



Statistisches Bundesamt

Neue Erhebungsinstrumente und Methodeneffekte

Band 15 der Schriftenreihe
Spektrum Bundesstatistik

METZLER
POESCHEL

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Ein Titeldatensatz für diese Publikation ist bei
Der Deutschen Bibliothek erhältlich.

Herausgeber: Statistisches Bundesamt, Wiesbaden



Fachliche Informationen zu dieser Veröffentlichung können Sie direkt beim Statistischen Bundesamt erfragen:

Gruppe I B, Telefon: 06 11 / 75 20 77 , Fax: 06 11 / 75 39 50 oder
E-Mail: christian.koenig@statistik-bund.de

Verlag: Metzler-Poeschel, Stuttgart

Verlagsauslieferung: SFG – Servicecenter Fachverlage GmbH
Postfach 43 43
72774 Reutlingen
Telefon: 0 70 71 / 93 53 50
Telefax: 0 70 71 / 93 53 35
Internet: <http://www.s-f-g.com>
E-Mail: staba@s-f-g.com

Erschienen im Juni 2000

Preis: DM 28,80 / EUR 14,73

Bestellnummer: 1030515-00900

ISBN: 3-8246-0619-4

Recyclingpapier aus 100 % Altpapier.



Informationen über das Statistische Bundesamt und sein Datenangebot erhalten Sie:

- im Internet: <http://www.statistik-bund.de>

oder bei unserem Informationsservice

65180 Wiesbaden

- Telefon: 06 11 / 75 24 05

- Telefax: 06 11 / 75 33 30

- E-Mail: info@statistik-bund.de

© Statistisches Bundesamt, Wiesbaden 2000

Für nichtgewerbliche Zwecke sind Vervielfältigung und unentgeltliche Verbreitung, auch auszugsweise, mit Quellenangabe gestattet. Die Verbreitung, auch auszugsweise, über elektronische Systeme/Datenträger bedarf der vorherigen Zustimmung. Alle übrigen Rechte bleiben vorbehalten.

Inhalt

	Seite
<i>Johann Hahlen</i> Begrüßung	4
<i>Jürgen Chlumsky</i> Einführung	7
<i>Erich Wiegand</i> Chancen und Risiken neuer Erhebungstechniken in der Umfrageforschung	12
<i>Thomas Petersen</i> Keine Alternativen: Telefon- und Face-to-Face-Umfragen	22
<i>Wolfgang Schulte</i> Folgen eines Wechsels von Face-to-Face- zu telefonischen Befragungen	42
<i>Rolf Pfeiderer</i> Methodeneffekte beim Umstieg auf CAPI-Techniken	57
<i>Marek Fuchs</i> Interviewsituation in computergestützten Befragungen – zur Wirkung von Technologieeffekten	71
<i>Heinrich A. Litzentroth</i> Testmarktforschung	92
<i>Bernd Schmidt</i> Registergestützte Datengewinnung	131
<i>Michael Hoppe</i> Aufbau und Organisation eines Access-Panels	145
<i>Wolfgang Bandilla/Michael Bosnjak</i> Perspektiven der Online-Forschung	166
Teilnehmerverzeichnis	175

Johann Hahlen *)

Begrüßung

Meine sehr geehrten Damen und Herren, ich freue mich sehr, nunmehr die dritte gemeinsame wissenschaftliche Tagung mit dem *Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM)* und der *Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI)* bei uns eröffnen zu können. Es ist schön, dass Sie heute so zahlreich gekommen sind.

Sie haben sich ein außerordentlich bedeutsames Thema – „*Neue Erhebungsinstrumente und Methodeneffekte*“ – gestellt, das auch für die weitere Arbeit der Statistischen Ämter sehr bedeutsam ist. Was die Nutzung neuester Techniken im Erhebungsprozess betrifft, hat die deutsche amtliche Statistik Nachholbedarf. Ich bin davon überzeugt, dass wir auf diesem Feld sowohl von den Marktforschungsunternehmen als auch von den sozialwissenschaftlichen Instituten lernen können.

Eine Begründung für die Behutsamkeit, mit der die Statistischen Ämter an neue Erhebungsinstrumente herangehen, wird im Titel dieser Tagung benannt: Methodeneffekte.

Die *Datenqualität* hat für uns traditionell und notwendigerweise einen sehr hohen Stellenwert. Wir produzieren in erheblichem Umfang über die Zeit vergleichbare Informationen, etwa lange Reihen. Für unsere Kunden ist dies überaus wichtig. Jeder Instrumentenwechsel, jede Änderung im Erhebungsprozess kann aber zu Brüchen in einer Reihe führen.

Qualität ist aber – auch für die amtliche Statistik – mehr als Genauigkeit in den Nachkommastellen und als Vergleichbarkeit über die Zeit. *Aktualität* der Daten und *Kosten* ihrer Gewinnung dürfen nicht vernachlässigt werden.

Mehr als bei den Themen der vorausgegangenen gemeinsamen Tagungen – 1995: Pretest und Weiterentwicklung von Fragebogen, 1997: Interviewereinsatz und -qualifikation – stehen dieses Mal Fragen der Effizienzsteigerung durch Berücksichtigung neuer Erkenntnisse und neuester Technik im Mittelpunkt der Diskussion.

Wir in der amtlichen Statistik wollen kostengünstiger produzieren und gleichzeitig die Verzerrung der Ergebnisse, den systematischen Fehler (der nie gänzlich vermeidbar ist), so gering wie möglich halten. Vernünftig eingesetzt bieten moderne Erhebungsinstrumente enorme Chancen bei der Lösung dieses Optimierungsproblems. Dabei sind immer auch die Risiken dieser Technologien verantwortungsvoll einzuschätzen.

Die Wahl des jeweiligen Erhebungsinstruments kann nicht losgelöst von anderen Aspekten eines umfassenden Qualitätsmanagements getroffen werden. Im Statistischen Bundesamt widmen wir uns diesem Thema mit einer – wie wir es genannt haben – „*Qualitätsoffensive Statistik*“. Lassen sie mich im folgenden kurz die zentralen Aspekte dieser Offensive verdeutlichen, um zu zeigen, was Sie diesbezüglich in naher Zukunft von uns erwarten können und um zu zeigen, welche engen Bezüge diese Tagungsreihe zu Fragen des Qualitätsmanagements hat.

*) Johann Hahlen, Präsident des Statistischen Bundesamtes, Wiesbaden.

Ausgangspunkt für Diskussionen zum Qualitätsmanagement in der amtlichen Statistik ist neben der parteiübergreifend akzeptierten und mit Priorität versehenen politischen Zielsetzung der *Verwaltungsmodernisierung* vor allem die spezifische *Situation der amtlichen Statistik*: Die Statistischen Ämter müssen mit sinkendem Budget ihr Datenangebot (im wesentlichen aufgrund europäischer Rechtsnormen) ausweiten und eine steigende Nachfrage nach statistischen Informationen befriedigen. Gleichzeitig verändern sich – wie bereits angedeutet – die Ansprüche an die Qualität statistischer Daten.

Vor diesem Hintergrund umfasst die „Qualitätsoffensive Statistik“ eine Vielzahl von Maßnahmen. Sie ist mit einem grundlegenden Wandel des Selbstverständnisses der Verwaltung verbunden, nach dem der Bürger nicht mehr als Antragsteller, sondern als Kunde angesehen wird. Dabei geht es im wesentlichen um zwei Ziele: Erstens sollen die staatlichen Dienstleistungen besser werden (d. h.: wir werden effektiver) und zweitens soll sich der Aufwand für die Erstellung dieser Dienstleistungen vermindern (d. h.: wir werden effizienter). Im Zentrum des Handelns stehen die Qualität der Produkte, die Wirtschaftlichkeit der Leistungserbringung sowie die Zufriedenheit der Kunden und der Mitarbeiter.

Zur inhaltlichen Strukturierung und Koordinierung der vielen Einzelmaßnahmen ist ein übergreifendes Konzept erforderlich. Im Statistischen Bundesamt wird das *Referenzmodell der Europäischen Stiftung für Qualitätsmanagement* (EFQM) eingesetzt, das international anerkannt ist und bereits erfolgreich in der Industrie angewendet wird. Ich kann in diesem Rahmen nur stichwortartig einige der vor uns liegenden Arbeitspakete nennen:

- Zur Prozess- und Ergebnisoptimierung wird ein standardisiertes Verfahren zur Dokumentation und kontinuierlichen Verbesserung unserer Prozesse sowie der internen und externen Produkte entwickelt. Im Bereich der Statistikproduktion werden *Qualitätsberichte* eingeführt. Für die übrigen qualitätsrelevanten Bereiche werden *Selbstbewertungen* nach dem EFQM-Modell durchgeführt. Der Qualitätsbericht zur Produktqualität und die Selbstbewertung werden derzeit in drei Pilotgruppen auf ihre Praxistauglichkeit geprüft.
- Zur Integration des Vorschlagswesens in die Qualitätsmanagementsystem-Struktur haben wir erfolgreich die *Ideenbörse* eingeführt. In den letzten zwei Jahren haben unsere Mitarbeiter über 400 Vorschläge gemacht. Über ein Drittel dieser Vorschläge konnten umgesetzt werden.
- Eine *Kosten- und Leistungsrechnung* mit flächendeckenden Zeitanschreibungen haben wir 1998 eingeführt. Die Kosten- und Leistungsrechnung wird zügig mit dem Ziel weiterentwickelt, steuerungsrelevante Kennzahlen für ein Kosten- und Leistungscontrolling zu erhalten.
- Die Verbesserung der Aktualität ist seit langem ein wichtiges Anliegen unseres Hauses, das wir in enger Zusammenarbeit mit den Statistischen Ämtern der Länder angehen. Im Jahre 1997 wurde gemeinsam mit den Statistischen Ämtern der Länder ein *Frühwarnsystem* eingeführt, um Terminverzögerungen rechtzeitig offen zu legen und Abhilfemaßnahmen einzuleiten. Dieses Frühwarnsystem wird im Rahmen der Qualitätsoffensive erheblich erweitert und mit weiteren umfangreichen Möglichkeiten zur zeitnahen Auswertung des Qualitätsstandes versehen.

- Im organisatorischen Bereich wird die Arbeit durch die Einführung von *Projektmanagement* unterstützt. Neben dem verstärkten Angebot von Schulungen für Projektleiter wird hierzu ein hauseigener Projektmanagement-Leitfaden erstellt.

Die genannten „Arbeitspakete“ beschreiben die in der Anfangsphase unseres neuen Steuerungsmodells erforderlichen Instrumente. Sie werden schrittweise durch weitere Aktivitäten in anderen Bereichen ergänzt. So arbeiten wir derzeit an der Erstellung eines Personalentwicklungskonzeptes, planen die rasche Einführung einer Budgetierung von Drittmitteln und arbeiten am weiteren Ausbau von Kontraktmanagement. Alle Einzelmaßnahmen werden im Rahmen unseres integrierten Gesamtkonzeptes eng verzahnt. Wir wollen das Entstehen isolierter „Insellösungen“ verhindern.

Die Qualitätsoffensive wird im Statistischen Bundesamt vieles verändern und – da bin ich sehr optimistisch – zu einem besseren, schnelleren und vor allem noch kundenorientierteren Dienstleistungsangebot führen.

Ein derartiges Qualitätsmanagement wird nur gelingen, wenn wir, d. h. die Statistischen Ämter des Bundes und der Länder, nicht im eigenen Saft schmoren. Unsere Fachleute müssen sich regelmäßig mit externen Fachleuten austauschen. Diese Tagung bietet hierzu eine gute Gelegenheit. Nützlich sind dabei nicht zuletzt die informellen Kontakte am Rande der Veranstaltung. Gerade sie können dazu beitragen, Verständnis für die Rahmen- und Produktionsbedingungen der anderen Beteiligten zu schaffen, gegenseitiges Vertrauen zu festigen und das eine oder andere Vorurteil zu korrigieren. Daher lade ich alle Beteiligten direkt im Anschluss an die Vorträge des heutigen Veranstaltungstages hier im Foyer ein. Mein herzlicher Dank gilt dem ADM, der freundlicherweise die Bewirtungskosten übernommen hat.

Schließlich schon jetzt allen Referenten herzlichen Dank für ihre Beiträge. Ich wünsche Ihnen zwei Veranstaltungstage voll interessanter Vorträge, spannender Diskussionen und anregender Gespräche.

Jürgen Chlumsky*)

Einführung

Meine sehr geehrten Damen und Herren, ich darf mich zunächst bei den verantwortlichen Kollegen des *Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (ADM)* und der *Arbeitsgemeinschaft Sozialwissenschaftlicher Institute (ASI)* für das mir entgegengebrachte Vertrauen bedanken, das Grundlage für die Ehre ist, diese wissenschaftliche Tagung zum Thema „Neue Erhebungsinstrumente und Methodeneffekte“ moderieren zu dürfen.

Die privatwirtschaftlich organisierte Markt- und Sozialforschung, die Einrichtungen der akademisch verfassten Sozialforschung und die Statistischen Ämter des Bundes und der Länder sind *Teile der statistischen Infrastruktur* in unserer Republik, Teile eines Gesamtsystems, dessen Ziel es ist, Planungs- und Entscheidungsgrundlagen für die Gestaltung der Gesellschaft zur Verfügung zu stellen.

Es ist keineswegs selbstverständlich, wie unverkrampft die drei Institutionen, die diese Veranstaltung repräsentieren, miteinander umzugehen gelernt haben. Die jährlich jeweils im November tagende *gemeinsame Arbeitsgruppe* und die von dieser Arbeitsgruppe initiierten gemeinsamen Vorhaben – diese Tagung ist nur eines der Projekte – haben daran erheblichen Anteil. Darauf dürfen wir gemeinsam stolz sein.

Zu Beginn der dritten gemeinsamen Tagung ist es erlaubt, von einer *Veranstaltungsreihe* zu sprechen. Ob es 1995 um den Pretest und die Weiterentwicklung von Fragebogen ging, 1997 um den Interviewereinsatz und die Interviewerqualifikation oder heute und morgen um neue Erhebungstechniken und Methodeneffekte – in dieser Veranstaltungsreihe dreht es sich jeweils um eine für uns alle gleichermaßen grundlegende und zentrale Frage: „*Wie können wir die Qualität unserer Erhebungen/unserer Umfragen weiter verbessern?*“ Und für uns alle geht es – wenn wir von Qualität sprechen – sowohl um Kriterien wie die Relevanz der Fragestellung, die Validität und Genauigkeit der Ergebnisse als auch um eine schnelle und kostengünstige Datenbereitstellung. Unsere Auftraggeber formulieren da – so vermute ich – ganz ähnliche Anforderungen.

Ob und inwieweit der Einsatz neuer Erhebungsinstrumente dazu beitragen kann, diese Anforderungen zu erfüllen, werden wir heute und morgen diskutieren. Gestatten Sie mir, Thema und Programm unserer Tagung an dieser Stelle ausschließlich aus dem Blickwinkel der amtlichen Statistik zu beleuchten. Aus drei Gründen:

- Erstens, weil ich annehme, dass dies aufschlussreich sein könnte für diejenigen von Ihnen, die über uns relativ wenig wissen,
- zweitens, um nichts von dem vorwegzunehmen, was Sie uns, Herr Wiegand, gleich im Anschluss über die Chancen und Risiken neuer Erhebungstechniken erzählen werden
- und drittens, weil mir diese Sicht der Dinge naturgemäß am vertrautesten ist.

*) Jürgen Chlumsky, Statistisches Bundesamt, Wiesbaden.

Ich hole ein wenig aus:

Meine Damen und Herren, die amtliche Statistik weiß, dass sie sich den Forderungen nach einer modernen aktivierenden Verwaltung stellen muss, d. h. weg muss vom Image der bürokratischen Buchhaltung. Dies wird ihr aber nur gelingen, wenn sie sich weit stärker als in der Vergangenheit an den Wünschen ihrer Kunden, den Datennutzern, orientiert. Für die Kolleginnen und Kollegen aus den Marktforschungsunternehmen mag es ganz unspektakulär klingen, wenn ich sage: Es reicht nicht aus, Daten zu erheben, aufzubereiten und zu veröffentlichen – gefragt ist eine umfassende Dienstleistung, zugeschnitten auf den ganz spezifischen Bedarf des jeweiligen Kunden.

Für die amtliche Statistik heißt dies, sie muss weg von der *Input-*, hin zur *Outputsteuerung*. Ich möchte Ihnen kurz diesen aus meiner Sicht für die amtliche Statistik notwendigen *Paradigmenwechsel* erläutern:

Im gegenwärtigen Verständnis der Aufgabe der amtlichen Statistik spielt das *Legalitätsprinzip* eine wesentliche Rolle. Und dies ist grundsätzlich auch voll in Ordnung. Wir leben von Steuergeldern. Sorge bereitet uns die *Art und Weise*, wie das Legalitätsprinzip *gehandhabt* wird und vor allem die *Schlussfolgerungen*, die dann z. T. gezogen werden.

Per Gesetz ist genau definiert, was wir wann zu erheben haben, detailliert, z. T. bis zu den Merkmalsausprägungen. Die Gesetze fixieren mithin den Input der Statistik (den Fragebogen, den Auskunftspflichtigen). In Zeiten knapper öffentlicher Kassen scheint es nun – zumindest auf den ersten Blick und bei sehr oberflächlicher Betrachtung – nahe zu liegen, das Aufgabenprogramm der amtlichen Statistik – auch hinsichtlich des Outputs – aus diesen Gesetzen (und nur aus diesen) abzuleiten. Alle Weiterverarbeitungen, Auswertungen, Analysen wären quasi Luxus, also Einsparpotential. Für Sie mag diese Vorstellung unwirtlich klingen, die amtliche Statistik war in den vergangenen Jahren und ist bis heute genau dieser Diskussion ausgesetzt.

Wir müssen rüberbringen, dass es in einer dynamischen Gesellschaft faktisch unmöglich ist, das, was an Informationen, an statistischem Output gebraucht wird, per Gesetz im Detail zu steuern. Dies führt zu einer völligen Erstarrung des Programms und gefährdet auf Dauer – da wenig attraktiv – die Zukunft der Statistik.

Was ist nun die Alternative? Wie eben angedeutet, die Outputsteuerung, das Denken vom Produkt aus:

- Welche Tabellen, Indikatoren, Mikrodatenfiles etc. stellt die amtliche Statistik mit welcher Qualität bereit?
- Welche Inputdaten sind dazu erforderlich? Welche davon per Erhebung zu beschaffen? Wo kann auf andere Quellen zurückgegriffen werden, wo sind Schätzungen, Modelle usw. ausreichend?

Für eine zeitgemäße Aufgabenerledigung reicht es aber natürlich nicht aus, wenn die Statistischen Ämter sich stärker als bisher gegenüber ihren Kunden öffnen, den Bedarf „beim Nutzer abholen“. Erforderlich ist ebenso eine ständige *Verbesserung des Produktionsprozesses*, was eine Intensivierung des Einsatzes von Informationstechnik und methodische Optimierungen der Statistikkonzepte notwendig macht.

Und da ist zunächst – wir sind hier mittendrin im Thema unserer Tagung – an den *Einsatz neuer Erhebungstechniken* zu denken.

Die Entwicklung der Umfrageforschung außerhalb der amtlichen Statistik ist seit einigen Jahren durch eine erhebliche Zunahme an telefonischen Befragungen gekennzeichnet. *Thomas Petersen* aus Allensbach und *Wolfgang Schulte* aus Mannheim werden uns heute Nachmittag über die *Folgen eines Wechsels von Face-to-Face- zu telefonischen Befragungen* berichten. Wir werden erfahren, wie sie das Risiko systematischer Verzerrungen, das mit jedem Instrumentenwechsel einhergeht, bewerten.

Bei mündlichen Befragungen lösen Laptop und Pentop immer mehr Papier und Bleistift ab. *Rolf Pfeiderer* von Infratest Burke wird uns nach der Kaffeepause über *Methodeneffekte beim Umstieg auf CAPI-Technik* aufklären.

Die *Vorteile computergestützter Umfrageforschung* sind beeindruckend: Sowohl bei CAPI als auch bei CATI erübrigt sich eine zusätzliche manuelle Erfassung, Signier- und Filterfehler können weitgehend vermieden werden, Daten werden bereits beim Befragten auf Plausibilität überprüft, so dass Fehler nicht wie bisher in mühevoller und zeitaufwendiger Arbeit in der Erhebungszentrale berichtigt werden müssen. Alle einschlägigen Untersuchungen zeigen eine hohe Akzeptanz bei den Befragten. Allerdings ist Akzeptanz zwar ein notwendiges, jedoch keineswegs ein hinreichendes Kriterium zur Bewertung des Computers als Hilfsinstrument. Daher wird *Marek Fuchs* von der Katholischen Universität Eichstätt für uns zum Abschluss des heutigen fachlichen Programms die *Interviewsituation bei computergestützten Befragungen* beleuchten.

Da jeder Interviewer ein eigenes Erfassungsgerät benötigt, ist die Einführung von CAPI selbst bei weiterhin stark fallenden Computerpreisen mit vergleichsweise hohen Anfangsinvestitionen verbunden. Bei der Größe der Stichproben, wie sie in der amtlichen Statistik gefahren werden, ein echtes Problem. Hier wäre aus meiner Sicht zu prüfen, ob in diesen Fällen die Feldarbeit nicht privaten Erhebungsinstituten übertragen werden könnte. Andernfalls muss ein Weg gefunden werden, die Feldarbeit innerhalb des Systems der amtlichen Statistik noch stärker zu professionalisieren.

Ein weiterer Innovationsschub könnte später einmal von der *Online-Forschung* ausgehen. Inwieweit sich das Internet als Befragungsmedium langfristig durchsetzen wird, ist zwar noch völlig offen. Die Perspektive aber ist faszinierend: Rückgriff auf ein bestehendes Netz und – anders als bei der telefonischen Befragung – geradezu unbegrenzte Visualisierungsmöglichkeiten. *Wolfgang Bandilla* und *Michael Bosnjak*, beide von ZUMA, werden morgen zum Abschluss der Veranstaltung ein wenig in die Zukunft schauen. Ein vielversprechendes Einsatzfeld sind möglicherweise Panel-Studien. Hier könnte auch das Interesse an Internet-Befragungen für die amtliche Statistik liegen und zwar sowohl bei Personen- und Haushaltsbefragungen als auch im Unternehmenssektor.

Mit Spannung erwarten wir zu Beginn des morgigen Tages den Bericht aus Nürnberg von *Heinrich Litzneroth*. Sein Thema ist die *Testmarktforschung*. Für die amtliche Statistik wäre zu prüfen, ob die hier eingesetzten Technologien nicht durchaus in modifizierter Form Anwendung etwa bei der Erhebung von Preisen finden könnten.

Die immer noch zunehmenden Erhebungswiderstände – eine Rahmenbedingung des Befragungsgeschäfts, mit der wir alle zu kämpfen habenwar für die Planer dieser Tagung Be-

gründung genug, *Michael Hoppe* von IPSOS aus Hamburg zu bitten, über *Aufbau und Organisation von Access-Panels* zu sprechen. Die amtliche Statistik hat mit diesem Instrument keine Erfahrung – und ob es für ihre Anwendungsfälle überhaupt in Betracht käme, wäre zu prüfen.

Für die Markt- und Sozialforschung vielleicht weniger relevant, zur Entlastung der Unternehmen von Statistik und gleichzeitig zur Rationalisierung der Arbeit in den Statistischen Ämtern jedoch ein zukunftsfähiges neues Erhebungsinstrument, ist der *elektronische Datenaustausch* (EDI). Daten, die bereits in elektronischer Form in den Informationssystemen der Unternehmen vorliegen und von Softwarepaketen verwaltet werden, die auf Buchhaltung, Lohnkonten, Lagerverwaltung u.ä. spezialisiert sind, sollten medienbruchfrei für die Statistik verfügbar sein.

Der morgen früh vor der Kaffeepause geplante Beitrag meines Kollegen *Bernd Schmidt* befasst sich mit einer weiteren Möglichkeit der Datenbeschaffung, die wir für die amtliche Statistik als äußerst zukunftsträchtig ansehen, weil sie belastungsarm und vergleichsweise kostengünstig ist – die *registergestützte Datengewinnung*. Es geht in dem Fall, der hier vorgestellt wird, um erhebliche zusätzliche Datenanforderungen der Europäischen Union (EU) im Bereich der Unternehmensstatistik, die ohne zusätzliche Primärerhebungen erfüllt werden sollen.

Die Statistik der natürlichen Bevölkerungsbewegung und die Wanderungsstatistik basieren auf Unterlagen der Standesämter und der Einwohnermeldeämter, die Arbeitsmarktstatistik auf Registerdaten der Bundesanstalt für Arbeit. Dies sind Beispiele für die schon gegenwärtig breite Nutzung von Verwaltungsunterlagen/-dateien für Zwecke der Statistik und den Verzicht auf aufwendige Primärerhebungen in hierfür geeigneten Fällen. Die Fortschritte beim Einsatz von Informationstechnik in den Behörden lassen vermuten, dass die Möglichkeiten, Daten aus dem Verwaltungsvollzug zu nutzen, noch lange nicht ausgeschöpft sind.

Auch bei einer erheblichen Ausweitung der registergestützten Datengewinnung wird die amtliche Statistik selbstverständlich weiterhin auf die Durchführung von Erhebungen angewiesen bleiben. Eine – neben der Nutzung neuer Erhebungsinstrumente – geeignete Maßnahme zur Optimierung des Produktionsprozesses besteht dann in der *Optimierung des Erhebungsumfangs*. Der Verzicht auf Vollerhebungen und die Reduzierung von Stichprobenumfängen entlasten die Befragten, sparen Kosten und ermöglichen darüber hinaus eine schnellere Bereitstellung der Ergebnisse.

Um die finanziellen und politischen Kosten der Datenbeschaffung zu senken, sollte die amtliche Statistik ferner ihr Augenmerk in Zukunft noch stärker als bisher auf eine *Effizienzsteigerung bei der Datenverwertung* lenken. Effiziente Bewirtschaftung der verfügbaren Daten heißt volle Ausschöpfung des Informationsgehalts von vorhandenem Zahlenmaterial.

Eine *Zusammenführung von Daten aus verschiedenen Erhebungen* kann nicht nur als Auswertungstechnik für den Datennutzer nützlich sein, sondern ebenso den Statistischen Ämtern als Erhebungssurrogat dienen. Damit tritt dann aber verstärkt die Forderung der Konsistenz von Variablen zum Zwecke ihrer Zusammenschau in einem Kreislaufschema neben die – derzeit im Vordergrund stehende – Forderung problemadäquater Begriffsdefinition im Hinblick auf den speziellen Zweck einer Einzelstatistik.

Ich wünsche der dritten gemeinsamen wissenschaftlichen Tagung von ADM, ASI und Statistischem Bundesamt einen erfolgreichen Verlauf und uns allen lehrreiche und dennoch kurzweilige Stunden.

Chancen und Risiken neuer Erhebungstechniken in der Umfrageforschung

1 Einleitung

Ein Referat über „Chancen und Risiken neuer Erhebungstechniken in der Umfrageforschung“ ist selbst eine Chance und ein Risiko. Es ist eine Chance, weil das Thema so breit angelegt ist, dass die verschiedenen Erhebungstechniken innerhalb der Rahmenbedingungen diskutiert werden können, die Umfrageforschung ermöglichen und begrenzen. Mit Rahmenbedingungen meine ich sowohl den Stand der wissenschaftlichen Erkenntnis und der forschungsrelevanten Technik als auch die für die Umfrageforschung relevanten gesetzlichen Regelungen.

Das Thema des Referats ist ein Risiko, weil es nicht allzu schwer ist, sich in seiner Breite zu verlieren. Deshalb ist zunächst eine Definition des Begriffs „neue Erhebungstechniken“ unumgänglich. Es wäre nicht angemessen, den Begriff „Erhebungstechnik“ eng zu definieren und allein auf den Einsatz technischer Hilfsmittel bei der Datenerhebung zu konzentrieren. Der Begriff muss sich vielmehr auf den Vorgang der Datenerhebung insgesamt beziehen. Die Verwendung der forschungsrelevanten Technik ist damit natürlich eingeschlossen. Insofern ist es vielleicht angemessener, von Erhebungsinstrumenten statt von Erhebungstechniken zu sprechen. Und der Begriff „neu“ kann sich nicht nur auf die neuesten Entwicklungen der letzten Jahre beziehen, wie z. B. die Durchführung von Umfragen mittels Internet. Er muss einen längeren Zeitraum und damit auch Entwicklungen einschließen, die wir vielleicht schon gar nicht mehr als neu empfinden, wie z. B. computerunterstützte Umfragen per Telefon.

Ich werde im folgenden fünf allgemeine Trends in der Markt- und Sozialforschung identifizieren, die – zum Teil direkt, zum Teil aber auch nur indirekt – mit der Entwicklung und dem Einsatz neuer Erhebungsinstrumente in Beziehung stehen und die Qualität der Forschungsergebnisse beeinflussen. Dabei handelt es sich im einzelnen um:

1. die Pluralisierung der Formen des Interviews;
2. die Technisierung und Computerisierung der Datenerhebung;
3. die Standardisierung der Forschungsinstrumente;
4. die Verstetigung des Forschungsprozesses und
5. die Internationalisierung der Forschung.

2 Erster Trend: Pluralisierung der Formen des Interviews

Noch vor nicht allzu langer Zeit wurden Umfragen fast ausschließlich mittels persönlich-mündlicher Interviews durchgeführt. Das hat sich im Verlauf der letzten zehn Jahre grundlegend geändert. Das Telefoninterview ist gleichberechtigt neben das persönlich-mündliche Interview getreten. Andere Formen des Interviews mittels interaktiver Medien schicken sich gerade an, langfristig das gleiche zu tun. Und das fast schon totgesagte schriftliche Interview feiert gegenwärtig eine Art Wiederauferstehung. Letzteres ist in erster Linie darauf zurückzu-

*) Erich Wiegand, Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM), Frankfurt am Main.

führen, dass in den letzten Jahren Untersuchungen zur Kundenzufriedenheit stark zugenommen haben. Das ist aber ein Bereich der Marktforschung, bei dem häufiger als in anderen Bereichen die Daten schriftlich erhoben werden.

Im Jahr 1998 haben die Mitgliedsinstitute des ADM zusammen rund zehneinhalb Millionen quantitative Interviews durchgeführt. Mit 41 % lag der Anteil telefonischer Interviews erstmals über dem der Face-to-Face-Interviews, die 39 % aller durchgeführten Interviews ausmachten. Der Anteil schriftlicher Interviews betrug 19 % und hat sich damit innerhalb von zwei Jahren fast verdoppelt. Der Anteil der Online-Interviews lag zwar 1998 nur knapp unter 1 %, ein Jahr zuvor war er allerdings noch kaum messbar.

In der Umfrageforschung vollzieht sich also gegenwärtig eine Pluralisierung der Interviewformen, die noch lange nicht abgeschlossen sein dürfte. Dabei wurde die frühere Dominanz des Face-to-Face-Interviews abgelöst durch das gleichberechtigte Nebeneinander verschiedener Interviewformen.

Die Auffassung, persönlich-mündliche Umfragen würden sich auch in Deutschland im Verlauf der Zeit zu einem quantitativ unbedeutenden „Nischenprodukt“ zurückentwickeln, vermag ich allerdings (noch) nicht zu teilen. Erstens wurde vor einigen Jahren, als telefonische Umfragen sich im stürmischen Aufwind ihrer Anfangsentwicklung befanden, schon einmal dasselbe behauptet. Eingetreten ist diese Entwicklung aber bisher noch nicht.

Und zweitens tut man sich mit solchen Langzeitprognosen ohnehin sehr schwer. Sollte beispielsweise im ständigen Spannungsfeld der Güterabwägung zwischen Freiheit der Forschung und Schutz der Privatsphäre die Rechtsprechung undifferenziert zugunsten der Privatsphäre entscheiden, dann träfe das telefonische Umfragen viel dramatischer als Face-to-Face-Umfragen. In dieser Situation wäre also eher mit einer Zunahme persönlich-mündlicher Interviews als mit deren allmählicher Bedeutungslosigkeit zu rechnen.

3 Zweiter Trend: Computerisierung und Technisierung der Datenerhebung

Parallel zur Pluralisierung der Interviewformen fand und findet eine zunehmende Technisierung und vor allem Computerisierung der Datenerhebung statt. Das ist der zweite feststellbare Trend in der Umfrageforschung. Die Computerisierung des Interviews zeigt sich natürlich an der zunehmenden Bedeutung des Internets für die Datenerhebung. Sie zeigt sich aber auch an den Veränderungen im Bereich der anderen Formen der Datenerhebung, wie z. B. bei telefonischen Interviews.

In den Anfängen des telefonischen Interviews hatte der Interviewer die Angaben der Befragten in den ihm in Papierform vorliegenden Fragebogen eingetragen. Häufig geschah dies sogar noch dezentral, d. h. von der Wohnung des Interviewers aus. Überspitzt formuliert haben sich bei dieser Vorgehensweise die spezifischen Nachteile persönlich-mündlicher und telefonischer Interviews addiert. Allerdings gehört diese Vorgehensweise größtenteils der Vergangenheit an. Die überwiegende Mehrzahl der telefonischen Interviews werden heute von zentralen Telefonstudios aus computerunterstützt durchgeführt. Allein in den Mitgliedsinstituten des ADM standen Ende 1998 dafür 2 738 sogenannte CATI-Plätze (Computer assisted telephone interviewing) zur Verfügung.

Abgesehen davon, dass die Stichprobe bei telefonischen Befragungen zunehmend durch das zufällige Generieren von Telefonnummern gebildet wird, beginnt die Computerunterstützung telefonischer Interviews heute bereits mit dem automatisierten Anwählen der für eine Untersuchung ausgewählten Telefonnummern. Wenn kein Kontakt zustande kommt, wird die Nummer später erneut automatisch angewählt. Wird ein Kontakt hergestellt, erscheint der „Fragebogen“ auf einem Monitor; der Interviewer wird in Abhängigkeit von den Antworten des Befragten durch eine entsprechende Programmierung automatisch durch das Interview geführt.

Die Vorteile computerunterstützter telefonischer Interviews liegen vor allem in den umfassenden Steuerungs- und Kontrollmöglichkeiten der Erhebungssituation. Durch die Möglichkeit der Aufzeichnung des Interviews erweitert der Einsatz moderner Technik darüber hinaus aber auch die Möglichkeiten der Auswertung. So ist man bei offenen Fragen nicht mehr auf die Mitschriften der Interviewer angewiesen, sondern kann auf die echten Verbatims, d. h. wortgetreuen Antworten, der Befragten zurückgreifen.

Unbestreitbar haben telefonische Interviews aber auch Nachteile: Listensätze, Kärtchenspiele und andere Befragungsmaterialien können dem Befragten einfach nicht vorgelegt werden. Dies macht sie für bestimmte Forschungsgegenstände wenig geeignet.

Neben den telefonischen Interviews – bei denen sich die Computerunterstützung bereits weitgehend durchgesetzt hat – wird sich auch die Durchführung persönlich-mündlicher Interviews computerisieren. Der Laptop beginnt zunehmend als „elektronischer Fragebogen“ den traditionellen Fragebogen in Papierform zu ersetzen. Im Jahr 1998 wurde bereits jedes siebte Face-to-Face-Interview mit Hilfe eines Laptops oder Pentops durchgeführt. Ende 1998 standen dafür in den Mitgliedsinstituten des ADM 4 512 sogenannte CAPI-Geräte (Computer assisted personal interviewing) zur Verfügung.

Auch die Vorteile der Computerunterstützung persönlich-mündlicher Interviews sind vor allem in der besseren Steuerung und Kontrolle der Erhebungssituation zu sehen. Der Interviewer kann – bewusst oder unbewusst – keine Fragen überspringen; er kann keine Filterfehler mehr machen; und die einzelnen Fragen müssen entsprechend der jeweiligen Antwortvorgaben eindeutig beantwortet werden. Listensätze und Kärtchenspiele erscheinen auf dem Monitor und können deshalb vom Interviewer nicht mehr „vergessen“ werden. Außerdem garantiert die Computerunterstützung, dass sie den Befragten tatsächlich auf die vorgesehene Art und Weise vorgelegt werden; und nicht wie vom Interviewer präferiert.

Da die Interviewer den „Fragebogen“ vom Institut per Modem auf ihren Laptop aufgespielt bekommen und die erhobenen Daten ebenfalls per Modem täglich an das Institut übertragen, können Fragebogen nicht mehr auf dem Postweg verloren gehen, wie das in der Papier- und Bleistift-Epoche des Face-to-Face-Interviews – aus welchen Gründen auch immer – häufiger der Fall gewesen sein soll. Dadurch wird es den Instituten möglich, Probleme in der Feldarbeit frühzeitig zu erkennen und durch eine entsprechende Feldsteuerung noch rechtzeitig zu beheben.

Sowohl bei telefonischen als auch bei persönlich-mündlichen Interviews bietet die Computerunterstützung in bestimmten Grenzen die Möglichkeit, das Antwortverhalten bereits während des Interviews auf Konsistenz zu überprüfen. Das nachträgliche Editing der erhobenen Daten wird also in dem Umfang entbehrlich, in dem das Editing in den Prozess der Datener-

hebung integriert werden kann. Diese Möglichkeit kann allerdings kein Grund dafür sein, auf ein nachträgliches Editing zu verzichten. Das wäre zweifellos mit einem wissenschaftlichen Qualitätsverlust der Forschungsergebnisse verbunden.

Schließlich ist auch eine Technisierung und Computerisierung schriftlicher Umfragen möglich. Denn was sind per Telefax durchgeführte Umfragen anderes als schriftliche Umfragen, bei denen der Postversand durch einen moderneren Kommunikationskanal ersetzt wird? Im Gegensatz zur Computerunterstützung persönlich-mündlicher und telefonischer Interviews sehe ich bei dieser Form der Technisierung schriftlicher Interviews aber überwiegend Nachteile. Wie Sie wissen, sind in Deutschland die Möglichkeiten des Telemarketing durch eine gefestigte Rechtsprechung sehr stark eingeschränkt. Telefonische Umfragen werden von diesen Verboten nicht erfasst. Ich bin aber sicher, dass dies im Fall einer zunehmenden Bedeutung bei Telefax-Umfragen nicht der Fall sein wird.

Abgesehen davon werden Telefax-Umfragen auf absehbare Zeit wegen der geringen Verbreitung der dazu notwendigen Faxgeräte in privaten Haushalten weitgehend auf den Bereich des Business-to-Business beschränkt bleiben müssen. Auch die Einsatzmöglichkeiten der sogenannten Disk-by-Mail-Umfragen sind wegen des dafür notwendigen PCs auf bestimmte Zielgruppen begrenzt. Bei dieser Erhebungstechnik, die man als Computerisierung schriftlicher Umfragen bezeichnen kann, wird den für die Befragung ausgewählten Personen oder Unternehmen per Post der auf einer Computerdiskette gespeicherte Fragebogen zugeschickt. Die Befragten füllen diesen „Fragebogen“ mit Hilfe ihres PCs aus und schicken die Diskette dann per Post an das Institut zurück.

Anders als die Technisierung schriftlicher Umfragen durch das Telefaxgerät weist die Computerunterstützung schriftlicher Interviews durch Disk-by-Mail keine rechtlichen Probleme auf und hat – ähnlich wie die Computerunterstützung persönlich-mündlicher und telefonischer Interviews – den Vorteil einer gegenüber herkömmlichen schriftlichen Umfragen verbesserten Steuerung der Datenerhebung. Natürlich kann weder bei Telefax- noch bei Disk-by-Mail-Umfragen sichergestellt werden, dass der Fragebogen auch tatsächlich von der jeweiligen Zielperson ausgefüllt wird. Und natürlich dürfen die eingeschränkten Einsatzmöglichkeiten beider Formen der Datenerhebung nicht verkannt werden. Kurz- und mittelfristig werden sowohl Telefax- als auch Disk-by-Mail-Umfragen nur in bestimmten Forschungsbereichen bzw. bei bestimmten Zielgruppen einsetzbar sein. Langfristig werden sich ihre Einsatzmöglichkeiten aber durch die zunehmende Ausstattung privater Haushalte mit den dafür notwendigen technischen Geräten ausweiten. Allerdings dürften sie dann bereits durch andere Formen der Datenerhebung abgelöst worden sein.

Die Grenzen zwischen der Computerisierung herkömmlicher Erhebungstechniken und dem neuen Instrument der Online-Forschung sind nicht immer leicht zu ziehen, weil die Übergänge fließend sind. Beispielsweise ist es denkbar, dass im Rahmen einer Disk-by-Mail-Umfrage ein Teil der Befragten den ausgefüllten Fragebogen nicht mehr auf Diskette per Post zurückschickt, sondern per E-Mail.

Bevor ich nun auf die Online-Forschung zu sprechen komme, ist eine Präzisierung des Begriffs notwendig: Online-Forschung kann zunächst Forschung über das Internet selbst bedeuten; über die Art und Häufigkeit seiner Nutzung sowie die Struktur und Interessen seiner Nutzer. Diese Form der Online-Forschung kann gegenwärtig erst zum Teil mit Hilfe des

Internets durchgeführt werden. Sie muss zum Teil noch auf die herkömmlichen Erhebungstechniken des persönlich-mündlichen, telefonischen oder schriftlichen Interviews zurückgreifen.

Die andere Bedeutung des Begriffs Online-Forschung meint Forschung mittels Internet, d. h. Forschung durch den Einsatz des Internets als Instrument der Datenerhebung. Nur in dieser Bedeutung werde ich im folgenden den Begriff Online-Forschung verwenden. Forschung mittels Internet setzt Forschung über das Internet voraus. Nur wenn man über ausreichende Informationen bezüglich Nutzer und Nutzung des Internets verfügt, können die methodischen Probleme dieses Erhebungsinstruments erkannt und die entsprechenden Fehler vermieden werden.

Die methodischen Probleme und Fehlerquellen der Online-Forschung konzentrieren sich auf die Frage der Repräsentativität. Es ist meines Erachtens eine Fehleinschätzung, wenn man glaubt, bevölkerungsrepräsentative Umfragen gegenwärtig schon mittels Internet durchführen zu können. Diese Einschätzung ist genauso falsch wie unmittelbar nach der Wiedervereinigung die Vorstellung, in den neuen Bundesländern telefonische Umfragen durchführen zu können. Allerdings weist diese Analogie auch darauf hin, dass es mittel- bzw. langfristig möglich sein wird, repräsentative Bevölkerungsumfragen im Internet durchführen zu können.

Gegenwärtig ist eine repräsentative Online-Forschung aber wegen der noch immer geringen und unterschiedlich hohen Verbreitung des Internets nur für bestimmte Gruppen der Bevölkerung möglich. Und selbst dann wird man wegen der unterschiedlichen Nutzung des Internets die Rekrutierung der zu befragenden Personen in den wenigsten Fällen im Internet selbst vornehmen können; sondern wird dazu häufig auf herkömmliche Techniken der Umfrageforschung – wie beispielsweise das Screening im Rahmen von Omnibus-Befragungen – zurückgreifen müssen.

Aber auch schon jetzt sind repräsentative Umfragen im Internet, die die Struktur der Besucher einer bestimmten Website korrekt abbilden, möglich. Sie sind sinnvoll, weil sie dem entsprechenden Diensteanbieter im Vergleich zu einer Befragung, die auf einer forcierten bzw. selbstselektiven Auswahl der Besucher seiner Website basiert, einen „added value“ bieten. Voraussetzung für die Bildung einer repräsentativen Stichprobe ist die Installation eines Mechanismus, der beispielsweise jeden n-ten Besucher der Website mit dem Fragebogen konfrontiert und ihm folgende Alternativen bietet:

Er kann den Fragebogen erstens gleich beantworten oder ihn sich zweitens zur späteren Beantwortung herunterladen. Er kann drittens seine E-Mail-Adresse angeben, damit ihm der Fragebogen auf diese Weise zugeschickt werden kann. Schließlich muss er viertens seine Teilnahme auch verweigern können und trotzdem die Möglichkeit haben, auf die gesuchte Website zu gelangen. Nicht zu vereinbaren mit den Anforderungen an eine repräsentative Stichprobe ist dagegen das „erzwungene“ Ausfüllen des Fragebogens, weil sonst ein Besuch der Website nicht möglich ist, und die Selbstselektion der Befragten, bei der die Teilnahme an der Befragung das aktive Anklicken eines entsprechenden Buttons erfordert.

Vereinfacht formuliert ist also Umfrageforschung mittels Internet gegenwärtig nur als Forschung über das Internet möglich – also über die Art der Nutzung bestimmter Online-Dienste und die Struktur ihrer Nutzer.

Zweifellos ist im Hinblick auf das Internet und seine Nutzung ein gesetzlicher Regelungsbedarf vorhanden, aber auch ein entsprechendes Regelungsbedürfnis der zuständigen Stellen. Inwieweit die Online-Forschung von solchen Regelungen langfristig tangiert wird, ist nicht zuletzt davon abhängig, wie es der Markt- und Sozialforschung gelingt, ihre ethischen Grundprinzipien – Freiwilligkeit der Teilnahme, Wahrung der Anonymität und Trennung von forschungsfremden Tätigkeiten – an diese neue Erhebungstechnik anzupassen und durchzusetzen.

Die zunehmende Technisierung und Computerisierung der Datenerhebung ist aber nicht nur auf das Interview beschränkt, sondern erfasst auch andere Formen der Datenerhebung. Beispielsweise gibt es kaum noch Gruppendiskussionen, die nicht per Video aufgezeichnet werden. Vom standesrechtlichen Standpunkt ist gegen die Aufzeichnung nichts einzuwenden, sofern die Teilnehmer vorab darüber informiert werden und damit einverstanden sind. Vom methodischen Standpunkt überwiegen die mit der Aufzeichnung verbundenen Vorteile erweiterter Auswertungsmöglichkeiten die möglichen Nachteile einer dadurch verursachten Verhaltensbeeinflussung der Teilnehmer.

Die Beobachtung von Gruppendiskussionen durch den Auftraggeber – hinter dem berühmten Einwegspiegel – ist ein durchaus übliches Verfahren und standesrechtlich nicht zu beanstanden, wenn die Teilnehmer damit einverstanden sind und man nicht annehmen muss, dass die Teilnehmer dem Auftraggeber aufgrund irgendwelcher Umstände – beispielsweise bestehende Geschäftsbeziehungen – bekannt sein könnten. Die moderne Videotechnik macht allerdings die physische Anwesenheit des Auftraggebers weitgehend überflüssig. Inzwischen verfügen die meisten Studios über die notwendige Technik, um die Aufzeichnung in Echtzeit dem Auftraggeber zu übermitteln. Umgekehrt können Gruppendiskussionen auch als Videokonferenz durchgeführt und damit zeit- und kostensparend Teilnehmer rund um den Globus zusammengebracht werden.

Ob allerdings Gruppendiskussionen per Videokonferenz als Instrument der Datenerhebung die herkömmlichen Gruppendiskussionen mit persönlicher Anwesenheit in allen Fällen vollständig ersetzen können, muss hier dahingestellt bleiben. Sie stellen aber zumindest eine Ergänzung des methodischen Instrumentariums für Gruppendiskussionen dar. Das gilt stärker noch für Gruppendiskussionen, die im Internet in eigens dafür eingerichteten Chat-Rooms durchgeführt werden.

Die Fortschritte in der forschungsrelevanten Technik haben aber die einzelnen Instrumente der Datenerhebung nicht nur verändert; sie haben darüber hinaus die Entwicklung und Anwendung verschiedener Erhebungsinstrumente überhaupt erst möglich gemacht. Ich denke hier beispielsweise an das TV-Meter im Rahmen der Fernsehnutzungs-Forschung oder an die Methode des adaptiven Conjoint Measurements, die ohne Computereinsatz bereits in der Phase der Datenerhebung nicht möglich ist.

4 Dritter Trend: Standardisierung der Forschungsinstrumente

Der dritte Trend in der Umfrageforschung ist die Standardisierung der Forschungsinstrumente. Das beginnt damit, dass bewährte Formulierungen und Abfolgen von Fragen aus einer Untersuchung in späteren Studien übernommen werden. Zweifellos wird dadurch die Vergleichbarkeit der Ergebnisse verschiedener Umfragen entscheidend verbessert. Allerdings ist

die Übernahme bewährter Frageformulierungen und Frageabfolgen nur der erste – und eigentlich selbstverständliche – Schritt auf dem Weg zur Standardisierung der Forschungsinstrumente.

Ein weiterer Schritt in dieser Richtung sind alle Bemühungen, die darauf abzielen, die Abfrage ganzer Themenblöcke zu vereinheitlichen. Solche Ansätze finden sich vor allem bei der Erfassung der soziodemographischen Variablen. Ein Beispiel hierfür sind die „Demografischen Standards“, die als gemeinsame Empfehlung von ADM, ASI und Statistischem Bundesamt entwickelt wurden. Es würde die Möglichkeiten kumulativer Forschung erheblich verbessern, wenn es gelänge, dieses standardisierte Erhebungsinstrument in der Umfrageforschung allgemein durchzusetzen.

Aber auch die Vereinheitlichung der Abfrage bestimmter Themenblöcke ist nur ein Zwischenschritt auf dem Weg zur Standardisierung der Forschungsinstrumente. In der Marktforschung ist seit einigen Jahren eine weit umfassendere Form der Standardisierung festzustellen. Ausgangspunkt dieser Entwicklung ist die Erkenntnis, dass sich die Probleme der Auftraggeber ähneln – zumindest die Probleme der Auftraggeber aus dem Bereich der privaten Wirtschaft. Häufig lassen sich die konkreten Probleme, die schließlich zur Erteilung eines Forschungsauftrags führen, übergeordneten Problembereichen zuordnen, wie beispielsweise: Markenpositionierung, Kundenbindung oder Werbewirkungskontrolle.

Es war für die Institute also naheliegend, für die verschiedenen Problembereiche Forschungsinstrumente zu entwickeln, bei denen nicht nur die Erhebung der Daten, sondern auch ihre Auswertung und die Präsentation der Ergebnisse weitgehend standardisiert erfolgt. Dabei muss die Betonung auf „weitgehend“ liegen, denn die Leistungsfähigkeit eines standardisierten Forschungsinstruments ist entscheidend davon abhängig, ob innerhalb einer vorgegebenen Struktur flexible Anpassungen an die spezifische Situation des jeweiligen Auftraggebers vorgenommen werden können.

Natürlich gibt es Forschungsprobleme, die einer maßgeschneiderten Lösung bedürfen und auf die standardisierte Forschungsinstrumente keine Anwendung finden können. Aber für viele Probleme der Umfrageforschung – insbesondere der Marktforschung – sind standardisierte Forschungsinstrumente ein angemessener Lösungsweg. Das gilt umso mehr, wenn die finanziellen Ressourcen der Auftraggeber begrenzt sind. Es ist in der Umfrageforschung wie in der Oberbekleidung: die qualitativen Unterschiede zwischen Maßanfertigungen und Konfektionsware verwischen sich zunehmend.

In Bezug auf die Qualität der Forschung wird häufig ein fehlendes Qualitätsbewusstsein der Auftraggeber bemängelt, was dazu führe, dass bei der Auftragsvergabe nicht auf die Qualität des Forschungsangebots geachtet wird, sondern ausschließlich auf den Preis. Die Institute würden sich diesem Preiswettbewerb weitgehend ohne Rücksicht auf die Forschungsqualität anpassen.

Ich möchte es an dieser Stelle dahingestellt lassen ob und in welchem Umfang diese Kritik an Auftraggebern und Forschungsinstituten berechtigt ist. Aber die Entwicklung standardisierter Forschungsinstrumente bietet den Instituten eine Möglichkeit, sich zumindest ein Stück weit aus dem reinen Preiswettbewerb zu entfernen und in einen Produktwettbewerb um die besseren Forschungsinstrumente einzutreten. Verdeutlicht wird dieses Bemühen durch die Tatsache, dass die Institute ihre standardisierten Forschungsinstrumente mit entsprechenden Bezeichnungen versehen und quasi als Markenartikel vertreiben.

5 Vierter Trend: Verstetigung des Forschungsprozesses

Der vierte Trend in der Umfrageforschung ist die Verstetigung des Forschungsprozesses. Dieser Trend ist vornehmlich eine Folge immer kürzerer Produktzyklen, die eine kontinuierliche Beobachtung des Marktgeschehens erfordern. Der Anteil des Umsatzes, den die Mitgliedsinstitute des ADM durch sogenannte Ad-hoc-Untersuchungen erzielt haben – also durch Einmal-Untersuchungen zur Klärung eines spezifischen und häufig aktuellen Problems –, ist folglich von 43 % im Jahr 1992 auf 36 % im Jahr 1998 zurückgegangen. Dagegen konnten kontinuierliche Forschungsprogramme – wie beispielsweise Werbetrackings – im gleichen Zeitraum einen – wenn auch moderaten – Zuwachs von 15 auf 17 % verzeichnen. Diese Form der Verstetigung des Forschungsprozesses und die zuvor beschriebene Standardisierung der Forschungsinstrumente stehen in einem engen Zusammenhang.

Die auffallendste Form der Verstetigung des Forschungsprozesses ist aber die zunehmende Bedeutung der Panelforschung, die im Zeitraum von 1992 bis 1998 ihren Umsatzanteil von 34 % auf 42 % steigern konnte. Allerdings darf diese Steigerung nicht der Panelforschung im klassischen Sinne, bei der dieselben Personen wiederholt befragt werden, zugeschrieben werden. Sie ist vielmehr auf die sogenannten Access- oder Convenience-Panel zurückzuführen. Bei diesem Forschungsinstrument werden die Namen und Adressen von Personen zusammen mit weiteren Merkmalen gespeichert – natürlich mit der entsprechenden Einwilligung der Betroffenen. Aus diesem Bestand befragungsbereiter Personen können dann für spezifische Studien Personen mit bestimmten Merkmalen und Merkmalskombinationen selektiert werden. Auf diese Weise ist es möglich, auch eng definierte Zielgruppen mit einem vertretbaren zeitlichen und finanziellen Aufwand zu erreichen. Entscheidend für die Qualität dieser Form der Umfrageforschung sind die Rekrutierung und die Pflege des Panels, d. h. des Bestandes befragungsbereiter Personen.

Auch wenn man sich meiner anfangs formulierten weiten Definition des Begriffs „Erhebungstechnik“ nicht anschließen und ihn auf die eigentlich technischen Aspekte der Datenerhebung begrenzen will, so wird man zumindest konstatieren müssen, dass aus der Verstetigung des Forschungsprozesses verschiedene Impulse zur Technisierung und Computerisierung der Datenerhebung ausgehen: Verbraucherpanel haben den Einsatz von Touchscreen und Scannertechnik in der Umfrageforschung erheblich gefördert und für Handelspanel werden durch die Ausweitung des sogenannten E-Commerce die Instrumente der Online-Forschung zunehmend wichtiger.

6 Fünfter Trend: Internationalisierung der Forschung

Der fünfte feststellbare Trend ist die Internationalisierung der Umfrageforschung. Diese Entwicklung wird sehr stark durch die Konsum- und Gebrauchsgüterindustrie beeinflusst, die in Deutschland – wie in anderen Ländern auch – mit einem Umsatzanteil von 55 % im Jahr 1998 der wichtigste Auftraggeber der privatwirtschaftlichen Markt- und Sozialforschungsinstitute ist.

Zu einem großen Teil haben sich die Unternehmen der Konsum- und Gebrauchsgüterindustrie schon längst zu Global Players entwickelt. Die Produkt- und Marketingstrategien dieser Unternehmen führen dazu, dass in der Marktforschung Untersuchungen zunehmend international – d. h. zeitgleich in mehreren Ländern – konzipiert und durchgeführt werden. Parallel zu

dieser Entwicklung gewinnen auch in der empirischen Sozialforschung international vergleichende Studien an Bedeutung. Und auch im Bereich der amtlichen Statistik ist eine zunehmende Internationalisierung festzustellen. Insofern können Mehr-Länder-Studien durchaus als ein neues Erhebungsinstrument verstanden werden.

Im Jahr 1998 haben die Mitgliedsinstitute des ADM nur noch 64 % ihres Umsatzes in Deutschland erzielt – davon 55 % durch in Deutschland ansässige und 9 % durch im Ausland ansässige Auftraggeber. Dementsprechend wurden 36 % des Umsatzes durch Tochtergesellschaften und Beteiligungen im Ausland erzielt; und dieser Anteil ist steigend.

Die Vergleichbarkeit der Ergebnisse internationaler Untersuchungen ist ein nicht zu unterschätzendes Problem, denn die inhaltliche Vergleichbarkeit der in den verschiedenen Ländern erhobenen Daten setzt methodische Vergleichbarkeit der Erhebungsinstrumente voraus. Die schon genannten standardisierten Forschungsinstrumente bieten diesbezüglich gewichtige Vorteile, denn sie führen zwangsläufig dazu, dass in den verschiedenen Ländern die Daten auf jeweils gleiche Art erhoben, analysiert und dargestellt werden.

Der Internationalisierung der Umfrageforschung entspricht zwangsläufig eine Globalisierung der Geschäftstätigkeit der Forschungsinstitute. Wenn die Auftraggeber zu Global Players geworden sind, müssen auch die Markt- und Sozialforschungsinstitute durch Gründung ausländischer Tochtergesellschaften oder durch Kauf bestehender Institute in den verschiedenen Ländern entweder selbst ein Global Player oder durch Allianzen und Kooperationen zumindest Teil eines Global Players werden.

Ich erwähne die Globalisierung der Geschäftstätigkeit der Markt- und Sozialforschungsinstitute im Zusammenhang mit den Chancen und Risiken neuer Erhebungstechniken, weil sie, wenn nicht sogar die Entwicklung, so doch zumindest den Einsatz standardisierter Forschungsinstrumente fördert. Diese wiederum erleichtern aber, wie schon gesagt, die inhaltliche Vergleichbarkeit der Ergebnisse internationaler Untersuchungen.

Globalisierung der Geschäftstätigkeit und die damit zusammenhängende Präsenz auf den relevanten Forschungsmärkten bedeutet aber wegen der neuen Erhebungstechniken nicht mehr zwangsläufig, jeweils auch physisch vor Ort vertreten sein zu müssen. Das ist ganz offensichtlich bei Umfragen mittels Internet der Fall. Es ist zunehmend auch bei telefonischen Umfragen der Fall, wenn internationale Untersuchungen zwar mit „native speakers“ in den jeweiligen Landessprachen, aber von einem zentralen Call Center aus durchgeführt werden.

7 Schlussbemerkungen

Ich werde zum Schluss meines Referats darauf verzichten, die fünf dargestellten allgemeinen Trends in der Umfrageforschung nochmals thesenartig zusammenzufassen. Stattdessen werde ich nur noch eine allgemeine Schlussbemerkung machen: Es gibt keinen „Königsweg“ der Datenerhebung in der Umfrageforschung mehr, sondern einen Pluralismus der Erhebungsinstrumente. Alle Erhebungsinstrumente – die „alten“ wie die „neuen“ – weisen spezifische Vor- und Nachteile auf, die sie jeweils für bestimmte Forschungsgegenstände und Forschungsansätze als besonders geeignet bzw. weniger geeignet erscheinen lassen. Die Entscheidung für eine bestimmte Technik der Datenerhebung sollte deshalb ausschließlich am

Forschungsgegenstand und dem gewählten Forschungsansatz orientiert sein und unabhängig von forschungsideologischen und wettbewerblichen Motiven erfolgen. Nur dann ist es möglich, die Chancen, die in den neuen Erhebungstechniken liegen, zu nutzen und die damit zweifellos auch verbundenen Risiken zu erkennen und zu vermeiden.

Mit der Entwicklung der Umfrageforschung bzw. ihrer Erhebungsinstrumente ist es ein wenig wie mit der Entwicklung der Gesellschaft insgesamt: Hätten wir jedes Risiko vermieden, würden wir immer noch in Höhlen leben; aber wollten wir jede Chance nutzen, würden wir vermutlich schon wieder in Höhlen leben.

Keine Alternativen: Telefon- und Face-to-Face-Umfragen

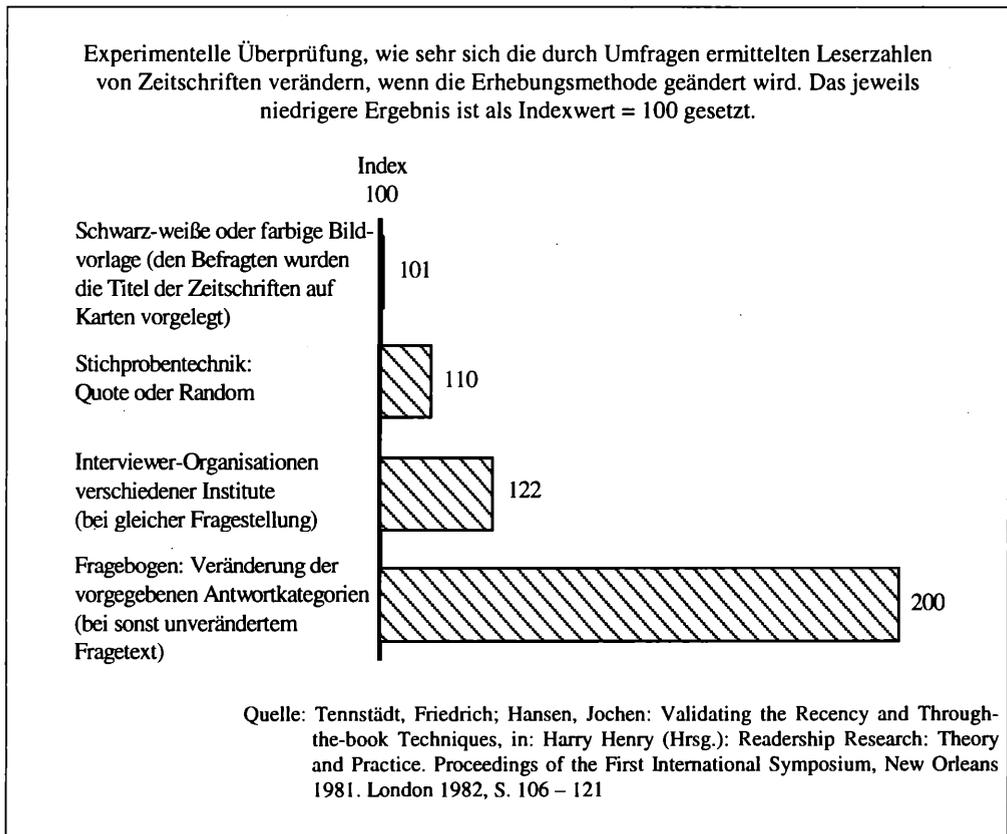
Der schnelle Aufstieg der Telefonumfragen seit Ende der siebziger Jahre wird begleitet von einer stetig wachsenden Zahl von Büchern und Artikeln in Fachzeitschriften, in denen die Vor- und Nachteile der Telefoninterviews mit denen der mündlich-persönlichen Interviews „von Angesicht zu Angesicht“ verglichen werden. Viele dieser Artikel listen die tatsächlichen bzw. mehr oder weniger begründet vermuteten Vorteile beider Verfahren auf, manchmal tabellenartig (z. B. Hippler/Schwarz 1990). Oft kommen solche Vergleiche zu dem Ergebnis, Telefonumfragen seien der herkömmlichen Face-to-Face-Befragung mindestens ebenbürtig, denn abgesehen vom Kostenvorteil könnten beim Telefoninterview die Interviewer wesentlich besser kontrolliert werden (Allerbeck 1985; Anders 1990). Eine Vielzahl von Untersuchungen beschäftigt sich mit Stichproben- oder Erhebungsfragen (z. B. Brückner 1985; Trometer 1990; Gabler/Häder 1997), hier wiederum liegt das Hauptgewicht der meisten Untersuchungen auf dem Thema Ausschöpfungsquoten und Nonresponse (z. B. Jung 1990; Porst 1991; Blasius/Reuband 1995; Kreiselmaier/Porst o. J.). Die Überprüfung der Validität der erhobenen Daten erfolgt in der Regel durch den Vergleich der soziodemographischen Merkmale der Befragten miteinander und gelegentlich mit den amtlichen statistischen Daten (z. B. Jung 1990). Die Tatsache, dass bei Telefonumfragen die Strichprobenziehung einfacher ist, wird gelegentlich als Argument dafür angeführt, dass Telefonumfragen den mündlich-persönlichen Interviews überlegen seien (von Gostomski u. a. 1997). Die hohen Verweigerungsquoten bei Telefoninterviews (Smith 1984) werden dagegen meistens als weniger schwerwiegend angesehen, zumal angenommen wird, dass das Face-to-Face-Interview stärker als das Telefoninterview anfällig für sozial erwünschte Antworten sei (Dillman 1978) – allerdings beruhen solche Aussagen meistens auf Vermutungen und nur selten auf experimentellen Daten.

