

Entscheidungsspielräume von Interviewern bei der Wahrscheinlichkeitsauswahl

Ein Vergleich von ALLBUS-Erhebungen

Interviewers' scope for influence on random sampling

A comparative analysis of ALLBUS surveys

Wolfgang Sodeur

Zusammenfassung

Eine ideale Wahrscheinlichkeitsauswahl würde genau festgelegten Regeln folgen. Beteiligte Personen müssten sich exakt an diese Regeln halten. Reale Wahrscheinlichkeitsauswahlen in der Sozialforschung jedoch werden auf der letzten Stufe meist von Interviewern durchgeführt, denen je nach Verfahren unterschiedliche Verhaltensspielräume entweder explizit zugestanden oder mangels hinreichender Kontrollen zumindest nicht verwehrt werden. Abhängig vom faktisch vorhandenen Spielraum benachteiligen Interviewer die schwerer erreichbaren Personen aufgrund des relativ höheren Aufwandes zur Kontaktaufnahme. Die dabei erzeugte "Wahrscheinlichkeitsauswahl" ist verzerrt: Schwerer erreichbare Personen fehlen systematisch in umso größerem Umfang, je größer der Spielraum der Interviewer bei der Auswahl war.

Diese Annahmen wurden anhand der 12 ALLBUS-Erhebungen von 1980-2000 überprüft: Bei direkter Auswahl von Personen aus Einwohnermeldeamtslisten (geringer Spielraum) fanden wir einen viel höheren Anteil der relativ schwer erreichbaren Personen als in

Abstract

Looking for hidden bias in random samples, we re-analyzed 12 ALLBUS Surveys (Germany's GSS, 1980-2000). In these surveys people had been selected by varying procedures: the interviewers had been allowed some amount of scope for own activities and thereby (hypothesis) induced bias of differing severeness.

The sampling bias is measured by means of internal criteria. By definition, the total population as well as the random sample are restricted to the end that known population parameters arise, in this case a total population of couples (one men, one women, same household, at least one child up to 14 years old), where the male person is working full-time outside the household and the female is not. As a result of these restrictions, an unbiased random sample is expected to consist of 50% women exactly, in spite of them being more easily accessed than their male partners.

Comparing the results of existing random samples (ALLBUS) to the expected values, the relative frequency of women differs systematically: in cases of sampling procedures

haushaltsbezogenen Auswahlen wie z. B. nach dem ADM-Design (relativ großer Spielraum).

Die Spielräume werden als partielle Rangordnung aus den Methodenberichten der 12 ALLBUS-Erhebungen abgeleitet. Die Auswahlverzerrungen zeigen sich durch den Grad der Abweichung von bekannten Verteilungsparametern, die definitorisch allein durch die Abgrenzung einer geeigneten Teilgesamtheit erzeugt werden.

prescribing the interviewers behavior in detail, we found a significant higher number of males (difficult to access) compared to samples which allowed more freedom.

1 Einführung: Befragungsbedingungen, Interviewerverhalten und Auswahlresultate¹

Wahrscheinlichkeitsauswahlen folgen zunächst theoretischen Konzepten. Versuche zur praktischen Umsetzung dieser Konzepte zielen auf eine möglichst große Annäherungen an das Ideal. Jedoch verfolgen sie dieses Ziel nicht bedingungslos, sondern müssen auch die entstehenden Kosten und den Bedarf an Zeit im Zusammenhang mit den verfügbaren Mitteln beachten.

Wie bei allen Kompromissen geraten Forscher auch hier in ein Spannungsfeld zwischen den erwünschten Eigenschaften der zu erhebenden Daten wie z. B. der Kalkulierbarkeit des Risikos bei ihrer Übertragung auf angebbare Grundgesamtheiten auf der einen und den verfügbaren Ressourcen an Geld und Erhebungszeit auf der anderen Seite.

Dilemmata der hier angedeuteten Art sind allgemein bekannt. Völlig aufzulösen wären sie nur unter den unrealistischen Bedingungen unbegrenzter Zeit- und Geldmittel sowie beliebiger Möglichkeiten zur Einflussnahme auf die Auswahl- und Erhebungsprozesse. Große Erhebungsinstitute bieten deshalb gezielt Mischlösungen² an, die auf die tatsächlichen oder vermeintlichen Bedürfnisse vermuteter Auftraggeber zugeschnitten sind. Damit werden einige Elemente einer

1 Der Aufsatz wurde ursprünglich eingereicht und angenommen für eine geplante Festschrift, die 2003 zum 70. Geburtstag von Erwin K. Scheuch erscheinen sollte. Diese Festschrift wird posthum nicht mehr erscheinen. Der Aufsatz bleibt dem Andenken an Erwin K. Scheuch gewidmet. Für Anmerkungen zu einer früheren Fassung danke ich Karl-Heinz Reuband und einem anonymen Gutachter.

2 Z. B. Auswahlen nach dem Random Route-Verfahren, bei denen erleichterte Auswahlbedingungen für die Interviewer zu Zeit- und Kosteneinsparungen führen, mögliche Auswahlverzerrungen dadurch aber mithilfe einer zusätzlichen Quotensteuerung begrenzt werden sollen.

Wahrscheinlichkeitsauswahl realisiert und gleichzeitig andere bewusst aufgegeben. Bei unbestrittener Notwendigkeit solcher Kompromisse erscheint die mangelnde Transparenz ihrer Folgen als problematisch.

Als potentiell besonders zeit- und kostenaufwändig erweist sich bei einer Wahrscheinlichkeitsauswahl der Umgang mit der unterschiedlichen Erreichbarkeit der Auswahleinheiten.³ Zur Illustration seien dazu zwei extreme Beispiele mit besonders hohem bzw. niedrigem Zeit- und damit Kostenaufwand beschrieben. Gemeinsam sei den Beispielen ein mehrstufiges Auswahlverfahren, das auf der 1. Stufe aus einer Gebietsauswahl nach dem ADM-Stichprobendesign (ADM/AG.MA 1999) besteht. Auf den folgenden Auswahlstufen nimmt ein Interviewer innerhalb des ihm zugewiesenen Gebiets von einem zufällig festgelegten Anfangspunkt aus einen Zufallsweg („random route“), dem entlang zunächst Häuser und darin jeweils Wohnungen bzw. Privathaushalte nach vorher festgelegten Regeln in eine bestimmte Reihenfolge gebracht werden. Aus dieser geordneten Reihe von Haushalten werden dann aufgrund einer systematischen Wahrscheinlichkeitsauswahl zunächst Haushalte und schließlich aus diesen Haushalten jeweils eine durch Wahrscheinlichkeitsauswahl (mittels Schwedenschlüssel) ermittelte Zielperson zur Befragung ausgewählt.

Probleme bei der Realisierung können auftreten,

- (a) weil entweder im ausgewählten Haushalt keine Person angetroffen wird oder
- (b) die Mitwirkung bei der Aufstellung einer Liste aller zum Haushalt gehörenden Personen aus der Grundgesamtheit verweigert wird,
- (c) weil die innerhalb des Haushalts aus der zuvor erstellten Liste ausgewählte Person nicht anwesend ist oder
- (d) die Zielperson die Befragung verweigert.

Die Realisationsversuche sollten nach allgemein anerkannten Regeln mit der ersten Verweigerung (b) oder (d) enden. Gestaltungsoptionen bleiben damit nur noch für die Fälle der Nicht-Erreichbarkeit (a) und (c). Von der Art der Behandlung dieser Fälle hängt der zeitliche und finanzielle Aufwand bei der Erhebung ebenso ab wie der Grad der Annäherung der realisierten Stichprobe an eine Wahrscheinlichkeitsauswahl.

(1) Variante mit hohen Kosten und relativ großer Näherung an eine Wahrscheinlichkeitsauswahl: Zur Durchführung von maximal 10 Befragungen wählt der Interviewer nach den oben genannten Regeln 10 Haushalte und darin je eine Person aus und versucht, notfalls unter Wiederholung von Kontaktversuchen, möglichst

3 Zum Problem der Erreichbarkeit vgl. Schnell (1997, Kap. 5).

viele der Befragungen zu realisieren. Ersatzadressen für nicht realisierbare Befragungen stehen ihm nur unter restriktiven Bedingungen (für stichprobenneutrale Ausfälle) zur Verfügung.

(2) Variante mit relativ geringen Kosten und starker Abweichung von einer Wahrscheinlichkeitsauswahl: Zur Durchführung von maximal 10 Befragungen darf der Interviewer 30 Haushalte bzw. Personen (siehe (1)) auswählen. Die Reihenfolge der Kontaktversuche steht ihm ebenso frei wie Wiederholungen dieser Versuche.

