

### 20 Jahre Feldarbeit im ALLBUS: ein Blick in die Blackbox

Koch, Achim

Veröffentlichungsversion / Published Version  
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:  
GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Koch, A. (2002). 20 Jahre Feldarbeit im ALLBUS: ein Blick in die Blackbox. *ZUMA Nachrichten*, 26(51), 9-37. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-207849>

#### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

---

# 20 JAHRE FELDARBEIT IM ALLBUS: EIN BLICK IN DIE BLACKBOX<sup>1</sup>

ACHIM KOCH

*If you love surveys or sausages, you should not watch either being made.*

*(J. Kochevar)*

Über die Details der Datenerhebung bei persönlich-mündlichen (face-to-face) Bevölkerungsbefragungen ist nur wenig bekannt. Mit dem vorliegenden Beitrag sollen schlaglichtartig einige ausgewählte Facetten der Feldarbeit beleuchtet werden. Datengrundlage bilden die zwölf Erhebungen, die für die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) zwischen 1980 und 2000 durchgeführt wurden. Betrachtet werden u.a. Anzahl und demographische Zusammensetzung der eingesetzten Interviewer, Wochentage und Tageszeiten der Interviewdurchführung sowie Informationen zur Feldzeit und zu den Fallpreisen in den einzelnen Erhebungen. Vor dem Hintergrund der Problematik der Ausschöpfungsberechnung wird zum Schluß auf die Frage des Zusammenhangs von Qualität, Zeitbedarf und Preis eingegangen.

The level of knowledge about the details of fieldwork in face-to-face interview surveys of the general population is rather low. The present paper focuses on certain particular aspects of fieldwork. The database consists of the twelve surveys of the German General Social Survey (ALLBUS), fielded between 1980 and 2000. We investigate the number and demographic structure of the interviewers, the days and times of interviewing, as well as both the length of the fielding period and the costs of fielding. The relationship between quality, fielding time, and costs is discussed against the background of the problems involved in determining response rates in a reliable manner.

---

<sup>1</sup> Überarbeitete Version eines Vortrages auf der Frühjahrstagung der Sektion Methoden der DGS am 30./31. März 2001 in Konstanz.

## 1. Einleitung

Die Datenerhebung bei persönlich-mündlichen (face-to-face) Bevölkerungsbefragungen erfolgt in Deutschland ganz überwiegend durch privatwirtschaftlich verfaßte Umfrageinstitute. Diese arbeiten mit einem Stab von Interviewern zusammen, die in der Regel als freiberufliche Mitarbeiter für die Institute tätig sind. Die Aufgaben der Interviewer bei der Feldarbeit umfassen die Mitwirkung an der Stichprobenbildung, die Kontaktierung der Zielpersonen und deren Motivierung zur Teilnahme an der Studie und schließlich die Durchführung der Befragung entsprechend den Regeln des standardisierten Interviews. Die Umfrageinstitute setzen u.a. durch ihre Verfahren der Interviewerschulung und –kontrolle, durch die verwendeten Honorierungsmodelle sowie weitere studienbezogene Vorgaben, z.B. hinsichtlich der Zahl und zeitlichen Terminierung von Kontaktversuchen, die Rahmenbedingungen für die Tätigkeit der Interviewer.

Sowohl über das Institutshandeln als auch über das tatsächliche Verhalten der Interviewer, über typische Probleme der Feldarbeit und deren Lösung unter Praxisbedingungen ist indes nur wenig bekannt. Die Gründe hierfür sind vielfältiger Natur: die Wettbewerbssituation der Umfrageinstitute und deren Angst vor einem möglichen Imageschaden durch die Offenlegung eventuell vorhandener Defizite; fehlende Informationen über das faktische Verhalten der Interviewer auch in den Instituten, weil dieses bei persönlichen – im Gegensatz zu zentral im Studio durchgeführten telefonischen – Umfragen schwieriger zu standardisieren und kontrollieren ist; mangelndes Interesse der Auftraggeber an Fragen der Datenerhebung, etc.

Auch mit dem vorliegenden Beitrag kann lediglich eine erste Beschreibung geleistet werden: Schlaglichtartig werden einige ausgewählte Facetten der Feldarbeit beleuchtet – wohl wissend, dass weite Bereiche weiterhin im Dunkeln verbleiben. Zugrunde liegen Informationen aus den zwölf Erhebungen, die für die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) zwischen 1980 und 2000 durchgeführt wurden. Um die Aussagen auf eine breitere Basis zu stellen, beschränken wir uns im folgenden auf Aspekte, über die jeweils für mehrere – im besten Fall für alle zwölf – Erhebungen Informationen vorliegen.<sup>2</sup> Die gewählte Vorgehensweise erlaubt es, an einigen Stellen Unterschiede und Veränderungen in der Feldarbeit – zwischen verschiedenen Umfrageinstituten und Stichprobenverfahren oder auch über die Zeit – zu identifizieren. Ursachen

---

<sup>2</sup> Den Auswertungen liegen einige wenige Merkmale zugrunde. Dies sind eine Kennung der primären Stichprobeneinheit (im folgenden auch als „Sample Point“ bzw. „Klumpen“ bezeichnet), die Interviewer-ID, Informationen zu Geschlecht, Alter und Bildung der eingesetzten Interviewer sowie Angaben zu Datum und Uhrzeit des Interviews. Es wäre wünschenswert, wenn zumindest diese rudimentären Informationen standardmäßig in sozialwissenschaftlichen Umfragen erfaßt würden.

und Gründe als auch mögliche Konsequenzen (z.B. was die Qualität der erhobenen Daten angeht) können nur cursorisch behandelt werden.

In Abschnitt zwei werden zunächst die Durchführungsmodalitäten der ALLBUS-Umfrage, speziell die verwendeten Stichprobenverfahren, skizziert. Daran schließen sich in Abschnitt drei Informationen zur Zahl der in den einzelnen Erhebungen eingesetzten Interviewer und zur Zahl realisierter Interviews pro Interviewer an. In Abschnitt vier wird die Zusammensetzung der eingesetzten Interviewer nach den Merkmalen Geschlecht, Alter und Bildung beleuchtet und analysiert, ob zwischen der Struktur des Interviewerstabes und dem demographischen Profil der realisierten ALLBUS-Stichproben ein Zusammenhang besteht. In Abschnitt fünf wird erläutert, an welchen Wochentagen und zu welchen Tageszeiten die ALLBUS-Interviews durchgeführt wurden. In Abschnitt sechs wird die Feldzeit der ALLBUS-Umfragen beschrieben. Eine Betrachtung auf Ebene der primären Stichprobeneinheiten („Klumpen“) bietet dabei einen detaillierteren Einblick in den zeitlichen Verlauf der Feldarbeit. In Abschnitt sieben wird anhand der durchschnittlichen Fallpreise ein Blick auf die Kosten der Umfragen geworfen. Vor dem Hintergrund der Problematik der Ausschöpfungsberechnung wird zum Schluß in Abschnitt acht auf die Frage des Zusammenhangs von Qualität, Zeitbedarf und Preis eingegangen.

## **2. Durchführungsmodalitäten der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften**

In der ALLBUS-Umfrage wird seit 1980 alle zwei Jahre eine Zufallsstichprobe der Bevölkerung in der Bundesrepublik Deutschland mit einem teils konstanten, teils variablen Fragenprogramm zu sozialwissenschaftlich relevanten Themenstellungen befragt.<sup>3</sup> Zwischen 1980 und 1990 umfaßte die ALLBUS-Stichprobe jeweils ungefähr 3.000 Personen aus der Grundgesamtheit der deutschen Bevölkerung ab 18 Jahren in Privathaushalten der alten Bundesrepublik (vgl. Tabelle 1). 1991 wurde aus Anlaß der deutschen Vereinigung

---

**3** Die Konzeption der Umfrage erfolgt - in Abstimmung mit einem wissenschaftlichen Beirat - durch die Abteilung ALLBUS am Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim. Zur Verbesserung der Infrastruktur in den Sozialwissenschaften stehen die erhobenen Daten sofort nach ihrer benutzergerechten Aufbereitung und Dokumentation im Zentralarchiv für empirische Sozialforschung (ZA) in Köln interessierten Forschern und Institutionen für Sekundäranalysen zur Verfügung. Der ALLBUS gehört zu einem wachsenden Netz von regelmäßig durchgeführten, unabhängigen nationalen Bevölkerungsumfragen, zu dem auch Studien wie der General Social Survey in den USA, die British Social Attitudes-Erhebungsreihe oder die laufenden Surveys des niederländischen Sociaal en Cultureel Planbureau gehören. Finanziert wird der ALLBUS vom Bund und den Ländern im Rahmen der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS).

zusätzlich eine Umfrage außerhalb des zweijährigen Turnus durchgeführt, bei der erstmals auch Bürger der neuen Bundesländer sowie deutschsprechende Ausländer in die Stichprobe aufgenommen wurden. Die Stichproben der gesamtdeutschen Erhebungen zwischen 1991 und 2000 waren mit circa 3.000 bis 3.500 Personen nicht oder nur geringfügig größer als in den 80er Jahren; Ostdeutschland wurde durch ein „oversampling“ in allen Studien bewußt überrepräsentiert.

**Tabelle 1: Die ALLBUS-Erhebungen 1980 bis 2000 im Überblick**

	1980	1982	1984	1986	1988	1990
<b>Grundgesamtheit</b>	Personen ab 18 Jahren mit deutscher Staatsangehörigkeit in Privathaushalten in Westdeutschland (incl. West-Berlin)					
<b>Stichprobenverfahren*</b>	ADM AR	ADM AR	ADM RR	ADM RR	ADM RR	ADM AR
<b>Nettofallzahl</b>	2955 West	2991 West	3004 West	3095 West	3052 West	3051 West
<b>Befragungsform</b>	PAPI	PAPI	PAPI	PAPI	PAPI	PAPI
<b>Erhebungsinstitut</b>	GETAS	GETAS	GETAS	Infratest	GETAS	Infas
	1991	1992	1994	1996	1998	2000
<b>Grundgesamtheit</b>	Personen ab 18 Jahren (Deutsche und Ausländer) in Privathaushalten in West- und Ostdeutschland					
<b>Stichprobenverfahren*</b>	ADM SR	ADM SR	Register	Register	ADM AR	Register
<b>Nettofallzahl</b>	1514 West 1544 Ost	2400 West 1148 Ost	2342 West 1108 Ost	2402 West 1116 Ost	2212 West 1022 Ost	2036 West 1102 Ost
<b>Befragungsform</b>	PAPI	PAPI	PAPI	PAPI	PAPI	CAPI
<b>Erhebungsinstitut</b>	Infratest	Infratest	Infratest	Infratest	GETAS	Infratest

\* AR = Adress-Random, RR = Random-Route, SR = Standard-Random

Insgesamt wurden in den zwölf ALLBUS-Umfragen zwischen 1980 und 2000 mehr als 38.000 Interviews realisiert. Alle Erhebungen fanden als persönlich-mündliche Befragung statt, bis 1998 als Paper and Pencil Interview (PAPI), im Jahr 2000 zum ersten Mal als Computer Assisted Personal Interview (CAPI). Mit der Feldarbeit waren drei Umfrageinstitute betraut: Infratest Burke (heute NFO Infratest), GFM-GETAS (heute IPSOS Deutschland) und Infas. Sämtliche Umfragen wurden auf der Basis von Zufallsstichproben, entweder in Form des ADM-Designs (bzw. einem äquivalenten Verfahren) oder als Stichproben aus Einwohnermelderegistern, realisiert.