Solche Vergleiche sind nützlich, notwendig und werden auch in diesem Aufsatz vorgenommen werden. Doch oft wird dabei unausgesprochen von der Annahme ausgegangen, bei Telefonumfragen und Face-to-Face-Umfragen handele es sich um Alternativen, um prinzipiell gleichermaßen leistungsfähige Verfahren, so dass nur ermittelt werden müsse, mit welcher Methode sich leichter repräsentative Stichproben bilden lassen, um zu entscheiden, ob Telefonumfragen oder mündlich-persönliche Umfragen grundsätzlich zu bevorzugen sind. Eine solche Annahme führt in die Irre. Bevor man sich der Frage zuwendet, welche spezifischen Verzerrungen die eine oder andere Interviewmethode bei unachtsamer Durchführung hervorbringen kann, muss man sich darüber im Klaren sein, dass das mündlich-persönliche Interview eine Vielzahl von Fragebogenmethoden und damit Analysetechniken zulässt, die das Telefon nicht erlaubt. Darum ist es höchst bedenklich, wenn Umfrageinstitute die Methode des Face-to-Face-Interviews vollständig zugunsten der Telefonumfrage aufgeben. Die Folge ist zwangsläufig eine Verminderung der Umfragequalität, nicht, weil die Stichprobe weniger repräsentativ wäre, sondern weil der Fragebogen verarmt. Und nichts hat einen größeren Einfluss auf die Ergebnisse von Umfragen als die Fragebogentechnik. Friedrich Tennstädt hat gezeigt, dass bereits eine rein formale Änderung der Antwortkategorien im Fragebogen das

*) Thomas Petersen, Institut für Demoskopie Allensbach.

Ergebnis mehr verändert als ein Wechsel des Stichprobenverfahrens (Tennstädt/Hansen 1982). Das Thema Fragebogen spielt dennoch in den meisten Artikeln und sogar in Handbüchern über Telefonumfragen nur eine geringe oder sogar gar keine Rolle (z. B. Lavrakas 1993).

Abbildung 1
Welchen Einfluss haben unterschiedliche Verfahren auf das Umfrageergebnis?



Eine Umstellung von Face-to-Face-Interviews auf Telefoninterviews bedeutet eine erhebliche Reduktion der methodischen Vielfalt. Es entfällt die Möglichkeit, im Interview Vorlagen oder visuelle Hilfen jeder Art einzusetzen. Das bedeutet, dass bei Telefonumfragen zahlreiche demoskopische Instrumente nicht oder nur sehr eingeschränkt eingesetzt werden können (Noelle-Neumann/Petersen 1998, S. 309 – 315), wie beispielsweise optische Skalen, Dialogbildblätter, viele projektive Tests, Tests von Verpackungsgrößen und -gestaltungen, viele Wissenstests, die Vorgabe von Kartenspielen mit Sortiervorgang, Vorführen von Gegenständen, Zeitschriften-Dummies (siehe hierzu ausführlich Noelle-Neumann 1998, S. 30 – 31).

Das hat zur Folge, dass sich wichtige Arten von Studien überhaupt nicht oder nur sehr eingeschränkt per Telefon durchführen lassen. Hierzu zählen unter anderem Markt-Media-Studien wie die Allensbacher Markt- und Werbeträger-Analyse (AWA) und die Media-Analyse (MA), sowie Umfragegutachten für die Rechtspraxis.

Darüber hinaus geht beim Telefoninterview auch die Möglichkeit verloren, den Interviewer als Beobachter in den Untersuchungsprozess einzubeziehen, Interviewerbeobachtungen einzusammeln, beispielsweise den Eindruck vom Befragten: Hat der Interviewer das Gefühl, der Befragte könnte anderen Menschen ein Vorbild sein? Wirkt er selbstsicher? Gepflegt? Wie interessiert oder gelangweilt wirkt der Interviewpartner? – ein Indikator für die Qualität des Fragebogens. Auch die Möglichkeit der Einschätzung der sozialen Schicht des Befragten durch den Interviewer geht verloren oder die Möglichkeit, Eindrücke über die Haushaltsführung oder die Nachbarschaft notieren zu lassen. Ebenso wenig ist am Telefon der Ausdruckstest durchführbar, der bei der Glücksforschung eine wesentliche Rolle spielt: Mit der knappen Interviewerbeobachtung „Der/die Befragte sieht fröhlich“, bzw. „nicht so fröhlich“ aus, kommt man zu etwa gleichwertigen Korrelationen mit dem vom Befragten geäußerten persönlichen Glücksempfinden wie mit den Ergebnissen mühsamer Frageserien zur emotionalen Stabilität, etwa Norman Bradburns „Effect-Balance Scale“ (Bradburn 1969; Noelle-Neumann 1974).

Das alles bedeutet, dass die Entscheidung, ob eine Untersuchung per Telefon oder als Face-to-Face-Befragung durchgeführt werden sollte, lange *vor* dem Zeitpunkt fallen muss, an dem man sich mit den Problemen beschäftigt, die die eine oder andere Methode in der technischen Abwicklung bereitet. Prinzipiell ist die Anwendbarkeit beider Verfahren – in den eben genannten Grenzen – unbestritten. Insbesondere die Validität der bei Face-to-Face-Befragungen ermittelten Ergebnisse ist durch zwölf Bundestagswahlprognosen des Allensbacher Instituts erwiesen (Noelle-Neumann/Petersen 1998, S. 25 – 27; Frankfurter Allgemeine Zeitung vom 26.9.1998, S. 3). Der Vergleich der Ergebnisse von Telefon- und mündlich-persönlichen Befragungen kann also nicht der Entscheidung dienen, welches Verfahren im Einzelfall oder gar grundsätzlich zu bevorzugen ist, sondern er dient der Feststellung, welche Fehlerquellen bei der Anwendung der beiden Erhebungsverfahren besonderes Augenmerk verdienen. Der Schwerpunkt der am Allensbacher Institut vorgenommenen Vergleiche zwischen Telefon- und Face-to-Face-Befragungen liegt daher nicht auf Stichprobenfragen, sondern auf der Frage, ob sich in den Ergebnissen von Umfragen beider Methoden, die sich teilweise beträchtlich voneinander unterscheiden, Muster herausbilden, anhand derer sich erkennen lässt, welche systematischen Verzerrungen der Ergebnisse die eine oder andere Methode begünstigt.

Man braucht eine größere Zahl von gleichartigen Feldexperimenten, um solche von der technischen Durchführung einer Umfrage abhängigen Grundmuster von Verzerrungen der Ergebnisse erkennen zu können. Die statistisch bedingten Schwankungen von Umfrageergebnissen können sogar dazu führen, dass man durch das Resultat einzelner Experimente in die Irre geführt wird (Petersen 1997). Doch gerade im Fall des Vergleichs zwischen Telefon- und Face-to-Face-Umfragen sind echte Experimente, deren Ergebnisse als uneingeschränkt gültig angesehen werden können, selten. Uns ist nur ein Fall bekannt, in dem exakt der gleiche Fragebogen in Telefoninterviews und mündlich-persönlichen Interviews gleichzeitig parallel verwendet worden ist (Geiger/Erben 1979). Dieses Experiment kam zu dem Ergebnis, dass das Interview als Telefoninterview nur etwa halb so lange dauerte wie als mündlich-persön-

liches Interview – obwohl der Fragebogen eine längere Item-Liste enthielt, die den Befragten im Face-to-Face-Interview vorgelegt wurde, während sie am Telefon Punkt für Punkt vorgelesen werden musste. Auch die Ergebnisse zeigten, dass am Telefon kürzer, flüchtiger geantwortet wurde. Insbesondere die Antworten auf offene Fragen fielen erheblich kürzer aus.

Darüber hinaus stellt das Allensbacher Institut seit einigen Jahren immer wieder gleichzeitig identische Fragen in Telefon- und mündlich-persönlichen Umfragen, um die Unterschiede in den Ergebnissen zu dokumentieren. Hierbei handelt es sich strenggenommen nicht um echte Experimente, weil das Frageumfeld nicht jeweils dasselbe ist, eventuelle Kontexteffekte also nicht kontrolliert werden. Auch ist die Zahl dieser Vergleiche noch lange nicht ausreichend. Doch die bisherigen Resultate können immerhin dazu dienen, empirisch gestützte Hypothesen aufzustellen, die in Zukunft der experimentellen Überprüfung bedürfen.

Der auffallendste Befund dieser Vergleiche ist der höhere Anteil der „Unentschieden“-Antworten in den „Face-to-Face“-Umfragen. Ein Phänomen, das auch andere Forscher beobachten (Reuband/Blasius 1996, S. 309 – 310). Nach Allensbacher Erfahrungen ist der „Unentschieden“-Anteil bei Telefon-Umfragen etwa ein Drittel niedriger als bei mündlich-persönlichen Befragungen.

Übersicht 1

Telefonumfragen: Weniger „Unentschieden“-Antworten

Basis: 18 Fragen mit zwei Antwortkategorien und der zusätzlichen Antwortmöglichkeit „Unentschieden“, gleichzeitig und mit jeweils identischer Formulierung parallel in mündlich-persönlichen und Telefonumfragen gestellt

Durchschnittlicher Anteil der Antwort „Unentschieden“

in den Telefonumfragen

15 %

in den Face-to-Face-Umfragen

22 %

Telefon: Jeweils etwa 500 Befragte,

Face to Face: Jeweils rund 2000 Befragte. Bevölkerung ab 16 Jahre. Untersuchungszeitraum 1994 – 1998.

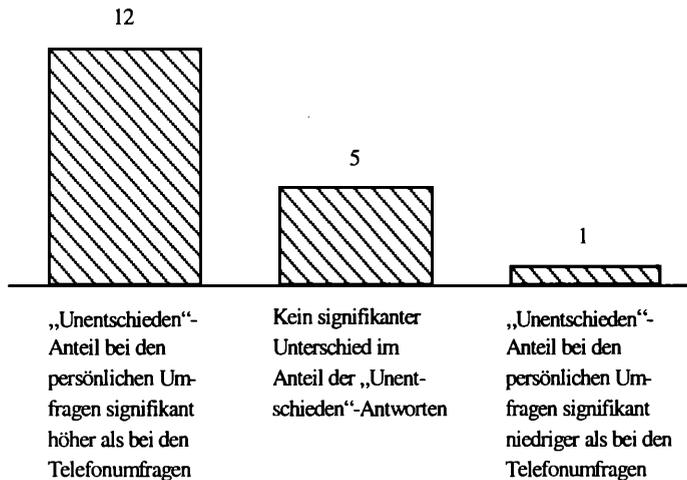
Quelle: Allensbacher Archiv, IfD-Umfragen

Abbildung 2 Telefonumfragen: Weniger „Unentschieden“-Antworten

Basis: 18 Fragen mit zwei Antwortkategorien und der zusätzlichen Antwortmöglichkeit „Unentschieden“, gleichzeitig und mit jeweils identischer Formulierung parallel in mündlich-persönlichen und Telefonumfragen gestellt

Richtung der Unterschiede in den „Unentschieden“-Antworten

– Zahl der Fälle –



Telefon: Jeweils etwa 500 Befragte,
Face to Face: Jeweils rund 2000 Befragte. Bevölkerung ab 16 Jahre. Untersuchungszeitraum 1994 – 1998.

Quelle: Allensbacher Archiv, IfD-Umfragen

Hinter der Kategorie „Unentschieden“ verbergen sich mehrere Dimensionen. „Unentschieden“ kann gewählt werden 1. als bequeme Auswegssituation für Befragte, für die es mühsam wäre, sich zu einer Entscheidung durchzuringen. 2. als differenzierte Antwort, die zunimmt, je mehr ein Befragter über eine Frage nachdenkt, sozusagen als Ausdruck des Zweifels auf den zweiten Blick, nachdem eine Antwort auf den ersten Blick einfach schien. 3. Als Fluchtkategorie unter moralischem Druck, die der Befragte wählt, wenn er fürchtet, sich mit einer klaren Meinungsäußerung sozial zu isolieren (für den theoretischen Hintergrund hierzu siehe Noelle-Neumann 1996). Welche Komponenten, also welche Verhaltensmuster der Befragten, sind für den höheren Anteil der „Unentschieden“-Antworten bei Face-to-Face-Umfragen verantwortlich? Einige Indizien lassen sich finden.

Es hat den Anschein, als wäre generell eine größere Verwirrung der Befragten beim mündlich-persönlichen Interview nicht für den höheren „Unentschieden“-Anteil verantwortlich. Stattdessen sieht es so aus, als würde „Unentschieden“ in den Face-to-Face-Umfragen überproportional oft in seiner Eigenschaft als inhaltliche Mittelposition gewählt. Bei Fragen mit drei Antwortkategorien, nämlich zwei Alternativen und einer Mittelposition wie „Kommt drauf an“ oder „Teils, teils“, ist es diese Mittelposition, auf die bei Telefonumfragen weniger Antworten entfallen, während der „Unentschieden“-Anteil bei beiden Methoden gleich groß ist.

Übersicht 2

Flüchtigeres Antworten im Telefoninterview schwächt die inhaltliche Mittelposition

Basis: 8 Fragen mit drei Antwortkategorien plus „Unentschieden“ gleichzeitig parallel und mit jeweils identischer Formulierung in Telefonumfragen und Face-to-Face-Umfragen gestellt

Durchschnittlicher Anteil der Antworten, die auf die inhaltliche Mittelposition („Teils, teils“, „Kommt darauf an“) entfielen

Telefon	Face to Face
28%	32%

Durchschnittlicher Anteil der Antwort „Unentschieden“

Telefon	Face to Face
12%	11%

Telefon: Jeweils etwa 500 Befragte,
Face to Face: Jeweils rund 2000 Befragte. Bevölkerung ab 16 Jahre. Untersuchungszeitraum: 1994 – 1998.

Quelle: Allensbacher Archiv, IfD-Umfragen

Einfacher als die Antwortvorgabe „Unentschieden“ lassen sich diese Mittelpositionen in zwei Gruppen einteilen: Zum einen gibt es die Fälle, in denen die Mittelposition für den Befragten als einfacher Ausweg erscheint, z. B. die Frage vom Frühjahr 1997: „Bei der Steuer wird sich ja einiges ändern. Was meinen Sie, werden Sie durch diese Änderungen alles in allem mehr Vorteile haben oder mehr Nachteile, oder wird sich da für Sie nicht viel ändern?“ Hier bietet die Antwort „Es wird sich nicht viel ändern“ dem Befragten die Möglichkeit, zu antworten ohne viel nachzudenken. Bei diesen Fragen wird die inhaltliche Mittelposition bei mündlich-persönlichen Umfragen nicht häufiger gewählt als bei Telefonumfragen. Anders sieht es bei den Fällen aus, bei denen die Mittelposition als Ergebnis eines gedanklichen Abwägungsprozesses erscheint, wie bei der Frage: „Haben Sie Angst, es könnte Ihnen etwas passieren, Sie könnten Opfer eines Verbrechens werden? Wie sehr fühlen Sie sich durch Verbrechen bedroht? Würden Sie sagen, Sie fühlen sich sehr bedroht, etwas bedroht, nicht bedroht?“ Hier erfordert die Entscheidung für die Mittelposition Nachdenken. Bei dieser und vergleichbaren Fragen wird die Mittelposition im Face-to-Face-Interview häufiger gewählt als beim Telefoninterview. Ein Ergebnis, das gut zu den oben berichteten Befunden von Geiger und

Erben (1979) passt: Im Telefoninterview wird flüchtiger geantwortet, der Befragte geht eher den Weg des geringsten Widerstandes. Die Tendenz vieler Befragter zum – wie es der amerikanische Sozialpsychologe Jon Krosnick ausdrückt – „Satisficing“, das heißt die Neigung, unter allen Antwortkategorien nicht unbedingt die zutreffendste auszuwählen, sondern die am leichtesten erreichbare der als hinreichend befriedigend empfundenen Kategorien (Krosnick 2000), scheint bei Telefonumfragen stärker zu sein als bei Face-to-Face-Interviews.

Übersicht 3 Beim Telefoninterview geht der Befragte den Weg des geringsten Widerstandes

Zwei Funktionen der inhaltlichen Mittelposition?			
Basis: 8 Fragen mit drei Antwortkategorien plus „Unentschieden“ gleichzeitig parallel und mit jeweils identischer Formulierung in Telefonumfragen und Face-to-Face-Umfragen gestellt			
Durchschnittlicher Anteil der Antworten, die auf die inhaltliche Mittelposition entfielen („Teils, teils“, „Kommt darauf an“)			
In den Fällen, in denen die Mittelposition als bequemer Ausweg erscheint (5 Fälle)		In den Fällen, in denen die Entscheidung für die Mittelposition Nachdenken verlangt (3 Fälle)	
Telefon	Face to Face	Telefon	Face to Face
30%	31%	24%	33%
Telefon: Jeweils etwa 500 Befragte, Face to Face: Jeweils rund 2000 Befragte. Bevölkerung ab 16 Jahre. Untersuchungszeitraum 1994 – 1998.			
Quelle: Allensbacher Archiv, IfD-Umfragen			

Tom W. Smith vom National Opinion Research Center (NORC) hat festgestellt, dass das Interview am Telefon von den Befragten als weniger interessant und als eine größere Störung und Belastung empfunden wird im Vergleich zum mündlich-persönlichen Interview (Smith 1984). Das Zusammenwirken dieser beiden Komponenten: Der Neigung der Befragten zu verkürzten Entscheidungsstrategien und der Tatsache, dass Telefoninterviews als besonders belastend empfunden werden, kann potentiell zu einem Validitätsproblem bei Telefonbefragungen führen.

Die Frage der Monotonie im Fragebogen und der Notwendigkeit, die Befragten immer wieder durch abwechslungsreiche Fragebogenkonstruktionen zu aktivieren und zu motivieren, spielt in der Methodendiskussion zu Unrecht derzeit nur eine geringe Rolle. Für den Befragten bedeutet es nicht dasselbe, wenn er eine größere Zahl von Items als Kartenspiel vorgelegt

bekommt, die er selbst sortieren kann, als wenn er alle diese Items nacheinander vorgelesen bekommt und zu jedem sagen muss, ob er ihm zustimmt oder nicht. Erp Ring hat bereits in den siebziger Jahren gezeigt, wie sehr die Validität von Umfragedaten, ihre Differenzierungskraft und damit ihre Brauchbarkeit für Analysezwecke durch zu lange und monotone Abfragen gefährdet wird (Ring 1976). Friedrich Tennstädt hat anhand eines Beispiels aus der Mediaforschung vorgeführt, dass flüchtiges Antworten als Folge belastender Fragebogen nicht nur Einflüsse auf das Antwortniveau hat, also den Anteil derer, die eine substantielle Antwort geben, sondern auch zu einer Verschiebung der Größenverhältnisse der Antwortkategorien zueinander führen kann. Tennstädt ermittelte in einer experimentähnlichen Untersuchung, bei der die Ergebnisse zweier ungefähr gleichzeitig durchgeführter Umfragen miteinander verglichen wurden, die Reichweiten verschiedener Zeitschriften mit einem in der Mediaforschung üblichen Fragemodell. In der einen Umfrage wurden insgesamt 28 Zeitschriftentitel abgefragt, in der anderen waren es 135.

Wie erwartet, waren die in der zweiten Umfrage gemessenen Reichweitenwerte geringer, aber nicht gleichmäßig. Unter dem Druck der vergleichsweise stärker ermüdenden Abfrage nannten die Befragten vorzugsweise nur noch solche Zeitschriften, die sie besonders regelmäßig lasen, das waren vor allem die wöchentlichen Publikationen, allen voran Programmzeitschriften. Die in längeren Intervallen erscheinenden Zeitschriften wurden dagegen überproportional seltener genannt (Tennstädt 1984). Ein Hinweis darauf, welche erheblichen Folgen für die Datenvalidität (und auch welche erheblichen finanziellen Folgen für die Verleger) es hätte, wenn man versuchen würde, die Reichweitenermittlung in der Mediaforschung von Face-to-Face-Befragungen auf stärker ermüdende Telefonumfragen umzustellen.

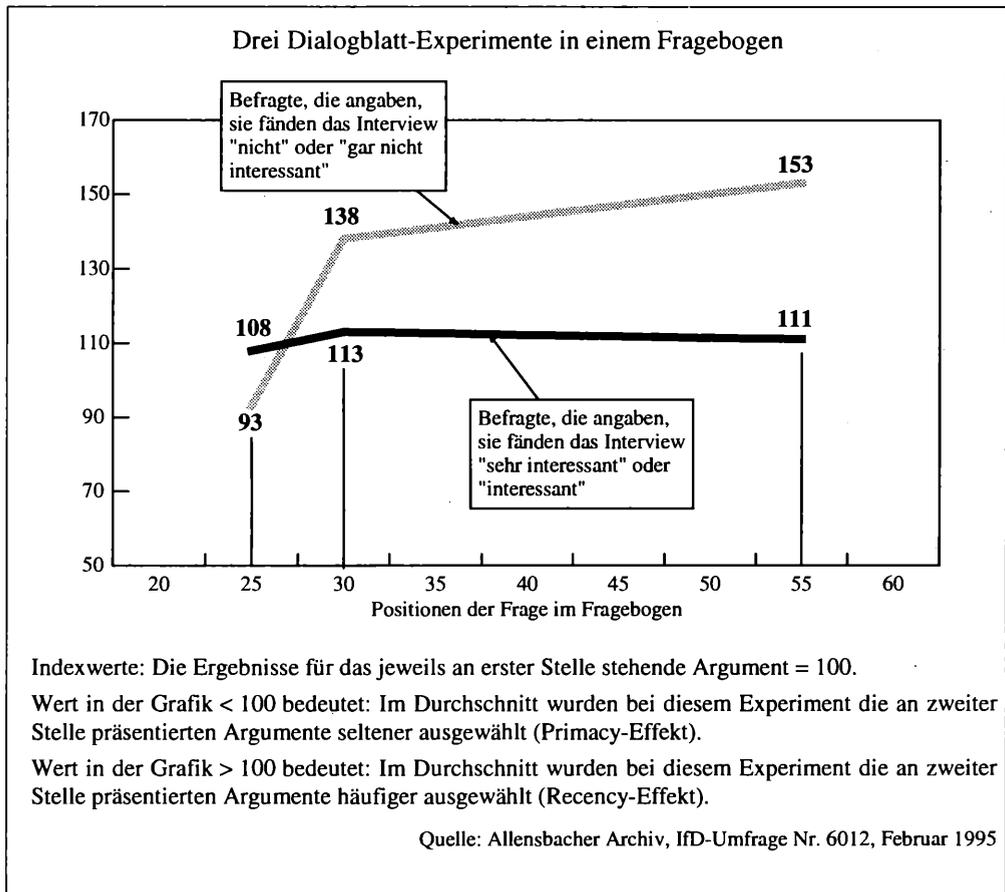
Übersicht 4
Die Folgen von Befragterermüdung in der Mediaforschung:
Verzerrungen in der Reichweitenmessung

Bei stärkerer Belastung gibt ein Teil der Befragten nur noch die häufig und regelmäßig gelesenen Titel an			
Es sagten, sie hätten innerhalb des letzten Erscheinungsintervalls gelesen oder durchgeblättert			
	Bei Abfrage von 28 Zeitschriften- titeln	Bei Abfrage von 135 Zeitschriften- titeln	Index
Durchschnittswert für 7 wöchentlich erscheinende Titel	15,5 %	12,6 %	81
Durchschnittswert für 7 zweiwöchentlich oder monatlich erscheinende Titel	13,2 %	8,3 %	63
n =	2 241	5 555	
Quelle: Tennstädt, Friedrich: Effects of Differing Methods on the Level of Magazine Readership Figures, in: Harry Henry (Hrsg.): Readership Research: Montreal 1983. Proceedings of the Second International Symposium, Amsterdam 1984, S. 229 – 241. Dort S. 235. Allensbacher Archiv, IfD-Umfragen Nr. 4686, Nov./Dez. 1981, 4007, 4009, April – Juli 1982			

Es gibt noch ein weiteres Indiz, das auf potentielle Unterschiede in der Datenvalidität zwischen Telefon- und mündlich-persönlichen Umfragen hindeuten könnte: Es spricht einiges dafür, dass das Telefoninterview etwas stärkere Reihenfolge-Effekte, konkret: Recency-Effekte produziert. Feldexperimente am Allensbacher Institut, bei denen die Reihenfolge der Präsentation von Argumenten in Dialogfragen variiert wurde, haben gezeigt, dass generell das zuletzt genannte Argument eine leicht erhöhte Chance hat, ausgewählt zu werden. Dieser rein auf die formale Gestaltung des Fragebogens zurückzuführende Effekt ist relativ gering, darum erst mit einer großen Zahl gleichartiger Experimente messbar. Er lässt sich wahrscheinlich auf die Organisation des menschlichen Gedächtnisses zurückführen: Während das zweite Argument noch im Kurzzeitgedächtnis des Befragten präsent ist, ist ihm das erste Argument bereits teilweise entfallen. Außerdem konnte festgestellt werden, dass der Recency-Effekt stärker ist, wenn dem Befragten die Argumente vorgelesen werden, als wenn ihm die Argumente auf einem Dialogbildblatt schriftlich präsentiert werden (Petersen 1997). Feldexperimente des Instituts für Demoskopie Allensbach zeigen, dass die Stärke von Recency-Effekten bei Face-to-Face-Interviews von der Position der betreffenden Frage im Fragebogen unabhängig ist – allerdings nur, wenn man die Gesamtergebnisse betrachtet. Bei

Befragten, die am Ende des Fragebogens sagten, für sie sei das Interview „nicht interessant“ oder „gar nicht interessant“ gewesen, lässt sich dagegen eine zunehmende Tendenz zu Recency-Effekten bei Fortschreiten des Interviews feststellen.

Abbildung 3
Gelangweilte Befragte neigen gegen Ende des Fragebogens zunehmend dazu, bei Dialogfragen das zweite Argument auszuwählen



Man muss also annehmen, dass Recency-Effekte verstärkt in solchen Interviews entstehen, die von den Befragten als besonders uninteressant und belastend empfunden werden – und damit in der Tendenz mehr in Telefon- als in mündlich-persönlichen Umfragen. Eine Auszählung von 20 Experimenten scheint diese Vermutung zu bestätigen: Bei jeweils gleicher Frageformulierung wurde in den Telefoninterviews die jeweils letztgenannte Antwortkategorie im Durchschnitt etwas häufiger ausgewählt als bei den mündlich-persönlichen Umfragen. Um hier Klarheit zu erhalten bräuchte man allerdings eine größere Zahl systematischer Reihenfolgeexperimente mit längeren Frageformulierungen.

Übersicht 5 Stärkere Recency-Effekte bei Telefonumfragen?

Basis: 20 Fragen mit zwei bis vier ausformulierten Antwortalternativen, die parallel und mit jeweils identischer Formulierung in Telefon- und mündlich-persönlichen Umfragen gestellt wurden

Durchschnittlicher Anteil der Antworten, der auf die im Fragetext jeweils letztgenannte Antwortalternative entfiel

bei Telefonumfragen

34 %

bei Face-to-Face-Umfragen

31 %

Telefon: Jeweils etwa 500 Befragte,

Face to Face: Jeweils rund 2000 Befragte. Bevölkerung ab 16 Jahre. Untersuchungszeitraum 1994 – 1998.

Quelle: Allensbacher Archiv, IfD-Umfragen

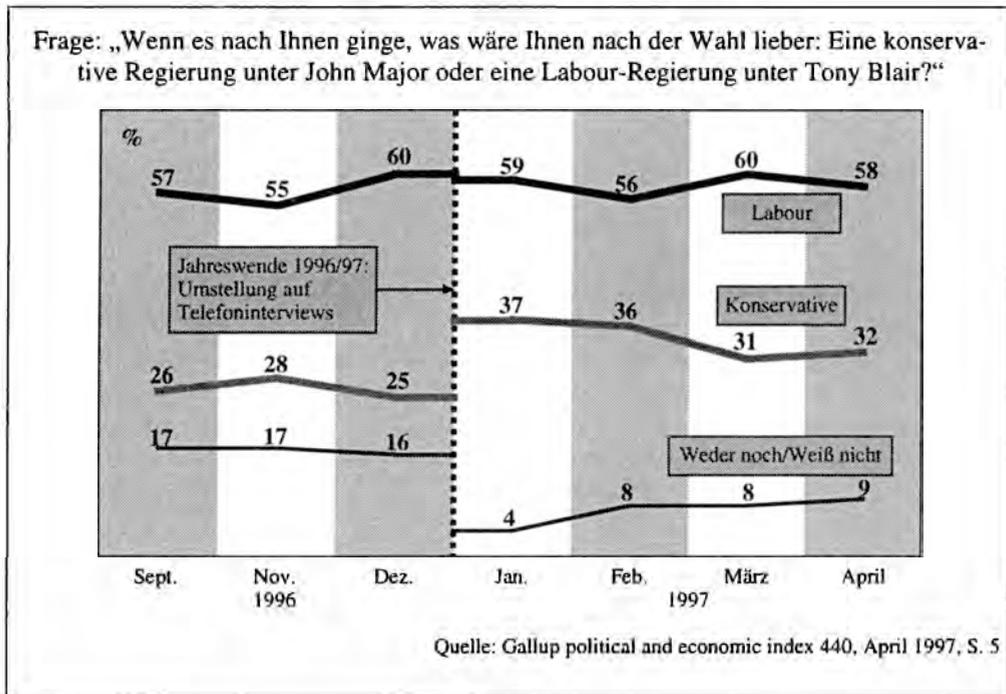
Schließlich könnte es noch einen weiteren Grund dafür geben, dass im Telefoninterview flüchtiger geantwortet wird als im mündlich-persönlichen Interview: Dieselbe Distanz zum Interviewer, die zu den im Durchschnitt höheren Verweigerungsraten am Telefon führt (Smith 1984) – es ist leichter, den Hörer aufzulegen als dem Interviewer die Tür vor der Nase zuzuschlagen – könnte auch hier zu bedenkenloseren Antworten führen. Es leuchtet ein, dass sich der Befragte im mündlich-persönlichen Interview mehr in einer Situation sozialer Kontrolle fühlt als beim Telefoninterview: Er muss den Interviewer in seine Wohnung lassen, er sitzt ihm persönlich gegenüber, blickt ihm ins Gesicht und muss jetzt diesem Menschen, mit dem er eine lockere aber doch persönliche soziale Beziehung eingegangen ist, seine Meinung offenbaren, nicht nur einer Stimme am Telefon.

Doch ein Phänomen ist noch nicht allein dadurch erklärt, dass es einleuchtend erscheint. Warum also könnte der soziale Druck im Face-to-Face-Interview stärker sein? Eine mögliche Erklärung bietet die Hirnforschung, wonach verschiedene Regionen des menschlichen Gehirns verschiedene Aufgaben erfüllen. Ganz grob vereinfacht kann man sagen, dass in der linken Hirnhälfte vorrangig Funktionen angesiedelt sind, die mit Rationalität, Verstand, abstraktem Denken zu tun haben. Hier sitzt auch das Sprachzentrum, werden gesprochene Worte besser aufgenommen, Sprache entschlüsselt. In der rechten Hirnhälfte ist dagegen vorrangig die Verarbeitung von Sinneseindrücken angesiedelt, eine stärker gefühlsbetonte Komponente. Dazu zählt die Aufnahme von Musik, die Wahrnehmung von Figuren und Formen (Vester 1975; vgl. auch Damasio 1995). Das bedeutet: Beim Telefoninterview wird in erster Linie die linke Hirnhälfte aktiviert, die für die Sprachentschlüsselung und die rationale Informationsverarbeitung zuständig ist. Beim Face-to-Face-Interview dagegen gibt es etwas Ungewohntes zu sehen, etwas zu greifen, zu sortieren. Hier werden vermutlich – es wäre eine reizvolle gemeinsame Aufgabe für die Sozial- und Hirnforschung, diese These empirisch zu überprüfen – beide Hirnhälften gleichermaßen aktiviert, so wie das mutmaßlich auch in der normalen Alltagssituation der Fall ist, in der sich Meinungen bilden und Entscheidungen gefällt werden.

Das heißt, der Interviewer ist beim mündlich-persönlichen Interview nicht nur physisch durch seine Anwesenheit stärker präsent, sondern wird vor allem vom Befragten auch gefühlsmäßig als Person stärker wahrgenommen, der gegenüber man sich den sozialen Normen entsprechend zu verhalten hat, der gegenüber man sich „falsch“ verhalten, blamieren kann. Das Telefoninterview regt dagegen nur das halbe Gehirn an, bildet nur die halbe Reaktion des Befragten ab.

Bisher handelt es sich bei diesen Ausführungen nur um – wenn auch begründete – Vermutungen. Doch es gibt Indizien, die diese Vermutungen zu bestätigen scheinen. Der vergleichsweise niedrige Anteil der „Unentschieden“-Antworten bei Telefonumfragen hat nämlich nicht zur Folge, dass die anderen Antwortkategorien gleichermaßen häufiger als bei Face-to-Face-Umfragen ausgewählt werden, sondern es ist vor allem die psychologisch schwächere Position, die Position, die unter moralischem Druck steht, die zu vertreten also Mut erfordert, die beim Telefoninterview stärker erscheint. Dies zeigen deutlich die Wahlumfragen des britischen Gallup-Instituts, das an der Jahreswende 1996/1997, und damit mitten im Unterhaus-Wahlkampf, seine Omnibus-Umfragen von Face-to-Face auf Telefon umstellte. Die Folge war, dass im Januar 1997 der Anteil der „Unentschieden“-Antworten stark zurückging, und zwar ausschließlich zugunsten der Positionen der konservativen Partei, die damals stark unter öffentlichem Druck stand. Die Zustimmungsraten zur Labour-Partei, ihren Vertretern und den von ihr vertretenen Positionen blieben nach dem Wechsel der Methode unverändert (mehrere Beispiele in: Gallup political and economic index 440, 1997).

Abbildung 4
Die unter Meinungsklimadruck stehende Position erscheint
bei Telefonumfragen stärker
– Wahlumfragen des britischen Gallup-Instituts –



Diesem Ergebnis scheinen in der Tendenz auch die Zahlen von Jung (1989, S. 19) zu entsprechen. Man kann annehmen, dass die Richtung und die Stärke des Meinungsklimadrucks sich in den Ergebnissen der Face-to-Face-Umfragen stärker niederschlagen als in den Telefonumfragen. Ein Teil derjenigen Befragten, die im Zweifel der konservativen Partei zuneigten, aber sich ihrer Meinung nicht sicher waren, zog es bei den mündlich-persönlichen Befragungen vor, sich unentschieden zu äußern und damit eine Position auszuwählen, die angesichts des vorhandenen Zweifels bei näherem Nachdenken doch angemessener erschien als eine klare Meinungsäußerung, und/oder auf die man sich zurückziehen konnte, ohne Angst zu haben, sich vor dem Interviewer zu blamieren.

Unsicherheiten der Befragten scheinen beim Telefoninterview schwerer sichtbar zu sein. Dies ist vor allem in der Panel-Forschung von Bedeutung: Meinungswechsel vollziehen sich nicht abrupt. Plötzliche vollständige Seitenwechsel vom einen Ende des Meinungsspektrums zum anderen sind selten. Tatsächlich beginnt eine Umorientierung mit Unsicherheit an der bisherigen Meinung und zeigt sich dementsprechend zuerst in einem Anstieg der „Unentschieden“-Angaben auf Kosten der schwächer werdenden Position, lange bevor die Gegenseite davon profitiert. Das Telefoninterview, in dem die „Unentschieden“-Position nur wenigen Befrag-

ten präsent ist, kann darum Wandlungen im Meinungsbild der Bevölkerung erst relativ spät abbilden, nämlich dann, wenn die Phase der ersten Verunsicherung bereits vorüber und der Vorgang der Umorientierung weiter fortgeschritten ist. In diesem Stadium allerdings reagiert dann die Telefonmethode hektischer, die Trends schwanken stärker (Tennstädt 1997).

Die Tatsache, dass die Befragten im mündlich-persönlichen Interview stärker auf sozialen, moralischen Druck reagieren als im Telefoninterview, wird gewöhnlich auf die unterschiedliche Interviewsituation zurückgeführt. Doch es ist denkbar, dass auch ein bisher unerkanntes Stichprobenproblem Anteil an diesem Befund hat. Nach Erfahrung des Instituts für Demoskopie Allensbach sind bei Stichproben von Telefonumfragen Personen mit hohem Selbstbewusstsein tendenziell überrepräsentiert. Auch Prozeduren, die eine Zufallsauswahl der Gesprächspartner im Haushalt ermöglichen sollen, wie beispielsweise die Frage nach der Person, die zuletzt Geburtstag hatte, hindern zahlreiche selbstbewusste Gesprächspartner nicht daran, sich selbst als die gesuchte Auskunftsperson auszugeben (Noelle-Neumann/ Petersen 1998, S. 313). Psychologisch bedingte Stichprobenverzerrungen, wie die mögliche Überrepräsentierung persönlichkeitsstarker Befragter bei Telefonstichproben, lassen sich durch Gewichtungsoperationen anhand von Faktoren wie Einkommen, sozio-ökonomischer Status oder Bildungsstand nur teilweise ausgleichen (vgl. Bien 1995), da sich besonders selbstbewusste Menschen – wie sie mit der Allensbacher Persönlichkeitsstärke-Skala identifiziert werden können (Noelle-Neumann 1985) – in allen sozialen und Bildungsschichten finden.

Oft machen sich psychologische Stichprobenverzerrungen in der Praxis der Umfrageforschung kaum bemerkbar. Doch dann, wenn der Untersuchungsgegenstand mit der die Stichprobe beeinflussenden Variable korreliert, kann ein solcher, unter Umständen lange un bemerkt gebliebener Faktor, zu Verzerrungen der Ergebnisse führen. Es wäre interessant, der Frage nachzugehen, ob, und wenn ja: wie weit, diese Komponente dazu beigetragen hat, dass sich die Allensbacher Umfrageergebnisse zu den Parteistärken vor der Bundestagswahl 1998, nicht aber vor der Bundestagswahl 1994 so sehr von den Resultaten anderer Institute unterschieden.

Aufschlussreich ist in diesem Zusammenhang eine Analyse mit Hilfe der eben erwähnten Skala der Persönlichkeitsstärke. Die „Persönlichkeitsstärke“ der Befragten wird mit der Frage ermittelt: „Hier sind verschiedene Eigenschaften. Könnten Sie die bitte einmal durchsehen und mir die Punkte nennen, wo Sie sagen würden: ‚Das passt auf mich, das trifft auf mich zu?‘“ Dazu werden zehn Kärtchen überreicht, auf denen die folgenden Aussagen stehen:

1. Gewöhnlich rechne ich bei dem, was ich mache, mit Erfolg.
2. Ich bin selten unsicher, wie ich mich verhalten soll.
3. Ich übernehme gern Verantwortung.
4. Ich übernehme bei gemeinsamen Unternehmen gern die Führung.
5. Es macht mir Spaß, andere Menschen von meiner Meinung zu überzeugen.
6. Ich merke öfter, dass sich andere nach mir richten.
7. Ich kann mich gut durchsetzen.
8. Ich bin anderen oft um einen Schritt voraus.
9. Ich besitze vieles, worum mich andere beneiden.
10. Ich gebe anderen öfter Ratschläge, Empfehlungen.

Diese zehn Aussagen wurden nicht ausgewählt, weil sie so einleuchtend klangen, sondern sie haben sich in einer Serie von Faktorenanalysen mit einer ursprünglich weit größeren Zahl von Items als zusammengehörig erwiesen. Jede dieser Aussagen wurde auf der Grundlage der „Faktorladungen“, also ebenfalls aufgrund der Ergebnisse der Faktorenanalysen, mit einem Punktwert versehen (detailliert beschrieben in Noelle-Neumann 1985). Je mehr Karten ein Befragter auswählt und sagt „das trifft auf mich zu“, ein umso größerer Punktwert wird ihm zugeordnet. Die 25 % der Befragten, die den höchsten Punktwert erreichen, bilden die Analysegruppe der „Persönlichkeitsstarken“ oder „Meinungsführer“.

Die Validität dieser Skala wurde unter anderem von Gabriel Weimann, Professor an der Universität Haifa, mit Netzwerkstudien in einem israelischen Kibbuz getestet. Weimann stellte fest, dass die mit der Skala als „persönlichkeitsstark“ identifizierten Personen tatsächlich eine zentrale Position im Netzwerk einnahmen und beispielsweise die Mehrzahl aller im Kibbuz geführten Gespräche über Politik anstießen (Weimann 1994). Inzwischen konnte auch anhand von Allensbacher Umfragedaten zu den Bundestagswahlen 1990 und 1994 gezeigt werden, dass eine Meinungsänderung in der Bevölkerung zu politischen Fragen mit einer deutlichen Verschiebung der Ansichten unter den Persönlichkeitsstarken beginnt. Der Rest der Bevölkerung folgt dann, mit einem Abstand von mehreren Monaten, nach (Noelle-Neumann/Kepplinger/Donsbach 1999).

Mit Hilfe der Persönlichkeitsstärke-Skala lässt sich nun simulieren, welchen Einfluss 1994 und 1998 eine Überrepräsentierung von persönlichkeitsstarken Befragten in der Stichprobe auf die Qualität von Wahlumfragen gehabt hätte. Gewichtet man den Anteil der persönlichkeitsstarken Befragten an der Gesamtstichprobe der Vorwahlumfragen aus dem Jahr 1998 hoch, so dass ihre Antworten mit einem Gewicht von 50 % statt mit 25 % in das Gesamtergebnis eingehen, verschlechtert sich dadurch die Prognoseauglichkeit der Daten deutlich. Nimmt man dagegen die gleiche Prozedur an den Vorwahlumfragen des Jahres 1994 vor, bleibt dies nahezu ohne Folgen: Eine Überrepräsentation von persönlichkeitsstarken Befragten hätte damals keine Verschlechterung der Prognosebasis bedeutet. Das Ergebnis stützt also die Vermutung, dass die Unterschiede zwischen den Ergebnissen der Allensbacher Face-to-Face-Vorwahlumfragen des Jahres 1998 auf der einen Seite und den Ergebnissen der Telefonumfragen anderer Institute auf der anderen Seite auf unterschiedliche Anteile von persönlichkeitsstarken Befragten in der Stichprobe zurückzuführen sein könnten.

Übersicht 6

**Bundestagswahl 1998: Eine Überrepräsentierung persönlichkeitsstarker Befragter
in der Stichprobe hätte zu einer schlechteren Prognosebasis geführt**

– Ergebnisse der letzten drei Befragungswellen vor der Wahl –

	Zweitstimmenwahlabsicht persönlichkeitsstarker Befragter	Zweitstimmenwahlabsicht anderer Befragter	Gesamtergebnis bei Anteil Persönlichkeitsstarker von 25 % (tatsächlicher Wert)	Gesamtergebnis bei Anteil Persönlichkeitsstarker von 50%	Wahlergebnis (Zweitstimmen)
			%		
CDU/CSU	37,6	35,3	35,9	36,5	35,1
SPD	37,0	42,3	41,0	39,7	40,9
Abweichung vom Wahlergebnis (Summe in Prozentpunkten)			0,9	2,6	

Quelle: Allensbacher Archiv, IfD-Umfragen Nr. 6085, 6066, 6067

Übersicht 7
**Bundestagswahl 1994: Eine Überrepräsentierung persönlichkeitsstarker Befragter
in der Stichprobe hätte die Prognosebasis nicht verschlechtert**
– Ergebnisse der viertletzten bis vorletzten Befragungswelle vor der Wahl –

	Zweitstimmenwahlabsicht persönlichkeitsstarker Befragter	Zweitstimmenwahlabsicht anderer Befragter	Gesamtergebnis bei Anteil Persönlichkeitsstarker von 25 % (tatsächlicher Wert)	Gesamtergebnis bei Anteil Persönlichkeitsstarker von 50 %	Wahlergebnis (Zweitstimmen)
			%		
CDU/CSU	41,3	38,3	39,1	39,8	41,5
SPD	32,7	35,0	34,3	33,9	36,4
Abweichung vom Wahlergebnis (Summe in Prozentpunkten)			4,4	4,2	

Quelle: Allensbacher Archiv, IfD-Umfragen Nr. 6004, 6005, 6007

Zusammenfassend lässt sich festhalten: Das Face-to-Face-Interview ist im Vergleich zum Telefoninterview das empfindlichere, aber auch das wesentlich differenziertere und variabelere Verfahren. Es ist die technisch und analytisch schwerer zu beherrschende Methode, deren Ergebnisse aber bei richtiger Anwendung und gleicher Frageformulierung mindestens so valide sind wie die des Telefoninterviews, und zu der es wegen der Vielseitigkeit ihrer Einsatzmöglichkeiten keinen Ersatz gibt.

Literaturhinweise

Allerbeck, Klaus (1985): Mündliche, schriftliche, telefonische Befragungen – ein Überblick, in: Max Kaase, Manfred Küchler (Hrsg.): Herausforderungen der Empirischen Sozialforschung. Beiträge aus Anlaß des zehnjährigen Bestehens des Zentrums für Umfragen, Methoden und Analysen, Mannheim, S. 56 – 65.

Anders, Manfred (1990): Praxis der Telefonbefragung, in: Forschungsgruppe Telekommunikation (Hrsg.): Telefon und Gesellschaft, Bd. 2, Berlin, S. 426 – 436.

Bien, Walter (1995): Gewichtung: Schönheitskorrektur von Stichproben mit dem Ziel, selber ruhiger zu schlafen und den Auftraggeber nicht mißtrauisch zu machen. Oder: Was nicht ist, ist nicht. Vortrag, gehalten auf der Herbsttagung der Deutschen Gesellschaft für Soziologie, Sektion Methoden, am 6.10.1995 in Marburg.

Blasius, Jörg; Reuband, Karl-Heinz (1995): Telefoninterviews in der empirischen Sozialforschung: Ausschöpfungsquoten und Antwortqualität, in: ZA-Information 37, S. 64 – 87.

Bradburn, Norman M. (1969): The Structure of Psychological Well-Being, Chicago.

Brückner, Erika (1985): Telefonische Befragungen – Methodischer Fortschritt oder erhebungswirtschaftliche Ersatzstrategie? in: Max Kaase, Manfred Küchler (Hrsg.): Herausforderungen der Empirischen Sozialforschung. Beiträge aus Anlaß des zehnjährigen Bestehens des Zentrums für Umfragen, Methoden und Analysen, Mannheim, S. 66 – 70.

Damasio, Antonio R. (1995): Descartes' Irrtum. Fühlen, Denken und das menschliche Gehirn, München und Leipzig.

Dillman, Don A. (1978): Mail and Telephone Surveys. The Total Design Method, New York u. a.

Gabler, Siegfried; Häder, Sabine (1997): Wirkung von Gewichtungen bei Face-to-Face- und Telefonstichproben. Eurobarometer-Experiment 1994, in: Siegfried Gabler, Jürgen H. P. Hoffmeyer-Zlotnik (Hrsg.): Stichproben in der Umfragepraxis, Opladen, S. 221 – 245.

Geiger, Herbert; Erben, Edgar (1979): Methodentest zur Überprüfung der Brauchbarkeit schriftlicher und telefonischer Umfragen zur Analyse des lokalen Bankenwettbewerbs, in: ESOMAR (Hrsg.): The Challenge of the Eighties (ESOMAR-Kongreß, Brüssel 2. – 6. September 1979), Amsterdam, S. 225 – 243.

von Gostomski, Christian Babka; Hartmann, Josef; Thum, Miriam (1997): Die Mannheimer Scheidungsstudie: Aspekte der Durchführung der telefonischen Befragung zu Determinanten der Ehescheidung, in: ZUMA-Nachrichten 41, S. 127 – 152.

Hippler, Hans-J.; Schwarz, Norbert (1990): Die Telefonbefragung im Vergleich mit anderen Befragungsarten, in: Forschungsgruppe Telekommunikation (Hrsg.): Telefon und Gesellschaft, Bd. 2, Internationaler Vergleich – Sprache und Telefon – Telefonseelsorge – Telefoninterviews, Berlin, S. 437 – 447.

Hyman, Herbert H. (1954): Interviewing in Social Research, Chicago.

Jung, Matthias (1989): Auf dem Weg zu einer besseren Datenqualität. Ein Zwischenbericht über die Erfahrungen mit telefonischen Umfragen. Arbeitspapiere der Forschungsgruppe Wahlen e. V., Mannheim, Nr. 1 – 26.

Jung, Matthias (1990): Ausschöpfungsprobleme bei repräsentativen Telefonumfragen, in: Forschungsgruppe Telekommunikation (Hrsg.): Telefon und Gesellschaft, Bd. 2, Berlin, S. 386 – 398.

Kreiselmaier, Jutta; Porst, Rolf (o. J.): Methodische Probleme bei der Durchführung telefonischer Befragungen: Stichprobenziehung und Ermittlung von Zielpersonen, Ausschöpfung und Non-Response, Qualität der Daten, ZUMA-Arbeitsbericht Nr. 89/12.

Krosnick, Jon A.; Fabrigar, Leandre R. (2000): Agree/Disagree, True/False, and Yes/No Questions: The Problem of Acquiescence, in: Jon A. Krosnick, Leandre R. Fabrigar: Designing Good Questionnaires, New York. Erscheint voraussichtlich 2000.

Lavrakas, Paul J. (1993): Telephone Survey Methods. Sampling, Selection, and Supervision, Newbury Park u. a.

Lazarsfeld, Paul F. (1935): The Art of Asking Why. Three Principles Underlying the Formulation of Questionnaires, in: National Marketing Review 1, No. 1, S. 32 – 43.

Noelle-Neumann, Elisabeth (1974): Lebensfreude – kein Thema für die Kirche? Fragen zu einem Test über Bewegungs- und Minenspiel von Katholiken, in: Herder Korrespondenz 28, S. 41 – 47.

Noelle-Neumann, Elisabeth (1985): Die Identifizierung der Meinungsführer. 38. ESOMAR-Kongreß, Wiesbaden, 1. – 5. September 1985.

Noelle-Neumann, Elisabeth; Köcher, Renate (1987): Die verletzte Nation, Stuttgart.

Noelle-Neumann, Elisabeth (1996): Öffentliche Meinung. Die Entdeckung der Schweigespirale. Frankfurt am Main und Berlin, vierte Auflage.

Noelle-Neumann, Elisabeth; Köcher, Renate (Hrsg., 1997): Allensbacher Jahrbuch der Demoskopie Bd. 10, 1993 – 1997. Demoskopische Entdeckungen, München und Allensbach.

Noelle-Neumann, Elisabeth (1998): Einsatz und Qualifikation der Interviewer als Qualitätsmerkmal der Umfrageforschung, in: Statistisches Bundesamt (Hrsg.): Interviewereinsatz und -qualifikation, Schriftenreihe Spektrum Bundesstatistik, Bd. 11, Stuttgart, S. 10 – 50.

Noelle-Neumann, Elisabeth; Petersen, Thomas (1998): Alle, nicht jeder. Einführung in die Methoden der Demoskopie, München, zweite Auflage.

Noelle-Neumann, Elisabeth; Kepplinger, Hans Mathias; Donsbach, Wolfgang (1999): Kampa. Meinungsklima und Medienwirkung im Bundestagswahlkampf 1998, Freiburg.

Petersen, Thomas (1997): The Last Shall Be First. Effects of question Order and Wording. Taking Stock of 50 Years of Split-Ballot Experiments. Paper Presented on May 18, 1997 at the 52nd Annual Conference of the American Association for Public Opinion Research (AAPOR) in Norfolk, Virginia/USA.

Porst, Rolf (1991): Ausfälle und Verweigerungen bei einer telefonischen Befragung, in: ZUMA-Nachrichten 29, S. 57 – 69.

Reuband, Karl-Heinz; Blasius, Jörg (1996): Face-to-Face-, telefonische und postalische Befragungen. Ausschöpfungsquoten und Antwortmuster in einer Großstadt-Studie, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie 48, S. 296 – 318.

Ring, Erp (1976): Questionnaire Monotony Endangers the Comparability of Results. It Should Be Avoided, in: Research That Works for Today's Marketing Problems. ESOMAR Congress Report, Amsterdam 1976, S. 429 – 446.

Smith, Tom W. (1984): A Comparison of Telephone and Personal Interviewing. 2nd Draft. Prepared for the General social Survey (GSS), Board of Overseers. Manuscript NORC, Chicago.

Tennstädt, Friedrich (1984): Effects of Differing Methods in the Level of Magazine Readership Figures, in: Harry Henry (Hrsg.): Readership Research: Montreal 1983. Proceedings of the Second International Symposium, Amsterdam 1984, S. 229 – 241.

Tennstädt, Friedrich (1997): General Election vom 1. Mai 1997 in Great Britain. Ein Rückblick auf die Opinion Polls. Manuskript im Allensbacher Archiv.

Tennstädt, Friedrich; Hansen, Jochen (1982): Validating the Recency and Through-the-Book Techniques, in: Harry Henry (Hrsg.): Readership Research. Theory and Practice. Proceedings of the First International Symposium. New Orleans 1981. London 1982, S. 106 – 121.

Trometer, Reiner (1990): Zur Durchführbarkeit von Allgemeinen Bevölkerungsumfragen als telefonische Befragung: Eine Analyse am Beispiel des ALLBUS 1988, in: ZUMA-Nachrichten 26, S. 72 – 78.

Vester, Frederick (1975): Denken, Lernen, vergessen. Was geht in unserem Kopf vor, wie lernt das Gehirn, und wann läßt es uns im Stich? Stuttgart.

Weimann, Gabriel (1994): The Influentials. People Who Influence People, New York.

Folgen eines Wechsels von Face-to-Face- zu telefonischen Befragungen

Telefoninterviews gehören heute zum Standard allgemeiner Bevölkerungsumfragen. Dazu haben primär die offensichtlichen Zeit- und Kostenvorteile gegenüber einer Face-to-Face-Befragung beigetragen. Weil das typische Stichprobendesign und der Einsatz der CATI-Systeme die Rolle der Interviewer systematisch verändert haben, wird häufig auch eine bessere Datenqualität reklamiert. Das trifft – genauso wie bei den neueren CAPI-Verfahren – sicherlich für die Befragung selbst zu, weil die Abfolge der Fragen mit komplexer Filterführung und zufallsvariiertes Item-Vorlage zentral gesteuert werden kann. Effekte der Stichprobenplanung und ihrer Realisierung sind nicht so eindeutig als Pro und Contra zuzurechnen, so dass die Vor- und Nachteile beider Verfahren abhängig von der Zielsetzung einer Erhebung und den verfügbaren Ressourcen abzuwägen bleiben.

Die Folgen des Wechsels von Face-to-Face- zu telefonischen Befragungen lassen sich unter zwei Aspekten getrennt analysieren und bewerten: Systematische Fehler können aus dem jeweiligen Stichprobenplan und aus seiner Realisierung resultieren. Obwohl die Realisierungsprobleme der Erreichbarkeit und Verweigerung schwerer wiegen, sollen zunächst die Design-Probleme kurz verdeutlicht werden.

CATI-Erhebungen sind an die simple Vorbedingung einer „ausreichenden“ Telefon-Versorgung der privaten Haushalte gebunden. Üblicherweise wird eine Telefondichte von 95 % als Schwelle angesehen. Zur Beurteilung der „Machbarkeit“ von Telefonstichproben ist diese Quote jedoch nicht das wichtigste Kriterium: Die Telefon-Haushalte dürften sich in ihren sozialstrukturellen Merkmalen nicht von der zu untersuchenden Bevölkerung unterscheiden. Zuverlässiger als Umfragen, die den Anteil der telefonisch Erreichbaren immer überschätzen, informiert darüber die „Einkommens- und Verbrauchsstichprobe“, die vom Statistischen Bundesamt alle fünf Jahre erhoben wird.

Als die Forschungsgruppe Wahlen im Herbst 1987 mit telefonischen Interviews für den monatlichen ZDF-Politbarometer experimentierte, war in Westdeutschland erst eine Telefondichte von 93 % erreicht. Lange Jahre war die Ausstattung mit Telefonen schichtspezifisch unterschiedlich; problematisch niedrig war damals die Erreichbarkeit der Einpersonenhaushalte, deren Anschlussquote sich erst 1993 den größeren Haushalten annäherte. Die damals gemachten Erfahrungen sind also nicht direkt auf heutige Verhältnisse zu übertragen.

*) Dr. Wolfgang Schulte, Forschungsgruppe Wahlen, Mannheim.

**Tabelle 1: Ausstattungsgrad der Haushalte mit Telefonen
nach Haushaltsgrößen und der sozialen Stellung des Haushaltsvorstands
in % der Haushalte**

	West 1973	West 1978	West 1983	West 1988	West 1993	Ost 1993	West 1998	Ost 1998
Insgesamt	51,5	69,5	88,1	93,2	97,4	48,8	97,3	94,3
Haushaltsgröße								
1 Person	40,9	60,0	78,8	86,8	95,0	42,8	94,9	88,8
2 Personen	51,6	71,9	90,7	95,3	98,6	53,9	98,7	96,2
3 Personen	56,6	75,0	93,4	97,5	98,6	48,9	98,5	96,6
4 Personen	56,7	73,9	93,9	97,6	98,8	49,9	98,9	97,9
5 Personen und mehr	52,6	72,1	93,1	97,7	98,5	43,9	98,9	96,8
Soziale Stellung des Haushaltsvorstands								
nicht erwerbstätig	43,4	64,2	84,3	91,5	97,1	50,1	97,2	94,5
erwerbstätig								
Arbeiter	34,4	58,3	86,2	93,7	96,6	39,2	96,9	92,6
Angestellte	68,6	81,7	93,0	96,0	98,8	55,5	98,3	96,5
Beamte	69,8	82,8	94,0	96,7	99,0	46,1	98,4	93,4
Selbständige	89,3	93,0	97,0	97,1	99,2	75,7	98,8	96,9
Landwirte	49,5	69,8	89,4	96,8	99,4	–	99,9	–

Quelle: Statistisches Bundesamt (Hrsg.): Statistisches Jahrbuch für die Bundesrepublik Deutschland, Ausgaben 1975 – 1994; Statistisches Bundesamt (Hrsg.): Wirtschaftsrechnungen. Einkommens- und Verbrauchsstichprobe 1998, Fachserie 15, Heft 1, Langlebige Gebrauchsgüter privater Haushalte

1 Stichprobendesign der beiden Erhebungsmethoden

Das Stichprobendesign der beiden Erhebungsmethoden unterscheidet sich in einem zentralen Punkt: Telefonstichproben können regional breit streuen, weil im Prinzip alle privaten Haushalte mit Telefonanschluß für einstufige Auswahl zur Verfügung stehen. Stichproben für Face-to-Face-Befragungen schalten aus ökonomischen Gründen des Interviewereinsatzes eine Flächenstichprobe davor und haben in der Regel wegen dieser Klumpung einen höheren Zufallsfehler.

Stichproben aus einer adäquat geschichteten Grundgesamtheit führen zu genaueren Schätzungen der Varianzen als einfache Zufallsstichproben. Wenn die Parameter für wichtige Variablen der Grundgesamtheit bekannt sind, ergibt sich ein leider selten genutzter Vorteil: die Qualität der Zufalls-Stichprobe könnte vor ihrem Einsatz überprüft werden. Doch das Kriterium für „sinnvolle“ Schichtung scheint beiden Verfahren abhanden gekommen zu sein.

Für die ADM-Flächenstichprobe ist die Gesamtheit der Wahlbezirke regional nur nach Boustedt-Regionen bzw. den heutigen BIK-Gemeindetypen (Behrens 1994) geschichtet, im wesentlichen also nach administrativen Ortsgrößenklassen, die nach den Gemeindereformen der siebziger und achtziger Jahre kaum noch die demographische Struktur der Bevölkerung indizieren. Anscheinend ist in Vergessenheit geraten, dass mit den Parteiateilen in den Wahlbezirken bessere Indikatoren für soziale Kontexte zur Verfügung stehen.

Für nationale Telefonstichproben ist eine Schichtung nach sozial definierten Kontexten nahezu unmöglich. Die Ortsnetze stimmen kaum mit den Grenzen der politischen Gemeinden überein; es gibt für sie keine empirisch gesicherten Informationen zur Bildung und Schichtung von Primäreinheiten. Somit bleiben nur regionale Klassifikationen als „Ersatz“. Unter Verweis auf diese Inkongruenz wird eine geeignete Schichtung in der Literatur de facto nicht thematisiert; vielmehr gilt die mikrostratifizierte Wahrscheinlichkeitsauswahl der Gemeinden und Stadtbezirke als vorbildlicher Standard (von der Heyde 1998, S. 134 f.). Der Nachteil dieses Verfahrens wiegt jedoch wegen der großen Anzahl der Sample Points nicht so schwer wie im ADM-Design.

Beide Verfahren benutzen den Haushalt nur als Vehikel, die zu befragende Person zu erreichen, ohne die Grundgesamtheit der privaten Haushalte genau zu kennen und für die Zufallsauswahl darüber zu verfügen. Für die Flächenstichprobe kann die Zahl der Privathaushalte je Gemeinde und anteilig auch je Sample Point hinreichend genau aus der Volkszählung und der Fortschreibung der Bevölkerung geschätzt werden. Für die zweite Stufe, die Auswahl der Haushalte, gilt beim Random Walk – auch mit Adressenvorlauf – wohl eher das Prinzip Hoffnung. Normalerweise gibt es keine Auflistung aller Haushalte; vielmehr steuert die Handlungsweise der Interviewer die Auswahl. Bereits die faktischen Schwierigkeiten, jeden Haushalt hinter einem Eingang oder Klingelschild zu identifizieren, lassen eine echte Zufallsauswahl im Sample Point nur mit Einschränkungen zu. Der übliche, knapp kalkulierte Stücklohn nur für realisierte Interviews bietet für die Interviewer keinen Anreiz, sich mit „schwierigen“ Adressen zu befassen; Abweichungen vom „Random“ Walk werden von den Instituten stillschweigend toleriert, häufig gar erwartet.

In der Praxis der Face-to-Face-Umfragen ist die geplante Zufallsauswahl der Haushalte also nicht unabhängig von der Realisierungschance; ein im Stichprobenplan nicht intendierter systematischer Fehler ist kaum zu vermeiden. Unter diesem Aspekt haben telefonische Umfragen eine bessere Qualität, weil der Spielraum der Interviewer bei der Auswahl minimal bleibt und effektiv kontrollierbar ist.

Ausreichende Telefondichte unterstellt, sind Telefonstichproben auf der Basis einer reinen Listenauswahl trotzdem nicht ohne grundsätzliche Probleme möglich. Außer der mangelnden Aktualität der Verzeichnisse und der unsicheren Identifikation von Privatanschlüssen haben die „anonymen“ Anschlüsse heute dazu geführt, dass die gelisteten Telefonnummern nicht mehr der intendierten Grundgesamtheit entsprechen. Seit 1992 ist der Eintrag im Telefonbuch freiwillig. Ende 1993 waren erst 3,7 % aller Anschlüsse nicht eingetragen (Häder 1994, S. 14), aber unter den überwiegend neuen Anschlüssen in Ostdeutschland nahm der Anteil sprunghaft zu. Nach einer neueren Untersuchung sind im Bundesdurchschnitt 13 – 14 % der Teilnehmer nicht eingetragen; mittelfristig ist eine Quote von 20 % zu erwarten (Follmer und Smid 1998). Weil der Nicht-Eintrag (im Westen stärker als im Osten) mit Alter und Bildung sowie der Gemeindegröße korreliert, würde das Ignorieren dieser Gruppe zu systematischen Auswahlfehlern einer Stichprobe führen.