Das letztgenannte Verfahren erlaubt eine relativ *kosteneffiziente* Durchführung der Befragungen, weil zeitraubende Wiederholungen von Kontaktversuchen und sogar die Wahrnehmung in Aussicht gestellter Befragungen mit ausgewählten, aber augenblicklich abwesenden Haushalts-Mitgliedern nur realisiert werden müssen, wenn die angestrebten 10 Interviews nicht auf anderem, zeitsparenderem Wege erzielt werden können. Es gibt Gründe zu befürchten, dass in diesem Fall die tatsächliche Auswahl im Hinblick auf eine unterschiedliche Erreichbarkeit der Zielpersonen gegenüber einer reinen Wahrscheinlichkeitsauswahl systematisch verzerrt ist.⁴

Das Ziel dieser Arbeit ist eine exemplarische Untersuchung der Zusammenhänge zwischen den jeweils festgelegten Auswahlregeln und dem Umfang der Auswahlverzerrungen. Damit wird ein wichtiger Grund für die systematischen Verzerrungen bei der Wahrscheinlichkeitsauswahl im Prozess der Auswahlhandlungen gesucht (Scheuch 1974: 69). Als empirische Basis dienen die 12 ALLBUS-Erhebungen aus den Jahren 1980-2000.

In Abschnitt 2 werden zunächst die Methodenberichte dieser 12 Erhebungen mit dem Ziel analysiert, den Umfang des Spielraums der Interviewer bei der Auswahl der zu befragenden Personen und bei der Realisierung der Befragungen mit den ausgewählten Personen einzustufen.

Abschnitt 3 beschreibt ein Verfahren zur Kennzeichnung von Auswahlverzerrungen anhand *interner Kriterien*, das keine empirischen Vorkenntnisse über die Grundgesamtheit benötigt. Damit wird die Anzeige spezieller Verzerrungen möglich, die das Interviewerverhalten im eigenen Interesse bei der Auswahl und Befragung unterschiedlich leicht erreichbarer Personen erzeugt.

Entsprechende Hypothesen über den Zusammenhang zwischen unterschiedlich großen Spielräumen der Interviewer und dem Umfang dadurch erzeugter Auswahlverzerrungen werden im abschließenden Abschnitt 4 mit den Daten der 12 ALLBUS-Erhebungen überprüft.

4 Vgl. dazu die Diskussion um die Berechnung von Korrektur-Gewichten zur Antreffbarkeit in den ALLBUS-Methodenberichten u. a. von 1991 (Abschnitt 6.3.1 Gewichtungen) und 1992 (Abschnitt 4.2.3).

2 Varianten praktizierter Auswahlverfahren

In diesem Abschnitt beschreiben wir kurz diejenigen Aspekte der Stichprobenziehung verschiedener ALLBUS-Erhebungen, die für die hier angestrebte Analyse bedeutsam sind. Wie bereits einleitend dargestellt, handelt es sich dabei um die Regeln bei der Auswahl auf der zweiten und gegebenenfalls dritten Stufe, nämlich bei der entweder direkten Auswahl von Personen oder der Auswahl von zunächst Haushalten und anschließend Personen innerhalb der Haushalte.

Leider sind die Beschreibungen dieser Stufen des Auswahlprozesses in den Methodenberichten zu den ALLBUS-Erhebungen nicht ähnlich detailliert wie die Beschreibungen der Gebietsauswahlen auf der ersten Stufe. Vor allem fehlen solche Einzelheiten über die Rahmenbedingungen der Erhebung, die Rückschlüsse auf das mögliche Verhalten der Interviewer innerhalb ihrer Gestaltungsspielräume – und damit indirekt auch die Einschätzung der Gefahr von Regelverstößen – erlauben würden: Zu nennen ist hier einmal die Höhe der Zahlungen für ein realisiertes Interview sowie gegebenenfalls die Höhe zusätzlicher Zahlungen für wiederholte Kontaktversuche, ferner die Art, wie in den Instituten die Einhaltung der vorgeschriebenen Verfahrensweisen kontrolliert wird. Dieses betrifft sowohl die laufende Erhebung wie auch frühere Erhebungen des Instituts, die auf dem Umweg über die Kenntnisse der Interviewer als *internalisierte Kontrollen* wirken könnten.

2.1 Bezahlungsform der Interviewer und Folgen für die Auswahl

In aller Regel werden Interviewer in Deutschland durch die Institute auf der Basis von Stückkosten entlohnt, d. h. durch einen festen Betrag für jede realisierte Befragung. Diese Form der Entlohnung führt zu leicht kalkulierbaren Erhebungskosten und erfordert weder weitergehende Kontrollen des zeiteffizienten Verhaltens der Interviewer noch besondere Maßnahmen zur Kostenkontrolle während der Erhebung.

Andererseits tragen bei dieser Form der Entlohnung, d. h. ausschließlich über den Stücklohn pro Befragung, die Interviewer das alleinige Risiko für jeden zusätzlichen Aufwand, der zum Beispiel durch eine wiederholte Nicht-Erreichbarkeit von Haushalten oder der im Haushalt ausgewählten Befragungspersonen zustande kommt oder auch durch besonders lange Befragungszeiten, die in der Person des Befragten oder der besonderen Datenlage begründet sind. Es ist davon auszugehen, dass Interviewer diese finanziellen Bedingungen beachten und ihr Verhalten auch an der erwarteten Höhe der Entlohnung orientieren.

Wenn die Bezahlung der Interviewer ausschließlich auf dem Stücklohn beruht, ist ihre Entlohnung (*ceteris paribus*) am relativ höchsten, wenn die Befragung unmittelbar im Rahmen des Zufallsweges durch das Gebiet geschieht. Das gilt für die Aufnahme der Haushalts-Liste, die Auswahl des zu befragenden Haushalts, den ersten Kontakt mit dem Haushalt, die Aufnahme der Liste aller zur Grundgesamtheit gehörenden Haushalts-Mitglieder und die Auswahl der zu befragenden Person (z. B. mittels Schwedenschlüssel). Bezogen auf jede einzelne Erhebungseinheit kann diese – mit optimaler Entlohnung verbundene – Bedingungskonstellation von Interviewern nicht beeinflusst werden. Anders verhält es sich aber mit einer Reihe entsprechender Erhebungseinheiten auf dem Zufallsweg durch das Gebiet: Da die auszuwählenden Haushalte in der Regel im Erhebungsgebiet dicht beieinander liegen, kosten sequentielle Kontaktversuche wenig Zeit und alle in derselben Sequenz realisierten Befragungen können als optimal entlohnt gelten.

Bereits schlechter sind die Realisationschancen für Befragungen in einem unmittelbar auf die erste Sequenz folgenden zweiten oder dritten Durchgang durch das Gebiet, wenn aufgrund einiger bereits realisierter Befragungen weniger Erhebungseinheiten zur Auswahl stehen. Sehr viel ungünstiger wird die Bezahlung jedoch, wenn Kontakte mit Haushalten oder Befragungen bereits ausgewählter Zielpersonen an anderen Tagen nach wiederholtem Besuch des Erhebungsgebiets versucht werden. Die erwartete Entlohnung wird im Extremfall u. U. negativ, wenn an einem Ort nur einzelne Haushalte übrig bleiben oder aufgrund vorangehender Erfahrungen nur wenig Hoffnung besteht, in den ausgewählten Haushalten jemanden anzutreffen.

Interviewer, welche die erwartete Höhe ihrer Entlohnung verhaltenswirksam beachten, werden sich also mit unterschiedlicher Intensität um die Realisierung von Befragungen bemühen. Sie werden ihr Bemühen ganz einstellen,

- (a) wenn die erwartete Entlohnung eine Mindestgröße unterschreitet oder
- (b) wenn Alternativen zur Durchführung anderer Befragungen unter günstigeren Bedingungen innerhalb derselben Erhebung oder
- (c) in einer anderen Erhebung bestehen.

Zur erst- und letztgenannten Bedingung (a,c) liegen uns keine Informationen über die untersuchten Erhebungen vor. Wir müssen daher annehmen, dass sie zwischen den zu vergleichenden ALLBUS-Erhebungen nicht variieren. Es gibt aber Zweifel an dieser Annahme: Die Interviewerstäbe der mit der Erhebung beauftragten Institute unterscheiden sich deutlich hinsichtlich ihrer Zusammensetzung nach Geschlecht und Alter und sind nicht einmal innerhalb eines Instituts identisch über die Zeit (vgl. Methodenberichte zu den ALLBUS-Erhebungen).

Bedingung (b) dagegen kann anhand der vorliegenden Beschreibungen beurteilt und zum Vergleich zwischen den Erhebungen genutzt werden. Im folgenden Abschnitt wird deshalb zusammengestellt, in welchem Umfang innerhalb der festgelegten Regeln Spielräume der Interviewer bei der Auswahl der Haushalte und damit Gestaltungsmöglichkeiten bei der Festlegung potentiell zu kontaktierender Haushalte bestehen.

Grundsätzlich kann jeder Interviewer seine Tätigkeit mehr oder weniger intensiv betreiben und – bei Erwartung unzureichender Entlohnung – jederzeit ganz einstellen. In diesem Sinne müssen auch Ausschöpfungsraten der Bruttostichprobe nicht notwendig den Befragten („Verweigerungen“) angelastet werden, sondern könnten teilweise auch Folgen der Abbruchentscheidungen der Interviewer aufgrund unzureichender Bezahlung von Befragungen bei erwarteten Mehrfachkontakten sein.

