archive for the ISSP and its website provides full information on archiving, data access, and study documentation (http://www.gesis.org/en/data_service/issp/index.htm). Starting April 2003, the ISSP Secretariat is located at the Norwegian Data Archive in Bergen, Norway (<http://www.issp.org/address.htm>).

SCIENTIFIC USE FILE DES MIKROZENSUS 2000 FÜR DIE WISSENSCHAFT ZUGÄNGLICH

Beim Statistischen Bundesamt können ab Oktober 2003 die faktisch anonymisierten Einzeldaten des Mikrozensus 2000 für Forschungszwecke bestellt werden. Wie für die bisher verfügbaren Mikrozensus-Scientific Use Files erfolgt die Lieferung der Daten zu Sonderkonditionen (Bereitstellungsgebühr von € 65,-), die durch eine Vereinbarung zwischen dem Statistischen Bundesamt und dem Bundesministerium für Bildung und Forschung (BMBF) ermöglicht wurden.

Die Daten der amtlichen Statistik sind für die empirischen Sozial- und Wirtschaftswissenschaften eine der wichtigsten Informationsquellen im Rahmen der Analyse wirtschafts- und gesellschaftspolitischer Zusammenhänge. Dies gilt insbesondere für den Mikrozensus, der als größte Mehrthemenumfrage der Bundesrepublik Deutschland vielfältige Informationen über die wirtschaftliche und soziale Lage der Bevölkerung enthält.

Das jährliche Grundprogramm umfasst folgende Themenbereiche: soziodemographische Angaben, Staatsangehörigkeit, Familien- und Haushaltszusammenhänge, Beziehung zum Lebenspartner (nichteheliche Lebensgemeinschaft), Haupt- und Nebenwohnung, Erwerbsbeteiligung und -tätigkeit, Schulbesuch, Ausbildungsabschluss, Quellen des Lebensunterhalts, Einkommen, gesetzliche Rentenversicherung. Darüber hinaus werden in jährlichen Ergänzungsprogrammen mit einem Auswahlsatz von 0,5 Prozent weitere Themen erfragt. Dies umfasst ergänzende Angaben zur Erwerbstätigkeit und zur früheren Erwerbstätigkeit, Retrospektivangaben zur Situation ein Jahr vor der Erhebung (Erwerbstätigkeit, Stellung im Beruf, Wirtschaftszweig, Wohnsitz) und berufliche und allgemeine Aus- und Fortbildung. Im Rahmen der vierjährigen Zusatzprogramme enthält der Mikrozensus 2000 Zusatzangaben zur beruflichen Ausbildung (Hauptstudienfachrichtung), Pendlermerkmale (Weg zur Arbeitsstätte, Lage der Arbeitsstätte, Entfernung, Verkehrsmittel, Zeitaufwand), Zusatzangaben für Ausländer (Anzahl der Kinder im Heimatland, Ehegatte bzw. Eltern im Heimatland) und für Erwerbstätigkeit (Tätigkeitsmerkmale, Abteilungen am Arbeitsplatzes, Arbeitszeit in der 2. Erwerbstätigkeit).

Nach dem Bundesstatistikgesetz von 1987 dürfen für die Durchführung wissenschaftlicher Vorhaben vom Statistischen Bundesamt und den Statistischen Ämtern der Länder Einzelangaben an Hochschulen oder sonstige Einrichtungen in Deutschland mit der Aufgabe unabhängiger wissenschaftlicher Forschung übermittelt werden, wenn sie den Befragten nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft ("faktische Anonymität") zugeordnet werden können. Diese Daten müssen zweckgebunden verwendet werden und sind zu löschen, sobald das wissenschaftliche Vorhaben beendet ist. Für die Weitergabe einer, dem konkreten Forschungsvorhaben entsprechenden Variablenauswahl steht das Scientific Use File als eine faktisch anonymisierte 70%-Substichprobe des Mikrozensus 2000 zur Verfügung, das Einzelangaben über rund 503.000 Personen bzw. rund 230.000 Haushalte enthält.