Eine der in der Literatur häufig beschriebenen Varianten der Zufallsauswahl besteht darin, die Endziffern der ausgewählten vorhandenen Telefonnummern zufällig zu ändern: „add a digit“ addiert eine Konstante oder Zufallsziffer, „randomize last digit“ (RLD) ersetzt sie. Dadurch entsteht wieder eine Zufallsstichprobe von Telefonnummern, mit der auch solche Teilgruppen zu erreichen sind, die in den verfügbaren Telefonverzeichnissen fehlen. Doch sie hat das Manko einer abhängigen Auswahl – abhängig davon, dass die Originalnummer im Telefonbuch steht. Mit den Vorschlägen von Gabler/Häder (1997 und 1998) wird dieses Problem lösbar. Ihr Algorithmus garantiert die gleiche, angebbare Auswahlchance aller gelisteten und nicht gelisteten Telefonnummern; mit der ISDN-Technologie und dem Mobiltelefon verschärft sich aber die ungleiche Inklusionswahrscheinlichkeit der Telefon-Haushalte.

2 Realisierung: Erreichbarkeit und Verweigerung

Die simple Frage, ob die ausgewählten Personen/Haushalte telefonisch besser als durch Interviewerbesuch zu erreichen und zu befragen sind, ist nicht eindeutig zu beantworten.

Eine hohe Ausschöpfungsquote ist für sich allein genommen kein Beweis für die Qualität, sie lässt sich nur im Zusammenhang mit den Feldprozeduren vergleichen und beurteilen. Im klassischen Random Route wird die zulässige Anzahl der Haushaltsadressen je Sample Point proportional zur Stichprobengröße festgelegt; preiswerter, deshalb häufiger wird nur ein großzügiges Maximum vorgegeben, und die Ausschöpfungsquote wird nicht vergleichbar nur auf die tatsächlich benutzten Adressen bezogen. Umfragen mit Adressenvorlauf oder Vorgaben aus Melderegistern haben zumeist geringere Ausschöpfung und können trotzdem höhere Übereinstimmung mit dem Mikrozensus ergeben (Koch 1998).

Ausschöpfungsquoten von Telefonstichproben sind damit nicht direkt vergleichbar, weil sich darin eine vom Stichprobendesign abhängige Anzahl von Geschäftsanschlüssen und von im RLD-Verfahren generierten, aber nicht angeschlossenen Telefonnummern befindet. Obwohl deshalb auch die Zahl der echten Nichtkontakte unsicher bleibt, ist am ehesten die „Netto“-Ausschöpfung der um diese neutralen Ausfälle bereinigten Stichprobe zu vergleichen. Normalerweise erzielen Telefonumfragen eine geringere Ausschöpfung; bei vergleichbarem Anteil der expliziten Verweigerungen resultiert sie aus der wegen der üblicherweise sehr kurzen Feldzeit geringeren Erreichbarkeit der Haushalte.

Zur Beurteilung einer realisierten Umfrage sind Erreichbarkeit und Verweigerung nur noch analytisch zu trennen, weil die erhobenen Daten keine oder zuwenig Informationen darüber liefern können. Deshalb wurde u. a. von Hartmann (1990) versucht, die Zusammensetzung der Nonrespondents aus einem Vergleich der ALLBUS-Daten mit dem Mikrozensus als Referenzstatistik zu erschließen. Die enormen Abweichungen beim Merkmal Haushaltsgröße lassen sich nicht mit der erhebungstechnisch verursachten „Mehrproduktion“ an Einpersonenhaushalten im Mikrozensus (Wedel 1989) wedgdiskutieren; auch der Verweis auf den in der Praxis nicht immer identisch definierten „Umfragehaushalt“ hilft nicht weiter. Entscheidend ist die in der Literatur durchgängig berichtete leichtere Erreichbarkeit größerer Haushalte. Bedingt durch das ADM-Design als Haushalts-Stichprobe führt die Transformation in eine Personenstichprobe dann bei allen mit der Haushaltsgröße korrelierten Variablen zu einer Verschlechterung der Schätzungen: Alte, insbesondere Frauen, bewohnen überdurchschnittlich oft kleine Haushalte. Für den beabsichtigten Vergleich konstatiert Hartmann zu Recht ein Dilemma: Die Auswahlwahrscheinlichkeit wäre als Haushalts-

gewichtung zu berücksichtigen, aber das an sich richtige Verfahren würde zu größeren Abweichungen in diesem Variablenbereich führen. In ihrer späteren Kritik haben Rendtel/Pötter (1993, S. 356) wohl übersehen, dass die Transformation als nachträgliche Schichtung einer Stichprobe ihre vollständige und fehlerfreie Realisierung voraussetzt. Wenn aber der durch Ausfälle verursachte systematische Fehler eindeutig überwiegt, ist es durchaus dem Problem angemessen, im Folgenden nur die ungewichteten Daten der Erhebungen 1987 und 1988 zu analysieren.

Zum Teil extreme Abweichungen ergeben sich im ALLBUS – wie in vielen anderen Studien auch – bei der mit Schulbildung und Beruf nur unzureichend beschreibbaren Schichtzugehörigkeit: Personen mit geringer Schulbildung sind unter-, diejenigen mit hoher Bildung dagegen stark überrepräsentiert. Unter den insgesamt unterschätzten Erwerbstätigen sind die Arbeiter deutlich weniger, Angestellte und Beamte zu stark vertreten. Hartmann wertet das in vorsichtiger Formulierung als Indiz für systematische Fehler, die mit einiger Sicherheit als Nonresponse-Effekt zu interpretieren seien. Seitdem die Daten des Mikrozensus für die Wissenschaft zugänglich sind, konnte dieser Effekt genauer analysiert werden. Ohne in die fragwürdige Logik der traditionellen Repräsentanz-„Beweise“ zu verfallen, belegen Hartmann/Schimpl-Neimanns (1992, S. 337) mit einer multivariaten Analyse des ALLBUS-West 1986 – 1990, dass dieser Bias im Kern ein Bildungsbias ist. In ihrer Interpretation ist er „theoretisch als Effekt kognitiver Ressourcen zu betrachten. Die partielle Verzerrung in der Variablen Stellung im Beruf lässt sich nach Dichotomisierung in erwerbstätig/nicht erwerbstätig als Effekt der Erreichbarkeit verstehen.“ Die Autoren benennen in ihrer Replik (Hartmann/Schimpl-Neimanns 1993, S. 364) selbst die Nachteile dieser indirekten Schätzung der Eigenschaften nichtteilnehmender Personen. „Insbesondere können nur solche potentiellen Ursachen des Nonresponse ermittelt werden, die sich in den groben Kategorien der amtlichen Statistik ausdrücken lassen.“

3 Wen haben wir eigentlich befragt?

Weil die Forschungsgruppe Wahlen (FGW) nicht die Standarddemographie (vgl. Statistisches Bundesamt 1988 und 1995) benutzt, kann eine vergleichbare Analyse ihrer Umfragen nicht durchgeführt werden. Trotzdem ist eine Gegenüberstellung der Umfrageergebnisse vor und nach dem Wechsel aufschlussreich für die simple Nachfrage, ob die beiden Erhebungsmethoden dieselben sozialen Gruppen erreichen. Die Zusammensetzung der tatsächlich Befragten kann aber die Effekte eventuell unterschiedlicher Erreichbarkeit und Verweigerung nicht trennen.

Die Face-to-Face-Daten wurden vom Marplan-Institut nach dem ADM-Design im klassischen Random Route monatlich in nur acht Tagen Feldzeit erhoben. Die CATI-Stichproben basieren auf den Telefonbüchern und nutzen nur die mit dem RLD-Verfahren generierten Telefonnummern. Regional nach Stadt/Land geschichtet, wurde disproportional nach dem Anteil der Wahlberechtigten ausgewählt, um Ungleichheiten der Telefondichte und Stadt-Land-Differenzen der durchschnittlichen Haushaltsgröße aufzufangen. Die Feldzeit betrug anfangs sechs Tage, wurde aber kurz danach auf vier Tage verkürzt.

In den Tabellen 2 und 3 (siehe S. 48 und 49) werden die ungewichteten Ergebnisse der FGW-Umfragen für den monatlichen ZDF-Politbarometer kumuliert für die Zeit vor/nach der Umstellung der Erhebungsmethode jeweils für August – Dezember 1987 und 1988 den An-

gaben von Hartmann/Schimpl-Neimanns (1992) zum ALLBUS 1988 und zum Mikrozensus 1987 gegenübergestellt, obwohl die FGW-Umfragen West-Berlin nicht einbeziehen und obwohl die Erhebungskategorien nicht ganz identisch sind. Die Untergruppen der „Berufe“ entsprechen weitgehend den ALLBUS-Tätigkeitsmerkmalen; sie konnten deshalb durch eigene Auswertungen des Datenfiles ergänzt werden. Schwierigkeiten bereiten die Kategorien der Schulbildung, die keineswegs der Beschriftung in der Publikation entsprechen: Die Kategorie „Hauptschule mit Lehre“ enthält alle beruflichen Abschlüsse inklusive einer Anlernzeit; „Hauptschule ohne Lehre“ enthält als Restkategorie auch die über 18-jährigen Schüler einer Oberstufe. Das mag die Vergleichbarkeit mit dem Mikrozensus erhöhen, erleichtert aber nicht die theoretische Interpretation als unterschiedliche kognitive Kompetenz.

Trotz der eingeschränkten Vergleichbarkeit scheinen die Face-to-Face-Daten der FGW dem Mikrozensus näher zu sein als der ALLBUS. Auffallend geringer ist der Bildungsbias im Schulabschluss Abitur/Fachhochschulreife, dem die Autoren in ihrer Erklärung den höchsten Stellenwert beimessen. Das gibt Anlass zur Vermutung, dass im Bemühen um eine hohe Ausschöpfung mit erheblich längerer Feldzeit eher die qua Bildung leichter zu Befragenden gewonnen wurden, ohne die angestrebte „Repräsentativität“ verbessern zu können.

Die Hartmann-Befunde können im folgenden als Kriterium dienen, die realisierten CATI-Umfragen im Vergleich mit den vorhergehenden Face-to-Face-Erhebungen zu beurteilen. Deutliche Differenzen gibt es beim Bildungsgrad: Personen mit Hauptschulabschluss sind, vorwiegend bei den über 35-Jährigen, in der CATI-Erhebung seltener, Personen mit Abitur in allen Altersgruppen häufiger befragt worden. Der von Hartmann diagnostizierte Bildungsbias wirkt sich also erst in der telefonischen Befragung aus: „Begünstigt“ durch die extrem kurze Feldzeit werden Gruppen mit höherer Teilnahmebereitschaft überproportional befragt.

Die Häufigkeiten der Berufsgruppen scheinen den Erwartungen zu entsprechen: Arbeiter (ungelernte und Facharbeiter) sind wesentlich schwächer, höhere Angestellte und Beamte entsprechend stärker vertreten. Doch die Interpretation als nach der Bildung verbleibende Effekt der Erreichbarkeit hilft hier nicht weiter: Erwerbstätige waren offensichtlich leichter zu befragen als Nichterwerbstätige. Der scheinbare Widerspruch beruht auf einem unerwartet hohen Defizit bei den über 60-Jährigen: In der telefonischen Befragung sind Einpersonenhaushalte mit 23 % deutlich schwächer vertreten als bei den Face-to-Face-Interviews mit 31 %. Da Alter damit (nicht linear) korreliert, wirkt sich das auf den Anteil der Älteren aus und trägt, historisch bedingt, mit zu den geringeren Häufigkeiten der Hauptschulbildung und der Arbeiterberufe bei.

Tabelle 2: Vergleich zwischen FGW-Umfragen, ALLBUS und Mikrozensus – West
%

	FGW ¹⁾ CATI 1988	FGW ¹⁾ FACE TO FACE 1987	ZUMA ALLBUS 1988	MIKRO- ZENSUS 1987
1-Personen-Haushalte	22,9	31,3	27,8	38,6
Geschlecht				
Männlich	49,2	47,8	44,4	46,9
Weiblich	50,8	52,2	55,6	53,1
Alter				
bis 34 Jahre	39,1	27,7	35,5	–
35 – 59 Jahre	42,2	42,4	38,1	–
60 Jahre und älter	18,7	29,9	26,4	–
Männer: 60 J. u. älter ..	7,4	12,2	10,7	–
Frauen: 60 J. u. älter ..	11,3	17,7	15,7	–
Schulbildung				
Hauptschule o. Lehre	13,1	19,9	19,8	29,2
Hauptschule m. Lehre	35,3	44,7	37,3	37,2
Mittlere Reife	28,6	24,4	22,9	19,3
Abitur, FH-Reife	22,8	11,0	20,0	14,3
Alter in Jahren und Schulbildung				
bis 34, Hauptschule	13,6	11,2	13,9	–
bis 34, Mittlere Reife	13,1	10,3	10,7	–
bis 34, Abitur	12,3	6,2	10,9	–
ab 35, Hauptschule	34,8	53,4	43,3	–
ab 35, Mittlere Reife	15,5	14,0	12,2	–
ab 35, Abitur	10,5	4,9	9,0	–

1) Ohne Angaben für Berlin-West.

Tabelle 3: Vergleich zwischen FGW-Umfragen, ALLBUS und Mikrozensus – West
 %

Stellung im Beruf	FGW ¹⁾ CATI 1988	FGW ¹⁾ FACE TO FACE 1987	ZUMA ALLBUS 1988	MIKRO- ZENSUS 1987
Beruf heute oder früher				
Arbeiter	20,3	31,6	27,3	–
Angestellte, Beamte	48,1	41,7	48,8	–
einfache und mittlere	37,7	37,6	46,3	–
leitende und höhere	10,4	4,1	2,5	–
Selbständige	9,7	7,3	7,4	–
nie tätig, keine Angaben	21,9	19,4	16,5	–
Erwerbstätige				
Arbeiter	24,7	32,8	28,9	35,9
un- und angelernt	5,9	6,6	9,8	–
Facharbeiter	18,8	26,1	19,1	–
Angestellte, Beamte	61,6	56,3	59,4	52,5
einfache und mittlere	48,4	50,5	55,7	–
leitende und höhere	13,2	5,8	3,7	–
Selbständige	13,5	10,4	11,7	11,6
Stellung im Beruf				
Arbeiter	14,7	15,9	12,8	18,3
un- und angelernt	3,5	3,2	4,3	–
Facharbeiter	11,2	12,7	8,5	–
Angestellte, Beamte	36,8	27,3	26,4	26,8
einfache und mittlere	28,9	24,5	24,7	–
leitende und höhere	7,9	2,8	1,7	–
Selbständige	8,1	5,1	5,2	5,9
Nicht erwerbstätig	40,3	51,5	55,6	49,0
Erwerbstätige	59,7	48,5	44,4	51,0

1) Ohne Angaben für Berlin-West.

Matthias Jung kam 1990 mit den etwas umfangreicheren Daten auch aus der Experimentierphase bereits zu dem gleichen Ergebnis: In Telefoninterviews sind „formal Höhergebildete und Befragte mit einem höheren beruflichen Status besser (zu) erreichen, während Personen mit Hauptschulabschluss und (damit) auch Arbeiter in den Telefonumfragen in weitaus ge-

ringern Umfang vertreten sind“. Dafür gibt Jung (1990, S. 397) zwei Gründe an. Für die Stichprobenziehung vermutet er „immer noch schichtabhängige Unterschiede bei der telefonischen Versorgung (1988)“; für deren Realisierung und Ausschöpfung wird vermutet, „dass formal Höhergebildete und Berufstätige aus dem Verwaltungsbereich mit dem Medium Telefon vertrauter sind“. Theoretisch unbefriedigend, müsste man auch von den Jüngeren annehmen, dass sie mit dem Medium Telefon vertrauter sind.

1988 waren zwar 93,2 % aller deutschen Privathaushalte mit Telefon ausgestattet, doch die Erreichbarkeit der Einpersonenhaushalte lag mit 86,8 % noch deutlich unter dem Gesamtwert. Mit geringerem Abstand dazu waren auch Arbeiterhaushalte und die der Nichterwerbstätigen schwerer zu erreichen. Das kann jedoch rein quantitativ betrachtet nicht die primäre Ursache für die Unterrepräsentierung dieser Gruppen unter den Befragten sein. Vielmehr muss man davon ausgehen, dass der für die Face-to-Face-Erhebungen von Hartmann belegte Bias zulasten der unteren sozialen Schichten durch Telefonumfragen, vorwiegend wegen der sehr kurzen Feldzeit, etwas verstärkt wurde. Diese Interpretation steht im Einklang mit den Überlegungen von Rainer Schnell (1997). Im Bezugsrahmen einer Rational-Choice-Theorie ist zu vermuten, dass telefonische Befragungen bei Erwerbstätigen, insbesondere bei den höheren Berufen, geringere Opportunitätskosten der zeitlichen Beanspruchung zu überwinden haben als Face-to-Face-Befragungen; gleichzeitig sinken am Telefon für Verweigerung aber auch die Interaktionskosten bei geringen kognitiven Ressourcen. Nur in diesem Sinne könnte man von einem „Medien“-Effekt der Telefon-Umfragen sprechen. Der folgende Vergleich mit Telefonumfragen in Ostdeutschland wird zeigen, dass das hier beschriebene Ausmaß und Muster solcher Differenzen (noch) nicht generalisiert werden darf.

4 Telefonumfragen in Ostdeutschland

Die alte Binsenwahrheit, sich vor allen Stichprobenplänen mit der Struktur des Telefonnetzes vertraut machen zu müssen, war anfangs der neunziger Jahre für Umfragen in Ostdeutschland eine wieder sehr aktuelle Notwendigkeit. Die Telefondichte erreichte 1993 kaum das West-Niveau von 1973 (vgl. Tabelle 1). Nach dem sozioökonomischen Status der Erwerbstätigen gab es dieselben Unterschiede wie damals in den alten Bundesländern. Das Telefonnetz wurde zwar in beeindruckendem Tempo ausgebaut, aber Anfang 1994 hatte die Telefondichte erst die 50 %-Marke erreicht.

Die regionalen Unterschiede in den Ländern und Ortsgrößenklassen der Gemeinden waren sehr ausgeprägt, denn in der ersten Phase des Netzausbaus wurden vorrangig gewerbliche Anschlüsse in größeren Städten bedient. 1993 wird das allgemeine Telefonnetz ausgeweitet, im Norden schneller als im Süden. Der flächendeckende Ausbau begann erst 1995. Bis dahin waren telefonische „Allgemeine Bevölkerungsumfragen“ mit dem Anspruch der Repräsentativität ausgeschlossen. Nach unbefriedigenden Experimenten im Frühjahr 1994 wurde der Politbarometer Ost erst im September 1995 auf telefonische Befragung mit anfangs sieben Tagen Feldzeit umgestellt. Der Stichprobenplan musste „nur“ sicherstellen, dass die regionalen Unterschiede und das immer noch vorhandene Stadt-/Land-Gefälle berücksichtigt wurden. Bis Sommer 1995 gaben in Umfragen 22 % der Telefonbesitzer an, nicht bzw. noch nicht im Telefonbuch eingetragen zu sein; ein RLD-Verfahren war also unverzichtbar.

Die Face-to-Face-Daten wurden vom Usuma-Institut Berlin nach einer von der FGW erstellten Flächenstichprobe mit klassischer Random Route monatlich in nur zwölf Tagen Feldzeit erhoben.

In den Tabellen 4 und 5 (siehe S. 52 und 53) werden die ungewichteten Ergebnisse der FGW-Umfragen für den monatlichen ZDF-Politbarometer Ost kumuliert für die Zeit vor/nach der Umstellung der Erhebungsmethode für Januar – Juli und September – Dezember 1995 dem Mikrozensus 1993 (Koch 1998) gegenübergestellt, ohne eine direkte Vergleichbarkeit zu beanspruchen. Der ALLBUS 1994 wurde nicht herangezogen, weil er auf einer Personstichprobe aus den Einwohnermelderegistern basiert und deshalb mit dem ADM-Design einer Haushaltsstichprobe nicht vergleichbar wäre. Stattdessen werden Daten des ZUMA-Wohlfahrts-Surveys 1993 mitgeteilt, der ein anderes Muster als ältere Umfragen von ZUMA aufweist: Abiturienten sind entgegen den bekannten Erfahrungen kaum überrepräsentiert, und die über 60-Jährigen sind nicht unterrepräsentiert. In erster Linie dienen diese Daten dazu abzuklären, dass die FGW-Umfragen keine unüblichen Verzerrungen aufweisen.

Die unterschiedliche Zusammensetzung der tatsächlich Befragten lässt sich wieder nicht den interessierenden Komponenten zurechnen: im Saldo sind unterschiedliche Erreichbarkeit und Verweigerung bereits verrechnet. Also kann wieder „nur“ die Frage beantwortet werden, wer in beiden Erhebungen tatsächlich befragt wurde.

Das Muster der Differenzen ist erstaunlich anders als bei der Umstellung 1987/88 im Westen. Zwar sind auch hier die über 60-Jährigen wieder geringer in den Telefonumfragen vertreten, aber die Differenz ist diesmal wesentlich geringer und betrifft anders als zuvor nur die älteren Männer. Eine geringere Erreichbarkeit der Einpersonenhaushalte kann anders als im Westen 1987/88 nicht die Hauptursache sein.

Unerwartet sind in den CATI-Umfragen Befragte mit Abitur nicht stärker vertreten. In beiden Erhebungen liegt der Anteil über dem fraglichen Wert des Mikrozensus, der bekanntlich ohne Auskunftspflicht zu den Bildungsfragen die niedrigen Schulabschlüsse unter- und die höheren daher überschätzt (Riede/Emmerling 1994).

**Tabelle 4: Vergleich zwischen FGW-Umfragen, ZUMA-Wohlfahrtssurvey
und Mikrozensus – Ost**

%

	FGW CATI 1995	FGW FACE TO FACE 1995	ZUMA W. SURVEY 1993	MIKRO- ZENSUS 1993
1-Personen-Haushalte	17,7	19,7	17,4	–
Geschlecht				
Männlich	45,6	48,6	48,2	46,7
Weiblich	54,4	51,4	51,8	53,3
Alter				
18 – 29 Jahre	14,5	15,8	12,6	19,7
30 – 39 Jahre	24,3	19,1	21,2	19,9
40 – 49 Jahre	18,9	17,7	15,9	15,9
50 – 59 Jahre	18,2	18,4	21,3	19,9
60 Jahre und älter	23,9	29,0	29,0	25,1
Männer: 60 J. u. älter ..	9,2	14,3	13,8	–
Frauen: 60 J. u. älter ..	14,7	14,7	15,3	–
Schulbildung				
Hauptschule	32,0	37,6	40,1	40,1
ohne Lehre	7,2	7,6	8,4	–
mit Lehre	24,8	30,0	31,7	–
Mittelschule, POS	47,1	41,6	43,2	45,9
Abitur, FH-Reife	20,8	20,8	16,7	14,0
Alter in Jahren und Schulbildung				
bis 34, Hauptschule	2,4	2,1	2,6	–
bis 34, Mittelschule	19,1	16,6	16,3	–
bis 34, Abitur	5,8	6,5	4,0	–
ab 35, Hauptschule	29,7	35,5	37,5	–
ab 35, Mittelschule	27,9	24,9	26,9	–
ab 35, Abitur	15,0	14,4	12,7	–

Tabelle 5: Vergleich zwischen FGW-Umfragen, ZUMA-Wohlfahrtssurvey und Mikrozensus – Ost

%

Stellung im Beruf	FGW CATI 1995	FGW FACE TO FACE 1995	ZUMA W. SURVEY 1993	MIKRO- ZENSUS 1993
Beruf heute oder früher				
Arbeiter	38,7	45,2	42,6	–
Angestellte, Beamte	47,1	43,2	49,6	–
einfache und mittlere	36,2	35,8	46,0	–
leitende und höhere	10,9	7,4	3,6	–
Selbständige	6,6	4,4	5,1	–
nie tätig, keine Angaben	7,6	7,2	2,5	–
Erwerbstätige				
Arbeiter	37,8	40,8	36,4	41,1
Angestellte, Beamte	52,0	51,0	54,9	51,9
einfache und mittlere	42,0	45,0	51,5	–
leitende und höhere	10,0	6,0	3,4	–
Selbständige	10,2	8,2	8,7	6,7
Stellung im Beruf				
Arbeiter	21,6	18,2	17,1	21,9
Angestellte, Beamte	29,5	22,6	25,7	27,5
einfache und mittlere	23,9	20,0	24,1	–
leitende und höhere	5,6	2,6	1,6	–
Selbständige	5,7	3,7	4,1	3,5
Nicht erwerbstätig	43,2	55,5	53,1	47,1
Erwerbstätige	56,8	44,5	46,9	52,9

Berufstätige (vgl. Tabelle 5) sind mit CATI auch wieder häufiger befragt worden, aber der Anteil liegt nur wenig über dem Sollwert aus dem Jahr 1993. Die Arbeiter sind – mit geringen Differenzen als früher – in CATI wieder weniger, die höher qualifizierten Angestellten und Beamten sowie die Selbständigen entsprechend häufiger vertreten.

Gemessen an den Hartmann-Befunden scheinen die Differenzen in Ostdeutschland also weniger ausgeprägt zu sein. Selbstverständlich erfassen die einfachen Klassifikationen nach Alter, Geschlecht, Schulbildung, Erwerbstätigkeit und Beruf nur unzureichend die sozioökono-

misch und kulturell bestimmte Lebenslage, die in der theoretischen Erklärung des Bias gemeint ist. Ein zusätzliches Indiz für die Richtigkeit der Interpretation ist jedoch die subjektive Einschätzung der eigenen wirtschaftlichen Lage der Befragten in Telefonumfragen: Sie „repräsentieren“ in Ostdeutschland 1995 in fast allen Kategorien die sozioökonomisch Bessergestellten stärker als die Face-to-Face-Umfragen. Dieser Effekt spiegelt die damals noch bestehenden Ungleichheiten in der Telefonversorgung und hat mit Erreichen der „Vollversorgung“ inzwischen wieder abgenommen.

Was könnte man aus den Erfahrungen der FGW beim Wechsel zu CATI-Umfragen heute noch lernen? Bei beiden Umstellungen war ja die Infrastruktur des Telefonnetzes noch kein optimaler Kontext für diese Erhebungen. Bei einem heutigen Wechsel wären deshalb geringere Effekte zu erwarten, zumal auch die Gewöhnung an und die Vertrautheit mit dem Medium Telefon immer weiter zunehmen. Erreichbarkeit ist bei Personen mit gleichem Lebensstil und gleicher Lebensphase eine ziemlich „stabile“ Variable, so dass der Erfolg der Erhebung eher von der üblicherweise kurzen Feldzeit beeinflusst wird. Dagegen ist das Ausmaß von Verweigerungen stärker situationsabhängig: In der vergleichsweise anonymen Beziehung zum Befragten wird die kritische Phase des Kontaktgesprächs für den Interviewer schwieriger als in der Face-to-Face-Situation. Intensive Schulung und Überwachung konnten deshalb die Qualität der Telefonumfragen in den letzten Jahren verbessern. Es gibt aber keinen Grund, sich weiterhin auf den komfortablen Befund von de Leeuw und van der Zouwen (1988) zu berufen, die damals nur „small mode differences“ zwischen beiden Erhebungsmethoden feststellten.

Literaturhinweise

Behrens, Kurt (1994): Schichtung und Gewichtung – Verbesserung der regionalen Repräsentanz, in: Siegfried Gabler; Jürgen H. P. Hoffmeyer-Zlotnik; Dagmar Krebs (Hrsg.): Gewichtung in der Umfragepraxis, Opladen, S. 27 – 41.

De Leeuw, Edith D.; van der Zouwen, Johannes (1988): Data Quality in Telephone and Face-to-Face Surveys, in: Robert M. Groves et al. (Hrsg.): Telephone Survey Methodology, New York, S. 283 – 299.

Follmer, Robert; Smid, Menno (1998): Nichteingetragene Telefonnummern. Ergebnisse eines Methodentests, in: Siegfried Gabler; Sabine Häder; Jürgen H. P. Hoffmeyer-Zlotnik (Hrsg.): Telefonstichproben in Deutschland, Opladen, S. 43 – 57.

Gabler, Siegfried; Häder, Sabine (1997): Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland, in: ZUMA-Nachrichten 41, S. 7 – 18.

Gabler, Siegfried; Häder, Sabine (1998): Probleme bei der Anwendung von RLD-Verfahren, in: Siegfried Gabler; Sabine Häder; Jürgen H. P. Hoffmeyer-Zlotnik (Hrsg.): Telefonstichproben in Deutschland, Opladen, S. 58 – 68.

Häder, Sabine (1994): Auswahlverfahren bei Telefonumfragen, ZUMA-Arbeitsbericht Nr. 94/03.

Häder, Sabine (1996): Wer sind die „Nonpubs“? Zum Problem anonymer Anschlüsse bei Telefonumfragen, in: ZUMA-Nachrichten 39, S. 45 – 68.

Häder, Sabine; Gabler, Siegfried (1998): Ein neues Stichprobendesign für telefonische Umfragen in Deutschland, in: Siegfried Gabler; Sabine Häder; Jürgen H. P. Hoffmeyer-Zlotnik (Hrsg.): Telefonstichproben in Deutschland, Opladen, S. 69 – 88.

Hartmann, Peter H. (1990): Wie repräsentativ sind Bevölkerungsumfragen? Ein Vergleich des ALLBUS und des Mikrozensus, in: ZUMA-Nachrichten 26, S. 7 – 30.

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

Hartmann, Peter H.; Schimpl-Neimanns, Bernhard (1993): Affirmative Repräsentativitäts-„Beweise“ oder Test konkreter Hypothesen zu Verteilungsabweichungen?, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie 45, S. 359 – 365.

von der Heyde, Christian (1998): InfraScope – Der Telefonbus von Infratest Burke, in: Siegfried Gabler; Sabine Häder; Jürgen H. P. Hoffmeyer-Zlotnik (Hrsg.): Telefonstichproben in Deutschland, Opladen, S. 133 – 140.

Jung, Matthias (1990): Ausschöpfungsprobleme bei repräsentativen Telefonumfragen, in: Forschungsgruppe Telefonkommunikation (Hrsg.): Telefon und Gesellschaft, Band 2, Berlin, S. 386 – 399.

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

Rendtel, Ulrich; Pötter, Ulrich (1993): „Empirie“ ohne Daten. Kritische Anmerkungen zu einer Repräsentativitätsstudie über den Allbus, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie 45, S. 350 – 358.

Riede, Thomas; Emmerling, Dieter (1994): Analysen zur Freiwilligkeit der Auskunftserteilung im Mikrozensus, in: Wirtschaft und Statistik, 9, S. 733 – 742.

Rösch, Günther (1998): Maßnahmen zur Reduktion von Stichprobenfehlern bei telefonischen Umfragen, in: Siegfried Gabler; Sabine Häder; Jürgen H. P. Hoffmeyer-Zlotnik (Hrsg.): Telefonstichproben in Deutschland, Opladen, S. 133 – 140.

Schnell, Rainer (1997): Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen, Opladen.

Schulte, Wolfgang (1997): Telefon- und Face-to-face-Umfragen und ihre Stichproben, in: Siegfried Gabler; Jürgen H. P. Hoffmeyer-Zlotnik (Hrsg.): Stichproben in der Umfragepraxis, Opladen, S. 196 – 206.

Statistisches Bundesamt (Hrsg., 1988): Merkmale einer allgemeinen Standarddemographie. Gegenüberstellung soziodemographischer Variablen aus dem Mikrozensus, der Einkommens- und Verbrauchsstichprobe, der Volkszählung und der Standarddemographie des Zentrums für Umfragen, Methoden und Analysen, Heft 4 der Schriftenreihe Ausgewählte Arbeitsunterlagen zur Bundesstatistik, Wiesbaden.

Statistisches Bundesamt (Hrsg., 1995): Demographische Standards, Ausgabe 1995, Methoden – Verfahren – Entwicklungen. Materialien und Berichte, Wiesbaden.

Wedel, Edgar (1989): Haushalte 1987 – Methode und Ergebnis der Volkszählung, in: *Wirtschaft und Statistik*, 5, S. 273 – 276.

Methodeneffekte beim Umstieg auf CAPI-Techniken

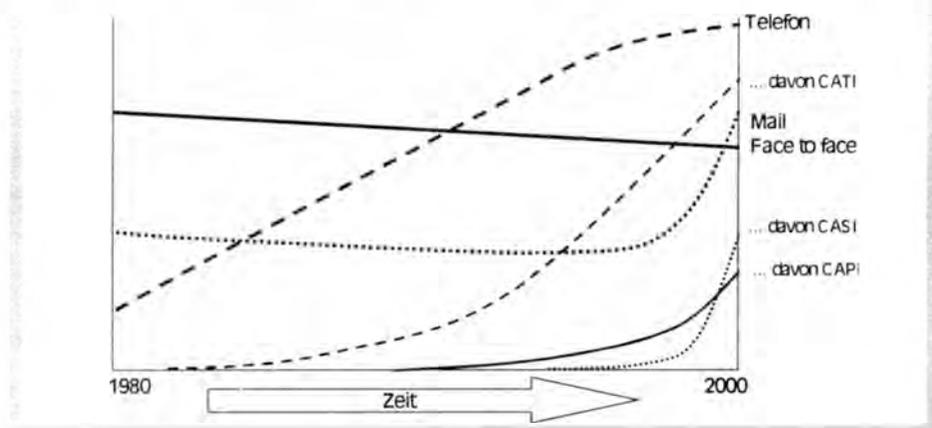
In der gestrigen Ausgabe der Süddeutschen Zeitung (vom 23.6.1999) leitet ein junger koso-
varischer Schriftsteller seinen Beitrag zur Frage „Können die Albaner verzeihen und verges-
sen?“ mit folgendem Satz ein: „Schon seit ein paar Jahren liegt Kafkas Brief an seinen Vater
auf dem Desktop meines Computers.“ Nicht der Schreibtisch, nicht das Bücherregal, nein,
der Desktop ist das gewöhnlichste Möbel, das hier als Ablagemöglichkeit zur Verfügung
steht. Ich will damit schon gleich zu Beginn darauf hinweisen – und damit in gewisser Weise
auch den vorgegebenen Titel dieses Referates relativieren –, dass sich die Zeiten geändert
haben, und dass immer auch bedacht werden muss, dass es möglicherweise auch dadurch
Methodeneffekte in der empirischen Sozialforschung geben kann, dass man nicht mit wich-
tigen gesellschaftlichen und kulturellen Änderungen Schritt hält.

Das Thema ist der Einzug des Computers in die Markt- und Meinungsforschung, und zwar
jetzt nicht mehr nur bei der Erfassung und Auswertung der Daten, sondern bei ihrer Erhebung
selber. Das ist noch keine besonders lange andauernde Entwicklung, aber eine, die gegen-
wärtig die Methodenlandschaft dramatisch verändert, ja revolutioniert.

Abbildung 1

Infratest Burke InCom

Trends zu Anzahl der Interviews nach Erhebungsmethoden in der empirischen Markt-, Meinungs- und Sozialforschung in der BRD



Die Anzahl klassischer Face-to-Face-Interviews, die in der Marktforschung geführt werden,
nimmt seit Jahren allmählich ab. Zu Beginn der neunziger Jahre hat die Zahl der Telefonin-
terviews die Anzahl der Face-to-Face-Interviews insgesamt wohl überstiegen – und das, ob-

*) Rolf Pfeleiderer, Infratest Burke, München.

wohl Telefoninterviews in der Markt- und Meinungsforschung erst seit Anfang der achtziger Jahre aufgrund der erst dann ausreichenden Verbreitung in den bundesrepublikanischen Haushalten für Forschungszwecke und insbesondere auch Repräsentativerhebungen akzeptabel wurden. Die Zahl der schriftlichen Interviews in der Markt-, Meinungs- und Sozialforschung ist ebenfalls längere Zeit eher zurückgegangen – die sehr idealtypische Darstellung auf der Grafik berücksichtigt dabei die übliche Oszillation bei solchen Entwicklungen nicht, z. B. auch nicht den von mir vermuteten Anstieg der schriftlichen Interviews bereits in den letzten zwei Jahren.

Am „Aufstieg“ der Telefoninterviews ist ganz zweifellos die Computerunterstützung (CATI) maßgeblich beteiligt. Bereits heute wird der größte Teil der Telefoninterviews computergestützt durchgeführt. Auch im Bereich der Face-to-Face-Interviews wird ein immer größerer Anteil computergestützt erhoben, in unserem Institut sind es beispielsweise schon ca. 70 % dieser Interviews. Und schließlich erleben wir gerade eben, wie die schriftlichen Interviews durch ihre computergestützte Form – Computer Self Administered Questionnaire (CSAQ) oder auch Computer Assisted Self Administered Interviewing (CASI) – in Form von Web-Interviews und E-Mail-Interviews eine Renaissance erleben. Es deutet sich sogar an, dass diese Interviewform, zumindest von der Zahl der Interviews in der Markt- und Sozialforschung, demnächst die vorherrschende Erhebungsart sein wird.

Es ist offensichtlich, dass das mit Hilfe eines gedruckten Fragebogens und eines Bleistifts oder Kugelschreibers durchgeführte Face-to-Face-Interview nicht mehr der „Königsweg“ der empirischen Markt- und Sozialforschung ist. Zumindest nicht mehr in quantitativer Hinsicht – aber vielleicht doch noch in qualitativer Hinsicht?

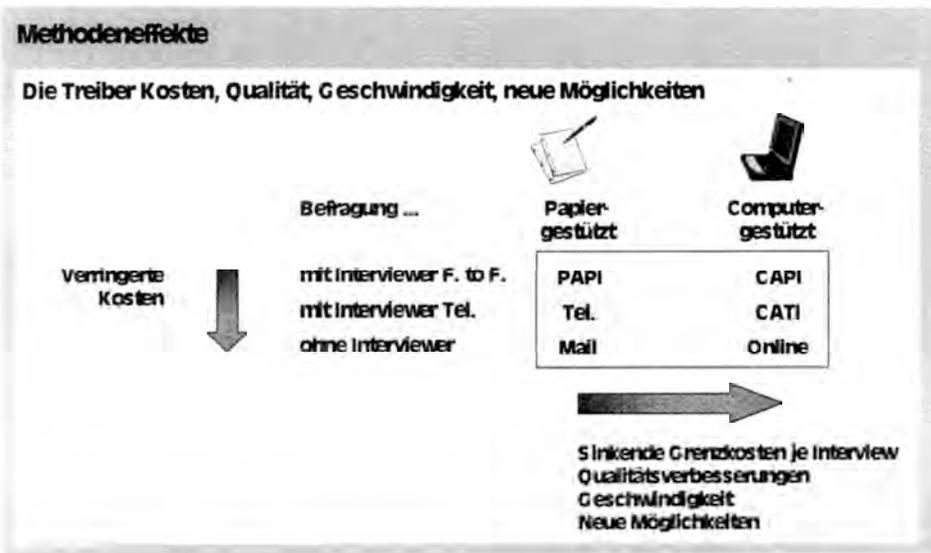
Um die Diskussion von Methodeneffekten durch den Einsatz von Interviewcomputern einordnen zu können, muss man sich vergegenwärtigen, was eigentlich die beschriebene Entwicklung vorangetrieben hat. Man kann da vor allem vier „Treiber“ identifizieren:

1. Kostengesichtspunkte;
2. Qualitätsgesichtspunkte;
3. Geschwindigkeitsvorteile und
4. Neue Erhebungsmöglichkeiten durch Computerunterstützung.

In einer vereinfachenden Weise lassen sich diese „Treiber“ mit dem Einsatz von Interviewcomputern so in Verbindung bringen, wie in der nachfolgenden Abbildung 2 (siehe S. 59) dargestellt.

Abbildung 2

Infratest Burke InCom



Das Face-to-Face-Interview war schon immer besonders teuer nicht nur wegen des teuren Einsatzes lebendiger Arbeitskraft, sondern auch wegen der üblicherweise anfallenden hohen Zeit- und Fahrtspesen. Zumindest letzteres kann beim Telefoninterview vermieden werden. Ganz generell wird außerdem erwartet, dass – zumindest bei Umfragen mit einer großen Anzahl Probanden – deutlich sinkende Grenzkosten je Interview zu besserer Ökonomie durch CAPI führen. Und das – so die Erwartung – soll gleichzeitig mit Qualitätsverbesserungen, mit erhöhter Geschwindigkeit zumindest bei der Feldarbeit empirischer Sozialforschung und mit überhaupt neuen Interviewmöglichkeiten verbunden sein. Ich möchte diese Erwartungen im Folgenden etwas genauer ansehen, wobei ich versuchen werde, soweit wie möglich „empirisch“ zu bleiben oder zumindest auf eigene Erfahrungen Bezug zu nehmen.

Der erste Treiber: Kosten

Man könnte zunächst einmal fragen, was denn Kosten überhaupt mit Methodeneffekten zu tun haben? Es wäre realitätsfern, wenn wir nicht einräumen würden, dass die Frage nach der „richtigen“, nach der je Problemstellung besten Methode, also die Frage nach der Qualität empirischer Sozialforschung, tatsächlich außerordentlich eng mit der Frage der Kosten von empirischer Sozialforschung verknüpft ist. Man kann sich das leicht klarmachen, indem man einmal die wesentlichen Qualitätskriterien bei der Erhebung demoskopischer Daten betrachtet:

Die Entwicklung des Fragebogens, also das Messkonzept

Einen guten, möglicherweise gar optimalen Fragebogen zu entwickeln – ganz gleich, ob für ein Face-to-Face-Interview, eine telefonische Befragung oder eine schriftliche bzw. Online-

Befragung – ist zeitaufwendig, kostet viele teure Forscherstunden. Und in der Regel sind diese Forscherstunden um so teurer, je erfahrener der Forscher in der Entwicklung von Fragebögen ist. Und auch der Handlingtest, der Pretest, der Feldtest eines Fragebogens kosten Geld. Je sorgfältiger gearbeitet wird, desto mehr Geld kostet es (unrühmliche Ausnahmen bestätigen diese Regel). Einsparmöglichkeiten ergeben sich vor allem dadurch, dass man auf schon bewährte Formulierungen, Fragenkomplexe usw. zurückgreift, also durch Standardisierung.

Die Bruttostichprobe, also die Qualität der Abbildung der jeweiligen Grundgesamtheit

Stichproben auf der Basis der Daten von Einwohnermeldeämtern sind aus verschiedenen Gründen inzwischen praktisch unbezahlbar geworden. Bei den in der Marktforschung üblichen „Random-Stichproben“ gibt es eine Reihe von unterschiedlichen Verfahren, die sich vor allem darin unterscheiden, wie sorgfältig versucht wird, wirklich allen Elementen der Grundgesamtheit die gleiche Auswahlchance zu geben. Von der möglichst vollständigen Adressauflistung für Face-to-Face-Untersuchungen über die Erfassung nur jeder x-ten Adresse bis hin zu nur noch schwer nachvollziehbaren „Auswahlprozeduren“ gibt es erhebliche Qualitätsunterschiede. Und für telefonische Interviews muss eben ein zusätzlicher Aufwand getrieben werden, wenn auch die nicht gelisteten Telefonnummern berücksichtigt werden sollen, wenn gleichzeitig die Auswahlwahrscheinlichkeit für Zielpersonen, die über mehrere Telefonnummern zu erreichen sind, gleich gehalten werden soll usw. Auch hier gilt: sorgfältiges Arbeiten kostet – sparen kann man am ehesten mit Hilfe erprobter, kontrollierter und standardisierter Verfahren.

Die Qualität der Nettostichprobe, insbesondere die Vermeidung responsebedingter Selektivität

Es bedarf in diesem Kreis keiner weiteren Erläuterungen, dass die Verbesserung der Ausschöpfung einer Bruttostichprobe grundsätzlich unter Qualitätsgesichtspunkten wünschenswert, aber leider immer mit Zusatzaufwendungen verbunden ist. Gleichzeitig muss man im Auge behalten, dass die Verbesserung der Ausschöpfung der Bruttostichprobe nicht immer eine Verbesserung der Nettostichprobe zur Folge hat: Insbesondere im Bereich telefonischer Umfragen gibt es Beispiele, dass die Verbesserung der Ausschöpfung nur bei ganz bestimmten Teilgruppen der Grundgesamtheit gelingt und damit gegebenenfalls die zunächst einmal (auf niedrigem Niveau) vorhandene Balance der Stichprobe im Hinblick auf die Abbildung wichtiger Merkmale der Grundgesamtheit eher beschädigen als stärken kann. Auch die Berücksichtigung solcher Effekte ist in der Regel nicht kostenfrei zu haben.

Die korrekte und vollständige Erfassung der Daten sowie ihre Konsistenz

Auch hier bedarf es keiner weiteren Erläuterungen, dass, insbesondere bei komplizierten Erhebungsinhalten, verbesserte Qualität direkt mit einem höheren Programmier- und Kontrollaufwand verbunden ist, und damit mit höheren Kosten. Auch hier liegt das „Heil“ vor allem in der Standardisierung von hochentwickelten Prüf- und Erfassungsroutinen – und natürlich vor allem in der Schulung der Interviewer. Keine Frage, dass Qualität hier Geld kostet.

Die Hoffnung beim Einsatz des Computers für Face-to-Face-Erhebungen ist also, den Grad der Standardisierung zu erhöhen, der Maschine Arbeiten zu übertragen, die vom Interviewer/ Erfasser mit einer größeren Fehlerrate erledigt werden und insgesamt so entweder gleiche Qualität billiger, oder aber verbesserte Qualität zu gleichen Preisen oder vergleichsweise geringen Zusatzkosten anbieten zu können.

Aber was sind nun die direkten Kosteneffekte durch CAPI?

CAPI spart Kosten

- für den Druck von Fragebogen, Vorlagenheften, Kartensätzen, Rückkuverts, usw.;
- für die Datenprüfung nach dem Interview – weil sie bereits weitgehend in den elektronischen Fragebogen eingebaut werden kann;
- für Interviewerführung und Interviewerkontrolle, weil vieles schneller geht, umständliche telefonische Kommunikation entfällt und Interviewerfälschungen noch mehr erschwert werden als ohnehin schon;
- für Porti (wenn nicht besonders umfangreiche „Datenpakete“ in Form von CDs nach wie vor postalisch verschickt werden müssen);
- für manche Arbeitsgänge, die nun elektronisch vermittelt schneller gehen (Zeit ist auch hier Geld).

CAPI kostet auf der anderen Seite mehr, weil

- die Interviewer neu (und zusätzlich) geschult werden müssen;
- mehr programmiert werden muss als vorher, und weil Programmierer teuer sind;
- neue Software bzw. Softwarelizenzen angeschafft werden müssen, und zwar nicht nur Fragebogensoftware, sondern auch Kommunikationssoftware;
- die Interviewer mit Computern ausgerüstet werden müssen, was auch für ein großes Marktforschungsinstitut wie unseres ein ungewöhnlich großes Investment war.

Unsere Erfahrungen nach zwei Jahren „Volllastbetrieb“ mit hunderten von Interviewcomputern zeigen, dass – bei wie wir hoffen, verbesserter bzw. mindestens gleich gebliebener Qualität – alles im allem die Kostenersparnisse überwiegen. Allerdings in einem Ausmaß, das durch verschiedene gesetzgeberische Aktivitäten der Bundesregierung gegebenenfalls weit überkompensiert werden kann.

Der zweite Treiber: Qualitätsverbesserungen

Es gibt eine ganze Reihe von qualitätsrelevanten Prozessen und Arbeitsschritten, für die CAPI ganz unzweifelhaft Verbesserungen bringt:

Im Bereich des *Fragebogens* sind das vor allem folgende Verbesserungen:

- Der Umstand, dass der zeitliche Ablauf des Interviews völlig transparent wird. Die Kenntnis von „Hängern“ in den Pretest-Interviews, der mögliche Vergleich zwischen unterschiedlich dargestellten bzw. auszufüllenden Vorgaben in zeitlicher Hinsicht und die Information über die Gesamtdauer durch die im Computer intern immer mitlaufende Uhr sind eindeutig Vorteile, die zu Verbesserungen im Fragebogen führen können.
- Filterführung, Rotation von Vorgaben, Rückbezüge auf zu Beginn des Interviews bereits gemachte Angaben werden den Interviewern durch die Technik abgenommen und entfallen daher als typische Fehlerquellen.
- Es sind datenbankgestützte Abfragen möglich, in denen dem Interviewer und dem Befragten Hilfen angeboten werden können, die weit über die übliche „Listenvorlage“ hinaus gehen.

- Das Fragebogendesign kann – insbesondere für Selbstausfüller, grundsätzlich aber auch für jeden einzelnen Interviewer – angepasst, „personalisiert“ werden, wodurch gegebenenfalls eine verbesserte Motivation beim Probanden erzielt werden kann. Dieser Aspekt muss allerdings durch einen erheblichen Mehraufwand bei der Fragebogenprogrammierung bezahlt werden.

Im Bereich der *Bruttostichprobe* bringt die Verwendung von Befragungscomputern naturgemäß keine Verbesserungen, da die Bildung der Bruttostichprobe der Befragung vorgelagert ist.

Im Bereich der *Nettostichprobe* bringt CAPI folgende Vorteile:

- Die Kontrolle darüber, wann genau ein Interview durchgeführt wurde, gegebenenfalls unterbrochen wurde und wie lange es insgesamt gedauert hat, ist erstmals überhaupt präzise möglich.
- Die Rücklaufkontrolle kann verbessert werden, weil viele manuelle Arbeiten präziser und schneller durch Programme erledigt werden können, insbesondere die Überprüfung gegebenenfalls vorgegebener Quotenvorgaben und Fallzahlvorgaben pro Interviewer.
- Entsprechend kann sowohl insgesamt als auch im Hinblick auf möglicherweise gewünschte zeitliche Gleichverteilung von Interviews die Aussteuerung der Feldarbeit computergestützt sehr viel aktueller und präziser erfolgen.
- Wegen des „elektronischen Tracking“ der Interviewertätigkeiten sind Teilfälschungen weiter erschwert bis nahezu unmöglich.
- Schließlich beobachten wir insgesamt auch eine verbesserte Motivation bei Interviewern wie Befragten, wenn die Ernsthaftigkeit der Interviewarbeit durch die Hereinnahme eines teuren Hightech-Gerätes unterstrichen wird. Auf diesen Aspekt soll später noch näher eingegangen werden.

Auf dem Gebiet der „*sauberen Daten*“ bietet CAPI folgende Vorteile:

- Es besteht die Möglichkeit, sofort logische und plausible Fehler zu hinterfragen und gegebenenfalls korrigieren zu lassen (programmgesteuert).
- Dadurch, dass Angaben zu offenen Fragen maschinengeschrieben sind entfällt eine erhebliche Fehlerquelle, nämlich die manchmal schwierige Interpretation von Interviewerhandschriften. Multimedia-Laptops, wie sie bei Infratest im Einsatz sind, können darüber hinaus auch dazu genutzt werden, die Antworten auf offene Fragen akustisch einzugeben, was eine Reihe weiterer Möglichkeiten eröffnet.
- Die Vollständigkeit von Befragtenangaben kann durch ein entsprechend geschriebenes Fragenprogramm garantiert werden, weil Auslassungen jeweils zur Wiederholung bzw. zur Bitte um Vervollständigung der Angaben führen.

Es gibt aber gleichzeitig eine Reihe von Aspekten, unter denen es zumindest zweifelhaft ist, ob CAPI da Qualitätsverbesserungen bewirkt.

Im Bereich des *Fragebogens* sind das folgende Punkte:

- Wir wissen nicht wirklich, wie sich die Verwendung von mehr Grafik, gegebenenfalls von Farbe und insbesondere von „personalisierten“ Fragebogenfassungen auswirkt. Solange

wir das nicht wissen, kann man zumindest daran zweifeln, dass es sich hier um Qualitätsverbesserungen handelt.

- Die gängigen Standardfragebogenprogramme für Markt- und Sozialforschung schaffen es nicht, das, was auf dem Bildschirm während des Interviews zu sehen ist, exakt in Form einer Hardcopy zu reproduzieren. Das erschwert bzw. entwertet die Fragebogensdokumentation. Insofern handelt es sich um einen – in Teilen in der Natur des Mediums selber liegenden – Qualitätsverlust gegenüber dem klassischen Face-to-Face-Interview mit Papier und Bleistift.
- CAPI-Fragebögen lassen dem Interviewer keinerlei Flexibilität: Er kann nicht rasch einmal ein paar Fragen vom Ende des Fragebogens vorziehen, um gegebenenfalls eine schwierige (meist, weil langweilende) Befragungssituation zu „retten“. Ein erhöhtes Maß an Standardisierung des Befragungsablaufs ist damit erkauft mit der Unmöglichkeit, durch Umstellungen zu Interviews zu kommen, die sonst abgebrochen würden. Obwohl Interviewer generell gehalten sein sollten, den vorgeschriebenen Frageablauf streng einzuhalten, ist im Einzelfall sehr wohl die Frage erlaubt, ob nicht die Rettung eines Interviews durch Umstellung des Frageablaufs (in einer möglichst nachvollziehbaren Form!) mehr Qualität bedeutet gegenüber der Alternative Interviewabbruch bei Beibehaltung des Standardverlaufs.

Mit Bezug auf die *Nettostichprobe* können folgende kritischen Punkte aufgeführt werden:

- Es bleibt zu klären (siehe dazu auch weiter unten), ob die Einführung des Computers in die Interviewer-Befragten-Interaktion nicht insgesamt auch einen speziellen CAPI-bedingten Sample-Bias erzeugen kann, beispielsweise durch zielgruppenspezifische und/oder interviewerspezifische Reaktanz. Gründe dafür können jeweils sein: Affinität zu bzw. Ablehnung von Computertechnologie, Aktualisierung von Datenschutzbedenken, sonstige psychologische Aspekte der Interviewsituation.
- Es gibt für CAPI-Interviews situative Einschränkungen, die beim Papier-und-Bleistift-Interview so nicht vorhanden waren. Man denke nur an das Treppenhaus-Interview, an die Notwendigkeit, für den Computer einen Aufstellplatz (Tisch) zu finden, meist auch die Notwendigkeit der „Stromversorgung vor Ort“ usw.
- Schließlich sollte nicht unterschätzt werden, dass es auch technische Fehler (z. B. ausnahmsweise Datenverlust bei der Übermittlung der Befragungsdaten ans Institut) und leider manchmal auch Programmierfehler geben kann. Kein Interviewer kann solche Fehler ad hoc korrigieren, vielmehr läuft das Frageprogramm mit Präzision falsch ab – meist für den Interviewer gar nicht erkennbar. Solche Fehler nehmen mit Übung ab, aber vollständig ausgeschlossen werden können sie nie – und manchmal ist es sogar eine richtige Kunst, sie überhaupt im nachhinein zu entdecken!

Im Hinblick auf das Qualitätskriterium „*saubere Daten*“ bringt CAPI zwar zweifellos hauptsächlich Qualitätsverbesserungen, dennoch bleibt ein Problem: Es werden gelegentlich zu saubere Daten produziert. Wenn der Fragebogen so programmiert ist, dass er keine logischen Fehler und Auslassungen duldet, dann sieht der Forscher auch nicht mehr, wo solche Fehler von den Befragten (und gegebenenfalls auch von den Interviewern) gemacht werden. Die Möglichkeit, daraus für weitere Fragebogenentwicklungen zu lernen, entfällt. Man sollte diese verminderte Lernfähigkeit des Systems als Problem erkennen und nicht als Qualitätsfortschritt verkaufen.

Die in der Markt- und Sozialforschung am meisten bis ausschließlich diskutierten Methodeneffekte im Hinblick auf CAPI beziehen sich darauf, ob bei identischer Bruttostichprobe möglicherweise gegenüber Befragungen mit Papier und Bleistift eine andere Nettostichprobe resultiert. Man findet dazu in der Literatur und den einschlägigen Fachzeitschriften ausgesprochen unterschiedliche Aussagen. Zwei davon seien hier stellvertretend zitiert:

„Sozial unsichere und technikferne Personen fühlen sich den Anforderungen der Laptop- und Telefoninterviews weniger gewachsen und verweigern häufig das Interview“, resümiert Helmut Mosmann in „Planung und Analyse“¹⁾ eine PAPI/CAPI/CATI Vergleichsuntersuchung des Instituts INRA.

Im Gegensatz dazu fassen Knobloch/Knobloch in „Rundfunk und Fernsehen“²⁾ ihre Literaturrecherche zu computergestützten Befragungen so zusammen: „Vor dem eigentlichen Interview ... führen computergestützte Verfahren ... nach bisherigen Erkenntnissen nicht zu Interviewverweigerungen.“ Sie führen dazu mindestens drei Belegstellen an.

Uns liegt aus dem Erfahrungsaustausch mit dem englischen Marktforschungsinstitut Taylor Nelson/AGB ein Bericht vor,³⁾ in dem zu den von diesem Institut bei der Einführung von CAPI durchgeführten Experimenten abschließend bemerkt wird: „No consistent differences have been found between the sample profiles of omnibus PAPI and CAPI data.“

Auch unser Institut hat anlässlich der Ausstattung des Interviewerstabs mit Befragungscomputern einen umfangreichen Vergleichstest durchgeführt, in dem bei vergleichbaren Bruttostichproben insbesondere die soziodemographischen Angaben der Befragten in der PAPI-Version mit einer entsprechenden CAPI-Version verglichen wurden. Die Ergebnisse sind in der folgenden Abbildung 3 (siehe S. 65) dargestellt.

1) Mosmann, Helmut: PAPI, CAPI, CATI: Ambivalenter technischer Fortschritt in der Datenerhebung, in: *Planung und Analyse*, 1/1999, S. 50.

2) Knobloch/Knobloch: Computergestützte Befragung: Der Computer in der Funktion von Interviewer und Fragebogen, in: *Rundfunk und Fernsehen*, 1/1999, S. 70.

3) Taylor Nelson/AGB: CAPI data analysis summary report 1996 (unveröffentlicht).

Abbildung 3

Infratest Burke InCom

Methodeneffekte durch CAPI im Hinblick auf die Stichprobe?

Infratest Vergleichstest „Omnibus“ (PAPI)	vs. „Wertesurvey“ (CAPI)	Stat. Bundesamt	
Feldzeit:	04.06.97-25.06.97	27.05.97-14.07.97	
Zielgruppe:	Deutsche ab 14 Jahren (hier ohne 14-17 Jahre)	Deutsche ab 18 J.	Deutsche ab 18 J.
Fallzahl:	2.473	3.089	
Auswahlmethode:	ADM-Netze, Random-Walk Schwedenschlüssel	ADM-Netze, Random-Walk Schwedenschlüssel	
Ergebnisse ungew./transformiert			
Männer	48,6 %	46,4 %	47,4 %
Alter 18 - 29 Jahre	15,8 %*	16,9 %	18,6 %
30 - 59 Jahre	60,1 %*	59,7 %*	53,4 %
60 Jahre und mehr	24,1 %*	23,1 %*	27,7 %
Erwerbstätige	54,3 %	51,9 %	53,8 %
1-Pers.-HH ¹⁾	26,5 %*	26,6 %*	34,8 %

¹⁾ ohne Transformation Haushaltsstichprobe \Leftrightarrow Personensichprobe

* statistisch signifikante Abweichung vom Referenzwert (95 % Sicherheitsniveau)

Um die Ergebnisse auch mit den amtlichen Daten vergleichen zu können, wurde die übliche Transformation der ursprünglich haushaltsrepräsentativ gezogenen Stichprobe in eine Personensichprobe vorgenommen (Korrektur der unterschiedlichen Antreffbarwahrscheinlichkeiten von Personen in den Haushalten je nach Anzahl der Zielpersonen). Das Ergebnis zeigt, dass die beiden Nettostichproben sich am ehesten im Hinblick auf den Anteil befragter Erwerbstätiger unterscheiden, in diesem Merkmal aber beide nicht in statistisch signifikanter Weise von den Referenzdaten des Statistischen Bundesamtes abweichen. Deutliche Abweichungen von den „offiziellen“ Daten gibt es sowohl bei der Altersstruktur als auch insbesondere beim Anteil von Einpersonenhaushalten. Die letztere Abweichung ist „üblich“ in der Umfrageforschung und reflektiert die besondere Schwierigkeit, mobile junge Einpersonenhaushalte mit vertretbarem Aufwand überhaupt zu erreichen, und auf der anderen Seite die Schwierigkeit, bei alten, einzeln lebenden Menschen die Misstruensbarriere zu überwinden. Insgesamt aber kann zumindest im Hinblick auf soziodemographische Merkmale nicht von einem methodeninduzierten statistischen Bias gesprochen werden.

Es könnte natürlich auch sein, dass die unterschiedlichen berichteten Erfahrungen von den Instituten mit der Auswahl der für CAPI eingesetzten Interviewer bzw. mit den unterschiedlichen Schulungsmaßnahmen zusammenhängen, oder aber dass sich Unterschiede nicht bei den soziodemographischen Merkmalen der Stichproben, sondern eher bei auf das Verhalten, und zwar insbesondere auf das Kommunikationsverhalten bezogenen Merkmalen finden lassen. Zumindest für die erste dieser beiden Vermutungen haben wir aus eigenen Befragungen Anhaltspunkte: In unserer jährlich stattfindenden Interviewerbefragung haben Mitte 1996 beispielsweise 37 % der CAPI-Interviewer dem Statement zugestimmt: „CAPI macht es schwieriger, Interviews auf der Basis von Zufallsstichproben zu bekommen.“ Ein Jahr später

waren es nur noch 21 % der CAPI-Interviewer, die sich so äußerten. Das zeigt klar einen Gewöhnungseffekt bei den Interviewern und signalisiert auch, wenn man die nach den Interviewerberichten durch CAPI z. T. deutlich verbesserte Motivation bei den Befragten berücksichtigt, dass CAPI zumindest insgesamt nicht zu einer schlechteren Ausschöpfung führen muss.

Auch im Hinblick auf andere als soziodemographische Merkmale ist aus der Literatur und aus den bereits zitierten Untersuchungen bzw. Übersichten keine klare Systematik der durch die Hereinnahme des Computers in die Interviewsituation möglicherweise bedingten Effekte auf das Antwortverhalten erkennbar. Das mögen einige kurze Zitate aus den bereits zitierten Quellen illustrieren:

- „Greater proximity to the Interviewer ... may result in a feeling of uneasiness ...“⁴⁾ Aber andererseits: „... or the reverse, for example if rapport is established.“⁵⁾
- Der „Goodwill“ der Probanden, über Fragen nachzudenken, „(ist) ... im Papierinterview größer.“⁶⁾ Andererseits aber: „CAPI-Befragte nehmen das Interview ernster, reagieren vorsichtiger.“⁷⁾
- „CAPI Respondants produce on average more spontaneous awareness responses than PAPI.“⁸⁾ Aber andererseits: „So wurden z. B. von PAPI über CAPI zu CATI erheblich verringerte Werte bei der ungestützten Markenbekanntheit gemessen.“⁹⁾
- Die Beispiele ließen sich fortsetzen: Da ist die Rede davon, dass Befragte und Interviewer buchstäblich näher zueinander rücken müssen, damit beide auf den gleichen Bildschirm sehen können, und konsequenterweise wird „mehr Nähe“ beim CAPI-Interview konstatiert – auf der anderen Seite ist von einer „Endpersönlichung“ der Interviewsituation durch die Dazwischenkunft des Computers die Rede. Oder es wird (bei Mosmann) in ein und dem selben Aufsatz darüber berichtet, dass die „sozial unverbindlichere Interviewform CAPI“ prestigeorientiertes Antwortverhalten begünstige, dass aber andererseits bei PAPI der direkte Kontakt zum Interviewer während der Befragung zu „reaktivem Verhalten der Befragten mit dem Wunsch nach sozialer Akzeptanz führt“.

Man wird wohl davon ausgehen müssen, dass es im Hinblick auf die Qualität von Befragungen, insbesondere im Hinblick auf die Nettostichproben, Methodeneffekte gibt bei dem Umstieg auf CAPI-Techniken, dass sie aber vielschichtig sind, nicht generell negativ oder positiv zu bewerten, sondern dass sie interviewerspezifisch, befragtenspezifisch, themen- und fragen-spezifisch sind. Da bleibt ein weites Feld für Forschung. Und man wird resümieren können, dass der Umstieg auf CAPI-Techniken einerseits mehr Freiheit durch Technik bringt, und andererseits auch ein doch gelegentlich spürbares Technikkorsett für Umfragen bedeuten kann.

