

Erprobung innovativer Erhebungskonzepte
für Haushalts-Panel-Stichproben

Erstbefragung 2000
der SOEP-Stichprobe F

Methodenbericht

Bernhard von Rosenblatt
Infratest Sozialforschung

München, Mai 2001
20.15090

Inhalt	Seite
1. Ziele und Anlage der Erhebung	3
2. Die Stichprobe	5
2.1 Statistische Anlage der Stichprobe	5
2.2 Adressenerhebung	9
2.3 Verfahren zur Repräsentation der ausländischen Wohnbevölkerung	10
2.4 Die realisierte Nettostichprobe	11
3. Durchführung der Erhebung	12
3.1 Zeitlicher Verlauf der Feldarbeit	12
3.2 Interviewereinsatz und Interviewmethoden	13
3.3 Stichprobenausschöpfung auf der Haushalts- und Personenebene	15
3.4 Nonresponse-Informationen	19
4. Datenprüfung und übergebener Datenbestand	23

1. Ziele und Anlage der Erhebung

Das Forschungsvorhaben „Erprobung innovativer Erhebungskonzepte für Haushalts-Panel-Studien“ ist ein Projekt des Deutschen Instituts für Wirtschaftsforschung (DIW), Berlin, unter Leitung von Prof. Dr. Gert Wagner, finanziert vom Bundesministerium für Bildung und Forschung. Das Vorhaben ist eng verknüpft mit dem Sozio-ökonomischen Panel (SOEP), das ebenfalls am DIW unter Leitung von Professor Wagner durchgeführt wird. Die Datenerhebung für beide Projekte ist im Unterauftrag an Infratest Sozialforschung, München, vergeben.

Ziel des Projekts ist es, ein Haushaltspanel nach dem Modell des SOEP aufzubauen und in diesem Rahmen bestimmte innovative Methoden erproben zu können. Die Projektarbeiten begannen im Oktober 1999. Im Jahr 2000 war die erste Befragung durchzuführen. Vorgesehen sind zunächst zwei Folgewellen der Befragung in den Jahren 2001 und 2002.

Es werden dieselben Erhebungsinstrumente verwendet wie in den parallel durchgeführten SOEP-Befragungen. Der entstehende Datenbestand kann mit dem des SOEP zu einem Gesamtdatenbestand integriert werden. Das neue Haushaltspanel wird daher auch als „Stichprobe F“ des SOEP bezeichnet.

Die Stichproben A – D des SOEP sind repräsentative Stichproben bestimmter Teilgruppen der Wohnbevölkerung Deutschlands:

- A Deutsche Haushalte West (Erstbefragung 1984)
- B Ausländerhaushalte West (Erstbefragung 1984)
- C Deutsche Haushalte Ost (Erstbefragung 1990)
- D Zuwanderer nach 1984 (Erstbefragung 1994/95)

Im Rahmen eines gemeinsamen Hochrechnungsrahmens bilden diese Teilstichproben zusammen die gesamte Wohnbevölkerung ab. Mit der Stichprobe E wurde dann erstmals eine einheitliche Querschnittsstichprobe der Wohnbevölkerung angestrebt:

- E Querschnitt '98 (Erstbefragung 1998)

Die neue **Stichprobe F** (Erstbefragung 2000) folgt dem Modell der Stichprobe E. Sie ist allerdings vom Umfang her erheblich größer: Ziel ist eine Haushaltsstichprobe von mindestens 6.000 teilnehmenden Haushalten. Außerdem wird mit der neuen Stichprobe ein innovatives Verfahren erprobt, um innerhalb eines einheitlichen Stichprobenansatzes die ausländische Wohnbevölkerung besser zu repräsentieren.

Der hier vorgelegte Methodenbericht konzentriert sich auf die Darstellung spezifischer Aspekte des Aufbaus einer neuen Haushalts-Panelstudie. Die methodischen Verfahren eines Haushaltspanels im Allgemeinen und des deutschen Sozio-ökonomischen Panels im Besonderen werden als bekannt vorausgesetzt, bzw. es wird auf die Literatur zum SOEP und die vorliegenden jährlichen Methodenberichte von Infratest zum SOEP verwiesen.

Das Jahr 2000 war das 17. Jahr des Sozio-ökonomischen Panels in Deutschland. Der Methodenbericht zu den Längsschnittstichproben A – E, die nunmehr auch „Altstichproben“ genannt werden, liegt vor.¹ Das Fragenprogramm und die Weiterentwicklung der Interviewmethoden im Erhebungsjahr 2000 sind dort beschrieben. Von einer erneuten Darstellung wird hier abgesehen.

Den Befragungspersonen gegenüber werden alle Stichproben des Sozio-ökonomischen Panels einheitlich unter dem Projekttitel LEBEN IN DEUTSCHLAND präsentiert. Das gilt auch für die neue Stichprobe F. Da es sich hierbei im Jahr 2000 aber um eine Erstbefragung handelt, unterscheidet sich das Erhebungsinstrumentarium teilweise von dem der Längsschnittbefragungen in den Stichproben A – E. Die Unterschiede sind minimal bei den eigentlichen Fragebogen, aber größer bei den begleitenden Materialien wie Intervieweranleitung, Anschreiben und Broschüren, Unterlagen zur Adressenerhebung usw. Alle Erhebungsinstrumente und Materialien für die neue Panelstichprobe wurden daher in einem eigenen **Anlagenband** zu diesem Methodenbericht zusammengestellt.

¹ Infratest Sozialforschung: SOEP 2000. Das Sozio-ökonomische Panel im Befragungsjahr 2000, Methodenbericht, München 2001.

2. Die Stichprobe

2.1 Statistische Anlage der Stichprobe

1. Grundgesamtheit:

Die Grundgesamtheit der Untersuchung besteht aus allen Privat-Haushalten in der Bundesrepublik Deutschland. (Lediglich die bewohnten Inseln ohne Zugang über einen PKW-Fahrdamm blieben ausgenommen).

2. Auswahlverfahren

Die Auswahl der Privathaushalte erfolgte über eine mehrfach geschichtete, mehrstufige Zufalls-Stichprobe auf der Basis des ADM-Stichprobensystems.