Die als Scientific Use Files zur Verfügung stehenden Mikrozensus-Daten können beim Statistischen Bundesamt bestellt werden. Nähere Informationen dazu unter:

*Statistisches Bundesamt
Zweigstelle Bonn
Gruppe IX B – Mikrozensus
Holger Breiholz
Postfach 17 03 77
D – 53029 Bonn
Tel.: 01888/643-8711, Fax: -8692
email: mikrozensus@destatis.de*

Darüber hinaus können sich interessierte Wissenschaftler an Ansprechpartner im Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim wenden. Die Abteilung Mikrodaten (German Microdata Lab) bietet zu den Scientific-Use-Files im World Wide Web (http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Daten/Abteilungsdaten/Mikrozensen/mz_daten_index.htm)

u.a. folgende Informationen an: Datendokumentation, Fragebogen, rechtliche Voraussetzungen der Datenlieferung und Bestellmöglichkeiten.

*ZUMA
Abteilung Mikrodaten
Postfach 12 21 55
D – 68072 Mannheim
Tel.:0621/1246-265, Fax: -100
email: mikrodaten@zuma-mannheim.de
<http://www.gesis.org/Dauerbeobachtung/Mikrodaten/>*

EUROPEAN SOCIAL SURVEY: ERSTE LÄNDERDATEN DER ERHEBUNG 2002/2003 VERFÜGBAR

Nach zweijähriger Projektlaufzeit ist der erste Teil der Daten aus dem Projekt European Social Survey (ESS) verfügbar. Seit Ende September 2003 können nach Abschluss einer äußerst umfangreichen Datenedition und -prüfung alle Interessierten über die vom Norwegian Social Science Data Service (NSD) eingerichtete Homepage <http://ess.nsd.uib.no> oder über einen Link auf der internationalen ESS-Projekthomepage <http://www.european-socialsurvey.org/archive> and data direkt auf die Daten zugreifen.

Bisher stehen Daten aus 15 Ländern zur Verfügung: Großbritannien, Finnland, Griechenland, Ungarn, Irland, Israel, Niederlande, Norwegen, Polen, Portugal, Slowenien, Spanien, Schweden, Schweiz, Tschechische Republik. Die Daten der noch ausstehenden Teilnehmerländer werden nach und nach zugänglich gemacht werden. Informationen zu neuen Datenveröffentlichungen werden jeweils auf den oben genannten Homepages verfügbar sein.

Nutzer der Homepage können einfache statistische Analysen direkt online durchführen oder sich die gewünschten Daten herunterladen. Dabei können entweder alle Länderdaten als Paket heruntergeladen oder Teildatensätze individuell nach den jeweiligen Nutzerwünschen zusammengestellt werden (z.B. nur ausgewählte Länder). Die Daten stehen in verschiedenen Formaten zur Verfügung (SPSS, SAS, STATA, DBASE). Bei der Homepage wurde insbesondere auf Benutzerfreundlichkeit geachtet. Auch Ungeübte sollen sich auf diesem Weg einen Einblick in Daten und Ergebnisse der ersten Welle des ESS verschaffen können.

Neben den Daten aus den Befragungen ist über die ESS-Homepage der Zugriff auf eine Reihe von Meta-Daten und eine umfangreiche Studiendokumentation möglich. Es sind z.B. Informationen über die politischen und die Bildungssysteme der Teilnehmerländer verfügbar, über das Altersprofil der Bevölkerung, die Durchführung der Feldarbeit, über Ausschöpfungsquoten und Ereignisse im Umfeld der Erhebungen.

Anlässlich des erfolgreichen Abschlusses der ersten Welle des ESS wird am 25./26. November 2003 eine Konferenz in Brüssel stattfinden. Eine Reihe renommierter SozialwissenschaftlerInnen werden erste methodische und substantielle Analysen der Daten präsentieren. Zwei Gastredner aus den USA werden eine weitere komparative Perspektive einbringen. Eine Veröffentlichung der Konferenzbeiträge in zwei Sammelbänden ist geplant. Weitere Informationen und Anmeldung: ess@natcen.ac.uk.