4) Taylor Nelson/AGB: CAPI data analysis summary report 1996 (unveröffentlicht).

5) Ebenda.

6) Mosmann, Helmut: PAPI, CAPI, CATI: Ambivalenter technischer Fortschritt in der Datenerhebung, in: Planung und Analyse, 1/1999, S. 50.

7) Ebenda.

8) Taylor Nelson/AGB, a. a. O.

9) Mosmann, a. a. O.

Der dritte Treiber: Geschwindigkeit

Mit CAPI lassen sich Umfragen in doppelter Hinsicht beschleunigen. Zum einen kann in der gleichen Zeit normalerweise ein etwas umfangreicheres Fragenpaket im Vergleich zur Papier- und-Bleistift-Umfrage bearbeitet werden, weil der Computer einen Teil der „Handlingprobleme“ übernimmt. Ergebnisse unserer Interviewerbefragungen zeigen allerdings, dass der Effekt – zumindest subjektiv aus der Sicht der Interviewer – nicht besonders bedeutend ist. 1996 meinten 50 % der CAPI-Interviewer, dass CAPI die Interviews schneller mache, während 26 % eine Verlangsamung beklagten. Ein Jahr später waren es nur noch 40 %, die das Gefühl hatten, mit CAPI würden die Interviews schneller, während 31 % die Gegenposition vertraten. Möglicherweise allerdings spiegeln diese Zahlen auch den Umstand wider, dass in dem Maß, wie sich die Projektleiter und Forscher des Instituts mit den Möglichkeiten der computergestützten Befragung beschäftigen, auch immer kompliziertere Anwendungen programmiert werden, die bei den Interviewern ebenso wie immer umfangreichere Fragebögen den Eindruck hinterlassen, dass es oft mit Computer eher langsamer geht. Insgesamt gehen wir im Institut davon aus, dass die durchschnittlichen Interviewzeiten bei ansonsten gleichem Fragenprogramm allenfalls minimal reduziert werden durch den Einsatz von Computern bei der Befragung.

Unter dem Gesichtspunkt der kommerziellen Marktforschung allerdings viel wichtiger ist die tatsächlich deutliche Zeitersparnis für die professionelle Abwicklung einer üblichen Marktforschungsuntersuchung. In Abbildung 4 (siehe S. 68) ist ein Beispiel für eine sogenannte Omnibusbefragung gegeben, wobei trotz der – wie schon beschrieben – verbesserten Steuermöglichkeiten von identischen Feldlaufzeiten im engeren Sinn – also die Tage, an denen Interviews gemacht werden – ausgegangen wurde. Der typische „Gewinn“ von ca. 5 Arbeitstagen kann für ein kommerzielles Marktforschungsinstitut im Wettbewerb mit anderen Instituten durchaus ein geldwerter Vorteil sein. Viele Auftraggeber jedenfalls legen hier Wert auf Geschwindigkeitsgewinn.

Abbildung 4

Geschwindigkeit- wieviel bringt CAPI?

Beispiel für Ablauf (Omnibusbefragung)	PAPI	CAPI
Frage schreiben/programmieren	1 Tag	3 Tage
Frage-Korrekturen/Programmtest	0,5 Tag	2 Tage
Produktion Erhebungsunterlagen	3 Tage	0,5 Tag
Versand/Postlaufzeit	3 Tage	1,5 Tage
Feldarbeit	x Tage	x Tage
Postlaufzeit zurück	2 Tage	--
Rücklaufkontrolle, Abrechnung ...	1 Tag	0,5 Tag
Datenaufnahme	1 Tag	--
Datenprüfung, Coden	3 Tage	2 Tage
	14,5 Tage +x	9,5 Tage +x

Der vierte Treiber: Neue Möglichkeiten

Multimedia-PCs bieten ohne Zweifel eine Reihe von Möglichkeiten, die für die klassische Papier-und-Bleistift-Befragung bisher nicht gegeben war (insbesondere, wenn der PC ein CD-ROM-Laufwerk hat). Zur Illustration, und um die Phantasie anzuregen, seien hier nur einige dieser Möglichkeiten genannt, die wir im Institut selber schon genutzt haben:

- Adaptive Conjointanalyse. Conjointanalysen konnten bisher mit Papier und Bleistift nur mit einer relativ geringen Anzahl von Merkmalen und Merkmalsausprägungen durchgeführt werden, da ein sogenannter „full profile approach“ mit einer letztlich aus Gründen der Befragbarkeit begrenzten Anzahl von Paarvergleichen notwendig war. Hier sind die Möglichkeiten durch den Computereinsatz außerordentlich erweitert.
- TV-Spotttests und auch Hörfunk-Tests mit „Vorlage“ des Testsujets im Rahmen von Stichproben, die durch Interviewerbesuch bei den Befragten zu Hause realisiert werden. Bisher waren solche Tests nur in Studios oder allenfalls in „rollenden Labors“ möglich. Verglichen mit dem letzten Fall sind auch die ökonomischen Ersparnisse erheblich.
- Design-Test für Verpackungen. Dabei kann vom Gesprächspartner mittels Cursortaste bzw. Maus die Testpackung auf dem Bildschirm gedreht, bewegt und bei Bedarf sogar auseinander genommen werden. Es sind Zoom-Effekte möglich und konkurrierende Verpackungen im Vergleich darstellbar.
- Audiovisuell gestützte Ermittlungen, beispielsweise im Hinblick auf die Milieuzugehörigkeit. Es ist etwas anderes, ob man die Präferenz für Schlager-/Heimat-/Popmusik usw. durch Vorgabe von Begriffen oder auch die Benennungen einzelner Titel oder Musik-

stücke ermittelt, oder ob die Titel oder Musikstücke angespielt und damit zu Gehör gebracht werden können. Es werden auf diesem Wege für die standardisierte Marktforschung ganz neue Möglichkeiten zur Skalen-Konstruktion eröffnet.

- „Virtual Reality“-Anwendungen erlauben es, um ein testweise neu designtes Auto herumzugehen, Orientierungsaufgaben an einem neu überarbeiteten Armaturenbrett zu stellen oder auch die Probanden aufzufordern, anhand eines elektronisch zur Verfügung gestellten „Bausatzes“ ihre Traumwohnung oder ihr Traumhaus einzurichten. Auch das sind Anwendungen, die für die Marktforschung neue Dimensionen eröffnen.
- Auch der Umstand, dass den Interviewern die Befragungsunterlagen einschließlich Fragebogen elektronisch überspielt werden können, eröffnet neue Möglichkeiten: Beispielsweise kann am Erscheinungstag einer Tageszeitung der Fragebogen auf die aktuelle Ausgabe hin formuliert werden, und noch am Abend des selben Tages können Interviews zur aktuellen Ausgabe durchgeführt werden. Damit wird das Ankreuzen mit Fettstift im Originalheft und die gelegentlich mühsame Vercodung solcher Markierungen im Institut weitgehend überflüssig.
- Ein jüngstes Beispiel neuer Möglichkeiten mit CAPI ist das sogenannte „Titelidentifikationsmodell“ in der Zeitschriftenforschung. Zu jedem Titel, der in der letzten Zeit vom Befragten einmal gelesen wurde, kann anhand der Vorlage beispielsweise der letzten 12 Titelbilder nachgefragt werden, welche der Ausgaben der Proband denn in der Hand hatte, bzw. an welche er sich noch erinnert. Und das ist eben neuerdings nicht mehr nur möglich für maximal zwei oder drei Titel – auch das war bisher schon schwierig, denn der Interviewer musste alle Titel vollständig bei sich führen und beispielsweise wöchentlich updaten –, sondern für dutzende Titel gleichzeitig.

Die Bedingungen, unter denen Umfragen durchgeführt werden, haben sich geändert

1964, als noch kein Zweifel war, dass das Face-to-Face-Interview mit Papier und Bleistift der „Königsweg der empirischen Sozialforschung bleiben wird“, wurde in der ersten Studie der Reihe „Massenkommunikation“, die unser Institut seither in Fünf-Jahres-Abständen durchführen konnte, ermittelt, dass 55 % der Haushalte einen Fernseher hatten und 25 % ein Telefon. Heute gibt es niemanden mehr ohne eigene Erfahrung mit Fernsehen und Telefon; 90 % der Wohnbevölkerung haben Erfahrungen mit Anrufbeantwortern oder Mailboxen, 40% mit Mobiltelefonen, 60 % mit PCs und 25 % mit dem Internet. All das gab es vor 35 Jahren überhaupt noch nicht. Es ist offenkundig, dass sich die technischen Voraussetzungen für Telekommunikation, und zwar insbesondere für Individualkommunikation, im Zeitraum gut einer mitteleuropäischen Generation komplett geändert haben. Die technischen Hilfsmittel, die heute weiten Teilen der Bevölkerung wie selbstverständlich zur Verfügung stehen und von ihr genutzt werden, haben auch das Kommunikationsverhalten selber geändert bzw. haben entsprechende Änderungen unterstützt. Diese Änderungen sollen hier nicht im einzelnen „bewiesen“ oder analysiert werden, es sei lediglich daran erinnert, was sich – und natürlich nicht nur im unmittelbaren Zusammenhang der Individualkommunikation – alles geändert hat, wobei in der folgenden kleinen Aufzählung nur solche Aspekte berücksichtigt sind, die im Hinblick auf das hier behandelte Thema des Einflusses von technischen Interviewhilfs-

mitteln auf Befragungsergebnisse unmittelbar relevant erscheinen, vor allem vor dem Hintergrund der oben angedeuteten Diskussion über Qualität von demoskopischer Forschung:

- Die individuelle Mobilität – und damit auch die Abwesenheitszeiten von zu Hause – hat sich enorm erhöht.
- Es ist inzwischen möglich, jederzeit überall erreichbar zu sein, und ein immer größerer Anteil der Bevölkerung hält das auch für wünschenswert und rüstet sich dementsprechend aus.
- Das Unterhaltungsangebot der Medien- und der Freizeitindustrie ist geradezu explosionsartig gewachsen in den zurückliegenden Jahrzehnten.
- Der Trend zu kleineren Haushalten und das Wachstum des Anteils von Einpersonenhaushalten ist ungebrochen.
- Werbung und Direktmarketing als ständige „Ansprache“ jedes einzelnen Verbrauchers sind mehr denn je allgegenwärtig.
- „Information overload“ ist nicht nur ein Modewort, sondern eine Zustandsbeschreibung nach der enormen Zunahme bildgestützter sowie akustischer und textvermittelter Informationsangebote und deren ebenfalls in großem Umfang gewachsenen Rezeption.
- Das Zeitbudget für technisch vermittelte Kommunikation, sei es Medienkommunikation oder Individualkommunikation per Telefon, via Internet usw., hat stark zugenommen auf Kosten der Zeit, die die Menschen für direkte, interpersonale, nicht-technisch vermittelte Kommunikation aufwenden.

All dies sind Faktoren, die in der einen oder anderen Weise mit den sinkenden Ausschöpfungsraten in der Markt- und Meinungsforschung zusammenhängen und diese z. T. bedingen. Hohe Ausschöpfungsraten, darauf wurde hier hingewiesen, sind keineswegs das einzige und in den meisten Fällen auch nicht das wichtigste Kriterium für die Qualität von Umfrageforschungsergebnissen. Aber sinkende Ausschöpfungsraten müssen unbedingt als Anlass genommen werden, Überlegungen anzustellen, wie unser Geschäft des Interviewens attraktiver, besser, schneller und vielleicht gleichzeitig sogar auch noch ökonomischer getan werden kann. So gesehen erscheint es mir kein falscher Gedanke, den Umstieg auf CAPI-Techniken auch als Übergang zu Web-gestützten Befragungen und E-Mail-Studien aufzufassen. Und auch die Frage zu stellen, inwiefern nicht unerwünschte Methodeneffekte dadurch entstehen, dass man sich dem Umstieg auf CAPI-Techniken verweigert und die veränderten Bedingungen für Befragungen nicht zum Anlass einer Überprüfung der Befragungstechniken nimmt. Ist es nicht schon längst so, dass die Verwendung von Papier und Bleistift als hauptsächliche Interviewhilfsmittel Rückständigkeit signalisiert?

Interviewsituation in computergestützten Befragungen – zur Wirkung von Technologieeffekten

1 Thema und Fragestellung

Computergestützte Befragungen gehören heute zu den Standardtechniken der Umfrageforschung (Couper et al. 1998). Nicht nur in telefonischen Befragungen – bei denen die Computerunterstützung in den siebziger und achtziger Jahren zuerst Einzug hielt (Fuchs 1994 und 1995) – sondern mehr und mehr auch in persönlichen Studien wird auf CAI-Techniken zurückgegriffen. Viele der großen amerikanischen Umfragen haben diesen Übergang zur Computerunterstützung bereits abgeschlossen; eine ähnliche Entwicklung beobachten wir zeitversetzt auch in Europa (Schneid 1991; Projektgruppe SOEP 1998). Die Vorteile der Computerunterstützung für Forscher wie Feldleiter liegen auf der Hand:

- Die Dateneingabe findet in der Interviewsituation statt. Dies erlaubt die Überprüfung der eingegebenen Codes im laufenden Interview ebenso wie Konsistenzprüfungen der Antworten auf miteinander in Beziehung stehenden Fragen. Zudem ist ein unproblematischer Rückgriff auf zuvor gespeicherte Daten des Befragten aus externen Datenbanken möglich, was insbesondere bei Wiederholungsbefragungen von großem Nutzen ist.
- Das Befragungssystem präsentiert dem Interviewer automatisch die nächste zutreffende Frage auf dem Bildschirm. Dies schränkt das versehentliche, aber auch das intentionale Überspringen und Auslassen von Fragen ebenso ein wie die Bearbeitung von Items, die auf Grund von Filterführungen für den jeweiligen Befragten nicht vorgesehen sind. Außerdem können die Befragungsinstrumente durch die automatische Filterführung komplexer gestaltet werden (weit verzweigte Filterführung, komplizierte Bedingungen), ohne den Interviewer zusätzlich zu belasten bzw. von der Administration der Fragen abzulenken.
- Die Zeit zwischen dem Abschluss der Feldphase und dem Erstellen eines Berichtes ist kürzer, da sowohl die Dateneingabe als auch eine erste Datenbereinigung bereits im Interview stattfinden. (Die Gesamtdauer eines Projektes scheint sich jedoch entgegen landläufiger Vermutungen nicht zu verkürzen – die Arbeitsschritte vor Beginn der Feldphase nehmen mehr Zeit in Anspruch.)

Die Entwicklung zu computergestützten Befragungen wird von Seiten der methodologischen Begleitforschung überwiegend positiv bewertet (siehe überblicksartig Couper et al. 1998). Andererseits haben die computergestützten Befragungsformen aber auch zusätzliche Probleme in die Interviewsituation eingeführt: In den ersten Jahren standen im Wesentlichen Hard- und Softwareprobleme im Mittelpunkt des Interesses (Couper et al. 1989; Weeks 1992). In den letzten Jahren sind eine Reihe von Studien hinzugekommen, die sich mit der Akzeptanz bei Interviewern und Befragten beschäftigen (Couper/Burt 1994), die Interviewdauer zum Gegenstand haben (Martin et al. 1993; Baker et al. 1995; Müller/ Kesselmann 1996; Fuchs et al. 2000) oder Fragestellungen aus dem Bereich Usability-Testing verfolgen,

*) PD Dr. Marek Fuchs, Katholische Universität Eichstätt.

also den Einfluss des Computerinterfaces auf das Verhalten der beteiligten Akteure und die Interviewsituation untersuchen (Couper et al. 1997; Hansen et al. 1997 und 1998; Fuchs 1999). Um diese letztgenannte Art von Effekten soll es im Folgenden ausführlicher gehen. Zuvor möchten wir jedoch eine terminologische Klärung vornehmen:

**Übersicht 1
Technologieeffekte und Modeeffekte**

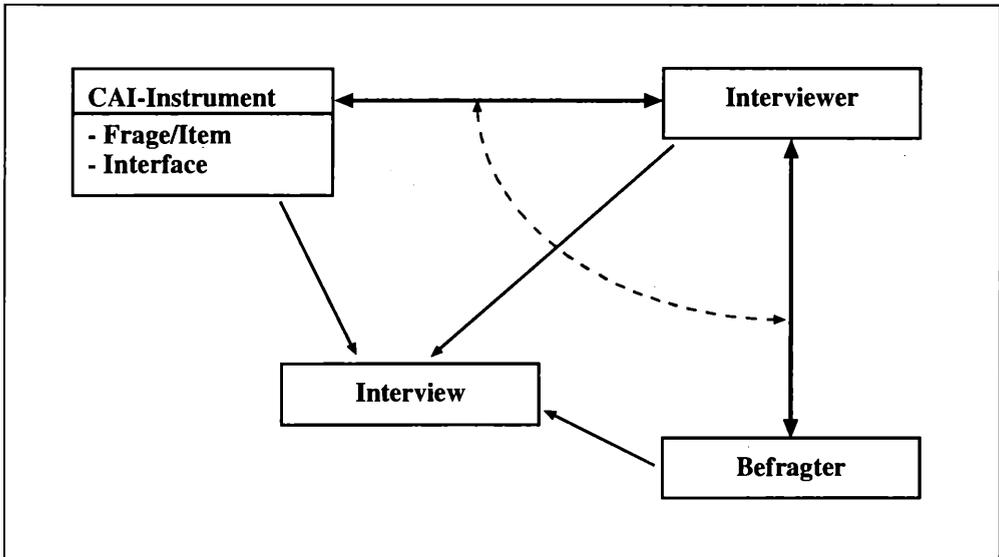
Befragungsmodus	Befragungstechnologie	
	„klassisch“	computergestützt
mündlich-persönlich	PAPI	CAPI
telefonisch	(PATI)	CATI
schriftlich	paper & pencil, postalisch usw.	CSAQ, Online

Als Ursache für mögliche Effekte der computergestützten Befragung auf die Datenqualität, auf die Dauer der einzelnen Interviews wie der Studie insgesamt und auf die Kosten eines Projektes wird auf die Präsenz des Computers in der Interviewsituation und die Gestaltung des Interfaces abgehoben. Damit unterscheiden sich diese Analysen vom Zugang der Untersuchungen zu Modeeffekten, die auf die Präsenz bzw. Abwesenheit des Interviewers bzw. auf den im Interview gewählten Kommunikationskanal (incl. der Unterstützungsmöglichkeiten durch Bildblattvorlagen usw.) abheben. Entsprechend sollen diese computerbezogenen Technologieeffekte von den Modeeffekten getrennt thematisiert und analysiert werden (vgl. Übersicht 1).

Als Ergebnis entsprechender theoretischer Überlegungen liegt eine Konzeption der Befragungssituation vor (vgl. Abbildung 1, S. 73), in der der Computer als zusätzlicher „Akteur“ in Erscheinung tritt: Interviewer und Befragter interagieren, um die Aufgabe (die Durchführung der Befragung) zu bearbeiten. Zusätzlich interagiert der Interviewer mit dem Computer, auf dem das Befragungsinstrument läuft. Auch der Computer ist in die Lösung der Aufgabe eingebunden, indem er Fragetexte und Antwortvorgaben präsentiert, die Dateneingabe entgegennimmt, Hilfenmenüs anbietet usw. Zentrale Annahme ist nun, dass die Interaktion zwischen Interviewer und Befragtem durch die Interaktion des Interviewers mit dem Computer in verschiedener Weise beeinflusst werden kann. Dies wird durch die gestrichelte Linie in der nachfolgenden Abbildung zum Ausdruck gebracht. Dabei gehen die aus Sicht der hier diskutierten Forschungslinie wichtigen Effekte nicht von der formalen Struktur des Fragebogens, also dem Frage-Antwort-Prozess aus, sondern von der Gestaltung des Bildschirminterfaces. D. h., selbst bei gleichen Frageformulierungen und identischen Antwortvorgaben wirken unterschiedliche Bildschirmdesigns differenziert auf das Verhalten des Interviewers dem Befragten gegenüber und damit möglicherweise auf die Datenqualität.¹⁾

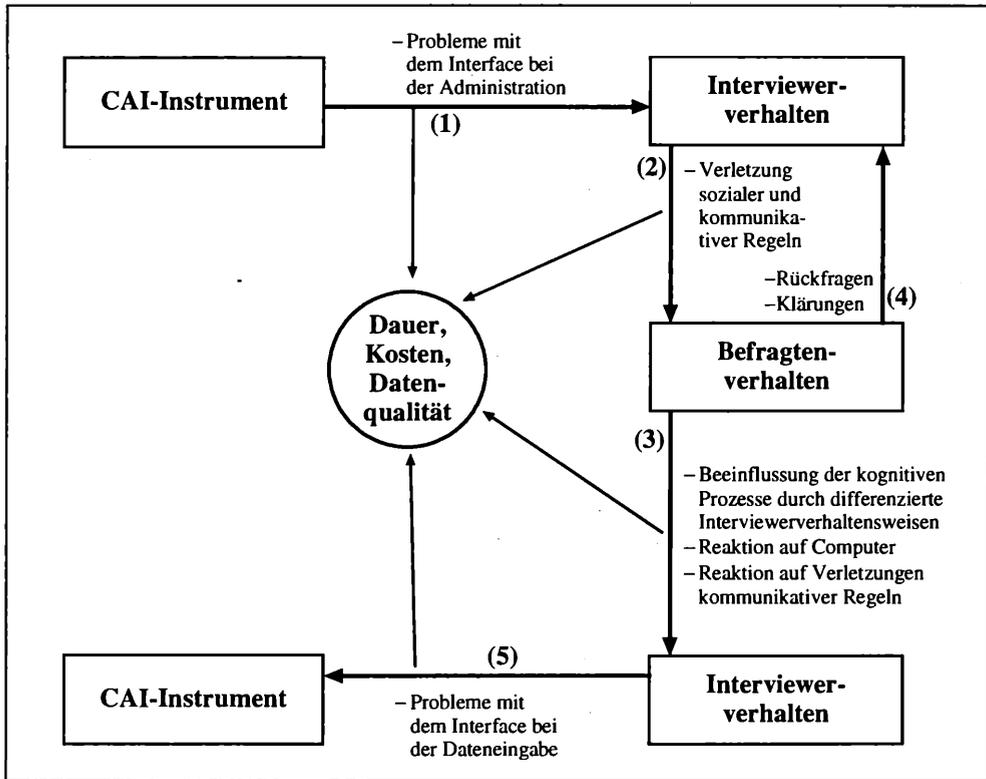
1) Häufig implizieren Variationen des Bildschirmdesigns zugleich Veränderungen des Frage-Antwort-Prozesses: Z. B. werden – je nach Design – zusätzliche Kontextinformationen gegeben, oder die Realisation des Frage-Antwort-Prozesses durch den Interviewer in einer konkreten Interviewsituation wird verändert.

Abbildung 1
Interviewer-Befragten-Interaktion im computergestützten Interview



Das Interface kann das Interviewerverhalten und – wie später gezeigt – auch das Befragtenverhalten in verschiedenen Stadien des Interviewflusses beeinflussen. Zerlegt man die Administration einer Frage bzw. eines Items aus analytischer Perspektive, so lassen sich fünf Segmente abgrenzen (vgl. die folgende Abbildung 2, S. 74): (1) Erkennen der Aufgabe durch den Interviewer (incl. Beachtung etwaiger Intervieweranweisungen), (2) Vorlesen der Frage, (3) Generierung einer Antwort durch den Befragten, (4) Interaktion des Interviewers mit dem Befragten, ggf. Beantwortung von Rückfragen des Interviewpartners, Entgegennahme der Antwort des Befragten, Feedback, sowie (5) Eingabe der Antwort in das CAI-Instrument.

Abbildung 2
Technologieeffekte in der Interviewsituation



In allen fünf Stadien der Abarbeitung eines Items kann die Gestaltung des CAI-Interfaces Effekte haben: Je nach Gestaltung des Bildschirms wird es dem Interviewer leicht oder schwer gemacht, (1) die Aufgabe bei der jeweiligen Frage zu erkennen, Intervieweranweisungen zu beachten und (2) die Frage so vorzulesen, wie sie auf dem Bildschirm steht. Zu denken ist etwa an überkomplexe Bildschirmdesigns, unklare Trennung zwischen Fragetext, Intervieweranweisungen und Antwortcodes, auf die Verteilung der für die Bearbeitung einer Frage wichtigen Informationen auf mehrere Bildschirme, die Notwendigkeit zu scrollen usw. Je nach Gestaltung des Bildschirms wird der Interviewer mehr oder weniger Zeit und Aufmerksamkeit auf den Bildschirm richten. Entsprechend wird er die Kommunikation mit dem Befragten aufrecht halten können, während er sich mit der anstehenden Frage beschäftigt oder aber er wird den Interviewpartner – kurzzeitig – vernachlässigen müssen. Daraus folgt entweder eine flüssigere oder eine segmentiertere Interviewadministration. (3) Diese differenzierten Verhaltensweisen des Interviewers bleiben nicht ohne Wirkung auf das Befragtenverhalten. Dabei sind zwei Effekte zu unterscheiden: (a) nämlich die Reaktionen des Interviewpartners auf eine nicht den Standardregeln folgende Administration eines Items, die auf ein problematisches oder unklares Bildschirmdesign zurückzuführen sind und (b) Reaktionen auf die korrekte, aber monotone, hochgradig segmentierte und jedes Item ohne Kontext „her-

unterleiernde“ Abarbeitung des Interviews.²⁾ Ähnliche Effekte des Interfaces auf die Interviewsituation können weiter (4) bei der Suche nach Zusatzinformationen für die Beantwortung von Rückfragen des Interviewpartners auftreten, wenn etwa das Hilfesystem des CAI-Instruments das Auffinden entsprechender Informationen erschwert, oder wenn (5) die Eingabe oder Änderung von Antworten durch die Gestaltung der Dateneingabemaske behindert wird.

Wir müssen also computerbezogene „Unfälle“ und Fehler im Interviewerverhalten trennen von den Folgen eines regulären Interviewerverhaltens, die gleichwohl durch die verwendete Technologie – hier: die Gestaltung des Interfaces – hervorgerufen werden. Beide Perspektiven verdienen gesonderte Aufmerksamkeit. Im Folgenden soll jedoch schwerpunktmäßig auf die Merkmale und Folgen des „normalen“, regelgerechten und scheinbar korrekten Interviewerverhaltens eingegangen werden.

Noch können nicht alle hier genannten Effekte experimentell belegt werden. Insbesondere der Nachweis, unter welchen Bedingungen diese Effekte nicht nur die Interviewsituation tangieren, sondern auch die Datenqualität beeinflussen, steht noch weitgehend aus. Dennoch, erste empirische Belege gibt es bereits für die zunächst nur in Laborexperimenten und Feldbeobachtungen identifizierten Effekte. Im vorliegenden Beitrag sollen exemplarisch erste vorliegende Ergebnisse der mit diesem Gegenstand befassten und unter dem Schlagwort „Usability-Testing“ operierenden Forschungslinie zusammengetragen werden.

2 Methoden

Die hier referierten Befunde basieren auf zwei verschiedenen Experimenten: (1) Zunächst berichten wir die Ergebnisse eines Vergleichs einer computergestützten und einer nichtcomputergestützten Variante des gleichen Fragebogens. (2) Danach gehen wir auf eine laborexperimentelle Studie ein, in der identische Fragen unter Verwendung verschiedener Bildschirmvarianten in einem Feldexperiment miteinander verglichen wurden.

(1) 1997 wurden im Usability-Labor des Survey Research Center der Universität Michigan (USA) 51 Interviews des National Health Interview Surveys (NHIS) durchgeführt (37 CAPI und 14 PAPI). An dieser Studie wirkten 9 Interviewer des Bureau of the Census der Vereinigten Staaten mit und 51 rekrutierte Versuchspersonen aus der Umgebung des Universitätsstandortes. Die Interviews wurden durch Videoaufnahmen dokumentiert (zwei Kameraperspektiven zeigen die Interviewsituation aus verschiedenen Blickwinkeln, das dritte Band enthält den Bildschirminhalt des CAI-Instruments aufgenommen durch einen Scann-Converter). Hinzu kommen ein Audiomitschnitt sowie Trace-Files und Ergebnisse einer Befragung der Teilnehmer zu ihren Erfahrungen. Die Methode sowie Ergebnisse dieser Tests sind ausführlich dokumentiert bei Hansen und Mitarbeitern (1997 und 1998) sowie Lepkowski und Mitarbeitern (1998).

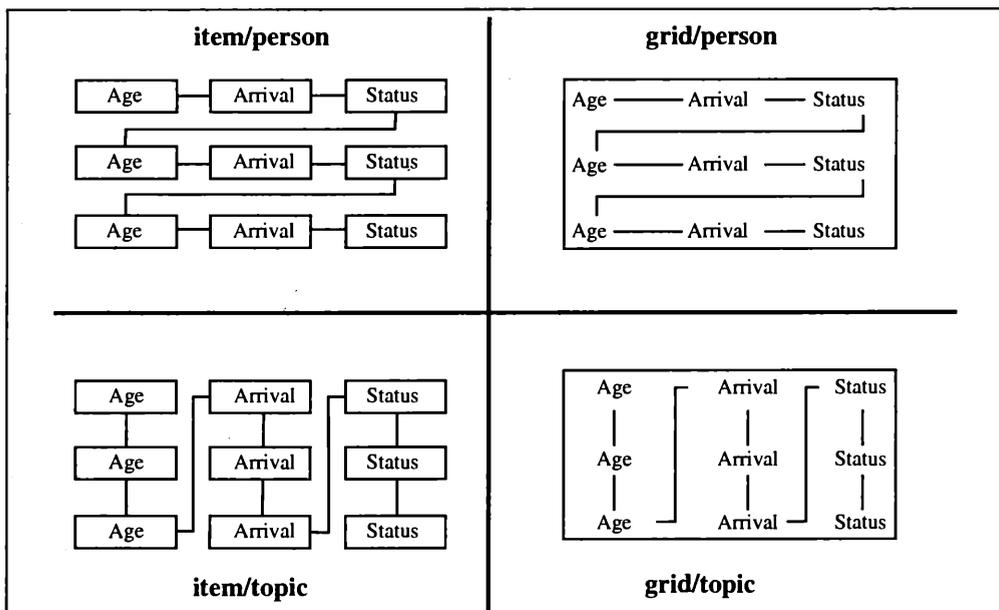
Für die nachfolgende beispielhafte Dokumentation von Technologieeffekten greifen wir auf diese Daten sekundäranalytisch zurück (Fuchs et al. 2000). Dabei wurden die Videobänder in einem zusätzlichen Analyseschritt – ähnlich wie beim Behavior-Coding – inhaltsanalytisch

2) Beide Effekte sind zu trennen von direkten Einflüssen des Computers auf das Befragtenverhalten, etwa auf Grund von Datenschutzbefürchtungen, Skepsis gegenüber der Technik, spezifischen Erwartungen über die weitere Verwendung der Daten usw. Diese letztgenannten Effekte haben zwar mit der Anwesenheit des Computers in der Interviewsituation zu tun, gehen aber nicht auf die Gestaltung des Interfaces zurück.

ausgewertet. Insgesamt wurden im soziodemographischen Modul des Fragebogens bei allen 51 Interviews zusammen 2 200 Items bzw. Fragen abgearbeitet, die durch ca. 7 000 codierte Aktionen (Segmente) erfasst wurden. Jedes Segment beschreibt eine spezifische Interviewer- oder Befragtenverhaltensweise (analog zu der in Abbildung 2 vorgenommenen Differenzierung) zusammen mit deren Dauer und weiteren Merkmalen. Diese Daten erlauben uns einen Vergleich der CAPI-Interviews mit den PAPI-Fällen auf einer sehr disaggregierten Ebene.

(2) Technologieeffekte treten jedoch nicht nur beim Übergang von „klassischen“ zu computergestützten Befragungsmethoden auf (z. B. PAPI vs. CAPI), sondern das Verhalten der beteiligten Akteure variiert auch in Abhängigkeit von der jeweiligen Gestaltung des CAI-Instruments. Daher haben wir identische Fragen mit verschiedenen Interfacevarianten verglichen. In einer im März 1998 an der Katholischen Universität Eichstätt durchgeführten CATI-Studie wurden 501 zufällig ausgewählte Interviewpartner mit einem Instrument befragt, in das ein 2-faktorielles Methodenexperiment integriert war. Die Interviewer wurden im soziodemographischen Modul bei jedem Interview zufällig mit einem von vier verschiedenen Bildschirmdesigns konfrontiert. Dabei haben wir die Dauer der Abarbeitung jeder einzelnen Frage gemessen, sowie von 234 zufällig ausgewählten Interviews das Verhalten des Interviewers anhand von Videoaufzeichnungen kodiert.

Abbildung 3
Vier verschiedene Bildschirmdesigns



Die vier Bildschirmvarianten unterschieden sich einerseits durch die Anzahl der gleichzeitig auf dem Bildschirm präsentierten Fragen und zum anderen durch die Fragereihenfolge. Auf die Wirkungen der unterschiedlichen Fragereihenfolge (vgl. ausführlich Fuchs 1999) soll hier nicht näher eingegangen werden.

Das in das Experiment einbezogene Fragemodul erfasste drei Merkmale/Variablen für alle im Haushalt lebenden Personen, so dass alle drei Fragen für jedes Haushaltsmitglied gestellt werden mussten. Integriert wurde (1) eine itembasierte und personenorientierte Version (item/person), (2) eine gridbasierte und personenorientierte Version (grid/person), (3) eine itembasierte und themenorientierte Version (item/topic) sowie (4) eine gridbasierte und themenorientierte Version (grid/topic). Die Abbildung 3 verdeutlicht die unterschiedlichen Vorgehensweisen: Während bei den itembasierten Versionen (links) auf jedem Bildschirm nur eine Frage gezeigt wird, sind bei den gridbasierten Versionen (rechts) alle Fragen bzw. Eingabefelder zu einer Tabelle zusammengefasst. Und während bei den personenorientierten Versionen (oben) zunächst alle drei Fragen für eine Person abgearbeitet werden, bevor die gleichen Items für die nächste Person administriert werden, verfolgen die themenorientierten Versionen (unten) einen Ansatz, bei dem zunächst eine Frage für alle Personen im Haushalt gestellt wird, bevor der Interviewer zum nächsten Merkmal übergeht und dieses für alle Personen erfragt (vgl. zum Hintergrund dieser Differenzierung die Überlegungen von Moore 1996; Moore/Moyer 1998 a; 1998 b). Dieses Vorgehen erlaubt einen Vergleich der Interviewdauer und der Verhaltensweisen der beteiligten Akteure in Abhängigkeit von der zufällig zugewiesenen Bildschirmvariante.

3 Befunde

3.1 Experiment (1), Vergleich PAPI vs. CAPI

In einem ersten Schritt konnte gezeigt werden, dass die eingesetzte Technologie bei der Datenaufzeichnung eine Rolle spielt und dadurch auf das Verhalten der Akteure einwirkt: Verdeutlichen möchten wir das anhand einer Frage, bei der Vor- und Nachname aller im Haushalt lebenden Personen – incl. Initial – zu erheben war (vgl. Abbildung 4, S. 78).

Vergleicht man die Aktionen der beteiligten Akteure bei PAPI- bzw. CAPI-Interviews, so stellt man eine Reihe von Unterschieden bei der Abarbeitung dieser Frage fest: Insgesamt dauert die Administration dieser Frage unter CAPI-Bedingungen drei Sekunden länger (+ 18 %) als im PAPI-Interview (20,6 vs. 17,5 Sekunden). Während sich keine signifikanten Unterschiede bei der Suche nach dem Fragetext (auf dem Bogen bzw. auf dem Bildschirm) sowie beim Vorlesen der Frage ergeben, dauern sowohl die Verhandlungen mit dem Befragten und auch die Erfassung der Antwort im CAPI-Interview länger (vgl. Abbildung 5, S. 79). Dies wird aus unserer Sicht durch zwei verschiedene Faktoren hervorgerufen:

Abbildung 4
Bildschirm des „What are the names ...“-Segments

Caseid: 005
Item: RPNAME@LNAME

-RPNAME-

"What are the names of all persons living or staying here?
Start with the name of the person, or one of the persons,
who owns or rents this home."

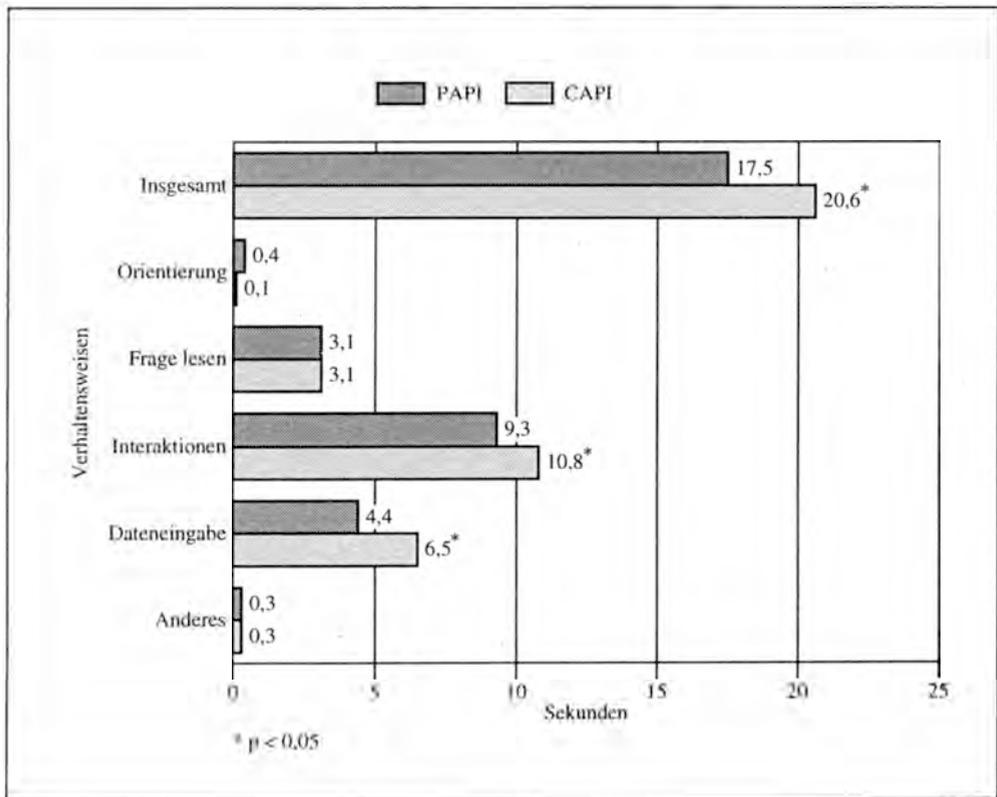
FR: BEFORE PROCEEDING, PLEASE MAKE SURE THE REFERENCE
PERSON IS NOT AN ACTIVE ARMED FORCES MEMBER.

PROBE FOR MIDDLE INITIAL IF NOT REPORTED.
PRESS "ENTER" TO SKIP TO LAST NAME IF NO MIDDLE INITIAL.

FIRST NAME: John _____ (H)
MIDDLE NAME: A
LAST NAME: Smith

(1) Wir glauben ableiten zu können, dass die Erfassung der Namen mit der Tastatur länger dauert als mit einem Stift: Schreiben scheint schneller zu gehen als Tippen (4,4 vs. 6,5 Sekunden). Zu berücksichtigen ist allerdings, dass die Antworten im PAPI-Interview vom Interviewer zunächst schnell und ggf. abgekürzt oder schlecht leserlich hingeschrieben werden können, da nach Abschluss des Interviews die Gelegenheit besteht, den Bogen zu editieren und entsprechende Korrekturen und Verbesserungen (etwa Druckbuchstaben an Stelle einer schlecht lesbaren Schreibschrift) vorzunehmen. CAPI-Interviews sind hingegen (aus Datenschutzgründen) nach Abschluss des Interviews nicht mehr für den Interviewer zugänglich. Jede alphanumerische Angabe muss also „auf den ersten Sitz“ richtig und vollständig erfasst werden. Möglicherweise trägt auch der höhere Editieraufwand mit der Tastatur – neben der Zeit für das reine Eintippen – zur längeren Dauer bei. Außerdem ist zu berücksichtigen, dass die EDV-Erfassung der Namen nach Durchführung des Interviews unter PAPI-Bedingungen zusätzliche Zeit in Anspruch nimmt, so dass der Saldo möglicherweise sogar zugunsten von CAPI ausfällt. Es bleibt jedoch festzuhalten, dass die Erfassung der Namen *im Interview* im CAPI-Interview mehr Zeit in Anspruch nimmt.

Abbildung 5
Dauer von Interviewer- und Befragtenverhaltensweisen
beim „What are the names ...“-Segment



(2) Weiter zeigt sich, dass Interviewer und Befragte *nach* dem Verlesen des Fragetextes im CAPI-Interview mehr Zeit auf Interaktionen verwenden als unter PAPI-Bedingungen. Dies rührt daher, dass das CAPI-Instrument eine starre Reihenfolge der Eingabefelder auf dem Bildschirm vorsieht. Während der Interviewer im PAPI-Interview die Angaben (Vor- und Nachname sowie ggf. Initialen) in der Reihenfolge erfassen kann, in der sie vom Befragten mitgeteilt werden, muss der Interviewer im CAPI-Interview nach Möglichkeit den Befragten dazu bringen, die Angaben in der vom Computer gewünschten Reihenfolge (hier: Vorname, Initialen, Nachname) vorzunehmen. Zwar bietet auch das CAPI-Instrument die Möglichkeit, zwischen den Feldern hin- und herzuspringen, jedoch kostet das ebenfalls Zeit und ist nur mit Mühe zu realisieren.³⁾ Daher wird der Interviewer einen Befragten, der die Antworten nicht

3) Es muss ja zumindest in das Feld für den Vornamen – anders bei den Initialen – etwas eingetragen werden, bevor der Interviewer zum nächsten Feld springen kann – dass leere Eingabefelder nicht zugelassen werden, ist ja eine wichtige Voraussetzung dafür, dass die Angaben im CAPI-Interview in der Regel vollständiger sind als bei PAPI.

in der vom Programm vorgesehenen Reihenfolge mitteilt, entsprechend instruieren (was durch eine Analyse der Videobänder bestätigt wird). Dies trägt zu der zusätzlichen Interviewdauer dieser Tätigkeit bei (9,3 vs. 10,8 Sekunden im Durchschnitt).

Diese Inflexibilität in der Administration finden wir aber nicht nur bei einzelnen Items, sondern sie lässt sich auch für größere Fragebogenmodule insgesamt aufzeigen. Betrachtet man die Dauer des soziodemographischen Moduls insgesamt, so stellt man fest, dass die computergestützte Variante 38 Sekunden länger dauert als die papiergestützte Variante (268 im Vergleich zu 230 Sekunden). An diesem Zuwachs sind Interviewer und Befragter proportional in gleicher Weise beteiligt. Analysiert man die Ursachen für diese zusätzliche Dauer (ausführlich bei Fuchs et al. 2000), so stellt man fest, dass das computergestützte Instrument – obgleich es identische Informationen erhebt – weniger flexibel ist und einen höheren Segmentationsgrad aufweist: Während der Interviewer in der papiergestützten Variante jeweils den gesamten Fragebogen oder zumindest diejenigen Fragen, die zum soziodemographischen Modul gehören, im Blick hat und sich bei der Abarbeitung der entsprechenden Fragen flexibel an den Reaktionen des Befragten sowie am Interviewverlauf orientieren kann, gibt das computergestützte Instrument eine rigide Fragereihenfolge und ein disaggregiertes, segmentiertes Design vor. Auf den meisten Bildschirmen findet sich nur ein Eingabefeld mit den zugehörigen Fragetexten und Antwortkategorien und erst wenn der Interviewer die Entertaste drückt, erscheint die nächste Frage. Beides, sowohl die rigide Fragereihenfolge wie auch die Segmentation (und die daraus folgende Redundanz), verlängern das Interview (vgl. hierzu auch das nachfolgende Experiment (2)).

Dies belegen die Befunde, wonach im CAPI-Interview mehr Fragen tatsächlich administriert werden als im PAPI-Interview (11,1 vs. 8,9 Fragen je Person im Durchschnitt), was aus unserer Sicht ein Hinweis darauf ist, dass der Interviewer im PAPI-Interview flexibler auf die vom Befragten gegebenen Antworten reagieren kann und z. B. Fragen dann nicht stellt (oder sich die Antwort bestätigen lässt), wenn der Befragte die entsprechende Information beim unmittelbar vorausgehenden Item bereits mitgeliefert hat. Die nachfolgenden Befunde deuten nämlich darauf hin, dass der Interviewer im PAPI-Interview leichter von solchen zusätzlichen Angaben des Befragten auf vorausgehende Fragen Gebrauch machen kann. Wenn also ein Interviewpartner bei der Beantwortung einer Frage zugleich z. B. das nachfolgende Item beantwortet (vor allem bei Faktfragen), kann der Interviewer im PAPI-Interview diese Informationen sofort aufnehmen, weil er das ganze Fragemodul vor Augen hat und muss die nachfolgenden Fragen u. U. nicht mehr administrieren oder kann sich die antizipierte Antwort bestätigen lassen.

An dieser Stelle wird deutlich, dass Interviewer häufig keine andere Wahl haben, als von den standardmäßigen Interviewerregeln abzuweichen: Gibt der Befragte bei der Beantwortung einer Frage Informationen, die als Antwort auf eine nachfolgende Frage dienen, hat der Interviewer zwei Optionen: (1) Notiert der Interviewer die Antwort auf die Frage, ohne das Item zu administrieren, so verletzt er grundlegende Interviewerregeln, aber er vermeidet die Belastung der Interviewsituation durch Redundanz. (2) Administriert der Interviewer diese Frage gleichwohl, vermittelt er dem Befragten den Eindruck eines unaufmerksamen Kommunikationspartners („Das hatte ich doch gerade schon gesagt“), was die Kommunikations-

situation mit dem Befragten belastet.⁴⁾ Auch der Kompromiss – sich die Antwort bestätigen zu lassen –, verletzt Interviewregeln und kann eine Irritation des Befragten nicht vollständig vermeiden.

Abbildung 6
Beispiel für Segmentation und Redundanz
(vier aufeinander folgende Bildschirme)

What race do you consider yourself to be? white/caucasian african-american asian ...
What race does JIM consider himself to be? white/caucasian african-american asian ...
What race does ANNE consider herself to be? white/caucasian african-american asian ...
What race does PAUL consider himself to be? white/caucasian african-american asian ...

Ein typisches Beispiel für derartige Redundanzen ist die in amerikanischen Studien übliche Frage nach der Rassenzugehörigkeit, die für jedes Haushaltsmitglied separat gestellt wird: „What race do you consider yourself to be?“ (vgl. Abbildung 6). In vielen Interviews – vor allem bei solchen in Haushalten mit mehr als zwei oder drei Personen – antwortet der Interviewpartner bei der zweiten oder dritten derartigen Frage u. U.: „We are all black!“ (vgl. auch Oksenberg et al. 1992 für entsprechende Beispiele). In dieser Situation gibt es für den Interviewer keinen Grund, die entsprechenden Fragen auch für alle weiteren Personen zu stellen – im Gegenteil: Unter Umständen würde er sogar dadurch die Kommunikation mit dem Befragten belasten, weil eine hochgradig redundante Interviewführung aus Sicht des Befrag-

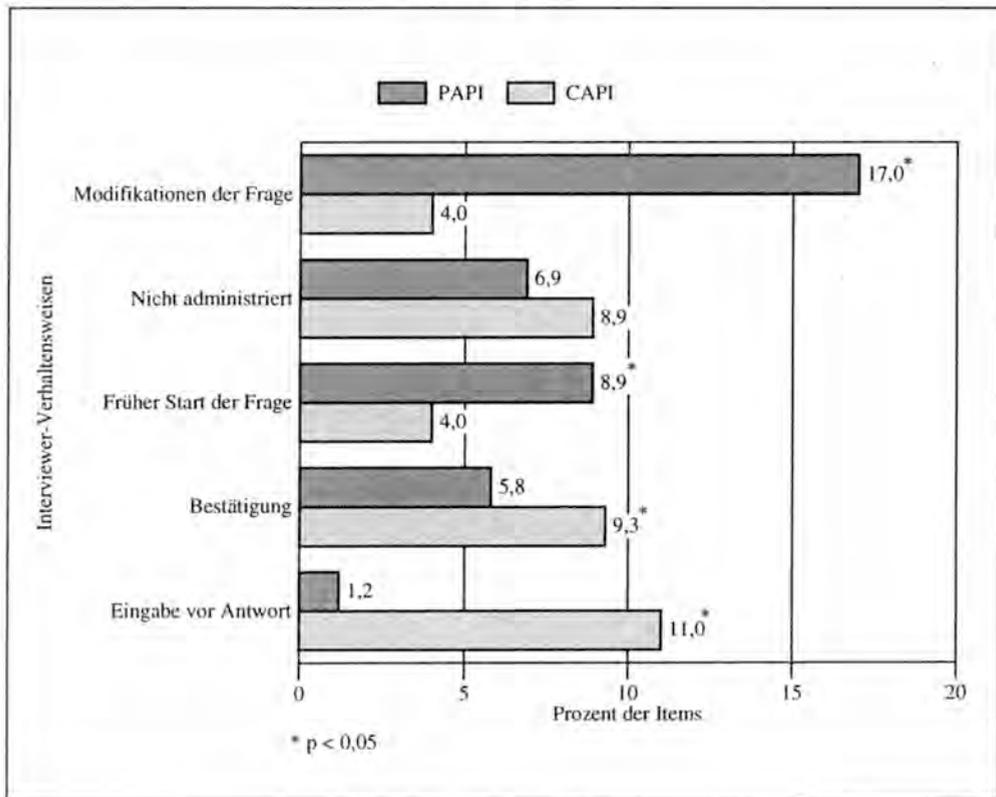
4) Und im Extrem kann dies zu einer Verschiebung der vom Befragten perzipierten Bedeutung der Frage führen: Wenn der Interviewpartner den Eindruck hat, eine Frage in der zunächst dechiffrierten Bedeutung bereits beantwortet zu haben, wird er möglicherweise nach weiteren Bedeutungen suchen, die es gerechtfertigt erscheinen lassen, dass der Interviewer die Frage trotz der bereits gegebenen Informationen stellt. – Vgl. hierzu die Überlegungen zu den kognitiven Prozessen bei der Beantwortung einer Frage z. B. bei Sudman et al. 1996.

ten als sozial unangebracht – weil Kommunikationsregeln verletzend –, unprofessionell und langweilig perzipiert würde. Signalisiert dies doch dem Interviewpartner mangelnde Aufmerksamkeit bzw., dass der Interviewer nicht adäquat auf die Mitteilungen des Befragten reagiert. Im PAPI-Interview wird es dem Interviewer leichter gemacht, derartige Redundanzen zu vermeiden, weil nicht nur die jeweils aktuelle Frage auf dem Bildschirm gezeigt wird, sondern auch einige der nachfolgenden Items sichtbar sind. So kann der Interviewer die Relevanz der vom Befragten gegebenen zusätzlichen Informationen für die nachfolgenden Fragen besser erkennen und sein Verhalten darauf ausrichten (= ggf. eine Frage nicht administrieren bzw. eine Bestätigung einholen). Dadurch verringert sich die Anzahl der vollständig administrierten Fragen. Unter den Bedingungen eines segmentierten CAI-Interviews (ein Item je Bildschirm) beobachten wir derartige Anpassungsleistungen des Interviewers an die soziale Befragungssituation seltener. Entsprechend höher ist die Anzahl der Items je Interview.⁵⁾

Eine entsprechende Analyse unserer Experimentaldaten zeigt, dass Interviewer – je nach Technologie – in unterschiedlichem Ausmaß derartige Abweichungen von den standardmäßigen Interviewerverhaltensweisen praktizieren: Während in der PAPI-Variante die Interviewer bei 17 % aller Items kleinere Modifikationen am Fragetext vornehmen, beträgt der vergleichbare Anteil bei CAPI lediglich 4 %. Andererseits ist der Anteil der nicht administrierten Items im CAPI-Interview um 2 Prozentpunkte höher und vor allem der Anteil der Fragen (+ 3,5 %), bei denen der Interviewer sich die vom Befragten zuvor im Kontext einer anderen Frage bereits genannte oder vom Interviewer sonstwie erschlossene Antwort bestätigen lässt (z. B. „Und Ihr Mann ist auch berufstätig?“). Weiter zeigt sich, dass Interviewer im CAPI-Interview bei jedem neunten Item bereits eine Antwort in den Computer eingeben, bevor der Befragte die Antwort gegeben hat. Der vergleichbare Anteil bei PAPI-Interviews ist mit gut einem Prozent zu vernachlässigen.

5) Wir werden später noch sehen (vgl. die Befunde zu Experiment (2)), dass man dieses Verhalten von Interviewer und Befragtem (!) durch ein entsprechendes Bildschirmdesign zulassen bzw. fördern kann.

Abbildung 7
Anteil des nichtstandardmäßigen Interviewerverhaltens
(Prozent der Items)



Wir können aus diesen Daten ablesen, dass die Interviewer auf die Segmentation des Interviewflusses in viele kleine Fragen, von denen jede ohne entsprechenden Kontext auf einer Bildschirmseite erscheint, mit einer Reihe von Strategien reagieren, damit die Interviewsituation nicht durch die dadurch produzierte Redundanz gestört wird: Sie beantworten Items zu einem größeren Anteil, ohne den Interviewpartner direkt zu fragen, sie lassen sich Angaben, die bereits bekannt sind, bestätigen, anstatt den kompletten Fragetext vorzulesen und sie antizipieren die Antworten der Befragten aufgrund der von ihnen zuvor gemachten Angaben (sie geben die Antwort ein, bevor der Befragte eigentlich geantwortet hat). Umgekehrt sind sie in CAPI bei weniger Items in der Lage, den nächsten Fragetext zu antizipieren und mit der Administration der Frage zu beginnen, während sie die Antworten auf das vorangegangene

Item eingeben: Der Anteil der Items ist im CAPI-Interview nur halb so groß. Trotz identischer Fragetexte variiert das Interviewerverhalten also ganz wesentlich mit der eingesetzten Technologie.⁶⁾

3.2 Experiment (2), Itemdesign vs. Griddesign

In den vorangegangenen Analysen konnten wir zeigen, dass sich beim Übergang zu CAI trotz identischer Fragen und gleicher Antwortvorgaben die Interviewersituation verändert und dass dies Konsequenzen für die Aktionen der beteiligten Akteure und für die Dauer der Befragung haben kann. Dabei war u. a. die vom CAI-Instrument ausgehende Redundanz als Ursachen für die Unterschiede herausgearbeitet worden. Im zweiten Experiment haben wir versucht, verschiedene Fragebogenvarianten im Feld gegeneinander zu testen, von denen wir aufgrund von Laborexperimenten annehmen konnten, dass sie es dem Interviewer in unterschiedlichem Ausmaß erlauben, die Folgen von Rigidität und Segmentation in der Interviewersituation zu vermeiden.⁷⁾

Zunächst zeigt sich, dass das Bildschirmdesign einen Einfluss auf die Dauer des Interviews hat. Diejenigen Varianten, die dem Interviewer in Tabellenform jeweils alle Eingabefelder und Informationen für einen Haushalt gemeinsam präsentieren (Griddesign), benötigen im Vergleich zu den beiden anderen Varianten weniger Zeit während der Befragung. Besonders krass fällt dieser Unterschied aus, wenn das gridbasierte Bildschirmdesign mit einer themenorientierten Fragereihenfolge verknüpft wird.

Berechnet man die durchschnittliche Dauer für jedes einzelne Item im Haushaltsroster (= Set von demographischen Fragen an jedes einzelne Haushaltsmitglied), so stellt man fest, dass die itembasierte, personenorientierte Version (item + person), die als Standardversion in computergestützten Befragungsinstrumenten üblicherweise zum Einsatz kommt, im Vergleich zu den anderen drei Versionen die meiste Zeit benötigt (6,6 Sekunden je Item). Im Kontrast dazu brauchen Interviewer in der gridbasierten themenorientierten Version mit 5,5 Sekunden im Durchschnitt die wenigste Zeit zur Bearbeitung. Dies entspricht einer Reduktion von etwa 17 %. Die beiden anderen Versionen bewegen sich hinsichtlich der Dauer zwischen diesen beiden Extremen. Auch wenn diese Zeitreduktion nicht auf die Gesamtdauer des Interviews extrapoliert werden darf, da das Griddesign nicht auf alle Module des Fragebogens anwendbar ist, ergeben sich aus seinem Einsatz dennoch Zeit- und Kostenersparnisse in erheblichem Umfang. Es ist wichtig, darauf hinzuweisen, dass sowohl die Fragereihenfolge als auch der Bildschirmaufbau zur Bearbeitungsdauer beitragen. Sowohl der gridbasierte Bildschirmaufbau als auch die themenorientierte Fragereihenfolge verkürzen jeweils die zur Bearbeitung eines Items notwendige Zeit. Beide Effekte kumuliert führen zur kürzesten Bearbeitungsdauer bei der gridbasierten, themenorientierten Version.⁸⁾

6) Einige der hier angeführten Beispiele für Technologie-Effekte (vgl. ausführlich Fuchs et al. 2000) mögen bei Benutzung anderer Befragungsprogramme nicht oder in anderer Art auftreten. Moderne Gestaltungsmöglichkeiten unter Windows können andere Effekte verringern – aber u. U. auch neue in die Interviewersituation einführen. Wir wollen mit diesen – singulären – Befunden lediglich darauf aufmerksam machen, dass der Einsatz von CAI solche Technologie-Effekte zeitigen kann. Dies soll nicht als Argument gegen CAI verstanden werden, sondern lediglich als Aufforderung für eine methodische Begleitforschung.

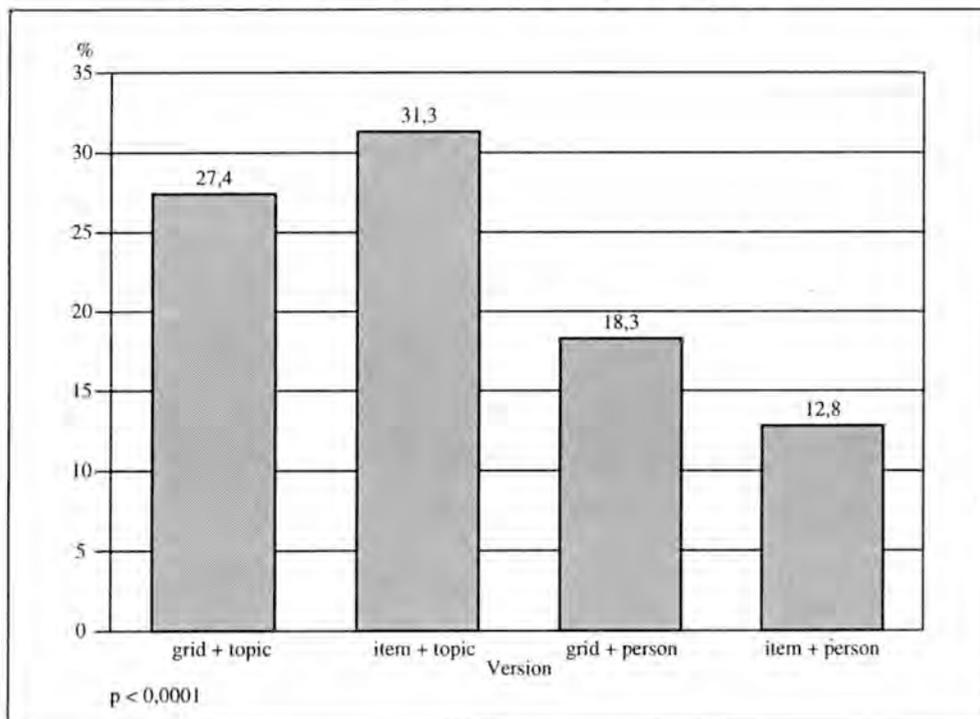
7) Die vier verglichenen Fragebogenvarianten sind im Abschnitt 2 über die Methoden beschrieben.

8) Eine Varianzanalyse ergibt, dass ein Drittel der erklärten Varianz auf das Bildschirmdesign zurückzuführen ist und zwei Drittel auf die Fragereihenfolge.

Zur Beantwortung der Frage, warum sich die vier Versionen hinsichtlich der Zeit unterscheiden, müssen wir zunächst auf ein Phänomen eingehen, das wir bereits bei früheren Untersuchungen feststellen konnten (Couper et al. 1997): Analysen haben gezeigt, dass der Interviewer für die Bearbeitung eines Items bei der ersten Person im Haushalt signifikant mehr Zeit benötigt als für die Bearbeitung des gleichen Items für die folgenden Personen in einem Haushalt. Dieses Phänomen haben wir mit dem Begriff Loopeffekt bezeichnet (Loop = Schleife). Der durchschnittliche Loopeffekt für alle Items beläuft sich auf 3,3 Sekunden für jedes einzelne Item, was einer Reduktion von etwa 38 % gegenüber der für die erste Person nötigen Dauer entspricht. Und wir können feststellen, dass sich für alle drei Items in unserem Haushaltsroster entsprechende Loopeffekte ergeben. In jedem einzelnen Fall benötigt der Interviewer zur Bearbeitung des Items für die erste Person im Haushalt substantiell mehr Zeit als für die Bearbeitung des gleichen Items bei den folgenden Personen im Haushalt. Interessanterweise können wir nun feststellen, dass der Loopeffekt bei den gridbasierten themenorientierten Versionen unseres Experimentes signifikant stärker ausgeprägt ist als bei den anderen Fragebogenvarianten, bei denen nur sehr geringe und zum Teil überhaupt keine Loopeffekte feststellbar sind. Die durchschnittlich geringere Dauer der Bearbeitung der gridbasierten themenorientierten Version ergibt sich daher – so unsere Interpretation – aus dem signifikant größeren Loopeffekt für diese Fragebogenvariante (genauere Daten und eine ausführliche Diskussion finden sich bei Fuchs 1999).

Diese Unterschiede sind aus unserer Sicht darauf zurückzuführen, dass sich sowohl Interviewer als auch Befragte bei Nutzung der gridbasierten themenorientierten Variante anders verhalten als unter den Bedingungen der Standardversion (itembasiert und personenorientiert). Der Unterschied besteht unseres Erachtens darin, dass die Segmentierung des Interviewverlaufs in Einzelfragen partiell aufgehoben wird: Betrachten wir zunächst das Befragtenverhalten, so stellen wir fest, dass die themenorientierte Fragereihenfolge zu einem höheren Anteil von Interviews führt, in dem der Befragte zumindest einmal im Haushaltsroster die Informationen für alle Personen im Haushalt oder für eine Gruppe von Personen im Haushalt gemeinsam mitteilt (z. B.: „Wir kamen alle im gleichen Jahr nach Deutschland“ oder „Wir sind alle Aussiedler“). Nach diesen Ergebnissen scheint der Bildschirmaufbau des Fragebogens also auch das Befragtenverhalten zu tangieren. Dies ist überraschend, da die Studie als computergestützte telefonische Befragung durchgeführt wurde und der Befragte eigentlich keine Gelegenheit gehabt haben sollte, den Aufbau des Bildschirms zur Kenntnis zu nehmen. Die einzige mögliche Erklärung besteht darin, dass der Interviewer sein Verhalten gegenüber den Befragten entsprechend dem Bildschirmaufbau modifiziert und dadurch unterschiedliche Stimuli an den Befragten sendet, auf die dieser mit entsprechenden Verhaltensweisen reagiert.

