

SOEP Survey Papers

Series B - Survey Reports (Methodenberichte)

SOEP 2002 – Methodenbericht Sonder- erhebung Hocheinkommensstichprobe zum Befragungsjahr 2002 (Welle 19) des Sozio-oekonomischen Panels

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 Sozialforschung. 2011. SOEP 2002 – Methodenbericht Sondererhebung Hocheinkommensstichprobe zum Befragungsjahr 2002 (Welle 19) des Sozio-oekonomischen Panels. SOEP Survey Papers 44: Series B. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)

Contact: DIW Berlin
SOEP
Mohrenstr. 58
10117 Berlin

Email: soepapers@diw.de

INFRATEST SOZIALFORSCHUNG

**SOEP 2002 –
METHODENBERICHT SONDERERHEBUNG HOCH-
EINKOMMENSSTICHPROBE ZUM BEFRAGUNGS-
JAHR 2002 (WELLE 19) DES SOZIO-OEKONO-
MISCHEN PANELS**

München, 2002

Reprint 2011

Inhalt	Seite
1. Ziele und Anlage der Erhebung	3
1.1 Zielsetzung	
1.2 Definition der "Haushalte im oberen Einkommensbereich"	4
1.3 Die Stichprobenbasis	5
1.4 Erhebungsmethoden und –instrumente	6
1.5 Zeitlicher Ablauf	7
2. Verfahren und Güte der Stichprobenbildung	8
2.1 Ausgangspunkt: Eine Zufallsstichprobe der Bevölkerung	8
2.2 Abbildung höherer Einkommen in der Ausgangsstichprobe	13
2.3 Bildung der Bruttostichprobe für die mündlich-persönliche Befragung	15
2.4 Teilnahmebereitschaft in der mündlich-persönlichen Befragung	20
2.5 Überprüfung der Einkommen	21
2.6 Struktur der Nettostichprobe nach sozioökonomischen Merkmalen	23
3. Datenprüfung und übergebener Datenbestand	30
SUMMARY	31

1. Ziele und Anlage der Erhebung

1.1 Zielsetzung

Ziel der Erhebung ist es, eine Datenbasis zur Analyse der Lebenslage, der Einkommen und der Vermögensbildung von Haushalten im oberen Einkommensbereich zu schaffen. Auftraggeber für Infratest ist das DIW Berlin, das auf dieser Datengrundlage ein Gutachten für die Bundesregierung erstellen wird. Anschließend an den von der Bundesregierung vorgelegten Armuts- und Reichtumsbericht¹ werden damit die Bemühungen fortgesetzt, Lebenslagen nicht nur am unteren, sondern auch am oberen Rand der Einkommensschichtung genauer zu beschreiben.

Im Rahmen des Sozio-ökonomischen Panels (SOEP) hatte das DIW seit einiger Zeit Überlegungen angestellt, wie eine Stichprobenbasis für "Reiche" oder für "einkommensstarke Haushalte" geschaffen werden könnte. In Zusammenarbeit mit Infratest Sozialforschung wurden verschiedene Ansätze konzipiert und erprobt. Der erfolgversprechendste Ansatz, der schließlich auch zu der hier vorliegenden Erhebung führte, wurde im Jahr 2001 im Rahmen einer vom BMBF geförderten "Feasibility Study" zu möglichen Datenerhebungen bei "einkommensstarken Haushalten" entwickelt und erprobt. Die Erfahrungen wurden in drei Berichtsbänden dokumentiert.² Die vorliegende Erhebung knüpft an diese Vorarbeiten an.

¹ Lebenslagen in Deutschland. Der erste Armuts- und Reichtumsbericht der Bundesregierung. Bundestagsdrucksache Nr. 14/11/117, Berlin 2001.

² Infratest Sozialforschung: Feasibility Study "Einkommensstarke Haushalte". Erster Zwischenbericht zur Testerhebung 2001: Die Stichprobenbasis. München 2001.
Infratest Sozialforschung: Feasibility Study "Einkommensstarke Haushalte". Zweiter Zwischenbericht zur Testerhebung 2001: Feldbericht und Hinweise zum Fragebogen, München 2001.
Infratest Sozialforschung: Feasibility Study "Einkommensstarke Haushalte". Dritter Bericht: Schlussfolgerungen aus der Testerhebung 2001. München 2001.

1.2 Definition der "Haushalte im oberen Einkommensbereich"

Bei der Abgrenzung der Haushalte im oberen Einkommensbereich sind drei konzeptionelle Entscheidungen zu treffen:

- a) welcher Einkommensbegriff verwendet wird
- b) ob Einkommen als solche oder als bedarfsgewichtete Äquivalenzeinkommen zugrunde gelegt werden
- c) und ab welcher Einkommensschwelle die Haushalte dem "oberen Einkommensbereich" zugerechnet werden.

Diese Entscheidungen müssen nach inhaltlichen Gesichtspunkten, aber auch unter Berücksichtigung pragmatischer Argumente getroffen werden. Die Lösung muss inhaltlich vertretbar, aber auch forschungspraktisch realisierbar sein.

a) Einkommensbegriff

Verwendet wird das "monatliche Haushaltsnettoeinkommen", und zwar nach Auskunft von Befragungspersonen, abgefragt in Einkommensklassen. Dies entspricht im Prinzip dem Erhebungsverfahren im Mikroszensus.³

Das erfragte monatliche Haushaltsnettoeinkommen ist, wie bekannt, eine recht grobe Schätzgröße. Einkommenskomponenten, die nicht monatlich anfallen, dürften weitgehend unberücksichtigt bleiben. Die Zuverlässigkeit der Angaben ist darüber hinaus nur begrenzt gegeben, insbesondere wenn nicht der Hauptverdiener, sondern – im Rahmen einer repräsentativen Personenstichprobe – andere Haushaltsmitglieder die Auskunftsperson sind.

Trotz dieser eingeschränkten Zuverlässigkeit ist das so erfragte Haushaltsnettoeinkommen die am häufigsten verfügbare statistische Information zur Einkommenslage der Haushalte. In der vorliegenden Erhebung dient diese Information als *Ausgangsgröße für die Stichprobenbildung*. Nimmt ein Befragungshaushalt dann am Interview der eigentlichen Erhebung teil, liegt eine Vielzahl neuer, differenzierterer Einkommensinformationen vor. Diese Informationen wurden im Zuge der Stichprobenbildung u.a. benutzt, um die ursprünglichen, weniger zuverlässigen Einkommensangaben der Ausgangsstichprobe zu überprüfen.

b) Einkommen oder Äquivalenzeinkommen?

Das Äquivalenzeinkommen ist eine errechnete Größe, die das tatsächlich verfügbare Haushaltseinkommen auf die Zahl der im Haushalt lebenden Personen (ggf. differenziert nach Alter) bezieht. Es ist also ein "bedarfsgewichtetes" Einkommen, das eine Bewertung des verfügbaren Einkommens im Hinblick auf Versorgungsaspekte beinhaltet. In der Armutsforschung ist die Verwendung von Äquivalenzeinkommen durchweg üblich. In der Reichtumsforschung wird teilweise analog vorgegangen (z.B. bei der Definition einer "200%-Schwelle").

³ nicht dagegen dem der EVS, die für Validierungszwecke hier daher nur bedingt geeignet ist.

Für die hier vorliegende Erhebung wurde demgegenüber das tatsächlich verfügbare Haushaltseinkommen zugrunde gelegt. Dies ist einerseits sachlich zu begründen: Es interessieren die Haushalte am oberen Rand der Einkommensschichtung in Deutschland. Hinzu kommt eine methodisch praktische Begründung, nämlich die Schwierigkeit, Äquivalenzeinkommen zu berechnen, wenn das Haushaltseinkommen nur in groben Größenklassen vorliegt.

Die Verwendung des tatsächlich verfügbaren Haushaltseinkommens ist sachlich vertretbar, hat aber strukturelle Effekte, die bei der Nutzung und Bewertung der Erhebungsergebnisse zu beachten sind. Der Haupteffekt ist, dass Ein-Personen-Haushalte bei diesem Vorgehen relativ selten im "oberen Einkommensbereich" zu finden sind, während Mehr-Personen-Haushalte – soweit sie Mehrfach-Verdiener-Haushalte sind – mit höherer Wahrscheinlichkeit über der gewählten Einkommensschwelle liegen.

c) Die Einkommensschwelle

Die oberen 5% der Einkommensschichtung in Deutschland, gemessen am Haushaltsnettoeinkommen nach Befragtenangaben, weist der Mikrozensus 2000 oberhalb einer Einkommensschwelle von 8.000 DM aus.⁴ Etwa die Hälfte davon haben ein Einkommen oberhalb von 10.000 DM (2,6%) und wiederum die Hälfte davon ein Einkommen oberhalb von 12.000 DM (1,4%).

