SOEP The German Socio-Economic Panel Study

33

# SOEP Survey Papers

Series B - Survey Reports (Methodenberichte)

SOEP – The German Socio-Economic Panel Study at DIW Berlin 2011

# SOEP 1998 – Methodenbericht Erstbefragung der Stichprobe E zum Befragungsjahr 1998 (Welle 15) des Sozio-oekonomischen Panels

DIW BERLIN

Running since 1984, the German Socio-Economic Panel Study (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing.

The SOEP Survey Papers is comprised of the following series:

**Series A** – Survey Instruments (Erhebungsinstrumente)

**Series B** – Survey Reports (Methodenberichte)

**Series C** – Data Documentation (Datendokumentationen)

**Series D** – Variable Descriptions and Coding

**Series E** – SOEPmonitors

**Series F** – SOEP Newsletters

**Series G** – General Issues and Teaching Materials

The SOEP Survey Papers are available at
**http://www.diw.de/soepsurveypapers**

**Editors:**
Prof. Dr. Gert G. Wagner, DIW Berlin and Technische Universität Berlin
PD Dr. Joachim R. Frick, DIW Berlin
Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:

Infratest Burke Sozialforschung. 2011. SOEP 1998 – Methodenbericht Erstbefragung der Stichprobe E zum Befragungsjahr 1998 (Welle 15) des Sozio-oekonomischen Panels. SOEP Survey Papers 33: Series B. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)

Contact: DIW Berlin
SOEP
Mohrenstr. 58
10117 Berlin

Email: soeppapers@diw.de

INFRATEST BURKE SOZIALFORSCHUNG

# SOEP 1998 – METHODENBERICHT ERSTBEFRAGUNG DER STICHPROBE E ZUM BEFRAGUNGSJAHR 1998 (WELLE 15) DES SOZIO-OEKONOMISCHEN PANELS

*München, 1998*

*Reprint 2011*

Inhalt | Seite


| | | |
|---|---|---|
| **1.** | **Erstbefragung der Stichprobe E: Methodische Konzeption** | **3** |
| 1.1 | Stichprobe und Auswahlverfahren | 3 |
| 1.2 | Kontrollierte Einführung computerunterstützter Interviews (CAPI) | 5 |
| 1.3 | Nonresponse-Informationen | 9 |
| **2.** | **Durchführung der Erhebung** | **11** |
| 2.1 | Instrumentenentwicklung | 11 |
| 2.2 | Erste Feldphase (Pretest) | 14 |
| 2.3 | Zweite Feldphase (Haupterhebung) | 15 |
| **3.** | **Die realisierte Nettostichprobe** | **17** |
| 3.1 | Ausgangsbrutto und realisierte Interviews auf Haushalts- und Personenebene | 17 |
| 3.2 | Der Methodensplit CAPI - PAPI | 20 |
| 3.3 | Verteilung der Interviews nach Feldphasen | 24 |
| 3.4 | Interviews pro Interviewer | 26 |
| 3.5 | Befragungserfolg nach Sample-Points | 27 |
| 3.6 | Nonresponse-Analyse | 28 |
| **4.** | **Datenprüfung und übergebener Datenbestand** | **33** |
| ANHANG | | 34 |

# 1. Erstbefragung der Stichprobe E: Methodische Konzeption

## 1.1 Stichprobe und Auswahlverfahren

Das SOEP besteht aus mehreren Teilstichproben, die seit Beginn des Projekts im Jahr 1984 die Entwicklung der Wohnbevölkerung widerspiegeln (siehe ***Abbildung 1***).

Im Jahr 1998 wird dieses Stichprobensystem ergänzt durch eine neue Stichprobe E, die die gesamte Wohnbevölkerung abbildet („Querschnitt 1998"). Die Funktion dieser Zusatzstichprobe ist in einem DIW-Diskussionspapier ausführlich dargestellt.[1]

Die vorgesehene Fallzahl der Stichprobe E umfaßt 1.000 Haushalte. Wie in den übrigen SOEP-Stichproben, werden im Haushalt alle Personen ab 16 Jahren befragt, so daß die Daten sowohl haushalts- als auch personenbezogen ausgewertet werden können. Da in den Haushalten im Durchschnitt etwa zwei Personen ab 16 Jahren leben, liegt die erwartete Zahl von Personeninterviews bei etwa 2.000.

Die Haushalte werden nach demselben Random-Verfahren ausgewählt, das auch für die SOEP-Stichproben A und C zur Anwendung kam. Es handelt sich um eine mehrfach geschichtete Zufallsauswahl aus den ca. 80.000 Stimmbezirken zur Bundestagswahl. In den ausgewählten Stimmbezirken (Sample-Points) wird eine zufällig ausgewählte Startadresse bestimmt, die Ausgangspunkt für einen Random-Walk ist. Der Interviewer listet auf einem - nach vorgegebenen Regeln bestimmten - Begehungsweg jeden fünften Haushalt als Befragungshaushalt auf.

Die Adressenauflistung für Stichprobe E wurde zeitlich getrennt *vor* der Durchführung der Interviews vorgenommen. Die für die Befragung eingesetzten Interviewer erhielten die Befragungsadressen dann als fest vorgegebene Adressen. Das Stichprobenverfahren kann daher als *„Random-Route mit Adressenvorlauf"* beschrieben werden.

Da mit einer erreichbaren Stichprobenausschöpfung im Bereich zwischen 50% und 60% gerechnet wurde, wurde die Bruttostichprobe der Befragungsadressen auf 2.000 Haushalte festgelegt. Bei genauer Replikation des methodischen Vorgehens in Welle 1 des SOEP im Jahr 1984 (Stichprobe A) wäre diese Stichprobe auf 200 Sample-Points mit je 10 Adressen zu verteilen gewesen. In diesem Punkt weicht die Vorgehensweise in Stichprobe E jedoch von der in Stichprobe A ab. Grund ist der Methoden-Split, der im folgenden Abschnitt erläutert wird.

---

[1] DIW-Diskussionspapier Nr. 163: Funktion und Design einer Ergänzungsstichprobe für das Sozioökonomische Panel (SOEP). Berlin, Februar 1998

Datenerhebung SOEP

# SOEP 1984 - 2000

| Stichproben | 1984 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A Deutsche West | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| B Ausländer | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| C Deutsche Ost | | | | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| D Zuwanderer I | | | | | | | | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Zuwanderer II | | | | | | | | | | | | 1 | 2 | 3 | 4 | 5 | 6 |
| E Querschnitt '98 | | | | | | | | | | | | | | | 1 | 2 | 3 |

| Befragte Personen | 84 | 90 | 97 |
|---|---|---|---|
| A+B | 12.240 | 9.519 | 8.467 |
| C | - | 4.453 | 3.844 |
| D | - | - | 972 |
| **Insgesamt** | **12.240** | **13.972** | **13.283** |

Abbildung 1

## 1.2 Kontrollierte Einführung computerunterstützter Interviews (CAPI)

Als langfristig angelegtes Projekt muß das SOEP sich mit dem Wandel von Befragungsmethoden und -technologien, der in der Umfragepraxis vor sich geht oder für die Zukunft absehbar ist, auseinandersetzen. In den vergangenen Jahren wurde in allen Ländern mit hochentwickelter Infrastruktur der Umfrageforschung die Umfragepraxis weitgehend auf eine neue Basis gestellt, und zwar durch Umstellung auf computerunterstütztes Interviewen (CAI). Dabei ging es zunächst um telefonische Interviews (CATI), dann aber auch um mündlich-persönliche Interviews (CAPI) oder Interviews mit Selbstausfüllen des Fragebogens durch die Befragten (CASI). Letztere werden durch die zunehmende Nutzung des Internet für die Durchführung von Befragungen in Zukunft verstärkte Bedeutung erhalten.

Beim SOEP kommt aus verschiedenen Gründen vorerst nur die Umstellung der mündlich-persönlichen Befragung auf Laptop-Interviews (CAPI) in Frage. Die möglichen Gewinne liegen in einer Verbesserung der Datenqualität, die mit Hilfe der Computerunterstützung des Interviews zu erreichen ist ***(Abbildung 2).***

Die Strategie für eine CAPI-Umstellung im SOEP wurde zwischen DIW und Infratest und im SOEP-Beirat ausführlich diskutiert. Dabei wurden auch internationale Erfahrungen berücksichtigt.[2] Infratest hat in bezug auf das SOEP lange Zeit eine „konservative" Position vertreten, also für die Beibehaltung der herkömmlichen Interviewmethode argumentiert. Dafür gibt es zwei gute Gründe:

(1) Die Umstellung einer laufenden Längsschnitterhebung verlangt in besonderem Maße Kontinuität. Interviewer und Befragte haben sich in den vielen Jahren der Panelbefragung mit bestimmten Verfahren aufeinander eingespielt. Änderungen hier bedeuten ein Risikio für die Stichprobenausschöpfung. Das gilt um so mehr, als der spezielle Interviewerstab für das SOEP bei einem Wechsel auf computerunterstützte Verfahren, also Laptop-Interviews, zu erheblichen Teilen ausgetauscht werden müßte.