Abbildung 8
Nichtstandardmäßiges Interviewerverhalten nach Version



Und dies bestätigt sich bei einer weiteren Analyse der durch Codierung von Videoaufzeichnungen gewonnenen Daten. Etwa 78 % aller Items im Haushaltsroster werden vom Interviewer mehr oder weniger einwandfrei administriert. Bei den verbleibenden 22 % der Items hingegen treten spezifische nichtstandardmäßige Verhaltensweisen auf: 9,3 % der Items werden vom Interviewer nicht verlesen; bei 5 % der Items verliert der Interviewer nicht die Frage, sondern statt dessen einen Stimulus, der sich auf die Verwandtschaftsbeziehung der in Frage stehenden Person zum Befragten richtet (z. B. „... und Ihre Frau?“). In weiteren 5,5 % der Fälle modifiziert der Interviewer die Frage dahingehend, dass sie als Bestätigung formuliert wird („... und Ihre Frau ist 32 Jahre alt, sagten Sie?“). Weitere unvollständige Frageformulierungen und Veränderungen des Fragetextes kommen hinzu. Dies summiert sich zu 22 % aller Items, die durch mindestens ein nichtstandardmäßiges Interviewerverhalten gekennzeichnet sind.⁹⁾

Diesbezüglich beobachten wir große Differenzen zwischen den einzelnen Varianten: In der gridbasierten themenorientierten Version sowie der itembasierten themenorientierten Version finden wir jeweils deutlich mehr nichtstandardmäßige Interviewerverhaltensweisen als in den

9) Dies ist – eingedenk der Tatsache, dass die Interviewer sich bewusst waren, dass ihr Verhalten während des Interviews auf Video aufgezeichnet wird – als erstaunlich hohen Wert anzusehen. Wir werden noch darauf zurückkommen, ob diese Verhaltensweisen die Datenqualität beeinträchtigen oder nicht.

beiden personenorientierten Varianten. Auch hier hat neben der Fragereihenfolge der Bildschirmaufbau einen eigenständigen Einfluss auf den Interviewverlauf.

Tabelle 1: Befragtenverhalten nach Fragebogenvariante

%

	grid + topic	item + topic	grid + person	item + person	Insgesamt
Befragter gibt die Antworten für alle Personen im Haushalt auf einmal	38,2	45,3	29,5	10,8	29,8
Phänomen tritt nicht auf	61,8	54,7	70,5	89,2	70,1
Insgesamt	100	100	100	100	100

p < 0,001

Aus unserer Sicht belegen sowohl die Befunde zum Interviewerverhalten wie auch zum Befragtenverhalten die These, dass die gridbasierte themenorientierte sowie die itembasierte themenorientierte Version einen Interviewverlauf begünstigen, bei dem beide Interaktionspartner von dem stark segmentierten Vorgehen der Standardversion (itembasiert personenorientiert) Abstand nehmen. Statt jedes einzelne Item unabhängig von den anderen Items zu bearbeiten, steht das Einzelitem bei der Administrierung im Kontext der zuvor und der nachfolgend zu stellenden Fragen.¹⁰⁾

4 Diskussion

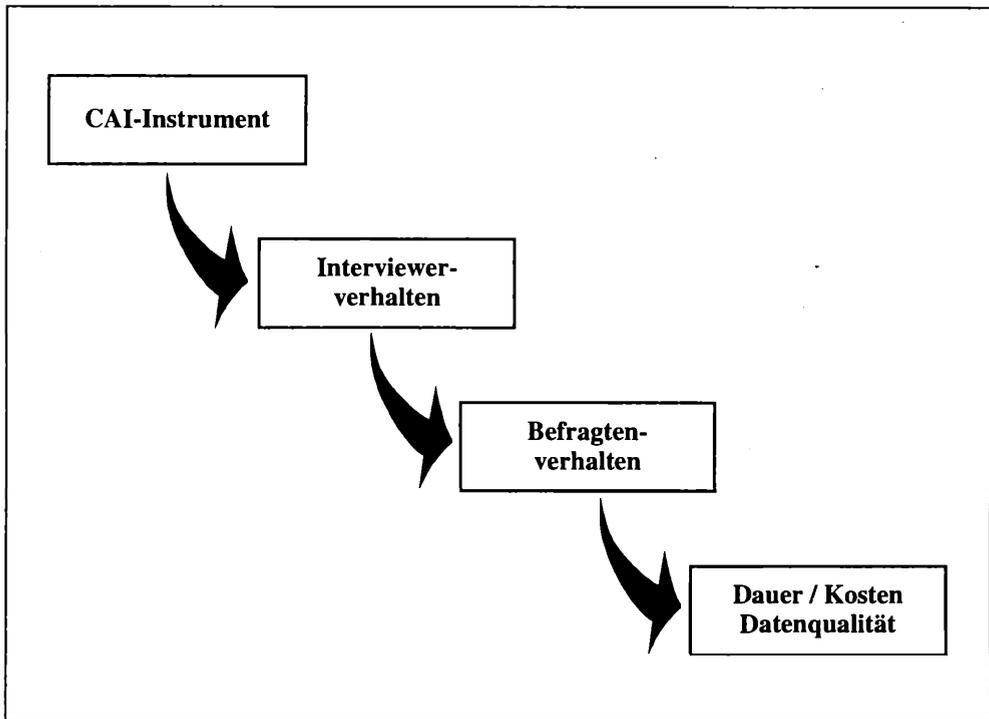
Die Ergebnisse zeigen, dass die Gestaltung des Bildschirminterfaces die Verhaltensweisen der beteiligten Akteure beeinflussen kann, was u. a. auch Konsequenzen für die Interviewdauer hat. Dabei gehen wir davon aus, dass einzelne Aspekte des Interfaces (nicht mit der Frageformulierung usw. zusammenhängende Merkmale) das Verhalten des Interviewers tangieren, indem sie in unterschiedlichem Ausmaß Aktionen zur Vermeidung von Redundanz und Segmentation ermöglichen bzw. stimulieren. Als Reaktion auf diese Variationen beobachten wir auch beim Befragten unter bestimmten Umständen Verhaltensänderungen. Im Kern besteht die Ursache für diese Modifikationen darin, dass sowohl Interviewer wie auch Befragter die Einzelitems stärker im Kontext der vorangegangenen und möglichen nachfolgenden Fragen sehen und der Interviewfluss dadurch glatter und mit weniger störenden, redundanten verbalen Beiträgen von beiden Seiten abläuft. Ob dies Auswirkungen auf die

10) Ein möglicher Nachteil dieses veränderten Interviewflusses besteht darin, dass die Datenqualität negativ beeinflusst wird. So wäre es denkbar, dass der Befragte über die Antworten bezüglich jeder einzelnen Person weniger intensiv nachdenkt, da ihm die vollständige und einwandfreie Frageformulierung seltener präsentiert wird als bei der gemäß Interviewregeln administrierten Fassung. Wir beobachten in unserem Datensatz nur sehr wenige Item Missing Values, weshalb die Berechnung von Standardindikatoren für die Datenqualität nicht in Frage kommt. Daher haben wir die Homogenität der Antworten für alle Haushaltsmitglieder berechnet. Darunter verstehen wir die durchschnittliche Zahl von unterschiedlichen Antwortausprägungen auf eine Frage für alle Haushaltsmitglieder. Dabei zeigen sich weder für das Ankunftsyear, das 19 verschiedene Antwortausprägungen annehmen kann, noch für den formalen Status, der vier Kategorien aufweist, signifikante Unterschiede bei der Homogenität. Wir ziehen daraus den Schluss, dass sich zumindest aus diesem Indikator keine Hinweise auf eine geringere Qualität derjenigen Daten ableiten lassen, die mit einer gridbasierten, themenorientierten Programmfassung erhoben werden.

Datenqualität hat – z. B. beim Item-Nonresponse oder bei den substantiellen Antworten – ist eine bisher weitgehend offene Frage. Die wenigen uns vorliegenden Daten (vgl. Fußnote 10 in diesem Beitrag, S. 87) geben jedoch keinen Hinweis darauf. Eine gründliche Analyse dieses Aspekts steht aber noch aus (vgl. Abbildung 9, S. 89).

Zur Beantwortung dieser und weiterer Forschungsfragen schlagen wir eine Perspektive vor, bei der die Gestaltung von computergestützten Fragebogen aus unterschiedlichen Blickwinkeln bewertet wird: neben dem Forscher sind dies der Interviewer und der Befragte. Natürlich muss ein computergestütztes Instrument in erster Linie den Bedürfnissen des Forschers entsprechen, damit valide und reliable Messungen durchgeführt werden können. Dazu ist es notwendig, dass der Frage-Antwort-Prozess für jedes einzelne Item detailliert geplant wird. Daneben aber muss auch die soziale Dimension der Interviewer-Befragten-Kommunikation berücksichtigt werden. Wenn das computergestützte Befragungsinstrument diese soziale Dimension stört, besteht die Gefahr, dass dies den Interviewverlauf, die Interviewdauer und möglicherweise auch die Datenqualität der erhobenen Daten beeinträchtigt. Daher gilt es, mehr über die sozialen und kommunikativen Anforderungen an einen Interviewer in der konkreten Befragungssituation zu lernen. Denn er oder sie hat nicht nur den Standardinterviewerregeln zu folgen, sondern zudem die Aufgabe, eine soziale Kommunikationssituation aufrechtzuerhalten; und eine solche Situation erfordert unter Umständen ergänzend zur neutralen Administration eines Frage-Antwort-Prozesses zusätzliche Verhaltensweisen, die je nach Gestaltung des Befragungsinstruments leichter oder schwieriger an den Tag gelegt werden können. Aber wir müssen auch die Perspektive des Befragten einnehmen, der sich durch stark segmentierte Instrumente unter Umständen nicht ausreichend ernst genommen fühlt und auf stupide und ermüdende Fragefolgen unter Umständen mit Abbrüchen oder mit wenig konzentriertem Antwortverhalten reagiert (Satisfying).

Abbildung 9
Technologieeffekte: Zusammenhang zwischen CAI-Bildschirmdesign
und Datenqualität



Zur Verfolgung dieser Analysestrategie ist aus unserer Perspektive ein kombiniertes Vorgehen bestehend aus stark kontrollierten Laborstudien und Feldexperimenten notwendig. Zwar bieten Laborstudien den Vorteil, einzelne Faktoren isolieren zu können. Aber ohne ein Design in einer echten Interviewsituation dem Druck einer realen Kommunikation auszusetzen, werden wir wenig darüber erfahren, in welcher Weise es das Verhalten des Interviewers – und indirekt auch des Befragten – tangiert. Daher sollte die Usabilityforschung als verschränkter Prozess von Laborexperimenten und Feldtests angesehen werden.

Literaturhinweise

Baker, R. P.; Bradburn, N. M.; Johnson, R. A. (1995): Computer-assisted personal interviewing: An experimental evaluation of data quality and costs, in: *Journal of Official Statistics*, 11 (4), S. 415 – 431.

Couper, M. P.; Groves, R. M.; Kosary, C. (1989): Methodological issues in CAPI, in: *American Statistical Association (Ed.): Proceedings of the Section on Survey Research*, S. 349 – 354.

Couper, M. P.; Burt, G. (1994): Interviewer attitudes toward computer-assisted personal interviewing (CAPI), in: *Social Science Computer Review* 12 (1), S. 38 – 54.

Couper, M. P.; Fuchs, M.; Hansen, S. E.; Sparks, P. (1997): CAPI Instrument design for the Consumer Expenditure (CE) Quarterly Interview Survey. Final Report, University of Michigan.

Couper, M. P.; Baker, R. P.; Bethlehem, J.; Clark, C. Z. F.; Martin, J.; Nicholls, W. L.; O'Reilly, J. (Eds., 1998): Computer assisted survey information collection, New York.

Fuchs, M. (1994): Umfrageforschung mit Telefon und Computer. Einführung in die computergestützte telefonische Befragung, Weinheim.

Fuchs, M. (1995): Die computergestützte telefonische Befragung. Einige Antworten auf Probleme der Umfrageforschung, in: *Zeitschrift für Soziologie*, 24, S. 284 – 299.

Fuchs, M. (1999): Screen design and interviewer behavior. Results from a field experiment. Paper presented at the 54. annual meetings of the American Association for Public Opinion Research, Mai 1999, St. Pete Beach, FL (USA).

Fuchs, M.; Couper, M. P.; Hansen, S. E. (2000): Technology Effects. Why does a CAPI interview take longer? in: *Journal of Official Statistics* (forthcoming).

Groves, R. M.; Mathiowetz, N. A. (1984): Computer assisted telephone interviewing: Effect on interviewers and respondents, in: *Public Opinion Quarterly*, 48, S. 356 – 369.

Hansen, S. E.; Fuchs, M.; Couper, M. P. (1997): CAI instrument usability testing. Paper presented at the Annual Meeting of the AAPOR, Norfolk, VA, May 1997.

Hansen, S. E.; Couper, M. P.; Fuchs, M. (1998): Usability Evaluation of the NHIS Instrument. Paper presented at the Annual Meeting of the AAPOR, St. Louis, MO, May 1998.

House, C. C. (1985): Questionnaire design with computer assisted telephone interviewing, in: *Journal of Official Statistics*, 1(2), S. 209 – 219.

House, C. C.; Nicholls, W. L. (1988): Questionnaire design for CATI: Design objectives and methods, in: *Groves, R. M.; Biemer, P. P.; Lyberg, L. E.; Massey, J. T.; Nicholls, W. L.; Waksberg, J. (Eds., 1998): Telephone Survey methodology*, New York, S. 421 – 436.

Laurie, H.; Moon, N. (1997): Converting to CAPI in a longitudinal panel study. Working Papers of the ESRC Research Centre on Micro-Social Change, 97 – 11, Essex.

Lepkowski, J. M.; Couper, M. P.; Hansen, S. E.; Landers, W.; McGonagle, K. A.; Schlegel, J.; Wright, T.; Chevarley, F. (1998): CAPI Instrument Evaluation: Behavior Coding, Trace Files, and Usability Methods. Paper presented at the annual meeting of the AAPOR, May 1998.

Martin, J.; O'Muircheartaigh, C.; Curtice, J. (1993): The use of CAPI for attitude surveys: An experimental comparison with traditional methods, in: *Journal of Official Statistics* 9 (3), S. 641 – 662.

Moore, J. C. (1996): Person- vs. topic-based design for computer-assisted household survey instruments. Paper presented at Inter CASIC '96, International conference on computer-assisted survey information collection, San Antonio, Tx.

Moore, J. C.; Moyer, H. L. (1998 a): ACS/CATI person-based/topic-based field experiment – final report. Center for Survey Methods Research, Bureau of the Census.

Moore, J. C.; Moyer, H. L. (1998b): Questionnaire design effects on interview outcomes. Paper presented at the Annual Meeting of the AAPOR, St. Louis, MO, May 1998.

Moyer, L. H. (1996): Which is better: Grid listing or grouped questions design for data collection in establishment surveys? in: *Proceedings of the American Statistical Association, Section on Survey Research Methods*, Alexandria, VA, S. 986 – 990.

Müller, S.; Kesselmann, P. (1996): Akzeptanz von computergestützten Erhebungsverfahren. Ein empirischer Vergleich mit der traditionellen Fragebogentechnik, in: *Marketing, Zeitschrift für Forschung und Praxis* 18 (3), S. 191 – 202.

Nicholls, W. L.; House, C. C. (1987): Designing questionnaires for computer-assisted interviewing: a focus on program correctness, in: *Bureau of the Census (Ed.): Third annual research conference, proceedings*, Washington, DC.

Oksenberg, L.; Beebe, T.; Blixt, S.; Cannell, C. (1992): Research on the design and conduct of the National Medical Expenditure Survey interviews. Final Report. Survey Research Center, Ann Arbor, M.

Projektgruppe SOEP (1998): Funktion und Design einer Ergänzungsstichprobe für das Sozio-ökonomische Panel. Diskussionspapiere des DIW, 163, Berlin.

Schneid, M. (1991): Einsatz computergestützter Befragungssysteme in der Bundesrepublik Deutschland. Ergebnisse einer Umfrage, ZUMA-Arbeitsbericht 91/20, Mannheim.

Schober, M. F.; Conrad, F. G. (1997): Does conversational interviewing reduce survey measurement error? in: *Public Opinion Quarterly* 61, S. 576 – 602.

Sudman, S.; Bradburn, N. M.; Schwarz, N. (1996): Thinking about answers. The application of cognitive processes to survey methodology.

Suchman, L.; Jordan, B. (1990): Interactional troubles in face-to-face survey interviews, in: *Journal of the American Statistical Association* 85, S. 45 – 54.

Weeks, M. F. (1992): Computer-assisted survey information collection: A review of CASIC methods und their implications for survey operations, in: *Journal of Official Statistics*, Jg. 8, Nr. 4, S. 445 – 465.

Testmarktforschung

1 Aufgabe und Umfeld der Testmarktforschung

Lassen Sie uns zu Beginn festlegen, was der Terminus Testmarktforschung im Folgenden beinhalten soll:

Unter dem Begriff Testmarktforschung bzw. Testmarkt werden allgemein jene Testverfahren subsumiert, die kurz vor der nationalen Markteinführung von echten Produktinnovationen, Relaunches, Brand- und Line-Extensions eingesetzt werden. Der Testmarkt bildet das letzte Glied einer ganzen Kette von Tests, die in den verschiedenen Phasen des Launches und Relaunches von Angeboten eingesetzt werden.

Bevor man an die Konzeptauswahl gehen kann, sollte man idealtypischerweise eine Vorstellung von der Struktur des zu bearbeitenden Marktes haben, wissen, welche Chancenpotentiale und Erfolgsfaktoren existieren und welche Positionierung man anstrebt. Erst dann geht es darum, zündende Ideen zu entwickeln, die das Verbraucherinteresse wecken. Die Bündelung mehrerer Ideen ergibt dann ein konkretes Konzept, das sich abtesten lässt.

Abbildung 1



*) Heinrich A. Litzenroth, GfK Testmarktforschung GmbH, Nürnberg.

Im Konzepttest wird die Tragfähigkeit der spezifizierten Idee für das neue Produkt untersucht. Der Produkttest, in dem das fertige Produkt hinsichtlich der subjektiven Qualität bei einer repräsentativen Auswahl von Konsumenten der Zielgruppe überprüft wird, parallel durchgeführte oder vorgelagerte Partialtests (z. B. Markenname, Verpackung, Geschmack, Handhabung), Preis- und Werbemittelpretests bilden weitere wichtige Glieder dieser Testkette.

Während in diesen vorgelagerten Tests einzelne Elemente des Marketing-Mix weitgehend isoliert voneinander getestet werden, findet im Testmarkt die Überprüfung der kompletten Marketingkonzeption statt, d. h., das simultane Zusammenspiel aller Marketingvariablen wird unter Einbezug des Wettbewerbs einer Generalprobe unterzogen. Er dient dazu, das Absatzvolumen bzw. den im nationalen Markt erreichbaren Marktanteil zu prognostizieren und/oder eventuelle Schwächen im Marketingprogramm aufzudecken, um so das Risiko eines Flops nach der Markteinführung von Neuheiten deutlich zu reduzieren.

Die Kosten für den Aufbau von Produktionskapazitäten, Distribution, Bekanntheit und Image neuer Marken übertreffen im Bereich von „fast moving consumer goods“ die eigentlichen Entwicklungskosten häufig um ein Vielfaches. Zweistellige Millionenbeträge allein für die Werbestreuung sind keine Seltenheit. Auch für die Listung im Handel sind ähnliche Summen bereitzustellen. Echt neue Marken werden deshalb immer seltener eingeführt. Die Mehrzahl der Neuerungen sind Line-Extensions, Brand-Extensions und Relaunches.

Abbildung 2



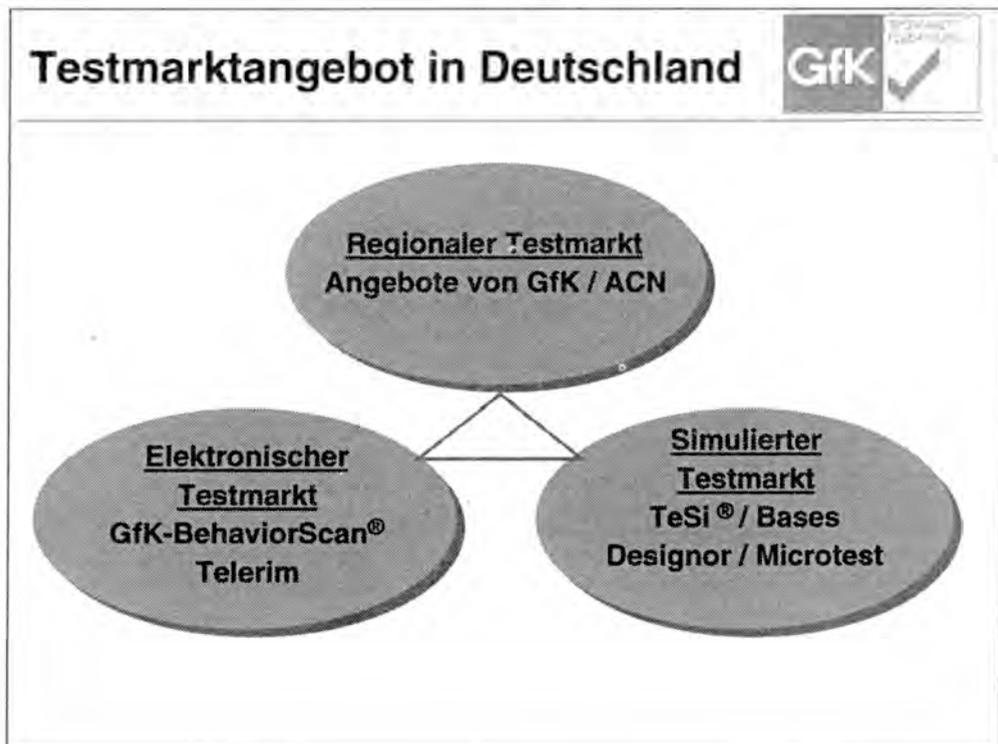
Bei einer weitgefassten Definition von neuen Produkten dürfte die Zahl der jährlichen Neueinführungen allein im deutschen Lebensmitteleinzelhandel zwischen 1 500 und 2 000 liegen. Rund 85 % dieser Neuheiten scheitern jedoch im Markt. Die Gründe für diesen hohen Anteil an Flops sind vielschichtig. Zu nennen sind beispielsweise:

- gesättigte Märkte,
- ein immer unberechenbarer werdender Konsument,
- die zunehmende Komplexität der Märkte,
- die sinkende Werbeeffizienz,
- die Informationsüberlastung,
- die Verkürzung der Produktlebenszyklen und
- die zunehmenden Restriktionen auf Handelsseite.

Die Reduktion des Risikos ist der Hauptnutzen, den die Testmarktforschung stiftet.

Der klassische Regionale Testmarkt, also die probeweise Vermarktung des zu überprüfenden Produktes in einem regional begrenzten Teilgebiet des Gesamtmarktes (z. B. Berlin, Saarland, Bremen, Hessen) hat aufgrund wachsender Probleme stark an Bedeutung verloren.

Abbildung 3



Deshalb konzentriert sich der Vortrag auf die in der Praxis sehr viel wichtigeren Verfahren, nämlich den Simulierten Testmarkt und den Elektronischen Testmarkt.

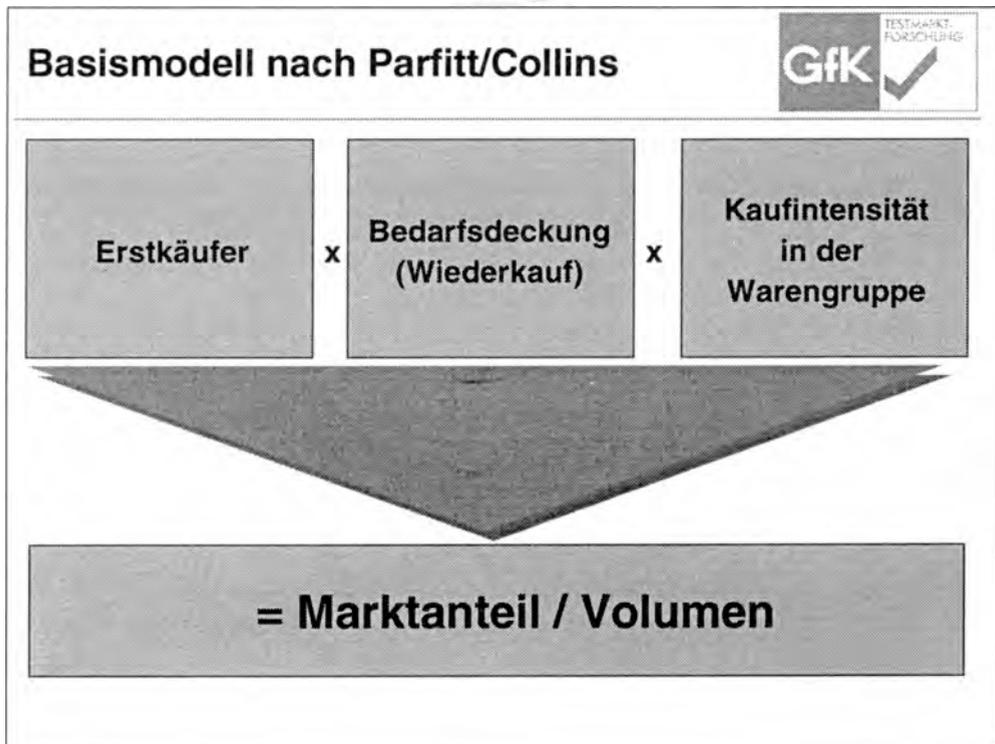
2 Der Simulierte Testmarkt

Beim Simulierten Testmarkt erfolgt die Datenerhebung nicht im realen Markt, sondern die Prognose des zu erwartenden Marktanteils bzw. Absatzvolumens sowie die Gewinnung weiterer diagnostischer Informationen erfolgt im Teststudio.

Dort wird für jede Testperson der Adaptionsprozess (Wahrnehmung \Rightarrow Erstkauf \Rightarrow Einstellungsbildung \Rightarrow Wiederkauf) simuliert.

Nahezu alle angebotenen Verfahren der Testmarktsimulation (BASES, DESIGNOR, Micro-Test, Sensor, TeSi[®]) basieren auf dem Modell von Parfitt/Collins und ermitteln einen langfristigen Marktanteil bzw. ein langfristig erreichbares Volumen.

Abbildung 4

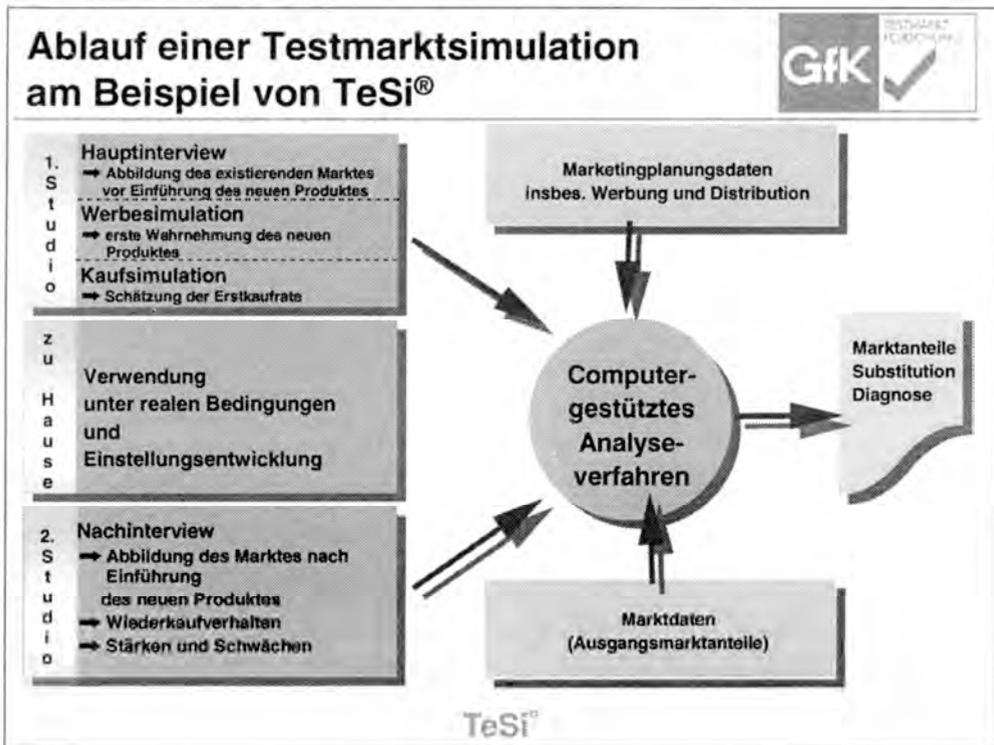


Der Ablauf einer Testmarktsimulation soll im Folgenden am Beispiel von TeSi[®] verdeutlicht werden.

TeSi[®] ermittelt den sogenannten Gleichgewichtsmarktanteil, den eine Neuheit nach der Stabilisierung des Penetrationsprozesses erreicht. Dies ist der Fall, wenn sich Distribution und gestützte Markenbekanntheit stabilisiert haben und das neue Angebot vom Wiederkauf getragen wird. Dabei werden die Marktanteilskomponenten Erstkaufpenetration, Bedarfsdeckung und Kaufintensität einzeln bestimmt.

- Das Erhebungsverfahren stellt eine Kombination aus Studio- und Home-Use-Test dar. In der Umgebung des Teststudios werden in der Regel 300 Verwender der betreffenden Produktkategorie angeworben und einzeln ins Teststudio eingeladen.
- Während des Hauptinterviews wird zunächst der existierende Markt vor Einführung des neuen Produktes abgebildet. Über die Ermittlung von Markenbekanntheit und Markenverwendung wird zunächst das sogenannte Relevant Set der Testpersonen bestimmt. Es beinhaltet jene Menge von Marken, die für die jeweilige Testperson bei der Kaufentscheidung relevant sind. Für diese Marken werden dann mit Hilfe eines sogenannten Chip Games (= Konstant-Summen-Methode) Präferenzen erhoben. Zudem werden Einstellungsdaten durch Eigenschaftsbeurteilungen gewonnen. Fragen zur Verwendungintensität in der betreffenden Warengruppe und zu den soziodemographischen Merkmalen der Testpersonen beschließen das Hauptinterview.
- Anschließend wird jeder Testperson ein Werbeblock vorgeführt, der Werbung für die wichtigsten Konkurrenzprodukte sowie auch einen Werbespot für das neue Produkt enthält. Um Stellungseffekte zu vermeiden, wird die Reihenfolge der Spots variiert. In dieser Werbesimulation erfolgt damit die erste Wahrnehmung des neuen Produkts.
- Die darauf folgende Kaufsimulation in einem Miniatur-Markt im Studio dient primär zur Schätzung der Erstkauftrate. Im Umfeld der relevanten etablierten Marken wird auch das neue Produkt angeboten. Bereits beim Empfang im Studio wurde jeder Testperson ein Geldbetrag ausgehändigt, der den Preis des teuersten Produktes übersteigt, um die erforderliche Kaufkraft sicherzustellen.

Abbildung 5



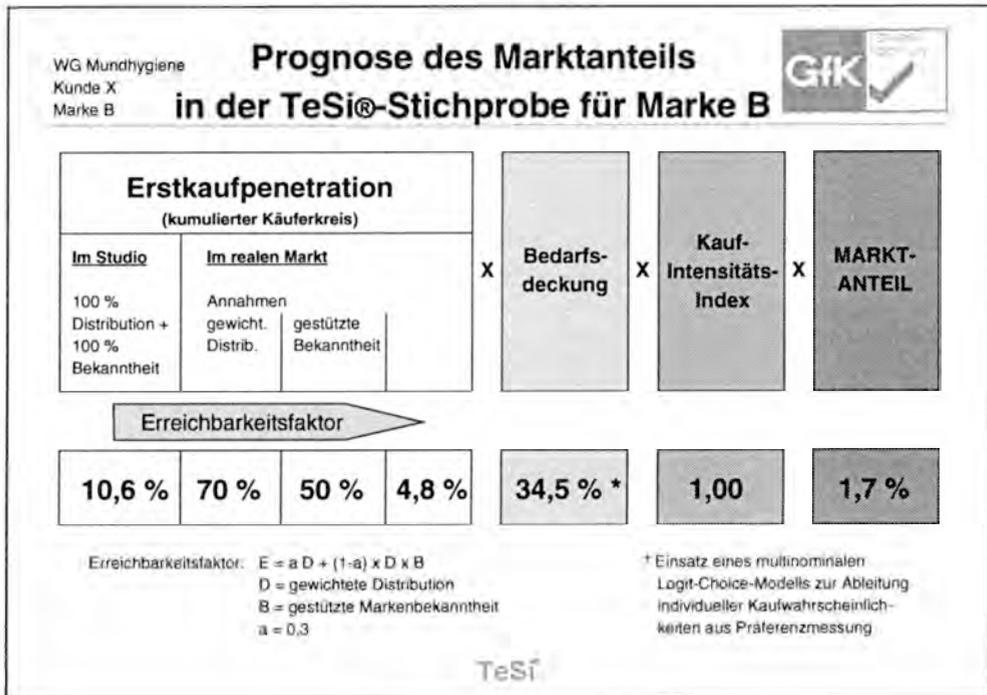
Nach dem Kauf wird die Testperson gefragt, welche Marke sie gekauft hätte, wenn die zuvor gewählte nicht vorhanden gewesen wäre. Dieser Vorgang wird sofort wiederholt, wie die Person Marken in ihrem Relevant Set hat. Dadurch wird u. a. die unterschiedliche Markentreue der Verbraucher berücksichtigt.

- Um die Verwendung des Testprodukts im direkten Vergleich mit der jeweiligen Stamm-marke zu ermöglichen, erhalten jene Konsumenten, die das Testprodukt nicht gekauft haben, dasselbe als Geschenk. Alle anderen bekommen die an zweiter Stelle gewählte Marke geschenkt. Die Testpersonen wissen nicht, welches das Testprodukt ist, geschweige denn, dass ein neues Produkt getestet wird. Sie verwenden die erhaltenen Produkte zu Hause über einen Zeitraum von einer bis zu mehreren Wochen, je nach Verwendungsdauer bzw. Kauffrequenz in der betreffenden Warengruppe. Dieser Home-Use-Test gibt den Konsumenten die Gelegenheit, das neue Produkt unter realen Bedingungen kennen zu lernen und ihm gegenüber eine Einstellung zu entwickeln.
- Nach Abschluss der Home-Use-Phase werden die Testpersonen ein zweites Mal ins Teststudio eingeladen. Im Nachinterview werden die Präferenz- und Einstellungsmessungen in gleicher Weise wie im Hauptinterview wiederholt, um so die Daten für die Prognose des Wiederkaufverhaltens zu ermitteln. Außerdem werden ganz am Ende des Tests offene Fragen zu Verwendungserfahrungen sowie Likes und Dislikes gestellt, um eine Diagnose der Stärken und Schwächen des neuen Produktes vornehmen zu können.

- Die im Rahmen dieses dreistufigen Testablaufs gewonnenen Daten werden mit Hilfe eines computergestützten Analyseverfahrens, welches auf zahlreiche Methoden und Modelle zurückgreift, verarbeitet. In Verbindung mit Marketingplanungsdaten und aktuellen Marktdaten gelangt man so zu Prognosen, die zeigen, wie sich der Markt durch die Einführung des neuen Produktes verändert und wie sich das neue Produkt darin behauptet. Insbesondere werden Marktanteile und Substitutionseffekte geschätzt. Diagnostische Informationen, die Stärken und Schwächen aufzeigen, Positionierungs-, Präferenz- und Segmentierungsanalysen runden das Bild ab.

Das folgende neutralisierte Beispiel zeigt die TeSi®-Marktanteilsschätzung für die Neueinführung einer Spezialzahncreme, die dazu beitragen soll, die Dachmarke zu revitalisieren. Die neue Marke B ist preislich eindeutig im Premiumsegment angesiedelt.

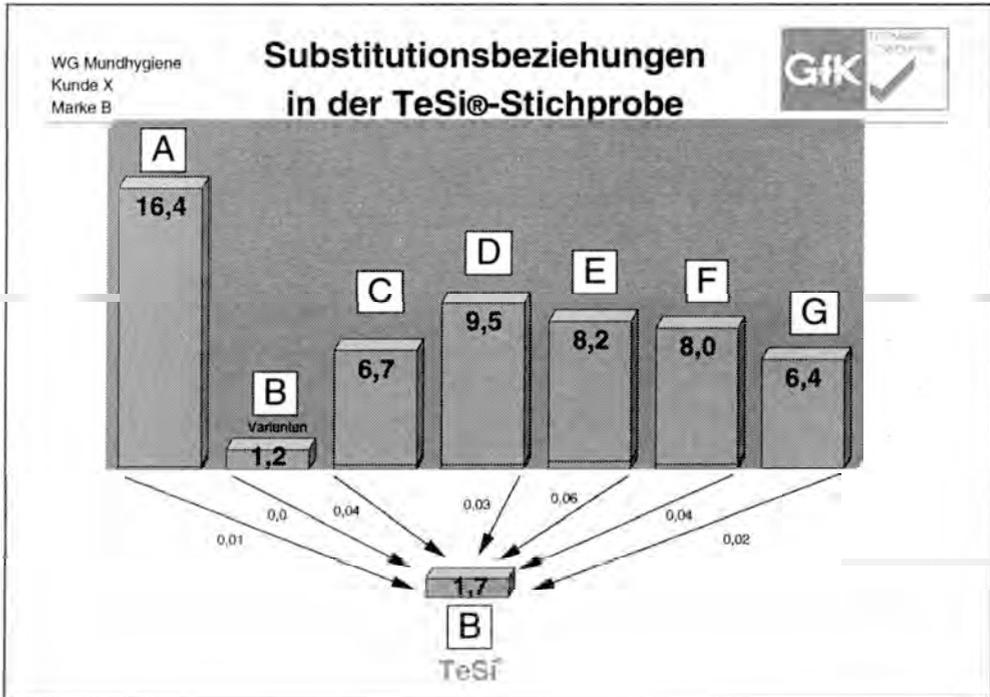
Abbildung 6



Im Studio, d. h. bei 100 % Distribution und 100 % Bekanntheit, erreichte die neue Marke einen weitesten Käuferkreis von 10,6 %. Geht man von realisierbaren Werten für die Distribution von 70 % und die gestützte Markenbekanntheit von 50 % aus, so ergibt sich ein weitester Käuferkreis von 4,8 %. Bei einer Bedarfsdeckungsrate von 34,5 % und einer mengenmäßig durchschnittlichen Verbrauchsintensität in der Warengruppe von 1,00 ergibt sich ein vorläufiger Marktanteil von 1,7 %.

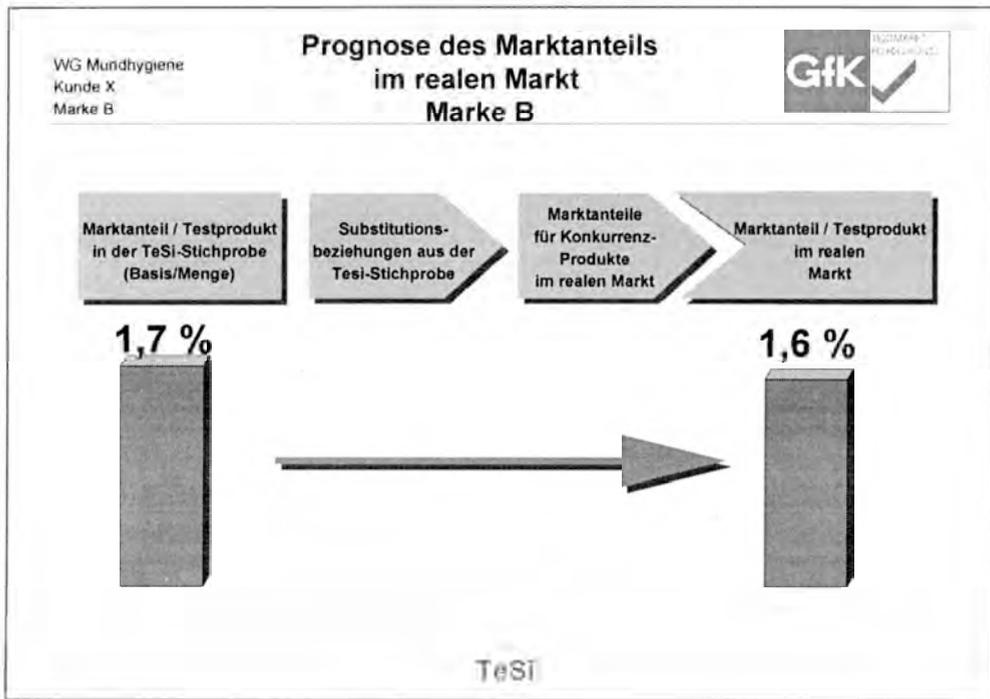
Aufgrund der Marktanteile im „TeSi®-Markt“, die sich vor und nach der Neueinführung ergeben, lassen sich die Substitutionsbeziehungen zwischen dem neuen Produkt und dem übrigen Angebot ermitteln.

Abbildung 7



Die aus der TeSi®-Stichprobe ermittelten Substitutionsbeziehungen lassen sich auf die „tatsächlichen“ Marktanteile der etablierten Marken übertragen, soweit sie aus Paneluntersuchungen bekannt sind. Dadurch wird eine Justierung der Marktanteilsschätzung am realen Markt möglich.

Abbildung 8



Im realen Markt ergibt sich im vorliegenden Fall ein mengenmäßiger Marktanteil von 1,6 %. Aufgrund der Premiumpreispositionierung erreicht das neue Produkt wertmäßig sogar einen Marktanteil von 3,1 %. Es wird deutlich, dass die vorhandenen Varianten der Marke B ihre Position halten können und das Testprodukt Marktanteile von fast allen anderen Marken abschöpfen kann.

Ergänzend zu den bisher üblichen Face-to-Face-Interviews nutzt TeSi[®] in Deutschland jetzt auch die Möglichkeiten der Multimedia-Technologie und setzt damit neue Standards in der Marktforschung. Die Datengewinnung erfolgt nun durch innovative Human-Interface-Technologie. Dabei werden die Aufgaben des Interviewers durch ein multimediales System übernommen, und die Dateneingabe erfolgt über Touchscreen und Voice-Recording.

Abbildung 9



Erfahrungen mit dieser Technik bestehen auf Konsumentenseite bereits durch den Umgang mit Geldautomaten, den PCs sowie den Check-In-Automaten am Flughafen.

Abbildung 10

Die Vorteile der Human Interface Technologie



- ▶ **Optimale Datenqualität**
 - höheres Involvement der Betrachter
 - Unterschätzung der Interviewdauer
 - elektronische Plausibilitätschecks
 - > automatische Filterführung
 - Randomisierung von Bildvorlagen, Statements etc.
 - > Unmittelbare Sprachaufzeichnung bei offenen Fragen

- ▶ **Schnellere Verfügbarkeit der Daten**

TeSi

Wie Befragungen von Testteilnehmern belegen, macht der Umgang mit der neuen Technik tatsächlich Spaß, und – viel wichtiger – die tatsächliche Länge des Interviews wird subjektiv extrem unterschätzt.

Dabei ist es aber nicht nur so, dass die individuelle Interviewdauer verringert wird und unsere Befragten mit Interesse bei der Sache sind, sondern die Möglichkeit der elektronischen Plausibilitätschecks garantieren zudem ein Höchstmaß an Datenqualität.

Da alle Daten mittels des gleichen Mediums erhoben und gespeichert werden, werden Übertragungsvorgänge in andere Medien (Eingeben von Daten etc.) größtenteils vermieden und die größte aller Fehlerquellen – nämlich der Mensch – weitestgehend eliminiert.

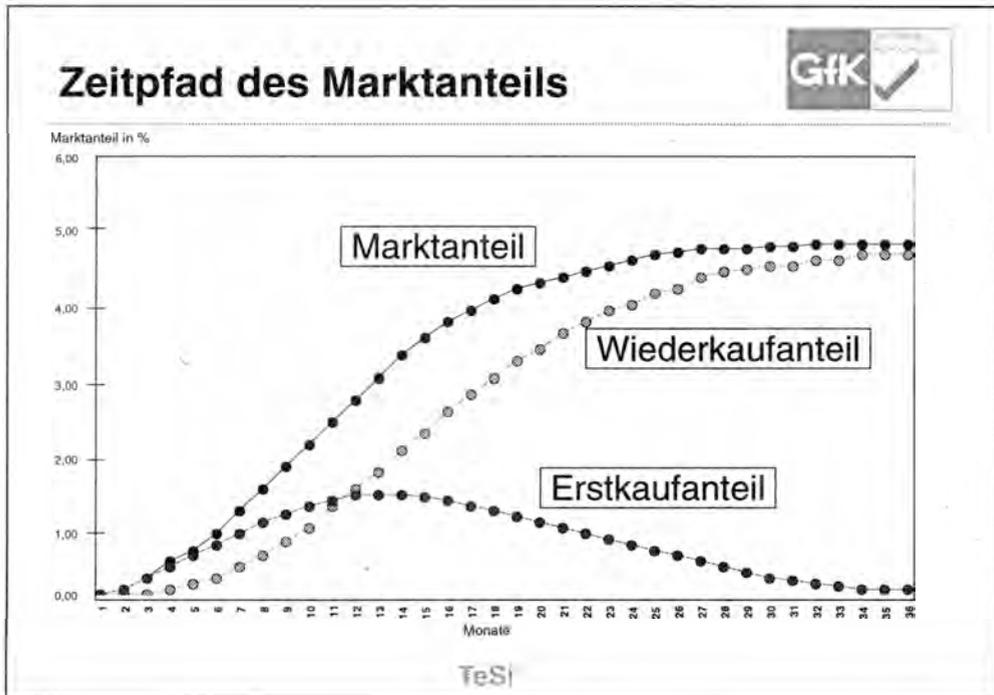
Natürlich nutzen wir noch weitere Optionen, die intelligente interaktive Systeme vom technischen Prozedere her bieten, wie z. B.

- automatische Filterführung,
- die Randomisierung von Statements oder von Bildvorlagen etc.

Auch die Datenauswertung wird entsprechend beschleunigt. Die Antworten zu geschlossenen Fragen werden per Online-Verbindung auf dem Datenhighway nach Nürnberg gesendet und die per Voice-Recording aufgezeichneten Antworten auf Wechselplatten gespeichert.

Aus diesen Ausführungen und Beispielen heraus darf aber keineswegs der Eindruck entstehen, dass für uns die technische Komponente im Vordergrund gestanden hätte. Diesbezügliche Potentiale und Optionen werden zwar konsequent genutzt, doch liegt unser Fokus primär auf der Weiterentwicklung unserer bewährten Forschungsmethoden.

Abbildung 11



Auch vom Modell her haben wir TeSi[®] in jüngster Zeit verfeinert. Durch die Abbildung der Marktanteilsentwicklung im Zeitpfad erfolgt eine bessere Abdeckung der ersten Einführungsmonate.

Dieser Schritt wird erst möglich durch die Modellierung der Erstkäufe, die vor allem in den ersten Einführungsmonaten die größte Bedeutung haben. Die Wiederkäufe in Form unserer Bedarfsdeckungsrate greifen erst später im Zeitpfad und bilden den Kern des bisherigen TeSi[®].

Das Füllen dieser „Lücke“ zwischen Launch und dem prognostizierten Marktanteil erlaubt zum einen die Integration verschiedener Marketing-Mix-Variablen in das Modell – wie z. B. Promotions, Distributionsaufbau und Werbung. Zum anderen ergibt sich hieraus auch die Möglichkeit, ein zeitgenaues Controlling der Entwicklung des neuen Produktes im Markt vorzunehmen, das frühzeitig Auskunft über die Erreichung des Marketingzieles gibt. Hierbei besteht dann für Marketing und Vertrieb die Möglichkeit, ein „Fein-Tuning“ in der Einführungsphase vorzunehmen, wenn Marketingziele wie Distribution eben nicht wie geplant erreicht werden.

Abbildung 12

Berücksichtigung der Promotion-Aktivitäten


RESEARCH
INNOVATION

Standardrechnung des Erreichbarkeitsfaktors

$$E = 0,3 * \text{Distri.} * + 0,7 * \text{Aided Awareness}$$

Promotions erhöhen die Erreichbarkeit

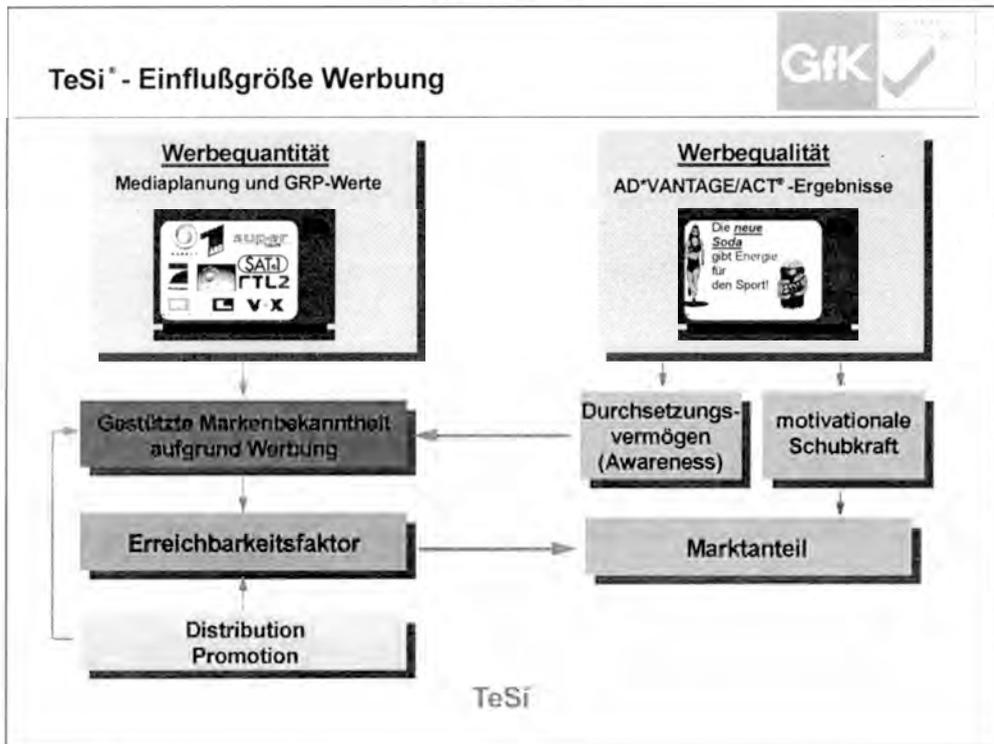
Neuer Ansatz für die Erreichbarkeit

$$E = 0,3 * \text{Distri.} * (1 - \text{gew. Distr. Dipl}) + \text{gew. Distr. Displ} + \\ + 0,7 * \text{Distr.} * \text{Aided Awareness} * (1 - \text{gew. Distr. Displ.})$$



Einen ersten Baustein, die dynamische Marktanteilsentwicklung exakter aufzeigen zu können, bildet die Integration von Promotionmaßnahmen; hierzu fließen neben den generell benötigten Plandaten – der gewichteten Distribution und der gestützten Bekanntheit – zusätzliche Informationen zu dem Themenkomplex Distribution von Displays über den Erreichbarkeitsfaktor in das Modell ein.

Abbildung 13

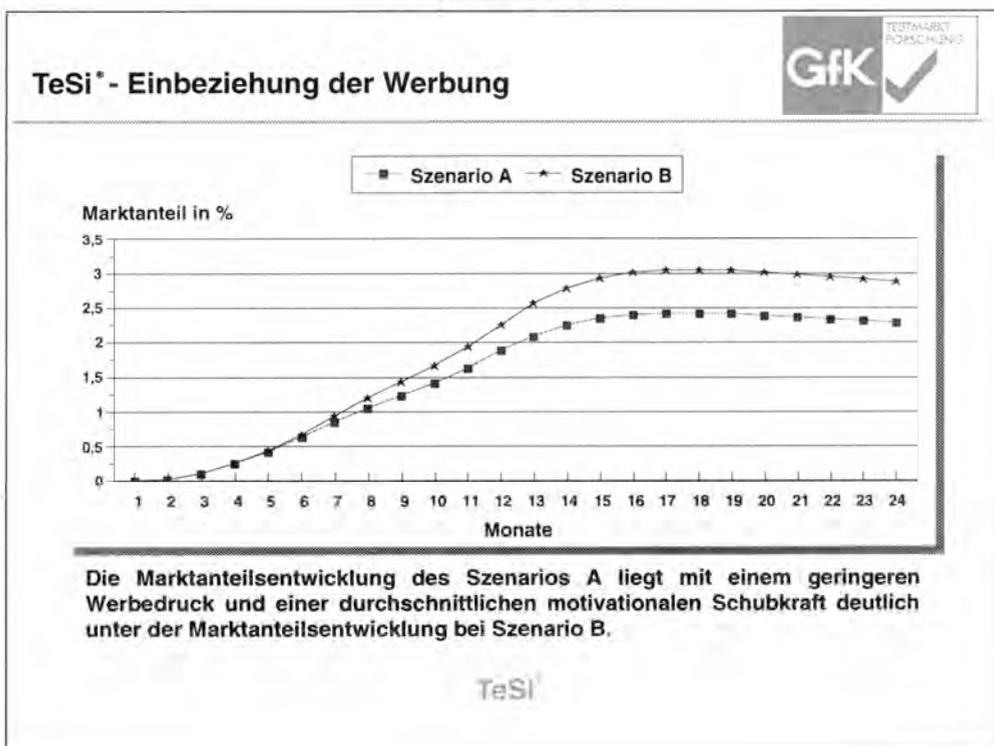


Weiterhin kann die Werbequalität und die Werbequantität Berücksichtigung im Modell finden.

Die Werbequantität – ausgedrückt durch GRPs (= Gross Rating Points) oder Werbespendings – wirkt über die Mediaplanung auf die Markenbekanntheit und damit auf den Erreichbarkeitsfaktor.

Daneben müssen noch Informationen über die Durchsetzungsfähigkeit und die motivationale Schubkraft des TV-Spots aus Werbemitteltests – wie z. B. AD*VANTAGE® – vorliegen. Zu beachten ist hierbei, dass die Einbeziehung von Werbequantität und Werbequalität nur gemeinsam möglich ist.

Abbildung 14



Auf dem vorliegenden Schaubild ist gut zu erkennen, dass die Marktanteilsentwicklung des Szenarios A mit einem geringeren Werbedruck und einer durchschnittlichen motivationalen Schubkraft deutlich unter dem Szenario B mit einem höheren Werbedruck und überdurchschnittlichen Werten bei Durchsetzungsfähigkeit und motivationaler Schubkraft liegt.

Übersicht

	Szenario A	Szenario B	Normwerte der Warengruppe Ad*Vantage/Act
Werbedruck in Mill. DM (2 Jahre)	19,1	65,1	/
Durchsetzungsfähigkeit	65	80	53
Motivationale Schubkraft	10	15	8

Abbildung 15



TeSi® - Einführung eines Konkurrenzproduktes

Modellvorstellung:
Der Erfolg einer Produktinnovation ist u.a. abhängig von:

- ▶ **Der Kraft der Marke (Markeneffekt)**
- ▶ **Grad der Innovation (Innovationseffekt)**

Zu berücksichtigende Konkurrenzprodukte müssen eine vergleichbare Innovation aufweisen wie getestetes Produkt

TeSi®

Da die Einführung eines Me-Too-Produktes der Konkurrenz (z.B. Persil Megaperls, Ariel Futur) den eigenen Markterfolg beeinträchtigen kann, kann im TeSi®-Modell dieses Szenario – die Einführung eines ähnlichen Konkurrenzproduktes als Antwort des Wettbewerbers auf die eigene Einführung – ebenfalls simuliert werden.

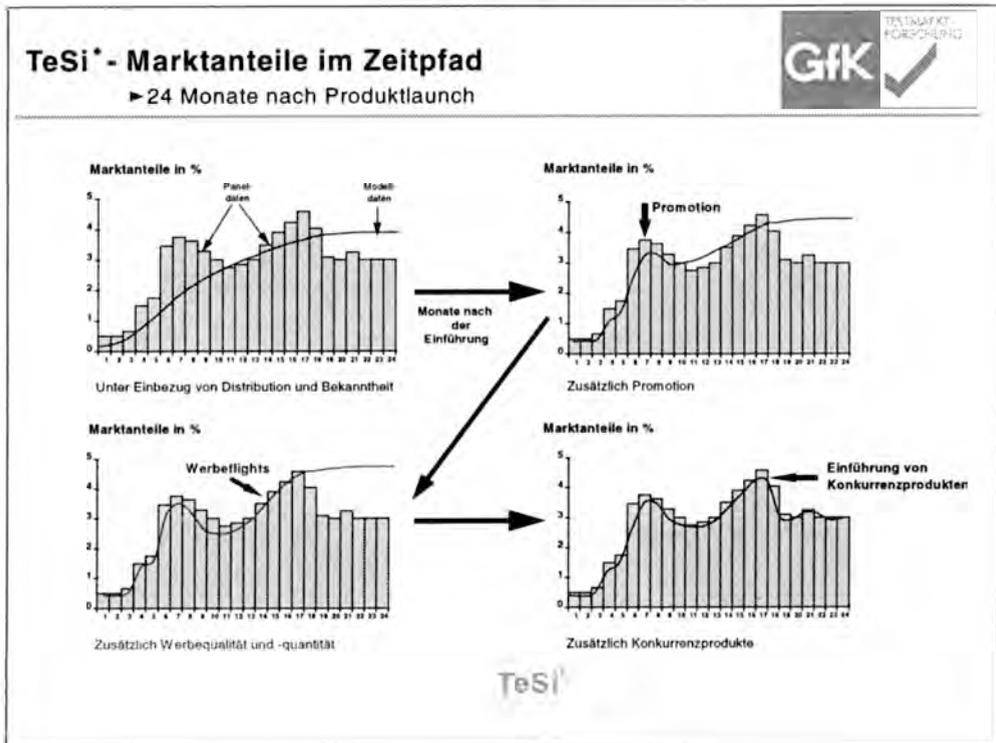
Hierbei geht die Modellvorstellung davon aus, dass der Erfolg einer Produktinnovation u.a. von den folgenden Komponenten abhängt:

- dem Markeneffekt – der Kraft der Marke und
- dem Innovationseffekt – dem Grad der Innovation.

Hierbei wirkt auf die Marke, für die das innovative Produkt eingeführt wird, sowohl der Marken- als auch der Innovationseffekt. Auf alle anderen Produkte im Referenzumfeld wirkt nur der Innovationseffekt. Über die Veränderung der Bedarfsdeckungsrate wird der Einfluss einer Konkurrenzeinführung auf das Testprodukt ex post ermittelt.

Zu berücksichtigende Konkurrenzprodukte müssen hierbei jedoch eine vergleichbare Innovation aufweisen wie das getestete Produkt.

Abbildung 16



Lassen Sie uns abschließend eine Prognose über 24 Monate, die aufgrund von TeSi®-Ergebnissen aufgestellt wurde, unter Einbezug der unterschiedlichen Variablen betrachten. Die Prognosedaten werden hierbei mit den tatsächlichen Paneldaten verglichen.

Bei der ersten Kurve, die Sie hier sehen, basiert die Prognose lediglich auf Distribution und Bekanntheit. Die Ergebnisse weichen noch beträchtlich von den realisierten Paneldaten ab.

Die zusätzliche Aufnahme von Promotions gleicht die Modelldaten im Prognosezeitraum bereits recht genau an.

Eine weitere Verbesserung der Genauigkeit bringt die zusätzliche Einbeziehung von Werbequalität und Werbequantität.

Und: Die Berücksichtigung der Einführung eines Konkurrenzproduktes nähert die Modellkurve den Paneldaten optimal an.

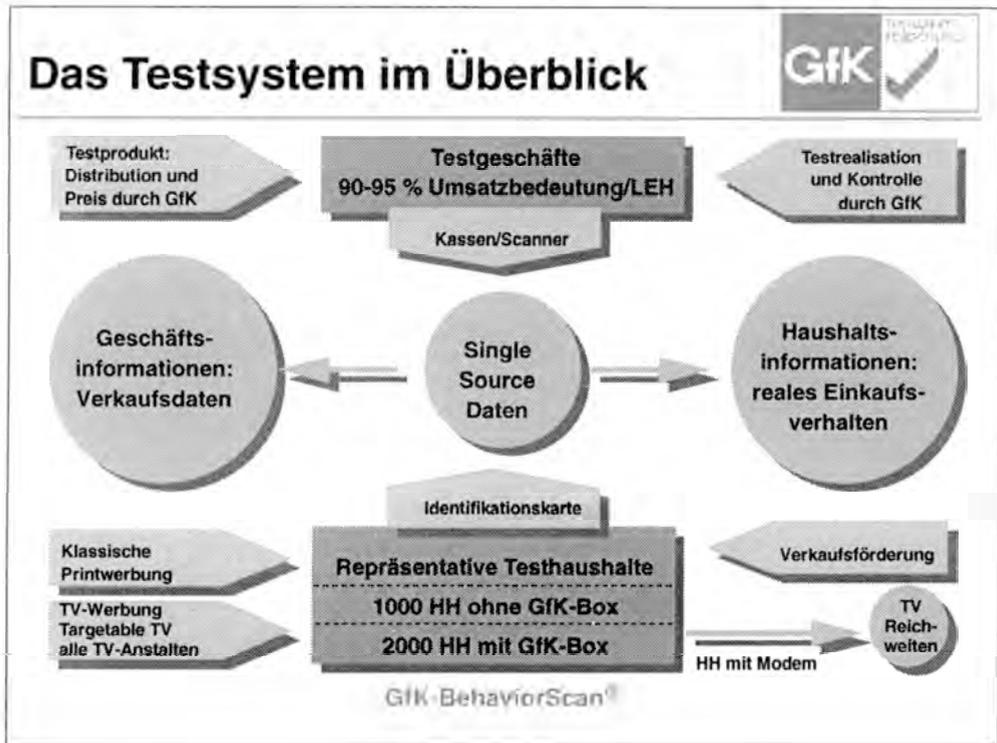
Erfahrene Nutzer von TeSi® erhalten die Möglichkeit, selbst unterschiedliche Marketing-Szenarien durchzuspielen. Eine CD-ROM mit einem benutzerfreundlichen Simulationsprogramm und den Daten aus dem Test hilft Ihnen, selbst weitere Marketingalternativen in ihrer Wirkung auf den Absatz des Testproduktes zu durchleuchten.

3 Der Elektronische Testmarkt

Der Elektronische Testmarkt ist ein lokaler Testmarkt, bei dem die Testgestaltung und Datenerfassung durch eine Kombination aus Haushalts- und Handelspanel erfolgt. Die Haushalte sind mit scannerlesbaren Identifikationskarten ausgestattet, so dass die Erfassung ihrer Einkäufe durch POS-Scanning in den Geschäften möglich wird. Die Einkäufe eines Haushaltes lassen sich somit komplett über die Zeit und verschiedene Einkaufsstätten aufzeichnen. Zudem besteht die Möglichkeit, die Konsumenten zu Hause mit Testwerbung zu versorgen.

Als Beispiel sei hier das Testsystem GfK-BehaviorScan[®] dargestellt.

Abbildung 17



Wesentliche Elemente des Testsystems sind die 3 000 repräsentativen Testhaushalte und die kooperierenden Testgeschäfte.

An allen mit grauen Pfeilen gekennzeichneten Stellen kann die GfK-Testmarktforschung experimentell „manipulierend“ bzw. kontrollierend Einfluss nehmen.

Neben den wichtigen Haushaltsinformationen, die das Kaufverhalten der Testhaushalte im Detail beschreiben, stehen insbesondere für die Promotionanalyse auch die Abverkaufsdaten aller Geschäfte zur Verfügung.

Haushaltsreichweiten gestatten zudem eine exakte Planung des Werbeeinsatzes im TV-Bereich.

Abbildung 18



Lokales Einzelhandelspanel mit hoher Umsatzabdeckung

Standort des Testsystems GfK-BehaviorScan® ist Haßloch im Rhein-Neckar-Raum in der Nähe von Ludwigshafen.

Alle relevanten Lebensmittel-Einzelhandelsgeschäfte bilden aufgrund entsprechender Kooperationsvereinbarungen die distributive Basis für die Durchführung von Marktexperimenten. Diese Geschäfte decken im Durchschnitt aller Warengruppen ca. 90 – 95 % des gesamten LEH-Umsatzes von Haßloch ab. Branchenspezifische Ergänzungen (z. B. Drogeriemarktsegment) sind ebenfalls verfügbar. Seit dem Frühjahr 1998 kann auch das Globus-SB-Warenhaus in Neustadt für Testprojekte genutzt werden.