Die Präferenz für die hier vorliegende Erhebung hätte bei einem möglichst hohen Schwellenwert, etwa bei 12.000 DM, gelegen. Für eine so enge Abgrenzung – bei dieser Einkommensschwelle erfasst man die oberen 1,4% der Haushalte – fehlt jedoch die erforderliche Stichprobenbasis. Festgelegt wurde

- eine Schwelle von mindestens 7.500 DM des Haushaltsnettoeinkommens im Jahr 2001
- mit einem disproportionalen, nämlich höheren Auswahlatz für Haushalte mit einem Einkommen von 10.000 DM und mehr.

1.3 Die Stichprobenbasis

Es gibt kein Verzeichnis aller Haushalte im oberen Einkommensbereich in Deutschland, aus dem man direkt nach Zufallsverfahren eine Stichprobe ziehen könnte. Die einzige Möglichkeit ist der Weg über eine allgemeine Bevölkerungsstichprobe, in der dann – im Rahmen eines Interviews – die Haushalte mit höherem Einkommen identifiziert werden.

Die Frage nach dem Haushaltseinkommen kann dabei nicht im Rahmen eines Kontaktinterviews – als sogenannte Screening-Frage zur Ermittlung der Zielhaushalte einer Untersuchung – gestellt werden. Dafür sind Einkommensfragen zu heikel. Die Frage nach dem Haushaltseinkommen kann bestenfalls am Ende eines Interviews gestellt werden, wenn sich eine gewisse Vertrauensbasis auf Seiten der Befragungspersonen herausgebildet hat.

⁴ Genauer: 5,6% bezogen auf alle Haushalte mit Einkommensangaben (Berechnungen des DIW).

Daraus folgt, dass eine Zufallsstichprobe einkommensstarker Haushalte nur im "Huckepack-Verfahren" möglich ist, also indem man eine große Bevölkerungsstichprobe nutzt, in der das Merkmal "Haushaltsnettoeinkommen" erfragt wurde und die datenschutzrechtlichen Voraussetzungen für eine Folgebefragung von Teilgruppen vorliegen.

Dies ist das Verfahren, das der vorliegenden Erhebung zugrunde liegt. Stichprobenbasis ist ein großer Datenbestand von 99.000 telefonischen Interviews, die im Rahmen der kontinuierlichen Mehrthemenbefragung *InfraScope* im Laufe des Jahres 2001 durchgeführt wurden. Die Standarddemographie dieser Befragung umfasst das Merkmal Haushaltsnettoeinkommen, so dass einkommensstarke Haushalte nach der oben genannten Definition abgrenzbar sind. Einzelheiten der Methodik und Vorgehensweise sind unten in Kapitel 2 dargestellt.

1.4 Erhebungsmethoden und –instrumente

Die Befragung der einkommensstarken Haushalte wurde im Rahmen der SOEP-Befragung des Jahres 2002 durchgeführt. Das heißt:

- Das Erhebungsprogramm ist mit dem des SOEP 2002 identisch. Pro Haushalt kommt ein HAUSHALTS-Fragebogen zum Einsatz, und zusätzlich wird mit jedem Haushaltsmitglied ab 16 Jahren ein individuelles Interview mit dem PERSONEN-Fragebogen durchgeführt. Letzterer enthält im Jahr 2002 als besonderes Schwerpunktthema auch eine individuelle "Vermögensbilanz".
- Die Erhebungsmethoden entsprechen ebenfalls dem SOEP-Standard. Die Interviewmethode ist im Prinzip die mündlich-persönliche Befragung. Ein Selbstausfüllen des Fragebogens durch die Befragungsperson ist zulässig, sofern die Befragungsperson dies bevorzugt. Die mündlich-persönlichen Interviews können entweder computerunterstützt als Laptop-Interviews (CAPI) oder in herkömmlicher Weise mit dem Papierfragebogen (PAPI) durchgeführt werden.

Für methodische Einzelheiten wird auf die von Infratest erstellten Methodenberichte zum SOEP verwiesen. Das Fragenprogramm ist dokumentiert im ANLAGENBAND zum Methodenbericht 2002.

1.5 Zeitlicher Ablauf

Der Auftrag für die Durchführung der Erhebung wurde Anfang Dezember 2001 erteilt. In den Monaten Dezember bis Februar wurden die Vorbereitungsarbeiten durchgeführt, also

- die Ziehung der Stichprobe
- die Durchführung von telefonischen Kontaktinterviews
- die Erstellung der Erhebungsinstrumente
- die Auswahl und Instruktion der Interviewer.

Die Feldarbeit mit der Durchführung der Interviews fand in der Zeit von März – Juli 2002 statt, wobei die Hauptfeldzeit in den Monaten März – Mai lag und die Monate Juni – Juli der Nachbearbeitung dienten.

Der geprüfte Datenbestand wurde dem DIW am 20. August 2002 übergeben. Die realisierte Stichprobe von Haushalten im oberen Einkommensbereich umfasst 1.224 Haushalte mit 2.671 Personeninterviews.

2. Verfahren und Güte der Stichprobenbildung

2.1 Ausgangspunkt: Eine Zufallsstichprobe der Bevölkerung

NFO Infratest verfügt über zwei Systeme kontinuierlicher Bevölkerungsbefragungen: das eine für mündlich-persönliche Befragungen (CAPI-Omnibus), das andere für telefonische Befragungen (InfraScope). In beiden Systemen werden kontinuierlich nach gleichbleibendem Verfahren unabhängige Bevölkerungsstichproben gezogen. Diese können über die verschiedenen Wellen hinweg zu größeren Stichproben kumuliert werden. Im Jahr 2001 wurden im CAPI-Omnibus in 24 Wellen insgesamt rd. 30.000 Interviews durchgeführt, im InfraScope-System insgesamt 99.000 Interviews.

Für die Bildung einer Stichprobe einkommensstarker Haushalte wurde aus Gründen der erforderlichen Fallzahl eine möglichst große Ausgangsstichprobe benötigt. Nur die InfraScope-Stichprobe bot eine hierfür ausreichende Basis.

Die InfraScope-Stichprobe 2001 ist eine Telefon-Stichprobe nach ADM-Standard, der auf die methodischen Verfahren von Gabler/Häder zurückgeht. Eine genauere Methodenbeschreibung ist auf den folgenden Seiten beigefügt.⁵

⁵ Die Methodenbeschreibung weicht in einer Reihe von Punkten von derjenigen ab, die im ersten Infratest-Bericht zur Feasibility Study, a.a.O., S. 5 ff., gegeben wurde. Grund ist, dass diese sich auf die InfraScope-Stichprobe 2000 bezog, in der noch nicht die Umstellung auf den ADM-Standard vollzogen war.

Telefonbefragungen Bevölkerung (ITMS nach ADM-Design)

1. Grundgesamtheit und Stichprobenanlage

Grundgesamtheit von Bevölkerungsbefragungen sind in der Regel alle in Privathaushalten lebenden deutschsprachigen Personen ab 14 Jahren. Weitere studienspezifische Einschränkungen sind möglich. Die telefonische Durchführung der Befragung reduziert die Grundgesamtheit auf die Auswahlgesamtheit „deutschsprachige Personen ab 14 Jahren in Telephonhaushalten“.

Die Befragung basiert auf dem Infratest-Telefon-Master-Sample (ITMS), das für derartige Untersuchungen aufgebaut wurde und zu verzerrungsfreien Stichproben (insbesondere weitgehende Vermeidung des not-at-home bias) ohne Klumpeneffekte führt. Das ITMS ist als multi-stratifizierte Haushaltsstichprobe auf Flächenbasis mit zufälliger Zielpersonenauswahl im Haushalt mit dem Schwedenschlüssel konzipiert. Es entspricht hinsichtlich des Random-Digit-Dialling dem ADM-Standard, der nach dem Gabler-Häder-Verfahren definiert wurde. Kennzeichnend ist dabei, dass das sog. „random last two digits (RL2D)“ der Telefonnummern nicht im Rahmen der einzelnen Stichproben durchgeführt wird, sondern dass dieser Randomisierungsschritt bereits im Rahmen der Erstellung der Auswahlgrundlage implementiert ist. Es garantiert, dass auch die Telefonanschlüsse in der Auswahlgrundlage enthalten sind und verzerrungsfrei gezogen werden können, die nicht in Verzeichnisse eingetragen sind.

2. Erstellung der ADM-Auswahlgrundlage und eindeutige regionale Verortung der Rufnummern im ITMS-System

ITMS Stichproben werden aus der Auswahlgrundlage der „Arbeitsgemeinschaft ADM-Telefonstichproben“ gezogen. Erstellungsbasis dieser Auswahlgesamtheit sind sämtliche Einträge aus Telefonverzeichnissen.