(2) Die SOEP-Befragung ist kein normales Interview. Die Anforderung, daß *alle* Personen ab 16 Jahren im Haushalt interviewt werden müssen, und die extrem hohen Ausschöpfungserwartungen haben zu einem komplexen Verfahrensmix geführt, in dem neben dem persönlich-mündlichen Interview auch das Selbstausfüllen des Fragebogens durch die Befragten zugelassen ist und eine erhebliche Bedeutung gewonnen hat (***Abbildung 3***). Diese flexiblen Anteile der Befragungsmethodik sind nicht über CAPI organisierbar. Bei CAPI muß wirklich mit allen Haushaltsmitgliedern ein mündlich-persönliches Interview geführt werden, was bei Mehr-Personen-Haushalten eine erhebliche Zeitbelastung bedeuten kann und die Organisation der Befragung (Termine, Anfahrtswege) erschwert.

---

[2] Vgl. etwa Moon, Nick and Heather Lauri (1997)

# CAPI: Wo die Computerunterstützung hilft

## Fehlerminimierung im Interview

- Filterführung
- Unzulässige Mehrfachnennungen
- Vollständige Angaben, z.B. in Verlaufsschemata (Tätigkeitskalender, Biographie)
- Numerische Werte: plausible Wertbereiche
- Konsistenz zwischen den Antworten auf verschiedene Fragen (ggf. auch von Welle zu Welle)
- Kommentar-Felder

## Steuerung im Panel-Haushalt

- Zuweisung der richtigen Instrumente zu jeder Person
- Identifizierung der Befragungsperson
- Kontrolle der vollständigen Bearbeitung, Protokollierung von Problemen

Abbildung 2

Datenerhebung SOEP

# Das SOEP-Interview: Besonderheiten

| Instrumente pro Haushalt | | Interviewer | | | Zentrale Bearb. | |
|---|---|---|---|---|---|---|
| | | F2F | | SA | Tel. | Mail / SA |
| • Haushaltsprotokoll | (Person X) | ✓ | | | ✓ | |
| • Haushaltsfragebogen | (Person X) | ✓ | oder | ✓ | | ✓ |
| • Personenfragebogen | (Person 1) | ✓ | oder | ✓ | | ✓ |
| | (Person 2) | ✓ | oder | ✓ | | ✓ |
| • Lebenslauf | (neue Person 3) | ✓ | oder | ✓ | | ✓ |

Abbildung 3

Die *neue Stichprobe E* bietet nun allerdings die Möglichkeit, eine computerunterstützte Durchführung der SOEP-Interviews zunächst im Rahmen einer *Erstbefragung* von Haushalten zu erproben. Die Längsschnittstichproben A bis D werden davon zunächst nicht berührt. Die in Stichprobe E gesammelten Erfahrungen können später als Basis für weiterreichende Entscheidungen in bezug auf das SOEP insgesamt dienen.

Unter dem Gesichtspunkt des Befragungserfolgs und der Verminderung von Risiken für die Längsschnittstichproben stellt der Start von Stichprobe E sozusagen ein „Fenster der Gelegenheit" dar, um computerunterstütztes Interviewen im SOEP einzuführen. Daher hat Infratest für diesen Schritt plädiert. Aus der Perspektive der *Datennnutzer* ist jedoch ein weiterer Gesichtspunkt zu berücksichtigen, nämlich die Frage möglicher Effekte der Befragungsform auf das Antwortverhalten (*mode effects*).

Der Wunsch, solche möglichen Effekte erkennen und analysieren zu können, führte zur Forderung nach einem „kontrollierten" Übergang zu computerunterstützten Interviews. Für Stichprobe E wurde daher ein *Split-Design* beschlossen:

- Stichprobe E wird geteilt in zwei strukturgleiche Teilstichproben E1 und E2.
- Stichprobe E1 wird mit herkömmlichen Befragungsmethoden bearbeitet, also mit „Paper-and-Pencil-Interviews (PAPI)"
- Stichprobe E2 wird mit der neuen Befragungstechnik bearbeitet, also mit „Computer Assisted Personal Interviews (CAPI)".

Die Strukturgleichheit der Stichproben wird dadurch hergestellt, daß verbundene Sample-Points („Zwillingspoints") für beide Stichproben gebildet werden:

- Die Bruttostichprobe von 2.000 Haushalten wird auf 125 Sample-Points à 16 Adressen verteilt. Von diesen 16 Adressen werden jeweils 8 der Stichprobe E1 (PAPI) und der Stichprobe E2 (CAPI) zugeordnet. Die Zuordnung der Adressen erfolgt nach dem „Reißverschluß", so daß die größtmögliche Identität beider Stichproben im Hinblick auf regionale und soziale Verteilungsmerkmale erreicht wird.

- Die Doppelpoints von zweimal 8 Adressen werden jeweils von *einem* Interviewer bearbeitet. Dieser muß also im Wechsel herkömmliche Interviews (PAPI) und Laptop-Interviews (CAPI) durchführen. Damit wird erreicht, daß mögliche Interviewereffekte im Einsatz von CAPI und PAPI als Störfaktor in einem Methodenvergleich ausgeschaltet sind.

Für jede Befragungsadresse ist vorab festgelegt, ob sie über CAPI oder PAPI bearbeitet werden soll. Eine gewisse Flexibilität in der Wahl der Interviewmethode wurde allerdings zugelassen. So durfte der Interviewer bei CAPI-Haushalten mit mehreren Befragungspersonen ggf. für einzelne dieser Personen einen Papierfragebogen zum Selbstausfüllen hinterlassen, wenn anders eine vollständige Teilnahme des Haushalts nicht zu erreichen war.

## 1.3 Nonresponse-Informationen

Die Teilnahme an der SOEP-Befragung ist freiwillig. In dem Maße, wie die Teilnahmebereitschaft mit bestimmten sozialen Merkmalen einhergeht (z.B. Haushaltsgröße, Einkommen, Bildung), kann die realisierte Nettostichprobe strukturell von den Verteilungen in der Grundgesamtheit abweichen. Dies kann ex post durch den Strukturvergleich mit externen Referenzstatistiken festgestellt werden („externe Validierung"). Oder man kann versuchen, Strukturmerkmale der Ausfälle in der Befragung zu erheben und auf dieser Basis Strukturverteilungen der Nichtteilnehmer mit denen der Teilnehmer zu vergleichen (Nonresponse-Informationen als Basis einer „internen Validierung").

Die Erstbefragung in der SOEP-Stichprobe E wurde so angelegt, daß Nonresponse-Informationen verfügbar sind. ***Abbildung 4*** zeigt die einbezogenen Analysemerkmale.

Abbildung 4

Infratest Burke®
Sozialforschung

# SOEP*E: Die Non-Response-Studie

## 1 Interviewergestützte Quartiersbeschreibung [1]

*Info für 100% der Bruttoadressen:*

1. Entfernung zur nächsten Großstadt
2. Art der Wohnbebauung
3. Art der Nutzung
4. Soziale Schicht der Bewohner

## 2 Haushaltsmerkmale

*Info für ca. 50 - 80% der Ausfälle:*

1. Haushaltsgröße

   *Angaben zum Haushaltsvorstand:*

2. Geschlecht
3. Altersgruppe
4. Nationalität (ggf. Deutschkenntnisse)
5. Berufstätigkeit / Stellung im Beruf

[1] nach ZUMA / Hoffmeier-Zlotnik

# 2. Durchführung der Erhebung

## 2.1 Instrumentenentwicklung

Die Erstbefragung in Stichprobe E erfolgt mit denselben Erhebungsinstrumenten wie die gleichzeitig stattfindende Befragung in den Altstichproben A bis D. Eingesetzt werden also

- das Adressenprotokoll mit Angaben zur Zusammensetzung des Haushalts
- der Haushaltsfragebogen[3]
- der Personenfragebogen[4]

Daneben gibt es eine Reihe „flankierender" Bestandteile des Erhebungsinstrumentariums. Diese mußten überwiegend für die spezifischen Anforderungen der Stichprobe E gestaltet werden. Eine vollständige Auflistung enthält ***Abbildung 5***.

Die CAPI-Befragung erfordert eine Umsetzung aller Fragebogen in ein DV-gestütztes Instrument. Infratest verwendet für die CAPI-Programmierung die spezielle Software *Intuitive* (auch *In2quest* genannt). Die Übermittlung des CAPI-Fragebogens an den Interviewer ebenso wie die Rückübermittlung der Daten bzw. der ausgefüllten Fragebogen an Infratest erfolgen über Modem.

Die CAPI-Programmierung stellt - wenn es um Fragebogen von vergleichsweise hoher Komplexität wie im SOEP geht - hohe Anforderungen an die Beteiligten. Es muß dafür eine enge Kommunikation zwischen DV-Programmierern und denjenigen, die inhaltlich mit dem Fragebogen vertraut sind, hergestellt werden. Die Voraussetzungen hierfür sind bei Infratest gut, nachdem die SOEP-Projektgruppe bei Infratest die Instrumentenentwicklung - in enger Abstimmung mit dem DIW - maßgeblich mitverantwortet und die Datenprüfung und -bereinigung ganz in den Händen dieser Projektgruppe liegt.