CALL FOR PAPERS

Due date: March 31, 2004

The Journal of Official Statistics (JOS) in cooperation with the Web Survey Methodology Portal

(<http://www.websm.org>)

is planning a

Special issue on Web surveys

SCOPE AND TOPICS: In recent years, a new way of collecting survey data has come into use: data collection through the World Wide Web. This innovation has many advantages, but also generates new methodological challenges. For this special issue we welcome manuscripts on various aspects of Web survey methodology and application, including (but not restricted to):

- coverage issues and respondent solicitation strategies
- sampling and weighting issues
- visual design effects including the use of multimedia
- nonresponse issues and participation patterns
- measurement errors
- mode effects and mixed mode surveys
- Web data collection in special contexts, such as employee surveys via corporate intranets, etc.
- questionnaire development and pretesting methods
- studies on the utility and effectiveness of Web surveys, such as cost and benefit issues
- software reviews and other technological issues.

JOURNAL INFORMATION: JOS is a scholarly journal published by Statistics Sweden, the national statistical office of Sweden. The journal publishes articles on statistical methodology and theory, with an emphasis on applications. For more information on JOS, please visit the JOS home page, www.jos.nu.

SUBMISSION GUIDELINES: Contributions should be submitted exclusively via email to jos@scb.se, preferably in MS Word. Detailed guidelines for preparation of a manuscript are available at the JOS home page. Manuscripts proposed for this special issue should not have been submitted or published elsewhere. All manuscripts are peer reviewed and evaluated with respect to quality, significance and clarity of presentation. Manuscripts should be as concise as possible without loss of clarity.

To be considered for the special issue, a submitted manuscript must reach JOS no later than March 31, 2004. The special issue is tentatively scheduled to appear in December 2004.

EDITORIAL COMMITTEE: The special issue is administered by a guest editorial committee, comprised of the following members:

- Dr. Michael Bosnjak (University of Mannheim, Germany)
- Dr. Gösta Forsman (Swedish National Road Administration and Linköping University, Sweden)
- Dr. Annica Isaksson (Linköping University, Sweden) -- Committee chair
- Dr. Katja Lozar Manfreda (University of Ljubljana, Slovenia)
- Dr. Lars Lyberg (Statistics Sweden, Sweden) -- Chief editor of JOS
- Dr. Tracy Tuten Ryan (Virginia Commonwealth University, Richmond, VA, USA)
- Dr. Matthias Schonlau (RAND, USA)

The committee members are looking forward to your submission.

* * *

GOR04

SIXTH INTERNATIONAL GOR CONFERENCE

GERMAN ONLINE RESEARCH '04

30. und 31. März 2004

Am 30./31. März 2004 findet an der Universität Duisburg-Essen, Campus Duisburg, die 6. Internationale Konferenz GERMAN ONLINE RESEARCH statt. Die Tagung wird vom Lehrstuhl für Sozialwissenschaftliche Methoden/Empirische Sozialforschung (Prof. Dr. Frank Faulbaum) zusammen mit der Deutschen Gesellschaft für Onlineforschung veran-

staltet. Die Konferenz ist die Größte ihrer Art nicht nur in der Bundesrepublik, sondern auch international. In über 100 wissenschaftlichen Beiträgen aus dem In- und Ausland werden Theorien, Methoden, Techniken und Ergebnisse der Forschung über das Internet sowie der Online- oder Mobilkommunikation behandelt. Ziel der GOR 04 ist es, den Stand der wissenschaftlichen Forschung, innovative Entwicklungen sowie praktische Erfahrungen zu dokumentieren und damit einen Transfer zwischen

- Wissenschaftlern
- Anwendern der Internet-Forschung
- Theorie und Praxis
- Universitäten und Unternehmen
- Nachfragern und Anbietern

zu ermöglichen. Die Konferenz wird Don Dillman (Washington State University) mit einem Keynote-Vortrag eröffnet. Einen besonderen Schwerpunkt wird die Online-Umfrageforschung bilden. Eine detaillierte Auflistung aller Themenbereiche sowie die Anmeldemodalitäten sind der folgenden Website zu entnehmen: <http://www.dgof.de/gor04/>