In der Mehrzahl der zwölf ALLBUS-Erhebungen kam das dreistufige Design des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (ADM) zum Einsatz, das in

Deutschland das Standard-Stichprobenverfahren bei persönlich-mündlichen Bevölkerungsumfragen darstellt (vgl. Arbeitsgemeinschaft ADM-Stichproben und Bureau Wendt 1994, Behrens/Löffler 1999, Koch 1997). Dieses Verfahren sieht eine dreistufige geschichtete Zufallsauswahl mit den Auswahlstufen ‚Wahlbezirke‘ – ‚Haushalte‘ – ‚Personen‘ vor. In der ersten Stufe wird eine Stichprobe von Wahlbezirken gezogen. In der zweiten und dritten Auswahlstufe ermitteln die Interviewer in den ausgewählten Wahlbezirken nach einem Zufallsverfahren die zu befragenden Haushalte und Zielpersonen. Da für einen Teil der folgenden Ausführungen insbesondere das Vorgehen in der zweiten und dritten Auswahlstufe von Relevanz ist, soll dieses etwas näher erläutert werden.

Für die Wahlbezirke sind im ADM-Design zwar die geographischen Grenzen bekannt, nicht jedoch die Adressen der Haushalte, die sie umfassen. Zur Bestimmung der Zielhaushalte wird deshalb den Interviewern anhand eines zufällig ausgewählten Startpunktes und Fortsetzungsregeln ein Begehungsweg vorgegeben. Entlang dieses Weges haben die Interviewer nach bestimmten Regeln eine festgelegte Zahl von Haushalten („Klingelschilder“) aufzulisten.<sup>4</sup> In der Praxis haben sich für die Ermittlung der Auswahlseinheiten der zweiten Stufe drei Verfahrensvarianten herausgebildet, die sich insbesondere darin unterscheiden, wieviel Entscheidungsspielraum den Interviewern eingeräumt wird bzw. wieviel Kontrolle von Institutsseite möglich ist. Beim restriktivsten Verfahren, dem sogenannten ‚Adress-Random‘, erfolgt die Auflistung der Haushaltsadressen durch die Interviewer zeitlich vor der Feldarbeit in einem getrennten Arbeitsschritt. Aus den aufgenommenen Adressen wird im Umfrageinstitut eine Substichprobe von Adressen für die eigentliche Befragung zufällig ausgewählt und den Interviewern zur Bearbeitung übergeben. Verbreiteter als diese Variante sind jedoch Verfahren, bei denen die Haushaltsauflistung in einem Zuge mit der Realisierung der Interviews erfolgt. Beim sogenannten ‚Random-Route‘ wird den Interviewern eine Bruttozahl von aufzulistenden Adressen vorgegeben, aus denen sie möglichst viele Interviews realisieren sollen. Beim weicheren ‚Standard-Random‘ wird von den Interviewern lediglich die Realisierung einer bestimmten Nettofallzahl von Interviews gefordert. Das Brutto von Adressen, das dafür angelaufen werden darf, wird entweder nicht begrenzt oder großzügig bemessen (z.B. indem nur fünf Interviews aus maximal 15 Adressen realisiert werden sollen).

Auch die Bestimmung der Befragungsperson im Haushalt – die dritte Auswahlstufe – ist im ADM-Design Aufgabe der Interviewer. Hierzu müssen sie zunächst alle Haushaltsmitglieder, die zur Grundgesamtheit der Studie zählen, in einer festgelegten Reihenfolge

---

<sup>4</sup> In den gängigen Vorgehensweisen werden entweder ab dem Startpunkt alle Adressen bis zur geforderten Anzahl aufgelistet oder nur jede x-te (z.B. jede dritte oder fünfte) Adresse bis zur geforderten Anzahl.

auflisten. Lebt nur eine Person der Grundgesamtheit im Haushalt, ist kein weiterer Auswahlsschritt notwendig. In allen anderen Fällen muß mit Hilfe eines Zufallsverfahrens die zu befragende Person ermittelt werden. In den ALLBUS-Erhebungen wurde den Interviewern dafür jeweils eine ‚Kish Tabelle‘ (resp. ‚Schwedenschlüssel‘) vorgegeben, bei der die Befragungsperson anhand einer vorbereiteten Zufallsziffernreihe ausgewählt wird.<sup>5</sup>

Das ADM-Design wurde in neun ALLBUS-Erhebungen verwendet: 1980, 1982, 1990 und 1998 als Adress-Random, 1984, 1986 und 1988 als Random-Route und 1991 und 1992 als Standard-Random. In den verbleibenden drei Erhebungen 1994, 1996 und 2000 dagegen wurden Stichproben aus Einwohnermelderegistern gezogen, ein Vorgehen, das aufgrund seines zeitlichen und finanziellen Mehraufwandes bei bevölkerungsrepräsentativen face-to-face Umfragen nur selten zum Einsatz kommt (vgl. Koch 1997).

Bei diesem Verfahren wird in einer zweistufigen geschichteten Zufallsauswahl zunächst eine Stichprobe von Gemeinden gebildet und anschließend werden in den Gemeinden Personenadressen aus den Einwohnermelderegistern zufällig ausgewählt. Die Bruttostichprobe wird damit völlig unabhängig von der Feldarbeit erstellt, den Interviewern werden die zu befragenden Personen mit Namen und Adresse vorgegeben. Die Hauptvorteile im Vergleich zum ADM-Design sind die Minimierung des Interviewereinflusses bei der Auswahl der Befragungspersonen, die Verfügbarkeit von Registerinformationen (Alter, Geschlecht, deutsche Staatsangehörigkeit: ja/nein) für Teilnehmer und Nichtteilnehmer der Umfrage sowie designbedingt gleiche Auswahlwahrscheinlichkeiten für die Zielpersonen.

### **3. Realisierte Interviews pro Interviewer**

Beginnen wir nun mit Informationen zu den zentralen Akteuren der Feldarbeit, den Interviewern. Sie stellen die wichtigste – und oft einzige<sup>6</sup> – Verbindung zwischen dem Forscher und den Befragten dar (vgl. Groves/Couper 1998: 191ff.). Im ALLBUS waren in den einzelnen Erhebungen zwischen 285 (ALLBUS 2000) und 719 (ALLBUS 1992) Interviewer an der Realisierung der jeweils 3.000 bis 3.500 Interviews beteiligt (vgl.

---

**5** Werden alle Auswahlvorschriften dieses Verfahrens korrekt eingehalten, erhält man eine Stichprobe, bei der jeder Haushalt die gleiche Chance hat, in die Befragung einbezogen zu werden. Die Auswahlchance des einzelnen Haushaltsmitgliedes hängt von der Haushaltsgröße ab. Dies hat zur Folge, dass bei der Auswertung der Daten auf Personenebene vom Design her eine Gewichtung mit der – um die nicht zur Grundgesamtheit zählenden Haushaltsmitglieder reduzierten – Haushaltsgröße notwendig ist, die sogenannte ‚Transformation‘ (zur Problematik dieser Gewichtung vgl. Hartmann/Schimpl-Neimanns 1992: 319ff.).

**6** Zumeist kommt als zusätzlicher Kommunikationsweg lediglich ein Anschreiben des Forschers an alle Zielpersonen in Betracht, vgl. z.B. Lynn et al. 1998.

Tabelle 2).<sup>7</sup> Damit lag die Zahl der von jedem Interviewer im Durchschnitt durchgeführten Interviews zwischen 4,9 (1991, 1992) und 11,0 (2000).

Warum die Anzahl der Interviewer und deren durchschnittlicher „workload“ zwischen den Studien variiert, kann hier nicht im einzelnen geklärt werden. Die Entscheidung über die Zahl der einzusetzenden Interviewer wird letztendlich vom Umfrageinstitut getroffen. Die Größe des Interviewerstabes, dessen Auslastung, aber auch die spezifischen Anforderungen der konkreten Studie spielen bei Entscheidungen über den Interviewereinsatz eine Rolle. Zu bedenken ist, dass in einem bestimmten Zeitraum von einem Institut in der Regel mehrere Studien parallel bearbeitet werden. Bei größeren Organisationen können durchaus ständig 15 bis 30 Projekte zeitgleich im Feld sein (vgl. Blyth 1998: 574), so dass die Planung und Steuerung des Interviewereinsatzes eine komplexe Aufgabe darstellt.

**Tabelle 2: Realisierte Interviews pro Interviewer<sup>\*</sup>**

Jahr	Anzahl realisierter Interviews	Anzahl Interviewer	Realisierte Interviews pro Interviewer (Mittelwert)	Maximale Zahl der Interviews pro Interviewer
1980	2955	430	6,9	24
1982	2991	415	7,2	48
1984	3004	438	6,9	30
1986	3095	611	5,1	31
1988	3052	324	9,4	145
1990	3051	364	8,4	58
1991	3058	625	4,9	20
1992	3548	719	4,9	22
1994	3450	608	5,7	23
1996	3518	451	7,8	52
1998	3234	369	8,8	55
2000	3138	285	11,0	78

\* nur Interviewer mit mindestens einem realisierten Interview

Datenbasis: ALLBUS 1980-2000

Für die Zukunft ist davon auszugehen, dass mit der zunehmenden Verbreitung von CAPI die Interviewerstäbe zahlenmäßig eher kleiner werden, die verbleibenden Interviewer

<sup>7</sup> Beteiligt heißt in diesem Fall: die betreffenden Interviewer haben mindestens ein Interview realisiert. Daneben haben in jeder Erhebung noch eine Reihe weiterer Interviewer mitgearbeitet, ohne jedoch wenigstens ein Interview zu erzielen. Deren Anteil betrug in den einzelnen Erhebungen zwischen einigen wenigen Interviewern und circa 15 Prozent aller Interviewer.



jedoch in größerem zeitlichen Umfang der Interviewertätigkeit nachgehen als bisher. Auf diese Weise können die Umfrageinstitute die hohen Investitionskosten (für Laptops, Software, etc.) bei CAPI am ehesten amortisieren (vgl. Collins et al. 1998: 42)<sup>8</sup>. Unter methodischem Blickwinkel ist eine solche Entwicklung allerdings nicht unproblematisch. Wenn Interviewer homogenisierende Effekte auf die resultierenden Daten haben, dann können in ähnlicher Weise Designeffekte (mit der Folge einer Reduzierung der effektiven Stichprobengröße) auftreten wie aufgrund der räumlichen Klumpung einer Stichprobe (vgl. Schnell 1997: 272ff.). Diese Designeffekte werden umso größer ausfallen, je höher die Zahl der realisierten Interviews pro Interviewer ist (vgl. Fowler/Mangione 1990: 25ff.; Groves 1989: 360ff.).