3. Schichtung

Dazu wurden sämtliche Gemeinden des Untersuchungsgebiets (ca. 14.600 mit Stand vom 01.01.1999) nach regionalen Kriterien und nach ihrer Größe in Schichten unterteilt. Für die Schichtung wurde der Stand der regionalen Abgrenzung vom 30. 6. 1996 zu Grunde gelegt. Als Schichtungsmerkmale werden

- Bundesländer
- Regierungsbezirke
- Kreise bzw. kreisfreie Städte
- Gemeindetypen

verwendet. Kriterium der Größe ist nicht die politische Gemeindegröße, sondern die Differenzierung der Gemeinden nach BIK-Typen, die zusätzlich zur Größe der Gemeinden deren Zentralisationsgrad einschließt. Die Zuordnung der Gemeinde erfolgte gemäß dem Agglomerationskonzept von BIK in folgender Weise:

Gemeindetypen (entsprechend den BIK-Stadtregionen)

0=	500.000 und mehr	Einwohner	(Kernbereich Zone 1)
1=	500.000 und mehr	Einwohner	(Randbereich Zone 2,3)
2=	100.000 bis unter 500.000	Einwohner	(Kernbereich Zone 1)
3=	100.000 bis unter 500.000	Einwohner	(Randbereich Zone 2,3)
4=	50.000 bis unter 100.000	Einwohner...	(Kernbereich Zone 1)
5=	50.000 bis unter 100.000	Einwohner	(Randbereich Zone 2,3)
6=	20.000 bis unter 50.000	Einwohner	
7=	5.000 bis unter 20.000	Einwohner	
8=	2.000 bis unter 5.000	Einwohner	
9=	unter 2.000	Einwohner	

- Zone 1 Kernbereich der Stadtregion
- Zone 2 Verdichtungsbereich der Stadtregion
- Zone 3 Übergangsbereich der Stadtregion

Die Größenordnung bezieht sich hierbei auf die Einwohnerzahl der jeweiligen Stadtregion (Summe Zonen 1-3). Bei Gemeinden, die keiner Stadtregion angehören, entspricht die Größenklasse der Einwohnerzahl in den politischen Grenzen (Typ 6 – 9). Unter Randbereich wird die Zusammenfassung des Verdichtungs- und des Übergangsbereichs verstanden, also die Zonen 2 und 3.

Ordnet man dem Bundesland Berlin die zwei historisch gewachsene "Quasi"-Regierungsbezirke zu, also den ehemaligen West- und den ehemaligen Ostteil, so ergeben sich insgesamt für das Gebiet der Bundesrepublik Deutschland 441 Kreiseinheiten, das sind 323 Landkreise und 118 kreisfreie Städte. Demzufolge führt das theoretische Schichttableau zu 441 Kreiseinheiten x 10 BIK-Typen, insgesamt also zu 4410 Schichten. Davon sind 1510 tatsächlich besetzt, d.h. in denen existiert mindestens eine Gemeinde.

4. Auswahlstufen

Alle Gemeinden eines bestimmten BIK-Typs in einem Kreis bilden eine der 1510 Schichten bzw. Zellen. Das Bedeutungsgewicht jeweils einer Zelle ist aus der Gesamtzahl der Privat-Haushalte in ihr definiert, jeweils mit dem Aktualitätsstand der BIK-Datei vom März 1997. Für die Stichprobe nun wurden mit Wahrscheinlichkeiten proportional zu den Bedeutungsgewichten 985 Sample Points so auf das Allokationstableau verteilt, dass auch für alle höheren Aggregationsniveaus proportionale Verteilungen möglich sind. Zusätzlich ergab sich ein Schichtungseffekt noch durch Sortierung der Auswahlseinheiten nach Kreisen und Gemeinden innerhalb jeder Schicht.

Danach wurden die so entstandenen gebrochenen Werte für die Sample-Points auf ganze Zahlen gerundet, wobei die Rundungen mehrstufig von "außen" nach "innen" erfolgt sind, d.h. von der höheren Aggregationsebenen Bundesland x BIK-Typ bis hin zu der niedrigsten Ebene Kreis x BIK-Typ.

Die Einheiten der ersten Auswahlstufe sind die (ggf. synthetisierten) Stimmbezirke zur Bundestagswahl 1994. Diese Auswahlstufe wurde pro Schicht als Intervallziehung mit Zufallsstart in zwei Schritten realisiert - als Auswahl der Gemeinden in Abhängigkeit von der Anzahl der Sample-Points und der verfügbaren Gemeinden in der jeweiligen Zelle und danach die Auswahl der Stimmbezirke innerhalb dieser Gemeinden.

Die Einheiten der zweiten Auswahlstufe sind die Privat-Haushalte innerhalb des jeweiligen Stimmbezirks. Ausgehend von der im Sample-Point liegenden und zufällig aus dem Straßenband ausgewählten Startadresse (die selbst nicht befragt werden darf), wurde ein Weg – nach Vorgaben formal beschrieben, der eine vom Interviewer unbeeinflusste Adressenauswahl gewährleistet. Die wichtigsten dieser Vorgaben lauten (jeweils bei der Startadresse beginnend):

- in absteigender Hausnummernfolge auf der gleichen Straßenseite
- in Mehrfamilienhäusern von oben nach unten, von vorne nach hinten.

- Im gleichen Stockwerk von rechts nach links

Die Interviewer hatten auf dem jeweiligen, so vorgegebenen Weg jeden dritten Privat-Haushalt jeweils in ihrer Adressenliste aufgelistet, bis die vorgegebene Zahl von 2*12 Adressen erreicht war. Mit diesen Haushalten haben die Interviewer Kontakt aufzunehmen, um sie für die Teilnahme an der Befragung zu gewinnen.

5. Vergleich Sample-Points – Sollvorgaben:

Insgesamt wurden für die alten Bundesländer einschl. des ehemaligen Westteils von Berlin 803 Sample-Points und für die neuen Bundesländer einschl. des ehemaligen Ostteils von Berlin 182 Sample-Points definiert. Das entspricht einem prozentualen West-Ost-Verhältnis von 81,5 zu 18,5.

	Anzahl der Privat-HH Lt. BIK-Datei		Sample- Points	
Alte Bundesländer u. Berlin-West				
Schleswig-Holstein	1.225.188	4,1%	34	4,2%
Hamburg	864.907	2,9%	24	3,0%
Niedersachsen	3.394.334	11,5%	92	11,5%
Bremen	340.010	1,2%	9	1,1%
Nordrhein-Westfalen	7.948.445	26,9%	216	26,9%
Hessen	2.640.754	8,9%	70	8,7%
Rheinland-Pfalz	1.719.069	5,8%	46	5,7%
Baden-Württemberg	4.597.483	15,6%	126	15,7%
Bayern	5.170.767	17,5%	141	17,6%
Saarland	498.562	1,7%	14	1,7%
Berlin ehemal. Westteil	1.133.931	3,8%	31	3,9%
Summe	29.533.450	100,0%	803	100,0%
Neue Bundesländer u. Berlin-Ost				
Berlin ehemal. Ostteil	631.646	9,4%	17	9,3%
Brandenburg	1.068.082	16,0%	30	16,5%
Mecklenburg-Vorpommern	747.735	11,2%	20	11,0%
Sachsen	2.002.725	29,9%	54	29,7%
Sachsen-Anhalt	1.179.927	17,6%	32	17,6%
Thüringen	1.061.505	15,9%	29	15,9%
Summe	6.691.620	100,0%	182	100,0%
BRD gesamt	36.225.070		985	