Diese werden zunächst von Redundanzen bereinigt. Im nächsten Schritt werden die letzten beiden Stellen aller eingetragenen Rufnummern abgeschnitten. Die so gewonnenen sog. Rufnummernstämme werden ebenfalls entdupliziert. Anschließend wird pro Nummernstamm ein Block mit der Ziffernfolge 00 bis 99 erzeugt. Somit erhält man ein Universum aller Telefonnummern, eingetragene wie auch generierte. Nur solche Telefonnummern sind in der Auswahlgrundlage nicht enthalten, die in Blöcken ohne einen einzigen Eintrag liegen.

Diese Auswahlgrundlage enthält lediglich Telefonnummern, jedoch keine Adressen, da diese für eine anonyme Befragung unerheblich sind. Aus den Einträgen werden jedoch Kennzeichen für die Art des Eintrags (geschäftlich/privat) sowie die Gemeindekennziffer übernommen.

Sofern eine Rufnummer nicht eingetragen ist, werden dieser bis zu drei verschiedene Gemeindekennziffern zugeordnet; und zwar die der drei häufigsten Gemeinden der eingetragenen Rufnummern desselben Nummernstamms. Kennzeichen für die Art des Anschlusses können für generierte Rufnummern natürlich nicht vergeben werden.

Die Nummern aller Teilnehmer, die auf keinen Fall im Zusammenhang mit einer Befragung angerufen werden wollen, werden in der Auswahlgrundlage gesperrt.

Da das ITMS als Flächenstichprobe konzipiert ist, muss jede Telefonnummer eindeutig regional verortet sein. Nicht eingetragene Telefonnummern sind nur dann eindeutig verortbar, sofern sich alle eingetragenen Rufnummern desselben Blockes in einer einzigen Gemeinde befinden. Ist dies nicht der Fall, wird im ITMS-System bei generierten Rufnummern mit mehreren möglichen Gemeindekennziffern eine der Gemeindekennziffern per Zufall ausgewählt. Diese Zufallsauswahl wird per Bedeutungsgewicht so gesteuert, dass die Häufigkeitsverteilung der Gemeindekennziffern der nicht-eingetragenen Nummern im jeweiligen Block der Verteilung der eingetragenen Nummern entspricht.

3. Schichtung, Ziehung und Stichprobenrealisierung

Die Schichtung der Haushaltsstichprobe erfolgt zum einen anhand von Kriterien der amtlichen Gebietseinteilung (Bundesländer, Nielsegebiete Regierungsbezirke, Kreise, ggf. – bei Schwerpunktstichproben - Gemeinden und Gemeindeteile) zum anderen anhand der BIK-Gemeindetypen (10er Skala). Das jeweilig verwandte Schichtungsmodell ist studienspezifisch wählbar und wird auf die angestrebte Nettofallzahl, die Optimierung der Feldarbeit und andere studienspezifische Gesichtspunkte ausgerichtet.

Das Nettosoll wird erstens mit dem reziproken Wert der erwarteten Ausschöpfung multipliziert und in einer Allokationsrechnung unter Verwendung des COX-Verfahrens auf die Schichtungszellen verteilt. Diese Brutto-Sollverteilung des Schichtungstableaus wird zweitens haushaltsproportional auf die jeweiligen schichtangehörigen Gemeinden verteilt und daraus dann das Ziehungsbrutto auf Gemeindeebene berechnet. Die Ziehung der Telefonnummern erfolgt pro Gemeinde per Zufallsstart und Intervall. Nicht-private Einträge, bereits gezogene sowie gesperrte Rufnummern werden dabei negiert.

Das ITMS besteht also aus einer mikrostratifizierten und ungeklumpten Stichprobe, die sich proportional zur Zahl der Privathaushalte auf die Mikrozellen (Gemeinden oder Gemeindeteile) aufteilt. Die Multistratifikation und Aufteilung der Stichprobe auf die Zellen erfolgt vollautomatisch über ein Allokationsprogramm.

Die Stichprobenrealisierung erfolgt nach dem Konzept der Nettosteuerung voll automatisch per Sample-Management-System (SMS). Dabei geht das Schichtungstableau der Allokationsrechnung als Sollstruktur in die Steuerung der Feldarbeit ein. Es ist somit gewährleistet, dass in jeder Zelle die erforderliche Zahl von Interviews durchgeführt wird. Von diesem Programm wird auch - falls nötig - die Gleichverteilung der Interviews auf Befragungstage und Tageszeiten gesteuert.

Innerhalb jeder Steuerungszelle sind die Datensätze der Telefonhaushalte nach Zufallszahlen sortiert. Somit bildet jede Zelle eine Urne im klassischen Sinne. Nicht erreichte Haushalte werden zurückgelegt und kommen in größerem zeitlichen Abstand zu anderen Tageszeiten zur Wiedervorlage. Die an einem bestimmten Tag nicht erreichten Haushalte werden durch solche substituiert, die an anderen Tagen nicht erreicht werden. Damit entfällt der sogenannte "not-at-home-bias" weitgehend (nur Haushalte, die auch nach dem 12. Kontakt nicht angetroffen wer-

den, werden angesteuert; nach unserer Erfahrung handelt es sich dabei i.d.R. um (noch) nicht geschaltete Telefonnummern, auf die keine Postansage aufgeschaltet ist).

Um mögliche Einflüsse der Tageszeit auf Untersuchungsergebnisse von vorneherein auszuschalten, wird die Stichprobe nach einem Verfahren der "dynamischen Repräsentativität" bezüglich der Besetzung der Zellen des Multistratifikationstableaus optimiert, so dass sich für jedes Stundenintervall vorgabenproportionale Teilstichproben ergeben.

Durch die letztlich nur noch aus den "harten" Verweigerern bestehenden Ausfälle und die optimale regionale Aussteuerung der Stichproben kann die abschließende Personengewichtung mit einer wesentlich kleineren Faktorenspannweite die Stichprobe an der Struktur der Wohnbevölkerung justieren, als dies mit den herkömmlichen Verfahren möglich ist.

4. Gewichtung

Nicht in allen von den Interviewern kontaktierten Haushalten kommt ein Interview zustande.

Diese Ausfälle können sich disproportional zur Grundgesamtheit verteilen und so Verzerrungen der Stichprobe hervorrufen. Derartige Verzerrungen werden durch aufeinanderfolgende Faktorengewichtungen ebenso ausgeglichen, wie die von der Haushaltsgröße und der Zahl der Telefonanschlüsse abhängende Auswahlchance für die Zielperson.

4.1 Telefonnummern- und Haushaltsstichprobe (Bereinigung der Auswahlchancen)

Das ITMS führt zu Stichproben, in denen jede Telefonnummer die gleiche Auswahlchance hat. Haushalte mit mehreren genutzten Anschlüssen haben daher eine der Zahl dieser Anschlüsse entsprechende Mehrfachchance bei der Auswahl. Zu deren Bereinigung wird die realisierte Stichprobe mit der reziproken Zahl der für Gespräche genutzten Anschlüsse – nur diese haben Einfluss auf die Auswahlchance der Haushalte - je Haushalt multipliziert.

4.2 Gewichtung der Haushaltsstichprobe nach Bundesländern und Gemeindetypen

Diese Stufe der Gewichtung soll Abweichungen vom ursprünglichen haushaltsproportionalen Sample-Ansatz korrigieren. Aufgrund der beschriebenen Stichprobensteuerung durch das ITMS, das zu vollständiger Proportionalität führt, kann eine solche Haushaltsgewichtung jedoch in der Regel unterbleiben.

4.3 Haushalts- und Personenstichprobe (Umwandlung)

Das beschriebene Auswahlverfahren führt zu einer haushaltsrepräsentativen Stichprobe, wobei – nach Bereinigung - jeder Haushalt die gleiche Chance hat, in die Auswahl zu kommen.

In jedem der ausgewählten Haushalte wird durch ein zufälliges, gleiche Auswahlchancen innerhalb eines Haushalts produzierendes Verfahren (z.B. Schwedenschlüssel) nur eine Person als Zielperson ausgewählt. Dies gilt unabhängig davon, wie viele zur Grundgesamtheit gehörende Personen in dem betreffenden Haushalt leben. Die Chancen für die in Privathaushalten

lebenden Personen der Grundgesamtheit, als Befragungsperson der Stichprobe ausgewählt zu werden, sind demnach umgekehrt proportional zur Zahl der zur Grundgesamtheit gehörenden Personen in ihren Haushalten.

Um eine repräsentative Personenstichprobe zu erhalten, wird die erstellte Stichprobe mathematisch im nachhinein so umgeformt, dass jede Person der Grundgesamtheit stichprobentheoretisch die gleiche Auswahlchance erhält.

4.4 Gewichtung der Personenstichprobe nach Bundesländern, Altersgruppen und Geschlecht

Die Gesamtstichprobe wird anschließend an die aus der amtlichen Statistik bekannten Sollstrukturen der genannten Merkmale angepasst. Als Datenbasis dient die Bevölkerungsfortschreibung.