Im übrigen ist darauf hinzuweisen, dass in manchen ALLBUS-Studien Interviewer vereinzelt weit mehr als die durchschnittliche Zahl von Interviews realisiert haben (vgl. Tabelle 2). Zwar wurde bei den Erhebungen des ALLBUS normalerweise eine Obergrenze für die Zahl der Interviews pro Interviewer definiert (diese betrug nie mehr als 50 Interviews pro Interviewer). Dies verhinderte jedoch nicht, dass in Einzelfällen dieses Limit überschritten wurde. In sechs der zwölf Erhebungen betrug die maximale Zahl von Interviews pro Interviewer zwischen 20 und 31, in vier Umfragen lag das Maximum bei 48 bis 58 Interviews. Höhere Werte weisen der ALLBUS 2000 (Maximum 78 Interviews) und insbesondere der ALLBUS 1988 auf. In dieser Erhebung hat ein Interviewer insgesamt 145 Interviews und damit knapp 5 Prozent aller Interviews der Studie durchgeführt. Auch wenn man in Rechnung stellt, dass bei der Feldarbeit mitunter Kompromisse zwischen dem theoretisch wünschenswerten und dem praktisch einlösbaren Vorgehen gemacht werden müssen, erscheint ein solch hoher Wert vor dem Hintergrund der oben angesprochenen Problematik von Interviewereffekten nur schwer tolerabel.

#### **4. Demographische Struktur der Interviewer**

Was wissen wir über die im ALLBUS eingesetzten Interviewer? Für die Erhebungen 1984 bis 1990 und 1994 bis 2000 liegen Informationen über das Geschlecht und das Alter, für die Umfragen 1994 bis 2000 zusätzlich Informationen zum Bildungsabschluß vor. Tabelle 3 gibt die Verteilungen der drei Interviewermerkmale wieder. Neben den Ergebnissen auf Ebene der Interviewer sind die Resultate auf Ebene der Interviews aufgeführt. Bei letzteren geht jeder Interviewer mit einem Gewicht proportional zur Zahl seiner durchgeführten

---

<sup>8</sup> Collins et al. nehmen im übrigen an, dass sich mit der weiteren Verbreitung von CAPI auch die Beziehung zwischen Interviewer und Umfrageinstitut, incl. deren vertraglicher Gestaltung, ändern wird: die Interviewertätigkeit bietet dann unter Umständen die Perspektive einer richtigen Berufslaufbahn (Stichwort: „interviewing as a career“).

Interviews in die Analyse ein. Außerdem sind zum Vergleich die Verteilungen für die Befragten der betreffenden Erhebungen dokumentiert.

Wie Tabelle 3 belegt, unterscheiden sich die Ergebnisse auf Ebene der Interviewer und auf Ebene der Interviews nur unwesentlich. Dies bedeutet, dass die betreffenden Kategorien von Interviewern im Durchschnitt ungefähr gleich viele Interviews realisiert haben.<sup>9</sup> Im folgenden werden wir uns im Text jeweils auf die Ergebnisse für die Ebene der Interviews beziehen.

**Tabelle 3: Demographische Merkmale von Interviewern und Befragten**

	<b>Interviewer*</b> (Ebene der Interviewer)	<b>Interviewer*</b> (Ebene der Interviews)	<b>Befragte</b>
<b>Geschlecht:**</b>	%	%	%
Männlich	66,4	67,9	47,7
Weiblich	33,6	32,1	52,3
N	3435	25321	25542
<b>Alter:**</b>	%	%	%
18-29 Jahre	14,8	12,3	20,7
30-44 Jahre	32,2	29,3	28,9
45-59 Jahre	34,0	37,0	24,4
60 Jahre u. älter	19,0	21,4	26,0
N	3431	25418	25500
Durchschnittsalter	45,8 Jahre	47,2 Jahre	46,3 Jahre
<b>Bildung:***</b>	%	%	%
Bis Hauptschul- abschluß	16,8	17,5	47,7
Mittlere Reife/ Realschulabschluß	33,3	35,0	29,7
Fachhochschulreife/ Abitur	17,2	15,7	9,6
Fachhochschul-/ Hochschulabschluß	32,7	31,8	13,1
N	1695	13187	13168

\* Nur Interviewer mit mindestens einem realisiertem Interview

\*\* Datenbasis ALLBUS 1984-1990, 1994-2000

\*\*\* Datenbasis ALLBUS 1994-2000

**9** Die Vermutung liegt nahe, dass die betreffenden Kategorien von Interviewern sich damit grosso modo auch kaum in ihren Erfolgs-, sprich Ausschöpfungsquoten unterscheiden. Zutreffend ist diese Folgerung dann, wenn die Interviews der unterschiedlichen Kategorien von Interviewern aus einem vergleichbar großen Bruttoansatz resultieren. Da entsprechende Informationen mit einer Ausnahme (ALLBUS 1986, vgl. Anmerkung 14) nicht vorliegen, kann dies hier nicht geprüft werden.

Zwei Drittel der ALLBUS-Interviews wurden von männlichen und ein Drittel von weiblichen Interviewern durchgeführt. Das Durchschnittsalter der Interviewer und der Befragten unterscheidet sich mit 47 bzw. 46 Jahren kaum, allerdings ist die jüngste und älteste Altersgruppe unterproportional, die Gruppe der 45-59jährigen dagegen überproportional unter den Interviewern vertreten. Die Interviewer verfügen über ein höheres Bildungsniveau als die Befragten: nahezu 50 Prozent aller Interviews wurden von Interviewern mit Fachhochschulreife/Abitur bzw. einem Fachhochschul- oder Hochschulabschluß durchgeführt, während von den Befragten nur 23 Prozent über einen solchen Bildungsabschluß verfügen.<sup>10</sup>

Dieses Profil der ALLBUS-Interviewer entspricht in der Grundtendenz den verfügbaren Informationen über die Interviewerstäbe bei anderen Erhebungen (vgl. zum Sozio-oekonomischen Panel z.B. Schräpler/Wagner 2001). Im internationalen Vergleich ist allerdings die überproportionale Beteiligung männlicher Interviewer auffällig. Nach Groves/Couper (1998: 196) dominieren bei face-to-face Befragungen weibliche Interviewer, so z.B. in den USA, aber auch in Kanada oder den Niederlanden. Dem ALLBUS vergleichbare Befunde werden dagegen für Alter und Bildung mitgeteilt. Für mehrere Studien aus den USA berichten die Autoren ein mittleres Alter der Interviewer von circa 50 Jahren sowie ein überdurchschnittliches Bildungsniveau.

Eine Differenzierung der Ergebnisse nach Erhebungszeitpunkt sowie nach West- und Ostdeutschland (nicht tabellarisch ausgewiesen) ergibt bei der Geschlechterverteilung kaum Unterschiede zwischen den Jahren.<sup>11</sup> Festzuhalten ist jedoch, dass in Ostdeutschland der ‚Männerüberschuß‘ bei den Interviewern stärker ausgeprägt ist als in Westdeutschland. Das Durchschnittsalter der Interviewer nimmt über die Zeit zu: Waren die ALLBUS-Interviewer in den 80er Jahren im Mittel zwischen 40 und 45 Jahren alt, sind sie in den 90ern circa 50 bis 55 Jahre alt.<sup>12</sup> In Ostdeutschland sind die Interviewer durchschnittlich zwei bis sieben Jahre älter als in Westdeutschland. Weit gravierender ist jedoch der Unterschied zwischen West und Ost im Bildungsniveau der Interviewer: In Ostdeutschland stammen circa 60 Prozent aller Interviews von Interviewern mit einem Fachhochschul- oder Hochschul-

---

**10** Inwieweit die demographische Zusammensetzung des Interviewerstabes von den Umfrageinstituten bewußt gesteuert wird, ist nicht bekannt. Das größte Umfrageinstitut in Deutschland – die GfK in Nürnberg – gibt als Zielsetzung an, den Interviewerstab entsprechend dem soziodemographischen Aufbau der Bevölkerung zu strukturieren (vgl. Niehoff 1998: 54).

**11** Auch bei dieser differenzierten Betrachtung unterscheiden sich die Resultate für die beiden Ebenen der Interviewer und der Interviews nicht gravierend. Im Text werden wieder die Ergebnisse auf Ebene der Interviews berichtet.

**12** Z.T. mag dies daher rühren, dass über die Zeit die gleichen Interviewer von den Umfrageinstituten für die Durchführung der ALLBUS-Studien eingesetzt wurden.

abschluß, im Westen sind es weniger als 20 Prozent. In dem beobachtbaren Zeitraum zwischen 1994 und 2000 haben sich diese Werte kaum verändert.<sup>13</sup>

Warum ist die Zusammensetzung des Interviewerstabes überhaupt von Interesse? Zum einen kann das Teilnahmeverhalten der Zielpersonen beeinflusst werden. Verschiedene Kategorien von Interviewern können unterschiedlich hohe Ausschöpfungsquoten erzielen<sup>14</sup> oder aber bei unterschiedlichen Befragtengruppen besonders erfolgreich sein. In Deutschland wurde die Bedeutung der demographischen Komposition der Interviewerstäbe u.a. im Kontext der Media-Analyse diskutiert (vgl. Context 1993). Ausgangspunkt war die Beobachtung, dass die an dieser Gemeinschaftsuntersuchung beteiligten Umfrageinstitute unterschiedliche Reichweiten für die erfragten Zeitschriften ermittelten. Man vermutete, dass die Abweichungen mit der unterschiedlichen Strukturierung der Interviewerstäbe der Institute zusammenhängen. Bei Analysen des Teilnahmeverhaltens stellte man ein „Homogamie-Prinzip“ der Gesellung fest: Männer haben eher Männer interviewt, Frauen eher Frauen, Ältere eher Ältere, usw. Erklärbar ist ein solches Ergebnis dadurch, dass bei einer Begegnung unter Fremden – wie es die Kontaktaufnahme für ein Interview typischerweise ist – eine Person mit ähnlichen Merkmalen sympathischer eingeschätzt wird und diese Sympathie die Kooperationsneigung erhöht (vgl. Groves/Couper 1998: 34). Auch im ALLBUS existieren Indizien für einen entsprechenden Zusammenhang zwischen Interviewer- und Befragtenmerkmalen.

Korreliert man in den einzelnen Erhebungen – für West- und Ostdeutschland getrennt – jeweils Geschlecht, Alter und Bildung der Interviewer mit dem betreffenden Merkmal der Befragten, dann erweisen sich 13 der 32 betrachteten Zusammenhänge als signifikant. In allen 13 Fällen besteht eine schwache positive Korrelation, das heißt auch hier läßt sich das in der Media-Analyse festgestellte „Homogamie-Prinzip“ beobachten.<sup>15</sup> Auffällig

---

**13** Die Ursache für das hohe Bildungsniveau der ostdeutschen Interviewer ist dabei nicht ganz klar. Von einem Umfrageinstitut wurde auf eine entsprechende Anfrage die Vermutung geäußert, dass in Ostdeutschland viele Personen, die aufgrund des Transformationsprozesses arbeitslos wurden (insbesondere auch Lehrer), sich als Interviewer betätigen.

**14** Für den ALLBUS 1986 berichtet z.B. Koch (1991), dass ältere Interviewer eine höhere Ausschöpfung erzielen als jüngere und dass weibliche Interviewer erfolgreicher als männliche sind. Die Wirkung solcher fixen Interviewermerkmale erfolgt z.T. vermutlich vermittelt über Faktoren wie die Interviewererfahrung, die Verfügung über ein breites Repertoire von Handlungsstrategien, u.ä., so dass bei Einführung entsprechender Kontrollvariablen die Zusammenhänge reduziert werden (vgl. Groves/Couper 1998: 198).