Die Konferenz sieht ebenfalls Ausstellungen vor. Interessenten werden gebeten, sich an folgende Adresse zu wenden: office.gor@uni-duisburg.de

Die Tagungsgebühren betragen für

- *Mitarbeiter von Hochschulen:* 120 Euro
- *Studierende:* 60 Euro
- *Aktive Teilnehmer:* 30 Euro (nur ErstautorInnen)
- *Angehörige nicht-universitärer Einrichtungen:* 290 Euro

Frühzeitige Anmeldungen bis zum 31.1.03 werden um 20% vergünstigt. In den Tagungsgebühren enthalten sind neben den Tagungsunterlagen zwei Mittagessen, das Dinner am 30.03.04 sowie Getränke während der Tagung. Die (WWW-basierte) Anmeldemöglichkeit zum Besuch der GOR04 ist ab dem 01. Dezember 2003 freigeschaltet. Weitere Informationen werden laufend unter <http://www.dgof.de/> gepostet.

* * *

BUCHBESPRECHUNG

Das Thema 'Umfragequalität' wird in dieser Ausgabe der ZUMA-Nachrichten mehrfach aufgegriffen (Mohler/Koch/Gabler; Schneekloth/Leven). In diesem Zusammenhang soll hier das erste einführende Handbuch in das Gebiet angezeigt werden. Den Autoren ist zuzustimmen, dass eine wichtige Vorbedingung für die Verbesserung der Umfragequalität gute Kenntnisse der Umfragemethoden und des Umfrageproduktionsprozesses sind. Es trifft auch zu, dass die akademische Ausbildung für 'survey workers' hinter dem Bedarf des Wirtschaftszweiges 'Umfrageforschung' weit hinterherhinkt. Vielen 'survey workers' fehle deshalb eine angemessene Ausbildung, um die komplexe Prozessstruktur einer Umfrage verstehen zu können. Die Autoren folgern daraus, dass eine Handreichung für Umfragequalität zugleich auch eine Einführung in die Methoden der Umfrageforschung sein müsse. Von daher geht ihr Anspruch wesentlich weiter, als Begriffe und Methoden der Prozess- und Produktionsqualität für die Umfrageforschung fruchtbar zu machen. Es handelt sich also um eine Darstellung der Umfragemethoden unter dem Aspekt der Qualitätsverbesserung. Insgesamt kann man sagen, dass beides gelungen ist.

BIEMER, PAUL P. & LYBERG, LARS E.

Introduction to Survey Quality

Wiley Series in Survey Methodology

Hoboken, New Jersey: John Wiley & Sons, 2003

ISBN 0-471-19375-5, 402 Seiten, 127,24 € oder 94.95 US\$

Die Autoren leiten die Leser dazu in zehn Schritten durch die verschiedenen, aufeinander bezogenen Themen, wobei sie sich auf ein profundes Wissen und die neueste Literatur stützen. Die einzelnen Kapitel bilden in sich abgeschlossene Einheiten, wodurch selektives Lesen und Nachschlagen ermöglicht wird. Im ersten Kapitel geben sie einen Abriss der Umfragegeschichte. Ihr Schwerpunkt sind dabei Umfragen der (amtlichen) Statistik, weniger Erhebungen von Meinungen, Einstellungen, Werten u.ä. Diese Schwerpunktbildung durchzieht den ganzen Band (dies bedeutet auch, dass wichtige psychologische und modelltheoretische Aspekte durchgängig vernachlässigt werden). Danach skizzieren sie, was man unter 'Qualität' und 'Qualitätsmanagement' verstehen kann. Im zweiten Kapitel