5. Fazit

Sowohl bei der Stichprobenanlage (Schichtung a priori) als auch bei der Gewichtung (Schichtung a posteriori) wird die Stichprobe nach der Verteilung der Privathaushalte bzw. den soziodemographischen Strukturen der deutschen Wohnbevölkerung und nicht etwa nach den Telefonhaushalten bzw. der deutschen "Telefonbevölkerung" ausgerichtet. Der Anteil der Telefonhaushalte an den Privathaushalten liegt inzwischen in West und Ost weit über 95%. Merkmalsunterschiede zwischen Telefonhaushalten und Privathaushalten sind deshalb ohnehin klein und können nach dem vorgestellten Stichproben- und Gewichtungsverfahren für die allermeisten Merkmale praktisch vernachlässigt werden.

2.2 Abbildung höherer Einkommen in der Ausgangsstichprobe

Die Standarddemographie der telefonischen InfraScope-Befragungen enthält, im Anschluss an die Frage zur Zahl der Personen im Haushalt, folgende Frage zur Höhe des Haushaltsnettoeinkommens:

Ich lese Ihnen jetzt sechs Einkommensgruppen vor. Zu welcher Gruppe gehört Ihr Haushalt? Gemeint ist das monatliche Haushaltseinkommen nach Abzug von Steuern und Sozialversicherung.

*bis unter 1.500 DM
1.500 b.u. 3.000 DM
3.000 b.u. 4.500 DM
4.500 b.u. 6.000 DM
6.000 b.u. 7.500 DM
über 7.500 DM*

Zur genaueren Abbildung höherer Einkommen wurde diese Frage ab Mitte März 2001 durch Nachfragen ergänzt, die eine Gruppierung der oberen Einkommensklassen in folgender Weise ermöglichen:

*7.500 b. u. 8.000 DM
8.000 b. u. 10.000 DM
12.000 DM mehr*

Nicht alle Befragten wollen oder können die Frage nach dem Haushaltsnettoeinkommen beantworten. Für 16% fehlt die Einkommensangabe.⁶

Interviews ohne Einkommensangabe sind für die Zwecke der vorliegenden Erhebung nicht brauchbar. Die Datenbasis, aus der die Stichprobe einkommensstarker Haushalte zu ziehen ist, verringert sich dadurch von insgesamt 99.175 Interviews auf 83.329 Fälle.

Anhand anderer Merkmale der Standarddemographie lässt sich überprüfen, welche Personengruppen auf diese Weise möglicherweise überproportional ausfallen. Problematisch wäre etwa ein überproportionaler Ausfall bei Personengruppen, bei denen höhere Einkommen zu erwarten sind. Entsprechende Auswertungen zeigen allerdings, dass dies nicht der Fall ist.

Wir haben diese Analyse bereits im 1. Bericht zur Feasibility Study, a.a.O. S. 9, anhand der InfraScope-Daten des Jahres 2001 vorgenommen. Der Anteil der Befragten mit Einkommensangaben betrug dort 80%⁷. Mehr als 3% unter diesem Durchschnittswert lagen nur drei Gruppen:

⁶ Ungewichtet. Der personengewichtete Anteil liegt etwas höher.

⁷ Nach Personengewichtung.

- Jüngere Befragte von 14-20 Jahren 53%
- Ältere Befragte von über 70 Jahren 75%
- Personen in größeren Haushalten von 5 und mehr Personen 72%
- Insgesamt 80%.

Die geringere Auskunftsfähigkeit in diesen Gruppen ist plausibel und deutet nicht auf spezifische Ausfälle bei einkommensstarken Haushalten hin. Im Gegenteil machen Personen aus den höheren Bildungs- und Berufsgruppen häufiger als der Durchschnitt Angaben zu ihrem Haushaltseinkommen:

- Personen mit abgeschlossenem Studium 84%
- gehobene/höhere Angestellte und Beamte 87%
- Freiberufler und Selbständige 81%

Der Ausfall von 16% in der InfraScope-Stichprobe 2001 aufgrund fehlender Einkommensangaben wirkt demnach eher in Richtung einer leichten Überrepräsentation höherer Einkommensgruppen.

Begrenzt man die Analyse auf befragte Personen mit Einkommensangabe, so zeigen sich im Vergleich von InfraScope mit SOEP und Mikrozensus (MZ) folgende Anteile von Haushalten mit höheren Einkommen:

Haushaltseinkommen	InfraScope	SOEP ⁸	MZ
	2001 %	2000 %	2000 %
7.500 DM und höher	7,7	6,3	7,0
8.000 DM und höher	5,5	5,3	5,6
10.000 DM und höher	2,7	2,3	2,6
12.000 DM und höher	1,3	1,0	1,4

Die drei Datenquellen weisen also weitgehend übereinstimmende Anteile von Personen im oberen Einkommensbereich auf, und zwar für alle hier interessierenden Einkommensschwellen.

Dies spricht dafür, dass Haushalte im oberen Einkommensbereich in unserer Ausgangsstichprobe gut abgebildet sind.

⁸ Zahlen für SOEP und MZ nach Berechnungen des DIW. SOEP-Zahlen gestützt auf Income-Screener im Haushaltsfragebogen.

2.3 Bildung der Brutto-Stichprobe für die mündlich-persönliche Befragung

Die Ausgangsstichprobe InfraScope 2001 enthält 6.330 Befragte, die ein Haushaltseinkommen von mindestens 7.500 DM angegeben haben. Dies ist die Ausgangsgröße für die Auswahl- und Ausfallprozesse, die zunächst zu der Brutto-Stichprobe für die mündlich-persönliche Befragung im Rahmen des SOEP 2002 führen.

Die Auswahl- und Ausfallprozesse werden im Folgenden beschrieben. Für den Datennutzer sind alle diese Prozesse im Datenbestand nachvollziehbar. Grundlage dafür ist das sogenannte "Bruttoband", das die genannten 6.330 Fälle dokumentiert. Der Bruttodatenbestand umfasst die Merkmale der InfraScope-Standarddemographie einschließlich ergänzender Regionaldaten sowie Informationen darüber, was mit dem jeweiligen Fall im Verlauf der Stichprobenbildung und Befragung passierte.

Schritt 1: Einverständniserklärung

Am Ende jedes Interviews in der telefonischen InfraScope-Befragung wird die Befragungsperson um ihr Einverständnis gebeten, dass die Telefonnummer gespeichert wird und das Umfrageinstitut für eventuelle weitere Fragen ggf. noch einmal anrufen darf. Dieses Einverständnis ist die datenschutzrechtliche Voraussetzung dafür, dass eine Folgebefragung möglich ist.

Von allen InfraScope-Befragten des Jahres 2001 erteilten 87% ihr Einverständnis in der beschriebenen Weise. Bei den Personen mit höherem Einkommen ist dieser Anteil sogar noch etwas höher:

Einverständnis zur Speicherung der Telefonnummer zum Zweck einer eventuellen weiteren Befragung

InfraScope 2001 gesamt	87,0%
darunter	
Personen mit Einkommensangabe	89,1%
Haushaltseinkommen > 7.500 DM	89,5%

Der Anteil von 11% Nicht-Zustimmung zu einer eventuellen weiteren Befragung variiert zwischen den Bevölkerungsgruppen nur geringfügig. Einen leicht erhöhten Anteil an Nicht-Zustimmung findet man bei älteren Befragten sowie bei Personengruppen, bei denen ohnehin Zweifel möglich sind, ob das von ihnen genannte Haushaltseinkommen von mindestens 7.500 DM zutrifft:

Anteil der Nicht-Zustimmung nach Personengruppen
Basis: Haushaltsnettoeinkommen > 7.500 DM

Durchschnitt	11%
Alter > 70 Jahre	15%
Arbeitslos	16%

Schüler	14%
Arbeiter	15%
einfache/mittlere Beamte	15%

Die Beschränkung der Stichprobe auf Personen mit Zustimmung zur Speicherung der Telefonnummer hat auf die Stichprobenstruktur damit nur einen vernachlässigswerten Effekt.

Die Größe der Stichprobe verringert sich durch die Einschränkung von 6.330 auf 5.663 Fälle.

Schritt 2: Regionale Klumpung mit disproportionalen Auswahlansatz

Anders als mündlich-persönliche Befragungen, bei denen mit Random-Route-Verfahren gearbeitet wird, sind Telefonstichproben nach ADM-Design praktisch ungeklumpt. Die Adressen der befragten Telefonhaushalte verteilen sich als Streuadressen bevölkerungsproportional über das gesamte Bundesgebiet. Die konkrete Haushaltsadresse ist in der Regel nicht einmal bekannt. Gearbeitet wird nur mit der Telefonnummer.