Natürlich gibt es prinzipiell auch die Möglichkeit der Regelverstöße, in diesem Fall also der Kontaktierung eigentlich nicht ausgewählter Haushalte, die tatsächlich oder vermeintlich leichter erreichbar sind und/oder der Befragung eigentlich nicht als Zielpersonen in einem Haushalt ausgewählter Personen. Solche Verstöße sind nur unter zusätzlichen Kontrollen aufzudecken, deren Umfang jedoch innerhalb der jeweiligen ALLBUS-Erhebung nicht immer klar dokumentiert ist.

Zusätzliche Kontrollen wären aber auch erforderlich, wenn Interviewer eine ergänzende, der Befragungszeit entsprechende Entlohnung ihres Zeitaufwandes für mehrfache Kontaktversuche eines Haushalts oder für die wiederholten Besuche eines Haushalts zur Befragung der ausgewählten Zielperson erhielten. Andernfalls wären die Erhebungskosten nicht im Vorhinein kalkulierbar und (unredliche) Interviewer erhielten die Möglichkeit zur direkten Manipulation ihres Einkommens. Alternativen zum reinen Stücklohn für realisierte Interviews durch zusätzliche, den Zeitaufwand berücksichtigende Elemente werden deshalb in der Erhebungspraxis nur äußerst selten gewählt und kommen unter den hier analysierten ALLBUS-Erhebungen unseres Wissens nicht vor.

Die genannten Überlegungen zum rationalen Erwerbsverhalten der Interviewer fassen wir in einer Hypothese zusammen:

Die Auswahl von Haushalten und darin von Zielpersonen durch die Interviewer wird umso größere Selektionseffekte aufgrund unterschiedlicher Erreichbarkeit aufweisen, je größer der Spielraum der Interviewer bei der Auswahl ist.

Ungeprüfte bzw. vereinfachende, der Realität nicht immer entsprechende Annahmen sind dabei insbesondere:

- Die Interviewerstäbe in allen ALLBUS-Erhebungen sind gleich hinsichtlich ihrer zeitlichen Disposition (z. B. Halb-/Ganztagstätigkeit als Interviewer, Verteilung der Kontakt- und Befragungsversuche auf Tageszeiten).
- Die Interviewerstäbe in allen ALLBUS-Erhebungen sind gleich hinsichtlich ihrer Berufserfahrungen als Interviewer, ihrer professionellen Ethik usw.
- Die Entlohnungsformen (Verhältnis von Stücklohn für Befragung versus zusätzlicher Entlohnung für besonderen zeitlichen Aufwand) sind bei allen Erhebungen gleich.
- Bei allen Erhebungen wurden in gleichem Umfang die Einhaltung der Auswahlregeln kontrolliert bzw. es bestanden gleiche Erwartungen der jeweiligen Interviewerstäbe gegenüber entsprechendem Kontroll-Verhalten.

Sowohl die hier verfolgte Hypothese zum Verhalten der Interviewer wie leider auch Zweifel an der Richtigkeit der vereinfachenden Hilfsannahmen werden durch einige frühere Analysen gestützt.

Diese Analysen leiden allerdings genau wie unsere eigene Studie unter einem Mangel an Daten. Die eigentlich benötigten Informationen über die Form der Bezahlung – dies trifft insbesondere den Fall von wiederholten Versuchen zur Kontaktaufnahme – und über den genauen Ablauf der Interviewer-Einweisung und -Kontrolle liegen in keinem Fall vor. In nur wenigen Fällen gibt es zumindest Daten mit indirektem Bezug zu den hier diskutierten Fragen, die zudem – mit Ausnahme vielleicht der Studie von Alt/Bien/Krebs (1991) – nur entweder Eigenschaften der Erhebung eines Instituts oder nur Eigenschaften der Interviewer bzw. ihres Verhaltens betreffen. Eine Interpretation im Sinne der hier diskutierten Fragen ist dann wieder nur unter ganz ähnlichen wie den oben genannten oder unter dazu komplexeren Hilfsannahmen möglich.

Schnell (1997: 58-59) verweist in seiner umfangreichen Untersuchung des „Nonresponse in Bevölkerungsumfragen“ auf den Einfluss des Ermessensspielraums, der Interviewern bei der Auswahl der zu Befragenden eingeräumt wird. Bis zu diesem Zeitpunkt lagen nur wenige Studien mit Einwohnermeldeamts-Stichproben (EWM) vor, bei denen ein „kleinerer Einfluss des Interviewers auf die Auswahl der Zielperson und die Art der Ausfälle erwartet“ wird (Schnell 1997: 59). Sie zeigten eine deutliche Verminderung des Anteils der nicht-erreichbaren Personen gegenüber vergleichbaren Auswahlen nach dem meist gewählten ADM Design (Koch/Gabler/Braun 1994; Alt/Bien/Krebs 1991).

In seiner eigenen Analyse von Erhebungsberichten aus 300 Bevölkerungsumfragen, die im Zentralarchiv für empirische Sozialforschung archiviert sind, findet er starke Unterschiede in der gesamten Ausschöpfungsquote wie auch im Anteil der

Verweigerungen und der Ausfälle aufgrund Nicht-Erreichbarkeit sowohl *zwischen den Instituten* wie auch über die Zeit (Schnell 1997: 76-100). Leider werden damit die oben genannten Zweifel an der Tragfähigkeit unserer Hilfsannahmen weiter gestärkt.

Über die starke Variation der Ausschöpfungsergebnisse zwischen Erhebungsinstituten hinaus berichtet Schnell auch über – zumindest indirekte – Hinweise auf die Folgen unterschiedlicher Ermessensspielräume der Interviewer. Insbesondere die Bevölkerungsumfragen im Rahmen der Media-Analysen zeigen über Institutsgrenzen hinweg eine höhere Ausschöpfung bzw. einen geringeren Anteil an Verweigerern und Nicht-Erreichbaren als andere Umfragen mit kommerziellen oder wissenschaftlichen Auftraggebern. Mangels direkter Informationen über die Kontrollen im Prozessverlauf der Erhebungen können diese Ergebnisse zwar nicht eindeutig den hier behandelten Spielräumen der Interviewer zugeschrieben werden. Es gibt aber plausible Erklärungen, warum vor allem die Bevölkerungsumfragen der Media-Analyse mit einem relativ hohen Kontrollpotential im Erhebungsprozess verbunden sein könnten:

„Die Media-Analyse ist hoch standardisiert und von der wahrgenommenen Güte der Feldarbeit hängt die weitere Zukunft des Umfrageinstituts innerhalb der Media-Analyse ab.“ Deshalb „sollten die wiederholten Ergebnisse eines Instituts nicht zu stark von den Ergebnissen der anderen Institute abweichen.“ (Schnell 1997: 79).

Auch innerhalb der Erhebungen zur Media-Analyse stellt Schnell jedoch (trotz der besonderen Standardisierung) große Unterschiede bei der Verteilung der Zahl der Kontaktversuche bis zur Realisierung des Interviews zwischen den Instituten fest:

„Damit muss für die Media-Analyse festgehalten werden, dass der Einfluss der Erhebungsinstitute auf die berichtete Erreichbarkeit der Befragten größer ist als der Einfluss der Befragten. Entweder unterscheidet sich also das Kontaktverhalten der Interviewer oder die Art der Berichterstattung zwischen den Instituten.“ (Schnell 1997: 223).

An anderer Stelle geht Schnell anhand einzelner Erhebungen (ALLBUS 1980,1988; vgl. Schnell 1997: 229-236)⁵ mit entsprechend günstiger Datenlage über das Interviewer-Verhalten auch einigen der möglichen Ursachen für unterschiedliche Anteile der Nicht-Erreichbaren nach. So vergleicht er z. B. die insgesamt für den ALLBUS 1988 benötigte Feldzeit von 10 Wochen mit entsprechenden

5 Die in unserer Arbeit analysierten ALLBUS-Erhebungen enthalten ebenfalls diese beiden Datensätze von 1980 und 1988. Umgekehrt sind die von uns analysierten ALLBUS-Erhebungen fast sämtlich (d. h. bis 1994) auch in den von Schnell analysierten 300 archivierten Studien enthalten.

Feldzeiten auf der Mikroebene, d. h. der von einzelnen Interviewern an bestimmten Sample-Points für ihre Befragungen verbrachten Zeit (Schnell 1997: 229-231). In mehr als der Hälfte aller Fälle beschränkt sich diese individuelle Feldzeit an einem Sample-Point auf etwa 3 Tage.

Wiederholte Kontaktversuche werden darüber hinaus (nun auf Basis der Kontaktprotokolle des ALLBUS 1980 ermittelt) zum größten Teil in kurzem zeitlichen Abstand zum ersten Kontaktversuch unternommen: „2/3 der Fälle weisen eine Differenz von maximal 4 Stunden auf, nur ca. 7% der Fälle besitzen eine Differenz von mehr als 8 Stunden“ (Schnell 1997: 233).