	Anzahl der Privat-HH Lt. BIK-Datei		Sample- Points	
Alte Bundesländer u. Berlin-West				
BIK 0	10.344.696	35,0%	282	35,1%
BIK 1	3.038.115	10,3%	82	10,2%
BIK2	3.181.440	10,8%	87	10,8%
BIK 3	1.868.197	6,3%	51	6,4%
BIK 4	997.308	3,4%	28	3,5%
BIK 5	124.163	0,4%	3	0,4%
BIK 6	2.722.968	9,2%	74	9,2%
BIK 7	4.298.632	14,6%	116	14,4%
BIK 8	1.699.527	5,8%	46	5,7%
BIK 9	1.258.404	4,3%	34	4,2%
Summe	29.533.450	100,0%	803	100,0%
Neue Länder u. Berlin-Ost				
BIK 0	1.627.339	24,3%	44	24,2%
BIK 1	271421	4,1%	8	4,4%
BIK 2	1.009.176	15,1%	27	14,8%
BIK 3	304.634	4,6%	9	4,9%
BIK 4	452.078	6,8%	12	6,6%
BIK 5	141.915	2,1%	4	2,2%
BIK 6	678.383	10,1%	18	9,9%
BIK 7	809.712	12,1%	22	12,1%
BIK 8	555.680	8,3%	15	8,2%
BIK 9	841.282	12,6%	23	12,6%
Summe	6.691.620	100,0%	182	100,0%
BRD gesamt	36.225.070		985	

2.2 Adressenerhebung

Das beschriebene Auswahlverfahren folgt dem Modell des „Random-Route“, das in Deutschland für mündlich-persönliche Befragungen von Zufallsstichproben der Bevölkerung das etablierte Verfahren ist. Innerhalb dieses Modells gibt es allerdings eine Bandbreite von Verfahrensweisen im Detail, die – je nach Positionierung am oberen oder unteren Rand – mit erheblichen Qualitäts- und Kostenwirkungen verbunden ist.

Für den Aufbau der neuen Panelstichprobe im Rahmen des SOEP-Stichprobensystems wurde eine Ausgestaltung des Random-Route-Modells gewählt, die deutlich höhere Qualitätsstandards ermöglicht als das sonst übliche Normalmodell. Kernpunkt ist der „getrennte Adressenvorlauf“. Während die Auswahl der Befragungshaushalte normalerweise in einem Arbeitsgang mit der Durchführung der Interviews erfolgt, wurden für die SOEP-Stichprobe F – ebenso wie 1998 schon für die Stichprobe E – beide Arbeitsvorgänge getrennt. Dies hat folgende methodische Vorteile:

- Für die eigentliche Befragung werden dem Interviewer feste Adressen vorgegeben. Die Abläufe und Kontrollmöglichkeiten entsprechen praktisch denjenigen einer Adressen-Random-Stichprobe, die beispielsweise mit Adressen aus Einwohnermelderegistern arbeitet.²
- Die Qualität der Adressen – Lage auf dem Begehungsweg, richtige Namen, korrekte Schreibweise – kann teilweise vorab geprüft werden und dann in der Befragungssituation noch einmal gegengecheckt werden. Gerade für eine Panelbefragung – im Hinblick auf die erneute Befragung im Folgejahr, Panelpflege, Adressenrecherche bei Umzügen – ist die Korrektheit der verwendeten Adressen wichtig.
- Die Adressenerhebung kann mit der Erhebung zusätzlicher Informationen verknüpft werden. Im vorliegenden Fall war das die „interviewergestützte Quartiersbeschreibung“ nach dem Konzept von Hoffmeyer-Zlotnik, wie sie schon 1998 für die Stichprobe E eingesetzt wurde. Dies eröffnet verbesserte Möglichkeiten einer Nonresponse-Analyse, also eines Strukturvergleichs von Teilnehmern und Nicht-Teilnehmern an der Befragung als Basis einer „internen Validierung“ der Stichprobe.
- Die Vorab-Fixierung der eingesetzten Adressen ermöglicht eine klare Beschreibung der Bruttostichprobe und der Bearbeitung jeder einzelnen Adresse im Feldverlauf. Infratest liefert diese Informationen in einem eigenen Datenbestand, dem sogenannten „Bruttoband“, ergänzend zu den eigentlichen Befragungsdaten, dem sogenannten „Nettoband“.

Die Adressenerhebung für die SOEP-Stichprobe F erfolgte im Januar/Februar 2000. Für alle 985 Sample-Points der Stichprobe wurde ein in der jeweiligen Region lebender Interviewer angeschrieben und mit der Adressenerhebung nach fest vorgegebenen Regeln beauftragt. Der Erhebungsbogen ist im **Anlagenband** wiedergegeben.

² Für ein Haushaltspanel ist eine Einwohnermeldeamtsstichprobe nicht geeignet, weil hier eine Haushaltsstichprobe benötigt wird, während aus Einwohnermelderegistern Personen-Adressen gezogen werden.

Die Interviewer hatten 24 Adressen pro Sample-Point aufzulisten. (In welcher Weise diese 24 Adressen später in die Bruttostichprobe eingehen, ist in Kapitel 3.3 näher beschrieben.) Die eingesetzten Interviewer waren zum Teil SOEP-Interviewer, die später möglicherweise denselben Point zu bearbeiten hatten, zum Teil andere Interviewer.

Der Gesamtbestand an Adressen, die auf diese Weise erhoben wurden, betrug 23.640 Adressen. Diese waren zusammen mit den Quartiersmerkmalen zu prüfen und zu erfassen. Auf dieser Grundlage wurden für jeden eingesetzten Interviewer Adressenlisten der Haushalte erstellt, die in der Feldphase dann zu bearbeiten waren.

2.3 Verfahren zur Repräsentation der ausländischen Wohnbevölkerung

Im Sozio-ökonomischen Panel wird auf die Repräsentation der ausländischen Wohnbevölkerung besonderer Wert gelegt. Zwei der bisherigen SOEP-Stichproben sind Sonderstichproben von Zuwanderergruppen (Stichproben B und D). Quantitativ sind die Zuwanderer (d.h. Ausländer und Aussiedler) im Gesamtsystem der SOEP-Stichproben damit sogar stärker repräsentiert als ihrem Bevölkerungsanteil entspricht.

In der neuen Panelstichprobe F soll die zugewanderte Wohnbevölkerung proportional – entsprechend ihrem Anteil und ihrer Verteilung in der Wohnbevölkerung – repräsentiert werden. Unter den in Deutschland gegebenen Bedingungen ist dies eine schwierig zu realisierende Zielvorgabe.