konzentrieren sie sich auf die Verflechtung zwischen dem Prozess der Umfrageherstellung und der resultierenden Datenqualität (Total Survey Error, Nonsampling Error, Mean Squared Error). Der sogenannte ‚Coverage Error‘ und Verzerrungen durch Nichtteilnahme (nonresponse) werden im dritten Kapitel erörtert. Besonders betont wird die Komplexität der Berechnung von Ausschöpfungsquoten, die, entgegen der landläufigen Meinung, mit zu den diffizilsten Qualitätsindikatoren der Umfrageforschung zu zählen sind. Im vierten Kapitel werden Probleme des Messens mittels Fragebögen dargestellt. Die Autoren verweisen hier u.a. auf die ‚kognitive‘ Wende in der Umfrageforschung. Man kann aber den Eindruck gewinnen, dass die damit verbundenen kommunikations- und entscheidungstheoretischen Annahmen nicht in ihrer vollen Dramatik für die Umfrageforschung gesehen werden. In diesem Zusammenhang sollte man auch das Fehlen einer Diskussion über das Messen latenter Konstrukte anmerken. Das ist sicherlich mit dem gewählten Schwerpunkt ‚amtliche Statistik‘ zu erklären. Es bleibt aber dennoch eine Schwachstelle dieses Bandes.

Im vierten Kapitel wird die Rolle der Interviewer beleuchtet. Entsprechend der Standardliteratur werden eine ganze Reihe von Merkmalen (soziodemographische, verhaltensbedingte, Alter, Geschlecht usw.) abgehandelt. Bezeichnenderweise wird der von Noelle-Neumann/Petersen (2000: 318ff) wichtigste Faktor ‚Talent‘ nicht angesprochen, stattdessen wird von ‚skills‘ im Sinne technischer Fertigkeiten gesprochen. Dies hat sicher seinen Grund darin, dass es zur Frage des Interviewertalents keine systematische Literatur gibt, was aber nichts an dessen Bedeutung mindern kann. Im sechsten Kapitel wird eine umfassende Darstellung der Effekte unterschiedlicher Erhebungsverfahren gegeben (face-to-face, telefonisch, computergestützt, schriftlich). Die unkritische Übernahme von möglicherweise kulturspezifischen Effekten als Universalien (etwa Dillmans Empfehlungen) entspricht zwar dem derzeitigen Stand der Forschung, sollte aber als Problem von den Lesern mitbedacht werden. Wichtig, weil so selten ausführlich behandelt, ist das siebte Kapitel zur Datenerfassung, -bereinigung und -verarbeitung. Dies ist eine in der Literatur stiefmütterlich behandelte Fehlerquelle, und es ist in der Tat ein besonderes Verdienst dieses Buches, dies in aller Ausführlichkeit darzustellen.

Damit kommen wir zu den letzten drei Kapiteln, die nicht mehr einer Nachzeichnung des Umfrageprozesses als Leitstruktur folgen.

Im achten Kapitel diskutieren die Autoren Verfahren der Bewertung von Fehlern (survey error evaluation). Dies ist der Teil, in dem die zuvor eingestreuten Hinweise auf Qualitätsmanagement systematisiert werden. Das neunte Kapitel ist eine breite Darstellung von Fragen des Stichprobenfehlers. Die Autoren widmen dabei dem Thema ‚simple random sampling‘ viel zu viel Platz, wenn man die geringe Zahl von Ländern bedenkt, in denen

derartige Stichprobenpläne zum Alltag zählen können. Das Kapitel wäre eigentlich nicht nötig oder hätte kurz zwischen Kapitel zwei und drei eingefügt werden können.