Dies ist der Grund, warum das für die vorliegende Befragung gewählte Erhebungsdesign ungewöhnlich ist: Es beinhaltet den Umstieg von einer Telefonstichprobe auf eine mündlich-persönliche Befragung, die gewöhnlich nach ganz anderen erhebungspraktischen Gesichtspunkten angelegt und optimiert wird.

Für den Erfolg einer mündlich-persönlichen Befragung ist es wichtig, dass die Anfahrtswege der Interviewer zu den Befragungsadressen nicht zu lang sind. Streuadressen, zu denen der Interviewer große Entfernungen zurücklegen muss, erhöhen nicht nur die Umfragekosten, sondern vermindern auch die erreichbare Stichprobenausschöpfung.

Im vorliegenden Projekt konnte eine regionale Klumpung, die die Anfahrtswege für die Interviewer verringert, vorgenommen werden, weil die Zahl der Telefonhaushalte in der Ausgangsstichprobe (N = 5.663) höher war als für die Durchführung der mündlich-persönlichen Umfrage benötigt. Knapp 2.000 Adressen konnten aus der Stichprobe herausgenommen werden. Dieser Arbeitsschritt wurde für eine gewisse regionale Klumpung zur Optimierung der Anfahrtswege für die Interviewer genutzt.

Zugleich erfolgte in diesem Schritt eine disproportional Schichtung der Stichprobe. Die Verringerung der Stichprobe um "überflüssige" Adressen erfolgte mit einem unterschiedlichen Auswahlansatz in vier Schichten, die nach dem Kriterium Einkommenshöhe und Ost/West definiert sind:

- Haushalte mit einem Einkommen von > 10.000 DM sollten möglichst vollständig in der Stichprobe bleiben. Das heißt, dass die Fallzahlreduzierung möglichst weitgehend bei den Haushalten mit Einkommen zwischen 7.500 und 10.000 DM vorgenommen werden sollte.
- Die neuen Bundesländer, die in der Ausgangsstichprobe der einkommensstarken Haushalte nur mit 8,4% vertreten sind, sollen durch einen leicht erhöhten Auswahlansatz auf einen Stichprobenanteil von mindestens 10% angehoben werden.

Welche Auswahlsätze bei diesem Arbeitsschritt in den vier Schichten der Stichprobe effektiv zustande kamen, zeigt die folgende Übersicht:

Anzahl der Telefonhaushalte ...	vor der Auswahl	nach der Auswahl	Auswahl- satz ¹⁾
Gesamt	5.663	3.672	64,8%
Haushalte mit ...			
Einkommen > 10.000 DM			
West	1.521	1.412	92,8%
Ost	<u>138</u>	<u>137</u>	99,3%
Gesamt	1.659	1.549	
Einkommen < 10.000 DM ²⁾			
West	3.666	1.865	50,9%
Ost	<u>338</u>	<u>258</u>	76,3%
Gesamt	4.004	2.123	

-
- 1) Prozentuierung Spalte 2 auf Spalte 1. Der Auswahlatz würde sich rechnerisch etwas verändern, wenn man als Basis alle Haushalte der Ausgangsstichprobe verwenden würde, einschließlich derer, die der Speicherung ihrer Telefonnummer nicht zugestimmt haben (N = 6.330 gesamt).
 - 2) einschließlich der Haushalte, für die differenziertere Einkommensklassen oberhalb von 7.500 DM noch nicht erfragt wurden (Befragungszeitraum Januar – Mitte März 2001).
-

Durch die disproportionalen Auswahlätze in diesem Arbeitsschritt erhöht sich in der Stichprobe

- der Anteil der Haushalte mit > 10.000 DM Einkommen von vorher 29,3% auf 42,2%
- und der Anteil der Ost-Haushalte von vorher 8,4% auf 10,8%.

Die in der Stichprobe verbleibenden 3.672 Haushalte bezeichnen wir als "Brutto 2". Sie bilden die Brutto-Stichprobe für den nächsten Arbeitsschritt, der den Versuch beinhaltet, die InfraScope-Befragungshaushalte für eine umfangreiche Folgebefragung zum "Leben in Deutschland" zu gewinnen.

Schritt 3: Kontaktinterviews zur Vorklärung der Befragungsbereitschaft

An diesem Punkt im Arbeitsablauf sind für die ausgewählten 3.672 Haushalte noch keine Namen und Adressen bekannt, sondern lediglich die Telefonnummern. Der Übergang zu einer mündlich-persönlichen Befragung durch Interviewer ist auf dieser Basis noch gar nicht möglich.

Erforderlich ist daher ein Zwischenschritt mit telefonischen Kontaktinterviews. Die Haushalte wurden angerufen und über das Vorhaben einer größeren Befragung zum "Leben in Deutschland" informiert. Wenn die angetroffene Kontaktperson sich vorstellen konnte, dass der Haushalt an einer solchen Befragung teilnehmen würde, wurde sie um die Angabe von Namen und Adresse gebeten, damit Informationsmaterial zu "Leben in Deutschland" per Post zugeschickt werden konnte.

Hat die Auskunftsperson mit Bezug auf diesen Zweck Namen und Adresse genannt, wurden diese notiert; wurden Name und Adresse nicht angegeben, gilt der Haushalt als Ausfall und kann in die weiteren Schritte nicht einbezogen werden.

Die telefonischen Kontaktinterviews wurden in der Zeit vom 28.1. – 21.2.2002 vom Infratel-Interviewerstab in München durchgeführt. Von den 3.672 Haushalten, die in diesem Schritt angesprochen wurden, erklärten 2.495 ihr grundsätzliches Interesse an der Befragung – zumindest insoweit, als sie ihre Anschrift für die Zusendung von Informationsmaterial angaben. 1.177 Haushalte konnten nicht erreicht werden oder äußerten kein Interesse an einer weiteren Beteiligung. Eine Übersicht über die Ausfallgründe im Einzelnen gibt die beigefügte *Rücklaufmeldung*.

Mit der erreichten Teilnahmequote von 72% war dieser Arbeitsschritt erfreulich erfolgreich. Es lagen damit 2.495 mögliche Teilnehmeradressen für eine mündlich-persönliche Befragung vor. Das waren rd. 300 mehr als die geplante Mindestzahl.

Rücklaufmeldung

Projekt: 08.20.27479 SOEP-High Income
Studienbereich: SoFo
Feldzeit: 28.01.02 - 21.02.02
Zielgruppe: Haushalte mit hohem Einkommen
Befragungsgebiet: BRD
Auswahlverfahren: Infrascopes
Telefon-Studio: Infratel München

Bruttoansatz	3.672	100,0%
Technisches Problem		-
Aufgelegt vor Kontakt		0,0%
Verbleibende Adressen	3.670	100,0%
Firmen-Anrufbeantworter		1,1%
Falsche Telefonnummer		2,7%
Fax/Modem		1,0%
Kein Privathaushalt		0,3%
Doppeladresse		0,2%
ZP nicht (mehr) unter der Adresse		-
ZP verstorben		-
KP: Kein Verständigung möglich		0,4%
ZP: Kein Verständigung möglich		0,1%
ZP geistig/körperl. n.i.d. Lage		0,0%
Max. Kontaktzahl erreicht (>= 12 Kont.)		-
Neutrale Ausfälle gesamt	213	5,8%
Bereinigtes Brutto	3.457	100,0%
Teilnehmer nimmt nicht ab		3,4%
Anrufbeantworter		0,9%
Abbruch Interview		0,2%
ZP vorübergehend krank		-
KP beschäftigt/nicht erreichbar		-
ZP beschäftigt/nicht erreichbar		-
KP: keine Auskunft; Projekt/Thema		3,9%
KP: keine Auskunft; sonst. Grund		17,2%
ZP: keine Auskunft; Projekt/Thema		0,1%
ZP: keine Auskunft; sonst. Grund		0,5%
Kein Termin möglich		1,4%
Ausfälle gesamt	962	27,8%
Durchgeführte Interviews	2.495	72,2%
Nettoansatz Soll	2.200	

2.4 Teilnahmebereitschaft in der mündlich-persönlichen Befragung

Die mündlich-persönliche Befragung ist die eigentliche Datenerhebung im Rahmen des SOEP 2002 (vgl. oben Kapitel 1.4).

Die Brutto-Stichprobe für diese Befragung – bezeichnet als "Brutto 3" – umfasst die 2.495 Adressen aus den telefonischen Kontaktinterviews (Kapitel 2.3). Diese Haushalte erhielten schriftliches Informationsmaterial zugesandt, und zwar die für Erstbefragungen verwendete Fassung der Broschüre "Leben in Deutschland" mit einem erläuternden Anschreiben.