Beide Ergebnisse sind im Sinne der hier verfolgten Hypothese als suboptimale Verhaltensweisen der Interviewer zu interpretieren, die möglicherweise durch strengere Vorgaben und Kontrollen der Erhebungsinstitute mit dem Ziel einer höheren Ausschöpfung verändert werden könnten.

Wir müssen aber nochmals darauf hinweisen, dass die Datenlage bei diesen letztgenannten Untersuchungen in positiver wie negativer Hinsicht komplementär ist zur Datenlage sowohl bei unserer eigenen wie auch der vorher genannten Untersuchung der 300 Erhebungsberichte durch Rainer Schnell (1997). Während dort die Variation der Ausschöpfung etc. nur auf der Makro-Ebene, d. h. über Erhebungen bzw. Institute geprüft wird, betreffen die zuletzt berichteten internen Verhältnisse in den ALLBUS-Erhebungen von 1980 und 1988 nur die Mikro-Ebene des Interviewerverhaltens unter den jeweils relativ konstanten Organisations- und Kontrollbedingungen allein einer Erhebung eines Instituts.

Etwas günstiger ist die Datenlage bei der abschließend diskutierten Untersuchung des Deutschen Jugendinstituts zum Thema ‚Partnerschaft und Familie heute‘ (Alt/Bien/Krebs 1991). Zur Prüfung der Konsequenzen unterschiedlicher Verfahren zur Wahrscheinlichkeitsauswahl wurde die Stichprobe teilweise nach dem ‚random route-Verfahren‘ (N=7.032) und teilweise aus den EWM-Registern (N=3.011) gezogen. Im ersten Schritt des mehrstufigen Auswahlverfahrens folgten beide Teilstichproben dem ADM-Design. Für die hier behandelten Fragen ist die Studie vor allem deshalb interessant, weil dasselbe Erhebungsprogramm zeitgleich in vergleichbaren Regionen vom gleichen Erhebungsinstitut mit zwei unterschiedliche Auswahlverfahren durchgeführt wurde.

Vergleiche der Verteilungen nach Geschlecht und Erwerbstätigkeit zeigen deutliche Unterschiede zwischen beiden Teilstichproben, wobei die EWM-Stichprobe besser mit den Ergebnissen des Mikrozensus übereinstimmt als die random route Stichprobe. Letztere enthält viel weniger Männer und Erwerbstätige. Als Gründe dafür vermuten Alt, Bien und Krebs, dass den Interviewern im Random Route-Ver-

fahren zumindest faktisch ein größerer Ermessensspielraum zur Verfügung stand. Folgerichtig suchen sie nach „Indikatoren für den unterschiedlichen Zugang zum Feld in den beiden Stichprobenverfahren“ (Alt/Bien/Krebs 1991: 67).

Während in der Random Route-Stichprobe die meisten Befragungen beim ersten Kontaktversuch realisiert werden, wird dieser Gipfel bei der EWM-Stichprobe erst im zweiten Versuch erreicht. Nach beiden Verfahren werden im ersten Versuch (gegenüber dem Mikrozensus) mehr Frauen und Nicht-Erwerbstätige befragt, also „leicht Erreichbare“ (S. 69). In beiden Stichproben wird diese Verzerrung mit zusätzlichen Kontaktversuchen geringer. Diese Tendenz ist aber in der EWM-Stichprobe stärker. Außerdem werden in der EWM-Stichprobe insgesamt viel mehr Befragte nach zwei oder mehr Kontaktversuchen befragt als in der Random Route-Stichprobe.

Alt, Bien und Krebs interpretieren den größeren Ermessensspielraum der Interviewer beim Random Route-Verfahren als Möglichkeit zur ‚Manipulation‘, verzichten aber aus guten Gründen darauf, den Inhalt möglicher Manipulationen genauer festzulegen. Im Rahmen unserer Arbeit folgen wir vor allem der Annahme einer (zeit-) ökonomischen Orientierung der Interviewer. Als Manipulation in diesem Sinne gelten vor allem die Entscheidungen der Interviewer über den jeweils nächsten Kontaktversuch. Im Random Route-Verfahren bestehen weniger Möglichkeiten zur Kontrolle solcher Entscheidungen. Verzerrungen entstehen nach dieser engeren Deutung nicht in erster Linie dadurch, dass die Interviewer „das Interview mit der Person durchführen, die sie antreffen“ (Alt/Bien/Krebs 1991: 69), sondern dass sie bei mangelndem Erfolg im aktuellen Versuch – vielleicht entgegen der Regeln – häufig keinen weiteren Kontaktversuch mehr unternehmen. Natürlich können wir auch weitergehende Manipulationen nicht ausschließen.

2.2 Praktizierte Auswahlregeln auf der Ebene von Gebietseinheiten

In diesem Abschnitt werden alle ALLBUS-Erhebungen 1980–2000 nach dem Gesichtspunkt geordnet,

- in welchem Umfang den Interviewern innerhalb der Auswahlregeln Spielräume zugestanden wurden und
- in welchem Umfang Kontrollen der von ihnen durchgeführten Auswahl-schritte möglich waren. Offen bleibt dabei jedoch, inwieweit Kontrollen tatsächlich durchgeführt bzw. von den Interviewern der Institute erwartet wurden.

Die Beurteilung von Spielräumen und Kontrollmöglichkeiten stützt sich ausschließlich auf die Methoden-Berichte der ALLBUS-Erhebungen. Einzelheiten über alle 12 Erhebungen sind in schematischer Form in einem gesonderten Anhang wiedergegeben, der vom Autor bezogen werden kann. Hier beschränken wir uns auf die zusammenfassende Einordnung der Erhebungen in 4 Gruppen:

(1) ALLBUS-Erhebungen 1994/1996/2000

Den geringsten Spielraum sehen wir in den ALLBUS-Erhebungen mit einer Wahrscheinlichkeitsauswahl der Zielpersonen aus den Verzeichnissen der Einwohnermeldeämter in zuvor ausgewählter Gebietseinheiten. Innerhalb der Regeln ergeben sich hier nur Spielräume der Interviewer bei den Entscheidungen über die Wiederholung von Kontaktversuchen bei zunächst nicht erreichten Zielpersonen. Auch außerhalb der Regeln sind die Spielräume gering bzw. nur bei sehr ernsten Verstößen zu realisieren, wenn nämlich andere als die ausgewählten Personen befragt oder die Befragungen ganz gefälscht werden.

(2) ALLBUS-Erhebungen 1980/1982/1990/1998

Einen größeren Spielraum sehen wir in jenen ALLBUS-Erhebungen, bei denen die Auswahl der Haushalte vor der eigentlichen Erhebung erfolgte: Den Interviewern wurde eine feste Zahl von Haushaltsadressen zur möglichst vollständigen Realisierung je eines Interviews übergeben. Innerhalb der Haushalte folgte die Auswahl zwar festen Regeln (Schwedenschlüssel), war aber vom Interviewer selbst vorzunehmen und deshalb im Ergebnis dem Institut vorher nicht bekannt.

Unterschiede zu den Erhebungen der Gruppe (1) liegen deshalb aus unserer Sicht vor allem in der geringeren Kontrollierbarkeit der Auswahl zu befragender Personen. Bei strikter Einhaltung der Regeln wären beide Gruppen demgegenüber gleich einzustufen im Umfang des Spielraums bei der Auswahl: Es steht kein *Überschuss* an Adressen zur Verfügung. Die Entscheidungen der Interviewer über die erneute Kontaktierung zunächst nicht erreichter Haushalte oder – innerhalb derselben – beim ersten Kontakt mit dem Haushalt nicht anwesender Zielpersonen bestimmen den Grad der Ausschöpfung der Stichprobe.

(3) ALLBUS-Erhebungen 1984/1986/1988

Bei dieser Gruppe von Erhebungen wurde sowohl die Auswahl von Haushalten wie auch der zu befragenden Personen von den Interviewern vorgenommen. Auf einem vorgeschriebenen Weg durch das Erhebungsgebiet (random route) waren Haushalte in einer bestimmten Reihenfolge aufzunehmen und von diesen Haushalten eine bestimmte Zahl formal bestimmter Haushalte (z. B. auf bestimmten Positionen der

Liste von Haushalten, die vorher auf dem Formular gekennzeichneten waren) zu kontaktieren. Innerhalb der Haushalte erfolgte die Auswahl wieder aufgrund des Schwedenschlüssels.

Bei strikter Einhaltung der Regeln besteht auch in dieser Gruppe von Erhebungen weder ein Spielraum bei der Auswahl von Haushalten und Befragungspersonen noch ein Überschuss an potentiell zu befragenden Personen. Die Kontrollmöglichkeiten sind jedoch gegenüber den beiden erstgenannten Gruppen (1) und (2) noch weiter vermindert, da den Instituten zunächst weder die auszuwählenden Haushalte noch die darin anzusprechenden Befragungspersonen bekannt sind.