Das letzte Kapitel nimmt die Themen der Kapitel eins bis acht auf und bettet diese in allgemeine Empfehlungen zur Minimierung des Total Survey Error ein. Wer sich für Umfragequalität interessiert, wird in den Kapiteln acht und neun fündig werden. Hier findet man auch wichtige Hinweise auf die Dokumentation des Produktionsprozesses und der Nutzung solcher Dokumentationen für die Qualitätsverbesserung. So gut diese Kapitel über die derzeit herrschenden Vorstellungen zum Qualitätsmanagement in der Umfrageforschung informieren, legen sie doch zugleich auch Zeugnis ab von dem eher miserablen Stand der wissenschaftlichen Durchdringung des Gebietes. Damit ist gemeint, dass es Standards, Leitlinien usw. zur Genüge gibt. Was fehlt, sind fundierte Untersuchungen der Effekte von Prozessveränderungen, die, wie man aus der Literatur zum Qualitätsmanagement in der produzierenden Industrie weiß, oft sehr subtil sein können. Um es deutlich zu machen, es geht hier nicht um die ‚großen‘ Entscheidungen ‚Incentive‘, ‚Registerstichprobe‘ oder ‚Substitution‘. Vielmehr geht es um den Alltag im Feldgeschehen: was tun, wenn von einer Studie auf die nächste alle Interviewer anstelle von durchschnittlich 7 Interviews nur noch 6 abliefern? Größere Institute führen in der Regel mehrere Studien gleichzeitig durch; welche Prozesse der Qualitätssteuerung sind dafür angemessen? Kurz gesagt, welches Handwerkzeug kann man Feldchefs an die Hand geben, das wissenschaftlich fundierte Wirkung verspricht? Hier gibt es offensichtlich noch vieles zu tun. Biemer und Lyberg haben dafür eine gute Grundlage gelegt.

Peter Ph. Mohler

Literatur

Noelle-Neumann, E./Petersen Th., 2000: Alle, nicht jeder. Einführung in die Methoden der Demoskopie. 3. Auflage. Berlin: Springer.

ZUMA-ARBEITSBERICHTE

Die ZUMA-Arbeitsberichte informieren über das gesamte Spektrum der Arbeit bei ZUMA und enthalten überwiegend Beiträge, die zur Veröffentlichung in Fachzeitschriften vorgesehen sind. Nachfolgend werden die ZUMA-Arbeitsberichte, die seit der letzten Ausgabe der ZUMA-Nachrichten publiziert worden sind, in Form von Abstracts kurz vorgestellt. ZUMA-Arbeitsberichte sind auch im Internet als PDF-Datei abrufbar (www.gesis.org/publikationen/berichte/zuma_arbeitsberichte/). Einzelhefte können unter folgender Adresse bestellt werden:

Zentrum für Umfragen, Methoden und Analysen
ZUMA-Publikationen
Postfach 12 21 55
68072 Mannheim

Georgios Papastefanou, Antje Springer, Sterios Tsioumanis & Kostadinos Mattas:
Cultural context and attitudes towards genetically modified food in Greece and West Germany. ZUMA-Arbeitsbericht 2003/01.

Bisherige Untersuchungen haben gezeigt, dass es innerhalb der Europäischen Union zum Teil erhebliche Unterschiede in den Einstellungen zu gentechnisch veränderten Lebensmitteln (GM food) gibt. Die Erklärungen für die gefundenen Einstellungsunterschiede basieren vor allem auf soziodemografische und Wissensvariablen als einstellungsdeterminierende Faktoren der Einstellung gegenüber GM food. Kulturelle Unterschiede werden hingegen als Erklärungsfaktoren kaum in Betracht gezogen. In dem vorliegenden Papier werden Hypothesen über den Einfluss des kulturellen Kontexts auf die Ablehnung von GM food formuliert. Diese Hypothesen basieren auf Überlegungen, die sich aus der Kombination des Ansatzes von Hofstede (1997) mit dem Einstellungsmodell von Eagly und Chaiken (1993) ergeben. Die Überlegungen beziehen sich auf Griechenland und Westdeutschland als Länder mit Stellvertreterpotential für einen bestimmten kulturellen Kontext (modern vs. traditional). Als Ergebnis kann festgehalten werden, dass das Wissen über GM food nur in Westdeutschland einen Effekt hat, in Griechenland sind hingegen die negativen Beliefs über GM food von großer Bedeutung für die Einstellung gegenüber GM food. Generell lässt sich festhalten, dass in Westdeutschland eine wesentlich strukturiertere Einstellungsstruktur zu finden ist als dies in Griechenland der Fall ist.