Das ADM-Stichprobensystem, wie es oben in Kapitel 2.1 beschrieben wurde, benutzt als kleinräumige Auswahlinheit die ca. 80.000 Stimmbezirke der Bundestagswahl. Da die Wählerverzeichnisse sich nur auf die wahlberechtigte, deutsche Wohnbevölkerung beziehen, gehen Wohnbezirke mit hohem Ausländeranteil mit einem zu geringen Gewicht in die Stichprobenbildung ein.

Im Ergebnis ist die ausländische Wohnbevölkerung selbst dann, wenn sie in die Befragung ausdrücklich einbezogen ist, in den Bevölkerungsstichproben nach dem ADM-System unterrepräsentiert. So hat sie etwa in der SOEP-Stichprobe E von 1998 einen Anteil von 4%, das ist etwa die Hälfte ihres Anteils an der Grundgesamtheit nach der amtlichen Statistik.³

Für die neue Stichprobe F hat Infratest daher ein Verfahren vorgeschlagen, das den Anteil der Ausländerbevölkerung in der Stichprobe verdoppeln, also etwa auf den proportionalen Bevölkerungsanteil von ca. 8% bringen soll. Man muss zunächst sagen „soll“, weil das Verfahren bisher nirgends angewendet und hier erstmals erprobt wurde.

³ Hauptgrund dafür sind nicht Ausfälle in der Befragung, etwa aufgrund von Sprachschwierigkeiten. Nach den verfügbaren Nonresponse-Informationen betrug der Ausländeranteil unter den Ausfällen ebenfalls nur 5%.

Vgl. Infratest Sozialforschung: SOEP '98 Erstbefragung der Stichprobe E, Methodenbericht, S. 32.

Das Verfahren besteht darin, den „Fehler“ des ADM-Stichprobensystems auf der *ersten* Auswahlstufe – das ist die Auswahl der Sample-Points – auf der *zweiten* Auswahlstufe zu korrigieren – das ist die Auswahl der Haushalte auf dem Begehungsweg im Sample-Point. Konkret wird die Länge des Begehungsweges für die Identifizierung von Ausländer-Haushalten verdoppelt. Damit wird die Wahrscheinlichkeit, auf Ausländer-Haushalte zu treffen, theoretisch um den Faktor 2 erhöht.

Zu beachten ist, dass damit die strukturelle Verzerrung in der Auswahl der Sample-Points nicht behoben wird. Ausländer, die in Wohnbezirken mit hohem Ausländeranteil leben, haben eine geringere Auswahlwahrscheinlichkeit als Ausländer, die in Wohnbezirken mit niedrigerem Ausländeranteil leben. Tendenziell dürfte dies in Richtung auf eine erhöhte Repräsentation von relativ besser integrierten Ausländern wirken.

Das realisierte Verfahren ist also keine perfekte Lösung des Problems, sondern eine pragmatische *second-best-Lösung*. Allerdings wird man mit der Umfragemethodik ohnehin nie ganz vermeiden können, dass nicht integrierte Ausländergruppen (z.B. Illegale oder in ethnischen Ghettos lebende Ausländer) unterrepräsentiert werden. Dies darf aber kein Argument dagegen sein, den über Umfragen erreichbaren Teil der Ausländer-Bevölkerung in Repräsentativstichproben der Wohnbevölkerung so gut wie möglich einzubeziehen.

2.4 Die realisierte Nettostichprobe

In den folgenden Kapiteln dieses Berichts werden Durchführung und Erfolg der Erhebung unter verschiedenen methodischen Gesichtspunkten näher beschrieben. Vorab soll kurz das Ergebnis dargestellt werden, also die letztlich realisierte Nettostichprobe in der ersten Befragungswelle dieses Haushaltspanels.

Die realisierte Stichprobe umfasst **6.060 Haushalte**. Damit wurde das vorgegebene Ziel erreicht, innerhalb des Stichprobenrahmens eine Zahl von mindestens 6.000 Haushalten für die Teilnahme zu gewinnen.

Unter den 6.060 Haushalten sind 445 Haushalte, in denen Personen ausländischer Staatsangehörigkeit leben. Das ist ein Anteil von 7,3%. Der angestrebte Anteil von 8% wurde damit zwar nicht ganz, aber doch näherungsweise erreicht.

In den 6.060 Teilnehmerhaushalten leben 11.543 Personen im Befragungsalter ab 16 Jahren. Jede dieser Personen sollte mit einem eigenen *Personenfragebogen* individuell befragt werden. Dies gelang mit einer Erfolgsquote von immerhin 94,4%. Auf Personenebene umfasst damit der Datenbestand der Erstbefragung **10.901 Interviews**.

Der Datenbestand der Gesamtheit der SOEP-Stichproben wurde damit nahezu verdoppelt. Unter Einschluss der neuen Panelstichprobe F umfasst der Gesamtdatenbestand des Jahres 2000 nunmehr **13.266 Haushalte** sowie **24.597 Personen**.

3. Durchführung der Erhebung

3.1 Zeitlicher Verlauf der Feldarbeit

Die Erstbefragung in der neuen Panelstichprobe F erfolgte zusätzlich zur Befragung 2000 in den SOEP-Altstichproben A – E. Dies stellte eine erhebliche Anforderung an die Interviewerkapazität und Einsatzplanung der Infratest-Feldorganisation dar.

Die Feldarbeit der beiden Projekte fand daher nicht zeitgleich, sondern zeitlich etwas versetzt statt. Die Befragung in Stichprobe F konnte – nach der Adressenerhebung im Januar/Februar – nicht vor Ende März beginnen. Um freie Interviewerkapazität dafür zu gewinnen, sollte die Befragung der Altstichproben bis dahin möglichst weit vorangetrieben sein, was hier eine erhebliche Beschleunigung der Abläufe gegenüber den Vorjahren erforderte. Diese Anstrengung war erfolgreich: Ende März waren rd. 6.000 Haushalte der Altstichproben fertig bearbeitet, die restlichen 1.200 verteilten sich auf die Monate April bis Oktober (siehe **Tabelle 1**).

Tabelle 1:
Monat des letzten Haushaltskontaktes bei Haushalten mit realisierten Interviews

Anzahl Haushalte nach Monate	Altstichproben A – E	Neue Stichprobe F
Januar	3.189	-
Februar	1.979	-
März	904	269
April	484	1.483
Mai	227	862
Juni	136	1.380
Juli	113	816
August	48	597
September	35	419
Oktober	<u>91</u>	<u>234</u>
Gesamt	7.206	6.060

Der größte Teil der Interviews in der neuen Stichprobe F wurde in den Monaten April bis Juni durchgeführt. Der Bearbeitungserfolg bis zu diesem Zeitpunkt blieb allerdings hinter den Erwartungen zurück. Nur mit sehr intensiven Nachbearbeitungsmaßnahmen in den drei Folgemonaten – und den entsprechend hohen Fallzahlen bearbeiteter Haushalte in dieser Phase der Feldarbeit – konnte das angestrebte Ziel von mindestens 6.000 Teilnehmerhaushalten bis Ende Oktober erreicht werden.