**15** Es bleibt weiterführenden Analysen vorbehalten zu klären, inwieweit diese Ergebnisse auch bei Beachtung möglicher methodischer Einwände Bestand haben. Zwei Kritikpunkte können gegen das gewählte Vorgehen geltend gemacht werden: Zum einen läßt die hier vorgenommene Analyse die Clusterung der Beobachtungen innerhalb der ‚workloads‘ der Interviewer außer Betracht, so dass die

dabei ist, dass die Zusammenhänge häufiger bei den ADM-Studien als bei den Registerstichproben auftreten: Bei ersteren erweisen sich zehn von vierzehn Zusammenhängen als signifikant, bei letzteren drei von achtzehn. Eventuell spielen für dieses Ergebnis die größeren Freiräume der Interviewer bei der Haushalts- und Zielpersonenauswahl im ADM-Design eine Rolle. Diese Hypothese deckt sich mit den Ergebnissen von Sodeur (1997: 81), der für mehrere ALLBUS-Stichproben anhand interner Prüfkriterien analysierte, inwiefern Abweichungen von der idealen Realisierung der Wahrscheinlichkeitsauswahl zu beobachten sind. Dabei schnitt der auf Basis einer Registerstichprobe durchgeführte ALLBUS 1994 besser ab als die ADM-basierten Studien, was Sodeur auf die geringeren Spielräume der Interviewer bei der Befragtenauswahl zurückführt.

**Tabelle 4: Korrelation zwischen Interviewer- und Befragtenmerkmalen**

Jahr	Erhebungsinstitut	Stichprobenverfahren	West			Ost		
			Geschlecht	Alter	Bildung	Geschlecht	Alter	Bildung
			Pearson's r			Pearson's r		
1984	GETAS	ADM	.09***	.15***				
1986	Infratest	ADM	.05**	.05**				
1988	GETAS	ADM	.05**	.09***				
1990	Infas	ADM	.02	.12***				
1994	Infratest	Register	.03	.02	.09***	.03	-.02	.05
1996	Infratest	Register	.01	.02	.04	-.01	.02	.06*
1998	GETAS	ADM	.02	.12***	.14***	.07*	.06	.05
2000	Infratest	Register	.03	.00	.05*	-.01	-.02	-.01

\* =  $p \leq .05$ , \*\* =  $p \leq .01$ , \*\*\* =  $p \leq .001$

Datenbasis: ALLBUS 1984-1990, 1994-2000

Neben dem Teilnahmeverhalten kann auch das Antwortverhalten der Befragten durch demographische Charakteristika der Interviewer beeinflusst werden (vgl. Reinecke 1998,

---

Berechnung der Standardfehler möglicherweise verzerrt ist. Zum anderen erfolgte die Zuweisung der Interviewer auf die Befragten nicht zufällig. Es kann damit nicht völlig ausgeschlossen werden, dass ältere Interviewer häufiger in Gegenden mit einem höheren Altersdurchschnitt tätig waren; Interviewer mit höherem Bildungsabschluß eher in „gehobenen“ Wohnvierteln mit überdurchschnittlichem Bildungsniveau, usw. Sehr wahrscheinlich erscheint dies allerdings nicht. Faktisch ist eine Zufallszuweisung der Interviewer auf die Befragten in bundesweiten face-to-face Umfragen zwar schwieriger als bei Telefonumfragen, aber nicht prinzipiell unmöglich. Ein Vorgehen, um Interviewereffekte und Effekte der Region trennen zu können, besteht etwa darin, dass Paare von Interviewern sich die Interviews in jeweils zwei Befragungsclustern teilen (vgl. Groves 1989: 361f und Schnell/Kreuter 2000).

der einen Überblick über einschlägige neuere Studien gibt).<sup>16</sup> Potentiell möglich ist dies insbesondere dann, wenn zwischen einem Interviewermerkmal und dem Inhalt einer Frage ein Zusammenhang besteht. Das bekannteste Beispiel in dieser Hinsicht sind Fragen, die sich mit Rassenbeziehungen beschäftigen. Für die USA wurde wiederholt festgestellt, dass sich die Antwortverteilungen zwischen schwarzen und weißen Interviewern bei Fragen, die sich mit der Diskriminierung der schwarzen Bevölkerung beschäftigen, unterscheiden (vgl. Fowler/Mangione 1990: 98ff.; Groves 1989: 398ff.).<sup>17</sup> Schanz/Schmidt (1984) allerdings fanden bei ihrer Analyse ausgewählter Fragen des ALLBUS 1980 nur inkonsistente und schwache Effekte soziodemographischer Merkmale der Interviewer auf das Antwortverhalten der Befragten.<sup>18</sup>

## 5. Zeitpunkt der Interviewdurchführung

Vorgaben für das Kontaktverhalten der Interviewer stellen einen wichtigen Bestandteil des Designs einer Umfrage dar; sie spielen insbesondere für das Erreichen der Zielpersonen eine entscheidende Rolle. Es ist gut belegt, dass Umfragen, die häufige Kontaktversuche der Interviewer, an unterschiedlichen Wochentagen und zu unterschiedlichen Tageszeiten, verteilt über einen längeren Zeitraum, vorsehen, überdurchschnittliche Ausschöpfungsergebnisse aufweisen (vgl. Groves/Couper 1998: 79ff.). Leider verfügen wir im ALLBUS nur für wenige Erhebungen über detaillierte Daten zur Zahl und zeitlichen Terminierung der Kontaktversuche der Interviewer. In allen Studien (mit Ausnahme von 1991) liegen jedoch bei erfolgreicher Interviewdurchführung Angaben zum Wochentag vor. Wie man Tabelle 5 entnehmen kann, wurden 84 Prozent aller ALLBUS-Interviews unter der Woche zwischen Montag und Freitag realisiert. Zwischen den einzelnen Werktagen bestehen dabei nur geringe Unterschiede: Die meisten Interviews fanden an einem Montag, Dienstag oder Mittwoch statt. Von Donnerstag bis Samstag nimmt der Anteil kontinuierlich etwas ab. Der Sonntag scheint für die Interviewertätigkeit weitgehend tabu zu sein, nur 3,6 Prozent aller ALLBUS-Interviews wurden an einem Sonntag realisiert.

---

**16** Es ist zu vermuten, dass der zugrundeliegende Mechanismus für solche Effekte eine Form sozial erwünschten Antwortverhaltens (z.B. im Sinne einer Angleichung an die vermutete Einstellung des Interviewers) darstellt.

**17** Z.T. macht man sich diesen Mechanismus auch bewußt zunutze, indem ein „race-“ oder auch „gender-matching“ zwischen Interviewern und Befragten unter der Annahme vorgenommen wird, dass bei Merkmalsgleichheit der Beteiligten validere Antworten erzielt werden (vgl. Groves 1989: 400ff.).

**18** Schräpler/Wagner (2001) berichten ein Analysebeispiel für das Sozio-ökonomische Panel, bei dem Alter und Bildung der Interviewer einen signifikanten Effekt auf das subjektive Wohlbefinden der Befragten aufweisen.

**Tabelle 5: Wochentag, an dem die Interviews realisiert wurden**

Wochentag	%
Montag	17,8
Dienstag	18,2
Mittwoch	17,8
Donnerstag	15,8
Freitag	14,4
Samstag	12,4
Sonntag	3,6
N	34952

Datenbasis: ALLBUS 1980-2000, ohne ALLBUS 1991

Dieses Muster ist recht stabil, zwischen den Erhebungen bzw. zwischen West- und Ostdeutschland sind nur geringe Unterschiede zu beobachten (nicht tabellarisch ausgewiesen). Die größten Abweichungen bestehen darin, dass 1980 besonders viele Interviews samstags (18,2 Prozent) durchgeführt wurden, während 2000 dieser Anteil mit 6,8 Prozent besonders niedrig ist. Im Gegenzug wurden im Jahr 2000 mit 8,7 Prozent überdurchschnittlich viele Interviews an einem Sonntag gemacht.

Für die Erhebungen ab 1991 liegen auch Informationen zu den Anfangszeiten der Interviews vor (vgl. Tabelle 6). Etwa ein Fünftel aller Interviews der ALLBUS-Erhebungen 1991 bis 2000 wurden in West- und Ostdeutschland vormittags vor 12 Uhr begonnen, zwei Fünftel nachmittags zwischen 12 und vor 17 Uhr und die restlichen zwei Fünftel abends um 17 Uhr oder später.

**Tabelle 6: Uhrzeit, zu der die Interviews begonnen wurden**

Jahr	N	West			N	Ost		
		vor 12 Uhr %	12 bis vor 17 Uhr %	17 Uhr oder später %		vor 12 Uhr %	12 bis vor 17 Uhr %	17 Uhr oder später %
1991	1500	12,0	38,1	49,9	1537	13,3	32,7	54,0
1992	2391	11,5	34,4	54,1	1148	14,3	30,6	55,1
1994	2264	15,4	45,4	39,2	1050	21,6	43,7	34,7
1996	2375	16,2	42,4	41,5	1108	23,9	42,5	33,6
1998	2167	22,9	44,7	32,4	1016	25,8	46,8	27,5
2000	1945	22,7	47,7	29,6	1044	31,1	44,3	24,5
Gesamt	12642	16,8	42,1	41,1	6903	21,0	39,4	39,6

Datenbasis: ALLBUS 1991-2000

Über die Jahre hat der Anteil der Interviews, die vormittags oder nachmittags durchgeführt worden sind, zu-, der Anteil der abends realisierten Interviews abgenommen. Während im ALLBUS 1991 und 1992 in West- und Ostdeutschland jedes zweite Interview um 17 Uhr oder später begonnen wurde, sind es im ALLBUS 2000 lediglich circa 30 Prozent in West- und 25 Prozent in Ostdeutschland. Kombiniert man die Informationen zu Wochentag und Uhrzeit, so ergibt eine Differenzierung zwischen ‚Wochenende‘, ‚Werktag tagsüber‘ und ‚Werktag abends‘ im wesentlichen eine Verlagerung der Interviews von ‚Werktag abends‘ zu ‚Werktag tagsüber‘, während der Anteil der Wochenend-Interviews kaum variiert. 1992 wurden in Westdeutschland 35,6 Prozent der ALLBUS-Interviews tagsüber zwischen Montag und Freitag durchgeführt, im Jahr 2000 waren es 57,9 Prozent. In Ostdeutschland stieg der betreffende Anteil von 33,1 Prozent auf 64,1 Prozent.