ZUMA-METHODENBERICHTE

In der ZUMA-Reihe „Methodenberichte“ werden Beiträge publiziert, die wichtige Aspekte der Arbeit von ZUMA dokumentieren, aber in der Regel nicht für eine Veröffentlichung in einer Fachzeitschrift geeignet sind, wie z.B. Dokumentationen, Bibliographien, Literaturberichte und Datenbeschreibungen. Die Berichte sind im Internet als PDF-Datei abrufbar (www.gesis.org/publikationen/berichte/zuma_methodenberichte/). Einzelhefte können auch bei folgender Adresse bestellt werden:

Zentrum für Umfragen, Methoden und Analysen
ZUMA-Publikationen
Postfach 12 21 55
68072 Mannheim

Antje Springer, Bettina Knapp & Georgios Papastefanou:

Bibliografie zu ausgewählten Analysen der Luxembourg Income Study (LIS) „Deutschland im internationalen Vergleich“ (Stand 9/2002). ZUMA-Methodenbericht 2003/04.

Antje Springer, Bettina Knapp & Georgios Papastefanou:

EVS-Bibliographie (2.Auflage, Stand: September 2002). ZUMA-Methodenbericht 2003/05.

Sabine Klein & Janet Harkness:

ISSP Study Monitoring 2000. ZUMA-Methodenbericht 2003/06.

Birgit Neugebauer:

Mitarbeiterbefragungen. Ein Literaturbericht. ZUMA-Methodenbericht 2003/07.

Nicole Kollinger & Annette Kohlmann:

Vergleich der Mikrozensus 1998 und 2000. ZUMA-Methodenbericht 2003/08.

Daniel Lehnert, Felix Weiss & Annette Kohlmann:

Mikrozensus 2000. Dokumentation und Datenaufbereitung. ZUMA-Methodenbericht 2003/09.

Bernhard Schimpl-Neimanns:

Mikrodaten-Tools: Umsetzung der Berufsklassifikation von Blossfeld auf die Mikrozensus 1973-1998. ZUMA-Methodenbericht 2003/10.

Peter Prüfer, Lisa Vazansky & Darius Wystup:

Antwortskalen im ALLBUS und ISSP. Eine Sammlung. ZUMA-Methodenbericht 2003/11.

ZUMA-WORKSHOPS

Erstes Halbjahr 2004

Thema	Datum
Der Mikrozensus als Datenquelle für die empirische Sozial- und Wirtschaftsforschung: Eine Einführung	17. Februar
Empirische Grundlagen der Sozialstrukturanalyse	9. - 10. März
Mitarbeiterbefragungen	16. - 17. März
Online Surveys	18. - 19. März
Einführung in die Clusteranalyse	23. - 24. März
Empirische Eliteforschung	6. - 7. Mai
Ansätze der computerunterstützten Inhaltsanalyse	11.- 12. Mai
Bevölkerungsumfragen am Beispiel des ALLBUS	13. - 14. Mai
Auswahl, Beurteilung und Anwendung sozialwissenschaftlicher Erhebungsinstrumente	26. Mai
Pretesting	27. Mai
HAMLET, Einführung in ein Programm der cui	15. - 16. Juni
Einführung in Datenmanagement und Datenrecodierung Teil 1 – Einstieg für Teilnehmer ohne Vorkenntnisse	22. Juni
Einführung in Datenmanagement und Datenrecodierung Teil 2 und 3 – Teilnehmer mit Vorkenntnissen	23. - 24. Juni
Datenanalyse bei komplexen Stichprobendesigns	6. Juli
Simulation for the Social Sciences	27. September - 1. Oktober

Detaillierte Hinweise zu den oben aufgeführten Veranstaltungen erhalten Sie unter: <http://www.gesis.org/Veranstaltungen/ZUMA/Workshops/index.htm> oder beim ZUMA-Tagungssekretariat (workshop@zuma-mannheim.de, Tel.: 06 21 – 12 46 221).