Die Feldzeit für die Erstbefragung in der neuen Stichprobe F betrug damit insgesamt 7,5 Monate. Das ist ein sehr langer Zeitraum, der auf die Schwierigkeiten verweist, eine Bevölkerungsumfrage dieses Umfangs – auf Random-Basis und mit den besonderen Anforderungen des SOEP – heute zu realisieren.

Ein weiterer Hinweis mag die erforderlichen Anstrengungen verdeutlichen. Um die geforderte Zahl von Haushalts- und Personeninterviews zu erreichen, waren ca. 38.000 Kontakte der Interviewer zu den Stichprobenhaushalten erforderlich.

- 20.282 Kontakte waren erforderlich für die 6.060 Haushalte, die schließlich erfolgreich bearbeitet werden konnten; das sind im Durchschnitt 3,3 Interviewerkontakte pro Haushalt.
- 17.759 Kontakte sind dokumentiert bei denjenigen Haushalten, die letztlich *nicht* erfolgreich bearbeitet werden konnten – sei es, weil niemand angetroffen wurde oder die Haushaltsmitglieder sich nicht zur Teilnahme gewinnen ließen oder weil die Adresse sich als Fehladresse erwies (z.B. unbewohnt, Geschäftsadresse usw.). Im Durchschnitt sind das 2,8 Kontakte bzw. Kontaktversuche bis zur (negativen) Klärung des Falles. Es können aber auch 9 und mehr Kontakte bzw. Kontaktversuche bis zur Feststellung eines Misserfolgs erforderlich sein – dies ist für immerhin 275 Haushalte dokumentiert.
- Nicht gezählt sind hier die erforderlichen Haushaltskontakte zur Identifizierung von Ausländer-Haushalten auf dem verlängerten Random-Walk (Adressen 13 – 24). Ein Adressenprotokoll mit einer Dokumentation der Kontakte wurde in diesem Teil der Bearbeitung nur in solchen Fällen angelegt, in denen ein Haushalt als „zur Zielgruppe Ausländer gehörend“ identifiziert wurde. Um dies festzustellen, mussten aber rd. 11.500 Haushalte kontaktiert werden!

3.2 Interviewereinsatz und Interviewmethoden

Die Befragungen im SOEP erfolgen durch Infratest-Interviewer, die in der Befragungsregion des jeweiligen Panelhaushalts leben. Bei den Längsschnittbefragungen der Altstichproben A – E spielt ergänzend mittlerweile die zentrale Bearbeitung von Panelhaushalten eine erhebliche Rolle, die über telefonische Kontaktinterviews und schriftlich-postalische Befragung erfolgt (vgl. Methodenbericht 2000, Kapitel 3.4). Bei einer Erstbefragung neuer Panelhaushalte ist jedoch der persönliche Interviewereinsatz unverzichtbar.

Im Jahr 2000 wurden für die SOEP-Altstichproben 429 Interviewer eingesetzt, die mit wenigen Ausnahmen über mehrjährige SOEP-Erfahrung verfügten. Für die neue SOEP-Stichprobe F mit ihren 985 Sample-Points musste eine größere Zahl von Interviewern eingesetzt werden:

Insgesamt eingesetzte Interviewer	616
darunter:	
- SOEP-erfahrene Interviewer, die auch in den Altstichproben eingesetzt waren	227
- Interviewer ohne SOEP-Erfahrung	389

Die neue Stichprobe des Haushalts-Panels konnte also nicht etwa von den erfahrenen SOEP-Interviewern einfach „mit erledigt“ werden. Erforderlich war vielmehr die Einarbeitung einer großen Zahl zusätzlicher Interviewer. Diese waren zwar alle in Face-to-Face-Bevölkerungsumfra-

gen geschult und erfahren, aber nicht mit den spezifischen Anforderungen des SOEP vertraut. Die projektspezifische Einarbeitung erfolgte anhand des umfangreichen „Interviewerhandbuchs LEBEN IN DEUTSCHLAND“ und einer Hotline zu Projektmitarbeitern bei Infratest in der Startphase der Feldarbeit.

Eine Besonderheit des Befragungsjahres 2000 im SOEP war es, dass der Einsatz der CAPI-Technik (CAPI = Computer Assisted Personal Interview) erstmals in allen SOEP-Stichproben freigegeben war. (Für nähere Ausführungen hierzu verweisen wir auf Kapitel 3.5 im Infratest-Methodenbericht 2000.) Der Einsatz computerunterstützter Interviewmethoden gehört zu den „innovativen Verfahren“, die in der neuen Panelstichprobe besonderen Stellenwert haben sollen. Es war daher vorgesehen, von Anfang an einen substantziellen Teil der Stichprobe mit CAPI zu bearbeiten.

Von den insgesamt 616 eingesetzten Interviewern waren etwa die Hälfte sogenannte Laptop-Interviewer, die für CAPI-Interviews ausgerüstet und geschult sind. Von diesen erhielten 241 (39% der eingesetzten Interviewer) die Erhebungsunterlagen per Modem als CAPI-Instrument. Ebenso wie in den Altstichproben war die CAPI-Nutzung jedoch nicht verpflichtend vorgeschrieben. Jeder Laptop-Interviewer erhielt die Befragungsinstrumente (HAUSHALTS-Fragebogen und PERSONEN-Fragebogen) zusätzlich als Papierfragebogen.⁴

Insgesamt haben die Laptop-Interviewer einen größeren Teil der neuen Panelstichprobe realisiert als die zahlenmäßig stärkere Gruppe der übrigen Interviewer. Dabei haben sie aber keineswegs alle Interviews wirklich mit CAPI durchgeführt:

	Laptop-Interviewer	übrige Interviewer	Gesamt
Eingesetzte Interviewer	241	378	616
Erfolgreich bearbeitete Haushalte	3.662	2.366	6.028 ⁵
darunter:			
CAPI-Interviews	1.916	-	1.916
CAPI-Anteil	52%	-	32%
Realisierte Interviews pro Interviewer	15,2	6,3	9,8

Von allen Haushalten, die erfolgreich durch Laptop-Interviewer bearbeitet wurden, ist etwa die Hälfte unter Nutzung der CAPI-Technik interviewt worden (52%). Das ist ein niedrigerer Wert als in den Altstichproben A – E, wo der entsprechende Anteil bei 68% liegt. In der Situation der Anwerbung und Erstbefragung für das neue Haushalts-Panel haben viele Interviewer offenbar die Arbeit mit Papierfragebogen vorgezogen.