(4) ALLBUS-Erhebungen 1991/1992

Die Auswahl von Haushalten und darin Befragungspersonen folgt im Prinzip denselben formalen Regeln wie bei den Erhebungen der Gruppe (3) mit einer Ausnahme: Durch den festgelegten Zufallsweg (random route) im Erhebungsgebiet wird wieder nach bestimmten Regeln eine Liste von Haushalten mit einer Auswahl derselben erstellt. Diese Auswahl von Haushalten ist aber größer als die Zahl maximal zu realisierender Interviews. Wenn z. B. 6 Interviews in 15 ausgewählten Haushalten zu realisieren sind, so steht den Interviewern ein gewisser Überschuss an Haushalten zur Verfügung, aus denen Sie (innerhalb der Regeln!) frei wählen können. Dieses reduziert offensichtlich die Notwendigkeit – oder die wie immer sanktionierte Erwartung der Institute an ihre Interviewer – zur Wiederholung von Kontaktversuchen und spart damit Zeit und Geld.

Mit diesem Versuch der Rangordnung von ALLBUS-Erhebungen nach dem Grad der Spielräume, die Interviewern bei der Auswahl von Haushalten und/oder zu befragenden Personen offen stehen, verbinden wir entsprechend der oben formulierten Hypothese die Annahme von Folgen für die dabei entstehenden Wahrscheinlichkeitsauswahlen: *Schlecht erreichbare* Personen der Grundgesamtheit müssten gegenüber *gut erreichbaren* Personen um so häufiger ausfallen, je weniger sich Interviewer im Fall nicht erreichter Haushalte oder Befragungspersonen um die Wiederholung von Kontakten bemühen (müssen). Die Vermutung ist dabei, dass mit größeren Spielräumen – hier steigend von Gruppe (1) bis Gruppe (4) – auch der Druck zur Wiederholung von Kontaktversuchen abnimmt. Relativ schlechter erreichbare Personen müssten deshalb – ceteris paribus – gegenüber besser erreichbaren Personen besonders häufig in der Gruppe (4) und besonders selten in der Gruppe (1) ausfallen.

Aufgrund dieser Sichtweise richtet sich das Interesse also auf den unterschiedlichen Umfang systematischer Ausfälle in verschiedenen, von Wahrscheinlichkeitsauswahlen repräsentierten Teilpopulationen.

3 Die Untersuchung von Auswahlverzerrungen: Abweichungen von den erwarteten Verteilungen

In diesem Abschnitt wird ein Verfahren zur Beurteilung der Ausfallraten unterschiedlich gut erreichbarer Personen in Wahrscheinlichkeitsauswahlen beschrieben. Diesem Ziel liegt ein allgemeineres Problem zugrunde: Wie sind Ausfälle in Wahrscheinlichkeitsauswahlen zu analysieren, wenn man die Grundgesamtheit nicht vorher kennt?

Die Idee für eine – teilweise – Lösung dieses Problems haben wir an anderer Stelle beschrieben (Sodeur 1997). Die dabei verfolgte Lösungsstrategie ist auch nicht neu: Statt nach der offensichtlich schwierigen Lösung für ein Problem sucht man nach einer maßvollen Änderung des Problems mit dem Ziel, es dadurch lösbar zu machen und gleichzeitig wichtige Beiträge zur Lösung des ursprünglichen Problems zu leisten.

Dieser Strategie folgend wird anstelle der direkten Beurteilung einer Wahrscheinlichkeitsauswahl aus einer gegebenen Grundgesamtheit der Versuch unternommen, eine veränderte, das heißt in aller Regel eingeschränkte Grundgesamtheit zu definieren, über die man *infolge der getroffenen Einschränkung* wichtige Eigenschaften kennt. Diese Eigenschaften nutzt man anschließend zur Beurteilung der nachträglich auf gleiche Weise eingeschränkten Wahrscheinlichkeitsauswahl. Hilfreich wird dieses Verfahren vor allem dann, wenn mit der getroffenen Einschränkung von Grundgesamtheit und Wahrscheinlichkeitsauswahl keine für den Beurteilungszweck wesentlichen Eigenschaften ausgeschlossen werden.

In dieser Arbeit (wie in der Bezugsarbeit) definieren wir als neue, *eingeschränkte Grundgesamtheit* alle Personen, die mit Partnern des jeweils anderen Geschlechts im selben Haushalt zusammenwohnen, wobei auch die Partner zur Grundgesamtheit gehören müssen. Dieses ist nicht die einzig mögliche Form einer geeigneten Einschränkung der Grundgesamtheit. Sie hat aber den Vorteil, dass eine für viele inhaltlich begründete Fragen bedeutsame Grundgesamtheit übrig bleibt. Bei den ALLBUS-Erhebungen, auf die sich diese Untersuchung bezieht, bleiben damit (zunächst) rund zwei Drittel aller Personen der ursprünglichen Grundgesamtheit erhalten. Ausgeschlossen sind nur Befragte, die nicht mit einem festen Partner des anderen Geschlechts im selben Haushalt wohnen oder deren Partner nicht zur Grundgesamtheit gehört. Tabelle 1 gibt Aufschluss über die Fallzahlen der ursprünglichen Stichproben aller ALLBUS-Erhebungen von 1980 bis 2000 und über die Folgen der schrittweisen Eingrenzung auf eine Teilstichprobe mit den gewünschten Beschränkungen.

Tabelle 1 Abgrenzung der ausgewählten Teilstichprobe

ALLBUS Jahr	Stichprobe Gesamt West	Befragte ohne Partner	Partner nicht im Haushalt	Befragte mit Partner im Haushalt	Partner < 18 oder gleiches Geschlecht	untersuchte Teilstichprobe
1980	2955	1102	9	1844	65	1779
1982	2991	1178	0	1813	40	1773
1984	3004	1220	22	1762	26	1736
1986	3095	928	154	2013	7	2006
1988	3052	1056	134	1862	12	1850
1990	3051	984	195	1872	26	1846
1991	1514	469	77	968	5	963
1992	2400	723	139	1538	8	1530
1994	2342	612	118	1612	13	1599
1996	2402	565	147	1690	11	1679
1998	2212	758	89	1365	9	1356
2000	2036	526	142	1368	24	1344

Die auf diese Weise eingeschränkte Grundgesamtheit hat eine *definitorisch erzeugte Eigenschaft*: Sie besteht – wie jedes einzelne der sie konstituierenden, geschlechtsheterogenen Paare – aus genau 50% Frauen und 50% Männern. Gleiches gilt für alle Untermengen der Grundgesamtheit, soweit sie nicht aufgrund von Eigenschaften definiert werden, die mit dem Auswahlprozess zusammenhängen. Mit dem Erwartungswert eines Anteils von 50% Frauen an der Wahrscheinlichkeitsauswahl oder vielen Untermengen dieser Wahrscheinlichkeitsauswahl ist also ein allgemeines Kriterium zur Beurteilung der Auswahl bekannt.

Jedoch stellt eine derart abgegrenzte Wahrscheinlichkeitsauswahl nicht immer eine unverzerrte Auswahl der geschlechtsheterogenen Paare mit entsprechenden Eigenschaften dar, wie Günther Rösch (Büro für Erhebungsdesign und Datenanalyse, Frauenberg) kürzlich zu Recht in einem mündlichen Tagungsbeitrag kritisiert hat. Den meisten der hier analysierten ALLBUS-Erhebungen liegt eine mehrstufige Auswahl zugrunde (vgl. Abschnitt 2), auf deren letzter Stufe aus allen zur Grundgesamtheit gehörenden Haushaltsmitgliedern genau eines ausgewählt wird. Setzt sich ein Haushalt nur aus dem Paar und Kindern unter 18 Jahren (die nicht zur Grundgesamtheit gehören) zusammen, so wird auf der letzten Auswahlstufe eine von 2 Personen ausgewählt: Damit ist immer einer der beiden Partner dabei. In Haushalten mit drei zur Grundgesamtheit gehörenden Personen (z. B. neben dem Paar aus einem bereits volljährigen Kind oder einem Elternteil), so wird die zu befragende Person unter drei Personen gewählt und das Paar ist nur in zwei von drei

Fällen beteiligt. Entsprechendes gilt auch für größere Haushalte (immer bezogen nur auf die zur Grundgesamtheit gehörenden Personen): Bei 4 Personen im Haushalt fällt die Auswahl der zu befragenden Person in 2 von 4 Fällen, bei 5 Personen in 2 von 5 Fällen (usw.) auf einen der beiden Partner. Je größer also der Haushalt ist, mit umso geringerer Wahrscheinlichkeit wird ein Mitglied des betreffenden Paares für die Befragung ausgewählt und damit im Rahmen der oben beschriebenen eingeschränkten Wahrscheinlichkeitsauswahl berücksichtigt. Will man auch diese Verzerrung ausschließen, dann kann man die Analyse entweder unter Konstanthalten der Haushaltsgröße weiter beschränken, also hier auf Haushalte mit genau zwei zur Grundgesamtheit gehörenden Personen, nämlich den beiden Partnern. Oder man muss die Auswahlverzerrung in den größeren Haushalten durch ausgleichende Gewichtung korrigieren. Eine der beiden Korrekturen wird immer dann erforderlich sein, wenn der Verdacht besteht, dass die jeweils untersuchten Sachverhalte von der Haushaltsgröße zumindest teilweise mitbestimmt werden könnten. In der vorliegenden Arbeit verzichten wir auf Korrekturen dieser Art und beschränken uns auf den Hinweis, dass beide Versionen einer möglichen Korrektur nur zu sehr unwesentlich veränderten Ergebnissen führen.