Die Arbeiten an der CAPI-Programmierung begannen bei Infratest Mitte Juli 1997 und wurden bis Mitte Januar 1998 mit dem Start des Pretests abgeschlossen. Mitte November 1997 fand bei Infratest ein CAPISOEP-Workshop statt, auf dem der DIW-Projektgruppe der Stand der Arbeiten vorgestellt und erläutert wurde. Die CAPI-Instrumente (Adressenprotokoll, Haushaltsfragebogen, Personenfragebogen) wurden anschließend dem DIW für eigene Tests als DV-Programm zur Verfügung gestellt.

---

[3] Abweichend von der Fragebogenfassung für Stichproben A bis D ist in der Fassung für Stichprobe E die Filterführung so gestaltet, daß die Fragen zur Wohnung komplett an *alle* Befragten gerichtet werden.

[4] Der ergänzende BIOGRAPHIE-Fragebogen wird - ebenso wie in den Altstichproben A bis D - erst zu einem späteren Zeitpunkt eingesetzt. Vorgesehen ist dafür Welle 3.

Abbildung 5

Infratest Burke ©

# SOEP '98 - Stichprobe E Methodensplit CAPI / PAPI Erhebungsunterlagen

Inhaltsverzeichnis


## ADRESSENERHEBUNG (1)

Adressenauflistungsbogen
Quartiersbeschreibung
Anweisung - Adressenermittlung

## SONSTIGE ERHEBUNGSINSTRUMENTE (2)

Interviewerhandbuch
Anschreiben mit Broschüre „Leben in Deutschland"
Adressenprotokoll E1 - PAPI Haushalte
Adressenprotokoll E2 - CAPI Haushalte
Liste 1 zum Adressenprotokoll
Datenschutzblatt

## FRAGEBOGEN

PAPI-Haushaltsfragebogen (3)
CAPI-Haushaltsfragebogen (Ausdruck auf Papier) (4)
Listenheft Haushaltsfragebogen
PAPI-Personenfragebogen (5)
CAPI-Personenfragebogen (Ausdruck auf Papier) (6)
Listenheft Personenfragebogen

Die Konzeption von Infratest lief zunächst auf eine „volle CAPI-Steuerung" des gesamten Interviewablaufs im Haushalt hinaus: Beginnend mit der Eingabe der Haushaltszusammensetzung im Adressenprotokoll und von dort ausgehend mit automatischem Aufruf der verschiedenen Erhebungsinstrumente für verschiedene Befragungspersonen. Das entsprechende Instrument konnte auch in ansprechender, funktionsfähiger Form programmiert werden. Gleichwohl wurde nach den ersten Pretesterfahrungen entschieden, auf den Einsatz dieses Instruments vorerst zu verzichten (vgl. Kapitel 2.2). Eingesetzte CAPI-Instrumente sind also

- der Haushaltsfragebogen
- und der Personenfragebogen.

Ein wichtiger Unterschied zwischen den CAPI- und den Papierfragebogen beim SOEP besteht darin, wie optische Hilfsmittel für den Befragten (Listenvorlagen, Skalen) gehandhabt werden. Bei einem normalen mündlich-persönlichen Interview wird für diesen Zweck ein „Listenheft" verwendet, das dem Befragten direkt vorgelegt wird. Der bisherige SOEP-Fragebogen (auf Papier) arbeitet demgegenüber *ohne* Listenheft, da er - im Hinblick auf die Selbstausfüller - ohnehin möglichst großzügig und ansprechend gestaltet ist. Man kann daher davon ausgehen, daß die Interviewer im Fall eines mündlich-persönlichen Interviews bei Fragen, wo dies erforderlich ist, die Befragungsperson einfach in den Fragebogen schauen lassen.

Bei den CAPI-Interviews soll hingegen *nicht* so verfahren werden. Der Befragte soll nicht mit dem Interviewer vor dem Bildschirm sitzen, sondern das Interview soll in einer echten Gesprächsform „face-to-face" geführt werden. Benötigte optische Hilfsmittel (Listen, Skalen) werden dem Befragten *auf Papier* vorgelegt. Erstmals nach 1984 wird im SOEP daher jetzt - in Stichprobe E - mit einem *Listenheft* gearbeitet. Um die Vergleichbarkeit der Interviewsituation in den zwei Teilstichproben E1 und E2 zu gewährleisten, wurden die Interviewer angewiesen, das Listenheft nicht nur bei CAPI-, sondern auch bei PAPI-Interviews zu verwenden.[5]

Die CAPI-Instrumente können über eine RTF-Ausgabe auf Papier ausgedruckt werden. Dem DIW liegt ein umfangreicher Materialband vor, in dem Infratest alle Erhebungsinstrumente in dieser Weise dokumentiert hat.

---

[5] Zu beachten ist, daß die Interviewsituation in E1 und E2 in diesem Punkt leicht abweichend gestaltet ist als in den Altstichproben A bis D. Dies kann auf die Antwortverteilungen insbesondere bei Skalenfragen (z.B. Zufriedenheitsfragen) durchaus einen Einfluß haben.

## 2.2 Erste Feldphase (Pretest)

CAPI-Instrumente müssen sehr sorgfältig auf Handhabbarkeit und Fehlerfreiheit getestet werden. Dies geschieht zunächst „trocken", d.h. durch die Bearbeiter selbst und mit ersten ausgewählten Testpersonen. Zweiter Schritt sollte ein Pretest unter „Ernstfall"-Bedingungen sein.

Für die SOEP-Stichprobe E wurde der Pretest so angelegt, daß die Interviews im Fall einer erfolgreichen Durchführung in den Gesamtdatenbestand integriert werden könnten. Dies wurde mit folgendem Vorgehen erreicht:

- Auf der Ebene der Auswahl der Sample-Points wurde als erstes eine einheitliche Gesamtstichprobe mit 125 Sample-Points gezogen.
- Von den 125 Points wurden 10 zufällig für den Pretest ausgewählt.
- Die Adressenauflistung in den Sample-Points fand zunächst nur in den 10 Pretest-Points statt (Dezember 1997).
- Von den 160 Befragungsadressen für den Pretest wurden je 80 den Stichproben E1 (PAPI) und E2 (CAPI) zugewiesen. Mit diesen Adressen wurden die Befragungsunterlagen Ende Januar an die 10 Pretestinterviewer versandt.

Wie sich zeigte, hatten die Interviewer erhebliche Startprobleme. In der zunächst benutzten CAPI-Version konnte das CAPI-Instrument nicht - wie gewohnt - per Modem auf die Laptops der Interviewer überspielt werden, sondern mußte auf einer CDRom übersandt und vom Interviewer selbst installiert werden. Die Handhabung des elektronischen Adressenprotokolls war ungewohnt. Die Terminvorgaben waren wegen der Pretestbedingungen sehr eng.

Im Ergebnis lagen Anfang März, am Ende der ersten Feldbearbeitungsphase des Pretests, enttäuschend wenige realisierte Interviews vor, nämlich 53 von 160, darunter 33 PAPI- und 20 CAPI-Interviews. Um die Ursachen der Probleme zu klären, wurden ausführliche Gespräche mit allen zehn Pretestinterviewern geführt. Es zeigte sich, daß die CAPI-Interviews selbst keine Schwierigkeiten bereiteten; die Interviewer waren im Gegenteil voller Lob über die Gestaltung der CAPI-Fragebogen. Die Startschwierigkeiten waren allein in den genannten Handling-Problemen begründet.

Die Konsequenz war die bereits erwähnte Entscheidung, das Adressenprotokoll und damit die Steuerung des gesamten Verfahrens im Befragungshaushalt aus dem CAPI-Instrument herauszunehmen. Für die CAPI-Stichprobe E2 wird nun ein Papier-Adressenprotokoll verwendet wie für die PAPI-Stichprobe E1. Der Einsatz von CAPI wird auf das eigentliche Interview (Haushalts- und Personenfragebogen) begrenzt.

Mit dieser Verfahrensmodifikation wurden die überarbeiteten Instrumente für die Nachbearbeitungsphase des Pretests Ende März an die Interviewer gesandt. Das Instrumentarium konnte damit vor Beginn der Haupterhebung noch einmal getestet werden. Es ergaben sich keine Hinweise auf Probleme oder nochmaligen Änderungsbedarf.

Insgesamt konnten aus den 160 Pretestadressen schließlich 70 Haushalte interviewt werden, darunter 39 mit PAPI und 31 mit CAPI. In bezug auf die erreichte Stichprobenausschöpfung war dies zwar noch kein befriedigendes Ergebnis. Die realisierten Interviews erwiesen sich nach genauer Prüfung der CAPI-Daten jedoch als fehlerfrei und auswertbar. An den programmierten CAPI-Fragebogen brauchten nur minimale Endkorrekturen vorgenommen zu werden.