Stellen diese Ergebnisse ein Indiz dafür dar, dass Interviewer in immer geringerem Umfang für Interviews zur Abendzeit zur Verfügung stehen? Wenn dem so wäre, wäre dies eine problematische Entwicklung, denn der Abend gilt gemeinhin – auch in der Sicht der Interviewer – als günstigster Zeitraum für die Interviewdurchführung.<sup>19</sup> Eine generelle Tendenz zur Verlagerung der verfügbaren Interviewzeit auf tagsüber läßt sich jedoch aus den vorliegenden Ergebnissen nicht zwingend ableiten, da die studienspezifischen Informationen für den ALLBUS keineswegs „repräsentativ“ für die Verwendung des gesamten Zeitbudgets der Interviewer sein müssen. Allerdings liefert ein Vergleich von Befragungsergebnissen der am ALLBUS 1986 und ALLBUS 2000 beteiligten Interviewer zusätzliche Hinweise auf einen entsprechenden Trend. In diesen Befragungen wurde u.a. ermittelt, an welchen Wochentagen und zu welchen Tageszeiten die Interviewer normalerweise versuchen, Interviews für Bevölkerungsbefragungen durchzuführen. Während im Hinblick auf die Wochentage zwischen 1986 und 2000 kaum Unterschiede bestehen (lediglich der Freitag wird im Jahr 2000 etwas seltener als 1986 erwähnt), sind bei den genannten Uhrzeiten klare Veränderungen zu erkennen. Die vorgegebenen Zeitspannen vormittags und nachmittags (vor 12 Uhr, 12-15 Uhr, 15-17 Uhr) finden 2000 bei mehr Interviewern Anklang als 1986; die Zunahme beträgt zwischen elf und vierzehn Prozentpunkten. Im Gegenzug nimmt der Anteil der Interviewer, die angeben, zwischen 17 und 20 Uhr zu interviewen, um sechs bis sechzehn Prozentpunkte ab. Nicht ganz in dieses Bild paßt, dass der

---

<sup>19</sup> In von Infratest durchgeführten Interviewerbefragungen wurde u.a. erfaßt, welche Tageszeiten nach Meinung der Interviewer für Bevölkerungsbefragungen am günstigsten sind. 1984, 1989 und 1991 nannte jeweils die Mehrheit (zwischen 58 und 63 Prozent) der Infratest-Interviewer die Zeitspanne zwischen 17 und 19 Uhr (vgl. Bliesch o.J.). Schnell (1997: 226ff.) nimmt auf der Basis von Auswertungen von Zeitbudget-Daten an, dass in der Bundesrepublik die günstigste Zeit für ein Interview an Wochentagen zwischen 17 und 19.30 Uhr liegt. Groves/Couper (1998: 100ff.) zeigen für Studien aus den USA, dass Kontaktversuche zur Zielperson abends am erfolgreichsten sind.

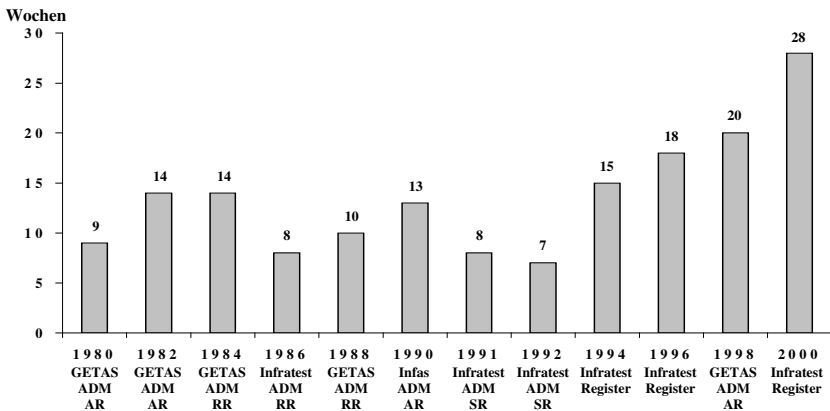


Anteil der Interviewer, die berichten, auch nach 20 Uhr zu arbeiten, im Jahr 2000 um sechs Prozentpunkte höher als 1986 liegt.

## 6. Feldzeit

Betrachten wir nun die Feldzeit der ALLBUS-Umfragen. Rein theoretisch sollte die Länge der Feldzeit einer Umfrage so gewählt werden, dass sowohl schwieriger zu erreichende Befragungspersonen kontaktiert als auch weniger kooperationsbereite Zielpersonen z.B. mithilfe eines Wechsels des Interviewers doch noch zur Teilnahme motiviert werden können. In der Praxis spielen allerdings für die Länge der Feldzeit einer Umfrage eine Vielzahl von Faktoren eine Rolle – angefangen von designbedingten Unterschieden wie dem verwendeten Stichprobenverfahren bis zu konkreten studienspezifischen Einsatzbedingungen wie etwa der Auslastung des eingesetzten Interviewerstabes.

**Abbildung 1: Feldzeit nach Erhebungsjahr: ALLBUS 1980-2000**



Wie Abbildung 1 zeigt, erfolgte die Realisierung der jeweils 3.000 bis 3.500 Interviews des ALLBUS in sehr unterschiedlich langen Zeitspannen. Im Minimum wurden sieben Wochen (beim ALLBUS 1992), im Maximum 28 Wochen (beim ALLBUS 2000) Feldzeit benötigt. Dabei lassen sich einige charakteristische Unterschiede zwischen den verwendeten Stichprobenverfahren ausmachen. So wurde die Studie mit der kürzesten Feldzeit – der ALLBUS 1992 – nach dem ADM-Standard-Random-Verfahren durchgeführt. Dieses Verfahren ist speziell darauf angelegt, eine bestimmte Fallzahl in einer kurzen Zeit zu erzielen. Das Grundprinzip dabei ist, durch die Vorgabe einer Nettofallzahl zu realisierender Interviews und eines großzügig bemessenen bzw. zahlenmäßig nicht begrenzten Bruttos von bearbeitbaren Adressen Nachbearbeitungen weitgehend zu vermeiden. Die Feldarbeit wird sozusagen in einem Anlauf erledigt. Dabei besteht allerdings die Gefahr, dass leichter erreichbare und besonders kooperationsbereite Zielpersonen in der realisierten Stichprobe überproportional vertreten sind. Auf der anderen Seite wurde für die Umfrage mit der längsten Feldzeit – den ALLBUS 2000 – eine Stichprobe aus Einwohnermelderegistern verwendet. Hier werden den Interviewern die zu befragenden Personen mit Namen und Adresse vorgegeben. Durch diese feste Definition der Bruttostichprobe besteht keine Möglichkeit zur Substitution schwer erreichbarer oder wenig kooperationsbereiter Zielpersonen. Eine befriedigende Ausschöpfung und die geplanten (Netto-)Fallzahlen können bei einer solchen Stichprobe nur erzielt werden, wenn jede ausgewählte Adresse intensiv bearbeitet wird. Dies ist ohne zeitaufwendige Nachbearbeitungen kaum möglich.<sup>20</sup> Insgesamt lassen sich die ALLBUS-Erhebungen nach den eingesetzten Stichprobenverfahren in eine Rangfolge bringen, bei der die „weicheren“, weniger geregelten und dafür mit mehr Ermessensspielraum für die Interviewer behafteten Verfahrensmodelle die kürzeste Feldzeit benötigen: Die beiden Umfragen nach dem Standard-Random-Modell weisen eine Feldzeit von durchschnittlich 7,5 Wochen auf, es folgen die drei Random-Route-Erhebungen (10,7 Wochen), die vier Adress-Random-Studien (14,0 Wochen) und die drei Registerstichproben (20,3 Wochen).

Neben solchen in der Verfahrenslogik begründeten Unterschieden spielen für die Dauer der Datenerhebung jedoch weitere, in den studienspezifischen Einsatzbedingungen liegende Faktoren eine Rolle: die zeitliche Plazierung der Feldzeit (Ferienzeit), die Größe und Auslastung des eingesetzten Interviewerstabes, aber auch solch idiosynkratische Ereignisse wie ein Poststreik können die Feldzeit beeinflussen. So ist etwa die extrem

---

<sup>20</sup> Daneben kann ein weiterer Aspekt für die Unterschiede in der Feldzeit zwischen ADM- und Registerstichproben eine Rolle spielen: die Wegestrecken für die Interviewer sind aufgrund der stärkeren Streuung der Adressen bei der Registerstichprobe länger als beim ADM-Design. Dadurch wird die konzentrierte Bearbeitung einzelner räumlicher Einheiten schwieriger, die Zahl der notwendigen Anfahrten in den Point erhöht sich.

lange Feldphase des ALLBUS 2000 nicht ausschließlich auf die Verwendung einer Registerstichprobe zurückzuführen. Eine solche Stichprobe kam auch beim ALLBUS 1994 zum Einsatz, bei dem die Feldzeit mit 15 Wochen nur knapp halb so lang war. Ein wesentlicher Unterschied zwischen beiden Erhebungen bestand freilich darin, dass der ALLBUS 2000 erstmals computerunterstützt durchgeführt wurde. Dadurch war die Zahl der beteiligten Interviewer nur halb so groß wie 1994; überdies war der kleinere CAPI-Stab im Frühjahr 2000 aufgrund des hohen Studienaufkommens bei Infratest stark belastet. Zusätzlich spielte eine Rolle, dass in der 2000er Erhebung aufgrund des schlechten Ergebnisses nach der regulären Feldzeit eine vierwöchige Phase angeschlossen wurde, in der eine Reihe von besonders gering ausgeschöpften Sample Points (insgesamt circa zehn Prozent aller Points) ausgetauscht und neu bearbeitet wurden (vgl. Koch et al. 2001: 53ff.).

Einen etwas detaillierteren Einblick in den Feldverlauf bietet eine Betrachtung auf Ebene der primären Stichprobeneinheiten<sup>21</sup> – bei den ADM-Studien sind dies Wahlbezirke, bei den Registerstichproben jeweils Adressklumpen von zehn Adressen, die innerhalb der Gemeinden aufgrund ihrer räumlichen Nähe zusammengefaßt und den Interviewern zur Bearbeitung vorgegeben wurden. Die Zahl dieser Klumpen variierte im ALLBUS zwischen 600 (1998) und 752 (1992).<sup>22</sup> Im Durchschnitt wurden in den Erhebungen pro Klumpen zwischen 4,6 (1986) und 5,4 (1994, 1996, 1998) Interviews realisiert. Darauf hinzuweisen ist dabei allerdings, dass das Ausgangsbrutto in den Studien nicht gleich groß war: es betrug „offiziell“ zwischen sieben und zehn Adressen pro Klumpen bzw. es war im Fall der Standard-Random-Studien 1991 und 1992 durch eine Obergrenze von 12 bzw. 15 Adressen definiert. Normalerweise wird ein solcher Klumpen von einem Interviewer bearbeitet; ein zweiter Interviewer wird lediglich dann tätig, wenn im Zuge der Nachbearbeitung ein Interviewerwechsel vorgenommen wird. Der Anteil der Klumpen, die von mehr als einem Interviewer bearbeitet wurden, betrug in den Erhebungen zwischen 5,1 Prozent (1984) und 36,7 Prozent (1994).

In Abbildung 2 sind zwei Informationen abgetragen: die durchschnittliche Zahl der Tage, bis das erste Interview im Klumpen vorlag, und die durchschnittliche Zahl der Tage zwischen erstem und letztem Interview im Klumpen. Richtet man die Aufmerksamkeit zunächst auf die Tage bis zum ersten Interview, so fällt auf, dass der offizielle Start der Feldarbeit keinesfalls gleichzusetzen ist mit dem Beginn der Bearbeitung sämtlicher

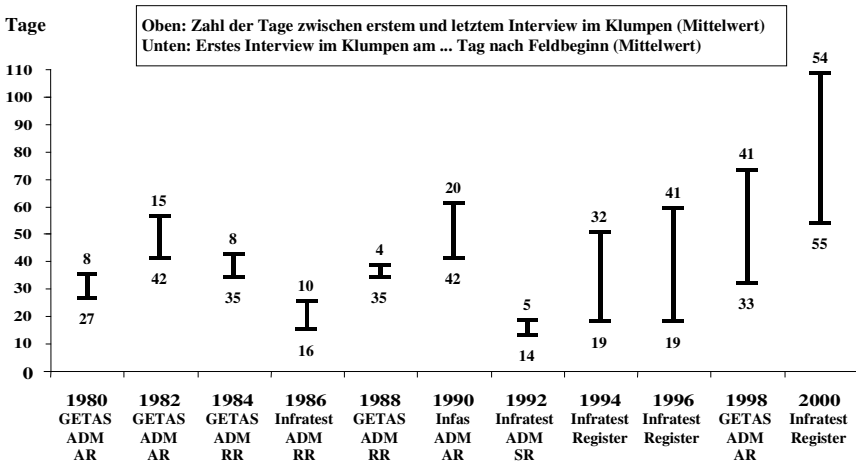
---

<sup>21</sup> Vgl. zu dieser Herangehensweise auch Schnell 1997 (229ff.), der die Feldarbeit für den ALLBUS 1988 näher beleuchtet.