⁴ Die Zusatzfragebogen kommen erst ab Welle 2 der Panelbefragung zum Einsatz.

⁵ Weitere 32 Interviews wurden im Zuge der Nachbearbeitungsmaßnahmen schriftlich-postalisch realisiert.

Dabei wurde von vornherein auch die Möglichkeit genutzt, zumindest die individuellen PERSONEN-Fragebogen von den Befragungspersonen selbst ausfüllen zu lassen. Für insgesamt 26% der vorliegenden PERSONEN-Fragebogen wurde diese Methode genutzt – im Einverständnis von Interviewern und Befragten, die das Selbstaussfüllen offenbar als die angenehmste Organisationsform der Befragung empfanden.

Der CAPI-Anteil an der Gesamtheit der realisierten Interviews beträgt 32%. Er ist in der neuen Stichprobe F damit höher als in den Altstichproben A – E, wo er im Jahr 2000 bei 21% liegt.

Die Laptop-Interviewer stellen den aktiven Kern des Infratest-Interviewerstabes. Dies schlägt sich darin nieder, dass in dieser Befragung jeder Laptop-Interviewer im Durchschnitt 15,2 Haushalte erfolgreich bearbeitet hat, während der Bearbeitungsumfang bei den übrigen Interviewern mit durchschnittlich 6,3 Interviews deutlich geringer ist.

Insgesamt hat in der neuen Panelstichprobe F jeder Interviewer im Durchschnitt 9,8 Haushalte erfolgreich bearbeitet. Der Vergleichswert für die SOEP-Altstichproben A – E liegt bei 14,0 Haushalten pro Interviewer.

3.3 Stichprobenausschöpfung auf der Haushalts- und Personenebene

Die Bruttostichprobe besteht auf der ersten Auswahlstufe, wie oben beschrieben (Kapitel 2.1), aus 985 Sample-Points. Auf der zweiten Auswahlstufe erfolgt über den Random-Walk die Auswahl der Befragungshaushalte (vgl. Kapitel 2.2 Adressenerhebung). Auf dieser Ebene gibt es eine Aufteilung in zwei Teilstichproben:

Hauptstichprobe

Diese umfasst für alle 985 Sample-Points die Adressen 1 bis 12 des Begehungsweges, insgesamt also 11.820 Haushalte, die zu befragen sind.

Ergänzungstichprobe

Diese umfasst ausgewählte Haushalte auf dem verlängerten Begehungsweg (Adressen 13 bis 24), und zwar

1. Alle Haushalte mit mindestens einer Person nicht-deutscher Staatsangehörigkeit
2. Ersatzadressen für qualitätsneutrale Ausfälle (QNA) in der Hauptstichprobe. QNA sind Haushaltsadressen, die sich als unbewohnt oder als Geschäftsadressen erweisen. Für jeden dokumentierten QNA-Fall kann der Interviewer eine Ersatzadresse verwenden, beginnend bei Adresse 13 und weiter in aufsteigender Reihenfolge.

Während die Bruttofallzahl in der Hauptstichprobe vorab feststeht, ergibt sich die Zahl gültiger Bruttohaushalte aus der Ergänzungsstichprobe erst im Verlauf der Feldarbeit. Die Ergebnisse sind folgende:

Ausländer-Haushalte

Auf dem verlängerten Random-Walk wurden von den Interviewern 423 Ausländer-Haushalte identifiziert. Das ist ein Anteil von 3,6% der zu kontaktierenden 11.820 Adressen.

Wie hoch die entsprechende Zahl in der Hauptstichprobe ist, lässt sich nicht genau feststellen, weil bei den Ausfällen eine entsprechende Information über die Nationalität der Haushaltsmitglieder nicht oder nur teilweise vorliegt. Man kann jedoch bei den *realisierten* Interviews mit Ausländer-Haushalten unterscheiden, wie viele jeweils aus den zwei Teilstichproben kommen:

aus der Hauptstichprobe	241 Haushalte
aus der Zusatzstichprobe	204 Haushalte

Nach dem theoretischen Erwartungswert müsste die Zahl in beiden Stichproben gleich hoch sein. Tatsächlich beträgt die Zahl in der Zusatzstichprobe nur 85% des Erwartungswerts. Man kann dies so erklären, dass die Interviewer auf dem verlängerten Begehungsweg, in dem es nur um die Ermittlung von Ausländer-Haushalten ging, nicht ganz so intensiv kontaktiert haben wie in der Hauptstichprobe.

Dies ist leicht nachvollziehbar, wenn man bedenkt, dass es darum ging, bei 11.820 Haushalten lediglich diese eine Kontaktfrage zu stellen – und dies bei einer „Erfolgswahrscheinlichkeit“ von nur 4%. Gemessen an dieser Anforderung ist die Zahl der ermittelten Ausländer-Haushalte sehr zufriedenstellend.

Qualitätsneutrale Ausfälle (QNA)

Die im Adressenvorlauf ermittelten Adressen waren „Klingelknopf“-Adressen, waren also nach Augenschein – nicht nach vorherigem Kontaktinterview – in die Adressenauflistung aufgenommen worden. Insofern ist es plausibel, dass die Interviewer in der folgenden Feldarbeit – ca. 3 Monate später – teilweise feststellen, dass es sich um unbewohnte Wohnungen oder Geschäftsadressen handelt. Bewohner zum Zeitpunkt der Adressenermittlung können in der Zwischenzeit auch verstorben oder verzogen sein.⁶

Insgesamt haben die Interviewer 625 Adressen in diesem Sinne als Fehladressen gemeldet, das ist ein QNA-Anteil von 5,3%. Die Möglichkeit zur Verwendung von Ersatzadressen haben die Interviewer allerdings nur teilweise genutzt, nämlich in 261 Fällen.

⁶ Sofern bereits ein Nachfolgehaushalt in die Wohnung eingezogen war, wurde dieser zum Befragungshaushalt.

Resultierende Bruttostichprobe

Die Zahl der Haushalte, die die effektive Bruttostichprobe bilden, errechnet sich folgendermaßen:

985 x 12 Adressen der Hauptstichprobe	11.820
abzügl. qualitätsneutrale Ausfälle (QNA)	<u>625</u>
Verwendete Adressen der Hauptstichprobe	11.195

Zuzüglich verwendete Adressen aus der Zusatzstichprobe:

1. Ausländerhaushalte	423
2. QNA-Ersatzadressen	<u>261</u>
Bereinigtes Brutto	11.879

Ausfälle und Interviews auf Haushaltsebene

Von den 11.879 Haushalten der Bruttostichprobe konnten 6.060 zur Teilnahme an der Befragung gewonnen werden. Das ist eine Stichprobenausschöpfung von 51,0%.