DURCHWAHL-RUFNUMMERN (STAND: NOVEMBER 2003)

Sie erreichen die Mitarbeiter von ZUMA unter der Nummer 0621-1246-(Durchwahlnummer); die Zentrale unter 1246-0. Sie ist von Montag bis Donnerstag von 8.30 bis 17.00 und freitags von 8.30 bis 15.30 besetzt. Die mit (S) bezeichneten Mitarbeiterinnen nehmen Sekretariatsaufgaben wahr.

DIREKTION

<i>Prof. Dr. Peter Ph. Mohler (Direktor)</i>	173
Carol Cassidy (Stellv. Direktorin)	146
Margit Bäck (S)	172
Elisabeth Bähr (S)	172
Maria Kreppe-Aygin	184

INTERNE INFRASTRUKTUR

Verwaltung

Dipl.-Kfm. Jost Henze	161
-----------------------	-----

Information & Kommunikation

Dipl.-Soz. Kerstin Hollerbach	174
-------------------------------	-----

EDV-Infrastruktur

Carol Cassidy	146
---------------	-----

Datenbanken

Joachim Wackerow	262
------------------	-----

WISSENSVERMITTLUNG & BERATUNG

Wissenschaftlicher Leiter

Prof. Dr. Ingwer Borg	151
-----------------------	-----

Projektberater

Dr. Wolfgang Bandilla	136
Dr. Michael Braun	176
PD Dr. Jürgen H.P. Hoffmeyer-Zlotnik	175
Dipl. Soz. Rolf Porst	228
Dr. Beatrice Rammstedt	155
Christa v. Briel (S)	231
Dagmar Haas (S)	152
Patricia Lüder (S)	221

Pretesting

Dipl.-Psych. Peter Prüfer	227
Margrit Rexroth, M.A.	230

Textanalyse, Verocoding

Alfons J. Geis, M.A.	222
Patricia Lüder (S)	221

Statistik

PD Dr. Siegfried Gabler	281
Dr. Sabine Häder	282
Dipl.-Math. Michael Wiedenbeck	279

Telefonumfragen

Dipl.-Soz. Michael Schneid	209
Dipl.-Soz. Angelika Stiegler	208

Online-Umfragen

Dipl.-Sozialw. Wolfgang Neubarth	205
----------------------------------	-----

Elektronische Handbücher ZIS/EHES

Dr. Angelika Glöckner-Rist	171
----------------------------	-----

Computerunterstützte Textanalyse,

Textpack, NSD-stat

Cornelia Züll	147
Juliane Landmann	144

DAUERBEOBACHTUNG

Wissenschaftliche Leiterin

Prof. Dr. Ursula Hoffmann-Lange	247
---------------------------------	-----

ALLBUS

<i>Dipl.-Soz. Achim Koch</i>	280
Dipl. Soz. Michael Blohm	276
Dipl. Soz. Alexander Haarmann	273
Dipl.-Soz. Martina Wasmer	273
Julia Khorshed (S)	274

International Social Survey Programme (ISSP)

<i>Dr. Janet Harkness</i>	284
Sabine Klein	272
Evi Scholz	283

Mikrodaten

<i>Dipl.-Soz. Bernhard Schimpl-Neimanns</i>	263
Dipl.-Soz. Jeanette Bohr	261
Dipl.-Sozialwiss. Helga Christians	266
Dipl. Soz. Nadia Granato	264
Dr. Annette Kohlmann	253
Dr. Paul Lüttinger	268
Dr. Heike Wirth	269
Dipl.-Soz. Immo Wittig	267
Irene Fischer (S)	265

Einkommen & Verbrauch

<i>Dr. Georgios Papastefanou</i>	278
----------------------------------	-----

Soziale Indikatoren

<i>Dr. Heinz-Herbert Noll</i>	241
Dipl.-Soz. Regina Berger-Schmitt	248
Dr. Caroline Kramer	244
Dr. Stefan Weick	245
Margit Bäck (S)	242

DRITTMITTELPROJEKTE

Nina Rother (PIONEUR)	285
Lars Kaczmirek (WebSM)	206