Die Information für die Befragten ist dieselbe wie bei anderen SOEP-Stichproben. Auf die Tatsache, dass die jetzt zu befragenden Haushalte eine Sondergruppe darstellen, nämlich Haushalte im oberen Einkommensbereich, wurde in keiner Weise Bezug genommen. Auch die Interviewer erhielten diese Information nicht – weder für die telefonischen Kontaktinterviews (siehe Kapitel 2.3) noch für die mündlich-persönliche Befragung. Die Information zum Einkommen aus der ursprünglichen InfraScope-Befragung wurde also – in anonymisierter Form – zur Stichprobenbildung verwendet. Sie wurde aber personenbezogen an niemanden – nicht einmal die in der Folgebefragung eingesetzten Interviewer – weitergegeben.

Das Anschreiben an die Stichprobenhaushalte enthielt die Vorankündigung, dass ein Infratest-Interviewer sich bei ihnen melden würde, um die Teilnahme an der Befragung abzuklären und ggf. einen Termin dafür zu vereinbaren. Die Freiwilligkeit der Teilnahme ist unberührt davon, dass der Haushalt bzw. eine Kontaktperson aus dem Haushalt ein grundsätzliches Interesse an dem Vorhaben signalisiert hatte.

Tatsächlich waren dann etwa 3 von 4 Haushalten zur Teilnahme bereit. Die Stichprobenausschöpfung in der mündlich-persönlichen Befragung beträgt 74,3% bezogen auf das bereinigte Brutto. Tabelle 1 zeigt die Struktur der Ausfälle im Einzelnen.

Ein Teil der Brutto-Stichprobe (ca. 12% der Haushalte) wurde von den Interviewern im Rahmen der vorgegebenen Feldzeit überhaupt nicht bearbeitet. Die Gründe hierfür lagen bei den einzelnen Interviewern (z.B. keine Zeit oder krank oder mit anderen Studien beschäftigt). Nachdem sich in der Feldarbeit abzeichnete, dass die angestrebte Zahl von Interviews auch ohne die noch unbearbeiteten Haushalte erreicht würde, wurde die Bearbeitung für diesen Teil der Stichprobe gestoppt. In Tabellè 1 sind diese Haushalte den qualitätsneutralen Ausfällen zugerechnet.

Tabelle 1
Stichprobenausschöpfung in der mündlich-persönlichen Befragung

	abs.	%	%
Fallzahl "Brutto 3"	2.495	100	
Neutrale Ausfälle:			
Verstorben	2	0,1	
Adresse falsch	3	0,1	
Feldbearbeitung gestoppt	<u>295</u>	<u>11,8</u>	
Bereinigtes Brutto	2.195	88,0	100
davon			
nicht erreicht	70	2,8	3,2
nicht interviewbar ¹⁾	6	0,2	0,3
nicht bereit	<u>488</u>	<u>19,6</u>	<u>22,2</u>
Summe Ausfälle	1.564	22,6	25,7
Interviews	1.631	65,4	74,3

1) z.B. wegen Krankheit, Alter usw.

2.5 Überprüfung der Einkommen

Nach Vorliegen der Interviews wurde für jeden einzelnen Haushalt überprüft, ob er tatsächlich zur definierten Untersuchungsgruppe gehört, d.h. nach Auskunft in der mündlich-persönlichen Befragung über ein Haushaltsnettoeinkommen von mindestens 7.500 DM verfügt. Nur wenn dies zutrifft, wird der Haushalt in die endgültige Netto-Stichprobe übernommen.

Bereits in der *Feasibility Study* hatte sich abgezeichnet, dass mit einem erheblichen Anteil von Haushalten in der Stichprobe gerechnet werden muss, bei denen die Einkommenseinstufung aus der ersten (telefonischen) Befragung im zweiten (mündlich-persönlichen) Interview nicht bestätigt wird. Die Gründe für eine solche Inkonsistenz können vielfältig sein:

- Zwischen beiden Befragungen liegt ein Zeitabstand, der im kürzesten Fall 3 Monate und im längsten Fall 18 Monate betragen kann. Es ist möglich, dass sich das Haushaltsnettoeinkommen in dieser Zeit tatsächlich verändert hat.

- Die Auskunftsperson für die Einkommensangabe ist im zweiten Interview oft nicht dieselbe wie im ersten Interview. Beispielsweise kann im telefonischen Interview ein Jugendlicher befragt worden sein, während der Haushalts-Fragebogen im SOEP von der Person, die sich mit den finanziellen Dingen der Familie am besten auskennt, beantwortet werden soll.
- Die Art der Einkommensabfrage ist in beiden Interviews unterschiedlich. Während im telefonischen Interview grobe Einkommensklassen vorgegeben werden, wird im Haushalts-Fragebogen des SOEP ein offener Einkommensbetrag erfragt; ergänzend stehen Angaben zu einzelnen Einkunftsarten zur Verfügung.

Beim Abgleich der zwei Einkommensangaben wird die Angabe aus dem SOEP-Interview als die zutreffende ("wahre") – mit einem höheren Grad an Aktualität und Zuverlässigkeit – gewertet.

Das Ergebnis des vorgenommenen Abgleichs zeigt erhebliche Diskrepanzen zwischen den Einkommensangaben im ersten und denen im zweiten Interview:

- Wenn man sich allein auf die offene Frage nach dem Haushaltsnettoeinkommen ("Income-Screener") in der SOEP-Befragung stützt, erreichen 32% der Haushalte nicht die Schwelle von 7.500 DM (jetzt definiert als 3.850 €, entspricht 7.530 DM). Hierunter sind Fälle mit fehlender Angabe enthalten.
- 5% unter den 32% erreichen die Schwelle allerdings, wenn man die individuellen Einkommen der Haushaltsmitglieder aufaddiert, die im jeweiligen PERSONEN-Fragebogen angegeben sind.
- Weitere 2% erreichen die Schwelle, wenn man haushaltsbezogene Einkunftsarten einbezieht, von denen nicht klar ist, ob Befragungspersonen sie bei einer pauschalen Angabe der Haushaltsnettoeinkommens mitrechnen würden, nämlich Sozialleistungen, Zinseinkünfte und Mieteinnahmen.
- Es verbleiben 25% der befragten Haushalte, die auch mit dem erweiterten Einkommensbegriff die definierte Einkommensschwelle *nicht* erreichen. Diese Haushalte werden als "ungültige Fälle" behandelt. Die Fallzahl der Nettostichprobe verringert sich dadurch von 1.631 auf 1.224 Haushalte.
- Innerhalb der verbleibenden Haushalte ist die Einkommensschichtung – verglichen mit den oben dargestellten Verteilungen aus InfraScope, SOEP 2000 oder MZ 2000 – nach oben verschoben. Dies erklärt sich aus der disproportionalen Schichtung der Stichprobe, also dem höheren Auswahlatz für die höheren Einkommen, wie oben beschrieben (vgl. Kapitel 2.3, Schritt 2). Tabelle 2 stellt die Einkommensschichtung in der Stichprobe im Vergleich zum Mikrozensus 2000 dar.

Tabelle 2
Einkommensschichtung im oberen Einkommensbereich

Einkommensklassen	Nettostichprobe 2002		Mikrozensus
	abs.	%	2000 %
7.500 b. u. 10.000 DM 3.850 b. u. 5.000 €	412	34	63
10.000 b. u. 12.000 DM 5.000 b. u. 6.000 €	321	26	17
12.000 DM und mehr 6.000 € und mehr	<u>491</u>	<u>40</u>	20
Summe	1.224	100	100

Für Analysen in der Gesamtgruppe der einbezogenen Haushalte muss die disproportionale Schichtung der Einkommensklassen rechnerisch korrigiert werden. Dabei sind im Detail einige Schwierigkeiten zu beachten, die sich in Tabelle 2 etwa auf folgende Punkte beziehen: Der Zeitabstand zwischen den Verteilungen beträgt zwei Jahre; die Einkommensklassen sind nicht ganz identisch definiert, da sie einmal in DM und einmal in Euro abgegrenzt sind; im Mikrozensus wurden Einkommensklassen vorgegeben, die den Effekt haben, dass Befragte sich an "runden" Beträgen orientieren, die nach einer Umrechnung in Euro gar nicht mehr "rund" sind.

2.6 Struktur der Nettostichprobe nach soziodemographischen Merkmalen

Teil des Datenbestands für alle Befragungshaushalte sind die Merkmale der Standarddemographie aus der ursprünglichen InfraScope-Befragung. Soweit es sich um personenbezogene Merkmale wie Alter, Geschlecht, Beruf etc. handelt, beziehen sie sich auf das ursprünglich befragte Haushaltsmitglied. Andere Merkmale sind haushaltsbezogen (z.B. Haushaltsgröße) oder regionale Zuordnungen (z.B. Bundesland, GGK).