Hinsichtlich der Haushaltsgröße folgen die Unterstichproben also im Prinzip Verzerrungen, die bereits in allen Gesamtstichproben mit Haushaltsbezug angelegt sind. Daneben gibt es aber auch andere Abweichungen zwischen der Gesamtstichprobe und der auf Paare bezogenen Teilstichprobe. Natürlich ist in der Teilstichprobe (nur Paare!) der Anteil der Verheirateten viel höher und der Anteil der Ledigen viel geringer als in der Gesamtstichprobe. Aber, etwas weniger offensichtlich, ist das Durchschnittsalter in der Teilstichprobe etwas höher (46,66 gegenüber 46,24 Jahre) und die Streuung etwas geringer als in der Gesamtstichprobe, was insbesondere an der geringeren Häufigkeit der jüngeren Jahrgänge unter den Befragten aus Paaren im gemeinsamen Haushalt liegt. Ebenfalls vor allem bei den jüngeren Personen haben unterschiedlich lange Ausbildungszeiten Auswirkungen auf die Paarbildung. So kommen in der Teilstichprobe wesentlich mehr Hauptschüler (56,5 gegenüber 50,1%) und weniger Personen mit höherem Schulabschluss (41,1 gegenüber 46,2%) vor als in der Gesamtstichprobe. Diese Unterschiede setzen sich fort bis zur jetzigen bzw. letzten beruflichen Stellung, wo Befragte aus den Paaren der Teilstichprobe seltener als Befragte aus der Gesamtstichprobe keine berufliche Position oder Ausbildungsposition nannten (5,7 gegenüber 10,7%).

Schließlich benötigen wir zur Prüfung der am Ende des vorangehenden Kapitels genannten Vermutungen nicht Erwartungswerte für den Anteil von Frauen oder Männern in der Wahrscheinlichkeitsauswahl, sondern entsprechende Erwar-

tungswerte für den Anteil relativ leicht gegenüber schwer erreichbarer Personen, um daran den Anteil der Ausfälle dieser Gruppen auf dem Weg von der Brutto- zur Nettostichprobe zu messen.

Ein entsprechendes Beurteilungskriterium gewinnt man durch Kombination unterschiedlicher Erreichbarkeiten mit dem – bei der hier gewählten Grundgesamtheit indikativen – Geschlecht der Befragten. Dazu muss man z. B. Paare auswählen, bei denen die Männer jeweils relativ schlecht und die Frauen relativ gut zu erreichen sind. Unter Anwendung von *Hilfstheorien* nimmt man z. B. Paare, bei denen der Mann berufstätig und die Frau nicht oder höchstens halbtags berufstätig ist. Die Unterschiede in den Erreichbarkeiten beider Partner lassen sich unter diesen Bedingungen verschärfen, wenn man die Auswahl der Paare zusätzlich einschränkt auf die Zugehörigkeit (kleiner) Kinder zum Haushalt und damit auf Anwesenheitsanfordernisse des nicht oder relativ wenig berufstätigen Partners.

Festzuhalten ist: die eben genannten Eigenschaften zur parallelen Beschränkung von Grundgesamtheit und Wahrscheinlichkeitsauswahl haben nichts mit dem Auswahlprozess zur Wahrscheinlichkeitsauswahl zu tun. Sie ändern deshalb auch nichts an dem Erwartungswert von 50% Frauen (bzw. Männern) in der Unterstichprobe. Die zusätzliche Beschränkung der Grundgesamtheit um bestimmte Konstellationen der Berufstätigkeit der jeweiligen Partner sowie um die Zugehörigkeit von Kindern zu ihrem Haushalt hat jedoch einen Preis (vgl. Kohler 2007: 59). Grundgesamtheit und zugehörige Wahrscheinlichkeitsauswahl werden dadurch weiter eingeschränkt, das zur Lösung ausgewählte Problem entfernt sich noch etwas mehr vom ursprünglichen Problem:

Angestrebt wird die Klärung der Zusammenhänge zwischen Auswahlverfahren, Erreichbarkeiten der Personen aus der Grundgesamtheit und ihrer tatsächlichen Aufnahme in die Befragung. Die Untersuchung beschränkt sich jedoch auf nur einen Teil der Grundgesamtheit, nämlich auf solche Paare im gleichen Haushalt, von denen die Männer aufgrund sehr spezieller Kriterien als relativ schlecht und die Frauen als relativ gut erreichbar eingestuft werden.

Definiert man z. B. eine Grundgesamtheit von geschlechtsheterogenen Paaren, bei denen der Mann ganztags berufstätig (relativ zur Frau schlecht erreichbar) ist und die Frau nicht oder höchstens halbtags berufstätig (gut erreichbar) ist und mindestens ein Kind bis zum Alter von 14 Jahren im Haushalt wohnt, so reduziert sich die Grundgesamtheit auf durchschnittlich 18,3% der ursprünglichen Größe⁶ oder in absoluten Zahlen: Von ursprünglich in den Allbus-Erhebungen der Jahre

6 Geschätzt anhand der entsprechenden Teilstichproben über alle 12 ALLBUS-Erhebungen 1980-2000 im Erhebungsgebiet West.

1980–2000 (im Erhebungsgebiet West) erfassten 31.054 Befragten gehören 19.461 zu den ausgewählten Paaren und darunter erfüllen 5.679 die genannten Bedingungen von Berufstätigkeit und Kindern.⁷

Im folgenden Abschnitt wird dieses Kriterium zur Prüfung der im Abschnitt 2 vermuteten, systematischen Auswahlverzerrungen benutzt. In einer Wahrscheinlichkeitsauswahl aus der eben beschriebenen Grundgesamtheit müssten Frauen und Männer ihre jeweiligen Paare mit gleicher Wahrscheinlichkeit *als Befragte* vertreten.

Abweichungen des Anteils befragter Frauen vom Erwartungswert (50%) indizieren also potentielle Verzerrungen der Auswahl, oder genauer: Aufgrund unterschiedlicher Erreichbarkeiten von Frauen und Männern in der oben definierten Grundgesamtheit müsste der Anteil der befragten Frauen einer verzerrten Wahrscheinlichkeitsauswahl systematisch und überzufällig höher als 50% liegen. Diese systematische Abweichung vom Erwartungswert müsste ferner mit dem Spielraum wachsen, der den Interviewern bei der Auswahl zugestanden und mutmaßlich in ihrem eigenen Interesse genutzt wird – entgegen den Regeln für die Erzeugung einer Wahrscheinlichkeitsauswahl.

4 Beurteilung der Ergebnisse

4.1 Vergleich der Rangordnungen nach Spielraum der Interviewer und Anteil befragter Frauen (leicht erreichbare Partner)

Tabelle 2 gibt Auskunft über den Umfang der Abweichungen im Anteil der befragten Frauen in den ausgewählten Untergruppen der 12 ALLBUS-Erhebungen. Spalte 2 enthält die Rangordnung der 12 Erhebungen nach dem Spielraum, der den Interviewern nach unserer Einschätzung bei der Wahrscheinlichkeitsauswahl der zu befragenden Personen zugestanden wurde (vgl. Abschnitt 2), Spalte 3 den Anteil der befragten Frauen (der Erwartungswert ist 50%!), und Spalte 5 die in z-Werte (d. h. mit dem Mittelwert 0 und der Standardabweichung 1) umgerechneten Anteile.

7 Beschränkt man sich zusätzlich aufgrund der oben diskutierten Auswahlverzerrungen bei unterschiedlichen Haushaltsgrößen auf Haushalte mit nur zwei erwachsenen Personen, so reduziert sich die Zahl der Paare weiter auf 5.040. Die folgenden Ergebnisse werden davon jedoch nicht nennenswert berührt.

Tabelle 2 Anteil befragter Frauen aus allen geschlechtsheterogenen Paaren, bei denen der Mann ganztags und die Frau höchstens halbtags berufstätig ist
(Nur Haushalte mit mindestens einem Kind unter 15 Jahren)

ALLBUS	Rang	% Frauen	N	z-Werte	Institut
1994	1	44,92	443	-2,1496	Infratest (Burke), München
1996	1	46,22	476	-1,6541	Infratest (Burke), München
2000	1	48,62	362	-0,5253	Infratest (Burke), München
1980	2	55,04	556	2,3890	GETAS, Bremen
1982	2	58,20	567	3,9587	GETAS, Bremen
1990	2	57,12	513	3,2585	INFAS, Bonn
1998	2	59,15	306	3,2562	GFM-GETAS (IPSOS), Hamburg
1984	3	54,16	517	1,8984	GETAS, Bremen
1986	3	55,97	620	2,9945	Infratest (Burke), München
1988	3	64,46	498	6,7419	GFM-GETAS (IPSOS), Hamburg
1991	4	60,82	319	3,9588	Infratest (Burke), München
1992	4	59,56	502	4,3644	Infratest (Burke), München

Die Erhebungen sind sortiert nach Einschätzung des Umfangs an Spielraum der Interviewer bei der Auswahl (Rang).