Im einzelnen verteilen sich Ausfälle und Interviews wie folgt:

	abs.	%
Bruttofallzahl	11.879	100,0
Haushalt nicht angetroffen	910	7,7
Teilnahme nicht möglich (gesundheitlich oder geistig nicht in der Lage; Sprachprobleme)	351	2,9
Teilnahme verweigert	4.057	34,2
Haushalt vom Interviewer nicht bearbeitet; ohne Ergebnis zurück	<u>501</u>	<u>4,2</u>
Summe der Ausfälle	5.819	49,0
Interviews	6.060	51,0

37% der angesprochenen Haushalte haben die Teilnahme an der Befragung verweigert oder waren dazu aus gesundheitlichen, geistigen oder sprachlichen Gründen nicht in der Lage. 8% der Haushalte wurden vom Interviewer nie angetroffen – dies trotz teilweise erheblicher Kontaktbemühungen, wie oben dargestellt (Kapitel 3.1).

4% der Adressen wurden vom eingesetzten Interviewer nicht bearbeitet oder ohne dokumentierte Bearbeitung zurückgesandt. Dabei handelt es sich teils um Einzeladressen und teils um ganze Sample-Points, bei denen der eingesetzte Interviewer – und ggf. auch ein Ersatzinterviewer – vor den Anforderungen des Projekts oder den zu bearbeitenden Adressen offenbar zurückschreckte.

Entwicklung der Stichprobenausschöpfung

Die Stichprobenausschöpfung ist damit im Vergleich zu früheren Erstbefragungen im Sozio-ökonomischen Panel erneut zurückgegangen. Vergleichbar sind etwa folgende Werte:

	Stichproben- ausschöpfung
SOEP-Stichprobe A im Jahr 1984	60,6%
SOEP-Stichprobe E im Jahr 1998	54,2%
SOEP-Stichprobe F im Jahr 2000	51,0%

Dieser Trend ist ähnlich aus anderen Umfrageprojekten, etwa dem ALLBUS, bekannt. Er spiegelt veränderte Einstellungen der Bevölkerung zur Teilnahme an solchen Umfragen, wahrscheinlich aber auch veränderte Arbeitsweisen und Einstellungen der Interviewer in den vorhandenen Stäben für mündlich-persönliche Befragungen. Diese veränderten Bedingungen sind durch noch so große Anstrengungen im Rahmen einzelner Projekte nicht zu überwinden.

Um so wichtiger wird es, die Strukturen der realisierten Stichproben zu prüfen („Validierung“) und eine eventuelle Über- oder Unterrepräsentierung bestimmter Bevölkerungsgruppen oder Regionen festzustellen. Soweit solche Effekte feststellbar sind, können sie ggf. über Gewichtungsmassnahmen im Zuge der Stichprobenhochrechnung korrigiert werden.

Einzelne Hinweise dazu werden weiter unten gegeben (Kapitel 3.4). Zunächst soll die realisierte Stichprobenausschöpfung auf der Personenebene dargestellt werden.

Stichprobenausschöpfung auf der Personenebene

Zur Methodik der Haushalts-Panels gehört die individuelle Befragung aller Haushaltsmitglieder ab 16 Jahren. Instrument dafür ist der PERSONEN-Fragebogen.

Die Bruttostichprobe auf Personenebene umfasst alle Personen ab 16 Jahren, die in den teilnehmenden Haushalten leben. Die Nettostichprobe umfasst alle Personen, von denen ein auswertbarer PERSONEN-Fragebogen vorliegt.

Die entsprechenden Zahlen in der neuen SOEP-Stichprobe F sind folgende:

	abs.	%
Teilnehmerhaushalte in Welle 1	6.060	
In diesen Haushalten leben...		
Personen insgesamt	14.536	
darunter:		
Kinder unter 16 Jahren	2.993	
Befragungspersonen (= Brutto)	11.543	100
darunter:		
- nie angetroffen	123	1,1
- zur Teilnahme nicht in der Lage, alt/krank	37	0,3
- zur Teilnahme nicht bereit oder Interview nicht durchführbar	<u>482</u>	<u>4,2</u>
Summe Ausfälle	642	5,6
 Interviews	 10.901	 94,4

Die angestrebte individuelle Information liegt damit für fast 95% der Haushaltsmitglieder in den Panelhaushalten vor. Dieser Wert entspricht ziemlich genau demjenigen in den SOEP-Altstichproben A – E.

3.4 Nonresponse-Informationen

Nonresponse-Informationen sind Informationen, die für Teilnehmer und Nichtteilnehmer („Nonrespondenten“) gleichermaßen vorliegen oder ggf. auch nur für Nichtteilnehmer bzw. Ausfälle erhoben werden. Der Strukturvergleich für Teilnehmer und Nichtteilnehmer gibt Hinweise auf eine eventuelle Selektivität des Teilnahmeverhaltens und damit auf eventuelle Verzerrungen in der Struktur der Nettostichprobe.

Im Sozio-ökonomischen Panel wurden Nonresponse-Informationen bei Erstbefragungen bereits mehrfach genutzt, um Hinweise zur Validität der Stichprobe zu erhalten.⁷

Leider gibt es in der Regel – wenn überhaupt – nur wenige Merkmale, die für Teilnehmer und Nichtteilnehmer einer Befragung gleichermaßen vorliegen. Für die neue SOEP-Stichprobe F stehen folgende Merkmale zur Verfügung:

⁷ Vgl. Infratest-Methodenberichte zu Stichprobe A/Welle 1 (1984) und zu Stichprobe E/Welle 1 (1998).

Regionale Verteilung

1. Bundesland
2. Gemeindegrößenklassen politisch
3. Gemeindegröße nach BIK (Stadtregionen)
4. Gemeindetypen nach BBR (Siedlungsstrukturtypen)

Interviewergestützte Quartiersbeschreibung

5. Lage zur nächsten Großstadt
6. Gebäudetyp
7. Umgebung
8. Soziale Schicht

Haushaltsmerkmale (unvollständig)

9. Haushaltsgröße
10. Geschlecht des HV
11. Altersgruppe des HV
12. Nationalität des HV, ggf. Sprachkenntnisse
13. Berufstätigkeit/berufliche Stellung des HV

Nur die Regions- und Quartiersmerkmale liegen vollständig für die gesamte Bruttostichprobe vor. Die Haushaltsmerkmale stützen sich auf Angaben der Interviewer zu Haushaltsausfällen. Da bei Ausfällen häufig jedoch keinerlei Information zu ermitteln ist, liegen die Angaben entsprechend unvollständig vor; der Aussagewert für Nonresponse-Analysen ist dementsprechend eingeschränkt.

Wir untersuchen im Folgenden Unterschiede in der Struktur der Teilnehmer und Nichtteilnehmer nach regionalen Merkmalen. Für verschiedene Regionen bzw. Regionstypen kann unmittelbar die jeweilige *Response-rate* angegeben werden (siehe **Tabelle 2**).