Datenerfassung am Point-of-Sale

Die Erfassung der zentralen Daten erfolgt über die Kassen der kooperierenden Einzelhandelsgeschäfte. Zudem werden alle Wettbewerbsaktivitäten exakt registriert.

Abbildung 19



Das Herz des Instrumentes: Das reale Einkaufsverhalten von 3 000 Haushalten

Im Zentrum des Testsystems GfK-BehaviorScan® befindet sich ein Haushaltspanel, das hinsichtlich seiner soziodemographischen Struktur ein verkleinertes Abbild aller Haushalte in der Bundesrepublik darstellt. Es umfasst 3 000 repräsentative Testhaushalte, davon

- 2 000 Haushalte mit GfK-Box und
- 1 000 Haushalte ohne GfK-Box.

Die GfK-Box erlaubt den Empfang spezieller Testwerbung.

Die Haushalte kaufen wie gewohnt ein: Sie wissen nicht, welche Artikel Testprodukte sind. Beim Bezahlen legen sie an der Kasse eine Identifikationskarte mit haushaltsspezifischer Nummer vor. Dadurch können ihre Einkäufe im Kassenspeicher als „Testdaten“ gekennzeichnet werden.

Auf diese Weise lässt sich das Einkaufsverhalten ohne Belastung der Haushalte über die Zeit hinweg verfolgen. Das Vorzeigen der Identifikationskarten bei Abschluss eines jeden Einkaufsvorganges an der Kasse wird durch ein System aus Motivations- und Kontrollelementen sichergestellt.

Jährliche Struktur-, Einstellungs- und Besitzstandserhebungen bei allen Testhaushalten liefern zusätzliche Informationen, die sich mit dem Einkaufsverhalten korrelieren lassen.

Abbildung 20



Testmarketingplan

Vor Beginn der Testdurchführung wird auf der Grundlage der aktuellen Distributions- und Preisstruktur in den Testgeschäften ein Testmarketingplan erstellt, der alle vorgesehenen Marketingmaßnahmen sowie deren zeitlichen Einsatz fixiert.

Dieser Testplan stellt dabei ein verkleinertes Abbild der in Aussicht genommenen nationalen Marketingkonzeption dar.

Präzise Realisation

Mitarbeiter der GfK sorgen vor Ort für die Realisation und die ständige Kontrolle der Testbedingungen. Diese Mitarbeiter steuern auch den Warenfluss zwischen Testgeschäften und dem zum Testsystem gehörigen Lager.

Zentrale Kennziffern, die das Verbraucherverhalten beschreiben, gestalten eine detaillierte Antwort auf die Testfragestellung.

Abbildung 21



Experimentelle Eingriffsmöglichkeiten

Die GfK-Testmarktforschung kann sowohl auf der Geschäftsebene (z. B. Platzierung von Testprodukten, Durchführung von Verkaufsförderungsmaßnahmen) als auch auf der Haushaltsebene (Werbung und Verbraucherpromotions) experimentell manipulierend eingreifen.

Bis auf den Hörfunk können alle wichtigen Medien als Träger für kommunikationspolitische Maßnahmen eingesetzt werden.

Im TV-Sektor stehen alle relevanten Kanäle zur Verfügung.

Alle 3 000 Korrespondenz-Haushalte bekommen wöchentlich die Programmzeitschrift „HÖRZU“ in die Briefkästen. Deshalb ist eine adressgenaue Auswahl jener Haushalte möglich, die mit Zeitschriftenwerbung angesprochen werden sollen.

Andere Printmedien sind ebenfalls einsetzbar, lassen sich jedoch nicht splitten.

Abbildung 22

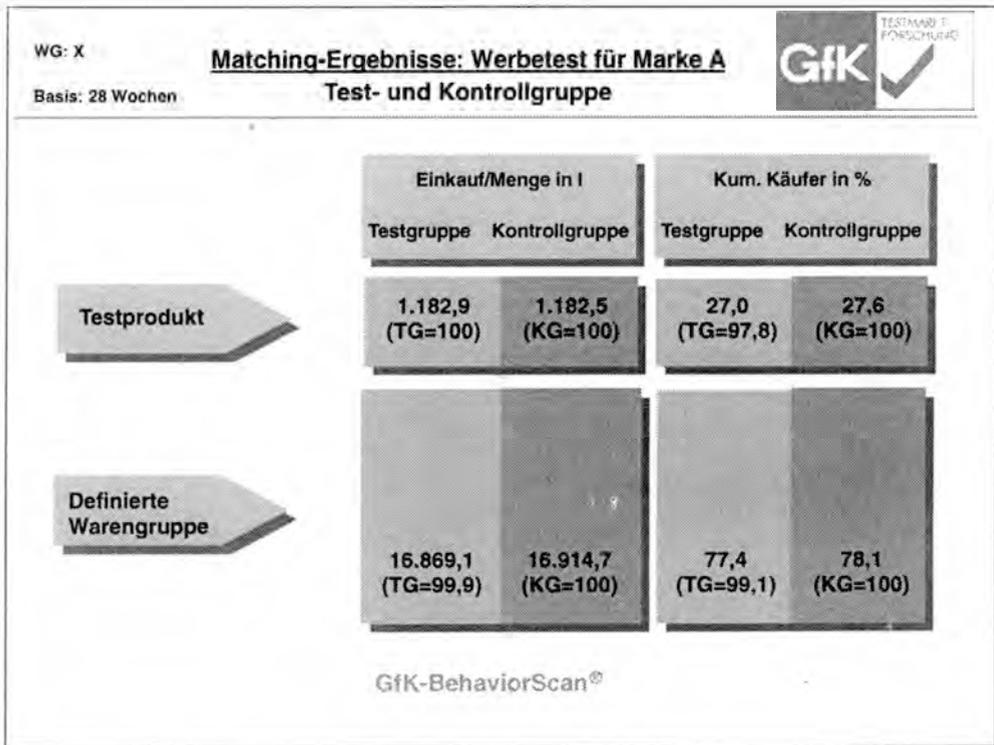


Maßgeschneiderte Splitgruppen

Insbesondere bei Werbewirkungsmessungen ist die Arbeit mit Test- und Kontrollgruppen erforderlich. Mit Hilfe eines mathematischen Optimierungsprogramms, des sogenannten Matchings, kann die Gesamtheit der beobachteten Haushalte individuell für jedes Projekt anhand zuvor definierter Kriterien in identische Test- und Kontrollgruppen aufgeteilt werden. Folglich weisen diese maßgeschneiderten Splitgruppen vor dem Einsatz der Testwerbung sowohl die gleiche Soziodemographie als auch das gleiche Kaufverhalten auf.

Beim Matching werden die Panelhaushalte zunächst im Zufallsverfahren in zwei oder mehr gleich große Gruppen aufgeteilt. Im zweiten Schritt werden die Gruppen durch den Austausch von Haushalten so lange optimiert, bis sie identisch sind.

Abbildung 23



Das vorliegende Beispiel zeigt anhand eines vereinfachten aber realistischen Beispiels, wie gut das Kaufverhalten der Test- mit dem der Kontrollgruppe nach der Matchingprozedur übereinstimmt.

Abbildung 24



Targetable TV: Gezielter individueller TV-Kontakt

Die Targetable TV-Technik erlaubt erstmals eine uneingeschränkte Kaufanalyse der Wirkung von Fernsehwerbung auf das reale Kaufverhalten der Konsumenten.

Am Kopfende des Kabelnetzes sorgt ein professionelles Fernsehstudio für die perfekte Ausstrahlung von Testwerbespots, der zentrale Studiorechner regelt die gezielte, individuelle Ansprache der gewünschten Splitgruppe.

Zum Empfang von Testwerbung ist an die Fernsehgeräte von 2 000 Testhaushalten ein Mikrocomputer (GfK-Box) angeschlossen. In diesen 2 000 Haushalten können regulär gesendete TV-Spots mit Hilfe der „Targetable TV“ genannten Technik durch Testspots überblendet werden. Der Einsatz regulärer Werbespots kann dabei in allen relevanten TV-Kanälen (ARD, Kabel 1, PRO 7, RTL, RTL II, SAT 1, Super RTL, VOX, ZDF) vorgenommen werden. Die Haushalte in der jeweiligen Kontrollgruppe sehen das allgemeine Werbeprogramm weiter.

Durch die Erfassung der Einkäufe und deren Zuordnung auf die Testhaushalte kann also genau festgestellt werden, ob und wieweit sich das Einkaufsverhalten der Testhaushalte gegenüber den Kontrollhaushalten durch die spezielle Werbung im Fernsehen verändert hat.

Abbildung 25



Die methodischen Anforderungen zur Werbeerfolgsmessung auf Basis ökonomischer Wirkungsmerkmale

- Eine experimentelle Wirkungsmessung von Werbung in Bezug auf das Kaufverhalten erfordert zuerst strukturgleiche Test- und Kontrollgruppen.
Diese Splitgruppen lassen sich entweder mit Hilfe des Matchingverfahrens (maßgeschneiderte Splitgruppen) definieren oder man kann auf die festinstallierten Gruppen zurückgreifen.
- Die definierten Gruppen müssen separat und gezielt mit Werbung ansprechbar sein. Dieser Forderung wird durch die Targetable TV-Technik bzw. der adressgenauen Verteilung der Testzeitschriften entsprochen.
- Zur Planung der differenzierten Werbeansprache von Test- und Kontrollgruppe müssen Informationen zum Seh- bzw. Leseverhalten verfügbar sein. Auf Basis einer Teilstichprobe von Haushalten werden Haushaltsreichweiten für Werbeblöcke analog zum GfK-Fernsehpanel elektronisch gemessen, außerdem werden die Lesegewohnheiten der Testhaushalte angelehnt an das Frageschema der Mediaanalyse (MA) erhoben.
- Zudem ist die exakte Erfassung des Einkaufsverhaltens der Splitgruppen erforderlich. Dies geschieht durch die Datenerfassung am Point-of-Sale mit Hilfe der Scannerkassen und der Identifikationskarte der Haushalte.

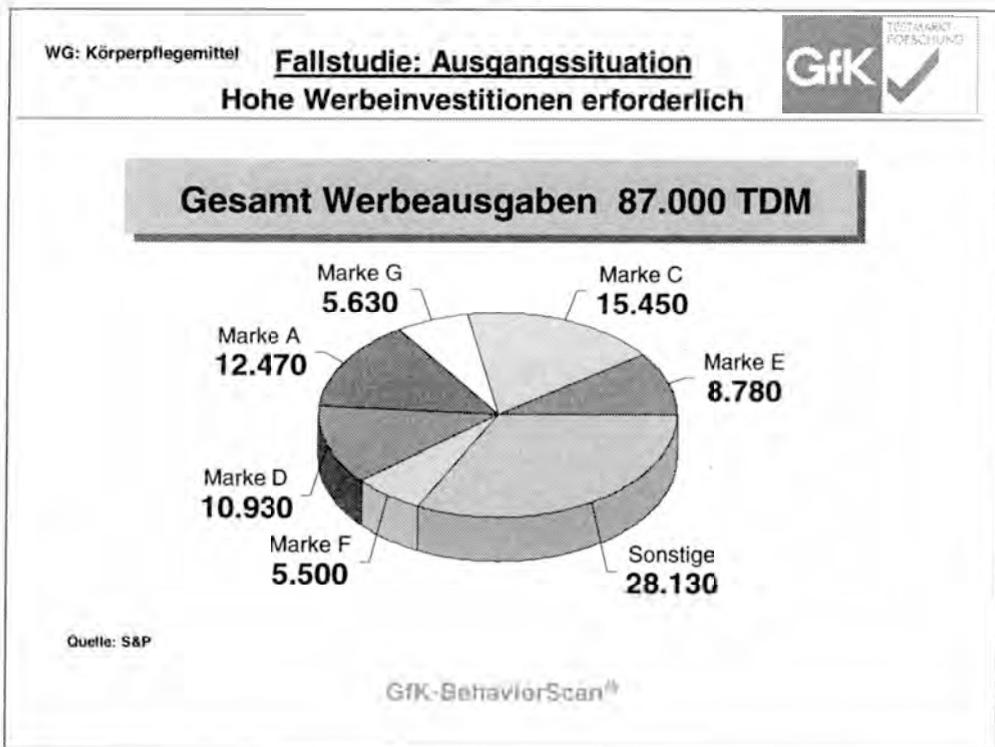
- Um über die am Point-of-Sale gemessenen Einkäufe eine exakte Bewertung der Effizienz von Werbemaßnahmen vornehmen zu können, müssen im Sinne einer validen, experimentellen Versuchsanordnung sämtliche bekannten und unbekannt potentiellen Einflussfaktoren (z. B. Einkaufsstättennutzung, Sortiments- und Preisgestaltung im Handel, Verkaufsförderungsmaßnahmen für die eigene Marke sowie für Konkurrenzprodukte etc.) neutralisiert werden. Die wichtigste Methode zur Neutralisierung von Störeinflüssen ist das sogenannte Splitting, d. h. die Bildung von strukturgleichen Test- und Kontrollgruppen innerhalb des Haushaltspanels.

Da sich die Haushalte der Test- und Kontrollgruppen im selben Ort befinden, nutzen sie dieselben Einkaufsstätten, verfügen somit über gleiche Wahlmöglichkeiten hinsichtlich Sortiment und Preis, erfahren eine gleichartige Verkaufsförderungsansprache etc.

Fallstudie: Alternativer Werbedruck für ein neues Produkt

Die folgende Fallstudie soll die Analysemöglichkeiten von Elektronischen Testmärkten verdeutlichen. Sie zeigt Ergebnisse eines Werbedrucktests für ein neues Produkt.

Abbildung 26



Hohe Werbeausgaben für eine Kategorie stellen eine besondere Eintrittsschwelle dar.

Abbildung 27

WG: Körperpflegemittel
Neuprodukttest
Testwochen 01-28

Fallstudie




Marktziel:

- Vergleichbare Marktstellung wie für die etablierten Marken B, F
- 5% Marktanteil

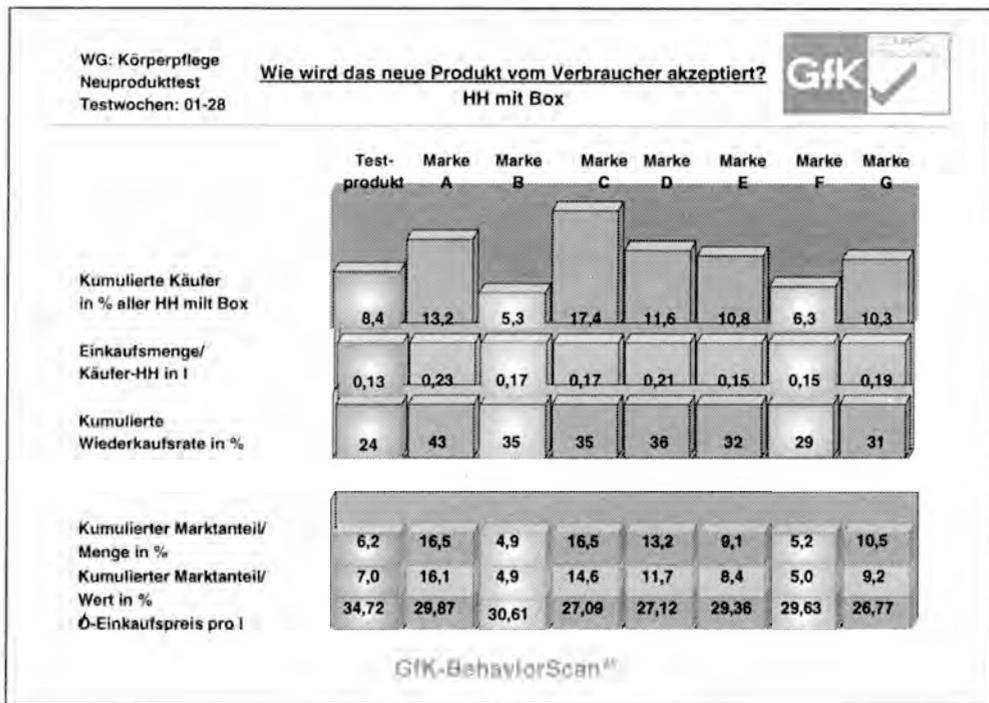
Testfragen:

- Welche Chancen und Risiken hat die neue Produktpersönlichkeit insgesamt?
- Welchen Erfolgsbeitrag leistet die TV-Kampagne überhaupt?
- Welcher von zwei alternativen Werbeetats unterstützt die Marke optimal?
- Was bewirken die eingesetzten Promotionmaßnahmen?

GfK-BehaviorScan®

Die Marktzielsetzung des Auftraggebers sah vor, einen Marktanteil von 5 % bzw. eine vergleichbare Marktstellung wie die Wettbewerber B und F auf dem nationalen Markt zu erreichen. Neben der Quantifizierung der Chancen und Risiken der neuen Produktpersönlichkeit auf dem deutschen Markt sollte der Test darüber Aufschluss geben, welcher von zwei alternativen Werbeetats die neue Marke optimal unterstützt und was die geplanten Promotionmaßnahmen zum Erfolg beitragen können.

Abbildung 29

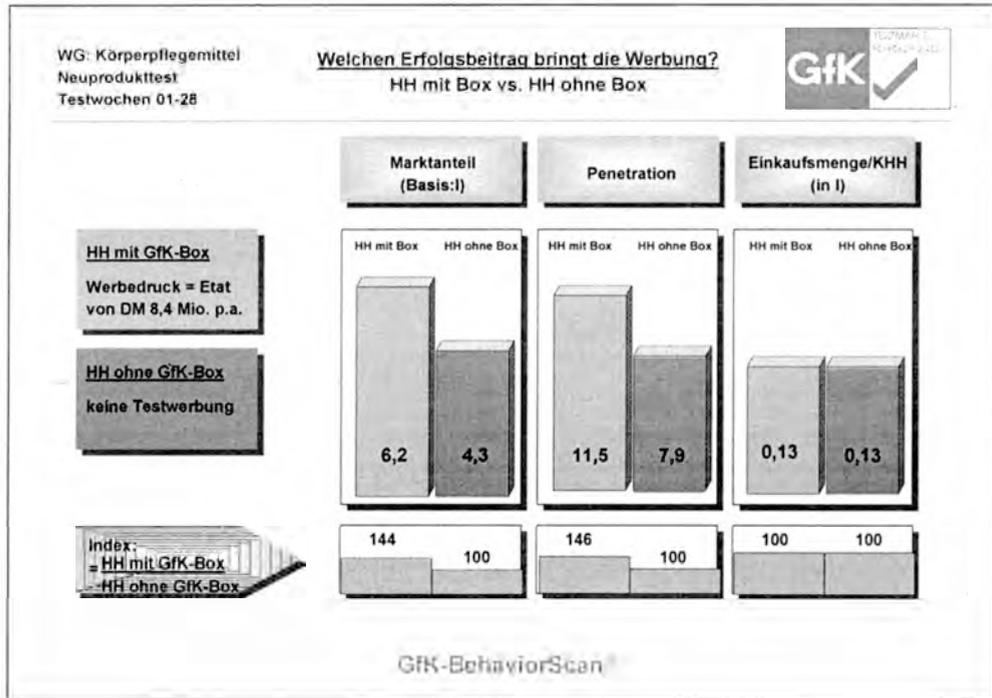


Obwohl für die neue Brand mindestens 13 %, z. T. aber auch über 20 % mehr als für vergleichbare Konkurrenzprodukte zu bezahlen ist, erreicht es unter den Marketingbedingungen im Test einen Marktanteil, der die Marken B und F übertrifft.

Ganz entscheidend dabei ist, dass es dem neuen Produkt gelingt, viele Käufer, nämlich mehr als 8 % der Haushalte, auf sich zu ziehen. Allerdings weist die neue Marke eine gewisse Schwäche im Bereich des Wiederkaufs auf: Nur 24 % Wiederkaufsrate bedeuten im Vergleich zu den etablierten Marken den schlechtesten Wert.

Dennoch: bei 100 % Distribution wird der anvisierte Marktanteil sogar deutlich übertroffen.

Abbildung 30



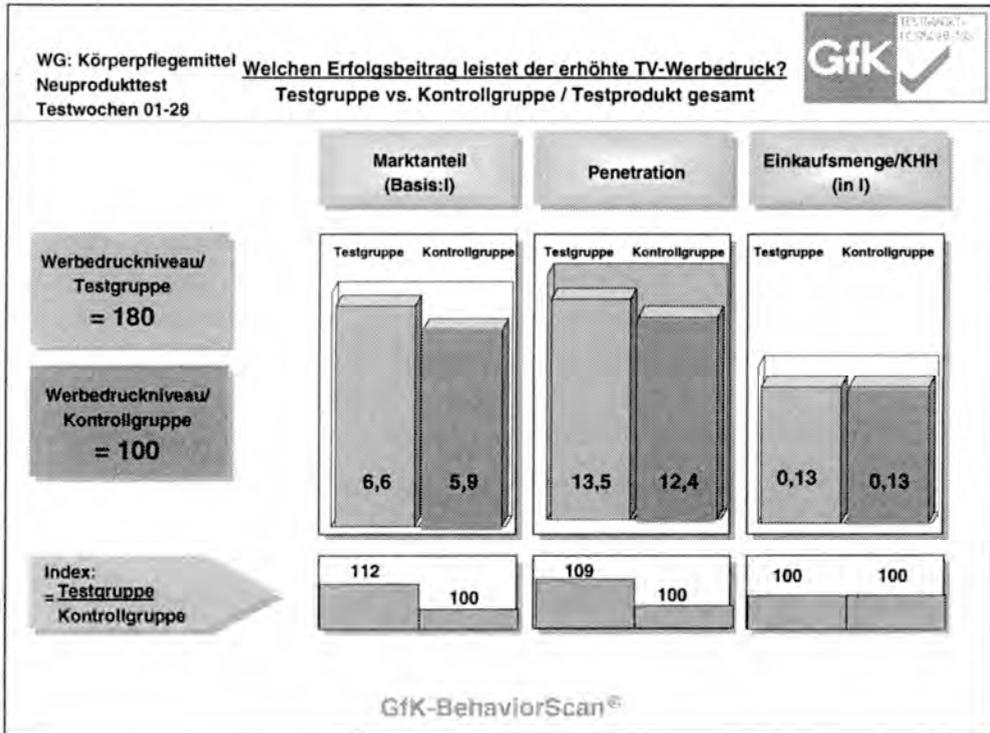
Wenden wir uns jetzt der Frage zu, welchen Erfolgsbeitrag die Werbung bei der Einführung dieser neuen Produktpersönlichkeit leisten kann.

Jene 2 000 Haushalte, die TV-Werbung für das neue Produkt erhalten haben, wurden im Durchschnitt einem Werbedruck von 8,4 Mill. DM pro anno ausgesetzt. Die Gegenüberstellung mit jenen Haushalten, die überhaupt keine Werbung erhalten haben, gestattet es, den Erfolgsbeitrag der TV-Werbung zu quantifizieren.

Die Grafik zeigt, dass der Erfolgsbeitrag der Werbung, bezogen auf das Kaufverhalten, mit gut 40 % (vgl. Marktanteil) zu Buche schlägt. Dabei ist entscheidend, dass es durch die Werbung gelingt, mehr Käufer an das neue Produkt heranzuführen (vgl. Penetration). Die durchschnittliche Einkaufsmenge je Haushalt bleibt hingegen gleich.

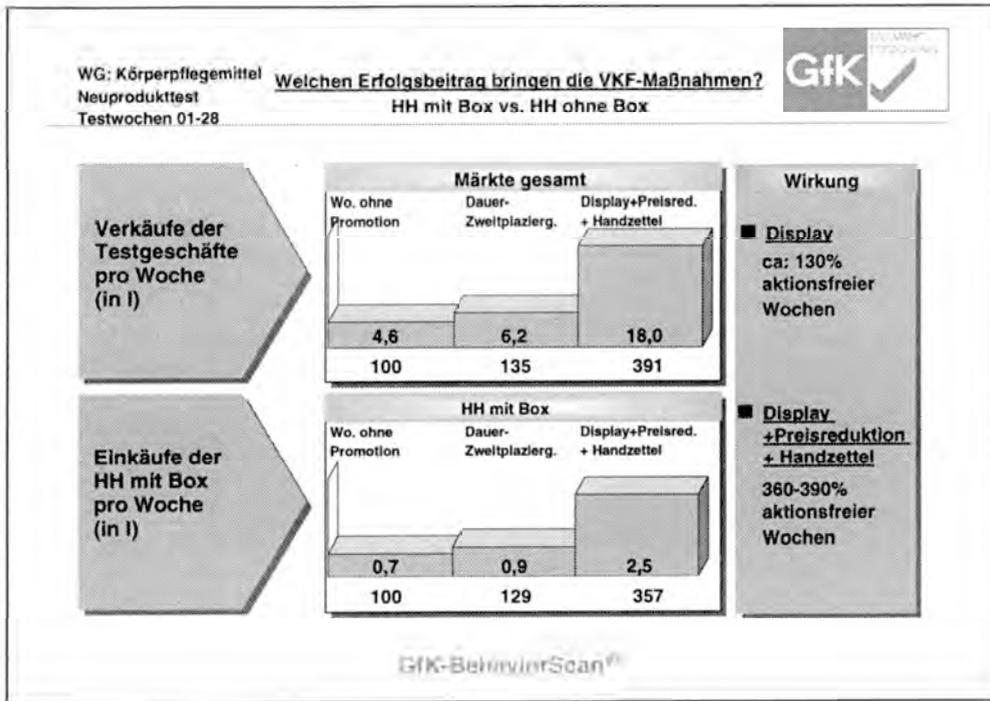
Diese Ergebnisse belegen, dass sich das Potenzial der eingesetzten Werbekampagnen unter realen Bedingungen und eingebettet in die anderen Marketingmaßnahmen voll entfalten kann. Die eingesetzte TV-Kampagne arbeitet also außerordentlich erfolgreich.

Abbildung 31



Diese Abbildung zeigt allerdings, dass ein um 80 % höherer TV-Werbedruck (10,8 Mill. DM pro anno gegen 6,0 Mill. DM pro anno) nach 28 Wochen nur zu einer Steigerung des Marktanteils um 12 % führt.

Abbildung 32



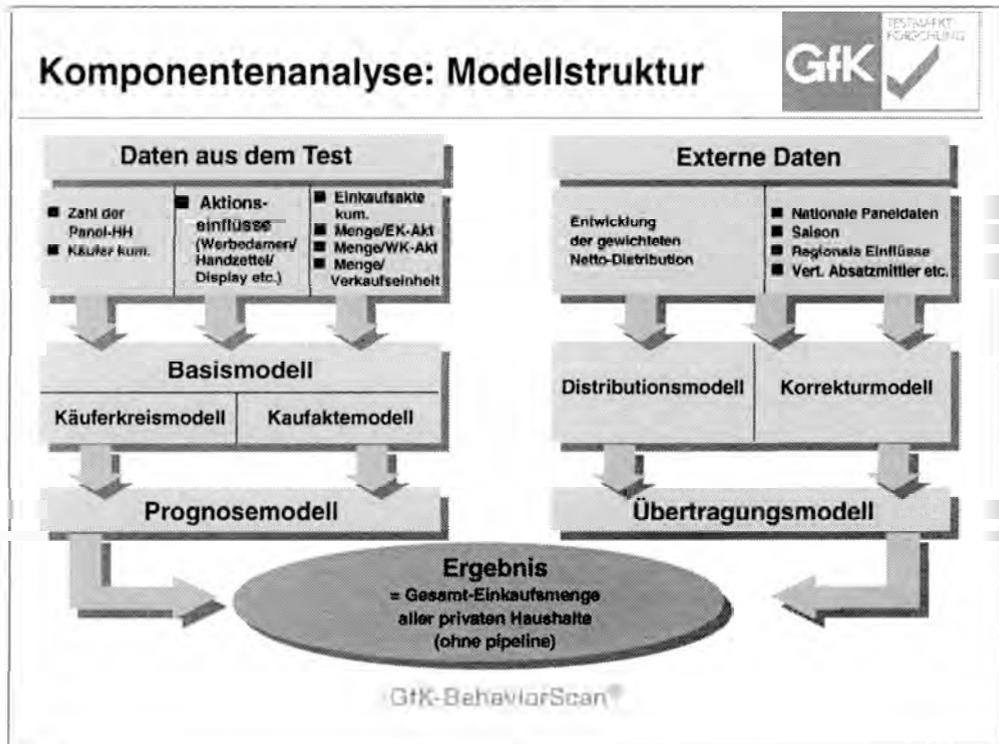
Die Analyse der Verkaufsförderungsmaßnahmen belegt, dass die Dauerzweitplatzierung für die Hauptgröße nur zu einer geringfügigen Absatzsteigerung für die gesamte Marke führt. Nach einer Preisreduktion um 12 % erhöhen sich die Abverkäufe auf das Vierfache aktionsfreier Wochen, wenn die Preissenkung durch einen Handzettel bekannt gemacht und der Konsument durch den Einsatz eines Displays im Geschäft auf das Angebot aufmerksam gemacht wird.

Aufgrund der Daten ergibt sich folgendes Fazit:

Die neue Marke besitzt klare Chancen auf dem deutschen Markt. Wenn national ein konzentrierter Marketingmitteleinsatz wie im Test durchgehalten werden kann, scheint das angestrebte Marktanteilsziel von rund 5 % erreichbar. Die Produktqualität sollte allerdings ebenso überprüft werden wie die Preisstellung im Markt (relativ geringe Wiederkaufsrate!).

Dabei empfiehlt sich jedoch, die Aufteilung des Kommunikationsbudgets auf die Bereiche klassische Werbung einerseits und Verkaufsförderung andererseits vor der nationalen Einführung zu optimieren. Die Ergebnisse deuten nämlich an, dass nicht der höhere Werbeetat von 10,8 Mill. DM optimal ist, sondern ein niedrigerer Etat, der Raum für weitere Promotionsaktivitäten lässt.

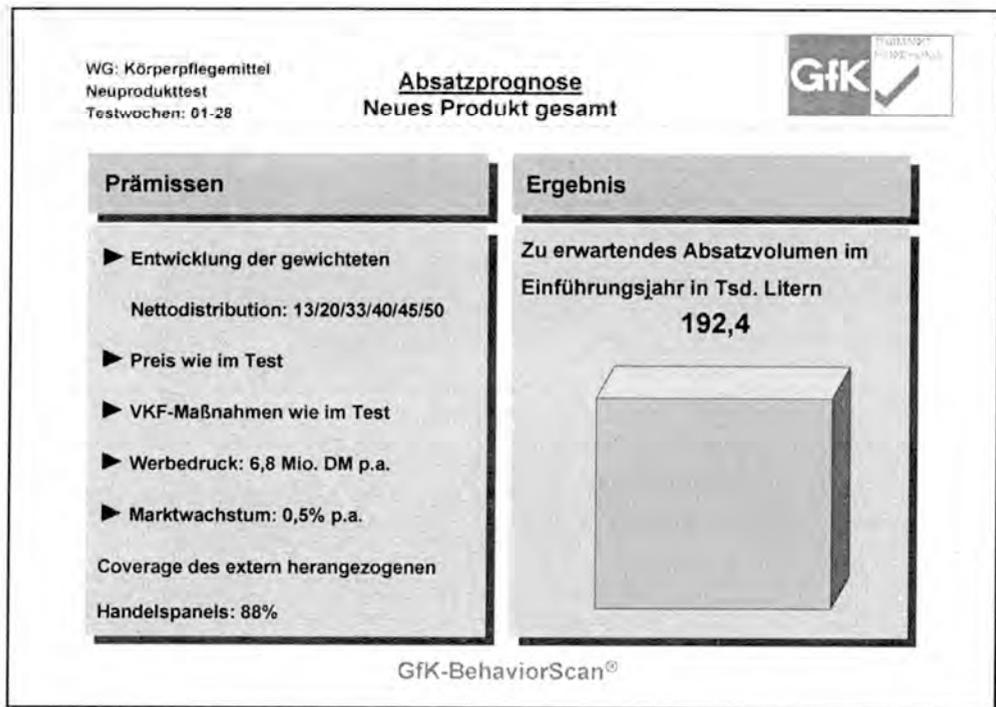
Abbildung 33



Zur Hochrechnung der Ergebnisse auf den Gesamtmarkt hat die GfK-Testmarktforschung ein Modell namens „Komponentenanalyse“ entwickelt. Dieses Modell basiert auf dem Four-Woodlock-Ansatz. Er wurde jedoch dahingehend ausgebaut, dass er eine Prognose auch dann erlaubt, wenn bei der nationalen Umsetzung von dem im Test realisierten Marketing-Mix abgewichen wird.

Die „Komponentenanalyse“ besteht im Wesentlichen aus zwei Teilen, dem Prognose- sowie dem Übertragungsmodell. Das Prognosemodell greift auf die im Test gewonnenen Daten zurück und extrapoliert das im Test erzielte Volumen auf ein volles Jahr. Das Übertragungsmodell berücksichtigt den zu erwartenden Distributionsaufbau, sonstige Elemente des nationalen Marketingplanes sowie regionale und saisonale Einflüsse.

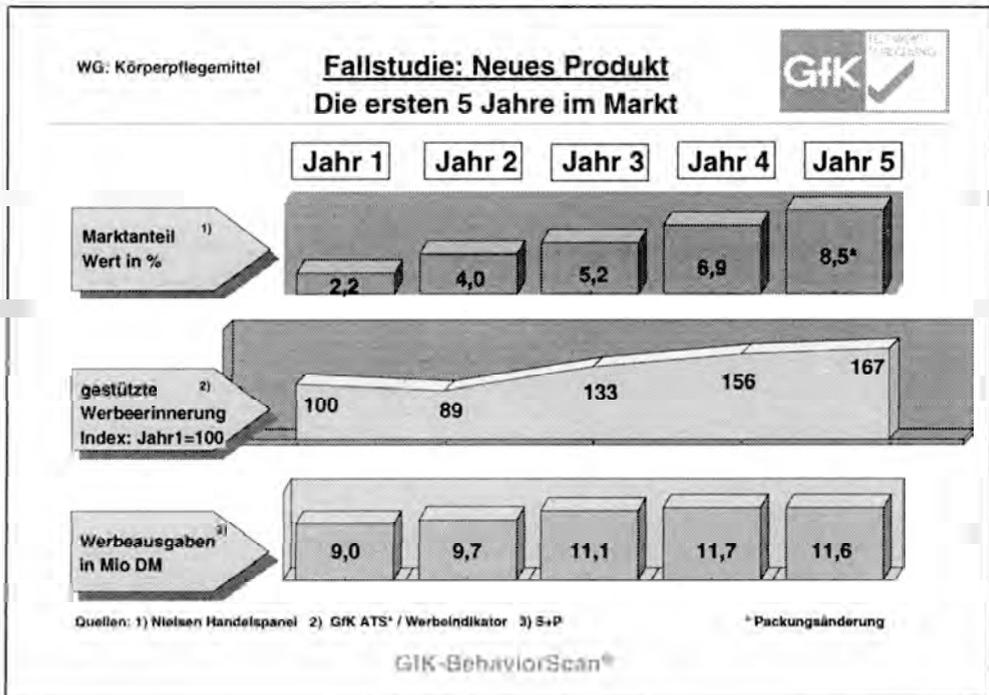
Abbildung 34



Für die Deckungsbeitragsplanung unseres Kunden wurde mit Hilfe der „Komponentenanalyse“ eine Volumensprognose erstellt.

Dabei wurde ein Werbebudget von 6,8 Mill. DM sowie ein zweimonatlicher Distributionsaufbau unterstellt, der am Ende des Einführungsjahres eine gewichtete Distribution von 50 % aufwies. Alle anderen Marketing-Mix-Variablen wurden national wie im Test geplant. Die Prognose ergab ein zu erwartendes Absatzvolumen im Einführungsjahr von 192 400 Litern.

Abbildung 35

*Die ersten 5 Jahre im Markt:*

Die neue Marke wurde mit einem nochmals optimierten Werbekonzept und einem Etat von 9,0 Mill. DM im ersten Jahr national eingeführt. Bereits im Jahr 3 konnte der angestrebte Marktanteil von 5 % übertroffen werden. Zwei Jahre später war man bereits bei 8,5 % Marktanteil angelangt.

4 Die verschiedenen Testverfahren im Vergleich

Lassen Sie uns im Folgenden den Versuch unternehmen, die verschiedenen Testmarktinstrumente direkt zu vergleichen.

Abbildung 36



Vergleichskriterien	Regionaler Testmarkt	Simulierter Testmarkt	Elektronischer Testmarkt
Testmöglichkeiten			
Produkt	++	++	++
Preis	(+)	++	+
Verpackung	+	++	+
Werbedruck	(+)	-	++
Kampagne	(+)	(+)	++
Promotion	+	-	++
Distribution	+	-	-
Messung	reales Kaufverhalten	simuliertes Kaufverhalten	reales Kaufverhalten
Validität	theoretisch hoch praktisch eingeschränkt	hoch, z.T. jedoch durch Laborsituation eingeschränkt	hoch
Realiabilität	sehr hoch	abhängig vom Stichprobenumfang	hoch
Diagnostisches Potential	gering	sehr hoch	gering
Internationale Vergleichbarkeit	gering	sehr hoch	hoch
Zeitbedarf	8 - 12 Monate	2 - 3 Monate	4 - 8 Monate
Kosten	sehr hoch	sehr gering	gering
Geheimhaltung	nein	ja	zum Teil

Pauschal lässt sich über die vorgestellten Testverfahren kein Urteil fällen. Die Auswahl des geeigneten Verfahrens ist immer abhängig von der konkreten Problemstellung.

Allerdings bleibt Folgendes festzuhalten:

- Der Regionale Testmarkt ist als einziges der diskutierten Testinstrumente in der Lage, die Handelsakzeptanz und damit die voraussichtliche Distribution für das Testprodukt zu ermitteln. Er ist aber auch mit Abstand das teuerste und zeitaufwendigste Verfahren, weist extreme Probleme beim Mediaeinsatz auf und bietet keinerlei Geheimhaltung.
- Die Geheimhaltung ist beim Simulierten Testmarkt fast zu 100 % gegeben, er weist das höchste diagnostische Potenzial auf, bietet ein/hohes Maß an internationaler Vergleichbarkeit und ist mit dem geringsten Kostenaufwand verbunden. Nicht erfassen kann er jedoch Variationen des Werbedrucks oder Verkaufsförderungsmaßnahmen am Point-of-Sale.
- Hinsichtlich der Wirkungsmessung von Werbung und Verkaufsförderung auf Basis realen Kaufverhaltens ist der Elektronische Testmarkt hingegen unübertroffen. Er benötigt dafür jedoch entsprechend mehr Zeit als der Simulierte Testmarkt und ist folglich mit etwas höheren Gesamtkosten verbunden. Bezüglich der Geheimhaltung bietet er zudem wesentlich mehr Diskretion als der Regionale Testmarkt.

Im Zweifelsfalle muss die individuelle Gewichtung von Gütekriterien durch den Auftraggeber den Ausschlag geben. Die vorliegenden Vergleichskriterien mögen dazu eine entsprechende Hilfestellung leisten.

5 Ausblick

Lassen Sie uns zum Abschluss einige Überlegungen zu der Frage entwickeln, welchen Weg die Testmarktforschung in Zukunft nehmen könnte.

Als Anbieter aller beschriebener Testmarktverfahren machen wir uns in der GfK natürlich Gedanken darüber, wie sich unser Dienstleistungsangebot dahingehend verbessern und weiterentwickeln lässt, dass wir den Anforderungen unserer Kunden noch besser gerecht werden können.

Aufgrund der kürzeren Produktlebenszyklen und der sich beschleunigenden Marktprozesse spielt das Thema „Zeit“ eine wichtige Rolle bei unseren Zukunftsüberlegungen. Vor diesem Hintergrund sind wir überzeugt, dass der klassische Regionale Testmarkt noch mehr an Bedeutung verlieren wird und die Relevanz von Elektronischen Testmärkten und Simulierten Testmärkten weiterhin zunehmen wird.

Zeitreserven liegen unseres Erachtens im Bereich der schnelleren Planung, Erhebung, Auswertung und Interpretation der Tests. In diesen Bereichen ist durch den Einsatz von adäquaten Informationstechnologien einiges zu erreichen. Allerdings dürfen diese Bemühungen nicht zu Lasten der Realitätsnähe unserer Instrumente gehen. Dort, wo reales Verbraucherverhalten in unsere Testansätze einfließt, d. h. immer wenn es um die Berücksichtigung von Verwendungsgewohnheiten, des realen Kaufverhaltens und um gewohnte Kaufzyklen in den jeweiligen Produktkategorien geht, werden den Beschleunigungsanstrengungen natürliche Grenzen gesetzt.

Wie die Vergangenheit belegt, kann die Entwicklung neuer Technologien entscheidende Impulse für die Verbesserung von Testmarktforschungsinstrumenten freisetzen. So ist beispielsweise die Einführung von Elektronischen Testmärkten erst durch die Verbindung der High-Tech-Bausteine, Scannerkassentechnologie, Kabelfernsehtechnologie und Mikrocomputertechnologie möglich geworden.

Die Grundidee für bestimmte Testverfahren existiert zum Teil schon sehr lange, doch die realistische Umsetzung dieser Ideen in praktikable Testinstrumente wird erst durch die Verfügbarkeit der entsprechenden Technologien möglich. Ein Beispiel, das diese Aussage belegt, ist die Renaissance der Conjoint-Analyse, deren Idee eigentlich schon 25 Jahre alt ist, die aber erst jetzt durch den Einsatz leistungsfähiger portabler PCs sowie durch die Möglichkeit der Verbindung mit CAPI-Interviews richtig zur Anwendung kommt.

Wir sind überzeugt, dass die rasanten technischen Entwicklungen, die sich mit Schlagworten „Multimedia“, „Communication Highway“ und „Interaktivität“ umschreiben lassen, mittelfristig Nutzen bringen und in entsprechende Testdesigns eingebaut werden können. Der erste Schritt zum Einsatz von Multimedia im Testmarktforschungsverfahren ist – zunächst im Offline-Bereich, spricht: im Studio – erfolgt. Multimediale Computerinterviews haben unseren Simulierten Testmarkt befruchtet. Dabei kommuniziert der Befragte direkt mit dem Computer, was den Einsatz von Interviewern ebenso wie den Interviewerbias reduziert und die Verwendungsmöglichkeiten realistisch gestalteter Befragungshilfen bis hin zu virtuellen Realitäten erweitern kann.

Bis aber der Online-Zugriff und die volle Interaktivität auf der Ebene des individuellen Haushalts in ausreichender Zahl möglich sein wird, wird wahrscheinlich das Jahr 2005 längst hinter uns liegen. Doch ist davon auszugehen, dass die starke Forcierung von Online-Diensten sowie die Erfahrungen der „Nintendo-Generation“ die Akzeptanz auf Konsumentenebene auch für innovative Testmarktforschungsdienste erhöhen wird. Dann dürfte jedoch die damit einhergehende Individualisierung der Medien zu einer neuen Herausforderung für die Testmarktforschung werden.

Die Internationalisierung unseres Testmarktangebotes ist ein weiteres Ziel, das wir mit Nachdruck verfolgen. Simulierte Testmärkte sind bereits heute ohne Probleme nahezu in allen Ländern der Welt in vergleichbarer Form einsetzbar. Bei Elektronischen Testmärkten gestaltet sich dieser Internationalisierungsprozess aufgrund der erforderlichen hohen Investitionen dagegen schwieriger. Derzeit sind entsprechende Angebote auf dem amerikanischen, dem deutschen und dem französischen Markt verfügbar. Weitere relativ große Länder wären bei entsprechendem Bedarf noch denkbar.

Auch wenn weitere Entwicklungstendenzen noch mit großer Unsicherheit behaftet sind, eines bleibt sicher: Die Reaktionen der Konsumenten werden auch in Zukunft die zentrale Rolle in der Testmarktforschung spielen.

Registergestützte Datengewinnung

Vorbemerkungen

In den Statistischen Ämtern des Bundes und der Länder wird gegenwärtig intensiv darüber diskutiert, wie Einzeldaten, die im Prozess des Verwaltungshandelns anfallen (bei Steuerbehörden, Arbeitsverwaltungen, Bundesaufsichtsämter usw.) und in umfassenden Dateien gespeichert werden, verstärkt als Datenquelle für die Wirtschafts- und Unternehmensstatistiken der Statistischen Ämter genutzt werden können.¹⁾

Diese Überlegungen sind auf den ersten Blick nicht neu: Schon jetzt werden in vielen Bereichen Sekundärstatistiken erstellt, die auf Daten aus administrativen Registern basieren (v. a. Steuerstatistiken, Beschäftigtenstatistik, Extrahandelsstatistik). Neu ist jedoch der Gedanke, dieses Nebeneinander an Primär- und Sekundärstatistiken über ein integratives Berichtssystem enger zu verzahnen. Für die Unternehmensstatistiken werden sich hier effiziente und weitreichende Möglichkeiten bieten, sobald das sich gegenwärtig noch im Aufbau befindliche wirtschaftsbereichsübergreifende Unternehmensregister voll einsatzfähig ist.

Die nachfolgenden Ausführungen sind in sechs Gliederungspunkte eingeteilt:

Im ersten Gliederungspunkt werden wichtige Rahmenbedingungen aufgezeigt, die die Arbeit der Statistischen Ämter beeinflussen. Anschließend werden spezifische Aufgaben des unternehmensstatistischen Berichtssystems kurz beschrieben.

Darauf aufbauend werden dann Überlegungen für einen Paradigmenwechsel hin zu einer verstärkten Nutzung von Verwaltungsdaten im Bereich der Unternehmensstatistiken skizziert. Hierbei ist es zunächst notwendig, auf Einflussgrößen einzugehen, die Argumente für einen entsprechenden Paradigmenwechsel liefern. Im Anschluss daran werden die Voraussetzungen und – zunächst allgemein – die Anwendungsmöglichkeiten vor dem Hintergrund des bestehenden komplexen Systems an Primärerhebungen diskutiert.

Im fünften Gliederungspunkt werden dann diese Überlegungen an einem Beispiel aus der Praxis – dem Projekt „Registergestützte Umsetzung der Konjunkturverordnung im Dienstleistungsbereich“ – konkretisiert. Den Abschluss bilden Ausführungen zu Chancen und Risiken eines integrativen, registergestützten Berichtssystems.

1 Rahmenbedingungen für die Arbeit der amtlichen Statistik

Der sich rasch und grundlegend vollziehende, alle Gesellschaftsbereiche umfassende Wandel hin zur Informationsgesellschaft, in der die „Information“ nicht nur ein immer stärker nachgefragtes Konsumgut, sondern für nahezu alle Bereiche des wirtschaftlichen Handelns ein an Bedeutung zunehmender strategischer Erfolgsfaktor darstellt, hat unmittelbare Auswirkungen auf das Datenangebot und die Arbeitsprozesse der Statistischen Ämter: Das Datenangebot ist qualitativ kontinuierlich zu verbessern (aktueller; bedarfsorientierter) und

*) Bernd Schmidt, Statistisches Bundesamt, Wiesbaden.

1) Der Bereich der Sozial-/Bevölkerungsstatistik und hier insbesondere die Frage der Durchführbarkeit einer registergestützten Volkszählung ist nicht Gegenstand der Ausführungen.

quantitativ auszudehnen (statistische Abbildung der internationalen Verflechtung der wirtschaftlichen Aktivitäten und der „Informatisierung“ der Gesellschaft). Darüber hinaus werden die Statistischen Ämter im Bereich der Wirtschaftsstatistiken mit komplexen Datenanforderungen der Europäischen Union (EU) konfrontiert, die über das bestehende statistische Berichtssystem nicht oder nur unzureichend befriedigt werden können.

Beiden Entwicklungen – sowohl der Wandel zur Informationsgesellschaft als auch die zunehmende EU-Ausrichtung der statistischen Berichterstattung – ist die Notwendigkeit zur Ausdehnung des Datenangebots inhärent.

Dem stehen jedoch Klagen der Befragten über zunehmende Belastung durch statistische Erhebungen gegenüber. Dies ist nachvollziehbar, wenn man bedenkt, dass zu den amtlichen Primärerhebungen eine Vielzahl an Erhebungen privater Institute hinzukommen, die in (un)regelmäßigem Turnus durchgeführt werden.

Ein anderer Aspekt, der jedoch tendenziell in die gleiche Richtung zielt, lässt sich subsumieren unter dem Begriff „Verwaltungsmodernisierung“. Gerade die amtliche Statistik ist seit Jahren dazu angehalten, systematisch und nachhaltig Kosten zu reduzieren.

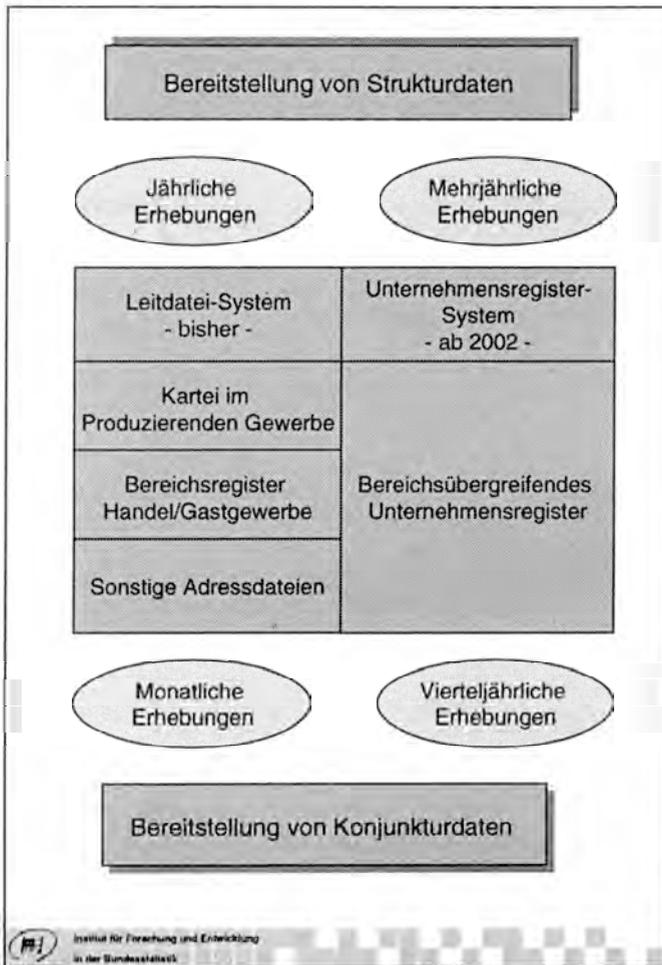
Die Bewältigung dieses „Spagats“ aus quantitativer Ausweitung/qualitativer Verbesserung des Datenangebots auf der einen Seite und dem Zwang zur Belastungs- und Kostenreduktion auf der anderen Seite ist im Bereich der Wirtschafts- und Unternehmensstatistiken nicht zuletzt aufgrund des hohen Stellenwerts dieser Statistiken in der Öffentlichkeit für die Statistischen Ämter eine essentielle Aufgabe.

2 Aufgaben des unternehmensstatistischen Berichtssystems

Was das Informationsangebot betrifft, muss das unternehmensstatistische Berichtssystem zwei grundlegende Aufgaben erfüllen:

Eine wichtige Aufgabe besteht darin, *Strukturinformationen* für und über die Aktivitäten der Betriebe und Unternehmen der einzelnen Wirtschaftsbereiche bereitzustellen. Hierbei handelt es sich um die Erfassung von tiefgegliederten Unternehmensdaten, die zuverlässige und vielschichtige Informationen insbesondere über die Kostenstruktur, (Investitions-)Tätigkeit, Wettbewerbsfähigkeit und Leistung der Unternehmen der einzelnen Wirtschaftsbereiche geben. Um dieses Informationsangebot bereitzustellen zu können, werden gegenwärtig eine große Zahl an Primärerhebungen in den Bereichen Bergbau, Verarbeitendes Gewerbe, Energie- und Wasserversorgung, Baugewerbe, Handel, Gastgewerbe und im übrigen Dienstleistungsbereich in periodischen Abständen (jährlich, mehrjährlich) teils als Totalerfassung, teils als repräsentative Stichproben und teils mit Abschneidegrenzen durchgeführt. In der Regel sind die amtlichen Primärerhebungen – seit dem Jahr 1998 auch alle Kostenstrukturerhebungen – mit Auskunftspflicht versehen.

Grundzüge des unternehmensstatistischen Berichtssystems



Auf der anderen Seite muss das statistische Berichtssystem *Konjunkturinformationen* bereitstellen. Hierbei handelt es sich um die Darstellung von kurzfristigen – in der Regel monatlichen oder vierteljährlichen – Verlaufsdaten. Die Daten sind Grundlage für die Analyse der konjunkturellen Entwicklung von Angebot und Nachfrage, Produktionsfaktoren und Marktpreisen. Auch diese statistischen Schlüsselgrößen werden gegenwärtig nahezu ausschließlich über monatliche und vierteljährliche Primärerhebungen bei Unternehmen und Betrieben ermittelt.

Steuerungsinstrument für die Planung und Durchführung dieser Erhebungen ist bisher ein bereichsspezifisches System von Leitdateien. Das im Aufbau befindliche bereichsübergreifende Unternehmensregister-System wird diese Aufgabe in Zukunft effizienter ausfüllen können.²⁾

3 Paradigmenwechsel zur registergestützten Datengewinnung

Das Gesamtsystem der Unternehmenserhebung ist durch hohe Komplexität gekennzeichnet. Es gibt nicht nur eine große Zahl an Primärerhebungen, diese Erhebungen sind darüber hinaus – sowohl was Methodik als auch was die Inhalte, Definitionen, Erhebungsperiodizitäten betrifft – nur teilweise aufeinander abgestimmt. Dies gilt insbesondere für die intersektorale Sichtweise. Eine einheitliche, *alle* Bereiche umspannende Strukturhebung im jährlichen Bereich bzw. Konjunkturerhebung im unterjährlichen Bereich gibt es gegenwärtig in der amtlichen Statistik nicht.

Der Grund für dieses komplexe Nebeneinander an Erhebungen liegt nicht zuletzt darin, dass das System je nach Entwicklung des (politischen) Datenbedarfs um neue Primärerhebungen ergänzt wurde. Die fast zwangsläufige Folge aus dieser sukzessiven, gewöhnlich nicht von den Statistischen Ämtern zu verantwortenden Angebotserweiterung, ist eingeschränkte Stringenz des Systems, die im ungünstigsten Fall dazu führen kann, dass identische Tatbestände mehrfach im Jahr bei denselben Unternehmen erhoben werden.

Die daraus resultierende starke Belastung der Auskunftspflichtigen ist primär im Produzierenden Gewerbe und im Handel und Gastgewerbe zu verzeichnen, und dort insbesondere bei den großen Unternehmen. Lücken im System gibt es dagegen bei den kleinen Unternehmen und generell im Bereich der unternehmensnahen, der öffentlichen und der personenbezogenen Dienstleister (NACE³⁾-Abschnitte K, M bis O), aber auch im Bereich der institutionellen Verkehrsstatistik. Hier gibt es nur für ausgewählte Teilbereiche Kostenstrukturhebungen, die jedoch nur alle 4 Jahre durchgeführt werden. Um in Zukunft diese zunehmend an Bedeutung gewinnenden Wirtschaftsbereiche statistisch angemessen abbilden zu können, ist die Einführung einer flächendeckenden Dienstleistungsstatistik, die zumindest jährlich die dringend benötigten Strukturdaten zur Verfügung stellt, unerlässlich. Diese Dienstleistungsstatistik ist bisher jedoch auch am Widerstand der Unternehmen gescheitert.

Vor diesem Hintergrund und bedingt durch die sich zukünftig über das Unternehmensregister bietenden Möglichkeiten der Verknüpfung verschiedener Datenquellen werden im Statistischen Bundesamt Überlegungen angestellt, aus dem bisher quasi eindimensionalen primärstatistischen System ein Berichtssystem zu entwickeln, das durch die möglichst optimale Verzahnung unterschiedlicher Datenquellen gekennzeichnet ist. Diese Überlegungen, die gegenwärtig im Bereich der Konjunkturstatistiken schon in die Praxis übertragen werden,⁴⁾ kommen einem Paradigmenwechsel der Datengewinnung gleich. Dieser ist jedoch notwendig, um die eingangs geschilderten, sich zum Teil diametral gegenüberstehenden Aufgaben bewältigen zu können.

2) Vgl. auch die Ausführungen auf S. 135 ff.

3) Allgemeine Klassifikation der Wirtschaftszweige in den Europäischen Gemeinschaften, revidiert (NACE rev. 1).

4) Vgl. die Ausführungen in Gliederungspunkt 5.

4 Voraussetzungen für eine registergestützte Datengewinnung

Leitidee ist die Entwicklung eines *integrativen*, eng verzahnten Berichtssystems bestehend aus

- stringenten, gut aufeinander abgestimmten Primärerhebungen und
- administrativen Dateien, die systematisch und einheitlich gepflegt werden.

Die Verzahnung der aus diesen Quellen kommenden Einzeldaten soll ein umfassendes, aktuelles Unternehmensregister ermöglichen, das in den Statistischen Ämtern geführt wird.

Wichtige Realisierungsvoraussetzungen sind

- (1) ein lückenloses, alle Wirtschaftsbereiche umfassendes und funktionsfähiges Unternehmensregister,
- (2) die Existenz geeigneter administrativer Dateien,
- (3) gesetzliche Grundlagen, die den Statistischen Ämtern den Zugriff auf Einzeldaten aus den administrativen Dateien erlauben und
- (4) Einsatzmöglichkeiten für administrative Daten im unternehmensstatistischen Berichtssystem.

Zu (1): Umfassendes und funktionsfähiges Unternehmensregister

Dem Unternehmensregister kommt im Rahmen des integrativen Berichtssystems eine Schlüsselrolle zu:

Zunächst ist das Unternehmensregister ein Instrument zur *Erhebungsunterstützung*. Es wird benötigt bei der Erhebungskonzeption (Auswahlplanung, Hochrechnung), bei der Erhebungsdurchführung (Versandunterstützung, Eingangskontrolle), bei der Datenaufbereitung (Plausibilitätskontrolle) und bei der Datenvalidierung (Fehlerrechnung). Diese eher konventionellen Aufgaben wurden bisher – soweit vorhanden – von bereichsspezifischen Dateien/Karteien⁵⁾ übernommen. Ein einheitliches Unternehmensregister wird diese Funktion jedoch in Zukunft effizienter ausfüllen können. Darüber hinaus bietet ein umfassendes, einheitlich geführtes Unternehmensregister Chancen für eine konzeptionelle Vereinheitlichung der Primärerhebungen und daher für eine Erhöhung der Stringenz des gegenwärtigen Berichtssystems.

Darüber hinaus soll das Unternehmensregister als *Bindeglied* zwischen primärerhobenen Daten und Daten aus administrativen Dateien fungieren. Ohne ein funktionierendes Unternehmensregister wird es nicht möglich sein, betriebliche Einzeldaten, die aus unterschiedlichen administrativen Dateien abgegriffen werden, sowohl untereinander als auch mit Betriebs- bzw. Unternehmensdaten aus amtlichen Erhebungen zu verknüpfen. Diese Schnittstellenfunktion kann im Prinzip nur von einem in den Statistischen Ämtern geführten Unternehmensregister ausgeübt werden.

Darüber hinaus kommt dem Unternehmensregister auch eine unmittelbare *Auswertungsfunktion* zu. Da das Unternehmensregister später in regelmäßigen Intervallen aktualisiert wird (einmal jährlich), kann es zur Abbildung von Zeitverläufen und zur Längsschnittanalyse (Unternehmensdemographie) genutzt werden.

5) Genannt werden muss insbesondere die Kartei im Produzierenden Gewerbe (KiPG).

Dieses für die registergestützte Datengewinnung essentielle Unternehmensregister wird in den Statistischen Ämtern aufgebaut. Gegenwärtig gibt es noch große Lücken insbesondere im sehr heterogenen Dienstleistungssektor. Um diese Lücken zu beseitigen, werden die in den Statistischen Ämtern vorhandenen Bereichsregister des Produzierenden Gewerbes, Handels und Gastgewerbes mit umfassenden administrativen Dateien verschmolzen. Dieses „Matchen“ der Dateien ist nicht zuletzt aufgrund der riesigen Datenmengen (je nach Datei mehrere Millionen Datensätze) und der nicht vorhandenen registerübergreifend einheitlichen Unternehmensnummer ein enormer Arbeitsaufwand und wird noch längere Zeit in Anspruch nehmen. Nach dem gegenwärtigen Stand der Planungen ist damit zu rechnen, dass ab dem Jahr 2002 auf ein funktionsfähiges, umfassendes Unternehmensregister zurückgegriffen werden kann, das dann jährlich anhand von Daten aus dem amtlichen Erhebungsprozess und aus geeigneten Verwaltungsdateien aktualisiert werden soll.

Rechtliche Grundlage sowohl für den Aufbau als auch für die spätere jährliche Registeraktualisierung ist die EU-Registerverordnung ⁶⁾ aus dem Jahr 1993, in der die Mitgliedstaaten verpflichtet werden, ein funktionsfähiges, wirtschaftsbereichsübergreifendes Unternehmensregister aufzubauen. Da für Aufbau und Pflege des Registers Informationen aus administrativen Dateien genutzt werden müssen, trat Mitte 1998 das Statistikregistergesetz (StatRegG)⁷⁾ in Kraft, das den Zugriff der Statistischen Ämter auf die benötigten Verwaltungsdaten und die Registerführung regelt.

Zu (2): Geeignete administrative Dateien

Zweite Voraussetzung für die Realisierung eines integrativen, auf primärerhobenen Daten und Registerdaten basierenden Berichtssystems ist die Existenz von administrativen Dateien, die sowohl aus inhaltlicher (erfasste Merkmale), berichtskreisbezogener (Deckungsgrad der erfassten Einheiten), erfassungstechnischer (Struktur und IT-Konzeption des Datenbankmodells) und zeitlicher (periodische Aktualisierung der Berichtskreise und -merkmale) Sicht für eine breit angelegte Nutzung im Rahmen der Statistikproduktion in Frage kommen.

Der Beantwortung dieser Frage ist das Institut für Angewandte Wirtschaftsforschung Tübingen (IAW) im Rahmen seines im Auftrag des Bundesministeriums der Finanzen (BMF) durchgeführten Gutachtens zu den „Möglichkeiten verstärkter Nutzung von Verwaltungsdaten für Zwecke der Wirtschaftsstatistik und zur Entlastung der Wirtschaft von statistischen Berichtspflichten“ nachgegangen.⁸⁾ In seiner umfangreichen Registerrecherche kommt das

6) Verordnung (EWG) Nr. 2186/93 des Rates vom 22. Juli 1993 über die innergemeinschaftliche Koordinierung des Aufbaus von Unternehmensregistern für statistische Verwendungszwecke; in: Amtsblatt der Europäischen Gemeinschaften Nr. L 196/1 vom 5. August 1993.

7) Artikel 1 des Gesetzes zur Durchführung der Verordnung (EWG) Nr. 2186/93 des Rates vom 22. Juli 1993 über die innergemeinschaftliche Koordinierung des Aufbaus von Unternehmensregistern für statistische Verwendungszwecke vom 16. Juni 1998; in: Bundesgesetzblatt Jahrgang 1998 Teil I Nr. 36, ausgegeben zu Bonn am 23. Juni 1998.

8) Vgl. Helbig, J., Engelage, Ch., Wiegert, R.: Möglichkeiten verstärkter Nutzung von Verwaltungsdaten für Zwecke der Wirtschaftsstatistik und zur Entlastung der Wirtschaft von statistischen Berichtspflichten, Gutachten für das Bundesministerium für Wirtschaft und Technologie, Tübingen, 1999.

Die nachfolgenden Ausführungen zu den Inhalten und Eigenschaften geeigneter administrativer Dateien basieren im wesentlichen auf dem IAW-Gutachten, S. 9 ff.