Die realisierte Netto-Stichprobe kann anhand dieser Strukturmerkmale mit der Ausgangsstichprobe verglichen werden. Selektivitätseffekte im Prozess der Stichprobenbildung, wie er in den vorangehenden Kapiteln 2.1 bis 2.5 beschrieben wurde, müssten sich in Verschiebungen der Stichprobenstruktur niederschlagen.

Der folgende Tabellensatz nimmt einen solchen Vergleich der Stichprobenstrukturen vor. Die jeweils erste Spalte zeigt die Struktur der Ausgangsstichprobe, die jeweils letzte Spalte die Struktur der realisierten Nettostichprobe. Die Spalten dazwischen zeigen die Struktur der "Aus-

fallgruppen", und zwar in der Reihenfolge der Arbeitsschritte im Prozess der Stichprobenbildung.⁹

Da die meisten Merkmale personenbezogen sind, erfolgt die Auszählung personengewichtet, unter Verwendung des Gewichtungsfaktors aus der ursprünglichen InfraScope-Stichprobe. Die Fallzahlen (ausgewiesen in der ersten Zeile) sind personengewichtet etwas höher als oben in den Kapiteln 2.1 – 2.5 berichtet.¹⁰

Eine Durchsicht der Tabellen zeigt, dass die soziodemographische Struktur der Stichprobe über die verschiedenen Schritte der Stichprobenbildung hinweg kaum verändert ist.¹¹

Einige Beispiele:

Anteil der ...	in der Ausgangs- stichprobe	in der Netto- stichprobe
Frauen	37%	36%
Erwerbstätigen	72%	72%
Personen in größeren Haushalten (4 u.m. Personen)	42%	42%
neuen Bundesländer inkl. Berlin	12%	12%
Ballungsregionen (BIK) ab 500.000 Einwohner	48%	50%

Leichte Verschiebungen im Sinne eines erhöhten Anteils gibt es bei den folgenden Strukturmerkmalen:

Anteil der ...	in der Ausgangs- stichprobe	in der Netto- stichprobe
Altersgruppen > 40 Jahre	64%	75%
Personen mit höherem Schulabschluss	62%	70%
Beamten	17%	24%

Umgekehrt sind etwas unterrepräsentiert:

Personen in Kleinstädten und ländlichen Gemeinden außerhalb der Ballungsregionen (31% vs. 23%).

⁹ Die Abkürzung "F2F" steht für "face-to-face", also die mündlich-persönliche Befragung.

¹⁰ Der Grund hierfür ist, dass die Haushalte im oberen Einkommensbereich eher größere Haushalte sind. Je größer der Haushalt, um so geringer ist die Auswahlwahrscheinlichkeit für das einzelne Haushaltsmitglied, befragt zu werden. Dies wird in der Designgewichtung, die Bestandteil der Personengewichtung ist, ausgeglichen. Wenn – wie im SOEP – alle Personen eines Haushalts befragt werden, entfällt diese rechnerische Umwandlung.

¹¹ Zu beachten ist, dass die Ausgangsstichprobe nicht etwa die bekannten Strukturen der Bevölkerung insgesamt wiedergibt, sondern die der Haushalte im oberen Einkommensbereich.

Sofern überhaupt Strukturverschiebungen vorliegen, muss jeweils untersucht werden, bei welchem Arbeitsschritt der Stichprobenbildung diese Zustände kamen – ob sie also zurückzuführen sind auf ...

- a) eine geringere Teilnahmebereitschaft bestimmter Gruppen (dies sind die Schritte "keine Zustimmung Tel.Nr." / "Kontaktinterview negativ" / "F2F Nicht-Teilnehmer")
- b) den ggf. disproportionalen Auswahlatz oder Effekte der regionalen Klumpung (Arbeitsschritt "nicht ausgewählt")
- c) oder ein negatives Ergebnis der Einkommensprüfung, d.h. ein Haushaltsnettoeinkommen unter 7.500 DM ("F2F Teilnehmer ungültig").

Der letztgenannte Grund c) spielt beispielsweise die maßgebliche Rolle bei den leichten Verschiebungen der Alters- und der Berufsstruktur:

Personen ab 40 Jahren und Beamte haben in der ursprünglichen Telefonbefragung zuverlässigere Einkommensangaben gemacht als jüngere Personen oder Nicht-Beamte. Es gibt bei ihnen daher weniger Ausfälle mit dem Grund "ungültiger Haushalt". Die Stichprobenselektivität ist hier also keine Verzerrung, sondern eine Bereinigung. Ein Ausgleich qua Gewichtung erscheint dafür kaum angebracht.

Wesentliches Ergebnis der Strukturprüfung scheint uns die Robustheit der Stichprobenstrukturen zu sein. Angesichts dieser Sachlage empfiehlt sich als das beste Vorgehen, auf Eingriffe in die Stichprobenstruktur qua Gewichtung zu verzichten.

Eine Ausnahme davon bilden lediglich die Verteilungen, die aufgrund der disproportionalen Schichtung der Stichprobe von den Sollverteilungen abweichen (Einkommen, Ost/West). Diese designbedingten Effekte sind rechnerisch durch Korrekturfaktoren auszugleichen.

SOEP 2002 - Brutto Stichprobe G - High Income

Prozentwerte senkr. PE-gew.	Gesamt	Bearbeitungsergebnis					
		Keine Zustimmung Tel.Nr.	Nicht ausgewählt	Kontakt- interview negativ	F2F Nicht- teilnehmer	F2F Teilnehmer ungültig	F2F Teilnehmer gültig
Basis abs.	7406	760	2370	1342	1016	483	1435
Geschlecht							
Männlich	63	65	60	64	62	65	64
Weiblich	37	35	40	36	38	35	36
Summe	100	100	100	100	100	100	100
Altersgruppen							
unter 20 Jahre	6	6	5	6	8	5	4
20 bis 29 Jahre	13	15	15	14	13	18	8
30 bis 39 Jahre	16	18	16	17	17	20	13
40 bis 49 Jahre	24	23	26	23	24	18	27
50 bis 59 Jahre	25	24	23	23	26	24	31
60 bis 69 Jahre	10	8	9	11	9	9	13
70 Jahre und älter	5	6	4	6	4	5	5
Keine Angabe	0	1	0	0	0	1	
Summe	100	100	100	100	100	100	100
Erwerbstätigkeit							
berufstätig	72	71	72	72	75	71	72
Lehrling	2	1	2	2	1	3	1
arbeitslos	1	1	1	0	0	1	1
Rentner oder Pensionär	10	10	9	10	8	9	12
Schüler	4	6	4	4	6	4	3
Fachschüler	0		0	0	0		
Student	4	4	4	3	3	4	3
Hausfrau	6	6	6	5	5	5	6
Wehr-/Zivildienstleistender	1	1	1	1	0	1	1
Sonstiges	1	1	1	2	1	2	1
Keine Angabe	0		0				
Summe	100	100	100	100	100	100	100

SOEP 2002 - Brutto Stichprobe G - High Income

Prozentwerte senkr. PE-gew.	Gesamt	Bearbeitungsergebnis					
		Keine Zustimmung Tel.Nr.	Nicht ausgewählt	Kontakt- interview negativ	F2F Nicht- teilnehmer	F2F Teilnehmer ungültig	F2F Teilnehmer gültig
Basis abs.	7406	760	2370	1342	1016	483	1435
Stellung im Beruf							
Arbeiter	3	4	3	5	2	4	1
Angestellter	53	53	55	50	52	57	50
Beamter (auch: Berufssoldat/Richter)	17	15	19	12	15	10	24
Landwirt	0		0	0			0
Freiberufler	7	8	5	10	7	9	8
Sonst. Selbstständiger/Unternehmer/PGH- Mitgl.	19	20	17	23	24	20	16
Mithelfender Fam.-ang.	0	0	0	0			0
Keine Angabe	0	0	0		0	0	0
Summe	100	100	100	100	100	100	100
Angestellte							
ausf. Ang	8	7	9	7	8	8	6
qualifiz. Ang.	29	26	32	28	27	34	24
hochqual. Ang.	23	24	24	22	21	23	26
leitende Ang.	39	41	35	42	43	35	43
Keine Angabe	1	1	1	1	0	0	1
Summe	100	100	100	100	100	100	100
Arbeiter							
un-/angelernte Arb.	25	25	32	18	7	23	41
Facharbeiter	62	75	61	55	75	70	53
Vorarb., Kol.-führer/Meister, Polier	10		7	20	7	6	6
Keine Angabe	3			7	11		
Summe	100	100	100	100	100	100	100
Beamte							
Einfacher/mittl. Dienst	11	19	12	10	11	12	7
Gehobener Dienst	37	30	42	30	32	49	37
Höherer Dienst	50	43	43	59	55	39	56
Keine Angabe	2	8	2	1	1		0
Summe	100	100	100	100	100	100	100