Ost / West

Die Stichprobenausschöpfung bzw. die *Response-rate* ist in den neuen Ländern um fast 6 Prozentpunkte höher als in den alten Ländern

Bei gleicher Teilnahmequote in Ost und West müsste die Nettostichprobe der alten Länder 100 Panelhaushalte mehr und die der neuen Länder 100 Haushalte weniger umfassen.

Diese leichte Stichprobenverzerrung kann durch eine Regionalgewichtung nach Bundesländern leicht korrigiert werden.

Tabelle 2:

Stichprobenausschöpfung nach Regionen / Gemeindegröße

	Inter-views <i>abs.</i>	Ausfälle <i>abs.</i>	QNA <i>abs.</i>	Gesamt <i>abs.</i>	Berein. Brutto <i>abs.</i>	Aus- schöp- fung %	Interv. Diff. Soll <i>abs.</i>	Anteil QNA %
1) Ost/West								
Alte Länder	4.874	4.878	497	10.249	9.752	50,0	-101	5,1
Neue Länder	1.186	941	128	2.255	2.127	55,8	101	6,0
Gesamt	6.060	5.819	625	12.504	11.879	51,0	0	5,3
2) Polit. GGK in 1000 Einw.								
7 500 u.m.	831	1.199	127	2.157	2.030	40,9	-205	6,3
6 100 -u. 500	1.176	1.116	140	2.432	2.292	51,3	7	6,1
5 50 -u. 100	557	550	47	1.154	1.107	50,3	-8	4,2
4 20 -u. 50	1.023	979	106	2.108	2.002	51,1	2	5,3
3 5 -u. 20	1.421	1.162	121	2.704	2.583	55,0	103	4,7
2 2 -u. 5	526	440	46	1.012	966	54,5	33	4,8
1 unter 2	526	373	38	937	899	58,5	67	4,2
Gesamt	6.060	5.819	625	12.504	11.879	51,0	0	5,3
3) BIK-Typ								
0 500 u.m. Kern	1.781	2.153	242	4.176	3.934	45,3	-226	6,2
1 500 u.m. Rand	564	526	52	1.142	1.090	51,7	8	4,8
2 100 -u. 500 K	673	684	74	1.431	1.357	49,6	-19	5,5
3 100 -u. 500 R	397	326	36	759	723	54,9	28	5,0
4 50 -u. 100 K	255	201	19	475	456	55,9	22	4,2
5 50 -u. 100 R	49	42	6	97	91	53,8	3	6,6
6 20 -u. 50	561	525	40	1.126	1.086	51,7	7	3,7
7 5 -u. 20	924	720	83	1.727	1.644	56,2	85	5,0
8 2 -u. 5	394	324	38	756	718	54,9	28	5,3
9 unter 2	462	318	35	815	780	59,2	64	4,5
Gesamt	6.060	5.819	625	12.504	11.879	51,0	0	5,3

Gemeindegröße nach politischen Gemeindegrenzen

Es gibt ein starkes Land/Stadt-Gefälle in der Stichprobenausschöpfung bzw. der Teilnahmebereitschaft der Haushalte: von 58,5% in den kleinsten Gemeinden bis zu lediglich 40,9% in Großstädten ab 500.000 Einwohnern.

Wäre die *Response-rate* in den Wohnorten über alle Gemeindegrößenklassen gleich, müsste die Stichprobe ca. 200 mehr Panelhaushalte in den Großstädten über 500.000 Einwohner umfassen und entsprechend weniger in den kleinen Gemeinden.

Auch diese Stichprobenverzerrung ist über eine Regionalgewichtung nach GGK leicht zu korrigieren. Denkbar wäre, dieser – im Prinzip bekannten – differenziellen *Response-rate* nach Gemeindegrößenklassen künftig durch eine entsprechende Anlage der Bruttostichprobe statt durch eine nachträgliche Stichprobengewichtung Rechnung zu tragen.

Die Verteilung nach BIK-Typ zeigt ein ähnliches Bild – mit den Kernzonen der großen Ballungsräume als Problemgebiet.

Weitere Nonresponse-Analysen

Die SOEP-Gruppe im DIW hat eine Nonresponse-Analyse zur neuen SOEP-Stichprobe F vorgenommen, die sich auf die Interviewerangaben zu Quartiersmerkmalen und auf die (unvollständigen) Interviewerangaben zur Struktur der Ausfälle bezieht.⁸ Die wesentlichen Ergebnisse und Schlussfolgerungen sollen hier wiedergegeben werden:

- Erreichbarkeit der Haushalte

Zunehmende Entfernung von der Großstadt und Wohnen in Einfamilienhäusern erhöht die Wahrscheinlichkeit, dass Haushalte erreicht werden; Leben in Mischgebieten (keine reinen Wohngebiete) senkt die Wahrscheinlichkeit.

Insgesamt ist allerdings die Nichterreichbarkeit weitgehend zufällig (Pseudo- $R^2 = 0.044$).

- Kooperationsbereitschaft der Haushalte

Am Stadtrand und in großer Entfernung zu einer Großstadt ist die Kooperationsbereitschaft höher.

Neben den signifikanten Effekten der sozialen Schicht weist auch die erhöhte Kooperationsbereitschaft in Ein- und Zweifamilienhäusern auf einen Mittelstandsbias hin.

Die Nicht-Kooperationsbereitschaft ist aber nahezu völlig zufällig (Pseudo- $R^2 = 0.005$).

⁸ Däubler, T., Pischner, R., Wagner, G.: Nonresponse-Analysen Stichprobe F. Handout zum SOEP-Workshop am 3. Mai 2001, Berlin.

- Fazit

Sinkende Stichprobenausschöpfung muss keinen Qualitätsverlust der realisierten Stichprobe darstellen. Implikationen für die Hochrechnung sind zu prüfen.

4. Datenprüfung und übergebener Datenbestand

Datenerfassung, Datenprüfung und Datenbereinigung erfolgen in der neuen SOEP-Stichprobe F nach denselben Verfahren wie in den SOEP-Altstichproben. Eine erneute Darstellung erübrigt sich hier; verwiesen wird auf den vorliegenden *Methodenbericht zum SOEP 2000* und die dortigen Hinweise.

Der Datenbestand zur Stichprobe F wurde dem DIW am 6.2.2002 übergeben. Er umfasst folgende Bestandteile:

	Satzlänge	Fallzahl
Nettodaten HAUSHALTS-Fragebogen	1.100	6.060
Nettodaten PERSONEN-Fragebogen	3.580	10.901
Bruttodaten Haushalte	63	12.504
Bruttodaten Personen ⁹	77	14.536

⁹ Nur Personen in Teilnehmer-Haushalten; einschließlich Kindern.