<sup>22</sup> Die genannten Zahlen beziehen sich auf die Klumpen, in denen mindestens ein Interview realisiert wurde (auch Abbildung 2 liegt diese Abgrenzung zugrunde). Die Zahl der Klumpen ohne realisiertes Interview lag in den einzelnen Erhebungen zwischen zwei (1996) und 48 (1992).

Klumpen. In den einzelnen ALLBUS-Erhebungen dauerte es im Durchschnitt zwischen 14 (1992) und 55 (2000) Tagen, bis das erste Interview in den Klumpen realisiert wurde. Die Studien unterscheiden sich also deutlich darin, wie intensiv in die Feldarbeit bei Feldbeginn eingestiegen wurde.<sup>23</sup>

**Abbildung 2: Feldbeginn und Feldzeit auf Klumpen-Ebene: ALLBUS 1980-2000\***



\* ohne ALLBUS 1991, bei dem keine Angaben zum Datum des Interviews verfügbar sind

Dabei sind Unterschiede zwischen den Umfrageinstituten erkennbar: Bei den von GETAS bzw. Infas durchgeführten Erhebungen wurde das erste Interview im Klumpen durchschnittlich nach 27 bis 42 Tagen realisiert; das heißt im Mittel lag erst vier bis sechs Wochen nach Feldbeginn ein greifbares Resultat der Feldarbeit vor. Schneller war dagegen – mit einer Ausnahme – Infratest: im Durchschnitt wurde das erste Interview in den Klumpen nach 14 bis 19 Tagen erbracht. Dabei machte es keinen Unterschied, ob eine ADM-Stichprobe (wie z.B. im Jahr 1986) oder eine Registerstichprobe (wie z.B. 1994) zugrundelag. Die Ausnahme bildet der von Infratest durchgeführte ALLBUS 2000: im Durchschnitt lag hier das erste Interview in den Klumpen nach acht Wochen vor. Dieser lange Zeitraum

<sup>23</sup> Darauf hinzuweisen ist, dass die hier vorgenommene Analyse sich auf die Interviewdurchführung bezieht. In vielen Fällen ist es möglich, dass dieser bereits Kontaktversuche zu früheren Zeitpunkten vorausgingen.

hängt mit der schon erwähnten geringeren Zahl von CAPI-Interviewern und deren großer Auslastung in dem betreffenden Zeitraum zusammen (deswegen wurde beispielsweise ein Viertel aller Adressen erst vier bis zehn Wochen nach Feldbeginn an die Interviewer zur Bearbeitung gegeben). Außerdem spielt der bereits angesprochene Austausch und die Neubearbeitung von Befragungsklumpen am Ende der Feldzeit eine Rolle: ohne diesen reduziert sich der Mittelwert der Tage bis zum ersten Interview von 55 auf 42.

Die Zahl der Tage zwischen erstem und letztem Interview in den einzelnen Klumpen vermittelt einen Eindruck von der effektiven Bearbeitungszeit, die für die Klumpen aufgewendet wurde. Man erkennt in Abbildung 2 zunächst, dass die faktischen Bearbeitungszeiten der Klumpen weit kürzer sind als die Feldphase insgesamt. Sie liegen zwischen durchschnittlich vier (1988) und 54 (2000) Tagen und betragen damit zwischen 1/17 (1988) und 1/3 (1996) der jeweiligen Feldzeit. Dies bedeutet, dass die Länge der Feldzeit einer Erhebung zu einem erheblichen Teil davon abhängig ist, inwieweit die Befragungsklumpen zeitgleich oder aber zeitversetzt bearbeitet werden.

Bei einigen der – face-to-face durchgeführten – ALLBUS-Erhebungen ist die faktische Feldphase kaum länger als die bei Telefonumfragen übliche Zeitspanne von wenigen Tagen. Den (Negativ-) Rekord halten in dieser Hinsicht die ALLBUS-Erhebungen 1988 und 1992. Im Durchschnitt lagen hier gerade vier bzw. fünf Tage zwischen erstem und letztem Interview in den Klumpen. In 77 bzw. 74 Prozent aller Klumpen wurden alle Interviews innerhalb einer Woche, in 39 bzw. 8 Prozent aller Klumpen sogar an einem Tag durchgeführt. Man kann mit einiger Berechtigung bezweifeln, ob bei diesen Erhebungen die faktische Datenerhebungsphase damit lang genug war, um alle Zielpersonen wirklich erreichen zu können (zum Zusammenhang von Länge der Feldzeit und Erreichbarkeit von Zielpersonen vgl. Groves/Couper 1998: 272ff.).

Auch in der Frage der effektiven Bearbeitungszeit sind wieder Unterschiede zwischen den eingesetzten Stichprobenverfahren zu beobachten. Die längste Feldzeit auf Klumpenebene weisen die drei Registerstichproben mit einer durchschnittlichen Bearbeitungszeit von 42,3 Tagen auf. An zweiter Stelle folgen die ADM-Stichproben mit einem separaten Adressenvorlauf mit 21,0 Tagen und das Schlußlicht bilden die Random-Route- bzw. Standard-Random-Studien mit 7,3 bzw. 5,0 Tagen. Dass gerade zwischen diesen beiden am weitesten verbreiteten Verfahrensmodellen kaum Unterschiede in der effektiven Bearbeitungszeit zu verzeichnen sind, ist womöglich ein Hinweis darauf, dass die Interviewer in der Feldpraxis nur wenig zwischen beiden Vorgehensweisen unterscheiden: das Random-Route-Verfahren, das sich eigentlich durch die feste Vorgabe einer bestimmten Zahl zu bearbeitender Bruttoadressen auszeichnet, kann durch die undokumentierte Substitution von nicht erfolgreich bearbeiteten Adressen faktisch dem Standard-Random-Modell

(mit einem zahlenmäßig nicht begrenzten oder eher großzügig bemessenen Brutto) sehr stark angenähert sein. Auf die Problematik der Adressensubstitution wird am Schluß des Beitrags nochmals einzugehen sein.

## 7. Fallpreise

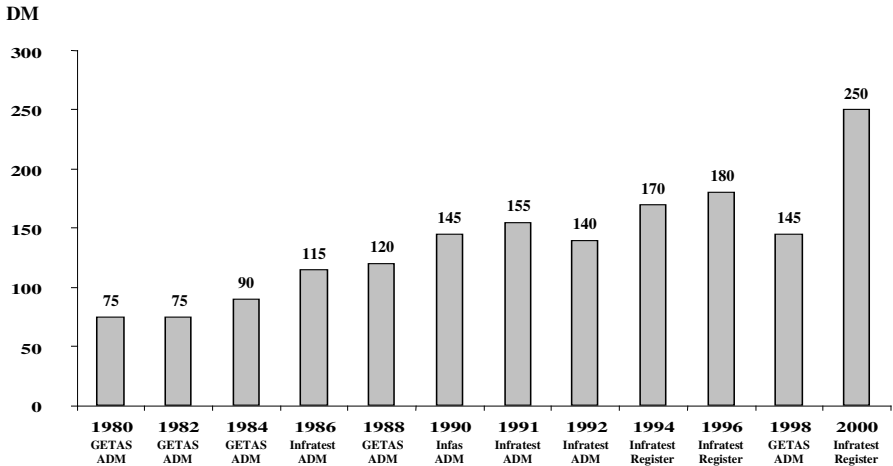
Eine Betrachtung der Feldarbeit ohne Berücksichtigung der entstehenden Kosten wäre unvollständig. Im folgenden wollen wir einen Blick auf die Fallpreise in den einzelnen ALLBUS-Erhebungen werfen. Diese wurden ermittelt, indem der bei dem Umfrageinstitut angefallene Nettopreis der Studie durch die Zahl der erzielten auswertbaren Interviews dividiert wurde. In die Fallpreise gehen damit neben den Feldkosten im engeren Sinne (wie Honorare und Spesen der Interviewer, Kosten von Feldsteuerung, Interviewerschulung und -kontrolle) auch alle weiteren Aufwendungen im Umfrageinstitut ein: für die Stichprobenbildung, Datenaufnahme und -prüfung, Projektleitung, usw.<sup>24</sup> In der Regel bilden die Feldkosten allerdings den größten Posten in der Kalkulation einer Studie.

Abbildung 3 zeigt die Entwicklung der Fallpreise im ALLBUS zwischen 1980 und 2000. Während 1980 das ALLBUS-Interview durchschnittlich DM 75,- kostete, betrug der Preis zwanzig Jahre später DM 250,-. Dies würde eine Erhöhung der Fallpreise um den Faktor 3,3 bedeuten – eine weit größere Steigerung als sie etwa für die privaten Lebenshaltungskosten im gleichen Zeitraum zu verzeichnen ist (Faktor 1,6; berechnet nach Informationen in: Statistisches Bundesamt 2000: 328ff.). Zu berücksichtigen ist jedoch, dass die Erhebungen 1980 und 2000 sich in mehrerlei Hinsicht unterscheiden. So wurde der ALLBUS 1980 von GETAS durchgeführt, der ALLBUS 2000 von Infratest; 1980 wurde das ADM-Design verwendet, 2000 eine Registerstichprobe; 1980 existierte das ISSP noch nicht, 2000 wurde der ISSP-Fragebogen als Selbstausfüller im Anschluß an das ALLBUS-Interview erhoben.<sup>25</sup> Zieht man deshalb die in Hinblick auf das durchführende Umfrageinstitut, das verwendete Stichprobenverfahren sowie die Befragungsdauer vergleichbaren Erhebungen 1980 und 1998 heran – beide wurden von GETAS nach dem ADM-Design (mit Adressen-Vorlauf) bei einer durchschnittlichen Interviewdauer von jeweils 50 Minuten realisiert –, dann ist mit der Zunahme des Fallpreises von DM 75,- auf DM 145,- nurmehr eine knappe Verdoppelung zu beobachten.

---

<sup>24</sup> Außen vor bleiben alle bei den Datenproduzenten ZUMA und ZA entstandenen Aufwendungen, etwa für die Fragebogenentwicklung, die weitergehende Datenprüfung und -dokumentation, usw.

<sup>25</sup> Das International Social Survey Programme (ISSP) ist ein weltweiter Forschungsverbund, der seit 1985 jedes Jahr sozialwissenschaftliche Umfragen mit wechselnden Themenschwerpunkten durchführt. In Deutschland wird die ISSP-Umfrage zumeist als schriftlicher Selbstausfüller im Anschluß an das mündliche ALLBUS-Interview erhoben.