Sehr deutlich ist der Zusammenhang zwischen den Spielräumen der Interviewer und den Abweichungen des Frauen-Anteils in den Extremgruppen: Insbesondere die Erhebungen mit sehr geringem Spielraum (1), das sind die drei Erhebungen mit direkter Auswahl der Befragten aus den Listen der Einwohnermeldeämter, zeichnen sich durch Anteile von Frauen aus, die dicht beim bzw. sogar unterhalb des Erwartungswertes von 50% liegen.

Die Erhebungen mit dem vergleichsweise größten Spielraum (4) liegen entsprechend im oberen Bereich der Abweichungen mit einem besonders hohen Anteil an befragten Frauen. Sie werden in dieser Hinsicht nur vom ALLBUS 1988 übertroffen, dem nach unserer Einschätzung ein geringerer Spielraum (3) der Interviewer zugeschrieben wurde.

Weniger klar ist das Bild auf den beiden Stufen mittlerer Spielräume: Ohne Berücksichtigung des ALLBUS 1988 (Ausreißer) liegen die Abweichungen des Anteils befragter Frauen in der Gruppe (3) sogar noch unter den Anteilen in der Gruppe (2).

Insgesamt stimmen die Rangordnungen von Spielräumen der Interviewer und Anteilen der befragten Frauen jedoch recht gut überein: Kendall's tau-c= 0,65 ($p < 0,05$).

4.2 Diskussion der Abweichungen

Neben einer generellen Tendenz zur Bestätigung unserer Vorhersagen muss im Detail ein beträchtliches Maß an Abweichungen festgestellt werden. Einige dieser Abweichungen werden in diesem Abschnitt kommentiert und einige davon versuchsweise durch zusätzliche Eigenschaften der ALLBUS-Erhebungen erklärt. Im Gegensatz zur relativ groben Einstufung der Spielräume in 4 Rangstufen wurden diese zusätzlichen Eigenschaften erst nachträglich ermittelt.

(a) In allen drei Auswahlen auf der Basis der Einwohnermeldeamts-Listen liegt der Anteil der befragten Frauen unter dem Erwartungswert von 50%, zum Teil – im ALLBUS 1994 – mit mehr als 2 Standardabweichungen sogar sehr deutlich darunter. Dieses ist erstaunlich, weil durch unsere spezielle Konstruktion von Grundgesamtheit und Wahrscheinlichkeitsauswahl die Frauen systematisch mit der Eigenschaft relativ leichter Erreichbarkeit verbunden sind. Die Erwartung war deshalb generell auf einen Frauen-Anteil von über 50% gerichtet, und die Analyse zielte nur auf die Frage, um wieviel und in welchem Zusammenhang mit dem Spielraum der Interviewer der tatsächlich realisierte Anteil der Frauen über dem Erwartungswert lag.

Theoretische Erklärungen dieses unerwarteten Ergebnisses sind u. a., dass bei der direkten Auswahl der Personen über die Einwohnermeldeamts-Listen andere Auswahlverzerrungen wie etwa Rollenzuschreibungen innerhalb der Paare (wer nimmt an Interviews teil?) mit Folgen für Verweigerungen sichtbar werden, die unter anderen Auswahlbedingungen durch die starken, in entgegengesetzter Richtung wirkenden Erreichbarkeits-Effekte überlagert werden (diesen Hinweis verdanke ich einem mündlichen Hinweis meines Kollegen R. Schnell):

Die hier gefundene negative Abweichung des Frauen-Anteils vom Erwartungswert wäre damit (wie eine Abweichung in anderer Richtung auch) als Mischung der Folgen gegenläufiger Einflüsse zu interpretieren, nämlich der Folgen

- unterschiedlicher Erreichbarkeiten, die allein und ohne andere Einflüsse den Anteil der befragten Frauen innerhalb der systematisch ausgewählten Paare über den allgemeinen Erwartungswert von 50% heben müssten. Verzerrungen in dieser Richtung sollten bei Wahrscheinlichkeitsauswahlen auf der Basis von Einwohnermeldeamts-Listen zwar geringer sein als bei allen anderen Auswahlformen der ALLBUS-Erhebungen, aber gleichwohl bestehen bleiben
- aller anderen Einflüsse, die systematische Ausfälle überwiegend der Frauen, teilweise aber auch der Männer bewirken. Die oben genannten Vermutungen (Rainer Schnell) über die Ausfälle der Frauen würden wohl überwiegend den Verweigerungen zuzurechnen sein.

Zu den letztgenannten Einflüssen mag auch zählen, dass bei den Personenstichproben aufgrund der Ziehungen aus Einwohnermeldeamts-Listen Adressen und Zielpersonen vorab bekannt waren. Ein Teil der ersten Kontakte wurde deshalb vermutlich per Telefon geknüpft. Im Zusammenhang mit den ausgedehnten Nachbearbeitungsphasen aller drei Erhebungen mag es sein, dass die Männer trotz ihrer häufigeren Abwesenheit vom Haushalt über dieses Medium leichter zu erreichen oder zu einem Interview zu überreden waren.

Die stetige Abnahme dieser Abweichungen über die drei Erhebungen (1994–1996–2000) gab zunächst Anlass zur Vermutung, dass in irgendeiner Weise Korrekturmaßnahmen zur Vermeidung verzerrter Auswahlen ergriffen wurden, die später aufgrund zunehmender Erfahrungen mit Einwohnermeldeamts-Stichproben ihrerseits korrigiert wurden. Die Suche nach entsprechenden Hinweisen in den Methodenberichten führte zwar auf einige Unterschiede zwischen den Auswahlverfahren der drei Erhebungen, aber auf keine Stützung dieser Vermutung:

Die Unterschiede zwischen den drei Einwohnermeldeamts-Auswahlen betreffen einmal die Behandlung von Ersatzadressen für stichprobenneutrale Ausfälle, die 1994 unter etwas anderen Schichtungsbedingungen gezogen wurden als die ursprüngliche Bruttostichprobe (vgl. ALLBUS Methodenbericht, Koch/Wasmer/Harkness/Scholz 2001). In den Jahren 1996 und 2000 dagegen wurden diese Ersatzadressen unmittelbar zusammen mit der Bruttostichprobe (also nach gleichen Regeln) gezogen und den Interviewern zusammen mit den Adressen der zu realisierenden Befragungen übergeben. Die Ersatzadressen durften nur zum Ausgleich stichprobenneutraler Ausfälle eingesetzt werden.

Eine zweite Abweichung betrifft die Erhebung 2000 gegenüber 1994 und 1996: Bei sehr schlechter Ausschöpfung von Adressklumpen (0–2 Interviews) wurden diese vollständig durch neue Adressklumpen ersetzt, wobei auch die eventuell bereits durchgeführten Interviews der ersetzten Klumpen unberücksichtigt blieben. In beiden Fällen ist jedoch kein systematischer Zusammenhang mit dem geringen Anteil befragter Frauen auszumachen.

(b) Die Einstufung der Spielräume der Interviewer in die Klassen 2, 3 und 4 trägt nur wenig zur Ordnung der Anteile befragter Frauen bei, erklärt also nicht die Auswahlverzerrungen aufgrund unterschiedlicher Erreichbarkeiten der Partner im Sinne der oben genannten Hypothese. Insgesamt lässt sich die Varianz des Anteils befragter Frauen zu 80,4% (η^2) auf die Gruppierung in 4 Klassen von Spielräumen zurückführen. Fasst man demgegenüber die beiden mittleren Klassen (2) und (3) zusammen, so sinkt der erklärte Anteil kaum (80,1%), und gleiches gilt, wenn die Klassen (2)–(3)–(4) zusammengefasst werden (77,7%).