IAW zu dem Ergebnis, dass es administrative Dateien gibt, die grundsätzlich für eine Nutzung im Rahmen der amtlichen Wirtschaftsstatistik in Frage kommen. Die wichtigsten sind

– *die Steuerregister der Länderfinanzverwaltungen*

Gegenwärtig verwalten 21 Oberfinanzdirektionen Steuerdaten der steuerpflichtigen Unternehmen, Organschaften und Personen. In den sogenannten Speicherkonten werden Grundinformationsdaten verwaltet, wie zum Beispiel Steuernummer, Anschrift, Bankverbindung, Umfang der steuerlichen Inanspruchnahme. In den Festsetzungsspeichern werden alle Angaben aus der jährlichen Steuererklärung von Unternehmen im steuerlichen Sinne und steuerpflichtigen Personen gespeichert. Aus Speicherkonto und Festsetzungsspeicher können folgende wirtschaftsstatistisch relevanten Informationen, die mit Daten aus Primärerhebungen korrespondieren, entnommen werden:

(1) Angaben über steuerpflichtige Unternehmen

Strukturinformationen (Anschrift, Rechtsform, Steuernummer, Gewerbekeznzahl, Gemeindegkeznzahl u. a.), Umsätze/Umsatzsteuer aus dem Umsatzsteuer-Voranmeldungsverfahren (monatlich/vierteljährlich), jährlicher Umsatz, jährliche Umsatzsteuer, Gewinn.

(2) Angaben über steuerpflichtige Personen

Strukturinformationen (Anschrift, Steuernummer, Angaben zur Person), zu versteuerndes Einkommen, Sozialversicherungsbeiträge (Altersvorsorge, Krankenkasse).

Von Bedeutung für die Unternehmensstatistik sind vor allem die aus dem sogenannten Umsatzsteuer-Voranmeldungsverfahren (UVV) kommenden Angaben zu den steuerbaren Umsätzen der Unternehmen. Gemäß § 18 Abs. 2 UStG sind alle steuerpflichtigen Unternehmen verpflichtet, monatlich⁹⁾ bzw. vierteljährlich¹⁰⁾ ihren steuerbaren Umsatz an die Oberfinanzdirektionen zu melden. Meldepflichtig sind gegenwärtig ca. 2,8 Millionen Unternehmer, deren Jahresumsatz im Vorjahr jeweils mehr als 32 500 DM betrug.¹¹⁾ Die Angaben sind sehr aktuell, da die Unternehmer verpflichtet sind, ihre Voranmeldung innerhalb von 10 Tagen nach Ablauf des Voranmeldungszeitraums abzugeben.¹²⁾ Die monatlichen/vierteljährlichen Umsatzdaten lassen sich zu Jahresdaten aggregieren.

– *die Beschäftigten-/Betriebsdatei der Bundesanstalt für Arbeit (BA)*

In der Versicherten- und Betriebsdatei der Bundesanstalt für Arbeit sind umfassende Einzelangaben über sozialversicherungspflichtig Beschäftigte und den ihnen zugeordneten Betrieben gespeichert. Die Angaben in den Dateien werden fortlaufend um An- und Abmeldungen aktualisiert.

Die Daten kommen aus dem sogenannten integrierten Meldeverfahren zur Sozialversicherung, d. h. ein Betrieb erledigt mit einer einzigen Meldung seine Meldepflichten (An- und Abmeldung, versicherungsrechtliche Änderungen) an die Krankenkassen, die

9) Falls die Umsatzsteuer des Vorjahres mehr als 12 000 DM betrug.

10) Falls die Umsatzsteuer des Vorjahres höchstens 12 000 DM betrug.

11) Neben den kleinen Unternehmen werden in den Umsatzsteuerdateien auch *Organgesellschaften*, die gem. § 2 Abs. 2 Nr. 2 UStG zusammen mit der *Organmutter* einen Steuerpflichtigen (*Organkreis*) bilden, nicht separat erfasst.

12) Die Fristen für die Abgabe der Voranmeldung und für die Entrichtung der Vorauszahlung können auf Antrag um einen Monat verlängert werden (§ 46 UStDV).

Rentenkassen und die Arbeitsverwaltung. Die ca. 27 Mill. Personendatensätze der Versichertendatei lassen sich über eine Betriebsnummer zu Betriebsdaten verdichten, die dann umfangreiche Einsatzmöglichkeiten in der Unternehmensstatistik bieten.

(1) Merkmale in der Versichertendatei (Auswahl)

Versicherungsnummer, Altersjahr, Geschlecht, Staatsangehörigkeit, Beginn der Beschäftigung (bei Anmeldung), Beschäftigungszeitraum (bei Abmeldung), Angaben zur Tätigkeit (ausgeübte Tätigkeit, Stellung im Beruf, Ausbildung), Vollzeit-/Teilzeitbeschäftigung, Bruttoarbeitsentgelt, Wohnort, Betriebsnummer.

(2) Merkmale in der Betriebsdatei und der Ortsdatei (Auswahl)

Betriebsnummer, Aufnahmedatum, Name und Anschrift des Betriebs, Ort des Betriebssitzes, Wirtschaftsklasse, Gemeindekennziffer.

Das Statistische Bundesamt erhält gegenwärtig vierteljährlich den aktuellen Bestand an sozialversicherungspflichtig Beschäftigten als anonymisierte Einzeldaten mit allen statistisch relevanten Merkmalen. Geregelt ist dies im Arbeitsförderungsreformgesetz, SGB III § 281, 282, 282a. Darüber hinaus wird jährlich das sogenannte Jahreszeitraummaterial mit sämtlichen Meldungen der Beschäftigung „von/bis“ und den zugehörigen Entgeltsdaten (Bruttoentgelte) im Rahmen der Beitragsbemessungsgrenze an das Statistische Bundesamt geliefert. Hinzu kommt die Lieferung von Betriebsdaten für die Zwecke des Unternehmensregisters gemäß Statistikregistergesetz (StatRegG).

– *die Dateien der Industrie- und Handelskammern, Handwerkskammern und der Berufsgenossenschaften*

Im Vergleich zu den beiden erstgenannten bereichsübergreifenden Dateien sind die Berichtskreise dieser Dateien weniger umfassend. In den Registern der Handwerkskammern zum Beispiel sind alle Unternehmen/Betriebe erfasst, die in die Handwerksrolle eingetragen sind. Von Interesse für die Unternehmensstatistik sind diese Register insbesondere für die effiziente Abgrenzung von Berichtskreisen. In Kombination mit den erstgenannten Dateien sind sie für eine registergestützte Datengewinnung von großer Bedeutung (komplementäre Aufgaben).

Neben diesen Dateien werden im IAW-Gutachten noch eine ganze Reihe anderer Register genannt, die für wirtschaftsstatistische Zwecke jedoch nur mit Einschränkungen nutzbar sind,¹³⁾ sei es aufgrund der Konzentration auf ganz spezifische Berichtskreise oder aufgrund der eingeschränkten regionalen Reichweite. Genannt werden

- Datenbanken der kommunalen Ämter öffentlicher Ordnung, Gewerbeämter,
- Kammerregister der freien Berufe,
- Daten des Verbandes Deutscher Rentenversicherungsträger,
- Datenbanken der städtestatistischen Ämter,
- Datenbestände der Deutschen Bundesbank bzw. der Bundesaufsichtsämter für das Kredit- bzw. Versicherungswesen.

13) Vgl. Helbig, J., Engelage, Ch., Wiegert, R.: Möglichkeiten verstärkter Nutzung von Verwaltungsdaten für Zwecke der Wirtschaftsstatistik und zur Entlastung der Wirtschaft von statistischen Berichtspflichten, Gutachten für das Bundesministerium für Wirtschaft und Technologie, Tübingen, 1999, S. 23 ff.

Die Ergebnisse der IAW-Registerrecherche werden auch bestätigt durch die umfangreichen Untersuchungen der Statistischen Ämter, die für den Aufbau und die spätere Pflege des Unternehmensregisters den Zugang zu den erstgenannten Dateien für notwendig erachtet haben.

Zusammenfassend kann festgehalten werden, dass es in Deutschland umfassende administrative Dateien gibt, die sich für eine Nutzung im Rahmen der amtlichen Unternehmensstatistik eignen. Einschränkend muss jedoch hervorgehoben werden, dass diese Dateien natürlich primär auf die Zwecke/Aufgaben der sie führenden Verwaltungen und nicht auf den Bedarf der amtlichen Statistik ausgerichtet sind. Um sie möglichst optimal auch für die Statistik nutzen zu können, wären umfangreiche Maßnahmen der Registertrimmung notwendig. Das heißt, die Register müssten sowohl um Merkmale und Auswertungstatbestände erweitert (z. B. Erweiterung der Umsatzsteuerdateien um das Merkmal „Betriebszusammenhang“, um auch Betriebsergebnisse abbilden zu können) als auch DV-technisch an den Bedarf der amtlichen Statistik angepasst werden.

Um ein modernes, in sich konsistentes, integratives Gesamtsystem unternehmensstatistischer Erhebungen erreichen zu können, müssten darüber hinaus umfangreiche Abstimmungsarbeiten und Umorganisationen im Datenbankmanagement und der angewandten Software im Bereich der Verwaltungen wie auch im Statistikbereich erfolgen. Von großer Bedeutung für eine effiziente Zusammenführung von administrativen Datenbankinformationen wäre ferner die Einführung einer registerübergreifend einheitlichen Unternehmenskennziffer. Gegenwärtig existiert ein einheitlicher Unternehmensidentifikator nicht.

Zu (3): Zugangsberechtigung zu administrativen Dateien

Dritte Voraussetzung für ein integratives, registergestütztes Berichtssystem ist eine gesetzliche Grundlage, die es den Statistischen Ämtern erlaubt, flexibel auf Einzeldaten aus geeigneten administrativen Dateien zuzugreifen. Aus Gründen des Datenschutzes ist die Übermittlung von Personen-, Betriebs-/Unternehmensdaten strikt von der Existenz einer entsprechenden gesetzlichen Regelung abhängig. Für deren Ausgestaltung gibt es letztendlich zwei grundsätzliche Strategien:

– Schaffung einer einzelgesetzlichen Grundlage

In diesem – bisher üblichen – Verfahren muss in Abhängigkeit vom jeweiligen Nutzungsvorhaben ein Gesetz verabschiedet werden, in dem bezogen auf die benötigten Dateien Verwendungszweck, Übermittlungsmodalitäten wie auch alle anderen damit zusammenhängenden fachlich-technischen Sachverhalte rechtlich geregelt werden. Dieses einzelgesetzliche Verfahren ist auch gängige Praxis der gegenwärtigen Registernutzung durch die Statistischen Ämter; jüngstes Beispiel ist die Regelung des Zugangs zu den für den Aufbau und die spätere Pflege des Unternehmensregisters benötigten administrativen Dateien im Rahmen des Statistikregistergesetzes (vgl. insbesondere §§ 2 bis 6 StatRegG).

– Schaffung einer allgemeinen Zugangsregelung für die Zwecke der amtlichen Statistik

Dies würde einer Art „Generalklausel“ gleichkommen, die es den Statistischen Ämtern ermöglicht, auf Anforderung Zugang zu administrativen Dateien zu bekommen. Eine entsprechende Regelung könnte in das Bundesstatistikgesetz integriert werden. Im IAW-Gutachten wird der Vorschlag gemacht, am §8 BStatG „Aufbereitung von Daten aus dem

Verwaltungsvollzug“ anzuknüpfen, wobei dieser ergänzt oder umformuliert werden könnte, weg von der Aufbereitungsregelung hin zu einer flexibleren Nutzungsregelung für statistische Zwecke.¹⁴⁾

Die Diskussion um die Ausgestaltung einer zukünftigen Zugangsregelung, die den Statistischen Ämtern eine effizientere Nutzung von Verwaltungsdaten ermöglichen würde, wird gegenwärtig intensiv geführt. Aus rein fachlicher Sicht ist eine allgemeine Zugangsregelung eindeutig zu präferieren.

Zu (4): Einsatzmöglichkeiten für administrative Daten

Vierte Voraussetzung ist die (Teil-)Substituierbarkeit von Primärerhebungen, d. h. innerhalb des primärstatistischen Berichtssystems müssen Erhebungen vorhanden sein, deren Merkmale zumindest teilweise durch administrative Daten ersetzt werden können.

Das IAW hat im Rahmen seines o. g. Gutachtens das komplexe System der amtlichen Primärerhebungen im Bereich der Wirtschaftsstatistik breitgefächert auf Einsatzmöglichkeiten für administrative Daten untersucht und kommt dabei zu folgenden Ergebnissen:

- Bei ungefähr 85 % der untersuchten Primärerhebungen sind generell Substitutionsmöglichkeiten gegeben.
- In 68 % der Fälle sind *Teils*substitutionen denkbar; d. h. es könnten nicht alle, aber zumindest einige Merkmale einer Primärerhebung durch Daten aus administrativen Dateien ersetzt werden.
- 17 % der Primärerhebungen im Bereich der Wirtschaftsstatistiken könnten nach IAW-Angaben vollständig ersetzt werden, was unmittelbar Kostenreduktion bei den Statistischen Ämtern und Belastungsreduktion bei den Befragten nach sich ziehen würde.

Als Beispiele nennt das IAW u. a. die vierteljährliche Handwerksberichterstattung und die Monatserhebungen im Handel und Gastgewerbe. Darüber hinaus wären auch mehrjährige Totalzählungen, wie die Arbeitsstättenzählung, die Handels- und Gaststättenzählung oder die Handwerkszählung in Zukunft substituierbar.

- 15 % der Primärerhebungen sind in keiner Form ersetzbar, da die vorhandenen administrativen Dateien hierzu keine geeigneten Informationen zur Verfügung stellen.

Das IAW weist in seinem Gutachten jedoch ausdrücklich darauf hin, dass Grundbedingung für einen effizienten Einsatz von Verwaltungsdaten vorher durchzuführende umfangreiche „Trimmungsmaßnahmen“ seien.¹⁵⁾ Das heißt, es müssten Änderungen und Erweiterungen in den Verwaltungsdateien (DV-Technik, Aufnahme von zusätzlichen Merkmalen, Änderungen in den Aufbereitungs- und Auswertungsmethoden und -periodizitäten usw.) vorgenommen werden. In diesem Zusammenhang wird auch betont, dass dieser Paradigmenwechsel bedeuten würde, dass sich die amtliche Statistik an die Realitäten der Registerinhalte anpassen müsste, sowohl aus fachlich-methodischer (Merkmalsdefinitionen, Berichtskreise usw.) als auch aus qualitativer Sicht (Akzeptanz von Proxi-Größen aus Verwaltungsdateien).

14) Vgl. Helbig, J., Engelage, Ch., Wiegert, R.: Möglichkeiten verstärkter Nutzung von Verwaltungsdaten für Zwecke der Wirtschaftsstatistik und zur Entlastung der Wirtschaft von statistischen Berichtspflichten, Gutachten für das Bundesministerium für Wirtschaft und Technologie, Tübingen, 1999, S. 201 f.

15) Vgl. Helbig, J., Engelage, Ch., Wiegert, R.: Möglichkeiten verstärkter Nutzung von Verwaltungsdaten für Zwecke der Wirtschaftsstatistik und zur Entlastung der Wirtschaft von statistischen Berichtspflichten, Gutachten für das Bundesministerium für Wirtschaft und Technologie, Tübingen, 1999, S. 135 ff.

Das IAW führt abschließend aus, dass die Chancen für eine veränderte Praxis der Datengewinnung in der amtlichen Wirtschaftsstatistik entscheidend von der politischen Bereitschaft zur Schaffung der hierfür notwendigen fachlichen und rechtlichen Rahmenbedingungen abhängig sei. Vor Konkretisierung so weitreichender Umgestaltungen im Bereich der Datengewinnung empfiehlt das IAW weitere Forschungsarbeiten mit dem Ziel einer umfassenden Evaluierung der Vor- und Nachteile.¹⁶⁾

5 Projekt „Registergestützte Umsetzung der Konjunkturverordnung im Dienstleistungsbereich“

Die Statistischen Ämter müssen an das Statistische Amt der Europäischen Gemeinschaften (Eurostat) ein umfangreiches Programm an Strukturmerkmalen (spätestens ab 2001) bzw. Konjunkturmerkmalen (spätestens ab 2003) liefern. Diese Datenanforderungen sind verbindlich in der sogenannten Strukturverordnung¹⁷⁾ und Konjunkturverordnung¹⁸⁾ festgelegt.

Von geringfügigen Änderungen/Ergänzungen einmal abgesehen, können die geforderten Strukturdaten aus dem bestehenden Berichtssystem bereitgestellt werden, vorausgesetzt, eine Dienstleistungsstatistik wird – wie derzeit geplant – eingeführt.

Probleme gibt es jedoch im unterjährigen Bereich: Die geforderten Konjunkturindikatoren für die Bereiche Verkehr und unternehmensnahe Dienstleistungen¹⁹⁾ können über das bestehende primärstatistische Berichtssystem *nicht* befriedigt werden. Da die Einführung einer *unterjährigen* Dienstleistungsstatistik, für die ca. 40 000 Unternehmen vierteljährlich auskunftspflichtig sein müssten,²⁰⁾ politisch nicht gewollt ist, wird im Statistischen Bundesamt gegenwärtig eine Strategie entwickelt, die die Bereitstellung der geforderten Informationen anhand der Nutzung von Einzeldaten aus administrativen Dateien zum Ziel hat.

Um die gewünschten Schlüsselvariablen „Umsatz“ und „Beschäftigte“ entsprechend den Vorgaben der Konjunkturverordnung bereitstellen zu können, sollen Einzeldatensätze aus der Betriebs-/Versichertendatei der Arbeitsverwaltung und aus den Umsatzsteuerdateien der Länderfinanzverwaltungen, mit einer aus dem Unternehmensregister gebildeten Leitdatei verknüpft, anschließend zu Unternehmensdaten verdichtet und hochgerechnet werden.

Zu diesem Zweck muss die Bundesanstalt für Arbeit den Statistischen Ämtern vierteljährlich Betriebsdaten zu den sozialversicherungspflichtig Beschäftigten übermitteln. Dies ist möglich, da die Bundesanstalt für Arbeit aus dem – schon skizzierten – integrierten Meldeverfahren zur Sozialversicherung monatlich Angaben über Zugänge und Abgänge von sozialversicherungspflichtig Beschäftigten bekommt, die in der sogenannten Versichertendatei systematisch gespeichert werden. Diese Versichertendaten können über eine Betriebsnummer zu

16) Ebenda, S. 137.

17) Verordnung (EG, Euratom) Nr. 58/97 des Rates vom 20. Dezember 1996 über die strukturelle Unternehmensstatistik, in: Amtsblatt der Europäischen Gemeinschaften Nr. L 14/1 vom 17.1.97, geändert durch Verordnung (EG, Euratom) Nr. 410/98 des Rates vom 16. Februar 1998, in: Amtsblatt der Europäischen Gemeinschaften Nr. L 52 U vom 21.2.1998.

18) Verordnung (EG) Nr. 1165/98 des Rates vom 19. Mai 1998 über Konjunkturstatistiken, in: Amtsblatt der Europäischen Gemeinschaften L 162/1 vom 5.6.98.

19) Gemeint sind die Abschnitte I (Verkehr, Nachrichtenübermittlung) und K (Grundstücks- und Wohnungswesen, Datenverarbeitung, Forschung und Entwicklung, Werbung, Rechts-, Steuer-, Unternehmensberatung, Markt- und Meinungsforschung, Architektur- und Ingenieurbüros, Reinigungsfirmen, u. a.) gemäß der NACE Rev. 1.

20) Falls eine 5%-Stichprobe zugrundegelegt wird.

Betriebsdaten aggregiert werden. Die Betriebsdaten wiederum lassen sich – ebenfalls über die Betriebsnummer – mit dem Unternehmensregister der Statistischen Ämter verknüpfen.

Was die Bereitstellung des Merkmals „Umsatz“ betrifft, ist die termingerechte Lieferung von Vierteljahresangaben aus den Umsatzsteuerdateien der Länderfinanzverwaltungen notwendig. Auch dies ist möglich, da die Unternehmen aus dem Umsatzsteuer-Voranmeldungsverfahren verpflichtet sind, monatlich (große Unternehmen) bzw. vierteljährlich (kleine und mittlere Unternehmen) ihre Umsätze an die Oberfinanzdirektionen zu melden. Diese Daten werden dann in den Umsatzsteuerdateien systematisch gespeichert und können über die Steuernummer mit den Daten aus dem Unternehmensregister verknüpft werden.

Das gesamte Vorhaben steht und fällt mit der Existenz eines umfassenden und funktionsfähigen Unternehmensregisters. Das Unternehmensregister wird nicht nur als Schnittstelle für die Verknüpfung der administrativen Dateien fungieren, sondern wird auch in Kombination mit der geplanten jährlichen Dienstleistungsstatistik Strukturdaten für die unterjährige Fortschreibung bereitstellen. Darüber hinaus soll das Unternehmensregister dafür sorgen, dass die Unternehmensdaten wirtschaftszweigsystematisch richtig zugeordnet werden.

Der zeitlichen Realisierung des Projektes sind durch die Vorgaben in der Konjunkturverordnung feste Termine gesetzt. Gemäß Konjunkturverordnung sind die Statistischen Ämter verpflichtet, Eurostat Konjunkturdaten für den Dienstleistungsbereich spätestens ab dem Jahr 2003 bereitzustellen. Aufgrund dieser Vorgabe muss das Konzept spätestens zu diesem Termin praxistauglich sein.

Folgende Aufgabenblöcke stehen bis dahin zur Erledigung an:

- Vor Gesetzesinitiierung ist es notwendig, Rückschlüsse über die Validität der zu erwartenden Ergebnisse und über den Arbeits- und Kostenaufwand in den registerführenden Stellen und in den Statistischen Ämtern zu bekommen. Zu diesem Zweck sind Testauswertungen mit Daten der Finanzverwaltungen und der Bundesanstalt für Arbeit geplant. Erste Ergebnisse aus diesen Tests sollen spätestens Mitte 2000 vorliegen.
- Schaffung der gesetzlichen Grundlage für die vierteljährliche Übermittlung von Einzeldaten aus den benötigten Verwaltungsdateien an die Statistischen Ämter bis spätestens Ende 2001.
- Entwicklung des fachlichen Konzepts für den Datentransfer und die Datenverknüpfung, die Datenaufbereitung (v. a. Entwicklung von Plausibilitätskontrollen) und die Datenauswertung (u. a. Entwicklung von Schätzmodellen zur Beseitigung von Registerinkongruenzen) bis Ende 2001.
- Durchführung eines umfassenden Praxistests sowohl des technischen Ablaufs als auch der fachlich-methodischen Projektbausteine im Jahr 2002. Da die skizzierte registergestützte Datengewinnung „Neuland“ für die amtliche Statistik darstellt, kommt diesem umfassenden Praxistest große Bedeutung zu.

Kritischer Faktor für eine termingerechte Umsetzung der Planungen ist die rechtzeitige Schaffung der gesetzlichen Grundlage für einen unterjährigen Zugang zu den benötigten administrativen Daten. Darüber hinaus ist für die fachliche und organisatorisch-technische Umsetzung des Projekts die intensive Kooperation zwischen Statistischen Ämtern und dateiführenden Stellen (Länderfinanzverwaltungen, Bundesanstalt für Arbeit) unentbehrlich.

6 Bewertung und Ausblick

Der Aufbau eines eng verzahnten, auf Registerdaten und primärerhobene Daten zurückgreifendes Statistikersystems ist ein wirkungsvolles Instrument, um den eingangs beschriebenen Zielkonflikt – Zwang zur Ausdehnung des Datenangebots auf der einen Seite und Zwang zur Belastungsreduktion bei den Berichtspflichtigen/Kostenreduktion in der Datengewinnung auf der anderen Seite – auflösen zu können.

Zum einen würde ein entsprechendes Berichtssystem zu unmittelbaren *Entlastungen der Wirtschaft von Berichtspflichten* führen. Ein den gesamten privaten Wirtschaftsbereich umspannendes und über primäre (amtliche Erhebungen) und sekundäre (Verwaltungsregister) Quellen kontinuierlich aktualisiertes Unternehmensregister würde in mehrjährigen Abständen stattfindende – sehr kostenintensive – Vollerhebungen (Zensen) überflüssig machen. Darüber hinaus könnten auch – zumindest wenn man den Ausführungen des IAW-Gutachtens folgt – bestehende Struktur- und Konjunkturerhebungen zur Disposition gestellt werden, ganz zu schweigen von der im vorigen Gliederungspunkt beschriebenen Möglichkeit, den *zusätzlichen* unterjährigen EU-Datenbedarf im Dienstleistungsbereich *ohne* neue Konjunkturerhebung erfüllen zu können.

Ausmaß der Substitutionsmöglichkeiten und Effizienz des Datengewinnungsprozesses hängen jedoch entscheidend davon ab, inwieweit es gelingt, die Verwaltungsregister und den technischen Datentransfer möglichst optimal auch auf die Belange der Statistik abzustimmen (Trimmung).

Unter der Voraussetzung eines effizienten, d. h. alle technischen Möglichkeiten ausschöpfenden IT-Einsatzes im Datenmanagement resultiert aus einem institutionalisierten, weitestgehend automatisierten Datentransfer von den dateiführenden Stellen an die Statistischen Ämter die Chance zur schnelleren Datenbereitstellung und damit zur Erhöhung der Aktualität der statistischen Berichterstattung.

Was die Trimmung der Verwaltungsregister aus methodisch-inhaltlicher Sicht anbelangt, muss man sich jedoch bewusst sein, dass Maßnahmen zur Verringerung der Inkongruenzen zwischen Statistik und Verwaltungsdaten, wie zum Beispiel die Umstellung auf andere statistische Einheiten und/oder Merkmalsanpassungen, in den Verwaltungsregistern eher enge Grenzen gesetzt sein werden. Um Substitutionsmöglichkeiten auszuschöpfen, würde die amtliche Statistik letztendlich nicht umhin kommen, im Rahmen ihrer Flexibilität und methodischen Anpassungsfähigkeit sich diesen Gegebenheiten zu stellen. Dies würde auch bedeuten, unter der Maßgabe der Entlastung von Berichtspflichtigen in gewissem Umfang Proxi-Größen aus Verwaltungsregistern akzeptieren zu müssen.

Die Nutzung von administrativen Daten kann jedoch auch qualitative Vorteile zeitigen: In den bestehenden Berichtskreisen sind kleine Unternehmen teilweise ausgeklammert. Mit der Nutzung von Angaben aus Verwaltungsregistern würden aktuelle, bereichsumspannende zentrale Unternehmensgrößen (Umsatz, Beschäftigte) *ohne* Abschneidegrenzen zur Verfügung stehen²¹⁾, die unmittelbar in den Auswertungsprozess einfließen und die statistische Datenbasis verbreitern könnten.

21) Einmal abgesehen von den Kleinstunternehmen mit weniger als 32 500 DM Jahresumsatz, die in den Umsatzsteuerdateien der Länderfinanzverwaltungen fehlen, und von (Klein-)Betrieben ohne sozialversicherungspflichtig Beschäftigte, die in den Dateien der Bundesanstalt für Arbeit nicht enthalten sind.

Es ist offensichtlich, dass eine registergestützte Datengewinnung in der Aufbauphase zusätzliche Kapazitäten in den Statistischen Ämtern binden würde. Darüber hinaus müssten auch die dateiführenden Stellen zusätzlichen Aufwand für Statistikbelange leisten. Dieser Aufwand muss jedoch in Relation zum Ressourceneinsatz gesetzt werden, der notwendig wäre, alle (zusätzlichen) nationalen und internationalen Datenanforderungen auf konventionellem Wege über Primärerhebungen abzudecken. Dies wäre – neben der (zusätzlichen) Belastung der Berichtspflichtigen – deutlich kostenintensiver als die Auswertung von Daten, die im Prozess des Verwaltungshandelns quasi „automatisch“ anfallen.

Nimmt sie die Bemühungen um „Verwaltungsmodernisierung“ ernst, kommt die amtliche Statistik an einer verstärkten Nutzung vorhandener Informationen aus administrativen Dateien in Zukunft wohl kaum vorbei.

Michael Hoppe*)

Aufbau und Organisation eines Access-Panels

1 Einleitung

Eingangs möchte ich erwähnen, dass das Instrument Access-Panel im Prinzip nicht neu ist. Ich glaube aber, dass das Access-Panel aktueller denn je ist. Darauf gehe ich später noch einmal ein. Insgesamt gibt es Access-Panels schon seit über 50 Jahren, in Deutschland seit über 20 Jahren. Wir haben das Panel vor ca. 10 Jahren aufgebaut und noch drei bis vier Mitbewerber in Deutschland bieten vergleichbare Panels an, davon auch einige, die europaweit oder weltweit agieren, u. a. auch IPSOS.

Der Vortrag wird unterteilt in vier verschiedene Schwerpunkte.

Abbildung 1

	Inhalt
	<ul style="list-style-type: none">- Was ist ein Access-Panel?- Wie wird ein Access-Panel aufgebaut?- Wie wird ein Access-Panel gepflegt?- Panel-Usage-Policy

*) Dr. Michael Hoppe, Ipsos Deutschland, Hamburg.

2 Paneldefinition

Der Panelbegriff ist im allgemeinen Sprachgebrauch eingengt auf eine Definition, die sich meist auf traditionelle Handelspanels bezieht.

Abbildung 2

The image shows a slide titled "Was ist ein Panel?" with the Ipsos logo in the top left corner. The main text is a definition in italics: "„Als Panel wird im Rahmen der Marktforschung ... ein bestimmter, gleichbleibender Kreis von Adressaten verstanden, bei dem in regelmäßigen zeitlichen Abständen Erhebungen zum (im Prinzip) gleichen Untersuchungsgegenstand durchgeführt werden.“". Below this is a citation: "Berekoven, Ludwig (1996): Marktforschung: Methodische Grundlagen und Anwendung, Wiesbaden". At the bottom, there is a paragraph: "Traditioneller Panelbegriff, bezogen auf Handels- und Verbraucherpanel, der die methodologischen Probleme der Panelsterblichkeit, des Paneffekts und der Panelerstarrung transportiert."

So wie im ersten Absatz beschrieben, verstehen wir ein Access-Panel *nicht*, denn mit diesem eher traditionellen Panel-Begriff sind ja auch viele Eigenschaften verbunden, die auf ein Access-Panel so nicht zutreffen. Das Access-Panel ist *nicht* verbunden mit Panelsterblichkeit in herkömmlicher Weise, bei der Haushalte auf ganz natürliche Art und Weise wegfallen; auch nicht mit dem Paneffekt, der beschreibt, dass je mehr man die Haushalte nutzt, diese ein anderes Kaufverhalten angeben, als es der Realität entspricht. Auch eine Panelerstarrung, die bedeutet, dass die Haushalte gleichmäßig immer älter werden, ist bei unserer Art von Panels nicht festzustellen.

Was ist dann ein Access-Panel?

Abbildung 3



Was ist ein Access-Panel?

- Ein Panel ist ein Pool vorrekrutierter Haushalte, die sich zur Zusammenarbeit bereit erklärt haben und antwortbereit sind.
- Access-Panel bedeutet dabei die zeit- und kosteneffiziente Auswahl eines bestimmten (begrenzten) Kreises von Testteilnehmern aus diesem Pool zum Zwecke der Zielgruppenerhebung und (einmaligen) Verbraucherbefragung.
- Von diesen Haushalten liegt eine Vielzahl an Informationen vor, die in einer Datenbank gespeichert sind.
- Verwendungsdaten, die nicht ad hoc vorliegen, können kostengünstig erhoben werden (z.B. via Omnibus-Screening)
- Diese Zielgruppen stehen ad hoc für Befragungen und Tests zur Verfügung.

Ein Access-Panel ist im Endeffekt ein Pool, und nur so möchte ich das auch bezeichnen, von vorrekrutierten Haushalten, die sich bereit erklärt haben, mit uns zeitweise und ad hoc, aber nicht kontinuierlich zusammenzuarbeiten. Das heißt, wir haben die Haushalte rekrutiert und diese haben gesagt: „Ja, ich bin bereit.“ Das bedeutet in der Praxis, dass wir ihnen Untersuchungsmaterial schicken, und dies kann ein Konzept, ein Produkt, ein U + A¹⁾-Fragebogen oder irgend etwas anderes sein, und die Haushalte erklären unverbindlich: „Ja, ich werde bei dieser Befragung mitmachen.“ Mehr ist es nicht. Die Haushalte verpflichten sich also nicht, regelmäßig daran teilzunehmen.

Im Endeffekt ist ein Access-Panel eine sehr effiziente Art der Befragung, denn die gesamte Befragungsdauer ist sehr kurz, auch die Selektionen gehen relativ schnell vonstatten. Man muss nicht vorrekrutieren, nicht lange nach Zielpersonen suchen, denn wir haben die meisten relevanten Daten in der Datenbank. In der Regel werden die normalen Demographien und die für unsere Kunden und potentiellen Kunden notwendigen Kriterien, d. h. Marken- und Produkt- oder Dienstleistungsverwendungen, abgefragt, z. T. auch Einstellungen oder die Bekanntheit von Marken und Produkten.

In der Praxis führen wir überwiegend Zielgruppenerhebungen durch. Der Kunde gibt uns Quoten vor, und danach wählen wir die Versuchspersonen aus der Datenbank aus. Auch bei Quoten mit sehr geringen Inzidenzraten ist es in der Regel kein Problem, diese sofort zu rekrutieren. Diese Zielgruppen stehen ad hoc zur Verfügung, d. h., wir nutzen sie nur dann oder ziehen sie nur dann, wenn wir einen entsprechenden Auftrag dafür haben. Das ist ein kontinuierlicher Prozess, der über Monate und Jahre hinweg immer wiederholt werden kann.

1) U + A = *usage and attitude*.

3 Aktualität der Access-Panels

Ich möchte nochmals auf die anfangs schon erwähnte Aktualität von Access-Panels eingehen. Die aktuellen Herausforderungen an die Marktforschung im Allgemeinen oder an diese Art von Panels im Speziellen sind durch mehrere Kriterien gegeben.

Abbildung 4



Wir stellen zum einen eine zunehmende Reserviertheit der Bevölkerung gegen Interviews fest, ich meine in diesem Fall Inhome-Interviews, bei denen die Interviewer zu Hause an der Tür klingeln und die Versuchsperson befragen möchten. Die Schwierigkeiten, in die Haushalte zu gelangen, werden immer größer. Das hat nicht nur mit der Angst vor Kriminalität zu tun, sondern einfach auch damit, dass man nicht gestört sein möchte in seiner Privatsphäre. In den USA, das haben Sie sicherlich auch schon mehrfach gehört oder wissen es, gibt es praktisch keine Inhome-Befragungen mehr. Es werden dort überwiegend Access-Panel- oder Telefonbefragungen durchgeführt, neuerdings auch Online-Befragungen. Sie können dort vielleicht einmal in einem Einkaufszentrum Probanden befragen und in ein Lokal hineinbitten, aber unser klassisches Inhome-Random-Route ist in den USA überhaupt nicht möglich. Ich glaube, dass es in Deutschland und Europa irgendwann einmal ähnlich sein wird, es dauert sicher noch. Aber wir merken zunehmend, dass es immer schwieriger wird, klassische Interviews durchzuführen.

Es wird auch immer schwieriger, Interviewer zu rekrutieren. In diesem Zusammenhang muss auch erwähnt werden, dass die deutsche Gesetzgebung es uns nicht gerade einfacher macht.

Aufgrund des Gesetzes zur Korrektur der Scheinselbständigkeit kündigen uns signifikant mehr Interviewer, als es früher der Fall war. Die Interviewer haben immer weniger Lust, bei relativ geringer Bezahlung und der Hemmschwelle, in fremde Haushalte zu gehen, für die Marktforschung zu arbeiten und sind jetzt noch zunehmend bzgl. ihres Selbständigenstatus und der möglichen Sozialabgaben verunsichert.

Die Zielgruppen fraktionieren sich zudem auch immer stärker. Bei der Übersättigung der Märkte suchen unsere Kunden z. T. immer kleinere und spitzere Zielgruppen, die dann auch mit speziellen Produkten und entsprechender Ansprache umworben werden. Bei diesen z. T. geringen Inzidenzen und sehr kleinen Zielgruppen hätten wir selbst mit unseren ca. 2 500 Inhome-Interviewern enorme Probleme, die Befragungen im angemessenen Zeitraum zu angemessenen Kosten durchzuführen. Dies ist eine hervorragende Aufgabe für ein Access-Panel, solche Zielgruppen zu finden. Bei ca. 30 000 – 35 000 Haushalten mit über 80 000 Versuchspersonen ist es in der Regel immer möglich (sofern die gesuchten Quoteninformationen gespeichert sind), innerhalb kurzer Zeit die Befragungen ordnungsgemäß durchzuführen. Selbst wenn die Zielkriterien nicht gespeichert sind, ist es in der Regel günstiger, bestimmte vorselektierte Haushalte mit höherer Trefferwahrscheinlichkeit vorab mit einer Postkarte anzuschreiben, als die gesamte Interviewerorganisation zur Rekrutierung schwieriger Zielgruppen einzusetzen. Aber die Zielgruppen sind nicht immer so schwierig zu bestimmen. Wir haben auch ganz normale Quoten, Verwender oder Nichtverwender einer bestimmten Marke, einkommensstarke Haushalte oder ähnliches. Bei diesen Normalquoten tritt dann insbesondere der Zeitvorteil in den Vordergrund, bei kurzen Befragungen auch der Kostenvorteil.

Die Konsistenz der Methoden in verschiedenen Ländern und die globale Standardisierung ist eine zunehmende Anforderung an die Marktforschung. Ein zunehmender Teil von Kunden reduziert die Anzahl der europa- oder gar weltweit eingesetzten Institute, um dann lieber mit einigen wenigen verbleibenden Instituten eine umfassende Standardisierung der Methoden, Fragebogen, Zielgruppen etc. ebenso wie des Ablaufes durchzuführen. Meistens bedeutet dies auch, dass unser Auftraggeber für eine Mehrländerstudie, gleich in welchen Erdteilen, auf der Institutsseite nur noch einen Ansprechpartner wünscht, der dann intern die Untersuchungen in den einzelnen Ländern organisiert. Unsere Kunden sparen somit die eigenen Organisationskosten für mehrere Länder, die Institute müssen aber andererseits in der Lage sein, ohne großen Mehraufwand diese Studien in mehreren Ländern durchzuführen.

Auch dies ist eine vorzügliche Herausforderung für die Access-Panels, da wir hier bereits in mehreren Ländern und Erdteilen mit einheitlichen Rekrutierungsfragebögen arbeiten, die Datenbanken international eingesetzt werden und die standardisierten Fragebogen bereits in mehreren Sprachen abgestimmt übersetzt vorliegen. Dadurch gibt es kaum noch Diskussionen über die Zielgruppe oder welche Stichprobe gezogen werden soll, wie viele Fragebogen oder Produkte, sondern es gibt ein ganz kurzes, eine Seite langes Briefing. Auftraggeber und Auftragnehmer kennen die Preise, wir haben die Methoden und alle Standards. Das heißt, es gibt keine lange Diskussion und Abstimmung, es ist überwiegend alles bekannt.

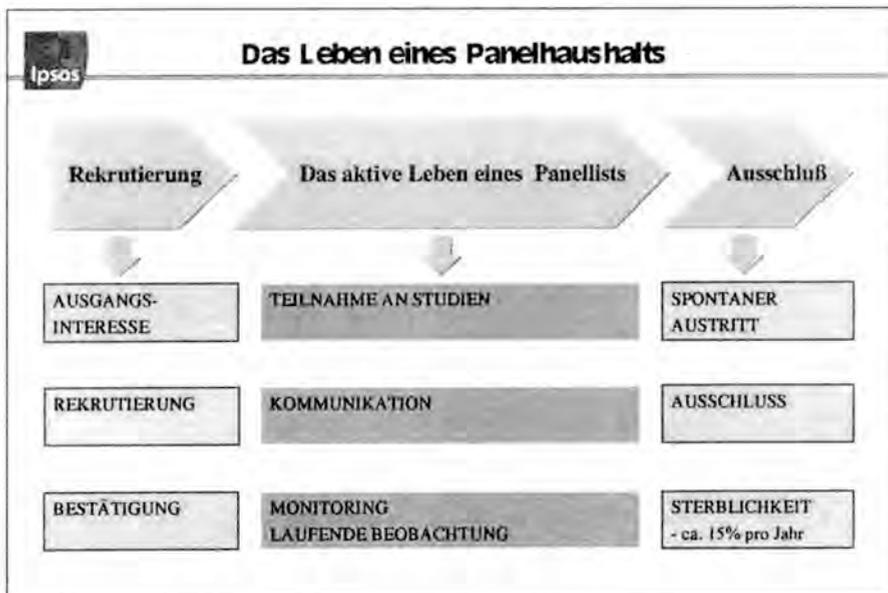
Das ist natürlich nur möglich, wenn man entsprechende Partner auf der Forscher- und Institutsseite hat, die diese Standardisierung mitmachen und akzeptieren. Wir sind aber als internationales Institut mit globalen Kunden zu dieser Arbeitsweise verpflichtet, denn dies ist ein zunehmender Trend der Globalisierung und der Standardisierung.

Der Kostendruck auf die Marktforschung und auf die Institute wird wie überall immer stärker. Wir müssen nicht nur schneller werden, wir wollen auch noch kostengünstiger arbeiten, und das Ganze möglichst verbunden mit einem verstärkten Dienstleistungsangebot. Was macht man also, wenn die Suche von Versuchspersonen immer schwieriger wird, die Rekrutierung der Interviewer teurer, die Durchführung kostenintensiver? Wir müssen andere Methoden anwenden. Eine davon ist mit Sicherheit das Access-Panel. Es ist insbesondere bei kürzeren Fragebögen, schwierigen Zielgruppen und zeitlichen Engpässen eine hervorragende Alternative.

4 Aufbau eines Access-Panels

Am Anfang steht die Rekrutierung der Probanden, danach folgt das aktive Panel-„Leben“ und irgendwann später, im Durchschnitt nach 5 bis 8 Jahren, der freiwillige, aber von uns gewollte Ausschluss.

Abbildung 5



In der Rekrutierungsphase bemühen wir uns, ein Anfangsinteresse zur Teilnahme bei den Haushalten zu provozieren. Dazu gibt es mehrere Möglichkeiten.

Abbildung 6



Rekrutierung von Haushalten

- Woher nehmen, wenn nicht stehlen?
 - Adressen kaufen (Adressverlage)
 - aus Telefonbüchern Adressen selektieren
 - andere Feldinstrumente nutzen wie z. B.
 - MTU-Umfragen
 - CATI-Befragungen
 - CAPI-Erhebungen
 - persönliche Rekrutierung durch Interviewer
 - Schneeball-Aktionen (Member-get-Member)

Wir können Adressen bei Adressverlagen kaufen, was insbesondere bei speziellen Zielgruppen (z. B. Mütter mit Babys, Gebissträger o. ä.) oder regionalen Schwerpunkten der Fall ist. Oder wir rufen die Haushalte entweder nach vorselektierten Telefonnummern oder nach Zufallsauswahl an.

Am kostengünstigsten sind immer Aktionen im Zusammenhang mit anderen Untersuchungen, die ohnehin stattfinden. Das heißt zum Beispiel, die Interviewer fragen am Ende des Telefon- oder mündlichen Interviews das Interesse an einer Panelteilnahme ab oder stecken Postkarten in die umliegenden Briefkästen.

Am erfolgreichsten sind die Schneeball-Aktionen bei bestehenden Panelhaushalten (member-get-member), die allerdings nicht zu häufig durchgeführt werden dürfen. Hierzu befragen wir unsere Teilnehmer, ob sie nicht noch andere Haushalte mit Interesse kennen. Bei nahezu 35 000 Haushalten mit ca. 85 000 Teilnehmern erhalten wir in der Regel mehr neue Haushalte, als wir alte im gleichen Zeitraum verlieren. Zunehmend passiert es auch, dass unser Panel von Interessierten angeschrieben wird, die von anderen Haushalten über unsere Aktivitäten gehört haben.

Die große „Kunst“ der Rekrutierung besteht in der richtigen ersten Ansprache, um überhaupt ein erstes Interesse und Vertrauen zu erwerben. Haben die Haushalte ein erstes Anfangsinteresse gezeigt, erhalten sie unseren ausführlichen Fragebogen zur Erfassung der Demographie, Marken- und Produktverwendung, Einschätzung der eigenen Haut- oder Haareigenschaften, Verwendung von PKW, Dienstleistungen usw.

Abbildung 7



Ermittelte Merkmale

- Mittels eines Rekrutierungs-Fragebogens werden zunächst grundlegende Merkmale erhoben wie
 - ➔ Haushaltsinformationen (z. B. Adresse, Ortsgrößenklasse, Einkommen, Haushaltsgröße)
 - ➔ Informationen über alle Haushaltsmitglieder (z. B. Alter, Vor- und Nachnamen, Berufstätigkeit)
 - ➔ Informationen zum Besitzstand (z. B. Fernseher, Kühlschrank, Computer, Handy, Camcorder, Tiere)
 - ➔ Informationen zur Verwendung von einigen Konsumgütern (was wird wie häufig verwendet)

Nach Rücksendung dieses Erfassungsbogens und der unterschriebenen Datenschutzerklärung gehen diese Daten in die Datenbank ein und der neue Haushalt ist in die „Panelfamilie“ aufgenommen.

Es folgt das aktive „Familienleben“, und ich sage bewusst Familienleben. Ich könnte Ihnen stapelweise Postkarten, Briefe und Fotos zeigen, in denen Mitglieder dieses Access-Panels uns aus dem Urlaub schreiben, bei der Geburt von Kindern die Anzeige und Fotos schicken und alle anderen persönlichen Veränderungen mitteilen. Von unserer Seite beantworten wir die Schreiben und versenden regelmäßig Informationen über aktuelle Themen und Teile von Studienergebnissen, um die Bindung aufrechtzuerhalten und die Teilnehmer zu motivieren. Diese Kommunikation machen wir bewusst, um die Fluktuation möglichst klein zu halten. Die gesamte Kommunikation erfolgt unter einem Pseudonym, so dass die Panelteilnehmer nach außen hin immer den gleichen Ansprechpartner haben und das Gefühl der Verbundenheit aufkommt.

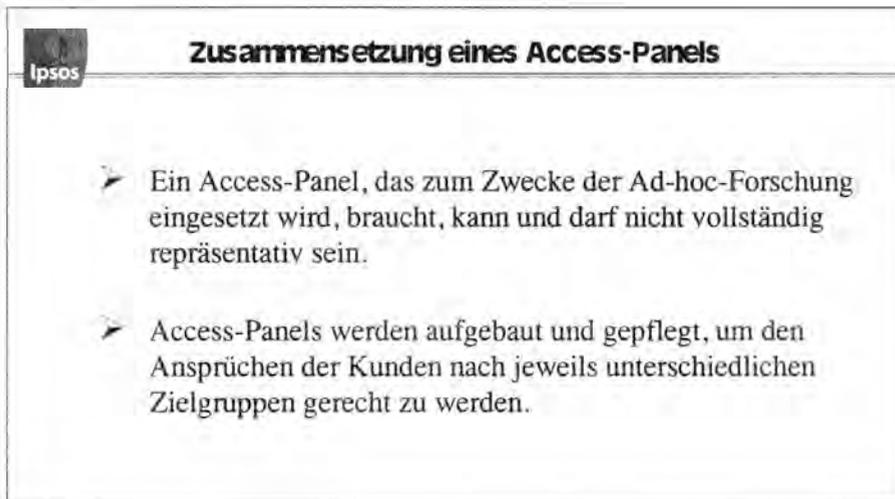
Auf der anderen Seite überprüfen wir die Haushalte auch in ihrem Antwortverhalten. Wenn die Fragebogen häufiger nicht oder nur teilweise ausgefüllt werden, bemühen wir uns um ein verändertes Verhalten. Falls dies nicht erfolgreich ist, beenden wir von uns aus die Zusammenarbeit. Genauso haben natürlich auch die Haushalte jederzeit die Möglichkeit zur Beendigung, indem sie einfach nicht mehr antworten oder uns dies schriftlich oder telefonisch mitteilen.

Die normale „Sterblichkeit“ liegt bei ca. 15 %, d. h., 15 % der Haushalte, also ca. 5 000, müssen jährlich neu rekrutiert werden (zuzüglich neuer Haushalte für spezielle Zielgruppen). Das heißt aber auch, dass uns die Haushalte im Durchschnitt ca. 6 bis 7 Jahre erhalten bleiben.

5 Repräsentativität

Immer wieder wird natürlich die Frage nach der Repräsentativität der Access-Panels gestellt. Es ist dabei die Frage, wie Repräsentativität definiert ist. Eine klassische Bevölkerungsrepräsentativität würde man nur erlangen, wenn alle Haushalte mit mehrschichtigen Auswahlmethoden random-route rekrutiert würden. Das haben wir nicht, wie aus den unterschiedlichen Rekrutierungsmethoden ersichtlich ist. Das brauchen wir auch nicht, um den Ansprüchen unserer Kunden bei Ad-hoc-Untersuchungen gerecht zu werden.

Abbildung 8



Zusammensetzung eines Access-Panels

- Ein Access-Panel, das zum Zwecke der Ad-hoc-Forschung eingesetzt wird, braucht, kann und darf nicht vollständig repräsentativ sein.
- Access-Panels werden aufgebaut und gepflegt, um den Ansprüchen der Kunden nach jeweils unterschiedlichen Zielgruppen gerecht zu werden.

In der Regel führen wir überwiegend Quotenstichproben durch und bemühen uns bei diesen Stichproben um eine regionale Verteilung nach Bundesländern und Ortsgrößen. Wichtiger ist, dass bei mehreren Teilstichproben (z. B. 4 verschiedene Konzept- oder Produkttests à 200 Versuchspersonen netto) diese untereinander nach den wichtigsten Demographiemerkmalen vergleichbar sind.

Für Teilzielgruppen können wir ein repräsentatives Abbild durch Quotenvorgaben schaffen, eine Bevölkerungsrepräsentativität können wir durch Vorgabe von verschiedenen Quotenmerkmalen (z. B. Bundesland, Ortsgröße, Haushaltsgröße, Alter, Einkommen, Beruf usw.) nur näherungsweise simulieren, aufgrund der unterschiedlichen Rekrutierungsmethoden aber nicht entsprechend z. B. einer einheitlichen Door-to-Door-Inhome-Befragung gewährleisten.

Abbildung 9

Ipsos

Repräsentativität

- Repräsentativität wird bei den Ipsos-Access-Panels nach folgenden Kriterien definiert:
 - Nielsen-Gebiet
 - Ortsgrößenklasse (4 Ortsgrößen)
 - Haushaltgröße (5 Kategorien)
 - Haushaltseinkommen (12 Kategorien)
 - Alter des Haushaltsvorstandes (5 Kategorien)
 - Beruf des Haushaltsvorstandes (5 Kategorien)

- Es werden in Europa Statistiken vom Statistischen Bundesamt, der Mediaanalyse oder landesspezifische Statistiken verwendet, um Quoten hierfür zu definieren.

Wie bei einer persönlichen Befragung erreichen wir sehr einkommenschwache ebenso wie sehr einkommensstarke Haushalte nur schwierig, anders als bei persönlichen Befragungen erreichen wir mobile Haushalte aber sehr gut. Um unsere Kundenbedürfnisse zu erfüllen, sind unsere Panels zweigeteilt.

Abbildung 10

Ipsos

Zusammensetzung der Ipsos-Access-Panels

- Die Ipsos-Access-Panels sind deswegen zweigeteilt:
- repräsentativer Teil
 - zur Durchführung repräsentativer Studien und Screenings (z. B. 10 000er Omnibus bei Ipsos)
- nicht vollständig repräsentativer Teil
 - gemäß des Kundenbedarfs an speziellen Zielgruppen (z. B. konsumstarke Haushalte, Haushaltsvorstand im Alter von 18 – 50 Jahren)
- beide Teile befinden sich im regelmäßigen Austausch unter Wahrung ihrer Bestimmung

Die sogenannte „repräsentative“ Omnibusbefragung bei 10 000 Haushalten wird von IPSOS in 4 europäischen Ländern (und auf Wunsch auch in den USA) 2- bis 4-mal jährlich durchgeführt (Euro-Select). Daneben bauen wir auf Kundenwunsch oder aus Eigeninitiative permanent weitere Spezialpanels auf bzw. nutzen den anderen Teil des Panels für „normale“ Quotenstichproben.

Abbildung 11



Zusammensetzung der Ipsos-Access-Panels

- Je nach Kundenwünschen können weitere Spezial-Panels daneben oder als Sub-Panels aufgebaut werden wie z.B.
 - ECAP European Chronic Ailment Panel
 - Baby-Panel
 - Mundhygiene-Panel (Zahnprothesenträger u.a.)
 - Haarpflege-Panel
 - usw.

Diese Haushalte gehen später z. T. auch in den allgemeinen Teil des Panels über und können auch für andere Aufgaben mitgenutzt werden.

6 Methodenvergleich

Abbildung 12



Zwei Fragen zur Methode

- 1 Liefere Panel-Stichproben identische Ergebnisse im Vergleich zu Stichproben, die durch Interviewer rekrutiert sind?
- 2 Liefere selbst-auszufüllende Fragen identische Ergebnisse im Vergleich zu konventionellen Face-to-Face-Interviews?

Zur Beantwortung dieser Fragen wurde eine vergleichende Studie durchgeführt

Die Untersuchung wurde so angelegt, dass zum einen die Ergebnisse einer Access-Panel-Stichprobe mit einer Random-Route-Stichprobe in Haushalten, bei denen der Fragebogen im Haushalt zur schriftlichen Beantwortung zurückgelassen wurde, verglichen wurde und zum anderen diese Random-Route/Selbstausfüller-Stichprobe mit einer klassischen Random-Route/Face-to-Face-Stichprobe.

Abbildung 13

 Testdesign			
	Studie A	Studie B	Studie C
Quelle	Access-Panel	Random	Random
Methode	postalisch	Hausbesuch, Fragebogen überlassen, postalisch zurück	Hausbesuch
Interview	Selbstausfüller	Selbstausfüller	Face-to-Face
Fallzahl	360	454	299

Abbildung 14

 Fragenkatalog	
Diverse Fragen zu folgenden Themen:	
<ul style="list-style-type: none">• Markenbekanntheit verschiedener Konsumgüter• Leserschaft verschiedener Magazine• Haushaltsausstattung• Bewertung eines ersten Konzeptes (Schokoriegel)<ul style="list-style-type: none">- Kaufbereitschaft- Preiseinstufung- Likes und Dislikes• Bewertung eines zweiten Konzeptes (Mobiltelefon)• Einstufung von 40 Statements zu Einstellungsfragen	

Diverse Fragen zur Markenbekanntheit, Leserschaft, Haushaltsausstattung etc. wurden von den Haushalten aus den unterschiedlichen Stichproben beantwortet. Die Ergebnisse zeigen deutlich, dass die Quelle der Stichprobe für die Interpretationen der Ergebnisse keinen Einfluss hat.

Abbildung 15

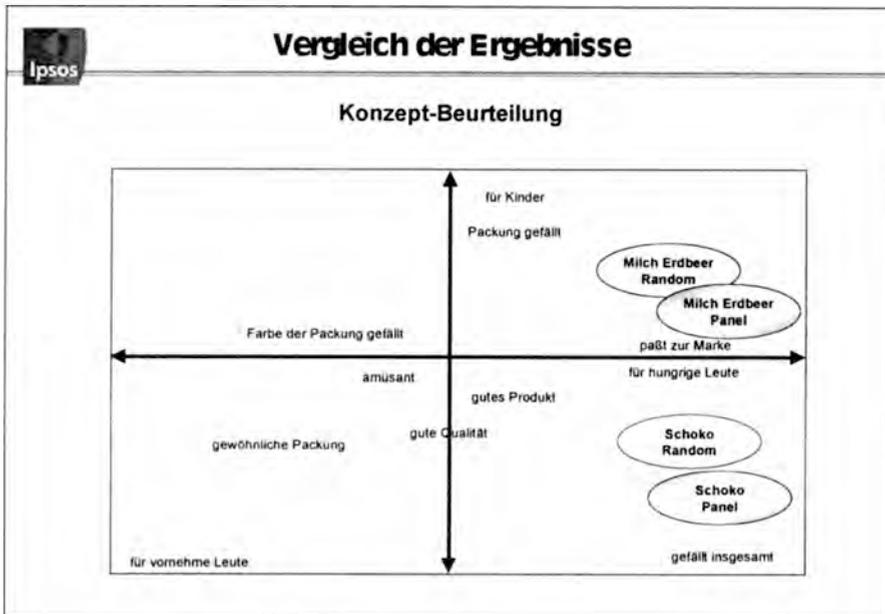


Abbildung 16

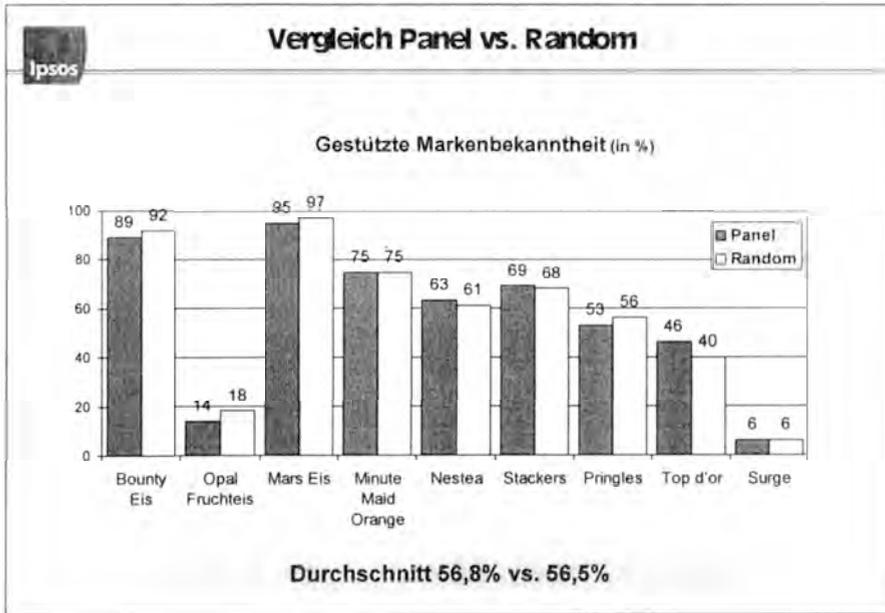
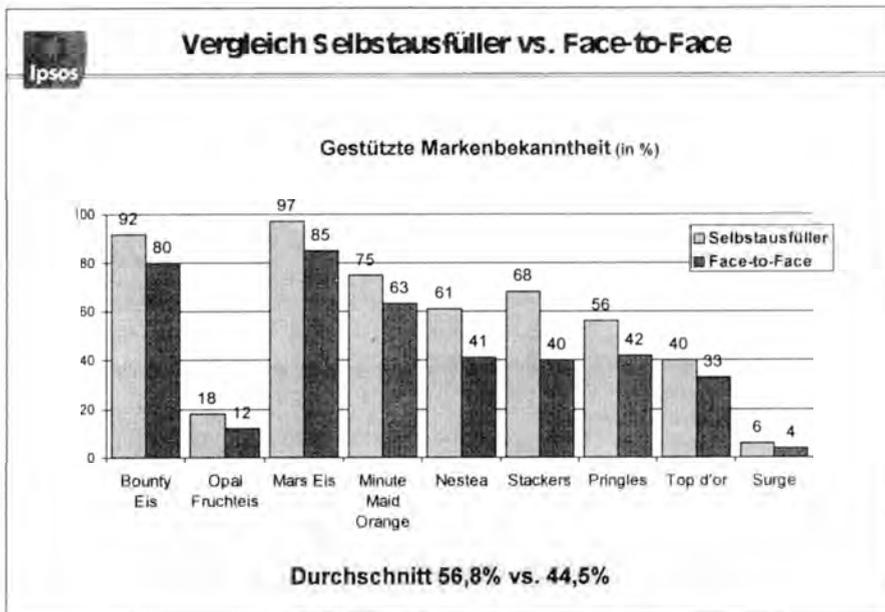


Abbildung 17



Es gibt keine signifikanten Unterschiede – die Empfehlungen sind die gleichen.

Anzumerken bleibt allerdings das unterschiedliche Antwortverhalten bei schriftlicher Beantwortung und mündlicher Befragung: In der mündlichen Befragungssituation tendiert der Befragte dem Interviewer gegenüber zu Gefälligkeitsantworten, um sein Image aufzubessern, ebenso wie die Befragten zu sozial erwünschten Antworten z. B. in Bezug auf Umwelt, Toleranz oder Solidarität neigen.

Andererseits ist die Antwortqualität in Bezug auf eine höhere Anzahl von Antworten (dies ist im übrigen noch viel extremer bei Online-Befragungen) im Panel besser (in der Face-to-Face-Situation kann der Interviewer nicht so schnell schreiben, es entstehen „peinliche“ Wartesituationen), man neigt hier auch zu „ehrlicheren“ Antworten und einige Fragen wie z. B. nach Inkontinenz, Sexualverhalten, Intimpflege u. ä. sind bei persönlichen Befragungen nur schwer oder gar nicht zu erhalten. Hier liegt auch ein deutlicher Vorteil der Access-Panels.

Abbildung 18

Ergebnisse der Studie

Ipsos

- Die Quelle der Stichproben spielt keine Rolle. Egal ob Panel oder Random - die Empfehlungen sind die gleichen.
- Unterschiede traten beim Vergleich verschiedener Datenerhebungsmethoden auf:
 - Gefälligkeitseffekt bei Interviewerstudien (mehr positive Antworten)
 - Befragte wollen ihr Standing/Image gegenüber dem Interviewer aufbessern
 - Antworten beim Face-to-Face-Interview tendieren zur „Political Correctness“ (z.B. Umweltschutz, Toleranz, Solidarität)
 - Ausfüllqualität bei offenen Fragen im Panel besser (höhere Anzahl von Antworten)
 - Im Panel insgesamt „ehrlicher“

7 Panelpflege

Wir haben ein sehr hohes Interesse, die Haushalte langfristig an unser Unternehmen zu binden, da die Neurekrutierung von Haushalten zusätzliche Kosten verursacht. Für die Panelpflege ist es neben der regelmäßigen studienunabhängigen Kommunikation notwendig, die Haushalte regelmäßig (also nicht nur einmal jährlich, ebensowenig wie 20-mal jährlich) mit Studien zu versorgen und mit unterschiedlichen Studientypen (z. B. Konzepttests, Produkttests, Omnibus) zu arbeiten. Wir bemühen uns um durchschnittlich 6 Studien jährlich, bei einem interessanten Studien- und Themenmix.

Abbildung 19



Das aktive Leben eines Panelhaushalts

- **Limitierung der Anzahl von Studienteilnahmen pro Haushalt**
 - nicht zu viele
 - nicht zu wenige
 - durchschnittl. 6 Studien pro Jahr
- **Teilnahme an Studienmix**
 - Konzepttests
 - Produkttests
 - Trackingstudien
 - Grundlagenstudien (Usage & Attitude, etc.)
 - Screenings
 - Omnibus

Wenn eine Versuchsperson zum fünften Mal einen Konzepttest zum gleichen Thema erhalten würde, auch wenn es dafür ein kleines Geschenk gäbe, müsste man nicht nur mit viel weniger Antworten rechnen, sondern auch mit einer Beendigung der Zusammenarbeit. Bei Produkttests reicht in der Regel das Produkt als „Geschenk“/Honorierung der Versuchsperson aus, bei Konzepttests gibt es meist Kleinigkeiten wie ein Nähset, Nageletui o. a. im Wert von DM 2,- bis DM 10,- bei U + A-Studien oder Trackings kann das Geschenk auch einen Wert von DM 20,- bis DM 50,- annehmen.

Daneben steht dem Haushalt im Rahmen der Panelpflege unser kostenloses Service-Telefon zur Verfügung, jeder Haushalt erhält zudem regelmäßig Newsletters und kommuniziert immer mit dem gleichen Pseudonym (Carola Friedrichs). Dazu gehört zur Panelpflege auch das regelmäßige Update der Haushaltsinformationen und das Beobachten der Teilnahmebereitschaft und Ausfüllqualität.