Die Interviewdaten aus den 70 Pretesthaushalten, die zur Teilnahme an der Befragung bereit waren, wurden daher wie angestrebt in den Gesamtdatenbestand eingestellt.[6]

## 2.3 Zweite Feldphase (Haupterhebung)

Die zweite Feldphase umfaßt die Bearbeitung des Hauptteils der Stichprobe, also 115 von 125 Sample-Points (1.840 Befragungsadressen).

Diese Phase begann im März 1998 mit der Adressenerhebung. Für jeden Sample-Point wurden 16 Adressen plus 4 Reserveadressen (für den Fall qualitätsneutraler Ausfälle) nach dem oben beschriebenen Random-Route-Verfahren aufgelistet. Zu jeder Adresse beantwortete der Interviewer darüber hinaus vier Fragen zur „interviewergestützten Quartiersbeschreibung":[7]

(1) Entfernung zur nächsten Großstadt
(2) Gebäudetyp
(3) Art der Nutzung
(4) Soziale Schicht

Der dafür entwickelte Erhebungsbogen hatte sich im Pretest voll bewährt und konnte unverändert eingesetzt werden. Die Interviewer erhielten jetzt lediglich einen deutlicheren Hinweis, daß unbewohnte Wohnungen - soweit erkennbar - bei der Auflistung nicht aufzunehmen sind. Im Pretest hatten sich 6% der aufgelisteten Wohnungsadressen als (derzeit) nicht bewohnt bzw. nicht von einem Privathaushalt bewohnt erwiesen.

Diese sogenannten „qualitätsneutralen Ausfälle (QNA)" werden dem Institut zurückgemeldet, und es wird dem Interviewer pro gemeldetem QNA eine Ersatzadresse zur Bearbeitung gegeben.

Die Unterlagen für die eigentliche Befragung wurden den Interviewern zusammen mit den aufbereiteten Befragungsadressen am 22. April 1998 übersandt. Damit konnte die Feld-

---

6 Sie sind im Bruttodatenbestand anhand der Variable ART=1 identifizierbar.
7 In Abstimmung mit ZUMA gestaltet auf Basis von Hoffmeyer-Zlotnik (1995)

arbeit beginnen. Sie war zunächst auf drei Monate, also bis Ende Juli veranschlagt. Tatsächlich mußte sie um zwei weitere Monate, d.h. bis Ende September, ausgedehnt werden.

Pro Sample-Point mit 16 Adressen wurde zunächst ein Interviewer eingesetzt. Tatsächlich kamen dann im gesamten Feldverlauf für die 125 Points (einschl. Pretest) 152 Interviewer zum Einsatz. 23 Interviewer nahmen die Arbeit an diesem Projekt gar nicht erst auf oder konnten keinen einzigen Befragungshaushalt zur Teilnahme gewinnen. Der Ersatz eines erfolglosen Interviewers durch einen anderen CAPI-Interviewer war teilweise nur möglich, indem lange Anfahrtswege in den Point in Kauf genommen wurden. Mitte Juli, d.h. kurz vor Abschluß der vorgesehenen Feldzeit, waren statt der angestrebten 1.000 bis 1.100 Teilnehmerhaushalte lediglich ca. 800 Haushalte erfolgreich interviewt. Es wurden daher drei Maßnahmen beschlossen, um die vorgegebene Zielzahl von Teilnehmerhaushalten zu erreichen:

- Verlängerung der Feldarbeit um weitere zwei Monate

- Nachbearbeitung von Sample-Points notfalls auch durch Interviewer, die nicht über die Ausstattung und Schulung für CAPI-Interviews verfügen; damit konnte der CAPI/PAPI-Split innerhalb der Sample-Points nicht voll durchgehalten werden.

- Ersatz von Sample-Points mit sehr niedriger Pointausschöpfung (drei oder weniger Teilnehmerhaushalte) durch neugezogene Points in derselben Region und Gemeindegrößenklasse. Es wurden also neue, statistisch äquivalente Stimmbezirke nach demselben Zufallsverfahren wie in der ursprünglichen Stichprobenziehung ausgewählt. In diesen wurde zunächst eine neue Adressenerhebung durchgeführt und dann 16 neue Adressen zur Bearbeitung gegeben. 20 von den 125 Sample-Points im endgültigen Datenbestand sind solche Ersatzpoints für nicht erfolgreich bearbeitete ursprüngliche Sample-Points. Die realisierten Interviews in den ersetzten Points (rund 50 Haushalte) wurden in den endgültigen Datenbestand nicht aufgenommen.

All diese notwendigen Maßnahmen sind ein Hinweis auf den enormen Schwierigkeitsgrad dieser Befragung, der weit über die üblichen Anforderungen an einen Interviewerstab hinausgeht. Bereits die SOEP-Befragung als solche stellt sehr hohe Anforderungen an die Interviewer (vgl. oben Abbildung 2). Durch den hinzukommenden Methodensplit, der ein wechselndes Arbeiten mit herkömmlichen und computerunterstützten Interviewverfahren verlangt, wurden die Anforderungen noch einmal erhöht.

Ende September konnte die Feldarbeit für die Erstbefragung der SOEP-Stichprobe E letztlich erfolgreich abgeschlossen werden. Die Datenaufbereitung und Datenprüfung erfolgte bis Ende Oktober. Das folgende Kapitel 3 gibt einen genaueren Überblick über die realisierte Nettostichprobe.

# 3. Die realisierte Nettostichprobe

## 3.1 Ausgangsbrutto und realisierte Interviews auf Haushalts- und Personenebene

Die Bruttostichprobe umfaßt, wie beschrieben, zunächst 2.000 Adressen. Von diesen erwiesen sich 60 (3%) als derzeit unbewohnte Wohnungen bzw. kein Privathaushalt (qualitätsneutrale Ausfälle). Soweit diese Information rechtzeitig vorlag, konnte eine Ersatzadresse eingesetzt werden; dies geschah in 52 Fällen. Daraus ergeben sich zwei Eckwerte für die weitere Analyse:

- Das *„erweiterte Brutto"* umfaßt alle 2.052 an die Interviewer zur Bearbeitung gegebenen Adressen. Dies ist die Fallzahl der Stichprobe, wie sie auf dem sogenannten „Bruttoband", das dem DIW als Datenbestand übergeben wird, dokumentiert ist.

- Das *„bereinigte Brutto"* umfaßt denselben Adressenbestand, jedoch ohne die qualitätsneutralen Ausfälle (QNA). Dies ist die Basis für die Berechnung der Stichprobenausschöpfung (n = 1.992).

***Tabelle 1*** zeigt, mit welcher Zahl von Ausfällen und Interviews die Befragung realisiert werden konnte.

Von den bearbeitbaren Befragungsadressen waren 4,6% Ausfälle, weil im Haushalt nie jemand angetroffen wurde. Der Hauptteil der Ausfälle mit 40,7% entfällt auf ausdrückliche Verweigerungen. Ein kleiner Teil der durchgeführten Interviews wurde nachträglich als nicht auswertbar gewertet (0,5%), etwa wegen Unvollständigkeit der zu einem Haushalt gehörenden Unterlagen.

Es verbleiben 1.079 auswertbare Interviews, d.h. Teilnehmerhaushalte in der Befragung. Die vereinbarte Zielgröße von 1.000 Haushalten in Stichprobe E wurde damit deutlich überschritten.

Die Stichprobenausschöpfung beträgt 54,2%[8]. Sie liegt damit erwartungsgemäß niedriger als in den vergleichbaren SOEP-Stichproben, deren Erstbefragung mehr oder weniger lange Zeit zurückliegt.

---

[8] Dies ist fast genau die Ausschöpfung, die auch bei den ALLBUS-Befragungen 1994 und 1996 erreicht wurde.