**Abbildung 3: Fallpreise in den ALLBUS-Erhebungen 1980 – 2000 (o. MWSt.)**

Generell ist davon auszugehen, dass Registerstichproben teurer als ADM-Stichproben sind, aber auch zwischen den verschiedenen Varianten des ADM-Designs existieren Preisunterschiede: am teuersten ist das Adress-Random mit der separaten Begehungsphase zur Ermittlung der Haushaltsadressen, darauf folgt das traditionelle Random-Route mit der festen Bruttovorgabe von zu bearbeitenden Adressen und am günstigsten ist das Standard-Random mit der Nettovorgabe zu realisierender Interviews. Neben solchen verfahrensbedingten Preisunterschieden sollte man überdies gewärtigen, dass auch zwischen verschiedenen Umfrageinstituten trotz vergleichbarer Leistungsangebote in der Preisgestaltung z.T. merkliche Differenzen zu verzeichnen sind.

Bedauerlicherweise existieren keine veröffentlichten Informationen zu den Fallpreisen anderer Bevölkerungsumfragen in der Bundesrepublik. Dennoch ist es wohl kaum verkehrt anzunehmen, dass z.B. der Fallpreis des ALLBUS 2000 mit DM 250,- am oberen

Limit dessen liegt, was in Deutschland bei ähnlichen Erhebungen bezahlt wird.<sup>26</sup> Es stellt sich die Frage, inwieweit dieser hohe Preis gerechtfertigt ist.

## 8. Qualität, Zeitbedarf und Preis

Selbstverständlich ist der Zusammenhang von Qualität und Preis in der Umfrageforschung zu komplex, als dass er im hier gegebenen Rahmen auch nur ansatzweise befriedigend behandelt werden könnte. Wenn das Thema zum Schluß dieses Beitrages dennoch aufgegriffen wird, dann um auf ein wichtiges praktisches Problem aufmerksam zu machen, dem sich eine qualitätsorientierte Surveyforschung in Deutschland gegenüberieht. Die Rede ist von der Zuverlässigkeit der Ermittlung der Ausschöpfungsquote, eines zentralen Qualitätsindikators in der Umfrageforschung.<sup>27</sup>

Das (Pseudo-)Motto eines amerikanischen Umfrageinstitutes faßt den Zusammenhang zwischen Qualität, Zeitbedarf und Preis so zusammen: „Quality, speed, price: pick two.“ (zit. nach Smith 1995: 167). Dieses Motto impliziert, dass bei hohen Preisen mit hoher Qualität in kurzer Zeit gerechnet werden kann. Im Gegenzug legt es nahe, dass bei niedrigen Preisen Abstriche in der Qualität oder der Geschwindigkeit in Kauf genommen werden müssen. In der Bundesrepublik stellt sich jedoch – zumindest bei oberflächlicher Betrachtungsweise – die Entscheidungssituation bei Bevölkerungsumfragen anders dar: Zieht man als Qualitätsindikator die Ausschöpfungsquote heran, besteht (vermeintlich) die Wahl zwischen einer hohen Ausschöpfung, kurzen Feldzeit und einem niedrigen Preis einerseits oder einer niedrigen Ausschöpfung, langen Feldzeit und einem hohen Preis andererseits. Ersteres erhält man, wenn eine Stichprobe nach dem ADM-Random-Route-

---

**26** Die ursprüngliche Kalkulation des ALLBUS 2000 sah einen etwas niedrigeren Fallpreis von DM 225,- vor. Dieser erhöhte sich dann allerdings auf DM 250,-, weil die angestrebte Nettofallzahl von 3.500 auswertbaren Interviews nicht realisiert wurde.

International nimmt der ALLBUS mit diesem Preis keine Spitzenposition ein. Nach einer mündlichen Auskunft von Tom Smith, dem Verantwortlichen für den amerikanischen General Social Survey (GSS), lag beispielsweise der Preis für ein Interview in der GSS-Erhebung 1996 um ein Mehrfaches höher.

**27** „For many years, it seemed as if the design of the sample and the response rate were the only two indicators of a survey’s quality.“ (Fowler/Mangione: 1990: 142). In der Praxis wird die Ausschöpfungsquote allerdings mitunter unreflektiert verwendet. Oft wird außer acht gelassen, dass sie lediglich ein Maß für die *potenzielle* Verzerrung von Daten durch einen Nonresponsebias ist. Die *tatsächliche* Stärke einer Verzerrung ist abhängig von der Höhe der Nonresponsequote einerseits und dem Ausmaß des Unterschiedes zwischen Teilnehmern und Nichtteilnehmern einer Umfrage andererseits, welches jeweils variablenpezifisch zu bestimmen ist (vgl. Groves/Couper 1998: 1ff.).



(bzw. Standard-Random-)Modell gewählt wird, letzteres stellt sich bei Verwendung einer Registerstichprobe ein.<sup>28</sup>

Zur Veranschaulichung dieser These sollen exemplarisch vier Bevölkerungsumfragen betrachtet werden, die von ZUMA in der jüngeren Vergangenheit betreut wurden und die alle für sich in Anspruch nehmen, hohen Qualitätsstandards zu genügen. Dies sind die beiden ALLBUS-Erhebungen 1996 und 2000, die auf Basis einer Registerstichprobe durchgeführt wurden, und der Wohlfahrtssurvey 1998 sowie der Sozialwissenschaften-Bus 1997, bei denen eine ADM-Stichprobe nach dem Random-Route-Modell zum Einsatz kam. Wie Tabelle 7 zeigt, beträgt bei dem auf Basis einer Registerstichprobe durchgeführten ALLBUS 2000 die Ausschöpfung lediglich 49 Prozent – und das bei einer Feldzeit von 28 Wochen und einem Fallpreis von DM 250,-. Etwas besser fällt das Ergebnis beim ALLBUS 1996 aus: hier wurden in 18 Wochen Feldzeit 54 Prozent Ausschöpfung erzielt; das einzelne Interview kostete durchschnittlich DM 180,-. Deutlich günstigere Kosten-Leistungs-Relationen werden dagegen für die beiden ADM-Studien mitgeteilt. So beträgt die berichtete Ausschöpfung des Wohlfahrtssurveys 1998, der wie die beiden ALLBUS-Erhebungen von Infratest durchgeführt wurde, 56 Prozent. Bei einem Fallpreis von DM 125,- ist er jedoch wesentlich billiger und auch die Feldzeit ist mit acht Wochen weit niedriger als in den ALLBUS-Studien. Noch mehr zu noch günstigeren Preisen bietet – vermeintlich – der Sozialwissenschaften-Bus 1997, der von GETAS erhoben wurde: Bei einem Fallpreis von DM 115,- und einer Feldzeit von lediglich vier Wochen wird hier eine Ausschöpfung von knapp 70 Prozent ermittelt.

Die Erklärung für dieses paradoxe Ergebnis liegt darin, dass die bei den ADM-Studien dokumentierten Ausschöpfungsquoten sehr unzuverlässig und in der Regel überhöht sind. Die Denkschrift der Deutschen Forschungsgemeinschaft zu „Qualitätskriterien der Umfrageforschung“ – an der auch Vertreter namhafter deutscher Umfrageinstitute mitgewirkt haben – führt zur Frage der Ausschöpfungsberechnung aus: „Die Aussagekraft der errechneten Stichprobenausschöpfung kann unterschiedlich sein. Sie hängt davon ab, wie die verwendeten Bruttoadressen ausgewählt und dokumentiert werden. In dieser Hinsicht gibt es zwischen verschiedenen Verfahrensmodellen erhebliche Unterschiede. Je „härter“ die Verfahrensregeln für die Bestimmung und Dokumentation des Stichpro-

---

**28** Das weitere Modell – ADM-Stichproben nach dem Adress-Random-Verfahren – bleibt im folgenden außen vor. Es wird in der Praxis eher seltener verwendet und nimmt – was den Ermessensspielraum der Interviewer bei der Auswahl der Zielpersonen sowie die Kontrollmöglichkeiten von Umfrageinstitut und Auftraggeber angeht – eine „mittlere“ Position zwischen ADM-RR-/ADM-SR-Stichproben einerseits, Registerstichproben andererseits ein.

benbrutto sind – was unter Qualitätsgesichtspunkten positiv zu werten wäre -, um so niedriger ist die errechnete Stichprobenausschöpfung.“ (Kaase 1999: 103f)

**Tabelle 7: Ausschöpfung, Feldzeit und Fallpreise in vier Bevölkerungsumfragen\***

Studie	Berichtete Ausschöpfungsquote	Feldzeit	Fallpreis (ca., ohne MWSt.)	Stichprobenverfahren
ALLBUS 2000	49,1%	28 Wo	250 DM	Register
ALLBUS 1996	54,2%	18 Wo	180 DM	Register
Wohlfahrts-survey 1998	56,1%	8 Wo	125 DM	ADM-Random-Route
Sozialwissen-schaften-Bus 1997	68,6%	4 Wo	115 DM**	ADM-Random-Route

\* Durchführendes Umfrageinstitut: ALLBUS, Wohlfahrtssurvey: Infratest; Sozialwissenschaften-Bus: GETAS

Fallzahl: ALLBUS 2000 N=3.138, ALLBUS 1996 N=3.518, Wohlfahrtssurvey N=3.042, Sozialwissenschaften-Bus N=2.044

Durchschnittliche Interviewdauer: ALLBUS 2000 ca. 65 Minuten, ALLBUS 1996 ca. 70 Minuten, Wohlfahrtssurvey ca. 50 Minuten, Sozialwissenschaften-Bus ca. 60 Minuten

\*\* Preis errechnet für eine Einschaltung von 55 Minuten Dauer

Bei ADM-Stichproben (zumindest bei dem gängigen Standard-Random- bzw. Random-Route-Modell) wird die Bruttostichprobe im Verlaufe der Feldarbeit durch die Interviewer erstellt. Durch die Vorgabe spezifischer Verfahrensregeln besteht dabei zwar theoretisch kein Spielraum bei der Bestimmung der Zielhaushalte und Zielpersonen. Faktisch aber werden den Interviewern von den Instituten – mehr oder weniger explizit – Kompensationsmöglichkeiten für Ausfälle eingeräumt, um eine schnellere und kostengünstigere Studienbearbeitung zu ermöglichen. Diese Substitution von Ausfällen wird jedoch häufig nicht oder nur eingeschränkt dokumentiert; überhöhte Ausschöpfungsquoten sind die Folge. In dem bereits zitierten DFG-Memorandum heißt es dazu: „Die Standardverfahren der Umfrageinstitute für persönlich-mündliche Befragungen auf random-Basis dürften daher durchweg mit Substitution arbeiten. ... Jeder dokumentierte Substitutionsfall erhöht allerdings die Bruttozahl der eingesetzten Adressen und verringert dementsprechend die errechnete Stichprobenausschöpfung. Es ist daher mit einer gewissen Tendenz zu rechnen, Substitution nicht vollständig zu erfassen und zu dokumentieren.“ (Kaase 1999: 103)<sup>29</sup>

<sup>29</sup> Ein einfaches Rechenexempel verdeutlicht, dass bereits relativ „kleine“ Manipulationen die Höhe der berichteten Ausschöpfung merklich beeinflussen können. Angenommen, für einen Inter-

Bei der Registerstichprobe wird die Bruttostichprobe ohne Mitwirken der Interviewer vor der Feldarbeit gezogen, eine unkontrollierte Substitution von Adressen ist nahezu ausgeschlossen und die Ausschöpfung dadurch sehr zuverlässig ermittelbar (zu den Möglichkeiten der Interviewerkontrolle bei Registerstichproben siehe Koch 1995). Die absurde Folge dieser Situation ist, dass die Qualitätsorientierung von Registerstichproben scheinbar doppelt bestraft wird: durch höhere Preise und längere Feldzeiten einerseits, und vermeintlich niedrigere Ausschöpfungsquoten andererseits.<sup>30</sup>

Was kann und sollte in dieser Situation getan werden? Notwendig ist eine offene und sachliche Diskussion von Fragen der Feldarbeit, Datenqualität und Kosten. Diese Aufgabe betrifft sowohl die Umfrageinstitute als auch die potentiellen Auftraggeber von Studien. Einen Ausgangspunkt für die Debatte können die bereits erwähnte DFG-Denkschrift und die „Standards zur Qualitätssicherung in der Markt- und Sozialforschung“ des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (1999) bilden. Bislang findet eine entsprechende Diskussion in der Bundesrepublik, aber auch in anderen Ländern, nur in Ansätzen statt (vgl. Schnell 1997: 265f). Zu oft überwiegt auf beiden Seiten das Desinteresse, sei es aus Angst, dass Schwachstellen offengelegt werden, oder aus Furcht vor Kritik, eine Umfrage entspräche nicht geltenden Qualitätsstandards.