SOEP 2002 - Brutto Stichprobe G - High Income

Prozentwerte senkr. PE-gew.	Gesamt	Bearbeitungsergebnis					
		Keine Zustimmung Tel.Nr.	Nicht ausgewählt	Kontakt- interview negativ	F2F Nicht- teilnehmer	F2F Teilnehmer ungültig	F2F Teilnehmer gültig
Basis abs.	7406	760	2370	1342	1016	483	1435
Bundesland							
Berlin West	3	3	4	2	2	1	2
Schleswig-Holstein	4	5	4	4	4	5	4
Hamburg	2	3	3	2	1	1	2
Niedersachsen	9	8	9	10	10	10	9
Bremen	1	1	1	1	1	0	1
Nordrhein-Westfalen	24	24	24	24	27	26	23
Hessen	10	9	10	9	9	10	11
Rheinland-Pfalz/Saarland	5	5	7	4	5	4	4
Baden-Württemberg	14	18	13	14	14	12	13
Bayern	18	16	20	15	18	17	18
Saarland	1	1	1	2	1	1	1
Berlin Ost	1	1		2	1	1	1
Mecklenburg-Vorpommern	1	1		2	3	2	1
Brandenburg	2	1	3	1	1	1	1
Sachsen-Anhalt	2	2	1	2	2	2	2
Thüringen	1	2	0	2	1	2	2
Sachsen	2	1	0	4	3	4	3
Summe	100	100	100	100	100	100	100
Gemeindetyp							
500000+ EW (Kern)	34	37	26	40	40	32	36
500000+ EW (Rand)	14	12	15	11	13	14	14
100000-499999 EW (Kern)	10	11	2	13	13	18	16
100000-499999 EW (Rand)	7	6	7	7	8	5	8
50000- 99999 EW (Kern)	3	3	3	5	4	3	3
50000- 99999 EW (Rand)	1	0	1	0	0	1	1
20000- 49999 EW	8	8	12	7	6	6	7
5000- 19999 EW	13	14	19	9	9	14	9
2000- 4999 EW	5	5	9	4	4	3	3
unter 2000 EW	5	4	6	3	4	3	4
Summe	100	100	100	100	100	100	100

Prozentwerte senkr. PE-gew.	Gesamt	Bearbeitungsergebnis					
		Keine Zustimmung Tel.Nr.	Nicht ausgewählt	Kontakt- interview negativ	F2F Nicht- teilnehmer	F2F Teilnehmer ungültig	F2F Teilnehmer gültig
Basis abs.	7406	760	2370	1342	1016	483	1435
Schulabschluss							
Haupt-/Volksschule	10	10	12	10	10	11	8
Weiterf. Schule (ohne Abschluss)	1	1	1	1	1	1	1
Mittlere Reife/Abschl. polytechn. Oberschule	27	27	30	25	26	26	22
Abitur, FHS-Reife*	62	61	56	63	62	62	70
Keine Angabe	1	1	1	1	1	0	1
Summe	100	100	100	100	100	100	100
Abgeschlossenes Studium (<i>Basis: *</i>)							
Ja	68	67	64	66	67	64	76
Nein	32	33	36	34	33	36	24
Summe	100	100	100	100	100	100	100
Haushaltsgrösse							
1	3	5	3	4	3	4	3
2	33	36	31	36	33	32	33
3	22	22	22	22	22	23	22
4	26	23	27	25	27	27	28
5	11	9	12	9	10	10	10
6	3	4	4	2	3	3	3
7	1	1	1	1	1	0	1
8	0	0	0	1	1	0	0
9	0	0	0	1	1	1	0
Summe	100	100	100	100	100	100	100
HHeinkom über 10000 DM							
Ja	37	34	8	54	52	47	52
Nein	56	56	83	39	41	49	44
K.A.	7	10	9	7	7	4	4
Summe (<i>Basis: > 7.500</i>)	100	100	100	100	100	100	100
HHeinkom über 12000 DM							
Ja	46	51	53	45	46	45	42
Nein	47	40	44	47	46	49	50
K.A.	8	9	2	8	8	6	7
Summe (<i>Basis: > 10.000</i>)	100	100	100	100	100	100	100

3. Datenprüfung und übergebener Datenbestand

Die Fragebogen der Sonderstichprobe "Haushalte im oberen Einkommensbereich" wurden nach SOEP-Standard erfasst, geprüft und aufbereitet. Für eine genauere Beschreibung der Verfahren wird auf die von Infratest erstellten Methodenberichte zum SOEP verwiesen. Der an das DIW übergebene Datenbestand ist in seinem Aufbau identisch mit dem Datenbestand für die SOEP-Stichproben A – F des Jahres 2002.

Der vollständige Datenbestand der Sonderstichprobe wurde dem DIW am 20.8.2002 auf einer CD-Rom mit folgenden Dateien übergeben:

HBRU2002.sav	Bruttoband Haushalte (SPSS)	6330	Fälle
PBRU2002.sav	Bruttoband Personen (SPSS)	3538	Fälle
H02G.dat	Nettodaten Haushalte (ASCII)	1224	Fälle
H02G.sav	Nettodaten Haushalte (SPSS)	1224	Fälle
H02G_U.dat	Nettodaten Haushalte ungeprüft (ASCII)	1224	Fälle
P02G.dat	Nettodaten Personen (ASCII)	2671	Fälle
P02G.sav	Nettodaten Personen (SPSS)	2671	Fälle
P02G_U.dat	Nettodaten Personen ungeprüft (ASCII)	2671	Fälle
5.Kind-G.xls	Mehr als 4 Kinder		

Code1p02-E.PMD Personenfragebogen/1.Teil (Fassung E) mit Variablen und Labels

Code2p02-E.PMD Personenfragebogen/2.Teil (Fassung E) mit Variablen und Labels

Codeh02-E.PMD Haushaltsfragebogen (Fassung E) mit Variablen und Labels

Codesnet.xls Codelisten für Haushalt-, und Personendaten

STR02_UR.xls Variablenlisten (mit Bspaltung) für Haushalt-, und Personendaten
(mit Kennzeichnung der Veränderungen zur Welle01)

Zusatzcodes.xls Nationen, Sorgen, Parteien und Bundesländer

Summary

The objective of the survey was to generate comprehensive data on living conditions, income and assets of high-income households in Germany. High income is defined to include the upper 6-7% of the income distribution, with oversampling of the upper 2-3%.

The sample was derived from a larger random population sample. The survey was implemented within the survey programme of the German Socio-economic Panel (GSOEP) in 2002.

The achieved sample comprises 1,224 households, of which 812 have a monthly net-income of 5,000 € and above.

Methodological details are shown in the following set of charts which are based on a presentation given to the GSOEP board in May 2002.

SOEP
Survey of
Oconomic
Panel

High-Income Sample

Feasibility Study (1), 2000/2001:
Sample of "wealthy" people ("Reiche"),
based on NFO Infratest Access Panel

→ Failed

Feasibility Study (2), 2001/2002:
Sample of the high-income segment
of the population, based on
telephone surveys

→ Feasible

Main Survey, 2002:
Implemented as subsample of
GSOEP 2002,
target size 1,000 households.

→ Achieved

19. November 2002 Chart 1 Infratest

SOEP
Survey of
Oconomic
Panel

High-Income Sample

Sampling:

(1) Random telephone sample (RLD) of current
population survey. All interviews in 2001: n = 99,175

(2) Of which: monthly net household
income > 7,500 DM n = 6,330 (* gross 1)

(3) Of which: agreed telephone no. to be used for
further interviews n = 5,663 (89%)

(4) Of which: live in a region where a SOEP
interviewer is around = regional clustering
with differential selection probability:
higher for income > 10,000 DM n = 3,670 (* gross 2)

➔ **recruiting for SOEP**

19. November 2002 Chart 2 Infratest

SOEP Survey of
Oconomic
Panel
Study

High-Income Sample

Recruiting for SOEP:

- (4) Basis: gross sample 2 n = 3,670
- (5) Brief telephone interview to check general co-operativeness → get name and address n = 2,490
- (6) Send information about SOEP to household (brochure "Living in Germany")
- (7) Send SOEP interviewer to household and try to get interviews according to standard SOEP procedure target: n = 1,300
achieved: n = 1,631
- (8) Check household income to make sure that income is > 7,500 DM target: n = 1,000 net sample
achieved: n = 1,224 net sample

19. November 2002 Chart 3 Infratest

SOEP Survey of
Oeconomic
Panel
Study

High-Income Sample

Quality:

- (1) Comparison to other data sources

Households with net income	> 7,500 DM	> 10,000 DM
Infratest telephone surveys 2001	7.7%	2.7%
SOEP 2000	6.3%	2.3%
Mikrozensus 2000	7.0%	2.6%
- (2) Analysis of sample structure

In the achieved sample no major selectivity in terms of sample structure can be discovered compared to the high-income segment of the larger population survey from which the sample is derived.

19. November 2002 Chart 4 Infratest