Abbildung 20



Das aktive Leben eines Panelhaushalts

- **Panelkommunikation**
 - kostenloses Service-Telefon
 - Newsletter (intern. Studien, häufig gestellte Fragen usw.)
 - nach außen hin: immer gleicher Ansprechpartner (Carola Friedrichs)
- **Monitoring**
 - regelmäßige Aktualisierung der Informationen über den Haushalt (Up-dating)
 - Beobachtung der Teilnahme/Rücksendung der Fragebogen und Ausfüllqualität
- **Aktualisierung von Haushaltsdaten**
 - laufende Aktualisierung der Informationen über den Haushalt (Änderungen der Haushaltsgröße durch Geburt oder Auszug, neuer Name durch Heirat, neue Adresse durch Umzug u.a.)
 - regelmäßiges Update von Verwendungsinformationen

Ein Ausschluss von Haushalten erfolgt entweder durch die Haushalte selbst oder bei langfristig unkorrektem Antwortverhalten.

Abbildung 21



Ausschluß von Haushalten

- **Freiwilliger Ausschluß**
 - wenn der Haushalt nicht mehr teilnehmen möchte
- **Ausschluß aus dem Panel, weil**
 - der Haushalt nicht antwortet,
 - Fragebogen nur selten zurückschickt,
 - Fragebogen schlecht ausfüllt

8 Panel-Usage-Policy

Um die Haushalte bzw. die Probanden nicht zu stark zu belasten und auch um eine Konditionierung zu vermeiden, haben wir intern weltweite Richtlinien aufgestellt. Es gibt automatische Ausschlusskriterien für die Ziehung von Stichproben, die in unserer Sampling-Software integriert ist.

Abbildung 22

 **Richtlinien für die Selektion von Stichproben**

Die Access-Panels von Ipsos haben automatische Ausschlußkriterien für die Ziehung von Stichproben in ihre Panel-Sampling-Software integriert. Hier ein Auszug:

- 3 Wochen zwischen jeder Studie
- 3 Monate zwischen Usage & Attitude-Studien in derselben Kategorie
- 4 Monate zwischen Konzepttests in derselben Kategorie oder bei Produkttests in derselben Kategorie
- 7 Monate zwischen identifizierten Produkttests in derselben Kategorie

Damit soll vermieden werden, dass ein Haushalt zu häufig insgesamt oder für verschiedene Studientypen oder Kategorien eingesetzt wird.

Die Frage nach der optimalen Panelgröße kann nicht einfach mit einer Zahl beantwortet werden, sondern ist abhängig von dem Umsatz/der Anzahl der Interviews ebenso wie der Anzahl der Kategorien, die befragt werden. Wenn wir den gleichen Umsatz mit nur einem Kunden und nur einer Produktkategorie hätten, bräuchte ich ein erheblich größeres Panel, sofern ich unsere Panel Usage Policy als verbindlich ansehe, da die Haushalte permanent gesperrt werden müssten. Für ein Panel ist es also sehr wertvoll, verschiedene Kunden und verschiedene Produktkategorien zu haben, um die Haushalte längerfristig zu halten.

Für unsere Umsatzgröße ist eine Panelgröße von ca. 35 000 Haushalten mit nahezu 85 000 Probanden eine angemessene Größe mit vernünftiger Auslastung und interessantem Produktmix. Ein kleineres Panel würde häufig nicht ausreichen, da z. T. mehrere Konzepte oder Produkte parallel getestet werden und wir bei schwierigen Quoten nicht genügend Haushalte vorhalten könnten.

Abbildung 23

Ipsos **Da capo**

- Die Größe eines Access-Panels muß mit deren Auslastung verknüpft werden:
 - nicht zu wenige Studien, um ein Abbrechen von teuer rekrutierten Haushalten zu vermeiden
 - nicht zu viele Studien, um eine Überlastung zu vermeiden

- Mit der Auftragslage wächst das Panel ...

- ... und das beständige Spiel von Rekrutierung, Pflege und Ausschluß beginnt von neuem.

9 Zusammenfassung: Stärken und Grenzen von Access-Panels

Abbildung 24

Ipsos **Stärken und Grenzen von Access-Panels**

- **Stärken**
 - antwortbereite Haushalte garantieren hohe Rücklaufquoten
 - ad hoc verfügbare Zielgruppen ersparen Zeit und Geld
 - Ermittlung auch von Low-Incidence-Zielgruppen
 - anonyme Befragung – besonders wichtig für sensible Themen
 - kosten-effiziente Methode
 - durch postalische Platzierung
 - kein Interviewereinsatz/-Bias
 - das Testen von Produkten erfolgt
 - in gewohnter Umgebung
 - in ausreichendem Zeitraum
 - in ausreichender Menge
 - regionale Streuung zu geringen Kosten
 - Erreichbarkeit auch von mobilen Haushalten/Personen

Die größten Stärken des Panels gegenüber Face-to-Face-Befragungen liegen in der Zeit- und Kostenersparnis, bedingt durch hohe Rücklaufquoten und die postalische Platzierung ohne Interviewereinsatz. Die Rücklaufquoten liegen bei 60 bis 65 % für normale Konzepttests, bei Produkttests kann man mit 80 bis 85 % Rücklauf rechnen, bei sehr beliebten Produkten, z. B. Babywindeln, sogar mit 95 bis 100 %. Das heißt auch, dass neben dem realen Kaufwert eines Testproduktes auch die emotionale Nähe zum Produkt den Rücklauf mitbeeinflusst.

Erhebliche Kosteneinsparungen ergeben sich natürlich auch dadurch, dass wir die wichtigsten Demographie- und Verwendungsdaten gespeichert haben, also nicht jedes Mal neu erheben müssen, und diese innerhalb kürzester Zeit aus der Datenbank abgefragt werden können. Selbst wenn Kriterien nicht gespeichert sind, ist eine präselektierte Postkartenaktion schneller und günstiger als ein Kontakt über den Interviewer. Dieser Kostenvorteil gilt insbesondere bei Low-Incidence-Zielgruppen.

Dadurch, dass wir keine Interviewer einsetzen müssen, entfallen natürlich auch die mit der Befragung und Rekrutierung zusammenhängenden Probleme.

Das Ausfüllen des Fragebogens erfolgt in gewohnter Umgebung zu einer Zeit, die der Versuchsperson passt. Dadurch sind auch die mobilen Zielpersonen besser zu erreichen und es entsteht kein Zeitdruck.

Besondere Vorteile weisen Access-Panels bei sehr sensiblen Themen auf, bei denen ein Interviewereinsatz nur selten in Frage kommt bzw. die Antwortbereitschaft sehr oft eingeschränkt ist. Die Probanden erweisen sich auch bei intimsten Fragen als sehr offen. Zudem ist die Antwortqualität ebenso wie die „Ehrlichkeit“ im Panel häufig besser als bei Face-to-Face- oder Telefonumfragen.

Abbildung 25



Stärken und Grenzen von Access-Panels

➤ **Grenzen**

- Ermittlung von spontanen Urteilen (Gefallen einer Werbung)
- Fragen zur „Wissens“-Überprüfung (z.B. unaided recall)
- Verständlichkeit und Eindeutigkeit der Fragen und Items (wird durch erfahrene Projektleiter sichergestellt)
- Reihenfolge der Fragen (Panelteilnehmer können Fragebogen zunächst einmal ganz durchlesen)
- Begrenzung des Fragebogenumfangs (es wurden Studien mit 70-seitigem Fragebogen erfolgreich durchgeführt)
- begrenzte Kontrolle, wer den Fragebogen im Haushalt ausfüllt (Identitätsproblem)

Die Grenzen des Access-Panels ergeben sich am meisten durch die Tatsache, dass der Proband nicht dabei überprüft werden kann, wie er den Fragebogen ausfüllt – ob er ihn erst ganz durchliest und dann von vorne anfängt oder ob ihm andere beim Ausfüllen assistieren. Zu diesen Themen durchgeführte interne Untersuchungen zeigen jedoch, dass der Fragebogen überwiegend von der Zielperson allein und auch in der Reihenfolge der angegebenen Fragen beantwortet wird.

Dennoch sind spontane Urteile oder z. B. die Abfrage von ungestütztem Recall in der Regel nicht möglich. Es bedarf beim Einsatz von Access-Panels mehr denn je des Einsatzes von erfahrenen Mitarbeitern, da die Fragen und Items eindeutig und verständlich formuliert sein müssen, eine Hilfestellung durch die Interviewer ist nicht möglich.

Theoretisch ist eine Begrenzung der Untersuchungslänge angezeigt, da alle Befragungen über 45 Minuten als eher problematisch in Bezug auf die Rücklaufquote anzusehen sind. Praktisch zeigt sich aber, daß auch 70-seitige Fragebögen oder Trackings über 2 Jahre erfolgreich durchgeführt wurden. Maßgeblich für den Erfolg ist hier die persönlich gehaltene Ansprache der Haushalte, eine ausreichende Kommunikation, ggf. auch eine telefonische Nachfassaktion, und natürlich auch die Attraktivität des Incentives.

Zusammenfassend ist festzustellen, daß Access-Panels eine sehr zukunftsweisende Erhebungsmethode darstellen – einige Kunden haben ihre Befragungen nahezu komplett auf das Access-Panel ausgerichtet und agieren äußerst erfolgreich am Markt.

Perspektiven der Online-Forschung

Weltweit werden im Juni 1999 etwa 180 Millionen Nutzer und Nutzerinnen des Internet geschätzt (Nua Internet Surveys, <http://www.nua.ie>). In Deutschland nutzen laut der dritten Welle des GfK-Online-Monitors (<http://www.gfk.de>), deren Ergebnisse im Februar 1999 veröffentlicht wurden, bereits jetzt über acht Millionen Bundesbürger im Alter von 14 bis 59 Jahren zumindest gelegentlich einen Online-Zugang (z. B. vom Arbeitsplatz, der Schule oder von zu Hause aus). Diese Zahl soll alleine in den letzten sechs Monaten des vergangenen Jahres um 40 % angestiegen sein.

Soviel zu den aktuellen Zahlen. Sie belegen, dass wir es beim Internet nicht mehr mit einem Randphänomen zu tun haben. Auch wenn gegenwärtig der weit überwiegende Teil der Bevölkerung das Internet nur vom Hörensagen kennt und über keinen Zugang zum Netz verfügt, stellt sich die Frage, ob es Sinn macht, schon jetzt die Möglichkeiten des Netzes auch zur sozialwissenschaftlichen Datenerhebung zu nutzen – also z. B. über das Internet Umfragen durchzuführen.

Die Vorteile eines solchen Unterfangens liegen auf der Hand, da bei Befragungen über das Internet im Vergleich zu postalischen Umfragen der finanzielle Aufwand geringer ist und sich innerhalb kurzer Zeiträume Befragungen mit extrem hohen Fallzahlen realisieren lassen. Diesen Vorteilen stehen jedoch gravierende methodische Probleme gegenüber, die es auch mittelfristig unwahrscheinlich erscheinen lassen, über das Internet z. B. allgemeine Bevölkerungsumfragen durchführen zu lassen. Zu nennen sind hier vor allem die Schwierigkeiten bei der Bestimmung der Grundgesamtheit, so dass (zumindest gegenwärtig) bei derartigen Befragungen eine Stichprobenauswahl nach einem wohldefinierten Design nicht möglich ist (vgl. Bandilla 1999). Trotz dieser speziell die allgemeinen Bevölkerungsumfragen betreffenden Einschränkungen zeigen sich schon heute Anwendungsbereiche, in denen die Vorteile des Netzes auch unter Wahrung der gängigen methodischen Standards genutzt werden können. Zu den bereits genutzten Anwendungsbereichen gehören z. B. Mitarbeiterbefragungen in global verteilten und vernetzten Unternehmen, die Befragung von Spezialpopulationen bzw. Experten sowie Befragungen registrierter Internetnutzer für bestimmte Inhalte, also in Bereichen, in denen die jeweiligen Grundgesamtheiten bestimmbar sind. Ein bereits in Grundzügen erkennbarer Anwendungsbereich liegt außerdem noch in der Durchführung von Pretests für allgemeine Bevölkerungsumfragen.

Im folgenden werden wir in einem kurzen Abriss auf die Entwicklung des Internet eingehen, in diesem Zusammenhang verschiedene Möglichkeiten zur Datenerhebung über das Netz vorstellen und daran anschließend einige technische Perspektiven thematisieren. Abschließen möchten wir diesen Beitrag mit einem kurzen Überblick zu den sich abzeichnenden Forschungsperspektiven im Bereich der sozialwissenschaftlichen „Online-Methodenforschung“.

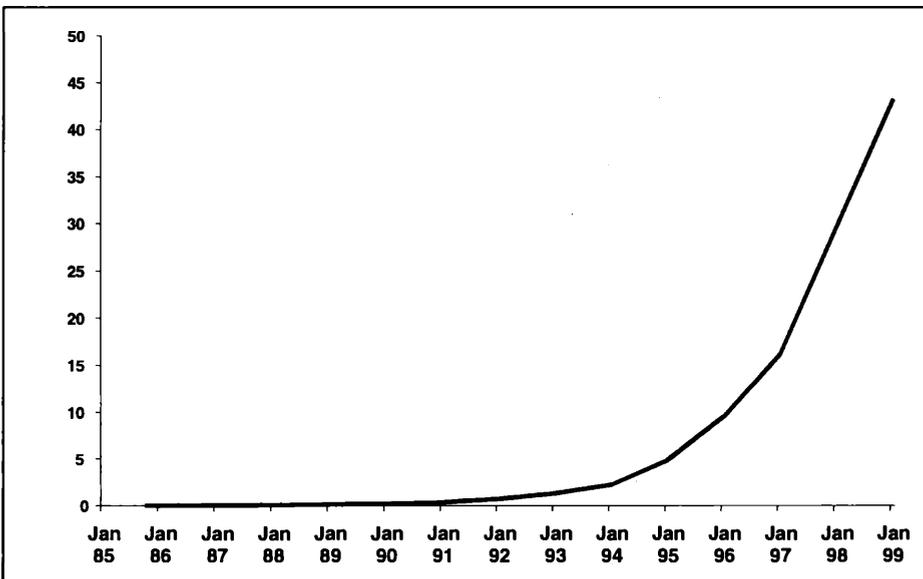
*) Dr. Wolfgang Bandilla, Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim.
Michael Bosnjak, Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim.

Mit Betonung möchten wir jedoch an dieser Stelle darauf hinweisen, dass sich die Internet-Forschung – verglichen mit dem Forschungsstand bei anderen computergestützten Datenerhebungsverfahren, wie z. B. CATI oder CAPI – noch im Anfangsstadium befindet. Diese Feststellung betrifft u.a. sowohl den akademischen als auch den kommerziellen Bereich.

Entwicklung des Internet

Obwohl das Internet schon seit Ende der sechziger Jahre existiert und seinen Ursprung im amerikanischen ARPANET hat (vgl. Musch 1997), spielt es als Kommunikations- und Informationsmedium für breitere Bevölkerungskreise erst seit einigen Jahren eine ernstzunehmende Rolle. Erst etwa 1993/1994 tritt das Internet aus seinem Schattendasein und beginnt aufgrund seiner vielfältigen Möglichkeiten auch für den kommerziellen Bereich interessant zu werden (vgl. Abbildung 1):

Abbildung 1
Entwicklung der Anzahl der an das Internet angeschlossenen Rechner in Millionen



Quelle: Network Wizards

So stieg die Zahl der Internetrechner weltweit zwischen Anfang 1993 von etwas über einer Million auf über 43 Millionen Anfang 1999. Verantwortlich für diese rasante Entwicklung sind u. a. staatliche Förderprogramme zum Ausbau des Netzes und parallel hierzu die starke Verbreitung von Arbeitsplatzrechnern mit zunehmend leichter zu bedienender Software. Gleichzeitig werden verstärkt Versuche unternommen, das Internet zum Zweck der Datenerhebung einzusetzen. Neben einer Vielzahl von Datenerhebungsverfahren (vgl. Batinic/Bosnjak 1997) handelt es sich insbesondere um Befragungen per E-Mail und Befragungen im World Wide Web (WWW-Befragungen).

Völlig neu ist die Idee nicht, über ein Netz Befragungen ohne den Einsatz von Interviewern durchzuführen: Schon Mitte der achtziger Jahre, also zu einer Zeit, in der das Internet nur eine randständige Rolle spielte, wurde z. B. in den Niederlanden das sogenannte „Telepanel“ realisiert. Hierbei wurde eine repräsentative Stichprobe von Haushalten regelmäßig mittels Videotext über Bildschirme befragt. Die TV-Monitore waren über Modem und Telefonleitungen mit einem Zentralrechner verbunden, so dass die technischen Voraussetzungen für interviewfreie Befragungen gegeben waren (vgl. Saris 1998). Den teilnehmenden Haushalten wurde das technische Equipment kostenfrei zur Verfügung gestellt, um auch Haushalte ohne entsprechende Ausstattungsmerkmale für das Panel zu gewinnen und somit eine bevölkerungsrepräsentative Stichprobe zu gewährleisten. Ähnliches – allerdings unter Ausnutzung der Möglichkeiten des Internet – wird gegenwärtig in den USA beim Projekt „Free-PC“ (<http://www.Free-PC.com>) versucht, in dessen Endstadium eine repräsentative Stichprobe von 30 000 Haushalten über kostenlos zur Verfügung gestellte Internetrechner regelmäßig befragt werden soll.

Befragungen per E-Mail

Die einfachste Variante einer Befragung mit Hilfe des Internet ist die Befragung per E-Mail. Hierbei können die besonderen Vorteile von E-Mail („elektronische Post“), wie Schnelligkeit, Asynchronität und Ökonomie (vgl. Bosnjak 1997) genutzt werden. Quasi analog zu einer postalischen Befragung wird an ausgewählte E-Mail-Adressen ein Fragebogen gesandt, der vom Empfänger ausgefüllt und auch per E-Mail zurückgesandt werden soll. Eine Modifikation dieser Vorgehensweise besteht darin, zuerst eine kurze Mail zu versenden, in der die Befragung angekündigt und um Teilnahme gebeten wird. Nur auf ausdrückliche Anforderung wird daraufhin der Fragebogen versandt.

E-Mail-Befragungen wurden schon in den achtziger Jahren durchgeführt (z. B. Kiesler/Sproull 1986) und sind methodisch unproblematisch, sofern auf der Grundlage von Designs für Zufallsstichproben gearbeitet werden kann. Diese Voraussetzung ist jedoch nur in Ausnahmefällen gegeben, wenn z. B. eindeutige Listen von gültigen E-Mail-Adressen zur Stichprobenziehung vorhanden sind. Im Regelfall dürfte es jedoch so sein, dass derartige Listen nicht verfügbar sind. Als weiteres Problem ist zu beachten, dass die Zielpersonen zunehmend nicht nur über eine, sondern über mehrere E-Mail-Adressen verfügen.

Im Vergleich zu den im Folgenden beschriebenen WWW-Befragungen haben E-Mail-Fragebögen auch den Nachteil, dass sie auf der Basis des ASCII-Zeichensatzes erstellt werden und der Text keine Formatierungsanweisungen wie z. B. Fettdruck enthalten darf. Von daher sind den Gestaltungsmöglichkeiten dieser Fragebögen enge Grenzen gesetzt und sie entsprechen nicht dem, was heutzutage als Standard bei schriftlichen postalischen Fragebögen gilt.

WWW-Befragungen

Die Konstruktion von WWW-Fragebögen erfolgt mittels der Seitenbeschreibungssprache HTML. Hierdurch bieten sich weitergehende Möglichkeiten als bei den schon beschriebenen E-Mail-Befragungen, wie z. B. Formulare mit selektierbaren Antwortvorgaben, Java- oder Javascript-Programme (vgl. Schuster 1997) und die Unterstützung durch Bild oder Ton. Diese Techniken machen es möglich, die auf dem Bildschirm präsentierten Fragebögen so zu

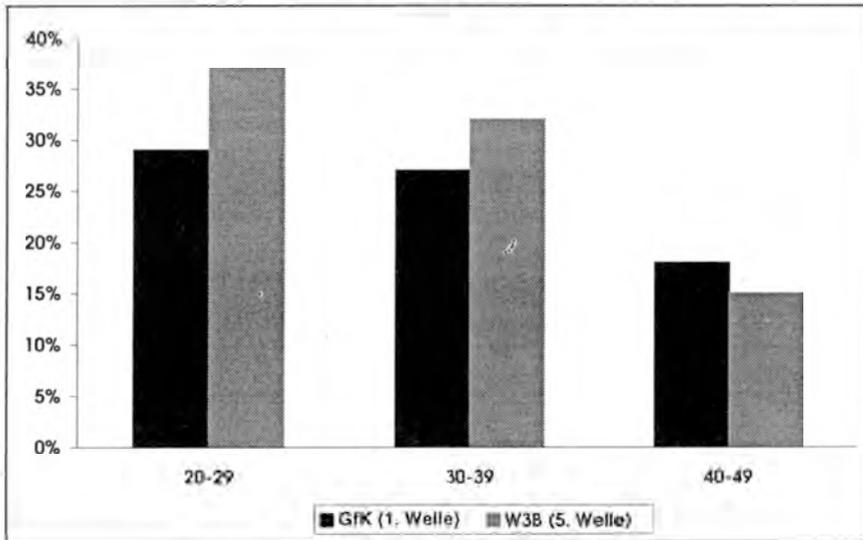
gestalten, dass sie für den Befragten besser und einfacher zu handhaben sind als ein gedruckter Fragebogen (z. B. durch automatische Filterführung, Ausblendung irrelevanter Fragen etc.). Neben den eigentlichen Inhaltsdaten fallen bei entsprechend eingerichteten WWW-Befragungen jedoch auch Daten über den Antwortprozess an, wie z. B. die Reihenfolge und die Dauer der Fragenbeantwortung. Diese Meta- bzw. Para-Daten können insbesondere im Rahmen von Pretests wertvolle Informationen über die Qualität des jeweiligen Fragenprogramms geben.

Einen hohen Bekanntheitsgrad erreicht haben die von der Georgia Tech Research Corporation zwischen Anfang 1994 und Ende 1998 durchgeführten Gvu-Surveys (http://www.cc.gatech.edu/gvu/user_surveys), die für ihre hohen Teilnehmerzahlen bekannt waren: Sie lagen in der Regel zwischen 10 000 und 20 000, teilweise sogar bei weit über 20 000 Teilnehmern. Die Ergebnisse dieser Umfragen fanden vor allem in den populären Computermagazinen Beachtung, wenn es darum ging, Veränderungen in der demographischen Struktur der das Internet nutzenden Personen aufzuzeigen.

Ein Merkmal dieser Umfragen war, dass an ihnen ohne weitere Einschränkungen jede Person teilnehmen konnte, die über einen Zugang zum Internet und die übliche Browser Software verfügte. Die Umfragen fanden alle 6 Monate zu festen Terminen statt, und zwar jeweils im April/Mai sowie Oktober/November eines jeden Jahres. Die Feldzeiten betragen 4 Wochen. Die Bekanntmachung der Umfragen erfolgte im wesentlichen über Werbebanner bei den großen Suchmaschinen und anderen häufig aufgesuchten Seiten. Ähnlich konzipiert ist im deutschsprachigen Raum die von Fittkau und Maaß seit 1995 durchgeführte W3B-Umfrage (<http://www.w3b.de>).

In der Rekrutierung der Teilnehmer liegt jedoch das methodische Hauptproblem dieser Umfragen, da keine aktive Stichprobenziehung erfolgt (Problem der Selbstselektion; – vgl. Bandilla 1999). Schon allein die Bekanntmachung der Umfragen über Werbebanner auf den großen Suchmaschinen ist äußerst problematisch, da sich weniger als 1 % derjenigen, die auf diesem Weg zur Teilnahme an einer Web-Untersuchung aufgefordert werden, letztendlich auch für eine Teilnahme entscheiden (vgl. Bandilla et al. 1999). Vergleicht man z. B. die Altersverteilung der Teilnehmer und Teilnehmerinnen der selbstselektierten W3B-Umfrage mit den Ergebnissen des in einem vergleichbaren Zeitraum durchgeführten GfK-Online-Monitors, also einer repräsentativen CATI-Befragung, so zeigen sich bei der W3B-Umfrage gravierende Abweichungen von der tatsächlichen Altersstruktur der das Internet nutzenden Personen (vgl. Abbildung 2, S. 170).

Abbildung 2
W3B- vs. GfK-Befragung – Alter der Befragten



Diese methodisch bedingten Ungenauigkeiten in den Ergebnissen sind auch der Grund dafür, dass die von GVU veranstalteten Umfragen in ihrer bisherigen Form zwischenzeitlich eingestellt wurden. Gerade in der Marktforschung und der zunehmenden Bedeutung des „E-Commerce“ sind in der gegenwärtigen Phase der geradezu explosionsartigen Ausbreitung des Internet genauere Daten über Struktur, Einstellungen und Verhalten der Internet-Nutzerschaft erforderlich. Diesbezüglich bieten Online-Panels oder auch Access-Panels (vgl. hierzu auch den Beitrag von Michael Hoppe in diesem Band, S. 145 ff.) eine Alternative zu den reinen WWW-Befragungen.

Online-Panels

Die einfache, preiswerte und auch immer häufiger praktizierte Variante ist ein Online-Panel, bei dem die Panel-Teilnehmer ausschließlich über das Netz rekrutiert werden. Das Grundprinzip ist das folgende: Die Teilnehmer gelangen per Zufall oder wie auch immer auf die Seiten eines Instituts, das für die Teilnahme am Online-Panel wirbt. Interessierte Personen füllen ein Formular aus, das in der Regel einige Fragen zur Demographie enthält und werden danach mehr oder weniger häufig aufgefordert, an bestimmten Befragungen teilzunehmen. Das Ganze ist häufig mit Incentives verbunden: Bonuspunkte, Verlosung von Sachpreisen etc. Im Vergleich zu einer reinen WWW-Befragung ist ein solches Online-Panel schon höher einzuschätzen: Z. B. ist eine Rücklaufkontrolle möglich, Teilnehmer mit bestimmten demographischen Merkmalskombinationen können gezielt befragt werden usw. Das grundsätzliche Problem der Selbstselektion ist bei dieser Vorgehensweise natürlich nicht aus der Welt, sondern nur verschoben, da auf der ersten Stufe der Rekrutierung keine aktive Stichprobenziehung, sondern lediglich eine Quotenauswahl erfolgen kann.

Dieses grundsätzliche Problem ist nur mit erheblich höherem Erhebungsaufwand zu lösen und wird z. B. von Nielsen/Netratings (<http://www.netratings.com/>) in den USA praktiziert. Die Rekrutierung der Teilnehmer erfolgt telefonisch über eine Zufallsstichprobe. Sofern bei diesen Anrufen Haushalte mit PC und Internetzugang angetroffen werden, wird versucht, diese als Panelteilnehmer zu gewinnen. Haushalte, die dieses Kriterium nicht erfüllen, werden im Abstand von sechs Monaten wiederholt befragt. Sollten diese Haushalte in der Zwischenzeit einen Internetzugang erhalten haben, wird versucht, auch sie in das Panel aufzunehmen. Angesichts der sich in quantitativer und qualitativer Hinsicht verändernden Nutzerstruktur im Internet wird hier also versucht, mit dem Panel ein jeweils aktuelles Abbild der Grundgesamtheit der Internet-Nutzer in den USA zu erhalten. Der Aufwand bei der Panelrekrutierung und Panelpflege ist jedoch enorm und nicht zu unterschätzen. Ausführlich werden Design und Vorgehensweise dieses Panels auf den Internetseiten von Nielsen beschrieben (<http://www.nielsen-netratings.com>).

Technische Perspektiven

Mit dem steigenden Interesse an Online-Forschung sowohl von kommerzieller als auch wissenschaftlicher Seite wächst auch die Nachfrage nach Werkzeugen zur Implementierung und Durchführung von Online-Umfragen. Während in der „Anfangszeit“ der Online-Forschung gute Kenntnisse der Seitenbeschreibungssprache HTML sowie zumindest einer serverseitigen Skriptsprache (z. B. zur Datenübermittlung und -speicherung) unerlässlich waren, wird inzwischen eine Reihe von Entwicklungswerkzeugen für Online-Umfragen angeboten. Diese Entwicklungswerkzeuge lassen sich grob in folgende drei Produktklassen einordnen:

1. *Statische Web-Survey-Generatoren*

In diese Klasse fallen Werkzeuge, mit Hilfe derer ein statisches Fragenprogramm für Web-Umfragen aufbereitet bzw. in HTML konvertiert werden kann. Idealerweise ermöglichen es diese Werkzeuge außerdem noch, den Prozess der Datensammlung und Speicherung ohne detaillierte Programmierkenntnisse vorzubereiten. Außer der Verwaltung von wiederverwendbaren Fragebogenkomponenten und der graphisch mehr oder weniger ansprechenden Gestaltung einer Fragebogenseite sind aufwendigere Befragungsprojekte, die z. B. Filterführungen und Ergebnismeldungen vorsehen, mit diesen Werkzeugen kaum möglich.

2. *Adaptive Web-Survey-Generatoren*

Zu dieser Kategorie gehören Werkzeuge, mit denen neben der zuvor genannten Grundfunktionalität auch komplizierte Filterführungen bzw. adaptive Befragungsverläufe umsetzbar sind.

3. *„Multi-Mode-Tools“*

In diese Klasse fallen Werkzeuge, mit Hilfe derer unter einer softwaretechnischen Oberfläche ein Fragenprogramm entwickelt wird, welches dann in unterschiedlichen Befragungsmodi appliziert werden kann, wie z. B. als Web-, E-Mail-, CATI-, CAPI- und Disk-by-Mail-Befragung. Diese Werkzeuggruppe gehört sicher zu den vielversprechendsten, allerdings existieren in der Praxis momentan noch keine einsatzfähigen Produkte.

Angesichts der sehr schnellen Entwicklung im Bereich der Softwareentwicklung soll in diesem Rahmen auf die Vorstellung bzw. Nennung einzelner Produkte verzichtet werden. Auf

der Homepage unserer Arbeitsgruppe *OnlineResearch* am Zentrum für Umfragen, Methoden und Analysen (ZUMA) halten wir dafür eine kommentierte und laufend aktualisierte Liste verfügbarer Werkzeuge zur Implementierung und Durchführung von Online-Umfragen bereit (<http://www.or.zuma-mannheim.de>).

Forschungsperspektiven

Im Bereich der sozialwissenschaftlichen „Online-Methoden-Forschung“ zeichnen sich zahlreiche Forschungstrends ab, die sich gegenwärtig auf vier Bereiche konzentrieren. Zu nennen sind hier Forschungsansätze, die (1) internetbasierte Verfahren der Stichprobenziehung, (2) die Determinanten der Teilnahmeentscheidung an Online-Untersuchungen, (3) befragungsmedienspezifische Mode- und Kontexteffekte sowie (4) präskriptive Modelle zur Gestaltung von Online-Untersuchungen zum Gegenstand haben.

Wie eingangs bereits erwähnt, können Online-Umfragen bei bestimmten Fragestellungen nur unzureichend mit „klassischen“ Befragungsverfahren konkurrieren, da allzu oft eine zufalls-gesteuerte Stichprobenkonstruktion kaum möglich ist. Dieser Umstand leitet sich in erster Linie aus den informationstechnischen Besonderheiten des Internet ab (vgl. hierzu Bosnjak 1997) und wird sich mittelfristig – z. B. zur Durchführung methodisch anspruchsvoller, allgemeiner Bevölkerungsumfragen – kaum zufriedenstellend lösen lassen. Eine Hauptaufgabe der sozialwissenschaftlichen Online-Forschung wird somit sein, internetbasierte Verfahren zur Stichprobenziehung weiterzuentwickeln und ggf. mit traditionellen Verfahren zu kombinieren. Das bereits weiter oben erwähnte Online-Panel von Nielsen/ Netratings, bei dem Online-Nutzer per CATI-Screening zufällig ausgewählt werden, sowie erste Arbeiten von Schaefer/Dillman (1998) zur Kombination gemischter Rekrutierungsstrategien (postalisch und per E-Mail) weisen in diese Richtung. Insbesondere bei netzspezifischen Fragestellungen, wie z. B. der Bewertung des Online-Auftritts eines Unternehmens, werden zudem vermehrt Event-Sampling-Verfahren eingesetzt, bei denen jeder „n-te Besucher“ (vgl. das Verfahren von Infratest Burke; Pfleiderer 1997) eines Informationsangebots um die Teilnahme an einer Online-Umfrage gebeten wird. Aus diesem Zufallsverfahren, bei dem über „Nutzungsereignisse“ und nicht über „Personen“ gesampelt wird, ergeben sich eine Fülle von grundlagenwissenschaftlichen Fragestellungen.

Eine dieser Fragestellungen – und hiermit sind wir bereits bei der zweiten Forschungsperspektive angelangt – bezieht sich auf die Determinanten der Teilnahmeentscheidung an solchen (auf einer ersten Auswahlstufe) „nicht selbstselektierten“ Online-Befragungen. Erste Untersuchungen zur Teilnahmeabsicht an E-Mail-Befragungen (z. B. von Bosnjak/Batinic 1999) sowie zu möglichen Faktoren zur Vorhersage und Erklärung der Teilnahmeentscheidung an Web-Befragungen (z. B. von Bosnjak/Bandilla/Tuten 1998) liegen bereits vor, bedürfen jedoch einer weiteren Validierung „im Feld“. Sollte es gelingen, diese Teilnahmeentscheidungen zufriedenstellend erklären und damit auch vorhersagen zu können, kann dies nicht nur von grundlagenwissenschaftlichem, sondern auch von angewandtem Interesse sein, da dann gezielte Interventionsstrategien zur Beeinflussung dieser Entscheidung – und damit der Ausschöpfungsrate – denkbar sind.

Ein weiteres, sehr breites Forschungsfeld sehen wir im Bereich der verfahrenvergleichenden Forschung (bzw. im Bereich der sog. Mode-Effects-Forschung). Bereits die Arbeiten von Kiesler/Sproull (1986) deuten darauf hin, dass in elektronischen Umgebungen „andere“

kommunikative Verhaltensweisen vorzufinden sind als in klassischen Befragungssituationen (wie z. B. per Telefon oder persönlich-mündlich). Für den Bereich der Online-Forschung stellt sich die Frage, inwieweit diese zeitlich-räumlich veränderte Befragungssituation einen Einfluss auf die Datenqualität hat. Theoretisch verwandt mit der Mode-Effects-Forschung ist der Bereich der Kontext- bzw. Fragenreihenfolgeeffekte. Auch diese Richtung könnte sich angesichts der erweiterten Möglichkeiten zur kontrollierten Applizierung von Befragungsstimuli bei Web-Untersuchungen in Zukunft weiterentwickeln.

Zuletzt wollen wir auf den Forschungsbedarf hinweisen, der sich aus den vielfältigen Gestaltungsmöglichkeiten insbesondere bei WWW-Befragungen ergibt. Während für den Bereich der postalischen und telefonischen Befragungen eine Vielzahl an theoretisch fundierten und empirisch geprüften „Guidelines“ zur Gestaltung und Administration von „klassischen“ Befragungssituationen vorliegen – zu nennen sind hier beispielsweise die Arbeiten von Dillman (1978 und 1991) – kann bei Online-Befragungen kaum auf eine vergleichbar große, empirisch fundierte Wissensressource zurückgegriffen werden. Doch auch hier zeichnen sich erste Entwicklungen ab, wie z. B. die Arbeiten von Gräf/Heidingsfelder (1999) sowie von Dillman, der seine diesbezüglichen Publikationen neuerdings auch online zur Verfügung stellt (siehe unter: <http://survey.sesrc.wsu.edu/dillman/>), belegen.

Insgesamt zeigt sich, dass sich der Bereich der Online-Forschung relativ schnell entwickelt und ausdifferenziert hat, wohl nicht zuletzt deshalb, weil die Vorteile des Einsatzes von Online-Techniken für sozialwissenschaftliche Untersuchungen von kommerzieller und grundlagenwissenschaftlicher Seite relativ schnell erkannt und darauf mit entsprechenden Forschungsaktivitäten reagiert wurde. Diese mittlerweile rege Forschungsaktivität zeigt sich nicht zuletzt darin, dass Themen aus dem Bereich der Online-Forschung bei nahezu allen internationalen sozialwissenschaftlichen Fachtagungen Eingang gefunden haben.

Literaturhinweise

Bandilla, W. (1999): WWW-Umfragen. Eine alternative Datenerhebungstechnik für die empirische Sozialforschung? in: B. Batinic; A. Werner; L. Gräf; W. Bandilla (Hrsg.): Online-Research. Methoden, Anwendungen und Ergebnisse, Göttingen, S. 6 – 17.

Bandilla, W.; Bosnjak, M.; Schneid, M.; Stiegler, A. (1999): Interaktive Medien als Gegenstand und Instrument der empirischen Sozialforschung, in: M. Berghaus (Hrsg.): Interaktive Medien, Opladen, S. 129 – 149.

Batinic, B.; Bosnjak, M. (1997): Psychologische Fragebogenuntersuchungen im Internet, in: B. Batinic (Hrsg.): Internet für Psychologen, Göttingen, S. 221 – 243.

Bosnjak, M. (1997): Internetbasierte, computervermittelte psychologische Fragebogenuntersuchungen, Mainz.

Bosnjak, M.; Bandilla, W.; Tuten, T. L. (1998): Anwendung des Elaboration-Likelihood-Modells auf per Banner beworbene Web-Untersuchungen [Tagungsbeitrag. German Online Research 1998]. Elektronische Publikation, URL am 30.7.99: http://infosoc.uni-koeln.de/gor98/abstracts/showabst.htm/a17_bosnjak.

Bosnjak, M.; Batinic, B. (1999): Determinanten der Teilnahmebereitschaft an internet-basierten Fragebogenuntersuchungen am Beispiel E-Mail, in: B. Batinic; A. Werner; L. Gräf; W. Bandilla (Hrsg.): *Online Research. Methoden, Anwendungen und Ergebnisse*, Göttingen, S. 145 – 158.

Dillman, D. A. (1978): *Mail and Telephone Surveys: The Total Design Method*, New York.

Dillman, D. A. (1991): The Design and Administration of Mail Surveys. *Annual Review of Sociology*, 17, S. 25 – 249.

Gräf, L.; Heidingsfelder, M. (1999): Bessere Datenqualität bei WWW-Umfragen. Erfahrungen aus einem Methodenexperiment mit dem Internet-Rogator, in: B. Batinic; A. Werner; L. Gräf; W. Bandilla (Hrsg.): *Online Research. Methoden, Anwendungen und Ergebnisse*, Göttingen, S. 113 – 126.

Kiesler, S.; Sproull, L. S. (1986): Response Effects in the Electronic Survey. *Public Opinion Quarterly*, 50, S. 402 – 413.

Musch, J. (1997): Die Geschichte des Netzes: ein historischer Abriss, in: B. Batinic (Hrsg.): *Internet für Psychologen*, Göttingen, S. 27 – 48.

Pfleiderer, R. (1997): Repräsentative Daten für Websites: der n-te Besucher. [Abstract/ Tagungsbeitrag-German Online Research 1997]. Elektronische Publikation, URL am 30.7.99: http://infosoc.uni-koeln.de/girlws/abstracts/fr_05.html.

Saris, W. E. (1998): Ten Years of Interviewing without Interviewers: The Telepanel; in: M. P. Couper; R. P. Baker; J. Bethlehem; C. Z. F. Clark; J. Martin; W. L. Nicholls II; J. M. O'Reilly (Eds.): *Computer Assisted Survey Information Collection*, New York, S. 409 – 430.

Schaefer, D. R.; Dillman, D. A. (1998): Development of A Standard E-Mail-Methodology. *Public Opinion Quarterly*, 62, S. 378 – 397.

Schuster, F. (1997): Programming Online-Questionnaires with JavaScript and Java, in: W. Bandilla; F. Faulbaum (Hrsg.): *SoftStat '97 – Advances in Statistical Software*, 6, Stuttgart, S. 159 – 166.

Teilnehmerverzeichnis

A

Abele, Franz; *Statistisches Amt der Stadt Stuttgart*, Stuttgart

Appel, Dr. Sibylle; *forsa – Gesellschaft für Sozialforschung und statistische Analysen mbH*,
Berlin

Arenz, Rainer; *BIK UMFRAGEFORSCHUNG GMBH*, Hamburg

B

Bajaja, Dr. Vladislav; *Statistisches Bundesamt*, Wiesbaden

Bandilla, Dr. Wolfgang; *Zentrum für Umfragen, Methoden und Analysen (ZUMA)*,
Mannheim

Bathelt, Anke; *Westdeutscher Rundfunk*, Köln

Bechtold, Dr. Sabine; *Statistisches Bundesamt*, Wiesbaden

Beckmann, Petra; *Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt
für Arbeit*, Nürnberg

Behrens, Kurt; *BIK ASCHPURWIS + BEHRENS*, Hamburg

Bien, Dr. Walter; *Deutsches Jugendinstitut e. V.*, München

Bihler, Wolf; *Statistisches Bundesamt*, Wiesbaden

Bosnjak, Michael; *Zentrum für Umfragen, Methoden und Analysen (ZUMA)*, Mannheim

Brugger, Pia; *Statistisches Bundesamt*, Wiesbaden

C

Chlumsky, Jürgen; *Statistisches Bundesamt*, Wiesbaden

D

Dankert, Martin; *INRA Deutschland GmbH*, Mölln

Drews, Hans-Peter; *INRA Deutschland GmbH*, Mölln

Dürr, Dr. Karin; *Deutscher Fachverlag*, Frankfurt/Main

E

Ehling, Dr. Manfred; *Statistisches Bundesamt*, Wiesbaden

Erbslöh, Barbara; *Stadt Essen – Amt für Entwicklungsplanung, Statistik, Stadtforschung und
Wahlen*, Essen

F

Fröhner, Dr. Rolf; *GfK AG*, Nürnberg

Frühwald, Markus; *GfK AG*, Nürnberg

Fuchs, Jürgen; *Bundesministerium für Familie, Senioren, Frauen und Jugend*, Bonn
Fuchs, PD Dr. Marek; *Katholische Universität Eichstätt*, Eichstätt

G

Gabler, PD Dr. Siegfried; *Zentrum für Umfragen, Methoden und Analysen (ZUMA)*,
Mannheim
Gärtner, Karla; *Bundesinstitut für Bevölkerungsforschung*, Wiesbaden
Gilberg, Dr. Rainer; *Infas Institut für angewandte Sozialwissenschaft GmbH*, Bonn
Götz, Katharina; *Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen*,
Düsseldorf
Grimm, Claus; *polis Gesellschaft für Politik- und Sozialforschung mbH*, München
Grohmann, Prof. Dr. Heinz; *Johann Wolfgang Goethe-Universität Frankfurt/Main*, Kronberg
Gruber, Stefan; *Statistisches Bundesamt*, Wiesbaden
Gruner, Erika; *Thüringer Landesamt für Statistik*, Erfurt
Grünheid, Dr. Evelyn; *Bundesinstitut für Bevölkerungsforschung*, Wiesbaden

H

Hahlen, Johann; *Statistisches Bundesamt*, Wiesbaden
Hartmann, Dr. Petra; *Universität Kiel*, Kiel
Heidenreich, Hans-Joachim; *Statistisches Bundesamt*, Wiesbaden
Helbig, Ina; *Statistisches Landesamt des Freistaates Sachsen*, Kamenz
Herz, Erika; *Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM)*,
Frankfurt/Main
Hofacker, Gabriele; *Hessisches Statistisches Landesamt*, Wiesbaden
Höfflin, Peter; *FIFAS – Freiburger Institut für angewandte Sozialwissenschaft e. V.*,
Freiburg
Hoffmeyer-Zlotnik, Dr. Jürgen H. P.; *Zentrum für Umfragen, Methoden und Analysen
(ZUMA)*, Mannheim
Hoppe, Dr. Michael; *Ipsos Deutschland GmbH*, Hamburg

I

Ignaczak, Jürgen; *ENIGMA – Institut für Markt- und Sozialforschung GmbH*, Wiesbaden

J

Jahn, Dr. Ingeborg; *Bremer Institut für Präventionsforschung und Sozialmedizin (BIPS)*,
Bremen

K

- Kieslich, Carolin; *GfK AG*, Nürnberg
- Klass, Johannes; *LINK + Partner GmbH, Institut für computergestützte Befragungen*, Frankfurt/Main
- Klimesch, Holger; *Institut für Marktforschung GmbH*, Leipzig
- Klumpe, Bettina; *ENIGMA – Institut für Markt- und Sozialforschung GmbH*, Wiesbaden
- Knobloch, Silvia; *Institut für Journalistik und Kommunikationsforschung*, Hannover
- Koltzk, Ingo; *Landesamt für Datenverarbeitung und Statistik Brandenburg*, Frankfurt/Oder
- König, Christian; *Statistisches Bundesamt*, Wiesbaden
- Kotzielski, Dr. Peter; *Sozialwissenschaftliches Institut der Bundeswehr*, Strausberg
- Kreuzholz, Werner; *Bayerisches Landesamt für Statistik und Datenverarbeitung*, München
- Krockow, Albrecht; *Statistisches Bundesamt*, Wiesbaden
- Krug, Prof. Dr. Walter; *Universität Trier*, Trier
- Kuchta, Peter; *Landesamt für Datenverarbeitung und Statistik Brandenburg*, Cottbus

L

- Landvogt, Werner; *Statistisches Bundesamt*, Wiesbaden
- Langenhoff, Georg; *sfs Landesinstitut Sozialforschungsstelle Dortmund*, Dortmund
- Linkert, Karin; *Statistisches Bundesamt*, Wiesbaden
- Litzenroth, Heinrich A.; *GfK Testmarktforschung GmbH*, Nürnberg
- Loreth, Dr. Hans; *Statistisches Landesamt Baden-Württemberg*, Stuttgart

M

- Mailänder, Josef; *Statistisches Landesamt Saarland*, Saarbrücken
- Maurer, Ulrich; *Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung*, Wiesbaden
- Meier, Dr. Gerd; *GfK AG*, Nürnberg
- Meißner, Thomas; *Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung*, Wiesbaden
- Metschke, Dr. Rainer; *Der Berliner Datenschutzbeauftragte*, Berlin
- Meulemann, Prof. Dr. Heiner; *Universität zu Köln*, Köln
- Modest-Wiegand, Angelika; *AMW Forschung und Beratung*, Mannheim
- Möhring, Wiebke; *Institut für Journalistik und Kommunikationsforschung*, Hannover
- Mueller, Prof. Dr. Dr. Ulrich; *MZG Medizinische Soziologie und Sozialmedizin*, Marburg
- Münzberg, Ronald; *Thüringer Landesamt für Statistik*, Erfurt
- Müssig-Trapp, Peter; *HIS GmbH*, Hannover

N

Nerb, Dr. Gernot; *ifo Institut für Wirtschaftsforschung e. V.*, München
Niedermann, Dr. Anne; *Institut für Demoskopie Allensbach*, Allensbach

P

Perry, Thomas; *Sinus Sociovision GmbH*, Heidelberg
Petersen, Thomas; *Institut für Demoskopie Allensbach*, Allensbach
Pfleiderer, Rolf; *Infratest Burke*, München
Porst, Rolf; *Zentrum für Umfragen, Methoden und Analysen (ZUMA)*, Mannheim
Prosteder, Hela; *Roland Berger Forschungsinstitut für Markt- und Systemforschung GmbH*,
München

Q

Quitt, Helmut; *Infratest Burke*, München

R

Rappl, Thomas; *Bayerisches Landesamt für Statistik und Datenverarbeitung*, München
Rexroth, Margrit; *Zentrum für Umfragen, Methoden und Analysen (ZUMA)*, Mannheim
Ritter, Dr. Gert; *Bundesministerium der Finanzen*, Bonn
Rösch, Günther; *Büro für Erhebungsdesign und Datenanalyse*, Frauenberg

S

Saul, Verena; *Dresdner Bank AG*, Frankfurt/Main
Schäfer, Hartmut; *Millward Brown Germany GmbH*, Frankfurt/Main
Scharrenbroch, Andrea; *GfK AG*, Nürnberg
Scheuch, Prof. Dr. Erwin; *Zentralarchiv für Empirische Sozialforschung*, Köln
Schlüer, Rüdiger; *interlogic GmbH*, Bielefeld
Schmidt, Bernd; *Statistisches Bundesamt*, Wiesbaden
Schmitt, Lars; Köln
Schneid, Michael; *Zentrum für Umfragen, Methoden und Analysen (ZUMA)*, Mannheim
Schönfeld-Nastoll, Ulrike; *Stadt Oberhausen*, Oberhausen
Schräpler, Jörg-Peter; *Ruhr-Universität Bochum*, Bochum
Schröder, Hermann-Dieter; *Hans-Bredow-Institut*, Hamburg
Schubert, Gudrun; *GfK AG*, Nürnberg
Schulte, Dr. Wolfgang; *Forschungsgruppe Wahlen e. V.*, Mannheim
Schupp, Dr. Jürgen; *Deutsches Institut für Wirtschaftsforschung*, Berlin

Schwickerath, Marco; *Statistisches Bundesamt, Wiesbaden*
Siegmond, Christiane; *Landesamt für Datenverarbeitung und Statistik Brandenburg, Potsdam*
Sodeur, Prof. Dr. Wolfgang; *Universität-Gesamthochschule Essen, Essen*
Spangenberg, Stefan; *Sozialwissenschaftliches Institut der Bundeswehr, Strausberg*
Stahl, Matthias; *Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V., Bonn*
Stich, Dr. Andreas; *Geschäftsstelle des Wissenschaftsrates, Köln*
Stiegler, Angelika; *Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim*
Stögbauer, Andrea; *Forschungsgruppe Wahlen, Mannheim*
Streufert, Volker; *Statistisches Landesamt Sachsen Anhalt, Halle/Saale*

V

Vollhüter, Sabine; *CZAIA Marktforschung GmbH, Bremen*

W

Wachter, Bernd; *psyma – arbeitsgruppe für psychologische marktanalysen GmbH, Rückersdorf*
Waldeck, Hans-Peter; *Statistisches Bundesamt, Wiesbaden*
Waldeck, Petra; *Statistisches Landesamt Sachsen Anhalt, Halle/Saale*
Wasmer, Martina; *Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim*
Wein, Elmar; *Statistisches Bundesamt, Wiesbaden*
Wenzel, Karsten; *Statistisches Landesamt Berlin, Berlin*
Wiegand, Erich; *Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM), Frankfurt/Main*
Wiegert, Dr. Rolf; *Universität Tübingen und Institut für Angewandte Wirtschaftsforschung Tübingen, Mössingen*
Wilsdorf, Dr. Steffen; *Universität Leipzig, Leipzig*

Z

Ziese, Dr. Thomas; *Robert Koch-Institut, Berlin*

Schriftenreihe Spektrum Bundesstatistik

herausgegeben vom Statistischen Bundesamt

Band 1

Internationaler Workshop 1992 zur Umgestaltung der Agrarstatistik in den Staaten Mittel- und Osteuropas

169 Seiten · DM 18,80 / EUR 9,61 · Bestellnummer: 1030501-93900 · ISBN 3-8246-0343-8

Beiträge zu einem Workshop des Instituts für Agrarsektoranalyse der Universität Bonn und des Statistischen Bundesamtes zur Umgestaltung der Agrarstatistik in den Staaten Mittel- und Osteuropas in Berlin 1992.

Band 2

Sabine Köhler

Bewertung des Waldes im Rahmen der gesamtwirtschaftlichen Vermögensrechnung

– Möglichkeiten und Grenzen –

82 Seiten · DM 13,80 / EUR 7,06 · Bestellnummer: 1030502-94900 · ISBN 3-8246-0344-6

In diesem Band wird untersucht, inwieweit eine Bewertung des Waldes (Waldboden und Baumbestände) auf der Basis vorhandener Statistiken im Rahmen der gesamtwirtschaftlichen Vermögensrechnung möglich ist und welche statistischen Lücken bestehen. Ausgehend von einer Darstellung der Methoden und Probleme der Waldbewertung in der forstwirtschaftlichen Theorie und Praxis sowie eine Analyse möglicher statistischer Quellen werden Schätzmodelle entwickelt und Größenordnungen aufgezeigt.

Band 3

Thomas Luh

Verbesserung der statistischen Erfassung der Unternehmensgewinne zur Berechnung des Bruttosozialprodukts von der Einkommenseite

91 Seiten · DM 15,80 / EUR 8,08 · Bestellnummer: 1030503-96900 · ISBN 3-8246-0350-0

In den Volkswirtschaftlichen Gesamtrechnungen der Bundesrepublik Deutschland wird das Bruttosozialprodukt über die Entstehungs- und die Verwendungsrechnung ermittelt. Eine eigenständige Berechnung von der Einkommenseite her ist nicht möglich, weil die Gewinne der Unternehmen nicht hinreichend erfaßt werden können. In diesem Band wird deshalb untersucht, welche Verbesserungen des statistischen Instrumentariums erforderlich sind, um die Unternehmensgewinne vollständig zu erfassen, so daß das Bruttosozialprodukt auch über die Einkommenseite berechnet werden kann.

Band 4

Finanzstatistische Kennzahlen für den Bildungsbereich

(In Vorbereitung.)

Band 5

Joachim Klaus u. a.

Umweltökonomische Berichterstattung

– Ziele, Problemstellungen und praktische Lösungsansätze –

244 Seiten · DM 24,80 / EUR 12,68 · Bestellnummer: 1030505-94900 · ISBN 3-8246-0358-6

Diese Studie, die im Rahmen des Aufbaus der Umweltökonomischen Gesamtrechnungen (UGR) in Auftrag gegeben wurde, stellt ausführlich die vielfältigen Ansätze für eine Umweltökonomische Berichterstattung auf nationaler und internationaler Ebene dar – auch unter dem Blickpunkt des nachhaltigen Wirtschaftens. Sie zeigt deutlich die bestehenden Schwierigkeiten und die Bandbreite der Möglichkeiten, an die Probleme heranzugehen, und bildet damit eine Grundlage für die Arbeiten zur Weiterentwicklung der UGR.

**Zu beziehen durch
den Buchhandel oder
den Verlag Metzler – Poeschel
70182 Stuttgart**

**Verlagsauslieferung:
SFG – Servicecenter Fachverlage GmbH
Postfach 43 43
72774 Reutlingen**

Schriftenreihe Spektrum Bundesstatistik

herausgegeben vom Statistischen Bundesamt

Band 6

Michael Reidenbach

Öffentliche Umweltschutzausgaben als Teil der Maßnahmen im Umweltschutz

118 Seiten · DM 15,80 / EUR 8,08 · Bestellnummer: 1030506–94900 · ISBN 3–8246–0378–0

Ergebnisse eines Forschungsprojektes, das im Rahmen der Umweltökonomischen Gesamtrechnungen (UGR) vom Deutschen Institut für Urbanistik (DIFU), Michael Reidenbach, bearbeitet wurde. Im Mittelpunkt steht die Erfassung der öffentlichen Umweltschutzausgaben der verschiedenen föderalen Ebenen der Bundesrepublik Deutschland. Um eine umfassende Information im Sinne der UGR zu ermöglichen, war es notwendig, ein gegenüber vorhandenen Methoden erweitertes und modifiziertes Konzept der Datensammlung zu entwickeln und empirisch anzuwenden.

Band 7

Effizienzbemessung der Hochschulausbildung auf statistischer Grundlage

85 Seiten · DM 14,80 / EUR 7,57 · Bestellnummer: 1030507–95900 · ISBN 3–8246–0460–4

Die aktuellen Bemühungen um eine Reform des deutschen Hochschulsystems zielen vor allem auf eine Erhöhung der Effektivität der Lehre und einen verstärkten Wettbewerb zwischen den Hochschulen. So sollen Lehre und Studium mit Hilfe von vergleichenden hochschulstatistischen Indikatoren transparenter gestaltet und damit ein Beitrag zu einer effizienteren Hochschulausbildung geleistet werden. Der vorliegende Band – mit Beiträgen aus Hochschulplanung, -politik und -statistik – untersucht die Möglichkeiten und Schwierigkeiten bei der Berechnung von Leistungsindikatoren für den Hochschulbereich insbesondere auf der Datengrundlage der amtlichen Statistik.

Band 8

Roland Sturm

Ergänzung der statistischen Berichterstattung über die öffentlichen Finanzen bei zunehmender Anwendung kaufmännischer Rechnungssysteme

218 Seiten · DM 34,80 / EUR 17,79 · Bestellnummer: 1030508–99900 · ISBN 3–8246–0471–X

Modernisierung und Neukonzeption des staatlichen Bereichs werden zunehmend von der Umwandlung kommunaler Einrichtungen in kaufmännisch rechnende Wirtschaftseinheiten begleitet, die zu einer Auslagerung von Teilbereichen aus den öffentlichen Budgets führen. Davon ist die Finanzstatistik betroffen, deren Aufgabe es ist, die monetäre Seite der Staatstätigkeit darzustellen.

Es müssen Wege gefunden werden, eine im Zeitablauf vergleichbare statistische Berichterstattung über den öffentlichen Bereich zu erhalten, auch wenn die kameralistische Rechnungslegung teilweise durch andere Formen der Buchführung abgelöst wird. Die Schwierigkeiten der Zusammenführung zweier unterschiedlicher Rechnungssysteme werden – ausgehend von den Grundcharakteristika dieser Systeme – dargelegt und praktische Lösungswege für die statistische Erfassung aufgezeigt.

**Zu beziehen durch
den Buchhandel oder
den Verlag Metzler – Poeschel
70182 Stuttgart**

**Verlagsauslieferung:
SFG – Servicecenter Fachverlage GmbH
Postfach 43 43
72774 Reutlingen**

Schriftenreihe Spektrum Bundesstatistik

herausgegeben vom Statistischen Bundesamt

Band 9 (vergriffen)

Pretest und Weiterentwicklung von Fragebogen

119 Seiten · DM 17,80 / EUR 7,57 · Bestellnummer: 1030507-96900 · ISBN 3-8246-0472-8

Pretestverfahren gewinnen als Instrument der Qualitätsverbesserung sowohl in der Umfrageforschung als auch in der Umfragepraxis zunehmend an Bedeutung. Sie werden zur Vorbereitung von Erhebungen nicht nur von Marktforschungs- und sozialwissenschaftlichen Instituten, sondern auch in der amtlichen Statistik eingesetzt.

Schwerpunkt dieses Bandes sind die einzelnen Beiträge der wissenschaftlichen Tagung „Pretest und Weiterentwicklung von Fragebogen“, die im Juni 1995 im Statistischen Bundesamt stattfand. Die Beiträge befassen sich einerseits mit Methoden zum Test von Erhebungsinstrumenten und andererseits mit der praktischen Umsetzung von Pretests. Ergänzt werden die Tagungsbeiträge um die Erfahrungen aus zwei Pretests, die das Statistische Bundesamt in Zusammenarbeit mit dem Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim, durchgeführt hat.

Band 10

Bernd Schmidt

Der Preisindex für die Lebenshaltung aller privaten Haushalte in Gestalt eines Kettenindex

– Beurteilung aus praktischer, empirischer und theoretischer Sicht –

131 Seiten · DM 19,80 / EUR 10,12 · Bestellnummer: 1030510-97900 · ISBN 3-8246-0528-7

Die Diskussion um „Festbasisindizes“ und „Kettenindizes“ hat auf dem Gebiet der Verbraucherpreisstatistik langjährige Tradition. Im Prinzip sind mit der Entscheidung für oder gegen eine der beiden Methoden unterschiedliche Auffassungen im Hinblick auf den Sachverhalt der Gewichtung verbunden.

In diesem Band wird untersucht, welche praktischen, empirischen und theoretischen Konsequenzen eine Umstellung des Preisindex für die Lebenshaltung aller privaten Haushalte von einem Laspeyres-Index mit über mehrere Jahre konstanten Gewichten auf einen Laspeyres-Kettenindex mit jährlicher Anpassung der Gewichte nach sich ziehen würde.

Band 11

Interviewereinsatz und -qualifikation

175 Seiten · DM 20,80 / EUR 10,63 · Bestellnummer: 1030511-98900 · ISBN 3-8246-0567-8

Die Qualität einer Umfrage hängt neben dem Fragebogen und der Stichprobe u. a. auch von Merkmalen der Interviewer und der Interviewersituation ab. Fragen des Interviewereinsatzes und der -schulung sowie der Ergebnisverzerrung durch Interviewer sind sowohl für die markt- und sozialwissenschaftliche Forschung als auch für die amtliche Statistik von großem Interesse.

Der vorliegende Band dokumentiert die Beiträge der wissenschaftlichen Tagung „Interviewereinsatz und -qualifikation“, die im Juni 1997 gemeinsam vom Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM), der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI) und dem Statistischen Bundesamt durchgeführt wurde.

Band 12 (vergriffen)

Internationale Hochschulstatistik

147 Seiten · DM 19,80 / EUR 10,12 · Bestellnummer: 1030512-98900 · ISBN 3-8246-0575-9

Neben den nationalen Daten stellen international vergleichbare hochschulstatistische Ergebnisse eine wichtige Informationsgrundlage zur Beurteilung der Attraktivität und Wettbewerbsfähigkeit des Hochschulstandorts Deutschland dar.

Zu beziehen durch
den Buchhandel oder
den Verlag Metzler – Poeschel
70182 Stuttgart

Verlagsauslieferung:
SFG – Servicecenter Fachverlage GmbH
Postfach 43 43
72774 Reutlingen

Schriftenreihe Spektrum Bundesstatistik

herausgegeben vom Statistischen Bundesamt

Band 13

Zeitverwendung in Deutschland – Beruf, Familie, Freizeit

(In Vorbereitung.)

Band 14

Möglichkeiten einer wissenschaftlichen Nutzung von Unternehmensdaten

aus der amtlichen Statistik

177 Seiten · DM 28,80 / EUR 14,73 · Bestellnummer: 1030514–99900 · ISBN 3–8246–0596–1

Bessere Zugangsmöglichkeiten zu Daten der amtlichen Statistik sind eine häufig formulierte und vieldiskutierte Forderung der Wissenschaft nicht nur in Deutschland. Der vorliegende Band versucht, den Stand der Diskussion über die wissenschaftliche Nutzung von Unternehmensdaten aus der amtlichen Statistik aufzuzeigen und will dieser Nutzung neue Impulse geben.

Die Beiträge geben die Referate wieder, die auf einem Workshop gehalten wurden, der am 24./25. Februar 1999 in Wiesbaden vom Statistischen Bundesamt in Zusammenarbeit mit Prof. Dr. Wagner von der Universität Lüneburg veranstaltet wurde. Es werden verschiedene Zugangsmöglichkeiten zu Daten der amtlichen Statistik erörtert, über Erfahrungen aus bestehenden Kooperationen berichtet und Vorschläge für die künftige Nutzung von Einzeldaten aus der amtlichen Statistik entwickelt. Schließlich wird aufgezeigt, welche wissenschaftlichen Fortschritte durch die Nutzung solcher Daten erreicht werden können.

Band 15

Neue Erhebungsinstrumente und Methodeneffekte

179 Seiten · DM 28,80 / EUR 14,73 · Bestellnummer: 1030515–00900 · ISBN 3–8246–0619–4

Wichtige Trends in der Umfrageforschung sind die Pluralisierung der Interviewformen und die Technisierung und Computerisierung der Datenerhebung. Die Kenntnis der Methodeneffekte neuer Erhebungsinstrumente ermöglicht es, für einen bestimmten Untersuchungsgegenstand die mit neuen Erhebungstechniken verbundenen Chancen zu nutzen, aber auch die damit zweifellos verbundenen Risiken zu erkennen und zu vermeiden.

Der Band enthält die Beiträge der wissenschaftlichen Tagung „Neue Erhebungsinstrumente und Methodeneffekte“, die im Juni 1999 gemeinsam vom Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM), der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI) und dem Statistischen Bundesamt durchgeführt wurde.

**Zu beziehen durch
den Buchhandel oder
den Verlag Metzler – Poeschel
70182 Stuttgart**

**Verlagsauslieferung:
SFG – Servicecenter Fachverlage GmbH
Postfach 43 43
72774 Reutlingen**