Als wirksam zur Erklärung von Ausfällen aufgrund unterschiedlicher Spielräume der Interviewer sind deshalb nur die beiden verbleibenden Klassen anzusehen:

- 2-stufige Auswahlen mit einer direkten Wahrscheinlichkeitsauswahl von Befragungspersonen aus den Einwohnermeldeamts-Listen gegenüber. (1)
- 3-stufige Auswahlen mit entweder Vorgabe ausgewählter Haushalte an die Interviewer oder Random Route-Verfahren zur Ermittlung der Haushalte durch die Interviewer auf der zweiten Stufe sowie einer Wahrscheinlichkeitsauswahl der Befragungspersonen innerhalb der Haushalte mittels Schwedenschlüssel durch die Interviewer auf der dritten Stufe. (2-3-4)

Dieses Ergebnis lässt sich auch nicht wesentlich dadurch verbessern, dass neben den bislang nur groben Einteilungen der ALLBUS-Erhebungen anhand weniger formaler Eigenschaften der Auswahlverfahren zusätzliche Kriterien zur Verfeinerung herangezogen werden. So wurden z. B. nicht in allen Erhebungen von den Interviewern Kontaktprotokolle geführt, die potentiell als Kontrollinstrumente eingesetzt werden könnten. Wo dieses geschah bleibt meist offen, welche Bedeutung die Institute der Dokumentation der Kontaktversuche und möglicher Ausfallgründe beigemessen haben und welche Konsequenzen daraus für Kontrollen und für die Steuerung des Interviewer-Einsatzes gezogen wurden. Eine wirklich verbesserte Einstufung der Spielräume wäre nur möglich, wenn detaillierte Informationen über die tatsächliche Vorgehensweise verschiedener Institute und – innerhalb gleicher Institute: über die Zeit – vergleichend zur Verfügung stünden.

Auf der anderen Seite haben wir vermutlich einigen formalen Kriterien zur Regelung des Auswahlverfahrens zu große Bedeutung beigemessen. So beruht z. B. die Abgrenzung zwischen den Klassen (3) und (4) auf folgenden Vorgaben an die Interviewer (hier nur typisiert wiedergegeben): In Erhebungen, die der Klasse (3) zugeordnet wurden, durften „maximal 10 Befragungen aus 10 Haushalts-Adressen“ realisiert werden, in Klasse (4) der gestuften Spielräume gab es die Erwartung von „6 Realisationen aus 15 Haushalts-Adressen“ (vgl. ALLBUS Methodenberichte).

Bei dieser Abgrenzung erschien die *explizite Freigabe von Spielräumen* an die Interviewer als entscheidender Gesichtspunkt. Demgegenüber muss vielleicht stärker betont werden, dass auch bei der scheinbar restriktiven Fassung „10 Realisationen aus 10 Adressen“ ein faktisch ganz ähnlicher, wenngleich nur implizit gewährter Spielraum bestand. Bei einer derzeit durchschnittlichen Ausschöpfung von rund 50% der Bruttostichprobe und einem entsprechend antizipierten Verhältnis von Brutto- und angestrebter Nettostichprobe ist jedem Interviewer klar, dass „5 Realisationen aus 10 Adressen“ keine schlechte Leistung bedeuten. Warum also sollten Befragungen in weiteren Haushalten versucht werden, wenn dieses erkenn-

bar, d. h. aufgrund zumindest eines vorangehenden Fehlversuches, nur mit größerem Aufwand erreichbar schien?

Zur realistischen Einschätzung der Situation von Interviewern müsste man also viel mehr wissen als nur die formalen Auswahlregeln:

- Welcher Art sind die Übereinkünfte zwischen Instituten und Interviewern über die Höhe der Bezahlung für realisierte Befragungen und welches die (übereinstimmenden?) Erwartungen über den Umfang der Nebenleistungen wie mehrfache Kontaktversuche, die mit der Bezahlung des Stücklohns als beglichen gelten?
- Wie werden mögliche Differenzen über solche Erwartungen zwischen Instituten und Interviewern abgestimmt?
- Gibt es über die formelle Dokumentation solcher Nebenleistungen (im Kontaktprotokoll) hinaus Prüfungen ihres Realitätsgehalts sowie tatsächliche oder von Interviewern erwartete Maßnahmen der Institute bei Verstößen?
- In welchem Umfang sind die eingesetzten Interviewer auf Beiträge in der laufenden Erhebung angewiesen bzw. stehen ihnen – möglicherweise günstigere – Befragungsalternativen offen?

Die Reihe von Fragen ließe sich leicht erweitern. Ihre Beantwortung verlangt Kenntnisse interner Verfahrensweisen der Institute bei der Feldorganisation, die für institutsübergreifende Vergleiche nicht leicht zu erlangen sein werden. Allerdings ließen sich mit solchen Erweiterungen der hier vorgeschlagenen Analyse wertvolle Einsichten in die Steuerung von Auswahlprozessen durch Interviewer gewinnen.

Literatur

- ADM Arbeitskreis Deutscher Markt und Sozialforschungsinstitute / AG.MA Arbeitsgemeinschaft Media-Analyse e.V. (Hg.), 1999: Stichproben-Verfahren in der Umfrageforschung. Opladen: Leske+Budrich.
- Alt, C., W. Bien und D. Krebs, 1991: Wie zuverlässig ist die Verwirklichung von Stichprobenverfahren? Random route versus Einwohnermeldeamtstichprobe. ZUMA-Nachrichten 28: 65-72.
- Bandilla, W., S. Gabler und M. Wiedenbeck, 1992: Methodenbericht zum DFG-Projekt ALLBUS Baseline-Studie 1991. ZUMA-Arbeitsbericht 92/04.
- Braun, M., R. Trometer und M. Wiedenbeck, 1989: Methodenbericht zur „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 1988. ZUMA-Arbeitsbericht 89/02.
- Braun, M., C. Eilinghoff, S. Gabler und M. Wiedenbeck (1993): Methodenbericht zur „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 1992.
- Brückner, E., H.-P. Kirschner, R. Porst, P. Prüfer und P. Schmidt, 1981: Methodenbericht zum „ALLBUS 1980“. ZUMA Arbeitsbericht 81/07.
- Erbslöh B. und M. Wiedenbeck, 1986: Methodenbericht zum „ALLBUS 1986“. ZUMA-Arbeitsbericht 87/04.
- Hagstotz, W., H.-P. Kirschner, R. Porst und P. Prüfer, 1982: Methodenbericht ALLBUS 1982. ZUMA-Arbeitsbericht 82/21.

- Kirschner, H.-P., 1984: ALLBUS 1980: Stichprobenplan und Gewichtung. S. 114-182 in: K. U. Mayer und P. Schmidt (Hg.): Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. Beiträge zu methodischen Problemen des ALLBUS 1980. Frankfurt, New York: Campus Verlag.
- Koch, A., S. Gabler und M. Braun, 1994: Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 1994. ZUMA-Arbeitsbericht 94/11.
- Koch, A., K. Kurz, H. Mahr-George und M. Wasmer, 1999: Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 1998. ZUMA-Arbeitsbericht 99/02.
- Koch, A., M. Wasmer, J. Harkness und E. Scholz, 2001: Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS), 2000. ZUMA-Methodenbericht 01/05.
- Kohler, U., 2007: Surveys from inside: An assessment of unit nonresponse bias with internal criteria. *Survey Research Methods* 1: 55-67. <http://w4.ub.uni-konstanz.de/srm/>.
- Porst, R., P. Prüfer, M. Wiedenbeck und K. Zeifang, 1985: Methodenbericht zum "ALLBUS 1984". ZUMA Arbeitsbericht 85/03.
- Scheuch, E. K., 1974: Auswahlverfahren in der Sozialforschung. S. 1-96 in: R. König (Hg.): *Handbuch der empirischen Sozialforschung*, Band 3a, 3. Aufl. Stuttgart: Ferdinand Enke Verlag.
- Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske+Budrich.
- Sodeur, W. 1997: Interne Kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen. *ZA-Information* 41: 58-82.
- Wasmer, M., A. Koch und M. Wiedenbeck, 1991: Methodenbericht zur "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 1990. ZUMA-Arbeitsbericht 91/13.
- Wasmer, M., A. Koch, J. Harkness und S. Gabler, 1996: Konzeption und Durchführung der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 1996.
- Wasmer, M., A. Koch, 2002: Konzeption und Durchführung der PAPI-Methodenstudie zur "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2000. ZUMA-Methodenbericht 02/01.
- Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim, und Zentralarchiv für empirische Sozialforschung an der Universität zu Köln: Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS):
- Einzelstudie 1980, PC-lesbares Codebuch, ZA-Nummer 1000
 - Einzelstudie 1982, PC-lesbares Codebuch, ZA-Nummer 1160
 - Einzelstudie 1984, PC-lesbares Codebuch, ZA-Nummer 1340
 - Einzelstudie 1986, PC-lesbares Codebuch, ZA-Nummer 1500
 - Einzelstudie 1988, PC-lesbares Codebuch, ZA-Nummer 1670
 - Einzelstudie 1990, PC-lesbares Codebuch, ZA-Nummer 1800
 - Basisumfrage 1991, PC-lesbares Codebuch, ZA-Nummer 1990
 - Einzelstudie 1992, PC-lesbares Codebuch, ZA-Nummer 2140
 - Einzelstudie 1994, PC-lesbares Codebuch, ZA-Nummer 2400
 - Einzelstudie 1996, PC-lesbares Codebuch, ZA-Nummer 2800
 - Einzelstudie 1998, PC-lesbares Codebuch, ZA-Nummer 3000
 - Einzelstudie 2000, PC-lesbares Codebuch, ZA-Nummer 3451

Korrespondenzadresse: Wolfgang Sodeur
 Ostlandstraße 72
 50859 Köln
 wolfgang.sodeur@t-online.de