Tabelle 1:
**Die Stichprobe: Ausgangsbrutto und realisierte Interviews auf Haushalts- und Personenebene**

| *HAUSHALTE* | | *abs.* | *%* |
|---|---|---|---|
| (1) | Ausgangsbrutto: | | |
| | 250 Sample-Points à 8 Adressen | 2.000 | |
| (2) | Qualitätsneutrale Ausfälle (QNA) [1] | 60 | |
| (3) | dafür eingesetzte Ersatzadressen | 52 | |
| (4) | Erweitertes Brutto, d.h. eingesetzte Adressen gesamt | 2.052 | |
| (5) | Bereinigtes Brutto, d.h. (4) ohne QNA | 1.992 | 100 |
| | davon: | | |
| (6) | nicht angetroffen | 92 | 4,6 |
| (7) | Verweigerung | 810 | 40,7 |
| (8) | nicht auswertbare Interviews | 11 | 0,5 |
| (9) | Auswertbare Interviews, d.h. Teilnehmerhaushalte | 1.079 | 54,2 |
| *PERSONEN* | | | |
| (10) | In den Teilnehmerhaushalten lebende Personen | 2.496 | |
| | darunter: | | |
| (11) | Kinder | 466 | |
| (12) | Personen ab 16 Jahren | 2.030 | 100 |
| | davon | | |
| (13) | ohne Interview | 73 | 3,6 |
| (14) | mit Interview | 1.957 | 96,4 |

1) Wohnung unbewohnt, Inhaber verzogen oder verstorben; kein Privathaushalt

**Stichprobenausschöpfung in Welle 1**

| | | % |
|---|---|---|
| Stichprobe A/West | 1984 | 60,6 |
| Stichprobe B/West | 1984 | 68,0 |
| Stichprobe C/Ost | 1990 | 70,0 |
| *Stichprobe E* | *1998* | |
| *gesamt* | | *54,2* |
| *West* | | *54,4* |
| *Ost* | | *53,3* |

In den 1.079 befragten Haushalten leben 2.496 Personen, darunter 2.030 Personen ab 16 Jahren. Diese bilden die Gesamtheit der befragbaren Personen, mit denen ein Personeninterview durchzuführen war. Bei 73 Personen (3,6%) konnte dies aus unterschiedlichen Gründen nicht realisiert werden. Es liegen insgesamt 1.957 auswertbare Personenfragebogen vor. Die „Vollständigkeitsquote" in der Abbildung der Teilnehmerhaushalte durch individuelle Interviews liegt mit 96,4% damit erfreulich hoch. Dies ist ein vergleichbares Ergebnis wie in den Altstichproben A bis D.

Unvollständige Teilnahme an der Befragung kommt insbesondere dann vor, wenn drei oder mehr Haushaltsmitglieder zu befragen sind. Auch in diesen größeren Haushalten liegen insgesamt jedoch für rd. 90% der befragbaren Personen auswertbare Personeninterviews vor.

**Teilnahme auf der Personenebene**

| *Haushaltsgröße* | *Personen* | *davon mit Interview* | |
|---|---|---|---|
| | *abs.* | *abs.* | *%* |
| 1 Befragungsperson | 327 | 327 | 100,0 |
| 2 Befragungspersonen | 1.200 | 1.173 | 97,8 |
| 3 Befragungspersonen | 327 | 300 | 91,7 |
| 4 und mehr Befragungspersonen | 176 | 157 | 89,2 |
| Insgesamt | 2.030 | 1.957 | 96,4 |

## 3.2 Der Methodensplit CAPI - PAPI

*Tabelle 2* zeigt die Ausschöpfungsrechnung getrennt für die zwei Teilstichproben E1 (PAPI) und E2 (CAPI). In E1 wurden 550 Haushalte zur Teilnahme gewonnen, in E2 529 Haushalte. Die Stichprobenausschöpfung beträgt 55,3% in E1 und 53,0% in E2.

Die geringfügig niedrigere Ausschöpfung in E2 ist nicht auf Vorbehalte gegenüber Laptop-Interviews zurückzuführen. Nach allem, was wir von den Interviewern über den Befragungsprozeß wissen, fällt die Entscheidung des Haushalts für oder gegen eine Teilnahme an der Befragung, bevor der Laptop überhaupt ausgepackt ist. Die speziellen Anforderungen der SOEP-Befragung machen jedoch, wie oben bereits ausgeführt, die Durchführung mit Laptop-Interviews schwieriger als mit Papierfragebogen. Mit letzteren kann, wenn der Fragebogen zum Selbstausfüllen geeignet ist, die Interviewsituation in einem Mehr-Personen-Haushalt flexibler gestaltet werden als mit Laptop-Interviews. Von daher ist es eher erstaunlich, daß die Stichprobenausschöpfung in E1 und E2 sich nur so geringfügig unterscheidet.

Ebenso erwartbar, aber in der Größenordnung ebenfalls geringfügig, ist der Unterschied zwischen beiden Teilstichproben in bezug auf die Vollständigkeit der Teilnahme aller Haushaltsmitglieder. Die Quote beträgt 97,3% in der Stichprobe E1 und 95,4% in der Stichprobe E2.

Der Stichproben-Split E1/E2 darf nun allerdings nicht gleichgesetzt werden mit der tatsächlichen Interviewmethode in den durchgeführten Haushalts- und Personeninterviews. Eine strenge Umsetzung der Splitvorgaben für PAPI- und CAPI-Haushalte wäre mit einem unvertretbaren Verlust an Stichprobenausschöpfung verbunden gewesen. Wie oben in Kapitel 2.3 beschrieben, wurden im Verlauf der Feldphase die Regeln in bezug auf die Interviewmethode gelockert, um die angestrebte Zahl an Teilnehmerhaushalten überhaupt zu erreichen.

*Tabelle 3* zeigt die tatsächlich angewendete Interviewmethode bei Haushalts- und Personeninterviews in den zwei Teilstichproben.

In Stichprobe E1 (PAPI-Split) haben die Interviewer weit mehr als in E2 von der Möglichkeit Gebrauch gemacht, Fragebogen von den Befragten selbst ausfüllen zu lassen. In beiden Stichproben kommt ein gewisser Teil von Interviews hinzu, die im Rahmen von Nachbearbeitungsmaßnahmen („conversion") schriftlich-postalisch bearbeitet wurden. Schließlich gibt es auch im PAPI-Split einzelne Interviews, die - aus welchen Gründen auch immer - mit CAPI durchgeführt wurden. Letztlich beträgt der Anteil der Interviews, die in Stichprobe E1 als mündlich-persönliches Interview mit PAPI durchgeführt wurden, 80,7% auf der Haushalts- und 66,2% auf der Personenebene.

In Stichprobe E2 liegt der entsprechende Anteil von Interviews, die als mündlich-persönliches Interview mit CAPI durchgeführt wurden, ähnlich hoch: nämlich 77,3% auf der Haushalts- und 74,8% auf der Personenebene. Ein erheblicher Teil der Interviews wurde in E2

mündlich-persönlich als herkömmliches Interview (PAPI) geführt. Dies war in der Regel keine Entscheidung des einzelnen Interviewers. Vielmehr ist es eine Folge der oben bereits erwähnten Entscheidung, in der letzten Feldphase auch Interviewer einzusetzen, die nicht zum CAPI-Stab gehören. Diese Interviewer konnten natürlich bei Haushalten, die eigentlich als CAPI-Haushalte (Stichprobe E2) definiert waren, nur herkömmliche Interviews auf Basis von Papierfragebogen durchführen.

Der Methodensplit mit den Teilstichproben E1 und E2 hatte vor allem den Zweck, eventuelle *mode-effects* beim Übergang von herkömmlichen auf computerunterstützte Interviews im SOEP analysieren zu können. Dies ist möglich, auch wenn das experimentelle Design des Stichprobensplits nicht exakt in die tatsächlich angewendete Interviewmethode umgesetzt werden konnte. Das Analysedesign im Hinblick auf *mode-effects* darf jedoch nicht auf einen Vergleich von Antwortverteilungen in den Teilstichproben E1 und E2 gestützt werden.[9] Es muß sich auf die tatsächlich angewendete Methode in den jeweiligen Interviews beziehen, also folgende drei Gruppen vergleichen:

| *Interviewmethode* | *Haushalts-fragebogen*[10] | *Personen-fragebogen*[11] |
|---|---|---|
| mündlich-persönlich PAPI | 534 | 832 |
| mündlich-persönlich CAPI | 425 | 726 |
| vom Befragten selbst ausgefüllt | 120 | 399 |
| Gesamt | 1.079 | 1.957 |

[9] Variable STICHU=10 bzw. 11

[10] Variable HFORM=1 bzw. 9 bzw. übrige

[11] Variable PFORM=1 bzw. 9 bzw. übrige.
Wenn man im Hinblick auf einen möglichen Interviewereffekt sichergehen will, könnte man die Analyse auf Interviews eingrenzen, die von CAPI-Interviewern durchgeführt bzw. betreut wurden (Variable CAPINT=1). Das wären 994 von 1.079 Haushalten.