Darüber hinaus ist jedoch auch eine stärkere Nachfrage nach anspruchsvollen Verfahrensmodellen unerlässlich, denn in der Umfragepraxis entscheidet letztlich der Markt über die Bedeutung von Qualitätskriterien. „Verfahrensmodelle mit hohen Qualitätsstandards werden auf seiten der anbietenden Institute nur entwickelt und angeboten, wenn es dafür eine Nachfrage gibt, also Auftraggeber, die auf die spezifischen Qualitätskriterien Wert legen und den dafür erforderlichen Aufwand auch zu finanzieren bereit sind.“ (Kaase 1999: 98) In der jüngeren Vergangenheit hat speziell der ALLBUS versucht, im Sinne eines „best practice“ die Stichprobenqualität zu maximieren (vgl. Kaase 1999: 99ff.). Es wäre wünschenswert, wenn weitere Studien auf diesem Weg mitgingen, denn die Imple-

---

viewer wären in einem Wahlbezirk acht Adressen als Brutto vorgesehen, aus denen er unter Beachtung aller Regeln vier Interviews erzielte. Dann läge seine Ausschöpfung bei 50 Prozent. Erzielte er noch ein weiteres Interview mit einer Person, die nicht zu seinem Brutto zählt, könnte er seine Ausschöpfung bereits auf 62,5 Prozent „verbessern“, wenn er bei der Dokumentation des Bruttos einen der vier Ausfälle unterschlägt.

**30** Ein Vergleich der Randverteilungen ausgewählter soziodemographischer Merkmale aus sechs allgemeinen Bevölkerungsumfragen, deren Ausschöpfung zwischen 50 und 80 Prozent variierte, mit Daten des Mikrozensus ergab keinen empirischen Hinweis darauf, dass die Stichprobenverzerrungen der Umfragen mit höherer (berichteter) Ausschöpfung geringer ausfallen (vgl. Koch 1998). Im Gegenteil: Tendenziell am besten schnitt der ALLBUS 1994 ab, der zwar lediglich eine Ausschöpfung von circa 54 Prozent aufwies, aber auf Basis einer Registerstichprobe durchgeführt worden war.

mentierung anspruchsvoller Verfahren der Stichprobenziehung und Feldarbeit wird umso erfolgreicher sein, je eher eine bestimmte „kritische“ Masse von Studien erreicht wird.

Damit Entscheidungen über das für eine Studie am besten geeignete Stichproben- und Felddesign auf einer fundierten Grundlage getroffen werden können, sind die Umfrageinstitute gefordert, stärker als bisher ihre Prozeduren und die damit gewonnenen Erfahrungen offenzulegen und zu dokumentieren. Außerdem ist methodische Forschung notwendig, in der unter kontrollierten Bedingungen unterschiedliche Modelle der Stichprobenbildung und Feldarbeit, die mit ihnen erzielbare Ergebnisqualität und der entstehende Kosten- und Zeitaufwand erforscht werden. Besonderes Augenmerk ist dabei auf die Frage der Verallgemeinerbarkeit der erzielten Resultate zu richten, denn: Zu unterscheiden ist zwischen der prinzipiell erreichbaren Qualität im jeweiligen Verfahrensmodell und der tatsächlich erzielten Qualität bei einer konkreten Umfrage. Mit anderen Worten: Nicht nur vom Verfahrensmodell, sondern auch von seiner Implementation hängt es ab, welche Ergebnisse im konkreten Fall erzielt werden (vgl. Kaase 1999: 107f). Die oben berichteten Ergebnisse vermitteln einen Eindruck davon, welche Unterschiede hinsichtlich Faktoren wie Ausschöpfung, Feldzeit und Kosten auch innerhalb des gleichen Verfahrensmodells auftreten können. Was die eigentliche Datenqualität und deren mögliche Variation innerhalb und zwischen den Verfahrensmodellen angeht, sind weitere Untersuchungen in der Zukunft erforderlich.<sup>31</sup>

## Literatur

Alt, Christian/Bien, Walter/Krebs, Dagmar 1991: Wie zuverlässig ist die Verwirklichung von Stichprobenverfahren? Random route versus Einwohnermeldeamtsstichprobe. ZUMA-Nachrichten 28: 65-72.

Arbeitsgemeinschaft ADM-Stichproben und Bureau Wendt 1994: Das ADM-Stichproben-System. Stand 1993. S. 188-202 in: Siegfried Gabler/Jürgen H.P. Hoffmeyer-Zlotnik/Dagmar Krebs (Hrsg.): Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag.

Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V., (Hrsg.) 1999: Standards zur Qualitätssicherung in der Markt- und Sozialforschung. Frankfurt/M.: Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute.

---

**31** Die bislang einzige methodische Vergleichsstudie beispielsweise zwischen einer Register- und einer ADM-Stichprobe in der Bundesrepublik fand 1988 im Rahmen des DJI-Familienurveys statt. Alt et al. (1991) liefern empirische Indizien dafür, dass die schlechteren Ergebnisse der ADM-Stichprobe – was die Verteilung der Merkmale Geschlecht und Erwerbstätigkeit angeht – auf Manipulationen der Interviewer bei der Auswahl der Befragungspersonen zurückzuführen sind.

Behrens, Kurt/Löffler, Ute 1999: Aufbau des ADM-Stichproben-Systems. S. 69-91 in: ADM Arbeitsgemeinschaft Deutscher Markt- und Sozialforschungsinstitute e.V. und AGMA Arbeitsgemeinschaft Media-Analyse e.V. (Hrsg.): Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis. Opladen: Leske + Budrich.

Bliesch, Uwe (o.J.): Die Interviewerorganisation in einem Markt- und Sozialforschungsinstitut. München: Infratest.

Blyth, Bill 1998: Current and Future Technology Utilization in European Market Research. S. 563-581 in: Mick P. Couper/Reginald P. Baker/Jelke Bethlehem/Cynthia Z. F. Clark/Jean Martin/William L. Nicholls II/James M. O'Reilly (eds.): Computer Assisted Survey Information Collection. New York: Wiley.

Collins, Martin/Sykes, Wendy/O'Muirheartaigh, Colm 1998: Diffusion of Technological Innovation: Computer Assisted Data Collection in the U.K. S. 23-43 in: Mick P. Couper/Reginald P. Baker/Jelke Bethlehem/Cynthia Z. F. Clark/Jean Martin/William L. Nicholls II/James M. O'Reilly (eds.): Computer Assisted Survey Information Collection. New York: Wiley.

Context 1993: Instituts-Handschriften: Was ist zu tun? Interviewer-Merkmale: Wie sie durchschlagen. Heft 10: 6-10.

Fowler, Floyd J./Mangione, Thomas W., 1990: Standardized Survey Interviewing. Minimizing Interviewer Related Error. Applied Social Research Methods Series, Volume 18. Newbury Park: Sage.

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

Groves, Robert M./Couper, Mick P., 1998: Nonresponse in Household Interview Surveys. New York: Wiley.

Hartmann, Peter H./Schimpl-Neimanns, Bernhard 1992: Sind Sozialstrukturanalysen mit Umfragedaten möglich? Analysen zur Repräsentativität einer Sozialforschungsumfrage. Kölner Zeitschrift für Soziologie und Sozialpsychologie 44: 315-340.

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

Koch, Achim 1991: Zum Zusammenhang von Interviewermerkmalen und Ausschöpfungsquoten. ZUMA-Nachrichten 28: 41-53.

Koch, Achim 1995: Gefälschte Interviews: Ergebnisse der Interviewerkontrolle beim ALLBUS 1994. ZUMA-Nachrichten 36: 89-105.

Koch, Achim 1997: ADM-Design und Einwohnermelderegisterstichprobe. Stichprobenverfahren bei mündlichen Bevölkerungsumfragen. S. 99-116 in: Siegfried Gabler/Jürgen H.P. Hoffmeyer-Zlotnik (Hrsg.): Stichproben in der Umfragepraxis. Opladen: Westdeutscher Verlag.

Koch, Achim 1998: Wenn „mehr“ nicht gleichbedeutend mit „besser“ ist: Ausschöpfungsquoten und Stichprobenverzerrungen in allgemeinen Bevölkerungsumfragen. ZUMA-Nachrichten 42: 66-90.

Koch, Achim/Wasmer, Martina/Harkness, Janet/Scholz, Elvira 2001: Konzeption und Durchführung der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 2000. ZUMA-Methodenbericht 2001/05. Mannheim.

Lynn, Peter/Turner, Rachel/Smith, Patten 1998: Assessing the Effects of an Advance Letter for a Personal Interview Survey. Journal of the Market Research Society 40: 265-272.

Niehoff, Renate 1998: Intervieweranforderungen und –auswahl. S. 51-65 in: Statistisches Bundesamt (Hrsg.): Interviewereinsatz und –qualifikation. Spektrum Bundesstatistik, Band 11. Stuttgart: Metzler-Poeschel.

Reinecke, Jost 1998: Regelmäßigkeiten des Interviewerverhaltens. S. 115-137 in: Statistisches Bundesamt (Hrsg.): Interviewereinsatz und –qualifikation. Spektrum Bundesstatistik, Band 11. Stuttgart: Metzler-Poeschel.

Schanz, Volker/Schmidt, Peter 1984: Interviewsituation, Interviewermerkmale und Reaktionen von Befragten im Interview: eine multivariate Analyse. S. 72-113 in: Karl Ulrich Mayer/Peter Schmidt (Hrsg.): Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. Beiträge zu methodischen Problemen des ALLBUS 1980. Frankfurt/M.: Campus.

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

Schnell, Rainer/Kreuter, Frauke 2000: Das DEFECT-Projekt: Sampling-Errors und Nonsampling-Errors in komplexen Bevölkerungsstichproben. ZUMA-Nachrichten 47: 89-102.

Schräpler, Jörg-Peter/Wagner, Gert G., 2001: Das Verhalten von Interviewern – Darstellung und ausgewählte Analysen am Beispiel des „Interviewerpanels“ des Sozio-oekonomischen Panels. Allgemeines Statistisches Archiv 85: 45-66.

Smith, Tom 1995: Trends in Non-Response Rates. International Journal of Public Opinion Research 7: 157-171.

Sodeur, Wolfgang 1997: Interne Kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen. ZA-Information 41: 58-82.

Statistisches Bundesamt, (Hrsg.) 2000: Datenreport 1999. Zahlen und Fakten über die Bundesrepublik Deutschland. Bonn: Bundeszentrale für politische Bildung.