Tabelle 2:
**Der Methodensplit: Teilstichproben E1 und E2 im Vergleich**

| | | *E1 PAPI* | | *E2 CAPI* | |
|---|---|---|---|---|---|
| *HAUSHALTE* | | *abs.* | *%* | *abs.* | *%* |
| (1) | Ausgangsbrutto: | | | | |
| | je 125 Sample-Points à 8 Adressen | 1.000 | | 1.000 | |
| (2) | Qualitätsneutrale Ausfälle (QNA) [1] | 34 | | 26 | |
| (3) | dafür eingesetzte Ersatzadressen | 28 | | 24 | |
| (4) | Erweitertes Brutto, d.h. eingesetzte Adressen gesamt | 1.028 | | 1.024 | |
| (5) | Bereinigtes Brutto, d.h. (4) ohne QNA | 994 | 100 | 998 | 100 |
| | davon: | | | | |
| (6) | nicht angetroffen | 40 | 4,0 | 52 | 5,2 |
| (7) | Verweigerung | 399 | 40,1 | 411 | 41,2 |
| (8) | nicht auswertbare Interviews | 5 | 0,5 | 6 | 0,6 |
| (9) | Auswertbare Interviews, d.h. Teilnehmerhaushalte | 550 | 55,3 | 529 | 53,0 |
| *PERSONEN* | | | | | |
| (10) | In den Teilnehmerhaushalten lebende Personen | 1.284 | | 1.212 | |
| | darunter: | | | | |
| (11) | Kinder | 238 | | 228 | |
| (12) | Personen ab 16 Jahren | 1.046 | 100 | 984 | 100 |
| | davon | | | | |
| (13) | ohne Interview | 28 | 2,7 | 45 | 4,6 |
| (14) | mit Interview | 1.018 | 97,3 | 939 | 95,4 |

1) Wohnung unbewohnt, Inhaber verzogen oder verstorben; kein Privathaushalt

Tabelle 3:
**Tatsächliche Interviewmethode in den zwei Teilstichproben**

| *Interviewmethode* | *E1 PAPI* | | *E2 CAPI* | |
|---|---|---|---|---|
| | *abs.* | *%* | *abs.* | *%* |
| ***HAUSHALTE*** | | | | |
| Mündlich-persönlich | 460 | 83,6 | 499 | 94,3 |
| darunter: | | | | |
| CAPI | 16 | 2,9 | 409 | 77,3 |
| PAPI | 444 | 80,7 | 90 | 17,0 |
| Selbst ausgefüllt | 90 | 16,4 | 30 | 5,7 |
| darunter: | | | | |
| mit Interviewer | 68 | 12,4 | 14 | 2,7 |
| schriftlich-postalisch | 22 | 4,0 | 16 | 3,0 |
| Haushalte gesamt | 550 | 100 | 529 | 100 |
| ***PERSONEN*** | | | | |
| Mündlich-persönlich | 698 | 68,6 | 860 | 91,6 |
| darunter: | | | | |
| CAPI | 24 | 2,4 | 702 | 74,8 |
| PAPI | 674 | 66,2 | 158 | 16,8 |
| Selbst ausgefüllt | 320 | 31,4 | 79 | 8,4 |
| darunter: | | | | |
| mit Interviewer | 299 | 29,4 | 60 | 6,4 |
| schriftlich-postalisch | 21 | 2,0 | 19 | 2,0 |
| Personen gesamt | 1.018 | 100 | 939 | 100 |

## 3.3 Verteilung der Interviews nach Feldphasen

In den Kapiteln 2.2 und 2.3 wurde der Verlauf der Feldarbeit bereits in groben Zügen dargestellt. ***Tabelle 4*** ergänzt diese Beschreibung durch genauere Zahlen.

Angegeben ist jeweils der Monat des letzten Haushaltskontakts. Im Fall der erfolgreichen Bearbeitung eines Haushalts ist das der Monat, in dem die Interviews im Haushalt (bzw. das letzte der Interviews) durchgeführt wurden. Bei Ausfällen ist es der Zeitpunkt, zu dem die Bearbeitung des jeweiligen Haushalts erfolglos eingestellt wurde. Teilweise klärt sich ein Ausfall dabei sehr rasch, etwa im Fall einer eindeutig leerstehenden Wohnung oder bei einer dezidierten Weigerung, an der Befragung teilzunehmen. Vielfach wird ein Ausfall jedoch zunächst als vorläufig bewertet, und es wird versucht, mit verschiedenen Formen der Nachbearbeitung noch zum Erfolg zu kommen; der letzte Kontakt liegt dann zeitlich eher in den späteren Feldphasen.

Die Zahlen in Tabelle 4 zeigen, daß etwa die Hälfte aller Interviews in zwei Monaten durchgeführt wurde: das war für die Preteststichprobe der Februar und für die Hauptstichprobe der Mai. Die wesentliche Nachbearbeitung lag für die Preteststichprobe im April, für die Hauptstichprobe in den Monaten Juli und August.

In der Nachbearbeitung wurde die Arbeit der Interviewer unterstützt durch eine zentrale Bearbeitergruppe. Diese führte telefonische Kontaktinterviews, um solche Haushalte doch noch zur Mitarbeit zu gewinnen, die zunächst eher ablehnend reagiert hatten oder bei denen die Situation ungeklärt war. So dienten rund 300 telefonische Kontaktinterviews zur Vorbereitung der Nachbearbeitung durch die Interviewer vor Ort in den Monaten Juli und August.

Teilweise lehnten die Befragungshaushalte in den Kontaktinterviews einen Interviewerbesuch weiterhin ab, erklärten sich jedoch zum Ausfüllen der Fragebogen bereit, wenn diese postalisch zugesandt würden. 33 Haushalte wurden auf diese Weise zur Teilnahme an der Befragung gewonnen.

Tabelle 4:
**Feldverlauf nach Monat des letzten Haushaltskontakts**

| | *Interviews* | *Ausfälle* |
|---|---|---|
| PRETEST | | |
| Januar | 2 | 3 |
| Februar | 51 | 16 |
| März | 5 | 1 |
| April | 12 | 77 |
| Mai | - | 3 |
| Gesamt | 70 | 100 |
| HAUPTERHEBUNG | | |
| April | 24 | 34 |
| Mai | 478 | 267 |
| Juni | 94 | 63 |
| Juli | 138 | 229 |
| August | 222 | 189 |
| September | 53 | 91 |
| Gesamt | 1.009 | 873 |

## 3.4 Interviews pro Interviewer

Die 1.079 Teilnehmerhaushalte der Stichprobe E wurden abschließend von 131 Interviewern bearbeitet. „Abschließend" heißt, daß diese Interviewer letztlich zum Erfolg gekommen sind - teilweise erst, nachdem ein anderer Interviewer den betreffenden Haushalt nicht zur Teilnahme gewinnen konnte. Unter den 131 Interviewern sind 2 aus der zentralen Bearbeitergruppe (Telefonkontakte mit schriftlich-postalischer Befragung) und 129 Interviewer vor Ort.

Diese 129 Interviewer waren sehr unterschiedlich erfolgreich. Fast jeder fünfte hat nur 1 bis 3 Haushalte erfolgreich bearbeiten können. Die meisten haben zwischen 4 und 12 Haushalten interviewt. Der „Schützenkönig" hat von 16 vorgegebenen Adressen 16 erfolgreich bearbeitet.

Einige Interviewer haben mehr als einen Sample-Point bearbeitet oder waren in verschiedenen Sample-Points zur Nachbearbeitung von vorläufigen Ausfällen eingesetzt. Die Maximalzahl der von einem Interviewer erfolgreich bearbeiteten Haushalte beträgt 37.

Tabelle 5:
**Zahl der Interviews (Haushalte) pro Interviewer**

| Erfolgreich bearbeitete Haushalte | Zahl der Interviewer |
|---|---|
| 1- 3 | 23 |
| 4 - 6 | 30 |
| 7 - 9 | 36 |
| 10 - 12 | 21 |
| 13 - 16 | 12 |
| 18 - 27 | 5 |
| 34 - 37 | 2 |
| | 129 |

## 3.5 Befragungserfolg nach Sample-Points

Der unterschiedliche Befragungserfolg der einzelnen Interviewer kann auf zwei Faktoren zurückgeführt werden: zum einen auf möglicherweise unterschiedliche *Fähigkeiten* der Interviewer selbst, zum anderen auf möglicherweise unterschiedlich leichte oder schwierige *Befragbarkeit* der Haushalte, die jeweils zu befragen waren. Aus dem Zusammenspiel beider Faktoren ergibt sich auch, daß der Befragungserfolg in den einzelnen *Sample-Points* sehr unterschiedlich ist. ***Tabelle 6*** zeigt dies auf der Ebene der „Zwillingspoints" à 16 Adressen (8 Adressen E1 plus 8 Adressen E2), die jeweils einem Interviewer zur Bearbeitung vorgegeben waren.

Die Bandbreite des sozusagen „normalen" Befragungserfolgs liegt im Bereich zwischen 5 und 12 teilnehmenden Haushalten pro Sample-Point (16 Adressen). 101 der 125 Sample-Points liegen in dieser Bandbreite. In 10 Sample-Points war der Befragungserfolg geringer, hier konnten nur 2 bis 4 Haushalte zur Teilnahme gewonnen werden. Auf der anderen Seite gibt es 14 Points mit besonders hohem Befragungserfolg, in denen 13 bis 16 Haushalte aus den 16 vorgegebenen Adressen zur Teilnahme gewonnen werden konnten.

Stichprobentheoretisch wäre die Idealbedingung eine Gleichverteilung der Stichprobe über die Sample-Points. Dies ist bei einem klassischen Stichprobendesign, wie es dieser SOEP-Befragung zugrunde liegt, praktisch jedoch nicht zu realisieren.

Tabelle 6:
**Bearbeitungserfolg nach Sample-Points**

| | Zahl der realisierten Interviews | Anzahl der Points |
|---|---|---|
| Basis: | 0 | - |
| Zwillings-Points | 1 | - |
| (E1 + E2) | 2 | 2 |
| à 16 Adressen | 3 | 4 |
| | 4 | 4 |
| | 5 | 9 |
| | 6 | 15 |
| | 7 | 16 |
| | 8 | 15 |
| | 9 | 12 |
| | 10 | 12 |
| | 11 | 9 |
| | 12 | 13 |
| | 13 | 4 |
| | 14 | 6 |
| | 15 | 3 |
| | 16 | 1 |
| | | 125 |

## 3.6 Nonresponse-Analyse

Wie oben dargestellt (Kapitel 1.3) umfaßt die methodische Anlage der Erstbefragung in Stichprobe E auch Informationen über Ausfälle (*nonresponse*). Daraus sollen Hinweise gewonnen werden, ob sich teilnahmebereite Haushalte strukturell von den nicht teilnahmebereiten unterscheiden. Je weniger dies der Fall ist, um so mehr kann man annehmen, daß die Ausfälle zufallsverteilt sind und die realisierte Nettostichprobe keine systematischen Strukturverzerrungen aufweist.

Die Nonresponse-Informationen wurden auf zweierlei Wegen ermittelt. Der erste ist die Ergänzung der Adresse durch *Merkmale des Wohnquartiers*. Die Daten werden vom Interviewer geliefert, der die Adressenerhebung durchführt („interviewergestützte Quartiersbeschreibung"). Zu diesem Zeitpunkt ist noch nicht bekannt, welcher Haushalt interviewt werden kann und welcher ein Ausfall werden wird. Die Informationen liegen für 100% der Bruttostichprobe vor.

Die zweite Art der Nonresponse-Informationen bezieht sich auf Merkmale der *Haushalte* selbst. Wenn der Interviewer einen Befragungshaushalt als Ausfall verbucht, soll er versuchen, zumindest einige wenige Merkmale über diesen Haushalt in Erfahrung zu bringen. Dies gelingt naturgemäß nur für einen Teil der Haushalte, da diese ja dadurch zum Ausfall werden, daß sie nicht kooperationsbereit sind (oder ggf. gar nicht angetroffen werden).

***Tabelle 7*** zeigt die Verteilungen der Quartiersmerkmale im Vergleich von Interviews, d.h. teilnehmenden Haushalten, und Ausfällen. Die Zahlen in der Spalte „Gesamt" (Interviews plus Ausfälle) stellen den Schätzwert der Merkmalsverteilung in der Grundgesamtheit dar. Die genauen Frageformulierungen sind im ANHANG wiedergegeben.

Die regionale Verteilung der Stichprobe, gemessen an der Entfernung zum Geschäftszentrum der (nächstgelegenen) Großstadt, weist erhebliche Unterschiede zwischen Interviews und Ausfällen aus - allerdings ohne ein klares Muster. Es gibt nicht etwa das häufig zu beobachtende Stadt-Land-Gefälle der Teilnahmebereitschaft. Vielmehr sind Stadtzentrumsbewohner in der realisierten Stichprobe eher überproportional vertreten, ebenso wie auf der anderen Seite die großstadtfernen Bewohner kleinerer Städte oder Landgemeinden.

Zur Bewertung dieses Ergebnisses ist zu beachten, daß durch die Verteilung der Stichprobe auf eine relativ geringe Zahl von Sample-Points (darüber hinaus noch mit sehr unterschiedlicher Pointausschöpfung) die regionale Struktur der Nettostichprobe nicht besonders stabil sein kann. Raumbezogene Analysen sind mit der SOEP-Stichprobe E designbedingt nicht sinnvoll durchzuführen. Die Frage ist vielmehr, inwieweit sich aus diesem Umstand Auswirkungen auf die Zusammensetzung der Stichprobe nach sozioökonomischen Kriterien ergeben.

Die drei weiteren Quartiersmerkmale, die für die Befragungshaushalte vorliegen, zielen unmittelbarer auf sozio-ökonomische Aspekte. Hier zeigen sich erfreulicherweise sehr geringe Strukturunterschiede zwischen realisierten Interviews und den Ausfällen:

- *Gebäudetyp*
  Haushalte, die in Ein- bis Zwei-Familienhäusern leben, sind in beiden Analysegruppen mit genau gleichem Anteil vertreten. Haushalte, die in verdichteten Wohngebieten leben (Hochhäuser, Blockbebauung usw.), sind in der realisierten Stichprobe nur geringfügig weniger vertreten als unter den Ausfällen.

- *Umgebung*
  Reine Wohngebiete im städtischen Bereich sind in der realisierten Nettostichprobe leicht unterrepräsentiert, ländliche Gebiete leicht überrepräsentiert.

- *Soziale Schicht der Bewohner des Viertels*
  Die Verteilungen, wie sie sich aus der Interviewereinschätzung ergeben, sind zwischen realisierten Interviews und Ausfällen fast deckungsgleich.

***Tabelle 8*** zeigt die entsprechenden Ergebnisse in bezug auf Haushaltsmerkmale. Die Analyse steht hier allerdings auf schwächeren Beinen, weil die Merkmale nur für einen bestimmten Anteil der ausgefallenen Haushalte ermittelbar waren:

| | |
|---|---|
| Haushaltsgröße | für 58% |
| Geschlecht des HV[12] | für 80% |
| Altersgruppe des HV | für 63% |
| Nationalität des HV | für 76% |
| Berufliche Stellung des HV | für 42% |

Die Nonresponse-Analyse muß notgedrungen diejenigen Ausfälle, für die eine Information vorliegt, als bestmögliche Schätzbasis für die Gesamtheit der Ausfälle hernehmen. Man macht also die - zugegebenermaßen unsichere - Annahme, daß die Ausfälle, für die keine Information vorliegt, in ihrer Struktur denjenigen gleichen, für die eine Information vorliegt. Die Zahlen in Tabelle 8 sind auf dieser Basis errechnet.

Die Ergebnisse zeigen insgesamt recht geringe Abweichungen in der Struktur der Ausfälle von der Struktur der Teilnehmerhaushalte:

- *Haushaltsgröße*
  Erwartungsgemäß sind die Ein-Personen-Haushalte in der realisierten Stichprobe leicht unterrepräsentiert. Größere Haushalte mit vier und mehr Personen sind demgegenüber verstärkt vertreten.

---

[12] HV = Haushaltsvorstand. Die genaue Formulierung der Fragen ist im ANHANG wiedergegeben.

- *Geschlecht und Alter des HV*
  Ältere Frauen sind, wie bekannt, eher weniger zur Teilnahme an Umfragen bereit. Dieser Effekt zeigt sich auch hier, jedoch in relativ schwacher Ausprägung.

- *Nationalität des HV*
  Unter den nichtteilnehmenden Haushalten sind 5%, unter den teilnehmenden Haushalten 4% Ausländerhaushalte.

- *Berufstätigkeit/berufliche Stellung des HV*
  Die Vergleichszahlen aus der realisierten Stichprobe konnten für diesen Bericht noch nicht ausgewertet werden. Die Zahlen für die Ausfälle weisen darauf hin - wenn man die Relation Arbeiter : Angestellte betrachtet -, daß Arbeiterhaushalte etwas verstärkt unter den nicht teilnahmebereiten Haushalten zu finden sind.

Zusammengenommen sprechen die Nonresponse-Informationen für eine relativ unverzerrte Abbildung der Grundgesamtheit durch die realisierte Stichprobe. Die „interne Validierung", wie sie mit Hilfe der Nonresponse-Informationen vorgenommen werden kann, muß allerdings noch durch eine externe Validierung, also den Vergleich mit Referenzstatistiken, ergänzt werden.

Tabelle 7:
**Nonresponse-Analyse: Quartiersmerkmale**

| | | *Interviews* % | *Ausfälle* % | *Gesamt* % |
|---|---|---|---|---|
| *Q1* | *Lage zur nächsten Großstadt* | | | |
| | In einer Großstadt: | | | |
| | - Stadtmitte (1-3)[1] | 14 | 8 | 11 |
| | - Stadtrand (4-5) | 27 | 35 | 31 |
| | Entfernung zur nächsten Großstadt: | | | |
| | - unter 25 km (6-7) | 23 | 22 | 22 |
| | - 25 b.u. 50 km (8) | 16 | 22 | 19 |
| | - 50 km und mehr (9) | 20 | 13 | 17 |
| | | 100 | 100 | 100 |
| *Q2* | *Gebäudetyp* | | | |
| | 1-2 Familienhaus (1-2) | 38 | 38 | 38 |
| | Mehrfamilienhaus (3) | 22 | 20 | 21 |
| | Reihenhaus u.ä. (7-8) | 20 | 19 | 19 |
| | Höhere Gebäude (4-5, 9-10) | 20 | 23 | 22 |
| | | 100 | 100 | 100 |
| *Q3* | *Umgebung* | | | |
| | Nur Wohngebäude (1) | 63 | 68 | 65 |
| | überwiegend Wohngebäude (2) | 23 | 23 | 23 |
| | überwiegend gewerblich (3) | 2 | 1 | 2 |
| | ländliche Bebauung (4) | 12 | 8 | 10 |
| | | 100 | 100 | 100 |
| *(4)* | *Soziale Schicht* | | | |
| | Unterschicht (1) | 4 | 5 | 4 |
| | Untere Mittelschicht (2) | 27 | 27 | 27 |
| | Mittlere Mittelschicht (3) | 63 | 59 | 61 |
| | Obere Mittelschicht (4) | 6 | 9 | 7 |
| | Oberschicht (5) | 0 | 0 | 0 |
| | | 100 | 100 | 100 |

1) Antwortkategorien in der Originalfrage. Vgl. ANHANG

Tabelle 8:
**Nonresponse-Analyse: Haushaltsmerkmale**

| | *Interviews* % | *Ausfälle*[1)] % | *Gesamt* % |
|---|---|---|---|
| *Haushaltsgröße* | | | |
| 1 Person | 28 | 36 | 32 |
| 2 Personen | 38 | 38 | 38 |
| 3 Personen | 16 | 15 | 15 |
| 4 und mehr Personen | 18 | 11 | 15 |
| | 100 | 100 | 100 |
| *Geschlecht des HV* | | | |
| Mann | 68 | 64 | 66 |
| Frau | 32 | 36 | 34 |
| | 100 | 100 | 100 |
| *Alter des HV* | | | |
| unter 29 Jahre | 8 | 8 | 8 |
| 30 - 49 Jahre | 39 | 37 | 38 |
| 50 - 69 Jahre | 38 | 33 | 36 |
| 70 Jahre u.älter | 15 | 22 | 18 |
| | 100 | 100 | 100 |
| *Nationalität des HV* | | | |
| deutsch | 96 | 95 | 96 |
| Ausländer | 4 | 5 | 4 |
| | 100 | 100 | 100 |
| *Berufstätigkeit des HV* | | | |
| Arbeiter | | 24 | |
| Angestellte | | 29 | |
| Beamte | | 2 | |
| Selbständige | | 5 | |
| Rentner | | 36 | |
| In Ausbildung/sonstige | | 4 | |
| | | 100 | |

1) Prozentuierung auf Fälle mit Angaben. Zur Höhe der fehlenden Angaben siehe Text.

## 4. Datenprüfung und übergebener Datenbestand

Haushalts- und Personenfragebogen für die Erstbefragung in Stichprobe E sind identisch mit denen für die Altstichproben A bis E. Dementsprechend sind auch die Programme der Datenprüfung und -bereinigung dieselben.[13]

Bei den CAPI-Interviews wird ein beträchtlicher Teil der Fehlermöglichkeiten, die Gegenstand der Datenprüfung bei herkömmlich geführten Interviews sind, bereits im Interview ausgeschlossen (Filterfehler, unzulässige Mehrfachnennungen, unzulässige Codes, numerische Werte außerhalb der als zulässig definierten Wertebereiche). Bei inhaltlich inplausiblen Angaben der Befragten wird teilweise bereits im Interview eine entsprechende Nachfrage eingeblendet, um den Sachverhalt zu klären.

Die Datenprüfung wird beim computerunterstützten Interviewen gegenüber dem herkömmlichen Verfahren im Arbeitsprozeß also nach vorne verlegt: in die Programmierung des CAPI-Instruments und in die Interviewsituation selbst. Eingriffe in die Daten *nach* Abschluß der Datenerhebung sollen dadurch überflüssig oder jedenfalls vermindert werden.

Im SOEP werden gleichwohl auch die CAPI-Daten ex post in derselben Weise geprüft wie die Daten der herkömmlich durchgeführten Interviews. Dazu werden die Daten aus der Datenstruktur des CAPI-Instruments zunächst in die Datenstruktur der übrigen Daten umgesetzt. In dieser Form durchlaufen sie denselben Prüfprozeß wie die übrigen Daten. Der Datensatz, wie er dem Datennutzer übergeben wird, ist für CAPI- und PAPI-Interviews identisch.

Für methodische Analysen enthält der Datensatz aber selbstverständlich eine Kennung der jeweils zugrundeliegenden Interviewmethode (Variablen HFORM und PFORM). Solche Analysen sollen nun, nachdem der Datenbestand verfügbar ist, vorgenommen werden. Sie können noch nicht Gegenstand dieses Berichts sein.

Der übergebene Datenbestand für die Erstbefragung von Stichprobe E umfaßt die folgenden Bestandteile:

(1) Nettodaten Haushaltsfragebogen (n = 1.079)
(2) Nettodaten Personenfragebogen (n = 1.957)

Beides jeweils in drei Versionen:
- Geprüfte Daten als ASCII-File
- Geprüfte Daten als SPSS-File
- Ungeprüfte Daten als ASCII-File

(3) Bruttodatensatz Haushalte (n = 2.052)
(4) Bruttodatensatz Personen (n = 2.496).

---

[13] Für eine genauere Darstellung siehe Infratest Burke Sozialforschung: Methodenbericht zum SOEP 1997. München 1998.

# ANHANG

# Anlage Q

# Quartiersbeschreibung

- Antwortcodes bitte in den Bogen „Adressenauflistung" eintragen.

## Angaben zu jeder Zieladresse

1. Liegt die Adresse in einer Großstadt von mehr als 100 Tsd. Einwohnern?

Ja ☐ ⇩

Wie weit ist die Adresse vom Geschäftszentrum der Innenstadt entfernt?

| | |
|---|---|
| Liegt im Geschäftszentrum | ❶ |
| Weniger als 1 km | ❷ |
| 1 - 2 km | ❸ |
| 3 - 5 km | ❹ |
| mehr als 5 km | ❺ |

Nein ☐ ⇩

Wie weit ist die Adresse vom Zentrum der nächstgelegenen Großstadt entfernt?

| | |
|---|---|
| Unter 10 km | ❻ |
| 10 - 25 km | ❼ |
| 25 - 50 km | ❽ |
| mehr als 50 km | ❾ |

2. Um was für ein Haus, das heißt welche Art von Wohnbebauung, handelt es sich bei dieser Adresse? Bitte schauen Sie sich die Rückseite mit den unterschiedlichen Gebäudetypen an und wählen Sie den Typ aus, der am ehesten zutrifft. →

Codes [ 1 - 10 ]

Nichts davon, sonstiges [ 11 ]

3. Besteht die unmittelbare Umgebung des Hauses ...

- nur aus Wohngebäuden? ..... ❶
- überwiegend aus Wohngebäuden, jedoch mit Läden/Kneipen/ Gewerbebetrieben? ..... ❷
- überwiegend aus gewerblich genutzten Gebäuden (Bürohäuser/Fabriken/ Geschäfte)? ..... ❸
- ländlicher Bebauung/Landwirtschaft? ..... ❹

4. Welcher sozialen Schicht würden Sie die Mehrzahl der Bewohner dieses Wohngebiets zurechnen?

| | |
|---|---|
| Unterschicht | ❶ |
| Untere Mittelschicht | ❷ |
| Mittlere Mittelschicht | ❸ |
| Obere Mittelschicht | ❹ |
| Oberschicht | ❺ |

1 - Villen

2 - freistehende Ein- oder Zweifamilienhäuser

3 - freistehende Mehrfamilienhäuser

4 - freistehende Gebäude, hohe und mittelgroße gemischt

5 - Wohnhochhäuser

6 - Terassenhäuser

7 - Reihenhäuser

8 - Zeilenbauweise, mehrgeschossig

9 - geschlossene Blockrandbebauung

10 - geschlossene Blockrandbebauung mit mehreren Höfen

## ANGABEN ZU DIESEM HAUSHALT (nur für Ausfälle)

Bitte versuchen Sie, möglichst vollständig folgende Angaben zu ermitteln oder zu schätzen:

1. Wie viele Personen leben in dem Haushalt?

   Personenzahl (einschl. Kinder) ☐

   *Keine Angabe möglich* ☐

2. Angaben zum Haushaltsvorstand (HV)

| | | HV |
|---|---|---|
| Geschlecht | Mann | ☐ |
| | Frau | ☐ |
| | *Keine Angabe möglich* | ☐ |
| Alter | Unter 16 Jahre | ☐ |
| | 16 - 29 Jahre | ☐ |
| | 30 - 49 Jahre | ☐ |
| | 50 - 69 Jahre | ☐ |
| | 70 Jahre u. älter | ☐ |
| | *Keine Angabe möglich* | ☐ |
| Nationalität | deutsch | ☐ |
| | Ausländer | ☐ |
| | *Keine Angabe möglich* | ☐ |
| | Deutsche Sprachkenntnisse: | |
| | gut | ☐ |
| | einigermaßen | ☐ |
| | schlecht | ☐ |
| Berufstätigkeit | Arbeiter | ☐ |
| | Angestellter | ☐ |
| | Beamter | ☐ |
| | Selbständiger | ☐ |
| | Schüler / Student / in Ausbildung | ☐ |
| | Rentner | ☐ |
| | Sonstiges | ☐ |
| | *Keine Angabe möglich* | ☐ |

Ich bestätige die Richtigkeit der gemachten Angaben:

| Datum | Abrechnungs-Nr. | Name des Interviewers | Unterschrift |
|---|---|---|---|